Enhancing information security and privacy by combining biometrics with cryptography

Sanjay Kanade

▶ **To cite this version:**

Sanjay Kanade. Enhancing information security and privacy by combining biometrics with cryptography. Other. Institut National des Télécommunications, 2010. English. NNT : 2010TELE0022 . tel-01057728

HAL Id: tel-01057728

https://theses.hal.science/tel-01057728v1

Submitted on 25 Aug 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse de doctorat de Télécom & Management SudParis dans le cadre de l'école doctorale S&I en co-accréditation avec l'Université d'Evry-Val d'Essonne

Spécialité :

Informatique

Par

M. Sanjay Ganesh Kanade

Thèse présentée pour l'obtention du diplôme de Docteur de Télécom & Management SudParis

Enhancing Information Security and Privacy by Combining Biometrics with Cryptography

Soutenue le 20 Octobre 2010 devant le jury composé de :

Prof. Gérard Cohen	TELECOM ParisTech	Rapporteur
Prof. Patrizio Campisi	University of Roma TRE	Rapporteur
Prof. Marinette Revenu	GREYC - ENSICAEN	Examinateur
Prof. Ramesh Pyndiah	TELECOM Bretagne	Examinateur
Dr. Nalini Ratha	IBM T.J. Watson Research Center	Examinateur
Dr. Tom Kevenaar	Priv-ID B. V.	Examinateur
Dr. Dijana Petrovska-Delacrétaz	TELECOM & Management SudParis	Encadrant de thèse
Prof. Bernadette Dorizzi	TELECOM & Management SudParis	Directeur de thèse

Thèse n° - 2010TELE0022

Abstract

Enhancing Information Security and Privacy by Combining Biometrics with Cryptography

by

Sanjay Ganesh Kanade

Doctor of Philosophy in Informatique

TELECOM & Management SudParis and l'Université d'Évry-Val d'Essonne

Dr. Dijana Petrovska-Delacrétaz and Prof. Bernadette Dorizzi

Securing information during its storage and transmission is an important and widely addressed issue. Generally, cryptographic techniques are used for information security. In cryptography, the general idea is to transform the information during a phase called encryption, before being stored or transmitted, based on a secret key. This secret key is required in order to retrieve the information from the transformed data during decryption. These secret keys are generally too long for a user to remember, and therefore, need to be stored somewhere. The drawback of cryptography is that these keys are not strongly linked to the user identity. In order to strengthen the link between the user identity and his cryptographic keys, biometrics is combined with cryptography.

Unfortunately, biometric systems possess problems of their own such as nonrevocability, non-template diversity, and possibility of privacy compromise which should be taken into consideration. Combining biometrics with cryptography in a secure way can eliminate these drawbacks. Thus, biometrics and cryptography can complement each other. The systems, in which, techniques from biometrics and cryptography are combined are called as crypto-biometric systems. The combined system can inherit the positive aspects of the two while eliminating their limitations.

In this thesis, first we present a through and systematic review of the crypto-biometric systems. The primary criterion for the classification is the main goal of the system. There can be two principal goals: (i) protecting biometric data, and (ii) obtaining cryptographic keys from biometrics. The systems in these two categories are further divided according to their working methodology. We propose crypto-biometric systems

from each of these two categories.

The first system we propose is a shuffling based cancelable biometric system. We propose a simple shuffling scheme which randomizes the biometric data with the help of a shuffling key. This shuffling scheme: (a) adds revocability to the biometric systems, (b) improves the verification performance (nearly 80% decrease in equal error rate) because it increases the impostor Hamming distance without changing the genuine Hamming distance, (c) adds template diversity, and (d) makes cross-matching impossible and thus protects privacy.

The Second proposal is to use Error Correcting Codes (ECC) to reduce variability in biometric data. We propose a novel method which can correct fractional amount of errors in biometric data. The error correction scheme reduces the variability in the test (query) biometric data with respect to the reference biometric data. When combined with the shuffling scheme mentioned above, more than 90% reduction in EER is observed as compared to the baseline biometric system.

The third contribution is a scheme for cryptographic key regeneration using biometrics. We propose a hybrid scheme to obtain high entropy cryptographic keys using biometrics. The shuffling scheme described above is first applied on the biometric data to make it revocable. This data is then used in a fuzzy commitment based key regeneration scheme. The generic scheme is then adapted to two biometric modalities: iris and face. The amount of errors (variability) in the biometric data for these two modalities is different. Therefore, different sets of error correcting codes are used for these modalities in order to cope with the variability in the biometric data. The entropy of keys obtained using the iris and face based key regeneration systems are 83 and 112 bits, respectively.

The fourth contribution is to use multi-biometrics to obtain high entropy cryptographic keys. In order to increase the security and increase the length and entropy of the keys, we propose a novel scheme to combine information from multiple biometric sources. A novel method called *FeaLingECc* (*Fea*ture *L*evel Fus*i*o*n* through Wei*g*hted *E*rror *C*orre*c*tion) is proposed for information fusion. It allows applying different weights to different biometric cues. Two systems are developed: a multi-unit type system that combines information from the left and right irises of a subject, and a multi-modal system which combines iris with face. A significant increase in the key entropy is observed

along with an improvement in the verification performance. The estimated entropy of the keys obtained with the two-iris system is 189 bits while that from the iris-face system is 183 bits.

Finally, we address the issue of sharing the crypto-bio keys. We propose a novel protocol to share the crypto-bio keys generated using our key regeneration scheme. The same crypto-bio key is shared in every run of the protocol. In order to have better security, we propose another novel protocol to generate and share biometrics based session keys. This protocol allows mutual authentication between the two parties - client and the server - without the need of trusted third party certificates. This protocol has a potential to replace existing key sharing protocols. Moreover, it can easily be integrated into existing key sharing protocols in order to have an additional layer of security.

Titre : La crypto-biométrie, une solution pour une meilleure sécurité tout en protégeant notre vie privée

(Enhancing Information Security and Privacy by Combining Biometrics with Cryptography)

Par
M. Sanjay Ganesh Kanade

Résumé

La sécurité de l'information au cours de son stockage et de sa transmission constitue un enjeu majeur de notre société. En règle générale, les techniques cryptographiques sont utilisées pour sécuriser l'information. En cryptographie, l'idée générale est de transformer (crypter) l'information avec une clé secrète. Cette clé secrète est nécessaire afin de récupérer les informations à partir des données transformées au cours du décryptage. Ces clés secrètes sont longues, donc difficilement mémorisables, et, par conséquent, doivent être stockés quelque part. L'inconvénient majeur de ce processus est le faible lien entre cette clé et l'utilisateur. Afin de renforcer ce lien la biométrie est un candidat de choix.

Pour assurer une utilisation pérenne des cette technologie et qui en plus préserve notre vie privée, il faudrait lui ajouter les propriétés de révocabilité et de diversité. La diversité permet que l'on puisse utiliser par exemple notre iris, dans plusieurs applications, sans que l'on sache que c'est la même personne. Et si toutefois notre iris est compromis, que l'on puisse révoquer le système (comme quand on change de mot de passe quand celui-ci a été compromis), et le réutiliser de nouveaux.

Un axe émergent de recherche se profile depuis 1998, qui en combinant biométrie et cryptographie apporte une solution pour lier de manière plus sécurisé l'identité de l'utilisateur et les clés de chiffrement, tout en apportant les caractéristiques qui manquaient à la biométrie, principalement la révocabilité et la diversité. Les avantages de cette combinaison, que nous appelons crypto-biométrie sont multiples. On arrive du coté biométrie, a apporter une solution algorithmique relativement aux problèmes de la protection des données biométriques et de la vie privée et a résoudre le problème de non-révocabilité des systèmes biométriques.

Dans cette thèse, tout d'abord un état de l'art complet de cette nouvelle technologie –la crypto-biométrie- est présenté. Un examen systématique des travaux existants est donné, ainsi qu'une proposition de classement des divers travaux publiés en deux catégories principales.

Le critère pour la classification est l'objectif principal du système : (i) la protection des données biométriques, ou (ii) l'obtention des clés cryptographiques à partir de la biométrie.

Nous proposons des nouveaux systèmes crypto-biométriques dans chacun de ces deux catégories. Notre première proposition permet de conférer à la biométrie de la révocabilité tout en apportant une amélioration des performances de vérification (diminution de près de 80% du taux d'erreur). Ces nouvelles fonctionnalités sont obtenues par un procédé simple (de permutations), qui en plus protège nos données biométriques et protège ainsi notre vie privée.

La deuxième proposition consiste à exploiter des codes correcteurs d'erreurs pour réduire la variabilité des données biométriques. L'originalité de cette approche est de pouvoir traiter la variabilité des données biométriques, d'en corriger une partie et de générer des nouvelles données « moins bruitées ». Le couplage de ces deux propositions permet d'améliorer encore d'avantage les performances biométriques, tout en assurant des systèmes biométriques révocables.

Notre troisième contribution est un système de régénération de clés cryptographiques en utilisant la biométrie, désignées comme clés « crypto-bio ». Nous proposons un système hybride pour obtenir des clés cryptographiques qui sont intimement liées à notre biométrie. La validité de notre proposition est testée tout d'abord sur la modalité iris, puis visage tout en utilisant des bases de données publiques. Différents ensembles de codes correcteurs d'erreur sont choisis pour ces modalités afin de faire face aux différentes sources de variabilité des données biométriques. L'entropie de clés obtenus est supérieure à celle rapportées dans la littérature : 83 et 112 bits pour les systèmes iris et visage, respectivement.

La quatrième contribution utilise la combinaison de plusieurs sources biométriques afin d'accroître la sécurité, donc d'augmenter encore d'avantage l'entropie des clés crypto-biométriques. Nous proposons une solution de type multi-ensemble, qui combine les données de l'iris gauche et droit d'un sujet, et un autre système multi-modal qui combine l'iris avec le visage. Une augmentation significative de l'entropie des clés crypto-bio est observée avec une amélioration des résultats de vérification. L'entropie estimée des clés obtenues avec le système à deux iris est de 189 bits tandis que celle du système iris-visage est de 183 bits.

Enfin, nous abordons la question d'utilisation de ces nouvelles clés crypto-biométriques dans des protocoles d'authentification classiques. Nos apportons une solution originale afin d'avoir une meilleure sécurité. Nous proposons un nouveau protocole pour générer et partager les clés

crypto-biométriques qui sont valables seulement pour la durée de cette session. Ce protocole permet une authentification mutuelle entre les deux parties - le client et le serveur - sans avoir besoin de certificats de confiance des tiers. Ce protocole a un potentiel pour remplacer les protocoles d'échange actuels. En outre, il peut facilement être intégré dans les principaux protocoles de partage afin d'avoir une couche de sécurité supplémentaire.

To my parents

Acknowledgments

It has been an auspicious journey for me to arrive at this point. I am really short of words now, but I take this opportunity to express my sincere thanks to my thesis advisor, Dr. Dijana Petrovska, with whom, I learned to analyze, criticize, and express myself as clearly as possible. I wouldn't be exaggerating to say that she took as much efforts as me for this research work. Particularly, she taught me how to think like a researcher, especially in the beginning of my thesis. Her belief in me throughout this work kept me inspired.

I thank my thesis director, Prof. Bernadette Dorizzi. Its been an honor and pleasure to work with her. Her continuous guidance, constant support, and invaluable advice was instrumental for the success of this work.

I express my sincere thanks to the French "Agence Nationale de la Recherche (ANR)" for providing me the financial support through the project BIOTYFUL (ANR-06-TCOM-018).

I would like to thank Prof. Gerard Cohen and Prof. Patrizio Campisi for accepting my request to be a reporter for this thesis. I also thank Prof. Marinette Revenu, Dr. Nalini Ratha, and Dr. Tom Kevenaar for honoring me by being a part of the jury.

I must thank Prof. Raghunath Holambe whose interactions with Dr. Petrovska and Prof. Dorizzi made my admission to TMSP possible.

I also thank Prof. Gerard Chollet for his interest in my work and the critical comments he gave which helped me improve this work.

I thank my former colleague, Dr. Danielle Camara, who was a post-doc researcher in our group during the first year of my thesis. Her comments and suggestions were quite helpful. I also thank my other former collegues, Dr. Emine Krichen, Mr. Aurelien Mayoue, and Dr. Mohamed Anouar Mellakh, for helping me with the OSIRISv1 system during my initial days. I also thank Mr. Ramachandra Raghavendra and Mr. Dianle Zhou for their help related to the baseline face recognition system. I also thank my other collegues, Nesma, Guillaume, Walid, and Veit-Anh, for their help and support.

I would like to thank Mr. Mohamed Abid and Prof. Hossam Afifi with whom I worked on the project BIOTYFUL. We worked on the iris based ePassport authentication scheme.

I would like to thank Ashish Gupta for his help, support, and encouragement throughout my stay here. I also thank Dr. Sandoche Balakrishnan, Dr. Mahendiran Prathaban, and Dr. Manoj Panda for making my stay in France easier.

I also thank my friends Bhushan Patil, Baghyesh Patil, Dhiraj Magare, and Bapusaheb Chavan for their help during various stages of my career. I also thank my friends Avinash, Yogesh, Prashant, Chandra, and Dipak for the emotional support.

Finally, I owe whatever I am to my parents Mr. Ganesh and Mrs. Mangala Kanade, and also to other members of my family, my grandmother Triveni, brothers Deepak and Ajay, sisters-in-law Savita and Snehal, nephews Dhruv and Avaneesh, and niece Krutika. Their constant support and sacrifices have made this thesis possible.

– Sanjay Ganesh Kanade

Contents

List of Figures

List of Tables

List of Abbreviations

AES	Advanced Encryption Standard
ANR	Agence Nationale de la Recherche
BCH codes	Bose, Ray-Chaudhuri and Hocquenghem codes
BIOTYFUL	BIOmetrics and crypTographY for Fair aUthentication Licensing
CBS	Casia BioSecure Database
ECC	Error Correcting Codes
EER	Equal Error Rate
FAR	False Acceptance Rate
FeaLingECc	*Fea*ture *L*evel Fus*i*on through Wei*g*hted *E*rror *C*orre*c*tion
FRGC	Face Recognition Grand Challenge
FRGC-Exp1*	FRGC Experiment-1 (controlled vs controlled) on our subset
FRGC-Exp4*	FRGC Experiment-4 (controlled vs uncontrolled) on our subset
FRR	False Rejection Rate
GAR	Genuine Acceptance Rate
HTTPS	Hypertext Transfer Protocol Secure
ICE	Iris Challenge Evaluation
ICE-Exp1	ICE Experiment-1 (right eye experiment)
ICE-Exp2	ICE Experiment-2 (left eye experiment)
NIST	National Institute of Standards and Technology
OSIRIS	Open Source Iris Recognition System
RS	Reed-Solomon
SudFROG	SudParis Face Recognition System
TLS	Transport Layer Security

Glossary

The most common terms used in crypto-biometrics are defined below:

1. Biometric template – Set of stored biometric features comparable directly to probe biometric features. It is a special case of a biometric reference, where biometric features are stored for the purpose of a comparison.

2. Identifier/authenticator/credential – Information provided by a user which is required to confirm his identity, e.g., password, token, and biometric characteristics.

3. Verification - One to one comparison of the captured biometric sample with a stored biometric template to verify that the individual is who he claims to be. The result of verification is a Yes/No response.

4. Identification - One to many comparison of the captured biometric sample against a biometric database in an attempt to identify an unknown individual. The result of identification is the identity of a user.

5. Authentication – A term generally used synonymously to verification. In this thesis, we make a distinction between verification and authentication. In addition to verifying the identity of a person based on his credentials, a secure session is opened between the two parties (generally a client and a server).

6. Repudiation – A user can willfully share his credentials and later claim that they were stolen.

7. Crypto-biometric system – A system that combines biometrics with cryptography in order to remove one or more drawbacks of either of the two techniques.

8. Crypto-biometrics – The field of study covering the design, development, evaluation, and analysis of crypto-biometric systems. The research in this field can be dated back since 1998.

9. Cancelable biometric template – The transformed data obtained by applying the cancelable transformation on the reference biometric data.

10. Crypto-biometric template – The template stored in a crypto-biometric system.

11. Helper data – A term used for the data stored in a crypto-biometric system which is required for key (re)generation during verification (e.g., locked code, information for binarization, etc).

12. Crypto-bio key – A key obtained from or with the help of biometric data.

13. Session key – A cryptographic key valid only during a single communication session.

14. BioHash – Quantized multiple projections of a biometric feature vector over a randomly generated ortho-normal matrix. The binary string obtained after quantization is denoted as BioHash. The BioHash may contain variability (i.e., Hamming distance ≥ 0).

15. BioHashing – The process of generating BioHash.

16. Hash key – A user specific key assigned to the user which is required to generate the random ortho-normal matrix for BioHashing.

17. Biometric Hash – Similar to a cryptographic hash. The Biometric Hash does not contain variability (i.e., Hamming distance $= 0$).

18. Stolen biometric scenario – Many crypto-biometric systems involve a secret parameter along with the biometric data (e.g., a Hash key in BioHashing). The stolen biometric scenario is a special case when it is assumed that the biometric data for all the subjects is compromised.

19. Stolen key scenario – Many crypto-biometric systems involve a secret parameter along with the biometric data (e.g., a Hash key in BioHashing). The stolen key

scenario is a special case when it is assumed that the secret parameter for all the subjects is compromised.

20. Biometric bottle-neck problem – The result of biometric comparison is one-bit (yes/no). When integrating them in secure authentication systems, this can result in a weak link. attackers can replace the biometric recognition module with a Trojan horse which can provide the required result. We define this situation as biometric bottle-neck problem.

21. Verification string – This is a bit-string stored in crypto-biometric systems. At the time of key (re)generation, another verification string is obtained and compared with the stored one. This comparison is with zero tolerance (i.e., Hamming distance = 0). Note that, this string is not used in the key (re)generation process.

22. Systematic error correcting code – An error correcting code is said to be systematic in nature if the input to the code is present in its original form in the output.

Chapter 1

Introduction

Biometrics and cryptography are two widely used techniques for providing information security. Biometrics is defined as automated recognition of individuals based on their behavioral and biological characteristics. Biometric recognition provides a strong link between the user's identity and the authenticator. Cryptography, on the other hand, deals with protecting information with the help of secret keys. In cryptography, it is understood that the keys are kept secret, i.e., it requires trust, and this trust is projected where it is required.

Each of these two techniques has some problems associated with it. Since the biometric data are permanently associated with the user, they cannot be replaced in case of compromise. Moreover, there are privacy risks associated with the use of biometrics. On the other hand, trust is required in cryptography and if this trust is broken, the cryptographic system cannot provide security.

Fortunately, these two techniques have complementary characteristics and can be combined to design better and more secure systems. The strong association of biometric characteristics with the user's identity can be utilized to provide the trust required in cryptography. Moreover, the cryptographic techniques can be employed to provide protection to the biometric data without compromising privacy.

In this thesis, we propose various schemes to enhance security and privacy by combining biometrics and cryptography. These two technologies, biometrics and cryptography, are introduced in Section 1.1 and 1.2, respectively. The concept of combining biometrics and cryptography is introduced in Section 1.3. The motivation and the de-

sired goals of this research work are summarized in Section 1.4. Finally, the main thesis contributions are described in Section 1.5. [1]

1.1 Introduction to Biometrics

1.1.1 Biometrics

Automated recognition of individuals based on their behavioral and biological characteristics is called biometrics. Some examples of biometric characteristics are fingerprint, iris, face (2D and 3D), retina, palm print, hand veins, ear, knuckles, DNA, voice, signature, gait, typing patterns, etc. These characteristics are denoted as biometric traits or modalities. Since the biometric traits are intrinsically bound to the person, they can be used to establish his identity with high degree of confidence.

A classical biometric system, as shown in Fig. 1.1, involves two distinct phases: enrollment and recognition/comparison. During enrollment, biometric information (such as fingerprint image or voice data) is captured using specific sensors. This information is processed using specifically designed algorithms to obtain pertinent features. These features are used to create a reference biometric template for the user. The features may be represented as a fixed dimension feature vector (e.g., iris code), or a feature set of variable dimension (e.g., fingerprint minutiae).[2] This reference biometric template is required at the time of verification for comparison purposes and hence, the biometric templates for all such registered users are stored in a central template database for further comparisons.

At the time of recognition/comparison, a fresh sample of the biometric measurement is captured and similar process, up to obtaining pertinent features, is followed. These features are compared with the stored templates.

Typically, biometric systems can operate in two distinctive modes: (a) identification mode – where the system answers the question, 'who is the user?', and (b) verification mode – where the system answers the question, 'is the user really who he is

[1] Some of the most commonly used terms in crypto-biometrics are defined in the glossary which is included before this chapter.

[2] In some cases, biometric model, which is a stored function (dependent on the biometric data subject) generated from biometric feature(s), is stored instead of the biometric templates, e.g., Hidden Markov Model. In this case, during comparison, the function is applied to the biometric features of a probe biometric sample to give a comparison score [58].

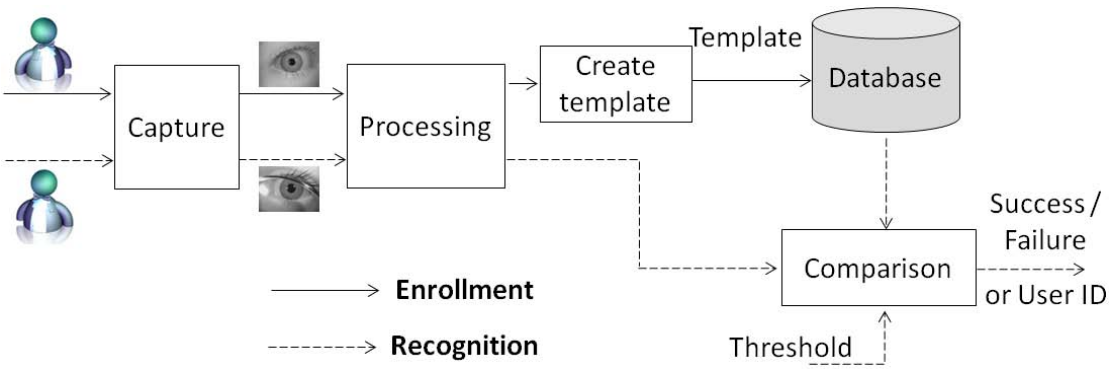


Figure 1.1: Basic idea of a biometric based person recognition system. In verification mode, the result of the comparison is either success or failure. In identification mode, the result of comparison is the User ID.

claiming to be?'. In other words, during identification, the information extracted from the fresh biometric data is compared with all the stored templates and the identity of the person to which the biometric data belongs is determined. In verification, the person who wants to get verified provides his identity along with his biometric data. A one-to-one comparison is carried out between the information extracted from the fresh biometric data and the stored template corresponding to the provided identity and the result of this comparison is either accept or reject.

A central template database is required for an identification system because all the templates are needed during comparison. If a system is intended for verification purpose only, then it is possible to store the reference templates on a personal storage device such as a smart card. In this way, a two factor (biometric and a smart card) scheme can be designed which provides increased security.

1.1.2 Multi-biometrics

An important development in the field of biometrics is to combine information from multiple biometric sources (i.e., cues). A system that consolidates the evidence presented by multiple biometric cues is known as a multi-biometric system. Depending on the sources of information, the multi-biometric system can be called as [111]:

- Multi-sensor – in which, more than one sensors are used to capture information from the presented biometric trait (e.g., capacitive and optical sensors for finger-

prints),

- Multi-sample – when more than one recording of the biometric trait is used (e.g., multiple face images can be used for creating the template),
- Multi-algorithmic – where the same biometric data is processed through multiple algorithms (e.g., minutiae and texture based features for fingerprints),
- Multi-unit or multi-instance – in which, multiple instances of the same biometric trait are used (e.g., information from images of left and right irises is combined),
- Multi-modal – when more than one biometric traits are used (e.g., a combination of iris and face).

The problem of consolidation of information presented by multiple biometric sources or cues from any of the types mentioned above is known as information fusion. The information fusion in a biometric system can be carried out at different levels such as [111]:

- sensor level – information coming from different sensors is combined,
- feature level – the biometric information extracted in form of features is combined,
- score level – match scores of individual biometric comparisons are combined,
- decision level – the results of individual biometric comparisons are combined,
- rank level – when the output of each biometric system is a subset of possible matches (i.e., identities) sorted in decreasing order of confidence, the fusion can be done at the rank level. This is relevant in an identification system where a rank may be assigned to the top matching identities.

The multi-biometric systems offer several advantages over uni-biometric systems some of which are discussed below.

- Multi-biometric systems can substantially improve the matching accuracy of the system.

- Having multiple information sources increases the size of the feature space available to individual users, thus making it possible to accommodate more individuals in a system.
- Multi-biometrics may address the problem of non-universality, e.g., in a speaker recognition system, the individuals who cannot speak cannot be enrolled. But, inclusion of another biometric such as iris may enable that person to enroll.
- When multiple biometric traits are involved, it becomes more difficult for an impostor to spoof the system.

However, the main disadvantage of multi-biometric systems is their increased complexity.

The biggest advantage of biometrics is that the biometric characteristics are permanently associated with the user. They can neither be stolen nor be shared. Moreover, the biometric characteristics possess permanence – they remain (reasonably) constant throughout the lifetime of the user. Because of the strong association of the biometric traits with the person's identity, they are widely being employed for identity verification. For example, biometric data is included in ePassports. Another example of a large scale deployment of biometric systems is the US-VISIT (United States Visitor and Immigrant Status Indicator Technology) program in which fingerprints are used for border control [3].

In spite of providing the advantage of a strong link between a person and his identity, biometric systems suffer from some drawbacks. These drawbacks are described in the next subsection.

1.1.3 Problems Associated with Biometrics

The biometric characteristics of a person are permanently associated with his identity. Though the property of permanent associativity of biometric data with the user makes biometric systems useful, it also raises some serious threats. There are two important issues related to biometric systems:

- **Non-revocability:** If the biometric data of a person stored in the database is somehow compromised, it cannot be canceled or replaced. Therefore, the person cannot use the same biometric characteristic in that system and possibly in all

other systems based on the same biometric characteristic. This is called non-revocability of biometrics. If it is a fingerprint based system, the person has an opportunity to use a different finger in that system but still this number of re-enrollments is limited. In case of face, it is not even possible.

- **Privacy compromise:** With an increasing use of biometric systems, the issue of protecting the privacy of a user is becoming prominent. User privacy is a complicated term. We define three types of privacy compromises:

  1. *Biometric data privacy compromise:* The raw biometric data of the user can be recovered from the stored templates [10, 32, 109, 33]. In some cases, the recovered biometric data can reveal certain biological conditions (e.g., fingerprints can reveal some skin conditions). Additionally, the synthesized data can be provided to the system to gain access.

  2. *Information privacy compromise:* When a person enrolls in different biometric systems with the same biometric trait, his templates in all these systems are reasonably similar (provided these systems are based on the same biometric algorithm). Therefore, templates from one database can be used to gain access to another system, and thus, the information stored in that system can be compromised.

  3. *Identity privacy compromise:* Since the templates stored in different databases of a user are reasonably similar, that person can be tracked from one system to another by cross-matching his templates from the two biometric databases. Similarly, when a system operates in identification mode, it can simply reveal that a person, to which the presented biometric belongs to, is enrolled in that particular system. This can be considered as a compromise of user's privacy. For example, consider an application of biometrics to the HIV (Human Immunodeficiency Virus) patients' (or any other sensitive group) social network. This network is a closed group of HIV patients, who share information only to the members. In this scenario, if the biometric recognition system works in identification mode, positive identification of a person based on the provided biometric data indicates that the person is a member of such sensitive group.

Other than biometrics, most commonly used credentials for establishing user identity can be considered to have one of the following two forms: (1) what you know (i.e., knowledge based, e.g., passwords, pass-phrases, etc.), and (2) what you have (i.e., possession based, e.g., token, smart card, etc.). A user is asked to present one or a combination of these two credentials, and upon their successful verification, access is granted. In such systems, the user identity is established based on the knowledge or possession of an assigned secret. Since these secrets are assigned to the user by the system, they are not intrinsically bound to the user, and hence, can be (more or less) easily stolen. These authenticators can also be shared willfully. Thus the system based on these two authenticators cannot guarantee the genuineness of the user.

Despite the weak link with the user identity, the knowledge and possession based authenticators can be easily revoked. They are assigned secrets, and hence, can be easily changed in case of compromise. Therefore, a combination of biometrics with one of these assigned authenticators can alleviate some of the problems described above. Various cryptographic techniques are used for such combination. Before we go into the details of combining biometrics and cryptography, a brief introduction to cryptography is provided in the next section.

1.2 Introduction to Cryptography

Cryptography is a process employed widely in order to secure the storage and/or transmission of electronic information. The basic idea of cryptography is shown in Fig. 1.2. It involves two phases: encryption and decryption. During encryption, the data, denoted as *plaintext* is transformed into unintelligible gibberish denoted as *ciphertext* with the help of an encryption key. The decryption process is the reverse of encryption, i.e., obtaining the plaintext from the ciphertext. The pair of algorithms that create the encryption and the reversing decryption is denoted as *cipher*.

1.2.1 Types of Cryptographic Systems – Symmetric-key and Public-key Cryptography

There are basically two types of cryptographic systems: symmetric-key cryptography and public-key (asymmetric) cryptography. In symmetric key cryptographic

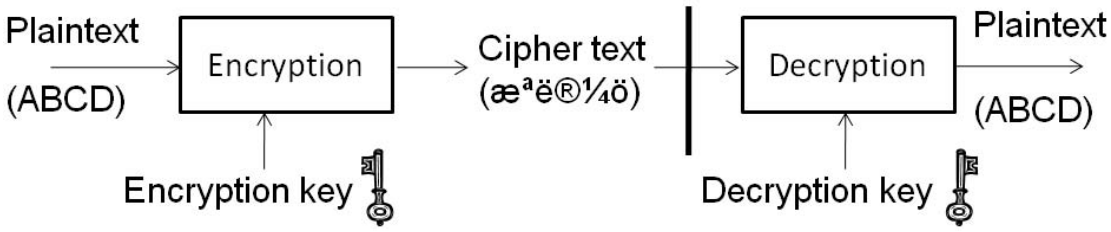


Figure 1.2: Basic idea of cryptography.

systems, the encryption and decryption keys need to be the same. Therefore, some other trusted secure mechanism must be employed in order to share the key. Examples of symmetric key cryptography systems include the Data Encryption Standard (DES) and Advanced Encryption Standard (AES) [4].

The advantages of symmetric-key cryptosystems is that they are fast and suitable for real-time applications. Moreover, the security provided by these systems is high as long as the key used for encryption/decryption is secret. However, the symmetric-key cryptosystems require additional key management techniques. The encryption and decryption key used in these systems is the same, and hence, it requires to be shared between all the entities participating in the communication. The keys can be shared through other trusted channels, e.g., by registered post or in person. Moreover, if a large amount of data encrypted with a single symmetric key is available, some cryptanalytic attacks can be made easier. Therefore, they need to be frequently renewed.

On the other hand, the public-key cryptographic systems involve a pair of mathematically related cryptographic keys – a public and a private key. The system is designed such that computation of the private key from the public key is computationally infeasible. The public key is accessible to everyone and is used for encryption. The private key, which is required for decryption, must be kept secret. Figure 1.3 shows a diagram of a generic public-key cryptographic system.

In this figure, there are two entities: Alice and Bob. Bob wants to send a message to Alice in a secure way. Alice has a public-private key pair. She sends the public key to Bob openly. Then Bob encrypts the message with this public key and sends the ciphertext to Alice. The message can be recovered from this ciphertext only with the help of the corresponding private key. Since Alice is the only one having access

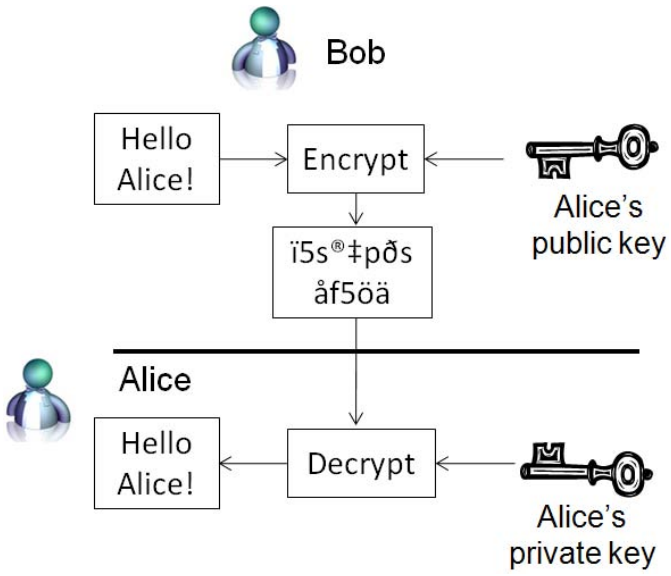


Figure 1.3: Basic idea of public-key cryptography.

to this key, only she can recover the plaintext.

Examples of public-key cryptosystems include the Diffie-Hellman key exchange protocol [43], the RSA (Rivest, Shamir, Adleman) algorithm [106], elliptic curve cryptography [88, 72], etc.

The advantage of the public-key cryptosystems is that they do not need additional key management techniques. A secure channel is not required for sharing the key since the only shared key is the public key. Unfortunately, the public-key cryptosystems are computationally expensive and hence too slow for practical purposes. Therefore, in practice, hybrid systems are employed in which a symmetric key is shared with the help of a public-key cryptosystem. The symmetric key is used for encryption/decryption in a faster symmetric-key cryptosystem. An example of such hybrid system is the widely used TLS (Transport Layer Security) protocol [41, 42]. This protocol is used in HTTPS to secure the world wide web traffic carried out by HTTP. HTTPS is used for secure e-commerce applications such as online payments through internet, online banking applications, etc.

Public-key cryptosystems are susceptible to man-in-the-middle attack. In this attack, a cryptanalyst, let's call her Eve, breaks the connection between Alice and Bob. Consider the same situation shown in Fig. 1.3 where Bob wants to send a message to Alice. In this attack (shown in Fig. 1.4), Eve replaces the public key of Alice with her

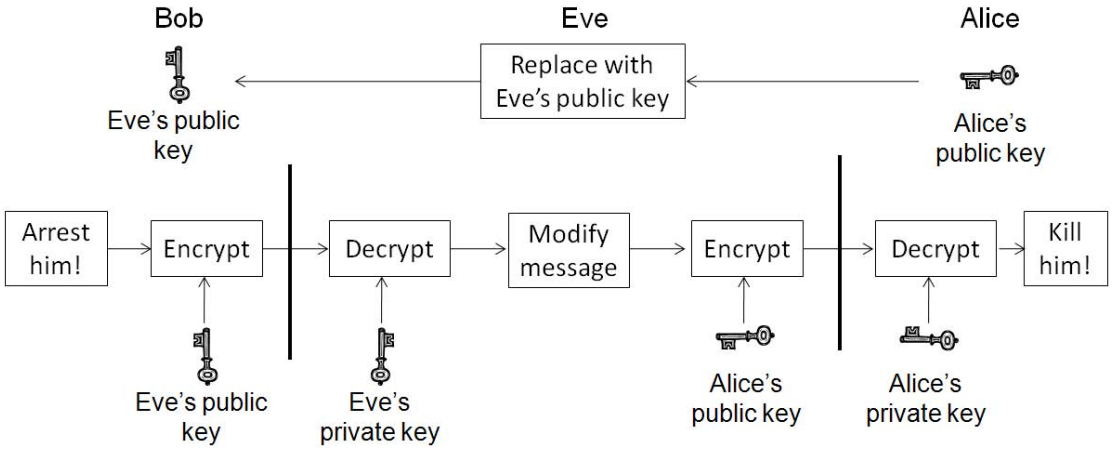


Figure 1.4: Man-in-the-middle attack on a generic public-key cryptosystem. Eve replaces the public key of Alice with her own public key and sends to Bob. Thus, she can access the message sent by Bob. Eve can also modify the text, e.g., here Bob sends the text "Arrest him!" which Eve changes to "Kill him!".

own public key and sends it to Bob pretending it to be of Alice. When Bob encrypts the message with Eve's public key and transmits it, Eve can decrypt it with her own private key. Then she can send a malicious message to Alice using Alice's public key. In this way, Alice can think that she is securely communicating with Bob and vice versa, while Eve knows and even can control all the communication between them. A schematic diagram showing the man-in-the-middle attack on a public-key cryptosystem is presented in Fig. 1.4. Note that, in the figure, the message sent by Bob to Alice is "Arrest him!" which is changed by Eve to "Kill him!".

In order to overcome such attacks, various authentication mechanisms are employed in public-key cryptosystems. Nowadays, the most common mechanism is to use trusted third party certificates which certify the authenticity of the public keys. The third party is an entity trusted by all other entities participating in the communication. Some systems also employ password based authentication. Sometimes, a combination of the two (certificates and passwords) is used. For example, in a typical online banking application, the authenticity of the server is confirmed with the help of third party certificates while that of the user is confirmed based on his secret credentials (password or PIN). Even though the third party certificates have advantages, this system can work only if the two communicating parties share a common understanding of trust over the

third party.

1.2.2 Problems with cryptography

According to the Kerckhoffs' principle, security of a cryptographic system lies entirely on the secrecy of the key [70]. Additionally, for security reasons, the cryptographic keys are required to be long. For example, the possible lengths of keys required in the AES are 128, 192, or 256 bits. For public-key cryptographic systems such as RSA, the key lengths are even higher (e.g., 512, 1024, or 2048 bits). Clearly, a user cannot remember such long keys and therefore, the keys need to be stored somewhere, e.g., on a smart card or in a computer.

In order to restrict access to these keys only to legitimate users, authentication mechanisms are used. Traditionally, authentication mechanisms employed in cryptography are knowledge based (e.g., passwords) or possession based (e.g., token, smart card, etc.). These authenticators are assigned to the user identity and do not necessarily indicate the presence of the person to which they belong. Therefore, they can be (more or less easily) stolen by an attacker, and in this situation, the system cannot distinguish between the attacker and a legitimate user.

Another issue related with these authentication mechanisms is repudiation. A user can willfully share his credentials and later claim that they were stolen. Thus, such a system can be easily cheated.

1.3 Introduction to Combination of Biometrics and Cryptography

From the description of problems associated with biometrics and cryptography (Section 1.1.3 and 1.2.2, respectively), it is clear that both of these techniques need improvements. Biometrics is good at identity verification with better protection against repudiation. Cryptography, which is great in providing security, privacy, and anonymity, requires better verification mechanisms. The common verification mechanisms used in cryptography are based on passwords and/or tokens which are vulnerable due to the weak link between the person's identity and the associated cryptographic keys.

In order to overcome these drawbacks of biometrics and cryptography, over

the last decade (since 1998), a new innovative multidisciplinary research field denoted as *crypto-biometrics* has emerged, that combines the two techniques. The systems in which biometrics is combined with cryptography are denoted as *crypto-biometric systems* in this thesis. With the help of these systems, revocability, privacy, and template diversity can be induced in biometric systems, as well as biometrics based cryptographic keys, denoted as *crypto-bio keys*, which are strongly linked to the user identity, can be obtained.

The biggest difficulty in combining biometrics and cryptography is that cryptography is precise whereas biometric data contain variability. Cryptography deals with binary keys which need to be exactly the same every time. Unfortunately, biometric data contain some variations in each measurement. Therefore, biometric data cannot be used directly as cryptographic keys leading to the need to develop specific techniques for designing the crypto-biometric systems.

A through review of such crypto-biometric systems found in literature, is presented in Chapter 2. We propose to divide these systems into two main categories according to their major goal as: (1) Protection of biometric data, and (2) Obtaining cryptographic keys with biometrics. The systems in each of these categories are further classified based on the applied methodologies for combining biometrics with cryptography. These two categories are described in Section 2.1 and 2.2, respectively.

Integrating these crypto-biometric systems in secure applications is another challenging problem. Especially, for any cryptographic application, the cryptographic keys (and in turn, crypto-bio keys) need to be shared. There are very few systems that facilitate secure sharing of these crypto-bio keys without the need of conventional public key infrastructure. A review of such systems is also presented in Section 2.3.

1.4 Motivation and Goals

The crypto-biometric systems should be designed in such a way that they inherit the advantages offered by biometrics and cryptography while eliminating their disadvantages. We set the following goals while designing the crypto-biometric systems:

1. **Identity verification and non-repudiation:** The system should be able to confirm the identity of the user with high degree of confidence. It also indicates

that the system should resist repudiation attempts carried out by the users. Involvement of biometrics helps achieve this property.

2. **Revocability:** If the stored user template is compromised, it should be possible to cancel that template and reissue a new one. Additionally, the newly issued template should not match with the previously compromised template. Thus revocability does not mean just to cancel the old template and issue a new one; it also means that, the authentication rights of the old authenticator are revoked. The system should be able to reject a person if he provides the authenticator linked with the old template. Note that, biometrics alone cannot provide this property because biometric characteristics cannot be changed while systems using passwords and tokens have excellent revocability.

3. **Template diversity:** It should be possible to issue different templates for different applications related to the same user. These templates should not match with each other and should make cross-matching impossible. Password and token based systems are good at that, though practically, password diversity can be argued. Biometrics, by itself, cannot have template diversity.

4. **Privacy protection:** Crypto-biometric systems should protect privacy. These systems should protect the biometric data privacy, information privacy, and identity privacy, defined in Section 1.1.3.

5. **Performance:** The crypto-biometric system should not degrade the verification performance of the underlying baseline biometric system.

6. **High key entropy:** If the goal of the crypto-biometric system is to obtain crypto-bio keys, the entropy of such keys should be high.

In this thesis, we propose various systems which satisfy these desired goals. We have proposed systems from each of the two classes of crypto-biometric systems mentioned before. This thesis is organized as follows: a detailed state of the art in crypto-biometrics is given in Chapter 2. The experimental evaluation strategies are explained in Chapter 3. In Chapter 4, we propose a shuffling based cancelable biometric system. In order to reduce the variability in biometric data, a novel approach based

on error correcting codes is proposed in Chapter 5. In Chapter 6, a hybrid system is introduced for obtaining biometrics based cryptographic keys. A multi-biometrics based cryptographic key regeneration system is proposed in Chapter 7. Protocols for sharing biometrics based session keys are described in Chapter 8. Finally, conclusions, perspectives, and future research directions are discussed in Chapter 9.

1.5 Summary of Thesis Contributions

In order to overcome the problems associated with biometrics and cryptography discussed in Section 1.1.3 and 1.2.2, respectively, we propose various crypto-biometric systems in this thesis. The main contributions of the thesis are:

- **Classification Scheme for Crypto-Biometric Systems:** We propose a classification scheme for crypto-biometric systems based on their main goals and methodologies. The two major categories of crypto-biometric systems are: (1) Protection of biometric data, and (2) Obtaining cryptographic keys with biometrics. The systems in the first category add revocability, template diversity, and privacy protection to classical biometric systems. The systems in second category provide biometrics based cryptographic keys and can optionally have the properties of revocability, template diversity, and privacy protection. The detailed state of the art in crypto-biometrics is given in Chapter 2.

- **Shuffling Based Cancelable Biometric System:** In order to tackle the problems of nonrevocability, non-template diversity and to provide privacy protection, we propose a simple shuffling based cancelable biometric scheme. This scheme randomizes the biometric data with the help of a shuffling key. The data to be shuffled is divided into blocks and these blocks are rearranged according to the bit values of the shuffling key. The original biometric data cannot be obtained from the shuffled data as long as the shuffling key is secure. The advantages of this scheme are: (a) it induces revocability in biometric systems, (b) adds template diversity, (c) makes cross-matching impossible and protects privacy, (d) helps improve the verification performance because it increases the impostor Hamming distance without changing the genuine Hamming distance. This system is described in details in Chapter 4.

- **Using Error Correcting Codes to Reduce Variability in Biometric Data:** Many crypto-biometric systems use Error Correcting Codes (ECC) to obtain keys from biometrics data. But these systems can correct either all the variabilities in the biometric data (treated as errors) or none of them. This puts a restriction on the recognition performance, and the recognition performance of the crypto-biometric systems generally degrades compared to those of the baseline biometric systems. In Chapter 5, we propose a novel method which can correct fractional amount of errors in biometric data. If the two biometric data being compared are b_1 and b_2, and b_2' is the b_2 after error correction, and D is the distance function, then $D(b_1, b_2') \leq D(b_1, b_2)$. When combined with the above mentioned shuffling scheme, it significantly improves the recognition performance.

- **A Hybrid Scheme to Obtain Cryptographic Keys Using Biometrics:** In Chapter 6, we propose a hybrid scheme to obtain high entropy cryptographic keys using biometrics which combines techniques from transformation based and key regeneration based crypto-biometric systems. The shuffling scheme described above is first applied on the biometric data to make it revocable. This data is then used in a fuzzy commitment based key regeneration scheme. We propose two adaptations of this scheme: one for iris and the other for face. With this scheme, we can obtain long keys with 83 to 93 bit entropy from iris while 110 to 112 bit entropy from face.

- **Using Multi-Biometrics to Obtain High Entropy Cryptographic Keys:** In order to increase the security against spoofing and to increase the length and entropy of the keys, we propose a novel scheme to combine information from multiple biometric sources. The information is combined in the feature domain. A novel idea of weighted error correction is proposed which allows applying different weights to different biometric information. We call this scheme as *FeaLingECc* (*Fea*ture *L*evel Fus*i*o*n* through Wei*g*hted *E*rror *C*orre*c*tion). We propose two systems: a multi-unit type system that combines information from left and right irises of a subject, and a multi-modal system which combines iris with face data. A significant increase in the key entropy is observed along with an improvement in the verification performance. The estimated entropy of the keys obtained with

the two-iris system is 189 bits while that from the iris-face system is 183 bits. This scheme is described in Chapter 7.

- **Biometrics Based Session Key Generation and Sharing Protocol:** The above systems are about obtaining cryptographic keys using biometrics. But in cryptography, the keys need to be securely shared in order to use them for secure communication or data storage. In Chapter 8, we propose two novel protocols: the first one is to share the crypto-biometric keys generated using the scheme described above. But there is one problem with using the same key for encryption of a large amount of data. It can make some cryptanalytic attacks easy. The second protocol is used to generate and share biometrics based session keys. These session keys are valid only for a particular session. These session keys are generated at the beginning of the session and shared between all the entities participating in the communication session. These two protocols allow mutual authentication between the two parties - client and the server - without the need of trusted third party certificates. It also allows to securely update the stored template online. The template stored in the database at the server is revocable (e.g., it is created by applying the shuffling scheme on the biometric data).

Chapter 2

Crypto-biometrics – State of the Art

In this chapter, we present a detailed overview of the research work going on in the field of crypto-biometrics. Since this field is relatively new (it started from 1998), it lacks a uniform classification of the various techniques found in literature. Many researchers call this research field as template protection and classify these systems as feature transformation and biometric cryptosystems. We would like to point out that template protection is not the only aim of the systems covered in this field. In fact, many systems classified as biometric cryptosystems were originally designed for obtaining cryptographic keys which are strongly linked to the person's identity. Many of these systems do not necessarily meet the criteria for a template protection system (e.g., revocability and privacy protection).

We call this research field as crypto-biometrics because all these systems combine techniques from biometrics and cryptography. As described in Section 1.1.3 (page–5) and 1.2.2 (page–11), biometrics and cryptography have certain limitations. Crypto-biometric systems attempt to eliminate these limitations by combining the two techniques. Based on their application and functionality, we classify the crypto-biometric systems into two main categories as: (a) Protection of biometric data, and (b) Obtaining cryptographic keys with biometrics.

In the first category, cryptographic techniques, such as encryption, hashing, transformation, etc., are used to protect the biometric data. The outcome of these

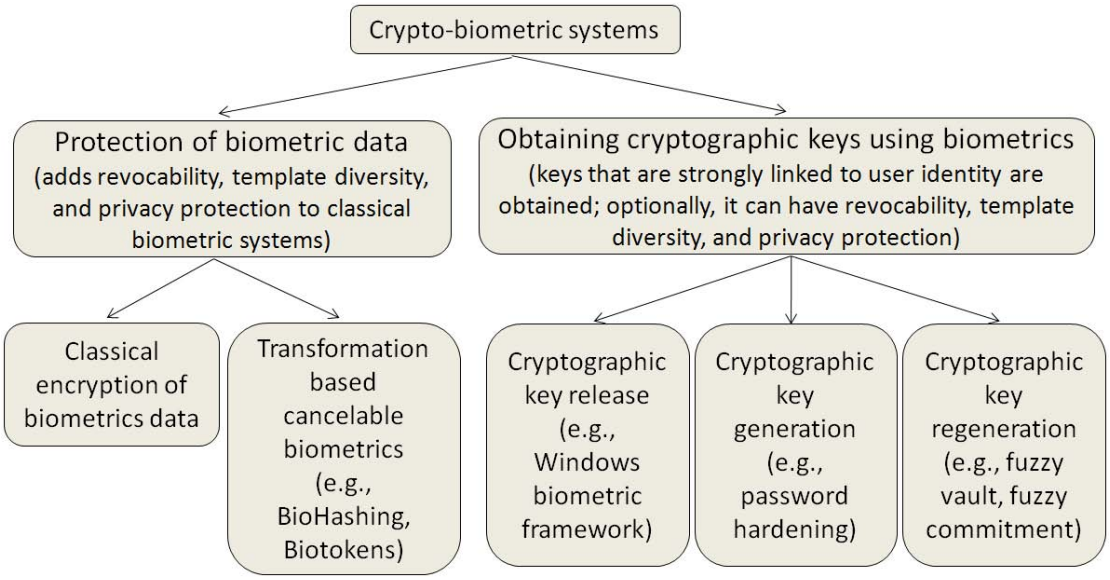


Figure 2.1: The proposed classification of crypto-biometric systems. Primary criterion for this classification is the main goal of the system. Secondary criterion is the methodology used in the system.

systems is a one-bit verification result similar to the classical biometric systems. On the other hand, in the systems from the second category, biometric data is used to obtain cryptographic keys (denoted as crypto-bio keys). The systems in these two categories are further divided depending on how these techniques are combined. A schematic diagram showing this classification of crypto-biometric systems is shown in Fig. 2.1. These two categories are discussed in Section 2.1 and 2.2, respectively. Note that, the techniques described in these categories can be combined to achieve additional features. Some other reviews can be found in [138, 34].

The focus of earlier classification attempts (e.g., Jain et al. [60]) is on the protection of biometric data by means of cryptographic techniques. Therefore, the issue of practical usability of the crypto-biometric systems, and in particular the employability of the crypto-bio keys, is not addressed. Since we also address the crypto-bio key sharing issue in this thesis work, a review of such key sharing protocols is presented in Section 2.3.

The systems described in this chapter are based on different biometric modalities. Hence, the performance comparison in terms of absolute values of False Acceptance Rate (FAR), False Rejection Rate (FRR), and /or Equal Error Rate (EER) is irrelevant. Therefore, we compare the performance of the crypto-biometric system with the

baseline biometric system it uses. In this way, we can get a fair idea of the effects of the modifications on the performance. Moreover, since the crypto-biometric systems involve some secret parameter (a key, or a password) along with biometric data, it is required to know the effect of compromise of one of the factors on the overall system performance.

The two categories introduced above are discussed in Section 2.1 and 2.2. The review of biometrics based key sharing protocols is presented in Section 2.3. The conclusions and perspectives are given in Section 2.4.

2.1 Protection of Biometric Data

The systems in this category use cryptographic techniques to add some of the desired characteristics (such as revocability, privacy protection, etc.) to biometrics based verification systems.

We divide these systems in two subcategories as: (1) systems using classical encryption of biometric data, (2) systems employing transformation based cancelable biometrics. These are discussed in the following subsections.

2.1.1 Classical Encryption of Biometric Data

The simplest solution is to encrypt the biometric template with a user specific password before storing it in the database. Classical encryption techniques, such as the Advanced Encryption Standard (AES) [4], can be used for the purpose. The comparison between the reference and test data cannot be performed in encrypted domain. Therefore, it is required to decrypt the data for comparison purposes at the time of verification. Thus, the biometric template is first recovered from the encrypted template and the comparison is carried out. Figure 2.2 shows a schematic diagram of a generic system of this category.

Since this system combines biometric information with an assigned secret, in case of compromise, a new template can be issued by changing the password achieving revocability. Moreover, different passwords can be used for different systems to issue different templates, thus achieving template diversity. The security can be enhanced by using a smart card for storing the encrypted template [16].

The drawback of this configuration is that the comparison between a query

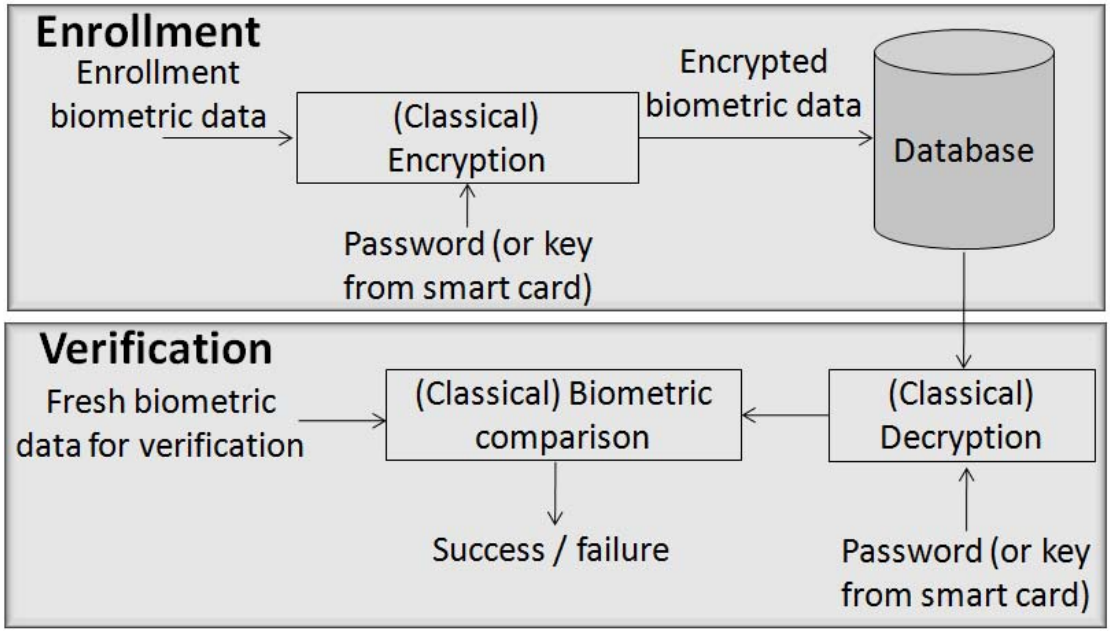


Figure 2.2: Use of classical encryption for protection of biometric data.

sample and the targeted template is carried out in a classical way. Therefore, such systems cannot improve the performance of the underlying biometric system. This factor can be seen as a disadvantage from the usability point of view because, in this system, a user is asked to provide password along with his biometric data but there is no performance enhancement. Moreover, once the biometric template is decrypted, it can be misused in different ways (e.g., to obtain the raw biometric data).

2.1.2 Transformation Based Cancelable Biometrics

In this category, we group the systems in which the biometric data (raw data such as face image or feature vector such as iris code) is transformed with user specific (generally one-way) transformations. Figure 2.3 shows a schematic diagram of a generic, transformation based, cancelable biometric system. Note that the transformation applied on the biometric data at the time of enrollment as well as verification is the same. The transformation should be such that the discriminative properties of the original biometric data are preserved in the transformed domain also. This allows the biometric comparison in transformed domain. The user verification decision is taken based on this

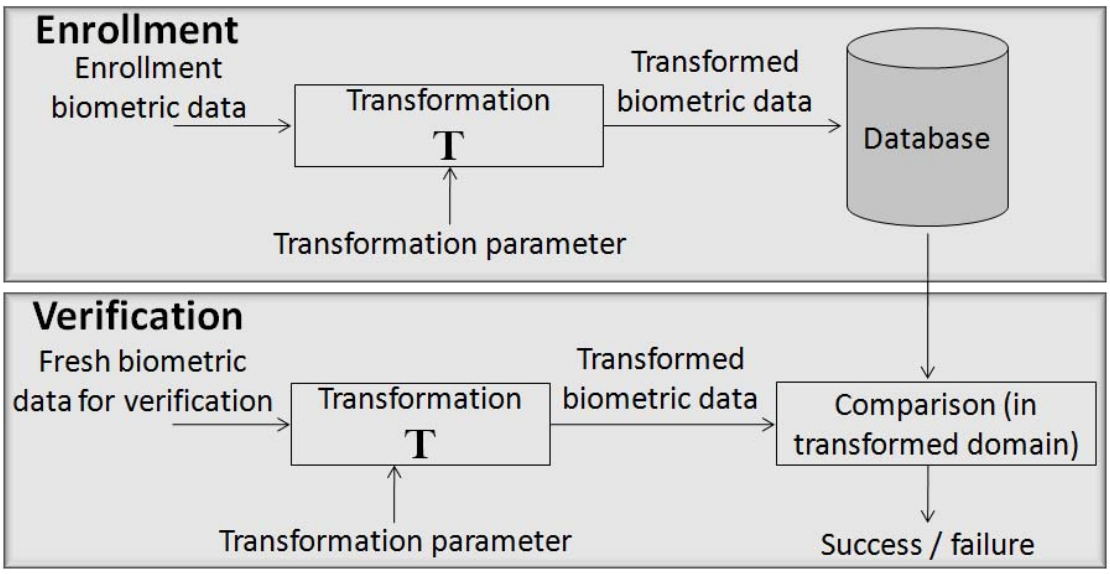


Figure 2.3: Transformation based cancelable biometrics.

comparison. The transformation is user specific which is generally controlled using a secret key or password. The transformation acts like an encryption scheme and protects the biometric data. In case of compromise, the transformation parameter (the key or password) can be changed to issue a fresh template, thus achieving revocability. Ideally, this newly issued template does not match with the old template. Similarly, template diversity can be achieved by using different transformation parameters across different systems.

In the following paragraphs, we take a brief look at some of the systems in this category. A summary of these systems is presented in Table 2.1.

Ratha et al. [104] (in 2001), introduced the term cancelable biometrics proposing transformation of the biometric signal (or features) using irreversible transformations. The transformation parameters are user specific. In their recent article (in 2007), Ratha et al. [105] proposed three different transformations (Cartesian, polar, and functional). These transformations provide different amount of security to the biometric data. They tested their system on a private database with 188 pairs of fingerprint images and reported that the performance of the underlying biometric system always degrades after transformation.

Another interesting and widely used technique called BioHashing [51] (2003)

was used by Jin et al. [62] (2004) for cancelability.[1] In BioHashing, a randomly generated, user specific key (denoted as hash key) is used to generate an ortho-normal matrix. The biometric feature vector is projected onto this matrix and after thresholding, a binary vector is obtained which is denoted as BioHash. Lumini and Nanni [78] (2007) proposed an improved version of this BioHashing scheme with modifications such as binarization threshold variation, space augmentation, feature permutation, and feature normalization. They reported that, in general, BioHashing scheme improves the performance of the underlying biometric system. But, the drawback is, in the stolen key scenario, the performance generally degrades compared to the baseline biometric system.

Teoh et al. [127] presented an analysis of the BioHashing method as a Random Multispace Quantization of the biometric data and the hash key. In their discussion, they argue that the performance degradation in the stolen key scenario is because of the quantization process.

In addition, we would like to comment on the improvement observed in BioHashing. From [62], it is seen that, the genuine Hamming distances decrease after application of BioHashing. The same case applies to the stolen key scenario where the same hash key is used for different biometric inputs from impostors and hence decreases the impostor Hamming distances. This results in decrease in performance in the stolen key scenario.

Other works based on the BioHashing technique are [129, 61, 73, 126, 128, 99, 17].

Savvides et al. [113] (in 2004) proposed cancelable biometric filters for face recognition where they use a random kernel (which can be obtained from a PIN) to encrypt facial images. They report that the performance of the baseline biometric system does not change when the transformation is applied.

Boult et al. [23] (in 2007) applied the *biotoken* scheme to fingerprint which they earlier proposed for face in [22]. The scheme is based on robust distance matching techniques. They reported 30% improvement in the verification performance for the fingerprint biotoken system.

Maiorana et al. [80] (in 2008) proposed a different way of transformation which

[1] The BioHashing technique was originally proposed for key generation.

is applied to Hidden Markov Model (HMM) based signature features. It makes use of a randomly generated sequence to divide the features into parts on which convolution is applied. In their later papers [83, 81], they show that the number of different templates that can be generated using this technique is limited. In order to increase this number, they proposed some improvements. The drawback of many of these cancelable biometric systems is that their performance degrades compared to the baseline biometric system. In some cases, the performance improves, however, the improvement is because of the additional parameter (such as password, PIN, key, token, etc.). Such systems should be analyzed for their verification performance in the stolen key (also called as stolen token) scenario. Such analysis is not reported in most of these works. For BioHashing based systems, the performance in the stolen key scenario degrades compared to the baseline biometric system.

As opposed to these observations, the performance of the Farooq et al. system [47] in the stolen key scenario remains equal to the performance of the baseline biometric system. This is an important property of this system. The cancelable biometric system that we propose in Chapter 4 possesses this property.

Some other works in the cancelable biometrics category can be found in [11, 131, 150].

2.2 Obtaining Cryptographic Keys with Biometrics

The systems in this category employ biometric recognition techniques to increase the security in a cryptographic framework. Typically, the keys needed for cryptographic applications are obtained using biometric data. These systems are classified in three categories: (a) key release, (b) key generation, and (c) key regeneration, described in following subsections.

2.2.1 Cryptographic Key Release Based on Biometric Verification

The easiest way to integrate biometric systems in a cryptographic framework is to store cryptographic keys securely and release them only after successful biometric verification. Thus, classical biometric user verification is involved in this configuration which provides the verification result and based on which the key (or some parameters

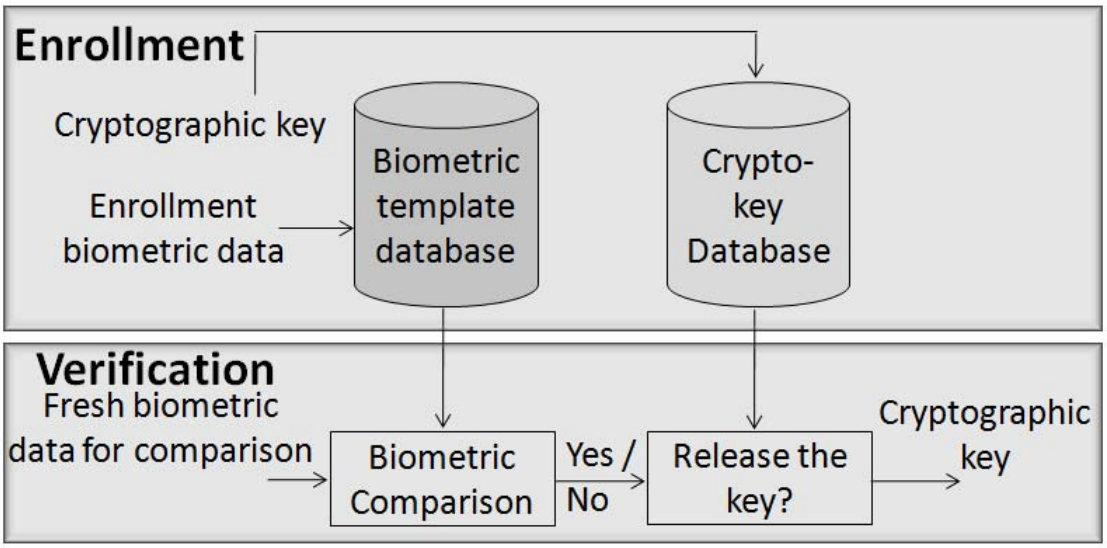


Figure 2.4: Cryptographic key release based on biometrics.

required to generate the key) are released. A schematic diagram of this configuration is shown in Fig. 2.4. Note that, these systems need to store classical biometric templates as required by the classical biometric system.

An example of this system can be found in Itakura and Tsujii [59] where some secret parameters are released upon successful biometric verification. The released parameters are required to obtain the cryptographic keys.

Another example of key release system is the Windows Biometric Framework [86] included in the Microsoft Windows 7. The Windows biometric framework allows users to login to their windows (and other) accounts with their biometric data (currently fingerprints).

The advantage of this approach is its simplicity. There is no need to design specific algorithms for integration of biometrics and cryptography. But, there are few problems associated with this configuration. The biggest and foremost concern is that the biometric template is stored in the system and thus the system inherits all the drawbacks of the biometric system such as nonrevocability, non-template diversity, and no privacy protection. These issues can be addressed by employing transformation based cancelable biometric system instead of classical biometric system in this configuration. Another problem is that the verification result of classical as well as the cancelable

biometric systems is a single bit information. Therefore, this configuration also suffers from the biometric bottle-neck problem. Because of these shortcomings, the systems in this category cannot achieve the desired enhancement in security. They are mentioned here for the sake of completeness.

2.2.2 Cryptographic Key Generation from Biometrics

From security point of view, a better solution than the key release is to generate a stable bit-string directly from the biometrics. Figure 2.5 shows a schematic diagram of a generic, biometrics based key generation system. These systems are sometimes denoted as template-free biometrics because, in some cases, they do not need storage of templates but they store only the verification string. As shown in Fig. 2.5, the biometric template is not stored in these systems but, only a verification string derived from the biometric data (or from the generated key) is stored. A similar verification string is extracted at the verification time and validity of the key is established by comparing the two verification strings. A summary of the key generation systems is presented in Table 2.2.

One of the earliest works in this category is that by Davida et al. [39] (1998) who proposed an off-line biometric verification scheme. In their proposal, a key is derived from iris data. They proposed to use majority coding and Error Correcting Codes (ECC) to stabilize the biometric data. But no experimental results are reported. From our experiments [66], we found out that the majority coding does not work with iris data. Moreover, the assumption in [39] that the Hamming distance between genuine iris codes is 10% is too restrictive.

Another work in this category is the *Hardened password* by Monrose et al. [90] (1999). They combined a typed password with a short string extracted from the user's typing patterns such as durations of keystrokes, and latencies between keystrokes. In a follow-up paper [89], they used voice biometrics instead of keystroke dynamics. They reported increase in entropy from 12 to 46 bits along with a considerable decrease in the FRR.

Vielhauer et al. [142] (2002) proposed a biometric hash generation scheme based on online signatures. The word hash in "biometric hash" is analogous to the cryptographic hash and should not be confused with the BioHash [62]. The biometric

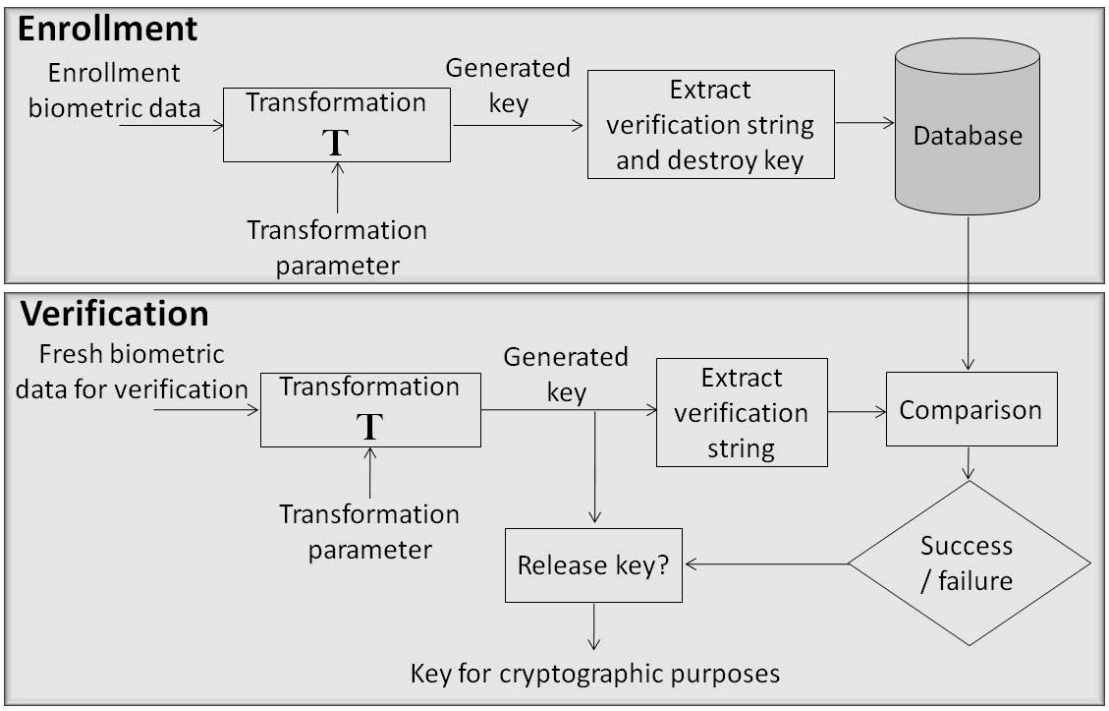


Figure 2.5: Cryptographic key generation using biometrics.

hash obtained from the same biometric source is a strictly constant bit-string (i.e., Hamming distance = 0) whereas, the BioHash does contain some variability (i.e., Hamming distance > 0).

The BioPKI system proposed by Hao and Chan [55] can generate private keys using biometrics. But, this system requires storage of classical biometric templates in order to generate the private key. Therefore, it inherits all the drawbacks of the classical biometric systems.

Goh and Ngo [51, 130] (2003) proposed a random projection based method called BioHashing. An inner product between the biometric feature set and a set of randomly generated vectors is calculated and quantized to obtain a binary string. The random projection space can be obtained from a seed stored on a token. Cryptographic interpolations of Shamir secret shares are then used to obtain a cryptographic key.

Fuzzy genetic clustering is shown to be effective in the work of Sheng et al. [118] (2008). They applied it to handwritten signature data. But, the false rejection rate is quite high (13.1%) for this system. Moreover, the keys have only 20 bits of effective

information which is quite low from the security point of view.

Argyropoulos et al. [12] (2009) used an ECC scheme similar to Davida et al. [39], with a side channel coding approach in which the enrollment biometric data is encoded using a systematic error correcting code.[2] Only the parity symbols of the encoded codeword are stored as a template. At the time of verification, these parity symbols are combined with the test data and decoded to obtain the original enrollment biometric data.

The theoretical construction, called as Helper Data System, proposed by Tuyls and Goseling [134] (2004) is a general idea of obtaining biometric based keys. It can be classified as key generation system. Principally, it is similar to the Davida et al. [39] construction. In the helper data system, a secret is extracted directly from the reference biometric data. The helper data is created such that the same secret can be reconstructed with the help of the helper data and a fresh biometric sample.

However, the term helper data is used in general for any auxiliary data stored in the crypto-biometric system which is required for extraction of the secret key during verification. For example, the locked code, which is a combination of a randomly generated secret bound to the reference biometric data, is also called as helper data.

The difficulty of the key generation approach is that the biometric data are not stable. Therefore, it becomes highly difficult to extract a stable string from such data without adversely affecting the verification performance. Scheidat et al. [114] experimentally analyzed the Biometric hash generation algorithm of [142] in both, cancelable biometrics and key generation mode. Their study clearly indicates the difficulty of the key generation approach as compared to the cancelable biometrics approach. For a particular test, in cancelable mode, the system has an EER of 3.1% but, for a similar test, the key generation system has an FRR of 23% and an FAR of 2.39%.

Another possible problem with the key generation approach is that the generated key can be considered as a reduced dimensional representation of the biometric data. Hence, unless the system involves some assigned secret (like a seed on a token or a PIN), the keys cannot be revocable.

[2] An ECC is systematic when its output contains the input in original form appended with parity symbols.

2.2.3 Cryptographic Key Regeneration Using Biometrics

The most widely studied approach for obtaining keys using biometrics is the key regeneration. Sometimes it is also called key binding [60]. The basic idea is that a randomly generated key is combined with the biometric data using cryptographic techniques and that key is later retrieved from the combined data at the time of verification. A schematic diagram of this approach is shown in Fig. 2.6. The key regeneration systems are summarized in Table 2.3.

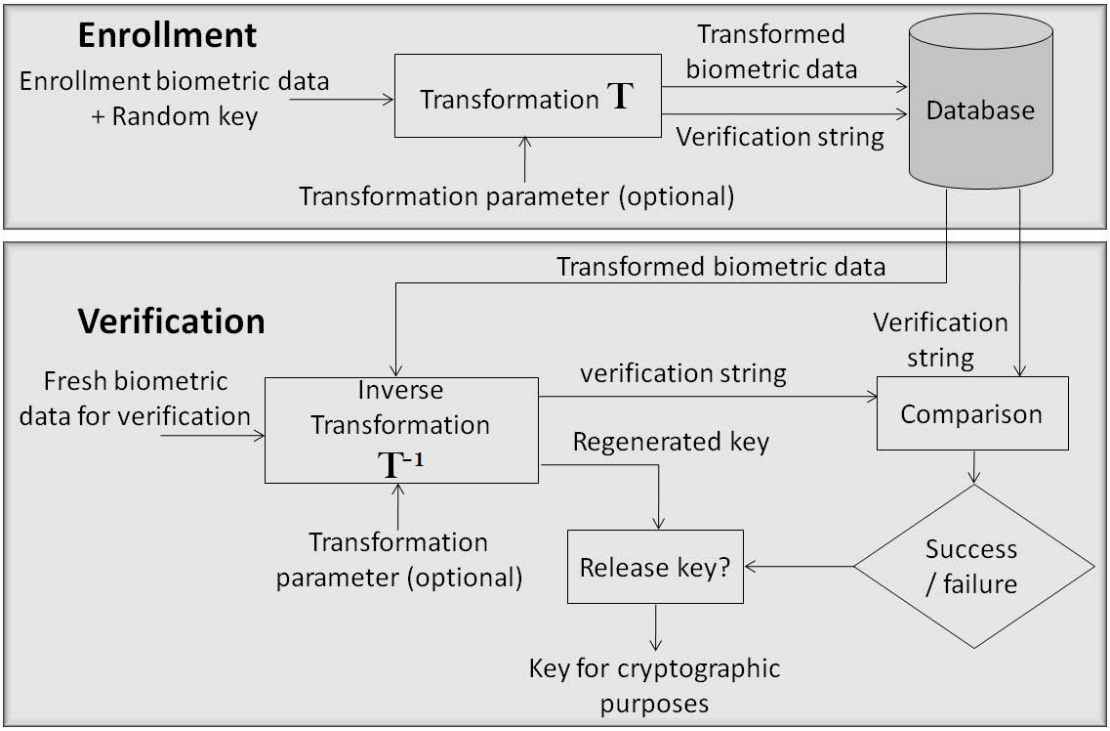


Figure 2.6: Cryptographic key regeneration using biometrics.

One of the earliest works in this approach is by Soutar et al. [120] (1999). They used a signal processing approach to bind the fingerprint data with a random phase-only function. Phase part of the Fourier transform of the fingerprint data is multiplied with the random phase-only function so that the combined data is cryptographically secure. This combined data is then linked with a random cryptographic key using a lookup table. This key is retrieved at the time of verification. Neither the experimental evaluation nor the security analysis of the proposed method are given in [120].

Juels and Wattenberg [64] (in 1999) proposed a theoretical scheme called *fuzzy commitment*. A random key is encoded using Error Correcting Codes (ECC) and is then XOR*ed* with the biometric data. The XOR*ed* data is cryptographically secure because neither the key nor the biometric data can be obtained from it without providing one of the two. The random key is retrieved at the time of key regeneration by providing fresh biometric data. This system requires ordered biometric data in binary form.

Juels and Sudan [63] (in 2002) developed another theoretical proposal called *fuzzy vault*. The requirement of ordered biometric set of the fuzzy commitment scheme was removed by the fuzzy vault, making it possible to use with biometric modalities such as fingerprint, the minutiae set of which does not have an order. A similar theoretical approach, called fuzzy identity based encryption was proposed by Sahai and Waters [112] (in 2005).

Dodis et al. [44] (in 2004) introduced the concepts of secure sketches and fuzzy extractors. They provide theoretical security analysis of the crypto-biometric systems which can be applied to the fuzzy commitment and fuzzy vault schemes. Boyen [24] presented a theoretical analysis of the fuzzy extractors and pointed out some of their shortcomings.

Based on these theoretical proposals, many key regeneration schemes are found in literature. Clancy et al. [36] applied the fuzzy vault to fingerprints. They also added some chaff points in the vault to obscure the minutiae. This approach was followed by Uludag and Jain [137] who employed fingerprint orientation field based helper data along with the vault to improve the performance. But, both these systems have high FRR (30% for [36] and 20% for [137]).

Nandakumar et al. [97] further improved the helper data extraction and fingerprint alignment algorithms presented in [137]. Moreover they used more than one fingerprint impressions for encoding/decoding. They applied mosaicing technique [110] to combine the minutiae and helper data from two fingerprint impressions which improves the performance. They achieved 30% overall improvement in Genuine Acceptance Rate (GAR=1-FRR) over the system in [140, 137, 139]. But, the GAR is still less than the baseline fingerprint system. Additionally, fingerprint based fuzzy vaults are vulnerable to attacks as described by Scheirer and Boult [117].

Based on the fuzzy commitment proposal, Hao et al. [54, 53] (in 2006) proposed

a key regeneration system for iris biometrics. Iris code extracted from an iris image is an ordered set of binary values and thus meets the fuzzy commitment scheme requirements. In order to cope with various errors that can occur in iris data, they proposed a two level error correction scheme. At first level, Hadamard codes are employed to correct random errors. The Reed-Solomon codes are later used in the second level to correct error bursts that can occur in the iris codes due to eye-lid occlusions, eye-lashes, reflections, etc. Overall error correction capacity of this scheme is 27%. The theoretical security analysis presented in [54] shows that the keys obtained with this system have 44-bit entropy.

Bringer et al. [26] proposed a system similar to [54] with a different ECC scheme: Reed-Muller codes with product codes. The errors in iris codes can contain bursts. In order to break the error bursts in iris data and distribute the errors uniformly throughout the iris code, they employed a random interleaver which increases the error correction efficiency. They obtained 42-bit keys but did not provide any security analysis in terms of entropy.

In the face-based fuzzy sketch of Sutcu et al. [124], user specific randomization is applied to the biometric feature vectors before processing them through the sketch creation or decoding process. The randomization matrix acts as a user specific secret and thus improves the performance. But, they did not provide analysis considering the compromise of this secret. Moreover, the entropy of the keys is 49-bits.

Maiorana et al. [82] used the fuzzy commitment approach to design their signature based key regeneration scheme. In their proposed approach, they employed adaptive selection of the error correcting codes. In particular, they used BCH codes and selected the parameters of BCH codes adaptively according to intra-user variations. The proposed idea of user adaptive ECC is interesting especially considering the fact that the intra-user variability can vary among different users. The performance shows a little improvement than the baseline biometric system.

There are various template protection schemes in literature that make use of reliable component selection methods to stabilize the biometric data [133, 71, 141, 69, 148, 145, 143, 49]. Basically, these systems, which their respective authors like to call as helper data systems, are fuzzy commitment based schemes. In addition to the commitment, these systems need to store some user specific helper data required for

the reliable bit selection. A critical issue with these systems is that the additional user specific information can be used for cross-database matching, and may compromise the privacy.

There are also proposals which combine the concepts from cancelable biometrics with key regeneration systems. Such hybrid systems attempt to achieve the advantages of both these classes (i.e., revocability, template diversity, privacy protection along with personal keys). An example of such system is the hardened fuzzy vault by Nandakumar et al. [98]. In this system, they applied a transformation to the biometric features before using them in the fuzzy vault construction.

Bringer et al. [28] also proposed a hybrid scheme for fingerprints. Their scheme applies cancelable transformation to fingerprints and then uses it in a fuzzy sketch. The fuzzy sketch used in this scheme is similar to the one in Bringer et al. [26].

Another example of a hybrid crypto-biometric system is face based template protection system by Feng et al. [48]. In this system, a cancelable transformation is applied on the biometric data followed by a discriminability preserving transform. The binary string obtained after these steps is protected by the fuzzy commitment scheme. The estimated entropy of the keys vary from 203 to 347 bits.

Boddeti et al. [20] proposed a key binding scheme using correlation filters on face images. They basically extended the correlation filters based cancelable biometric system approach of Savvides et al. [113] to obtain cryptographic keys. A password is used to generate a random kernel which is convolved with face images to make them revocable. The reported entropies, which vary depending on the system parameters, are 74-bits and 103-bits for two different settings.

A drawback of these hybrid systems from the verification performance point of view is that the key generation system performance degrades than the baseline system even though it involves an additional parameter (password or key).

Some other systems in this category can be found in [146, 144, 147, 91, 98, 132, 77, 149]. There are some works related to the security evaluation of template protection schemes which can be found in [87, 117, 13, 92, 121, 122]. Various attacks against the crypto-biometric systems can be found in these works, including attacks regarding our work (reported in [121, 122]). These attacks are based on the Error Correcting Codes (ECC) output statistics.

An interesting and emerging development in this research domain is to use multi-biometrics for cryptographic key regeneration. Though there are a large number of multi-biometric systems [107, 111, 108], very few proposals are found in literature which employ multi-biometrics for key regeneration. The reason behind this may be the fact that, in order to obtain high entropy keys, the multi-biometric fusion must be carried out such that the size of the biometric information being used in the crypto-biometric system will increase. In most of the crypto-biometric systems, this biometric information is a set of features (ordered or unordered), and therefore, feature level fusion is a convenient choice. But, the feature level fusion has its own difficulties such as the curse of dimensionality [111] and difference in feature representations of different modalities. The curse of dimensionality imposes limits on the number of features used in a pattern classification system, and hence, it generally needs to be followed by a feature selection process. Score level fusion, which is a popular way of information fusion in multi-biometric systems, is not possible in key regeneration systems because the reference biometric templates required to calculate individual scores cannot be stored in order to be compliant with the template protection scheme.

Sutcu et al. [123] proposed a method to combine fingerprint and face features in a fuzzy sketch scheme. But, they did not carry out real tests with the fused biometric information but rather predicted the results for the multi-biometric system from the two uni-biometric system results.

Nandakumar and Jain [96, 95] (in 2008) proposed a fuzzy vault scheme which combines fingerprints with iris. A significant improvement in verification performance is observed (e.g., from a GAR of 88% and 78.8% for individual iris and fingerprint systems, respectively, to 98.2% for the multi-biometric system). But, the total entropy of the multi-biometric vault (49-bit) is still low. Additionally, the security analysis provided in [95] does not consider the effect of addition of zeros to the iris codes. If zeros are added to the iris code, and error correction is applied to this data, the amount of errors in the biometric data corrected by the ECC is more than the error correction rate of the ECC. In fact, this concept is used in our work (Section 6.2) in order to increase the error correction capacity of Hadamard codes which is otherwise fixed. The increased error correction capacity must be given due attention during estimation of the key entropy. Considering these facts, an analysis of the multi-biometrics based scheme

in [95] decreases the entropy from 49 to 23 bits.

Recently, Fu et al. [50] proposed theoretical models describing multi-biometric cryptosystems. They proposed fusion at the biometric and cryptographic levels and then derived four models adopted at these two levels. However, this work is theoretical and no actual evaluation of verification performance as well as key entropy is carried out.

Combination of different techniques described above can also be employed. For example, Nagar et al. [93, 94] proposed a hybrid crypto-biometric system based on fingerprints. They combined techniques from fuzzy vault and fuzzy commitment in order to make the system more secure and increase the verification performance. Minutiae descriptors, which capture ridge orientation and frequency information in a minutia's neighborhood, are embedded in the vault construction using the fuzzy commitment scheme. Though this system performs better than the fuzzy fingerprint vault scheme [97], a lower performance of the proposed hybrid scheme compared to the baseline fingerprint system is reported.

2.3 Review of Biometrics Based Secure Cryptographic Protocols

The crypto-biometric systems described in the previous sections (Section 2.1 and 2.2) try to remove the limitations of the biometric and/or cryptographic systems. The cancelable biometric systems (Section 2.1.2) add revocability, template diversity, and privacy protection to the biometric systems. On the other hand, the systems described in Section 2.1 try to derive user specific cryptographic keys, the authenticity of which is confirmed with the help of biometrics. A strong link is established between the user's identity and his cryptographic keys when the cryptographic keys are derived from biometrics. If properly designed, these systems can also have the important properties of revocability, template diversity, and privacy protection.

As mentioned in Chapter 1, cryptography is used to secure the information during storage and/or transmission. Cryptography is broadly divided into two types: symmetric-key cryptography in which the encryption and decryption keys are the same, and public-key (also called asymmetric) cryptography where the encryption and de-

cryption keys are different but are mathematically related. In order to establish a cryptographically secure communication channel between all the entities participating in the information exchange, the correct cryptographic keys must be shared between them. Most of the crypto-biometric systems described in the previous section do not mention any specific key management/sharing methodologies. Those systems rely on conventional cryptography for the key sharing purpose and one side is still required to trust the other.

In this section, we present an overview of protocols that are specifically designed for sharing the crypto-bio keys or to create secure authenticated sessions based on biometrics. Note that we have not considered the protocols in which classical biometric comparison is used for authentication.

Boyen et al. [25] (in 2005) proposed a biometrics based remote authentication protocol in which the fuzzy extractors [24] are used. The problem with this protocol is that it stores the reference biometric template along with the protected crypto-biometric template. Though this reference biometric template is not shared, it can still be considered as a privacy compromise.

The one-time biometric authentication protocol proposed by Ueshige and Sakurai [136] (2006) creates biometric authentication based secure sessions but it requires storage of classical biometric templates. In this protocol, a one-time transformation is generated which is unique to the session. This transformation is applied to the stored templates as well as the fresh biometric data. The comparison of the two transformed templates is carried out to establish the authenticity of the subject.

Bringer et al. [27] (in 2007) employed the Goldwasser-Micali cryptosystem [52] for biometric authentication. This system allows the biometric comparison to be carried out in the encrypted domain. The proposed system requires storage of classical biometric templates. In order to protect the privacy, the system makes sure that the biometric data stored in the database cannot be explicitly linked to any user identity, but it only detects whether the data belonging to an identity is present in the database.

The "Secure Ad-hoc Pairing with Biometrics: SAfE" protocol proposed by Buhan et al. [30, 29] (2007) is a protocol which can be used to establish a secure link between two parties. Keys are obtained from biometrics with the help of the fuzzy extractor scheme. The drawback of this protocol is that it shares the biometric data

between the two parties and requires mutual trust among them. Moreover, it also requires a secure channel for exchanging the biometric data.

Tang et al. [125] proposed an authentication protocol based on fuzzy extractor. This protocol provides security by employing the ElGamal public-key cryptosystem [45]. Cryptographic keys can be obtained and shared with this protocol while preserving some aspects of the user privacy. The drawback of this proposal is that it requires storage of biometric templates in a database. Additionally, it needs a secure communication link between the parties for exchanging the information.

Recently Barni et al. [15] proposed a scheme for privacy preserving authentication based on fingerprints. This scheme employs the ElGamal cryptosystem which facilitates biometric comparison in encrypted domain. However, this scheme can only be used for authentication and not for key sharing.

Abid and Afifi [8] (in 2009) proposed a protocol for ePassport authentication based on elliptic curve cryptography. They proposed to employ biometrics, specifically fingerprints, to securely generate the parameters of the elliptic curve. These parameters are used for the ePassport bearer's authentication. This proposal is theoretical and no experimental evaluation is reported. The difficulty of this approach is that it requires a stable input from biometrics.

In [9], we integrated our work on iris based cryptographic key regeneration [65] with the Abid and Afifi scheme [8]. In this proposal, first a stable key is obtained from biometrics which is used to obtain the security parameters of the elliptic curve. This proposal can help integrate crypto-biometrics in ePassports. This system is described in Chapter 8.

The drawback of the authentication protocols in [136, 27] is that they can only authenticate the subject. But they cannot produce the cryptographic keys required for secure communication. The protocols in [25, 125, 8] can share keys but these keys are the same for all the sessions. Using the same key for encryption of a large amount of data can make some cryptanalytic attacks easier. Therefore, most of the practical systems, e.g., the Transport Layer Security (TLS) protocol [42], employ a session specific symmetric key for secure communication. The session key is temporarily generated in every session. Public-key cryptographic protocols are used to share this key.

Following the same concept of session keys, Scheirer and Boult [115, 116] (in

2008) proposed "bipartite biotokens". They combined their earlier proposal of revocable biotokens [23] with fuzzy vaults [63] which enables to securely share keys using biometrics. In this scheme, a series of transformations is shared between the client and the server. A new transformation (in succession) is applied in every communication session. The bipartite biotokens are session specific and make it possible to share session specific data between two parties.

2.4 Conclusion and Discussion

A number of systems, denoted as crypto-biometric systems, are found in literature that combine biometrics with cryptography in order to remove the artifacts present in the two techniques. A comprehensive review of such crypto-biometric systems is presented in this chapter. The review is presented such that the systems are classified according to their purpose and their basic working methodology. Clear distinction is made between the various classes along with their representative schematic diagrams. Moreover, a review of crypto-bio key sharing protocols is also included.

The comparison (from verification performance point of view) of various crypto-biometric systems is a critical issue. This comparison involves systems developed for different modalities, on various databases. In order to have a common ground for comparison, we propose to compare the performance of the crypto-biometric systems with the respective baseline biometric systems employed in them. Moreover, if more than one authenticators are involved in the system, the performance in case of compromise of one of those factors should also be reported. Additionally, the entropy of the keys obtained by the crypto-biometric systems should also be considered as an important parameter for comparing different systems.

Table 2.1: Summary of cancelable biometric systems; The verification performances are reported in terms of FAR, FRR, and EER in %.

Ref	Technique	Database	Results	Stolen key	Remarks
[104, 105]	Various one-way transformations	Proprietary, 188 x 2 fingerprint images	Verification rate 5% less than baseline system	-	-
[62]	BioHashing	Fingerprint, FVC2002-DB1	EER = 0%	-	-
[113]	Cancelable filters	CMU-PIE face	100% verification result; same as baseline	-	-
[127]	BioHashing	FERET face	Better than baseline system	Degrades than baseline system	-
[78]	BioHashing	Fingerprint-FVC2002, face-ORL and Yale-B, signature-SUBCORPUS-100 MCYT	Better than baseline system	Degrades than baseline	-
[47]	Transformation	Proprietary database; 1000 fingerprints	EER = 0%; better than baseline	Same as baseline system	-
[23]	Revocable biotokens	Fingerprints; FVC2000,2,4	Nearly 30% better than baseline	-	-
[80]	Transformation	MCYT signature	EER =13.30%; EER baseline=10.29%	-	-
[83]	Transformation	Proprietary; signature	EER=14%; EER baseline=11%	-	-
–	Shuffling	Iris; NIST-ICE	EER=0.23%; EER baseline=1.71%	Same as baseline system	This work; Chapter 4
[67]	ECC & Shuffling	Iris; NIST-ICE	EER=0.057%; EER baseline=1.71%	Same as baseline system	This work; Chapter 5
Advantages: Add revocability, template diversity, and privacy protection to biometrics					
Limitations: The verification result is one-bit (yes/no). Therefore, they suffer from the biometric bottle-neck problem.					

Table 2.2: Summary of biometrics based cryptographic key generation systems; The verification performances are reported in terms of FAR, FRR, and EER in %.

Ref	Technique	Database	Results	Entropy	Stolen key	Remarks
[39, 40]	ECC, majority coding	-	Theoretical	-	-	-
[90]	Password hardening	Keystroke dynamics; empirical analysis	FRR = 48.4%	12-bit	-	-
[142]	Statistical features	Online signatures; empirical tests	FAR = 0% at FRR = 7.05%	-	-	Small data set
[51]	BioHashing; Shamir-secret sharing	Spacek's Faces94	Better than baseline system	-	-	-
[89]	Password hardening	Voice; proprietary database	FRR = 20%	46-bit	-	-
[118]	Fuzzy genetic clustering	Signatures; Proprietary database	FRR = 13.1% at FAR = 0%	20-bit	-	
[114]	Biohashing	Online handwriting; proprietary	FRR = 23% at FAR = 2.39%	-	-	Performance degrades than baseline
[12]	ECC	Gait; proprietary database from HUMABIO project and CASIA	Better than baseline system	-	-	-
Advantages: It can be designed to be template free system, hence no information leakage about the biometric data.						
Limitations: Difficult to design, low performance. If no additional parameter (such as password) is used, the system cannot be revocable. The key is not generated randomly.						

Table 2.3: Summary of biometrics based cryptographic key regeneration systems; The verification performances are reported in terms of FAR, FRR, and EER in %.

Ref	Technique	Database	Results	Entropy	Stolen key	Remarks
[64]	Fuzzy commitment	-	Theoretical	-	-	-
[120]	Signal processing	Fingerprints	-	-	-	-
[63]	Fuzzy vault	-	Theoretical	-	-	-
[36]	Fuzzy vault	Fingerprints, empirical study	30% FRR	69-bit	-	-
[44]	Fuzzy extractors	-	Theoretical	-	-	-
[137]	Fuzzy vault	Fingerprints; FVC2002, DB2	20% FRR	-	-	-
[54]	Fuzzy commitment	Iris; proprietary	FRR = 0.47% at FAR = 0%	44-bit	-	Good quality database
[26]	Fuzzy commitment	Iris; NIST-ICE	FRR = 5.62% at FAR = 10^{-5}	-	-	-
[97]	Fuzzy vault	Fingerprints; FVC2002–DB2† and MSU-DBI‡	9%† and 15%‡ FRR	-	-	-
[28]	Cancelable and then fuzzy sketch	Fingerprints; FVC2000	FRR = 3% at FAR = 5.53%	-	-	Degraded performance
[96, 95]	Fuzzy commitment and fuzzy vault	Fingerprints (MSU-DBI) and iris (CASIAv1)	FRR = 1.8%	49-bit	-	Our entropy estimate ≈ 23 bits
[93, 94]	Fuzzy commitment and fuzzy vault	FVC2002-DB2	FAR = 0.01% at FRR = 5%	49-bit	-	Our entropy estimate ≈ 36 bits
[82]	User adaptive fuzzy commitment	Online signatures; private database	EER = 17%	-	-	Improved performance
[20]	Key binding using cancelable filters	Face; CMU-PIE database	FRR = 1.4% at FAR = 0%	74 to 103 bits	-	Degraded performance
[65]	Fuzzy commitment and shuffling	Iris; NIST-ICE	FRR = 1.04% at FAR = 0.055%	83-bit	FAR = 14.06%	This work; Improved performance; Chapter 6
[68]	Fuzzy commitment and shuffling	Iris NIST-ICE and face NIST-FRGCv2	FRR = 0.91% at FAR = 0% FAR	183 bit	FAR = 36.43%	This work; Improved performance; Chapter 7
Advantages: if properly designed, these systems can have the properties of revocability, template diversity, and privacy protection; the keys are randomly generated; high entropy keys can be obtained.						
Limitations: The verification performance generally degrades compared to that of the baseline biometric system.						

Chapter 3

Performance Evaluation Strategies of Crypto-biometric Systems

In the previous chapter, we have presented a through review of crypto-biometric systems. In this thesis, we are going to propose some novel crypto-biometric systems. The systems that we are going to propose need a biometric system in order to extract features from the biometric data. Additionally, our crypto-biometric systems require the biometric features in binary format. These biometric systems are briefly described in Appendix A. In this work, our focus was on developing crypto-biometric systems. Therefore, we relied upon open-source baseline biometric systems. We have developed crypto-biometric systems for iris and face modalities along with a multi-modal system combining the two. The biometric databases along with their associated experimental protocols are described in Appendix A.

We have followed a strategy that the development and evaluation data sets have zero overlap. Different parameters, such as error correction capacities, decision thresholds, etc., are tuned on the development set and with those parameters, the performance is evaluated on the evaluation data sets.

In this chapter, we define various metrics to measure the verification performance of the biometric as well as crypto-biometric systems. The performance metrics for biometric systems are discussed in Section 3.1 while those for the crypto-biometric sys-

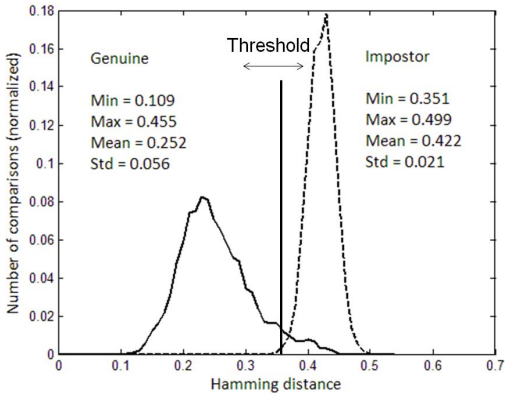


Figure 3.1: An example of Hamming distance distribution plots. The threshold is decided such that the number of genuine Hamming distances above the threshold and the number of impostor Hamming distances below it are minimum.

tems are discussed in Section 3.2. Additional strategies required for security evaluation of the crypto-biometric systems is given in Section 3.3.

3.1 Performance Evaluation of Biometric Systems

The biometric systems used in this work generate binary feature vectors. The binary feature vectors obtained from the reference and test biometric samples are compared using the Hamming distance. In general, if the two feature vectors are extracted from images belonging to the same person (genuine comparison), the Hamming distance between them is close to zero. On the other hand, if they are from different persons (impostor comparison), then the feature vectors are random and differ by a Hamming distance of nearly 0.5. An example of a plot of genuine and impostor Hamming distance distributions is shown in Fig. 3.1.

Comparison of the Hamming distance with a threshold τ leads to an accept/reject decision. There can be two errors in the decision process: *false acceptance* when an impostor is accepted and *false rejection* when a genuine user is rejected by the system. Based on these two errors, following verification error rates are defined:

- False Acceptance Rate (FAR): It is the probability that a non-authorized person is accepted as authorized. It is calculated as a ratio of the number of impostor attempts successfully accepted by the system to the total number of impostor

trials.

$$\text{False Acceptance Rate (FAR)} = \frac{\text{Number of impostor attepts accepted}}{\text{Total number impostor trials}}. \quad (3.1)$$

- Genuine Acceptance Rate (GAR): It is the probability that an authorized person is successfully accepted. It is calculated as a ratio of the number of genuine attempts successfully accepted by the system to the total number of genuine trials. It is equal to 1-FRR.

$$\text{Genuine Acceptance Rate (GAR)} = \frac{\text{Number of genuine attepts accepted}}{\text{Total number genuine trials}}. \quad (3.2)$$

- False Rejection Rate (FRR): It is the probability that an authorized person is rejected access. It is the ratio of the number of genuine verification attempts rejected by the system to the total number of genuine trials. It is equal to 1-GAR.

$$\text{False Rejection Rate (FRR)} = \frac{\text{Number of genuine attepts rejected}}{\text{Total number of genuine trials}}. \quad (3.3)$$

A plot of FRR (or GAR) as a function of FAR is denoted as Receiver Operating Characteristic (ROC) [18]. If the axes of such plot are on the normal deviate scale, this plot is denoted as Detection Error Trade-off (DET) [85]. The performance of a biometric system can also be summarized with a single number, where the FAR=FRR. this point is denoted as Equal Error Rate (EER).

The Hamming distance distribution plots (such as the one shown in Fig. 3.1) show the user discrimination capability of the biometric system. If the genuine and impostor distributions are well separated, then the threshold can be set in between the two curves. In this case, there will be zero FAR and FRR. If an overlap between the two curves exists, it results in FAR and FRR.

In order to calculate the verification error rates and confidence intervals, we used the Biosecure Performance Evaluation Tool [5], [103]. This tool employs the parametric confidence interval estimation procedure described by Bolle et al. [21]. The tool calculates the values of EER, operating point (FRR at a given value of FAR) along with their 90% confidence intervals from the similarity scores for genuine and impostor comparisons.

3.2 Performance Evaluation of Crypto-biometric Systems

In our proposed crypto-biometric systems, the biometric data are combined with a shuffling key which is obtained using a password. The shuffled data is compared using the same matcher used in the classical biometric system. Every user is assigned a unique and secret shuffling key which ideally is known only to the user. For all the tests, the data pair, consisting of biometric data and the shuffling key, acts as a unit. When a user presents his biometric data for verification, he also provides his unique shuffling key. This implies that, in case of genuine users, the key provided during verification is the same as that used during enrollment of that user, while the biometric data may contain intra-user variability. In case of random impostor tests, an impostor provides his biometric data along with his key.

The verification performance of the crypto-biometric system largely depends on the baseline biometric system it is based upon. From the literature review presented in Chapter 2, we found out that, there are systems for which the performance degrades, improves, or remains unchanged. Therefore, it is important to study the change in performance of the system when the biometric information is combined with another secret. In general, a verification system that combines biometric information with other assigned secret, should have better performance than the baseline biometric system because the user is required to carry (e.g., a token) or remember (e.g., a password) another authenticator in addition to the biometrics.

In case of cancelable biometric systems, a similarity (or dissimilarity) score is generated and the verification decision is in form of accept/reject. Therefore, the Biosecure Performance Evaluation Tool [5], [103] can be used for estimating the FAR, FRR, and EER values along with the confidence intervals. In case of key (re)generation systems, the system does not yield a score. Therefore, the Biosecure Performance Evaluation Tool cannot be used to calculate the confidence intervals. Other methods, such as making multiple partitions of the evaluation data and using the mean and standard deviation of the performance on these partitions can be employed in such cases.

Scheidat et al. [114] proposed two measures for calculating verification performance of key generation systems. These measures are Collision Rate and Reproducibility Rate. But in fact, these two measures are just another names for False Acceptance Rate

(FAR) and Genuine Acceptance Rate (GAR) applied to the key generation systems. When the hashes (cryptographic keys in this case) generated from biometric samples of two different users are the same, they call it as collision (which is generally denoted as false acceptance). Collision Rate is the ratio of total number of collisions to the total number of such comparisons. Reproducibility rate is the ratio of number of successful hash generation attempts to the number of comparisons. Since these terms have the same meaning as the conventional terms (FAR and GAR), we use the conventional terms.

3.3 Security Evaluation of Crypto-biometric Systems

The crypto-biometric systems are supposed to increase the security, and therefore, it is required to carry out theoretical as well as experimental security analysis of such systems. Since the systems proposed in this work have two factors: biometric and shuffling key, it is necessary to evaluate the system performance when one of the two factors is compromised. Hence, two security scenarios are considered: stolen biometric: when the biometric data for all the users are compromised; and stolen key: when the shuffling keys of all the users are compromised. So, in the stolen biometric scenario, we assume that the impostor has the biometric data of the user. The information he does not have is the shuffling key. Hence, he tries to get verified by providing the stolen biometric data along with his shuffling key.

In the stolen key scenario, the assumption is made that the shuffling keys for all the users are known to the impostors. The other factor - biometric data - is assumed to be secret. In this case, the impostor provides his biometric data along with the stolen shuffling key of the claimed identity.

These two security scenarios are two extreme hypothetical situations, assuming compromise of one of the two factors for all the users. Since the secret information in these scenarios is less than that in the ideal case (where both the factors are secret), the performance of the system is bound to degrade. The question we investigate is that whether the performance degrades beyond the baseline biometric system or not.

Moreover, for the crypto-biometric systems which are used to obtain crypto-bio keys, the length of the key does not really indicate the level of security offered by

the system. Though the lengths of the keys can be quite high, the entropy significantly reduces due to the redundancy added by the ECC. A theoretical estimation of the entropy of the key must be carried out by taking into considerations the redundancy added by the ECC. In this thesis, we estimate the entropy of the crypto-biometric keys against brute force attacks. We followed the entropy estimation methodology given by Hao et al. [54]. In order to estimate entropy with this approach, the number of degrees of freedom in the biometric data which is used in the key regeneration system must be calculated first. It is done as follows:

Let μ and σ be the mean and standard deviation of the binomial distribution fitting the impostor Hamming distance distribution. Then the number of degrees of freedom is estimated to be:

$$N = \mu(1-\mu)/\sigma^2. \tag{3.4}$$

Hao et al. [54] use sphere packing bound [79] to estimate the number of brute force attempts required to guess the key. It is based on the degrees of freedom and the error correction capacity of the system. Let's consider that the number of degrees of freedom is N and the error correction capacity is e. The fraction of N corresponding to the error correction capacity e is P. Then the number of brute force attempts required is:

$$\begin{aligned} BF &\approx \frac{2^N}{\binom{N}{P}}, \quad \text{and} \\ \text{Entropy} \quad H &= log_2(BF). \end{aligned} \tag{3.5}$$

This method is frequently used, wherever applicable, in the forthcoming chapters.

3.4 Template Diversity Test

In order to prove that a cancelable biometric system adds template diversity, we propose a specific test. In this test, one biometric feature vector is transformed with 100,001 (or even more) transformation parameters. This results in 100,001 different templates from a single feature vector. The first such cancelable template is compared with the remaining 100,000 cancelable templates.

If the Hamming distance (or whichever distance is applicable) distribution of these comparisons is close to the impostor Hamming distance distribution, it indicates that a large number of independent templates can be obtained from a single biometric feature vector using the cancelable biometric system in consideration.

3.5 Summary

In this chapter, the various performance metrics, with which, the verification performances of the biometrics as well as crypto-biometric systems are measured, are defined. The performance evaluation strategy applied for evaluation of the biometrics as well as crypto-biometric systems is also developed. Particularly, two security scenarios are defined: stolen biometric and stolen key. The crypto-biometric systems proposed in forthcoming chapters are evaluated using this experimental performance evaluation strategy.

The forthcoming chapters will discuss various crypto-biometric systems proposed in this thesis. The information provided in this chapter is required for a detailed understanding and analysis of the proposed systems.

Chapter 4

Cancelable Biometric System

With the increasing use of biometrics, more and more concerns are being raised about the privacy of biometric data. In the existing biometric systems that we denote as 'classical biometric systems', the information needed for further comparisons, denoted as biometric reference or template, is stored in a database. This information remains substantially similar across databases if the modality and the biometric algorithm are the same, e.g., for minutiae based fingerprint systems, minutiae sets extracted from the same fingerprint in different systems are similar. If such template is compromised, it is not possible to replace it with a new one because the biometric characteristics (from which this information is extracted) are permanently associated with their owners. In other words, it is not possible to revoke or cancel a template. This phenomenon is called as lack of revocability.

The permanent association of biometric data with the user leads to another problem. Since the templates in all the systems based on the same biometric characteristic and using same biometric algorithms are similar, a compromised template from one biometric database can be used to access information from another system. This can be referred to as cross-matching between databases. This can be considered as a threat to privacy. Moreover, in some cases, the stored information can be used to create a dummy representation of the biometric trait which can be used to access the system [10, 32, 109, 33]. For example, a dummy finger can be constructed from a fingerprint image.

Because of these reasons, the property of cancelability or revocability is becom-

ing a necessity. In order to induce revocability into biometric systems, cryptographic techniques are a good candidate. Many systems that induce these characteristics are proposed in literature. A review of these systems in presented in Section 2.1.2 (page–20). In this chapter, we propose a simple shuffling scheme to create cancelable templates from biometric data. This scheme involves two factors: biometrics and a shuffling key. Because of this additional parameter, the proposed scheme significantly improves the verification performance of the baseline biometric system. A distinct advantage of this scheme is that its performance in stolen key scenario remains equivalent to that of the baseline biometric system.

This chapter is organized as follows: the proposed shuffling based cancelable biometrics scheme along with its advantages is described in Section 4.1. Verification performance of this scheme is carried out on iris and face modalities. Details about the evaluation are presented in Section 4.2. Section 4.3 sets out conclusions and perspectives.

4.1 A Biometric Data Shuffling Scheme to Create Cancelable Biometric Templates

The shuffling scheme described in this section was first applied to iris biometrics and then to face biometrics. In general, it can work with any biometric modality provided the biometric features are represented as an ordered set. In this scheme, a randomly generated shuffling key is used to shuffle the biometric data. The shuffled biometric data represents the cancelable template. It is not feasible to recover the original biometric data from this cancelable template. This scheme can be considered analogous to classical symmetric encryption technique because, as in encryption, a key is used to protect the biometric data. But contrary to classical encryption, the user discrimination properties of biometric data are retained by the transformed data, and hence, comparison between two such transformed biometric data can be carried out in the transformed domain. The shuffling technique is explained in details in the next subsection.

4.1.1 The Proposed Shuffling Technique

The shuffling scheme that we introduce requires a binary shuffling key $\mathbf{K}_{sh}$ of length L_{sh}. Since this key is a long bit-string, it is stored on a secure token or it

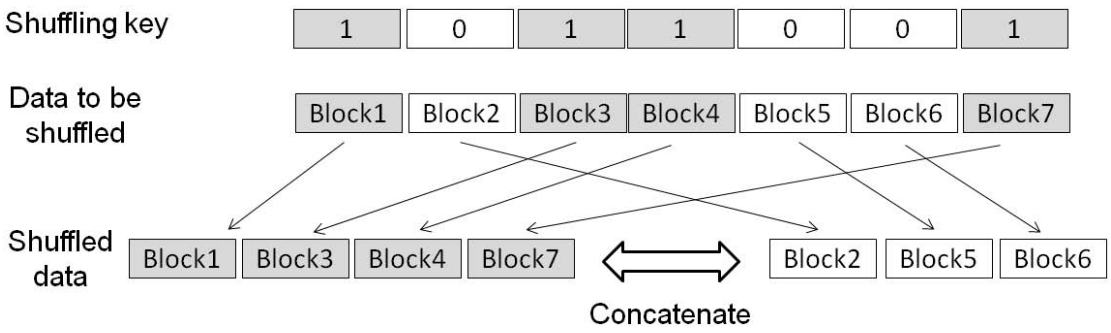


Figure 4.1: The proposed shuffling scheme.

can be obtained using a password. The biometric feature vector is divided into L_{sh} blocks each of which has the same length. To start the shuffling, these L_{sh} blocks of the feature vector are aligned with the L_{sh} bits of the shuffling key $\mathbf{K}_{sh}$. In the next step, two distinct parts containing biometric features are created: the first part comprises all the blocks corresponding to the positions where the shuffling key bit value is one. All the remaining blocks are taken into the second part. These two parts are concatenated to form the shuffled biometric feature vector which is treated as a revocable template. Figure 4.1 shows a schematic diagram of this shuffling scheme.

The original and shuffled feature vectors have one-to-one correspondence. A block from the original vector is placed at a different position in the shuffled vector. Thus, only the alignment of the feature blocks is changed by the scheme with no change in the actual values of the features. The length of the biometric feature vector does not change because of the shuffling. Hence, the matching algorithms used for calculating the similarity (or dis-similarity) score between two biometric feature vectors are still applicable for the shuffled data.

Note that the effectiveness of this scheme is because it changes the alignment of the feature vectors. If the feature vectors do not require any particular order (e.g., fingerprint minutiae sets), this system is ineffective. This system can work only if the biometric data is in form of an ordered set.

4.1.2 Advantages of Using the Proposed Shuffling Scheme

The proposed shuffling scheme has the following advantages:

1. **Revocability:** The shuffled feature vector, which is treated as a cancelable template, is a result of combination of an intrinsic identifier (i.e., a biometric characteristic) and an assigned identifier (the shuffling key). Therefore, in case of compromise, it can be canceled and a new template can be generated by changing the shuffling key $\mathbf{K}_{sh}$ (the assigned credential).

2. **Performance improvement:** Another advantage of using the shuffling scheme is that it improves the verification performance. The shuffling process changes the alignment of the feature vector blocks according to the shuffling key. When two biometric feature vectors are shuffled using the same shuffling key, the absolute positions of the feature vector blocks change but this change occurs in the same way for both of the biometric feature vectors. Hence, the Hamming distance (in case of binary vectors) between them does not change. On the other hand, if they are shuffled using two different keys, the result is randomization of the feature vectors and the Hamming distance increases. In fact, the shuffling process acts like a randomizer and moves the average Hamming distance for such cases close to 0.5.

   A unique shuffling key is assigned to each subject during enrollment and he has to provide that same key during every subsequent verification. This means, in ideal case, that the genuine users always provide the correct shuffling key and hence, the Hamming distance for genuine comparisons remain unchanged. On the contrary, in case of random impostor attempts where an impostor tries to get verified with his own credentials, he provides his biometric data along with his shuffling key (or a random shuffling key) to match against other users. The feature vectors for such impostor comparisons are shuffled with two different shuffling keys and the result is that the Hamming distances increase. This effect can be seen in Fig. 4.2. The separation between the genuine and impostor Hamming distance distributions shows the ability of the system to distinguish genuine users from impostors. As can be seen from Fig. 4.2, shuffling increases the separation between the two distributions. In this way, the shuffling scheme improves the verification performance of the system.

3. **Template diversity:** With the help of the shuffling technique, different templates

can be issued for different applications by using different shuffling keys with the same biometric data. This particularly helps to avoid cross-database matching. In order to make the template-diversity effective, it is suggested that the shuffling key should be generated randomly and protected by a password.

4. **Protection against stolen biometric data:** If a feature vector is shuffled using two different shuffling keys, the resulting shuffled vectors appear to be originating from two different subjects. They can be seen as comparing two random sequences and hence they do not match. Therefore, if a stolen biometric data of a legitimate person is used by an impostor to get verified, the system can still resist such attack due to the use of shuffling key.

5. **Biometric data protection:** It is not computationally feasible to recover the original biometric feature vector from the shuffled data without the proper shuffling key. However, as in classical encryption, the security depends on the secrecy of the shuffling key.

These effects can be better understood from the experimental results and analysis presented in the next section.

4.2 Experimental Results and Security Analysis of the Proposed Cancelable Biometrics Scheme

The cancelable biometric system is based upon an underlying baseline biometric system. Therefore, for fair comparison, first the biometric verification performance of the baseline biometric system is reported followed by the performance of the proposed cancelable system.

The proposed cancelable biometric system is evaluated on two biometric modalities: iris and face. For iris, the CBS database [103] is used for development and the NIST-ICE database [101] is used for evaluation purposes. For face, the development and evaluation data sets are derived from the NIST-FRGCv2 database [100]. Details about these databases along with the associated experimental protocols are given in Appendix A. The experimental evaluations of the proposed cancelable system on iris and face modalities are given in the following subsections.

4.2.1 Results and Security Analysis on Iris Modality

Experimental Setup

The iris databases and the associated experimental protocols are described in details in Section A.2.1 (page–186).

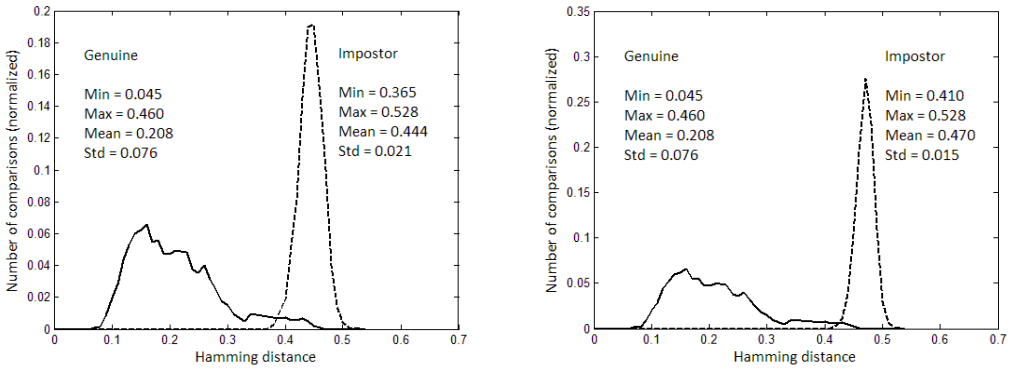
In case of iris modality, the development database, the CBS database, has two parts: CBS-BiosecureV1 and CBS-CasiaV2. On each of these parts, we carried out 6,000 genuine and 6,000 impostor comparisons. The experimental protocol is designed such that it allows comparisons between images obtained in different sessions and different illumination conditions, and between images of eyes with and without glasses.

The proposed cancelable biometric system is then evaluated on the NIST-ICE iris database. The parameters tuned on the development database (in case of cancelable system, this parameter is the length of the shuffling key) are used for evaluation. The shuffling key is the assigned parameter in the cancelable biometric system which provides protection to the biometric data. Therefore, the length of this key should be long enough from security point of view to avoid brute force attack.

There are two separate experiments defined (and commonly used in the research community) for the NIST-ICE database: ICE-Exp1 consisting of comparisons of right eye images, and ICE-Exp2 which consists of left eye image comparisons. All possible comparisons are carried out for genuine as well as impostors. In total, 12,214 genuine and 1,002,386 impostor comparisons are carried out in ICE-Exp1, whereas in ICE-exp2, 14,653 genuine, and 1,151,975 impostor comparisons are performed.

Results on Iris Modality

The genuine and impostor Hamming distance distributions for the CBS-BiosecureV1 data set before and after shuffling are shown in Fig. 4.2. As described in Section 4.1.2, the shuffling process increases the impostor Hamming distances while the genuine Hamming distances remain unchanged. This can be seen from the Fig. 4.2. In this figure, the mean of the impostor Hamming distance distribution of the baseline system shifts from 0.44 to 0.47 when the shuffling scheme is applied. Note that, the genuine Hamming distance remains unchanged. This reduces the overlap between the genuine and impostor distribution curves which improves the user discrimination capacity of the



(a) Baseline iris biometric system

(b) Proposed cancelable biometric system (baseline iris system with shuffling)

Figure 4.2: Normalized Hamming distance distributions for genuine and impostor comparisons on the CBS-BioSecureV1 [103] development data set.

system thereby increasing the verification accuracy. The Hamming distance curves for the CBS-CasiaV2 data set are shown in Fig. C.1.

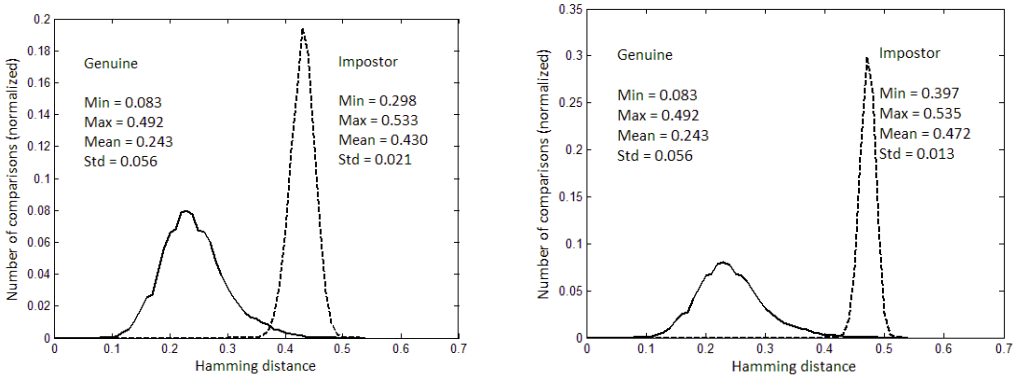
The better separation between genuine and impostor Hamming distance distribution curves improves the verification performance of the system. The verification performance in terms of Equal Error Rate (EER) on the development database (CBS database) is reported in Table 4.1. A clear improvement in performance can be seen by comparing the EER of the baseline system with the proposed cancelable system. For example, on the CBS-BiosecureV1 data set, the EER for the baseline system is 2.63% which reduces to 0.93% when the cancelable scheme is applied. Similarly, on the CBS-CasiaV2 data set, the EER reduces from 3.03% to 0.56% because of the shuffling scheme. For the sake of comparison, the EER values reported in the documentation of the OSIRISv1 on these two data sets are also reported in this table.[1]

The proposed shuffling based cancelable biometrics scheme is then evaluated on the NIST-ICE database. As noted before, we carried out separate experiments according to the common protocol for ICE evaluation for right (ICE-Exp1) and left (ICE-Exp2) iris comparisons. The Hamming distance distributions for the ICE-Exp1 experiment are shown in Fig. 4.3.

[1]The baseline iris system is based on OSIRISv1; the difference is that the matching module is re-implemented to cope with the iris rotations.

Table 4.1: Verification results of the baseline biometric system (which is based on the OSIRISv1) and the proposed cancelable system on iris modality; development data sets (CBS database [103]); in terms of EER in %. Values in bracket indicate the error margins for 90% confidence intervals.

Experiment	CBS-BiosecureV1	CBS-CasiaV2
Baseline	2.63[±0.34]	3.03[±0.36]
Proposed cancelable	0.93[±0.20]	0.56[±0.16]
OSIRISv1 [103]	2.83[±0.35]	2.12[±0.31]



(a) Baseline iris biometric system (ICE-Exp1)

(b) Proposed cancelable biometric system (baseline iris system with shuffling)

Figure 4.3: Normalized Hamming distance distributions for genuine and impostor comparisons on the NIST-ICE [101] evaluation database, for ICE-Exp1 (right-eye experiment).

Similar to the experiments on development sets, a better separation between genuine and impostor Hamming distance curves is seen on the NIST-ICE evaluation data sets. The improvement caused by such change in the Hamming distance distributions is evident from the results given in Table 4.2. The Equal Error Rate (EER) of the system decreases considerably because of the application of shuffling. e.g., on the evaluation database for iris modality, the NIST-ICE database, the EER reduces from 1.71% for the baseline system to 0.23% for the proposed cancelable system for the ICE-Exp1. Similarly, for ICE-Exp2, the EER reduces from 1.80% to 0.37%. Thus, there is nearly 80% reduction in the EER of the cancelable system when compared to the baseline system.

Table 4.2: Verification results of the baseline biometric system (which is based on the OSIRISv1) and the proposed cancelable system on iris modality; evaluation database (NIST-ICE [101]); in terms of EER in %. Values in bracket indicate the error margins for 90% confidence intervals.

Experiment	ICE-Exp1	ICE-Exp2
Baseline	1.71[±0.11]	1.80[±0.10]
Proposed cancelable	0.23[±0.04]	0.37[±0.05]
OSIRISv1 [103]	1.52[±0.12]	1.71[±0.12]

Security Analysis of the Proposed System on Iris Modality

The cancelable biometric system proposed in this chapter has two factors: biometrics and a shuffling key. In order to test the robustness of the system, as described in Section 3.3, we carried out the performance evaluation in two extreme hypothetical impostor scenarios: (i) stolen biometric and (ii) stolen key.

In the stolen biometric scenario, we consider a hypothetical extreme situation when the biometric information for all the users is stolen. Here, an impostor will try to provide the stolen biometric data along with a wrong shuffling key. In this situation, the EER increases compared to that of the cancelable system with both factors secret. But, it is still less than the EER of the baseline biometric system. For example, as shown in Table 4.3, for ICE-Exp1, the EER of the cancelable system is 0.23% when both the factors are secret. Considering that the iris image is stolen for all the users, the EER increases to 0.27% which is still less than the EER for baseline system (1.71%). Thus, use of the shuffling scheme prevents the impostors from being successfully verified using stolen biometric data.

Table 4.3: Security analysis of the proposed cancelable system on iris modality in terms of EER in %. Two scenarios are considered: (i) stolen biometric and (ii) stolen key. Values in bracket indicate the error margins for 90% confidence intervals.

Test	Development data set		Evaluation data set	
	CBS-BiosecureV1	CBS-CasiaV2	ICE-Exp1	ICE-Exp2
Baseline	2.63[±0.34]	3.03[±0.36]	1.71[±0.11]	1.80[±0.10]
Cancelable	0.93[±0.20]	0.56[±0.16]	0.23[±0.04]	0.37[±0.05]
Stolen biometric	1.50[±0.26]	0.71[±0.18]	0.27[±0.08]	0.44[±0.09]
Stolen key	2.63[±0.34]	3.03[±0.36]	1.71[±0.11]	1.80[±0.10]

In the stolen key scenario, we consider another extreme situation when the

shuffling keys of all the users are compromised. As in the stolen biometric scenario, the EER increases compared to that of the cancelable system having both parameters secret. But, the EER is equal to the EER of the baseline biometric system meaning that the system in this stolen key scenario is still as good as the baseline biometric system (see Table 4.3). In fact, the proposed shuffling scheme is such that, it increases the Hamming distance between two iris codes if and only if they are shuffled with different keys. If the same key is used to shuffle two codes, the Hamming distance remains intact. Thus in the stolen key scenario, the Hamming distance distribution is exactly the same as that for the baseline system, and hence, yields the same result as that of the baseline biometric system. This is a distinct advantage of our system over other cancelable systems found in literature. For most of the cancelable systems found in literature, the performance degrades if the keys (or the cancelable parameters used) are compromised. Only the Farooq et al. [47] system is shown to have the performance equal to the baseline biometric system in the stolen key scenario. See the Table 2.1 for a detailed comparison.

Detection Error Tradeoff (DET) curves for the proposed cancelable system along with the security threats are shown in Fig. 4.4 for the iris modality. These curves show the performance on the evaluation database – the NIST-ICE database – for the ICE-Exp1 experiment. The DET curves for the baseline system and that for the stolen key scenario overlap with each other which indicates that the performance of the system in stolen key scenario is same as the baseline system.

The stolen biometric scenario also proves the template diversity concept. It shows that, if the biometric feature vector is shuffled with two different keys, the two shuffled codes appear to be random. The impostor Hamming distance distributions for the random impostor case (when both, biometric data and key are secret), stolen biometric scenario, and the stolen key scenario are shown in Fig. 4.5 for the iris modality on the NIST-ICE database. Clearly, the distribution for stolen biometric scenario, which is obtained by comparing shuffled iris codes of the same users shuffled with different shuffling keys, lies near the random impostor distribution. This indicates that two iris codes of the same user shuffled using two different shuffling keys, are as different as two shuffled iris codes of two random impostor. Thus, by changing the shuffling key, different templates can be issued for the same user.

We carried out an additional test to prove that the proposed shuffling based

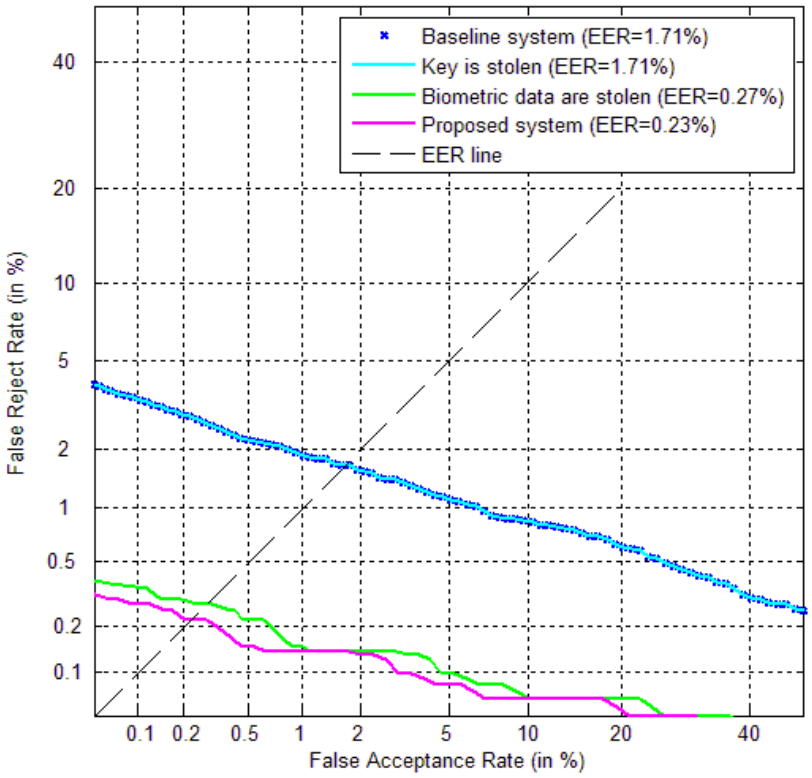


Figure 4.4: DET curves for the proposed system performance along with the possible security threats for iris modality on the NIST-ICE database (evaluation data set) [101]; ICE-Exp1.

cancelable biometric system adds template diversity. We shuffled one iris code with 100,001 randomly generated shuffling keys. The first shuffled iris code is compared with the remaining 100,000 shuffled iris codes. The distribution of Hamming distances obtained from these comparisons is shown in Fig. 4.6. This distribution is also close to the random impostor distribution which validates our claim of template diversity.

In case of compromise, the cancelable template can be revoked. In order to revoke the template, the user is asked to re-enroll into the system. The fresh biometric data is shuffled with a newly generated random shuffling key. Since this shuffling key is different than the one used in earlier enrollment, the old template and the newly issued template cannot match with each other. If an attacker obtains an iris code of the user from previously compromised template or from another biometric system, that iris code cannot be used by the impostor to get verified because the new shuffling key resists such attacks.

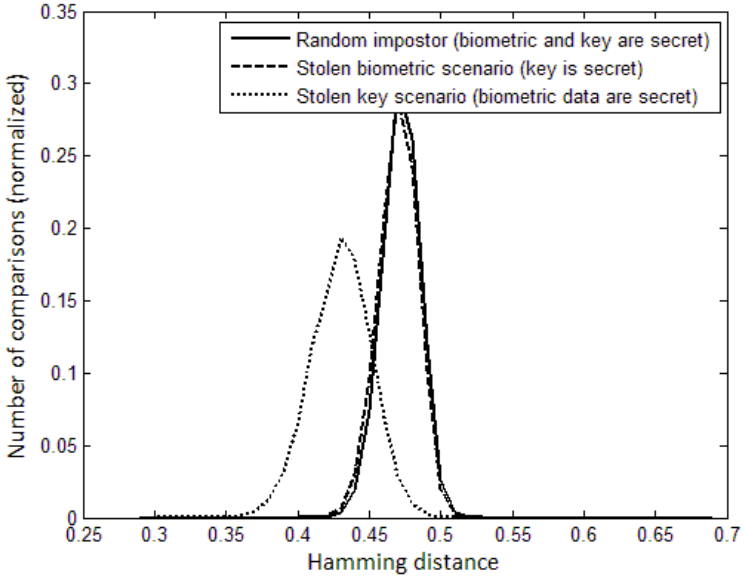


Figure 4.5: Impostor Hamming distance distributions for the proposed system along with the possible security threats for iris modality on the NIST-ICE database [101] (ICE-Exp1).

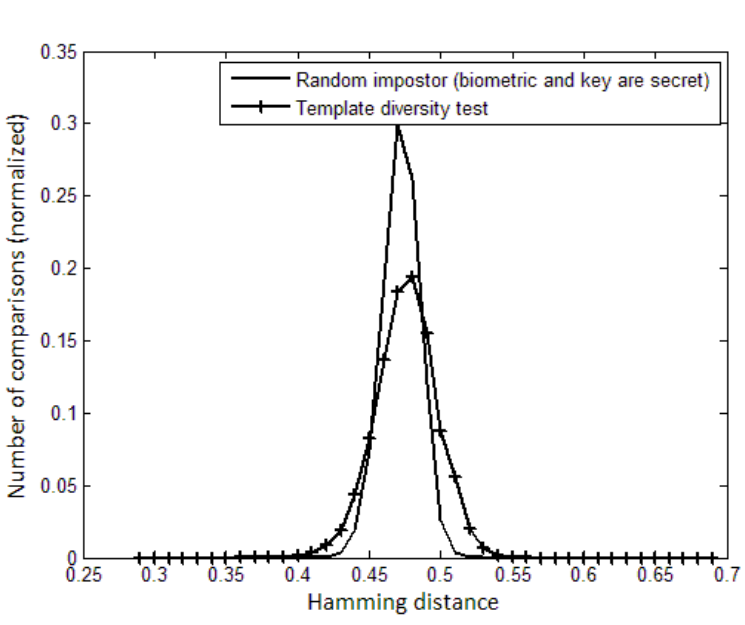


Figure 4.6: Impostor Hamming distance distributions for the proposed system along with the Hamming distance distributions for the template diversity test on iris modality on the NIST-ICE database [101] (ICE-Exp1).

4.2.2 Results and Security Analysis on Face Modality

Experimental Setup

Details about the data sets used for our experiments on face modality along with the associated experimental protocols are given in Section A.2.2 (page–186).

We have derived a subset of the FRGCv2 face database [100] for our experiments. We have used this subset instead of the complete FRGCv2 database in order to reduce the time required to run the full evaluation. The subset used in our experiments is composed of 250 subjects each of which has 12 images. Data from the first 125 subjects are used for development and the remaining 125 subjects are used for evaluation.

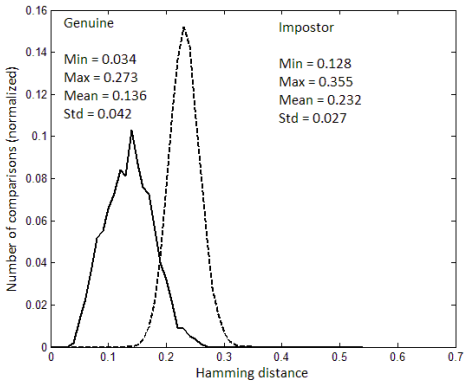
Two separate experiments are carried out during development as well as evaluation: FRGC-Exp1* – where the enrollment as well as test images are captured under controlled conditions, and FRGC-Exp4* – in which the enrollment images are from controlled conditions while the test images are from uncontrolled conditions. For the FRGC-Exp1*, 3,500 genuine and 496,000 impostor comparisons are carried out while for FRGC-exp4*, 4,000 genuine and 496,000 impostor comparisons are performed.

Results

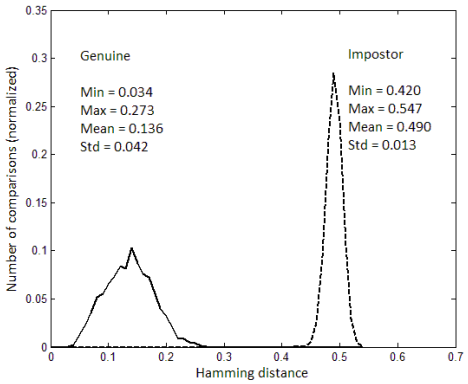
The Hamming distance distribution curves for genuine and impostor comparisons before and after shuffling on the development data sets are shown in Fig. 4.7. The curves for both, FRGC-Exp1* and FRGC-Exp4*, experiments are shown.

As was observed in case of iris, the impostor Hamming distances increase because of the shuffling process. Note that the genuine Hamming distances remain unchanged. A clear separation between genuine and impostor Hamming distance distributions is observed for both the experiments. This complete separation results in zero EER. The results of the proposed cancelable system for the FRGC-Exp1* and FRGC-Exp4* on the development data sets are reported in Table 4.4.

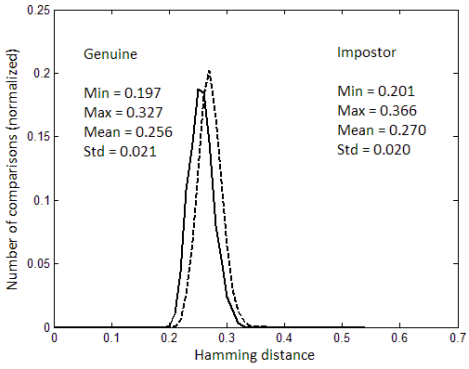
Note that, the improvement in performance is because of the increase in impostor Hamming distances. The shuffling scheme works as a randomization process which shifts the mean of the impostor Hamming distance distribution close to 0.5. Therefore, if the mean of the original (un-shuffled) impostor Hamming distance distribution is small, the improvement in performance will be more prominent. This can be visualized by



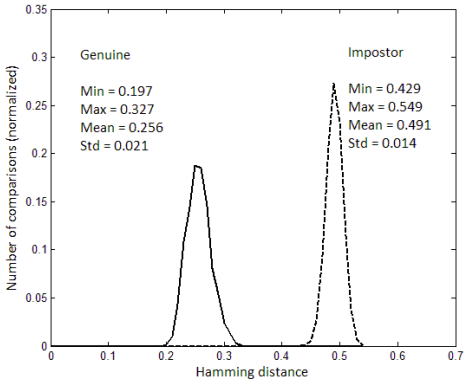
(a) Baseline face biometric system (FRGC-Exp1*)



(b) Baseline face system with shuffling (FRGC-Exp1*)



(c) Baseline face biometric system (FRGC-Exp4*)



(d) Baseline face system with shuffling (FRGC-Exp4*)

Figure 4.7: Normalized Hamming distance distributions for genuine and impostor comparisons on the NIST-FRGCv2 development data set for FRGC-Exp1* and FRGC-Exp4*.

Table 4.4: Verification results of the proposed cancelable system on face modality on development data sets in terms of EER in %. The values in bracket indicate confidence intervals.

Test	**Development set**	
	FRGC-Exp1*	**FRGC-Exp4***
Baseline	8.10[±0.41]	35.90[±0.68]
Proposed cancelable	0	0

comparing the improvements for iris and face modalities. For example, on the development data set CBS-BiosecureV1 for iris, as shown in Fig. 4.2, the average impostor

Hamming distance for iris is 0.44, which after shuffling, increases to 0.47. Similarly, for face, on the development data set Exp1 (Fig. 4.7), the average impostor Hamming distance is 0.23, which moves to 0.49 after shuffling. Thus, the increase in the separation between genuine and impostor Hamming distance curves is more in case of face than for iris. Therefore, the improvement in performance is higher in case of face than in case of iris.

The proposed cancelable system is then evaluated on the evaluation data sets. The two experiments defined earlier, FRGC-Exp1* and FRGC-Exp4*, are carried out. The Hamming distance distributions for these two experiments on the evaluation data sets are given in Fig. 4.8.

As it is seen for the experiments on development sets, a clear separation is obtained on the evaluation sets also. The outcome of this separation is zero EER as reported in Table 4.5.

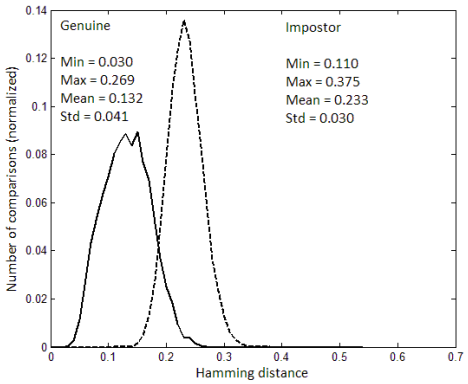
Table 4.5: Verification results of the proposed cancelable system on face modality on evaluation data sets in terms of EER in %. The values in bracket indicate confidence intervals.

Test	**Evaluation set**	
	FRGC-Exp1*	**FRGC-Exp4***
Baseline	7.65[±0.40]	35.00[±0.68]
Proposed cancelable	0	0

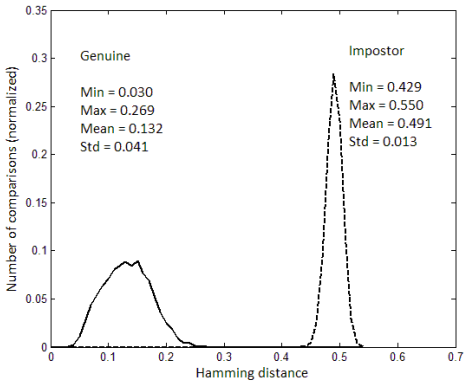
Security Analysis of the Proposed System on Face Modality

The experimental security analysis of the proposed system carried out for the iris modality is performed for the face modality also. The two scenarios: (i) stolen biometric scenario and (ii) stolen key scenario, are followed. During this tests, it is observed that the proposed cancelable system behaves in a similar way as it did on iris. The performance in case of the stolen biometric case remains unchanged. In the stolen key scenario, the performance is exactly the same as that of the baseline biometric system. The results for these tests in terms of EER are reported in Table 4.6.

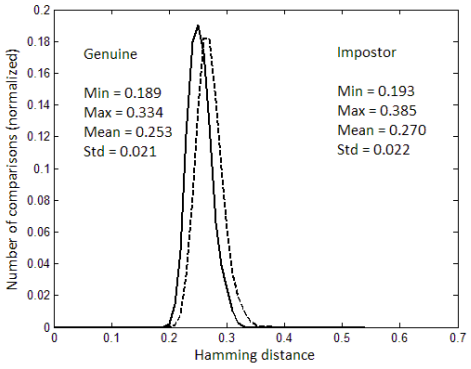
As reported in Table 4.6, the EER of the cancelable system on face modality is 0%. Therefore, it cannot be shown using the DET curves. The ROC curves for the face system (evaluation data set) are shown in Fig. 4.9. Similar to the iris modality, the



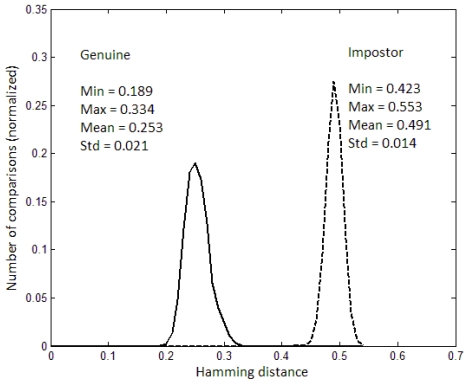
(a) Baseline face biometric system (FRGC-Exp1*)



(b) Baseline face system with shuffling (FRGC-Exp1*)



(c) Baseline face biometric system (FRGC-Exp4*)



(d) Baseline face system with shuffling (FRGC-Exp4*)

Figure 4.8: Normalized Hamming distance distributions for genuine and impostor comparisons on the NIST-FRGCv2 evaluation data set for FRGC-Exp1* and FRGC-Exp4*.

Table 4.6: Verification results for the cancelable system on face modality in terms of EER in % along with the experimental security analysis. Case-1 – face image is stolen; Case-2 – Shuffling key is stolen.

Test	**Development set**		**Evaluation set**	
	FRGC-Exp1*	**FRGC-Exp4***	**FRGC-Exp1***	**FRGC-Exp4***
Baseline	8.10[±0.41]	35.90[±0.68]	7.65[±0.40]	35.00[±0.68]
Proposed cancelable	0	0	0	0
Stolen biometric	0	0	0	0
Stolen key	8.10[±0.41]	35.90[±0.68]	7.65[±0.40]	35.00[±0.68]

curves for baseline system and that for the stolen key scenario overlap. Moreover, the curves for the cancelable system and the stolen biometric scenario also overlap indicating that the system performance mostly remains unaffected when the face image is stolen.

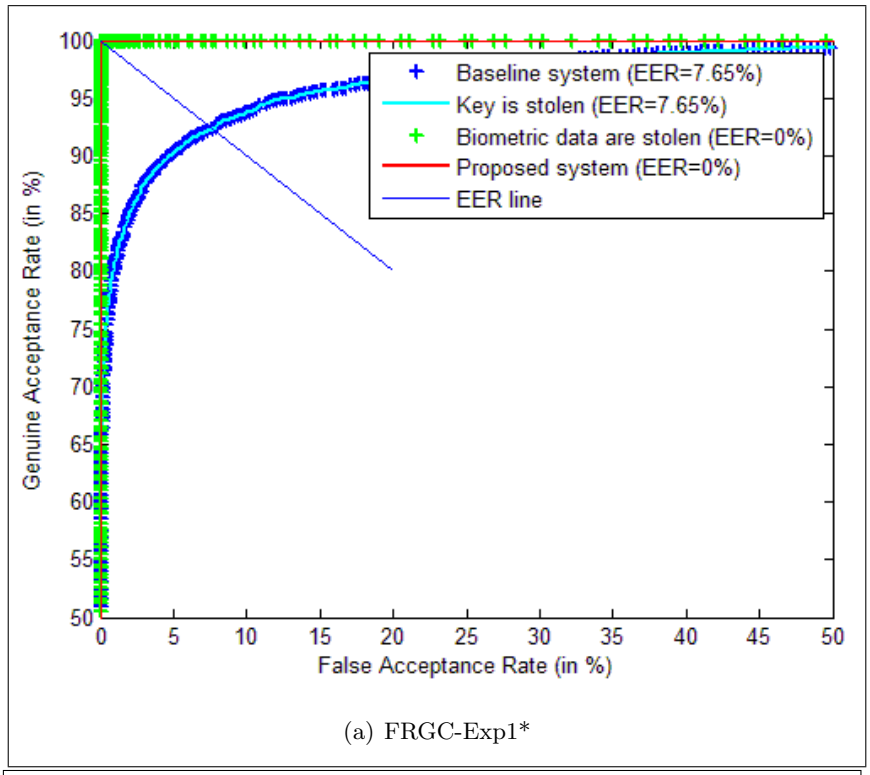
4.3 Conclusions and Perspectives

Classical biometric systems lack the important properties of revocability and template diversity because the biometric traits are permanently associated with the user. Cancelable biometric systems overcome these drawbacks of classical biometric systems. The shuffling scheme proposed in this chapter employs a randomly generated shuffling key to randomize the biometric feature codes. The shuffled feature vectors act as cancelable templates. The system can issue different templates for different applications using the same biometric which preserves privacy. If the stored template is compromised, it can be canceled and a new template can be issued by changing the shuffling key. Such use of shuffling key prevents an attacker from getting verified by providing the compromised template or stolen biometric data. One distinct advantage of this system is that the performance of the baseline system increases by more than 80% due to shuffling. And even if one of the two secret factors, the biometric data and the shuffling key, is compromised, the EER of the system in such scenario still remains less than or equal to that of the baseline biometric system.

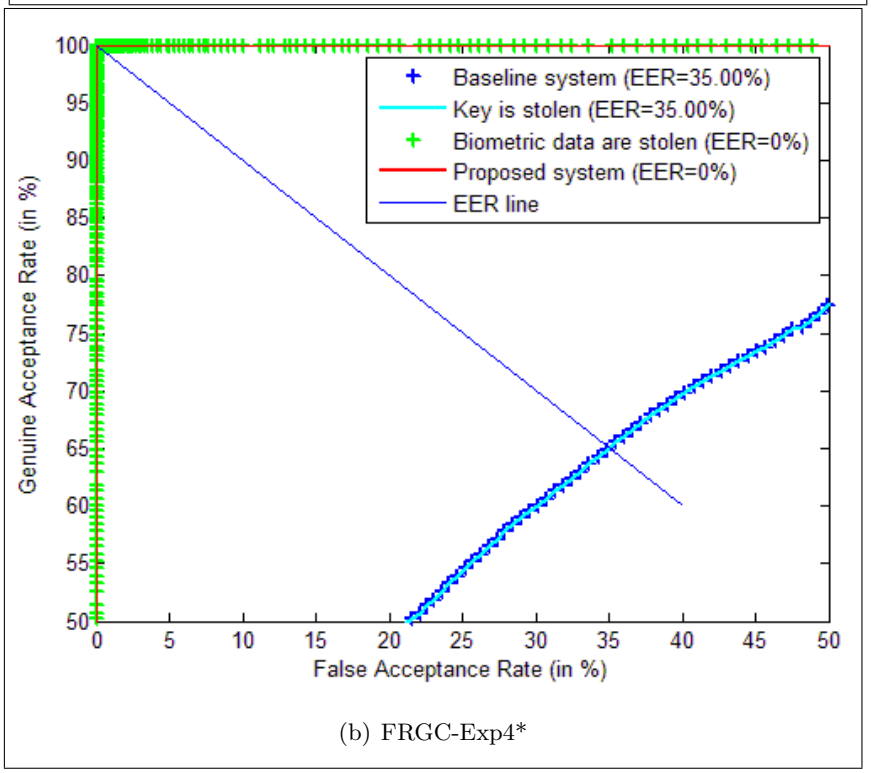
The drawback of this shuffling scheme is that it is not noninvertible. Practically, it works as a classical symmetric encryption where data can be encrypted by a key and the encrypted data can be decrypted by providing the same key. If an attacker succeeds to obtain the shuffling key, he can de-shuffle the cancelable template to obtain the reference biometric data. However, when such compromise is detected, the system can revoke the old template and issue a new one and the earlier attack becomes irrelevant.

A limitation of this shuffling scheme in its current form is that it can only be applied to biometric systems when the templates are in form of an ordered set. It cannot be applied to unordered sets such as a set of fingerprint minutiae.

The proposed shuffling scheme is very effective and therefore it is used as a means to induce revocability in all our proposed key regeneration systems in the following chapters.



(a) FRGC-Exp1*



(b) FRGC-Exp4*

Figure 4.9: ROC curves for the proposed system performance along with the possible security threats for face modality on the evaluation subset, NIST-FRGCv2 database.

Chapter 5

Using Error Correcting Codes to Reduce Variability in Biometric Data

5.1 Introduction

Among the biometrics research community, it is a well known fact that two measurements of a biometric source are not the same. The variations in such two recordings can be a result of various sources such as temporal, conditional, presentational, or random variations. For example, speech recordings may have high temporal variations, different positioning of fingerprints, lighting conditions for face, and presentation of iris resulting in rotational effect. Traditionally, signal processing techniques are applied in classical biometric systems to eliminate or reduce these variations. In this chapter, a novel method to deal with these variations with the help of Error Correcting Codes (ECC) is presented.

In the previous chapter, we proposed a shuffling scheme which increases discriminability of the biometric system by increasing impostor scores while keeping the genuine scores unchanged. This results in better separation between genuine and impostor distributions and therefore improves the verification performance. In order to further improve the verification performance, the separation between genuine and impostors should be increased. In this chapter, we treat the intra-user variability in bio-

metric data as errors and try to reduce it with the help of ECC. The proposed ECC scheme can correct partial errors in the biometric data and thus reduces the genuine Hamming distances. When the shuffling scheme is applied on such error corrected data, it only increases the impostor Hamming distances thus improving the separation between them. The outcome of the ECC scheme and shuffling scheme is the improvement in the verification performance.

ECC have already been used to cope with the biometric data variability in a number of crypto-biometric systems such as [39, 64, 54, 26, 12]. In these systems, the biometric data variability is treated as errors. All of these systems use ECC to remove *all* the errors that occur in biometric data. If the test data contain variabilities as compared to the reference data, the ECC treat those variabilities as errors and try to remove those errors. In such cases, the ECC (if successful) will remove all the errors from the test biometric data and recover the biometric data used during enrollment. But, if the errors are more than the error correction capacity, the recovered data is random. Thus, these systems can either correct all the errors or none of them. This condition to correct all the errors becomes too restrictive which decreases the verification accuracy of the system. This condition is relaxed in the system proposed in this chapter.

The proposed system can correct partial errors in the biometric data which helps improve the verification performance. Additionally, the cancelable biometric scheme described in Chapter 4 is applied on the error corrected biometric data to make the system revocable.

Note that the biometric data variability is treated as errors in the ECC based schemes. This variability means the differences between the reference and test biometric data. This can be different than actual noise in biometric signal. For example, even if a test biometric sample is free of noise but the enrollment sample is noisy, the ECC based schemes treat the test data for error correction.

This chapter is organized as follows: in Section 5.2, the basic idea of biometric data matching as a problem of communication through noisy channel is presented. The proposed system is based on this scheme. The proposed scheme for reducing biometric data variability is detailed in Section 5.3. Verification performance of the proposed system is evaluated on publicly available iris databases. Details about these results along with the security analysis are reported in Section 5.4. Finally, Section 5.5 sets out

conclusions and perspectives.

5.2 Biometric Data Matching as a Problem of Communication Through a Noisy Channel

The crypto-biometric key regeneration systems based on the fuzzy commitment scheme [64] described in Chapter 2 treat the biometric data matching issue as a problem of communication through a noisy channel. This model is shown in Fig. 5.1. Some examples of such systems are [54, 26]. This model considers transmission of data through a virtual channel. The differences in the biometric data from one acquisition to another are treated as noise. This noise causes errors in the data being transmitted.

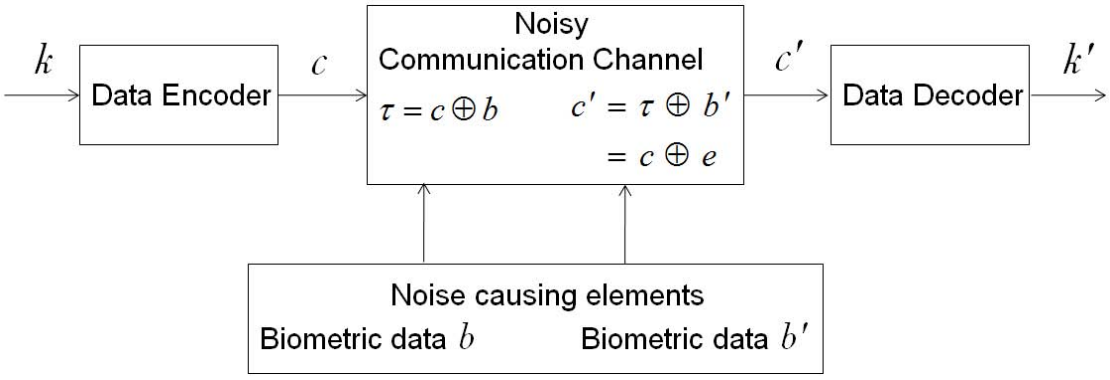


Figure 5.1: Biometric data matching as a problem of communication through noisy channel. Biometric data act as noise causing elements. This model applies to fuzzy commitment based key regeneration systems, e.g., [64, 54, 26]. In this figure, k = random key, k' = regenerated key, c = encoded codeword, and c' = corrupted codeword.

In this model, the aim is to transfer a random key $k, k \in [0,1]^m$ to the other end of the virtual channel. This key is first encoded by the ECC to obtain a codeword c, $c = ECC(k), c \in [0,1]^n$. The biometric feature vector b is XOR*ed* with c to obtain an enrollment template τ, $\tau = c \oplus b$. At the time of key regeneration, the test biometric feature vector b' XOR*ed* with τ to get c' which is actually c containing errors between b and b'. Thus, the errors e between b and b' are transferred onto the codeword c.

$$\tau = c \oplus b, \quad \text{and} \tag{5.1}$$

$$\begin{aligned} c' &= \tau \oplus b', \\ &= c \oplus b \oplus b', \\ &= c \oplus e. \end{aligned} \tag{5.2}$$

If the errors e between the biometric samples b and b' are less than the error correction capability of the ECC, the decoding process results in the exact key k and if not, the output of the decoding function is a random element belonging to $[0,1]^m$. This process is shown in Fig. 5.1. In this way, a random key k can be protected with the help of the enrollment biometric data and later recovered by providing another set of biometric data which is close to the enrollment data. This scheme can further be extended to obtain the reference biometric data b.

This model can treat either all the errors in the biometric data or none of them. The regenerated key k' is expected to have zero tolerance (i.e., Hamming distance = 0). This condition is too restrictive. Instead, if this condition is relaxed to allow some tolerance (i.e., Hamming distance $\leq$ threshold) the verification performance can improve. In order to relax this condition, this scheme is modified so that it can correct as many errors as possible. The total amount of errors corrected by our proposed system can be equal to the total amount of errors present in the biometric data or its fraction. The proposed scheme is described in the next Section.

5.3 Reducing Intra-user Variability in Iris Codes and Cancelable Template Generation

Crypto-biometric systems based on the fuzzy commitment scheme [64] use ECC to eliminate the errors caused by the biometric data in the encoded random key. None of the systems really performs error correction on the biometric data in order to reduce the variability among them. Here we introduce a novel way to use ECC by which we can successfully reduce the biometric variabilities. In this thesis work, the proposed system is developed for iris modality.

Iris codes obtained from two images of an iris are generally not the same. They have two different types of variabilities which we refer to as errors [54]: (a) *background*

errors, which are random in nature, occurring due to camera noise, image capture effects, iris distortions etc., and (b) *burst errors* which generally occur due to eyelids, eye lashes, specular reflections, etc. Hao et al. [54] proposed a concatenated scheme using Reed-Solomon codes to correct burst errors and Hadamard codes to correct background errors. The aim of this scheme is to obtain a key to be used in cryptographic systems. Hence, this system needs to correct all the errors in the iris codes.

Since burst errors occur due to eye-lids, eye lashes, and specular reflections, the probability of occurrence of these errors is nearly the same in genuine as well as impostor cases. The other type of errors, random errors, occur due to camera noise, image capture effects, etc. in genuine cases. In impostor cases, the random errors are due to the randomness of iris structures, i.e., these errors are due to the inter-personal variabilities. Therefore, in general, there are more random errors in impostor comparisons than in genuine comparisons.

Using this hypothesis, we propose a scheme which will correct only the random errors up to a certain limit such that the Hamming distance between genuine comparisons will decrease by a greater amount than in impostor cases. Later, the shuffling scheme is applied to add revocability. A combined effect of the error correction and shuffling is a better separation between genuine and impostor distributions which improves the verification performance. We use only Hadamard codes to correct the errors. A brief introduction to Hadamard codes is presented in the following subsection. The proposed algorithm for correcting errors in iris codes is then described in Section 5.3.2. The procedure to obtain revocable templates is then presented in Section 5.3.3.

5.3.1 Hadamard Codes

In this section, Hadamard codes are briefly introduced (details can be found in [79]). Hadamard codes are obtained from a Hadamard matrix generated by the Sylvester method. Hadamard matrix is a square orthogonal matrix with elements '1' or '-1'. The Hadamard code $HC(k$ is constructed from the Hadamard matrix $H(k)$ as:

$$HC(k) = \begin{bmatrix} H(k) \\ -H(k) \end{bmatrix}. \tag{5.3}$$

The codewords are obtained by replacing -1 with 0 in $HC(k)$. The Hadamard code of size $n = 2^k$ has $2n$ codewords each of which is n bits long. The code has a

minimum distance of 2^{k-1} and hence can correct up to $2^{k-2}-1$ errors (i.e., $\approx 25\%$). During encoding, an input value i is encoded into a codeword w. Here i is $(k+1)$ bits and w is $n=2^k$ bits. The matrix $HC(k)$ has $2n$ rows which are considered as codewords. The input value i is considered as a row index and the corresponding row is taken as an output codeword. Thus an input block of $(k+1)$ bits is converted into an output block of 2^k bits.

At the time of decoding, every 0 in the received codeword w is replaced by -1 to obtain w'. Then the product,

$$w'HC^T(k) = (a_0, a_1, ..., a_r, ..., a_{2n-1}), \tag{5.4}$$

is calculated. The position r, where a_r is maximum, is the decoded value. If at most $2^{k-2}-1$ errors have occurred, the decoded value is equal to the input value, i.e., $r=i$.

5.3.2 Correcting Errors in Iris Data

In this section, an algorithm to correct errors in iris data is proposed. The scheme involves two distinct phases namely enrollment phase and verification phase. In the enrollment phase, as shown in Fig. 5.2, a revocable template is generated from the enrollment iris image. In the verification phase, as shown in Fig. 5.3, a test iris image is provided by the user for verification.

The proposed system works on blocks of iris codes. The reference as well as test iris codes are divided into blocks of equal length. The length of these blocks is determined by the size of the output of the error correcting code used. In this case it is Hadamard code of size $n=2^{m-1}$. Let's consider the reference and test iris codes are denoted as $\mathbf{X}$ and $\mathbf{Y}$, respectively, where,

$$\mathbf{X} = \{x_1, x_2, ..., x_p\}, \quad \text{and} \tag{5.5}$$

$$\mathbf{Y} = \{y_1, y_2, ..., y_p\}. \tag{5.6}$$

Here x_i and y_i are binary strings each having $n=2^{m-1}$ bits. We need to correct errors in the *test iris code* $\mathbf{Y}$ with respect to the *reference iris code* $\mathbf{X}$. A $p \times m$ bit random bit-string $\mathbf{K}$, called a *random key*, is divided into p blocks of m bits each

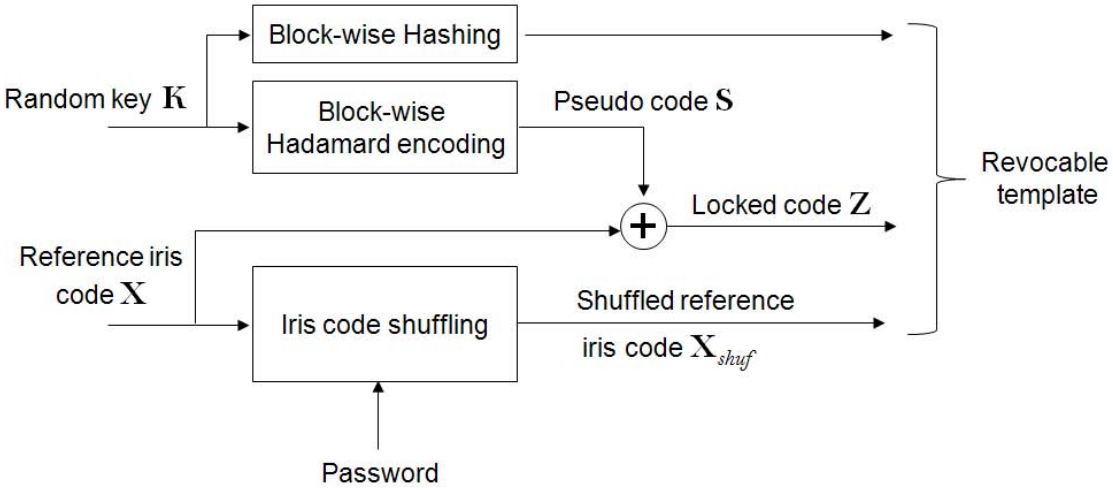


Figure 5.2: Block diagram showing the enrollment process for the proposed scheme. Here, $\mathbf{K}$ is a random key, $\mathbf{X}$ = reference iris code, $\mathbf{S}$ = pseudo code, $\mathbf{Z}$ = locked iris code, and $\mathbf{X}_{shuf}$ = shuffled reference iris code.

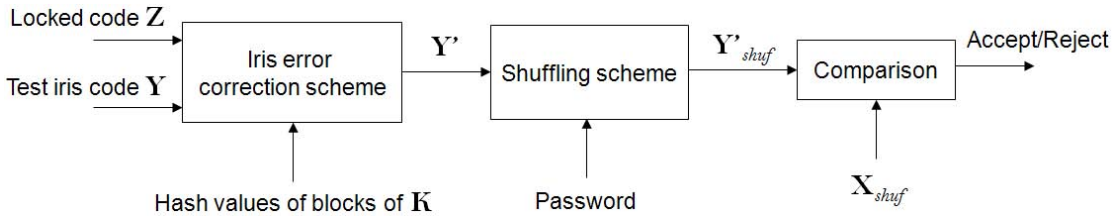


Figure 5.3: Block diagram showing the verification process for the proposed scheme. Here, $\mathbf{K}$ is a random key, $\mathbf{Y}$ = test iris code, $\mathbf{Z}$ = locked iris code, $\mathbf{Y}'$ = modified (error corrected) test iris code, $\mathbf{X}_{shuf}$ = shuffled reference iris code, and $\mathbf{Y}'_{shuf}$ = shuffled modified test iris code.

such that,

$$\mathbf{K} = \{k_1, k_2, ..., k_p\}. \tag{5.7}$$

Each block of $\mathbf{K}$ is encoded with a Hadamard code of size $(m-1)$. The output of the Hadamard encoding is a set of encoded codewords denoted as *pseudo code*:

$$\begin{aligned} \mathbf{S} &= \{s_1, s_2, ..., s_p\} \text{ where,} \\ s_i &= had\_enc(k_i). \end{aligned} \tag{5.8}$$

Each of the s_i is a binary string having $n = 2^{m-1}$ bits. This *pseudo code* is

XOR*ed* with the reference iris code $\mathbf{X}$ to form a *locked iris code* as:

$$\begin{aligned} \mathbf{Z} &= \mathbf{S} \oplus \mathbf{X}, \\ \text{and} \quad \mathbf{Z} &= \{z_1, z_2, ..., z_p\}, \quad \text{where,} \\ z_i &= s_i \oplus x_i. \end{aligned} \tag{5.9}$$

The shuffling scheme described in Chapter 4 is applied on the reference iris code to obtain revocable template. The *reference iris code* $\mathbf{X}$ is shuffled with a randomly generated user specific shuffling key to obtain a *shuffled iris code*, $\mathbf{X}_{shuf}$.

Hash value of each block of the key $\mathbf{K}$, i.e., each of the $k_i's$, is obtained. These hash values along with the *shuffled iris code* $\mathbf{X}_{shuf}$, and the *locked iris code* $\mathbf{Z}$, together form the reference template. The shuffling key can either be obtained from a password or protected by it. Alternatively, the complete template can be protected by a password. This constitutes the user enrollment phase as shown in Fig. 5.2.

The most distinct feature of the proposed scheme compared to the existing schemes (e.g., Juels and Wattenberg [64], Hao et al. [54], Bringer et al. [26], etc.) is in the decoding part. The decoding and error correction are carried out block-wise by processing one block at a time. While in the previous proposals, the decoding is done on the complete data. The flowchart of this process is shown in Fig. 5.4. The i^{th} block of $\mathbf{Y}$, y_i, is XOR*ed* with the i^{th} block of the locked code $\mathbf{Z}$, z_i, producing s'_i. This s'_i is a corresponding block of pseudo code $\mathbf{S}$, s_i, with the errors e_i between reference iris code block x_i and test iris code block y_i transferred onto it.

$$\begin{aligned} s'_i &= z_i \oplus y_i, \\ s'_i &= s_i \oplus x_i \oplus y_i, \\ s'_i &= s_i \oplus e_i. \end{aligned} \tag{5.10}$$

The s'_i is decoded by the Hadamard code to obtain k'_i as:

$$k'_i = had\_dec(s'_i). \tag{5.11}$$

The hash value of this k'_i is compared with the hash value of the corresponding block k_i of the original random key. If the two hash values are equal, it means that $k'_i = k_i$. If so, then the k'_i is re-encoded using Hadamard code to obtain s''_i. Since

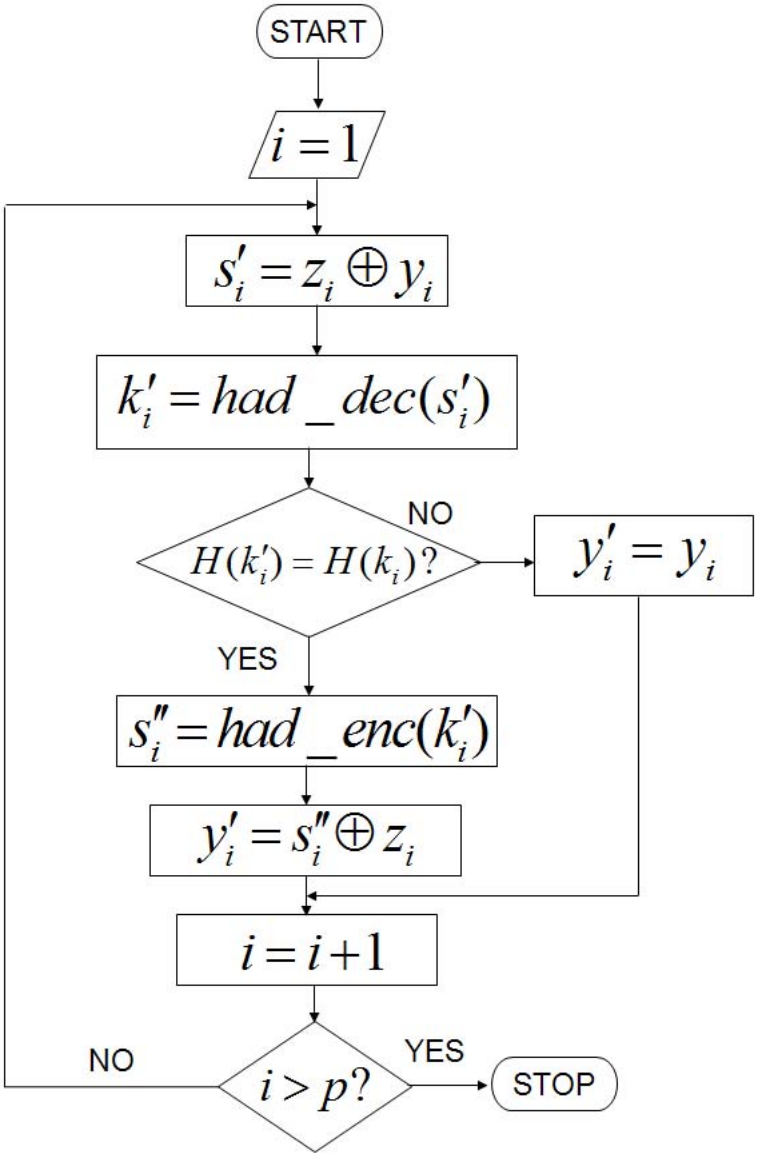


Figure 5.4: The proposed algorithm for applying ECC to reduce variability in iris codes (the "Iris error correction scheme" block shown in Fig. 5.3), where z_i = locked iris code block; y_i = test iris code block; y_i' = modified test iris code block; k_i = random key block; *had_enc* = Hadamard encoding; and *had_dec* = Hadamard decoding.

$k_i' = k_i$, $s_i'' = s_i$. This s_i'' is XOR*ed* with z_i to obtain y_i'. Since, $s_i'' = s_i$, $y_i' = s_i'' \oplus z_i = s_i'' \oplus s_i \oplus x_i = x_i$.

$$
\begin{aligned}
If H(k_i') &= H(k_i), \\
k_i' &= k_i, \\
\text{and, } s_i'' &= had\_enc(k_i'). \qquad (5.12) \\
\text{But, } s_i &= had\_enc(k_i). \\
\therefore s_i'' &= s_i. \qquad (5.13)
\end{aligned}
$$

$$
\begin{aligned}
y_i' &= s_i'' \oplus z_i, \\
y_i' &= s_i'' \oplus s_i \oplus x_i, \\
\therefore y_i' &= x_i. \qquad (5.14)
\end{aligned}
$$

If $k_i' \neq k_i$, then k_i' is not re-encoded and without further processing, $y_i' = y_i$. Note that, Equations (5.12)–(5.14) are irrelevant in this case. This process of block-wise error correction is carried out for all blocks $i, (i = 1, 2, ..., p)$ resulting in $\mathbf{Y}' = \{y_1', y_2', ..., y_p'\}$ which is the modified test iris code. This code is the test iris code with some blocks replaced by the regenerated reference iris code blocks. Consequently, when this modified test iris code is compared with the reference iris code, the Hamming distance is decreased. In case of genuine users, the random errors are generally less than 25% (which is the Hadamard code error correction capability) whereas for impostors, the errors are generally more than 25%. Hence, this error correction scheme helps reduce the genuine Hamming distance by a significant amount which ultimately helps in a better user separation.

In the next step (see Fig. 5.3), the error corrected test iris code $\mathbf{Y}'$ is shuffled using the shuffling key and this shuffled code $\mathbf{Y}'_{shuf}$ is compared with the shuffled reference iris code $\mathbf{X}_{shuf}$. The final verification decision is taken based on the score of this comparison. This verification process is shown in Fig. 5.3.

The biometric data can be recovered from the locked iris code if an impostor knows the key $\mathbf{K}$. In order to provide security to the biometric data and protect user's privacy, the key $\mathbf{K}$ is not stored in the system. Instead, a one-way hash function is used to hash each block of the key and the hash values are stored as part of the template.

Moreover, the iris code comparison is not carried out in the classical way. In conventional iris code matching algorithm, the iris noise masks represent the locations of possible errors in the iris codes. This helps to suppress the possible error bits from the iris code comparison resulting in symmetric error correction assuming that both, the reference and test iris codes, may contain errors. The Hamming distance considering masks (HD_{mask}) between two iris codes is calculated using the following formula:

$$HD_{mask} = \frac{\|(Code_1 \oplus Code_2) \cap Mask_1 \cap Mask_2\|}{\|Mask_1 \cap Mask_2\|}, \tag{5.15}$$

where, $Code_1$, $Code_2$ represent the reference and test iris codes respectively and $Mask_1$, $Mask_2$ are their respective noise masks. It is clear from equation (5.15) that both the iris masks are needed to be logically AND*ed* and thus the reference mask must be stored in the system. This mask can leak vital information about the iris code making the system weaker (in case of attacks). Also using such masks will enable an impostor to select only a certain number of bits from the iris code and he can get more easily accepted by the system. Hence from a security point of view, we prefer not to use the masks. In fact, using the noise masks improves the performance of the system. Thus by opting not to use the masks, we are making the verification process more difficult. Similar approach was followed by Hao et al. [54].

5.3.3 Use of the Shuffling Scheme to Obtain Cancelable Templates

The biometric data shuffling scheme described in Section 4.1 is used in the proposed system which further improves the user separation capability of the system. In this shuffling scheme, an iris code is shuffled with a user specific random shuffling key. The iris code is divided into blocks and these blocks are aligned with the shuffling key bits. If a bit in the shuffling key is 1, the corresponding block is taken into part 1 and if the bit is zero, the corresponding block is taken into part 2. The concatenation of the two parts gives a shuffled iris code. The Hamming distance between shuffled reference iris code $\mathbf{X}_{shuf}$ and shuffled modified test iris code $\mathbf{Y'}_{shuf}$ is considered in order to make the verification decision. This shuffling scheme is described earlier in Section 4.1 (page 48) in details.

As discussed in Section 4.1.2, the most distinctive feature of this shuffling scheme is that it increases the Hamming distance for impostor comparisons but the

Hamming distance for genuine comparisons remains intact. Thus, the ECC scheme decreases the Hamming distance for genuine comparisons and the shuffling scheme increases the Hamming distance for impostor comparisons which results in improvement in the verification performance. This effect can be seen from the Hamming distance distributions in Fig. 5.5. Note that the key regeneration systems (e.g., Juels and Wattenberg [64], Hao et al. [54], Bringer et al. [26], etc.) do not attempt to change the Hamming distance distributions, and hence, they cannot improve the performance of the underlying biometric system. The shuffling based cancelable scheme from Chapter 4 improves the distribution but the improvement is achieved only due to increase in impostor Hamming distances. The scheme proposed in this chapter works on both, the genuine as well as the impostor Hamming distances and hence the improvement is significantly higher than the other schemes as shown in following section.

5.4 Experimental Results and Security Analysis of the Proposed System

5.4.1 Experimental Setup

Following the strategy used in Chapter 4, we developed our system on the publicly available CBS database [103]. In order to prove the portability of our system, the system is then evaluated on the NIST-ICE database [101] with the parameters obtained from the CBS database tests. These databases and their associated protocols are described in Section A.2.1 (page–186). The tuning parameters obtained from the development sets are the error correction capacity and the length of the shuffling key. The error correction capacity is decided by observing the genuine and impostor Hamming distance distributions on the development data set shown in Fig. 5.5. The length of the shuffling key is the same as that used in the previous chapter.

On the NIST-ICE database (evaluation database), two separate experiments are defined: ICE-Exp1 consisting of comparisons of right eye images, and ICE-Exp2 which consists of left eye image comparisons. All possible comparisons are carried out for genuine as well as impostors. In total, 12,214 genuine and 1,002,386 impostor comparisons are carried out in ICE-Exp1, whereas in ICE-exp2, 14,653 genuine, and 1,151,975 impostor comparisons are performed.

The proposed system involves error correction and shuffling. Therefore, we also perform tests when only the error correction scheme is applied without shuffling. The results for only shuffling without error correction are the same as those presented in the previous chapter (Section 4.2.1, page 52). We also performed tests when one of the two factors, biometrics and shuffling key (password), is compromised for all the persons and compared it with the baseline biometric system's verification performance.

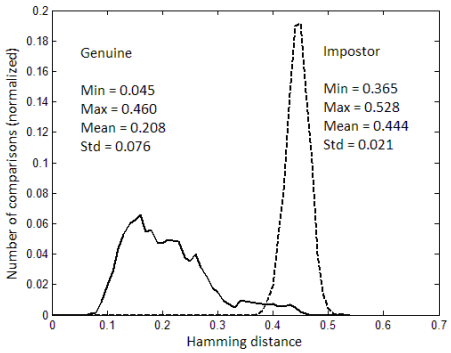
5.4.2 Experimental Results

We first present the results on the CBS database [103]. Three tests are carried out: (1) comparisons using the baseline biometric system (described in Section A.1), (2) baseline system with error correction, and (3) the proposed system (baseline system with error correction and shuffling). The Hamming distance distribution plots for these three tests on the CBS-BiosecureV1 data sets are shown in Fig. 5.5. These plots clearly indicate that when the proposed error correction scheme is applied to the biometric data and then the shuffling is performed, a significantly better separation between genuine and impostor distributions is obtained. This better separation indicates greater ability to distinguish between genuine and impostors and improves the verification performance.

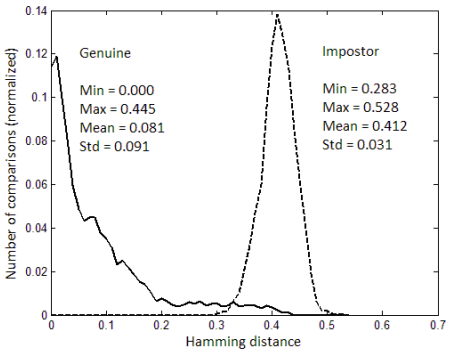
The verification results of the proposed system on the development data sets (CBS database) are reported in Table 5.1. These results are reported in terms of EER in %. From this table, it can be seen that, after applying the ECC and shuffling scheme, the EER decreases by more than 90%, e.g., on the CBS-BiosecureV1 data set, the EER decreases from 2.63% for the baseline biometric system to 0.14% for the proposed system. The Detection Error Trade-off (DET) curves plotted in Fig. 5.6(a) and 5.6(b) clearly show these improvements achieved by the proposed system over the baseline system. Note that, some security scenarios discussed in the next subsection are also shown in these figures.

Table 5.1: Verification results of the proposed system on iris development data sets (CBS database [103]); in terms of EER in %. Values in bracket indicate the error margins for 90% confidence intervals.

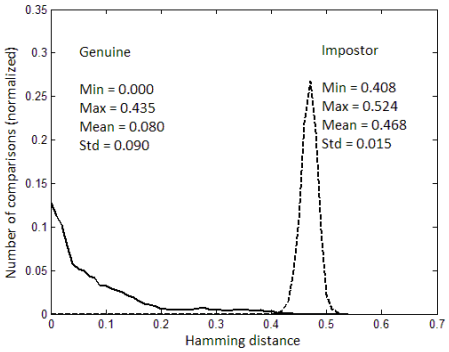
Experiment	**CBS-BiosecureV1**	**CBS-CasiaV2**
Baseline	2.63[±0.34]	3.03 [±0.36]
Proposed	0.14[± 0.08]	0.10 [±0.07]
Baseline+Cancelable (Table 4.1)	0.93[±0.20]	0.56[±0.16]



(a) Baseline iris biometric system



(b) Baseline iris system with error correction

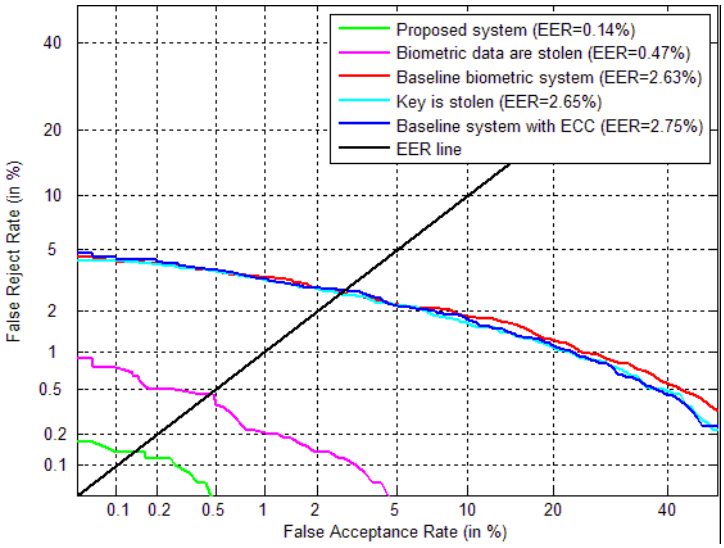


(c) The proposed system (baseline iris biometric system with error correction and shuffling

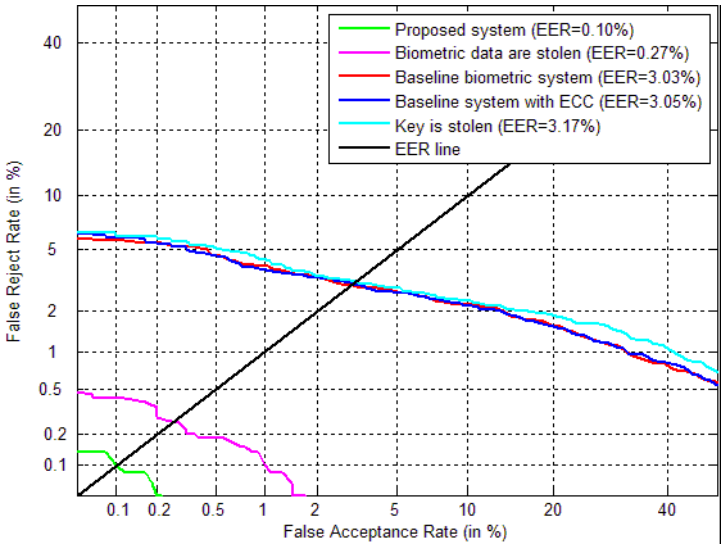
Figure 5.5: Normalized Hamming distance distributions for genuine and impostor comparisons on the development data set (CBS-BiosecureV1 data set [103]).

Using the parameters from the development sets, the proposed system is further evaluated on the NIST-ICE database. These results are reported in Table 5.2. As observed on the development database, more than 90% reduction is achieved with the proposed system on the NIST-ICE database for both the experiments. On the ICE-Exp1, the EER decreases from 1.70% for the baseline biometric system to 0.06% for the proposed system. Similarly, for the ICE-Exp2, the EER reduced from 1.80% to 0.13%. The Detection Error Trade-off (DET) curves shown in Fig. 5.7. These curves clearly show the improvement achieved by the proposed system over the baseline system.

Another popular way to report the biometric system performance is to report the values of FRR at fixed values of FAR. The results of the NIST-ICE evaluations [101] were reported in this manner. In order to facilitate the comparison, the results on ICE



(a) CBS-BiosecureV1 data set



(b) CBS-CasiaV2 data set

Figure 5.6: DET curves showing the performance comparison of the proposed system with the baseline biometric system on the development database (CBS database [103]).

database are reported in Table 5.3. The performance of the proposed system[1] is better than the best reported result (0.1-0.2% FRR at 0.1% FAR for the SAGEM algorithm) in ICE [101].

[1]Note that the performance of the proposed system is with a combination of biometrics and an assigned secret (shuffling key)

Table 5.2: Verification results of the proposed system on iris evaluation data sets (NIST-ICE database [101]); in terms of EER in %. Values in bracket indicate the error margins for 90% confidence intervals.

Experiment	ICE-Exp1	ICE-Exp2
Baseline	1.70 [±0.11]	1.80 [±0.10]
Proposed	0.06 [± 0.02]	0.13 [±0.03]
Baseline+Cancelable (Table 4.2)	0.23[±0.04]	0.37[±0.05]

Table 5.3: Verification results in terms of FRR at specified values of FAR for the ICE database, (all values are in %); (a) baseline biometric system, (b) proposed system.

Experiment	FRR in % at		
	FAR=0	FAR=0.001	FAR=0.1
ICE-Exp-1 (a)	13.62	8.24	3.51
ICE-Exp-1 (b)	0.15	0.12	0.06
ICE-Exp-2 (a)	25.80	13.78	4.35
ICE-Exp-2 (b)	0.44	0.29	0.13

5.4.3 Security Analysis

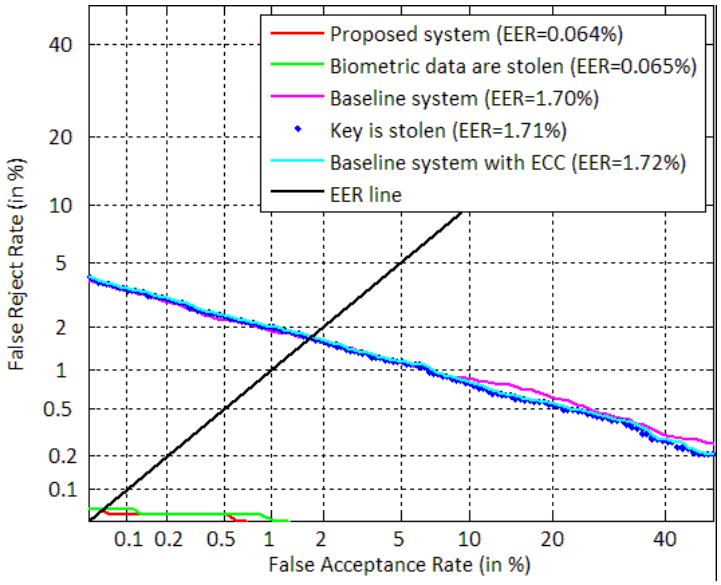
Theoretical Security Analysis

The proposed system provides protection to the templates and here we estimate the security theoretically. The system needs two inputs from the user: an iris image and a password. We propose to use an 8-character randomly generated password which can have 52-bit entropy [31]. The iris code itself contains some correlations and following the procedure given by Daugman [37], we estimate the degrees of freedom in the 1,188-bit iris codes as follows:

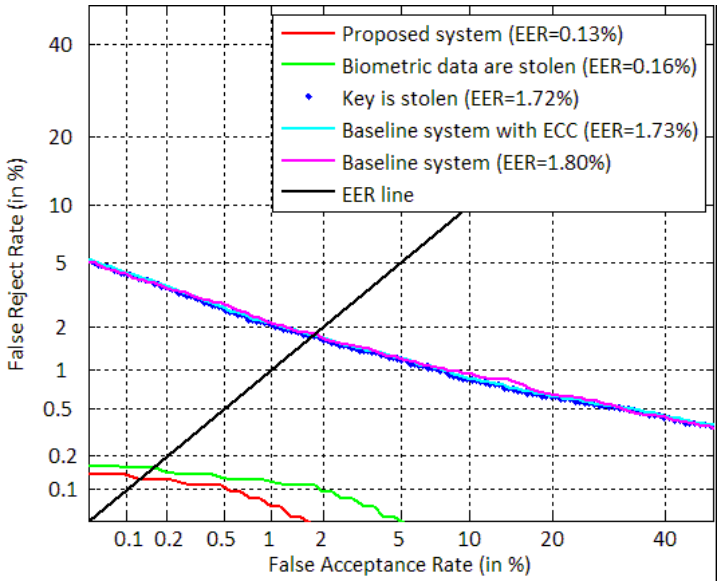
Let μ and σ be the mean and standard deviation of the binomial distribution fitting the impostor Hamming distance distribution. Then the number of degrees of freedom is estimated to be:

$$N = \mu(1-\mu)/\sigma^2. \tag{5.16}$$

With this procedure, we estimate the number of degrees of freedom to be 561. The Hadamard code acts on 32-bit blocks of the 1188-bit iris code thus, on average, each block has $f \approx 15$ degrees of freedom. The Hadamard code can correct at most 7 error bits occurring in 32-bit block resulting in 22% error correction. Considering this error tolerance (i.e., $g = 0.22 \times f \approx 3$) and following the procedure given in Hao et al. [54],



(a) ICE-Exp1 (right eye experiment)



(b) ICE-Exp2 (left eye experiment)

Figure 5.7: DET curves showing the performance comparison of the proposed system with the baseline biometric system along with security scenarios on the NIST-ICE database [101].

an impostor will need:

$$BF \approx \frac{2^f}{\binom{f}{g}} \approx \frac{2^{15}}{\binom{15}{3}} \approx 72, \tag{5.17}$$

brute force calculations to successfully guess the random key block. But, since the

random key block is only 6 bits long, the maximum number of calculations required per block will be 64. To guess the random key completely, $64 \times 37 \approx 2^{11.2}$ calculations will be required. Thus the overall security provided by the system is $52 + 11 = 63$ bits. This security can be significantly increased by limiting the maximum number of login attempts.

Experimental Security Analysis

In Chapter 3, we have defined experimental evaluation strategy for biometric and crypto-biometric system. For the crypto-biometric systems, two security scenarios are considered: (1) stolen biometric scenario – when an impostor obtains the biometric data of a genuine person but not the shuffling key, and (2) stolen key scenario – when an impostor obtains the shuffling key but not the biometric data.

Analysis of the proposed system is carried out for these two security scenarios. The results, in terms of EER, are reported in Table 5.4. The DET curves corresponding to these tests are shown in Fig. 5.6 and 5.7 on the development (CBS) and evaluation (NIST-ICE) databases, respectively. An additional test, denoted as "Baseline+ECC" is also reported in this table. In this test, the error correction scheme is applied to the biometric data but the shuffling is not performed. This test is for comparison purposes only.

Table 5.4: Security analysis of the proposed system; the values of EER are in %. Two scenarios are considered: (i) stolen biometric and (ii) stolen key. Values in bracket indicate the error margin for 90% confidence intervals.

Experiment	Development Set		Evaluation Set	
	CBS-BiosecureV1	**CBS-CasiaV2**	**ICE-Exp1**	**ICE-Exp2**
Baseline	2.63[±0.34]	3.03[±0.36]	1.70[±0.11]	1.80[±0.10]
Baseline+ECC	2.75[±0.35]	3.05[±0.36]	1.72[±0.11]	1.73[±0.10]
Proposed	0.14[±0.08]	0.10[±0.07]	0.06[±0.02]	0.13[±0.03]
Stolen biometric	0.47[±0.14]	0.27[±0.11]	0.06[±0.04]	0.16[±0.05]
Stolen key	2.65[±0.34]	3.17[±0.37]	1.71[±0.11]	1.72[±0.10]

In the stolen biometric scenario, when an impostor always provides a stolen iris image of a genuine person, it is observed that, the EER increases compared to that of the proposed system. For example, for the ICE-Exp2, the EER of the proposed system is 0.13% which increases to 0.16% in the stolen biometric scenario. But, this increase is

not significant. On the other hand, in the stolen key scenario, for the same experiment, the EER significantly increases to 1.72%. Even though this EER is much higher than the proposed system, it is still less than the baseline biometric system. From this discussion, we can say that if one of the two factors (biometric and shuffling key) in the system is stolen, the system is at least as secure as the baseline biometric system.

The shuffling scheme employed in the system adds revocability, template diversity, and privacy protection to the system. If a template of a user is compromised, the system can issue a new template for that user with the same biometric trait by changing the shuffling key and password. If the same attacker, who got the compromised template in the previous attempt, tries to access the system after revocation, he has to crack the template again. This is because, as shown in the stolen biometric scenario, the new shuffling key protects the system.

In Section 4.2.1 (page 55), we have already shown that the shuffling scheme provides template diversity. The same shuffling scheme is used in the system proposed in this chapter. Therefore, the proposed system possesses the important property of template diversity.

As defined in Chapter 1 (page–5), there are three aspects of privacy compromise: (i) cross database matching, (ii) raw biometric data recovery, and (iii) identification (proof of enrollment in a particular system). In the proposed system the stored template is revocable. It is infeasible to recover the biometric data from it without the shuffling key. Moreover, since it possesses template diversity, cross database matching is not possible. Finally, the involvement of password forces the system to operate in verification mode only. Therefore, identification of a person is not possible. Thus, the proposed system protects user's privacy.

The drawbacks of the shuffling scheme, such as invertibility if shuffling key is available, applicable to only ordered set of biometric features, etc., are inherited by this system.

5.5 Conclusions and Perspectives

In this chapter, we proposed a novel approach to use Error Correcting Codes (ECC) for reducing variabilities (which are treated as errors) in iris codes. After care-

fully studying the causes of errors in iris codes, we proposed an ECC scheme that can correct more errors in genuine iris codes than in impostors. Moreover, we use an iris code shuffling scheme which shuffles the iris data with a user specific randomly generated shuffling key. The shuffling scheme increases the Hamming distance for impostor comparisons whereas for the genuine comparisons, the Hamming distance remains the same. The combination of the two techniques enables the system to distinguish genuine users from impostors with higher accuracy as compared to the baseline biometric system. The shuffling key is protected by a password which makes the system revocable. The templates do not store personal biometric data in its usual form thereby protecting the user privacy. The use of password (shuffling key) does not allow access to the system even if a stolen iris image is provided.

The proposed system improves the verification performance of the underlying biometric system by reducing the EER by more than 90%, e.g., for ICE-Exp1, the EER reduces from 1.70% to 0.064%. Even for high security range, where FAR = 0%, the FRR on for ICE-Exp1 was 0.15%. Only 18 of the 12,214 genuine comparisons were falsely rejected while there was no false acceptance for more than a million impostor tests. This system is also suited for high security applications.

The idea of using ECC to *reduce* the biometric variability, in general, can be applied to any biometric modality having ordered binary feature sets provided the ECC is tuned according to the nature of errors in that biometric data.

Chapter 6

Cryptographic Key Regeneration Using Biometrics

In Chapter 2, we have classified the crypto-biometric systems in two categories: (a) protection of biometric data, and (b) Obtaining cryptographic keys with biometrics. From the first category, we proposed two systems in Chapter 4 and 5 which add revocability, template diversity, and privacy protection to the classical biometric systems along with improving the verification performance. In this and the following chapter, we propose various systems which deliver strong crypto-bio keys from biometric data.

Obtaining cryptographic keys using biometrics is a remarkable concept because it offers distinct advantage over classical methods of generating cryptographic keys. Classical cryptographic systems rely on identifiers such as passwords or tokens, that are assigned to the users by system administrators, in order to authenticate the user and generate secure keys for that user. Clearly, these assigned secrets have their own disadvantages like they can be stolen or shared, and hence, are insufficient to prove the user's identity. Using biometrics to obtain cryptographic keys, which are denoted as crypto-bio keys in this thesis, can provide a better solution as far as identity verification is concerned. Biometrics can be employed for obtaining crypto-bio keys in different ways as discussed in Section 2.2 (page–23), which are: cryptographic key release, key generation, and key regeneration. The systems proposed in this chapter are from the third category, i.e., cryptographic key regeneration using biometrics.

From the literature review presented in Chapter 2, it is observed that not

many systems achieve all of the goals that we defined for crypto-biometric systems in Section 1.3 (page–11). These goals are: identity verification and non-repudiation, revocability, template diversity, privacy protection, and performance improvement. For most of the systems, the performance gets degraded compared to the baseline system while for those with performance improvement, the performance degrades in the stolen key scenario. Some of the systems do not possess the property of revocability.

The system proposed in this chapter achieves all these goals. The most important goal of this proposal is to obtain high entropy keys which are strongly linked to the user's identity. The key regeneration system proposed in this chapter is based on the fuzzy commitment scheme [64]. As discussed in Section 5.2 (page–67), the fuzzy commitment scheme converts biometric data matching into an error correction problem. Hao et al. [54] proposed an adaptation of this fuzzy commitment scheme for iris biometrics. Their system employs a two-level error correction scheme. Our proposed system is a modified and improved version of this scheme. The shuffling based cancelable biometric system proposed in Chapter 4 is applied on the biometric data before using it in the fuzzy commitment scheme. This makes the system revocable and also improves the verification performance.

As it is done in all fuzzy commitment based schemes, the proposed system treats the biometric data variability as errors. The system copes with these errors with the help of Error Correcting Codes (ECC). The system described in this chapter is evaluated on two different modalities: iris and face. Since the nature and amount of errors (variability) that can occur in iris are different than that in face, the ECC needs adaptations according to the modality.

This chapter is organized as follows: a generic proposal for biometric based key regeneration is presented in Section 6.1. One of the most important point to consider while designing this key regeneration system is the selection of ECC. General description about selection of ECC is given in Section 6.1.2. The iris based key regeneration system along with experimental results is proposed in Section 6.2. In Section 6.3, we propose another key regeneration system to obtain keys from face biometrics. An extension of the generic scheme in Section 6.1 is proposed in Section 6.4. This extended scheme can obtain the reference biometric data and improves the security of the system. Finally, the Section 6.5 sets out our conclusions and perspectives.

6.1 Biometrics Based Key Regeneration Scheme

The system proposed in this section is a hybrid system which combines the ideas from the transform based cancelable biometric system (Section 2.1.2) with the biometrics based key regeneration approach (Section 2.2.3). The shuffling based cancelable biometric system described in Chapter 4 is applied on the biometric data before using them in the key regeneration scheme. As said earlier, the key regeneration scheme is an improved version of the Hao et al. [54] scheme which is an adaptation of the fuzzy commitment scheme (which was originally proposed by Juels and Wattenberg [64]) for iris biometrics.

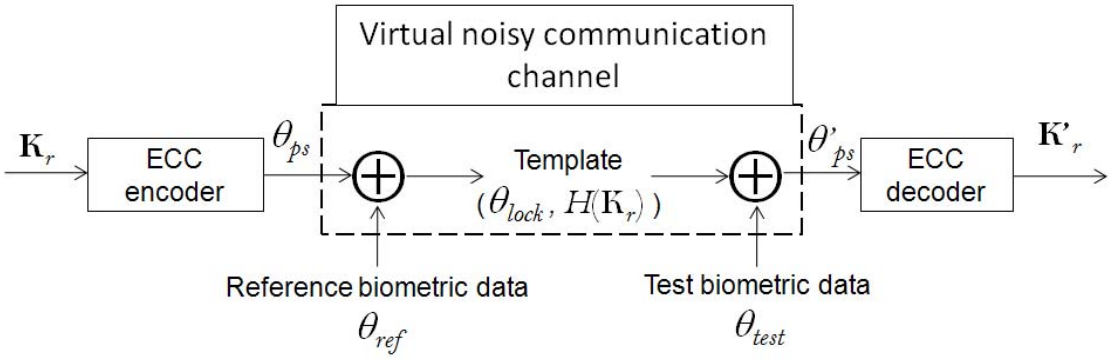


Figure 6.1: A simplified diagram of the fuzzy commitment [64] based key regeneration scheme.

A simplified diagram showing the fuzzy commitment based key regeneration scheme is presented in Fig. 6.1. As discussed in Section 5.2, the fuzzy commitment scheme models biometric data matching as a noisy channel communication problem. A randomly generated key $\mathbf{K}_r$ is treated as the information that needs to be transmitted through the virtual channel which can induce noise in the transmitted data. The noise is caused by the variations in biometric data provided during enrollment and verification. The variations between the reference and test biometric data are treated as errors.

In order to cope with these errors, the key $\mathbf{K}_r$ is first encoded using Error Correcting Codes (ECC) to obtain an encoded codeword θ_{ps} before passing through the virtual channel. The encoded codeword θ_{ps} is called as pseudo-code since its size and nature is similar to that of the biometric data. Bit-wise XOR*ing* is used to combine the reference biometric data θ_{ref} with the pseudo-code θ_{ps}. The XOR*ing* process acts as an

encryption algorithm so that neither the key $\mathbf{K}_r$ nor the reference biometric data θ_{ref} can be recovered from the combined data without the presence of either of them. Therefore, the combined data is referred to as locked code θ_{lock}. XOR*ing* the test biometric data θ_{test} with the locked code θ_{lock} transfers the variations between the two biometric data onto the encoded key. The error-transferred encoded key is decoded by the ECC to recover the key $\mathbf{K}'_r$. If the variations between biometric samples are within the error correction capacity of the ECC, the recovered key $\mathbf{K}'_r$ is the same as the original key $\mathbf{K}_r$. In order for the XOR*ing* operation to work, the lengths of the encoded key (i.e., θ_{ps}) and that of the biometric data being XOR*ed* (i.e., θ_{ref} or θ_{test}) must be the same. If they are not, then the biometric data is truncated to match their lengths.

In our proposed scheme, in order to induce revocability, the shuffling based cancelable biometric system described in Chapter 4 is applied on the biometric data before using it in the fuzzy commitment based scheme. Additionally, we use a two-level error correction scheme first proposed by Hao et al. [54]. The schematic diagram of the proposed system is shown in Fig. 6.2.

6.1.1 Revocability in the Key Regeneration System

Many systems have been proposed in literature based on the fuzzy commitment scheme described above. One of the most important of these schemes is that proposed by Hao et al. [54] using iris biometrics. Hao et al. [54] proposed a concatenated ECC scheme to cope with the errors (variability) in iris data. In such systems, if it is found out that the original crypto-biometric template is compromised, it can be canceled and a new one can be created by changing the random key $\mathbf{K}_r$. The drawback of this scheme is that the compromised biometric data can still be used to obtain the new crypto-biometric key, since the user is required to provide only the biometric data during verification in this scheme. Moreover, though the system can issue different crypto-biometric templates for different applications by using different random keys $\mathbf{K}_r$, such keys can be recovered from all of them by using a single set of compromised biometric data. This drawback is inherited by the Hao et al. system.

In order for the system to possess the properties of revocability and template diversity, biometric data must be combined with some kind of assigned secret which a user needs to provide at the time of key regeneration. One way to achieve this is by

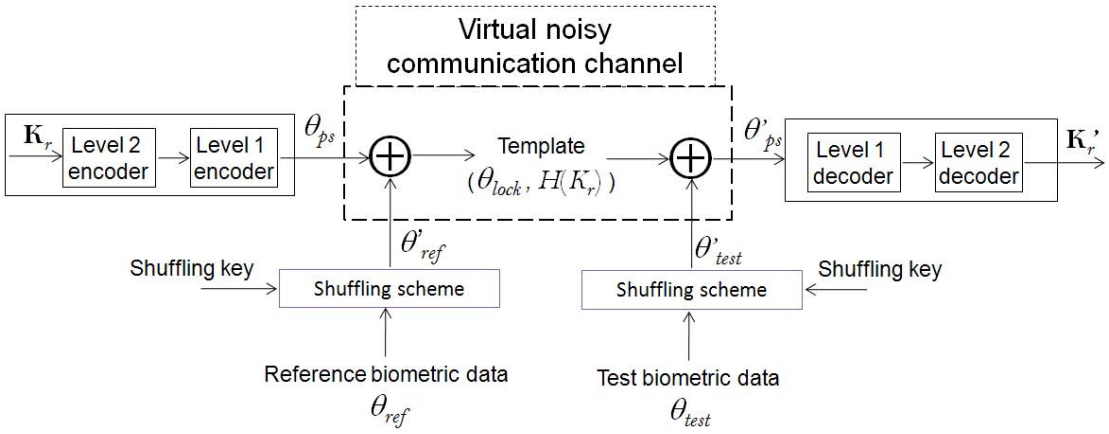


Figure 6.2: The proposed hybrid system for biometrics based cryptographic key regeneration. It combines the shuffling based cancelable biometric scheme with the fuzzy commitment based key regeneration model.

storing the template on a personal smart card. But, this solution limits the job of the attacker (who already has the compromised biometric data) to stealing the smart card. A better solution is to combine the biometric data with a password so that the attacker has to break the password every time the template is revoked. In order to achieve better revocability, we propose a hybrid scheme by modifying the Hao et al. scheme [54]. The cancelable scheme described in Chapter 4 is first applied on the biometric data to induce revocability in the system. Thus, all the advantages of the shuffling scheme as described in Section 4.1.2 are inherited by this key regeneration scheme. A schematic diagram of this proposed system is shown in Fig. 6.2.

The shuffling scheme in this system can be replaced by another cancelable scheme provided that the error correcting codes are adapted according to the error distribution in the cancelable biometric data. The factors affecting the selection of ECC are described in the next section.

6.1.2 Finding Appropriate Error Correcting Codes

The selection of the ECC depends on two criteria: (1) the key $\mathbf{K}_r$ should have a length as big as possible, and (2) ideally, the error correction capacity of the ECC should be such that it can correct *all* of the intra-user variations in the biometric samples. The first criterion indicates that the ECC should have a high code rate. Code rate of an

error correcting code is defined as the ratio of the input size to the output size of the ECC. The second condition requires that the error correction capacity is at least as high as the amount of intra-user variations that need to be corrected.

Generally, in order to increase the error correction capacity, the error correcting codes add a large number of parity symbols to the input data which decreases the input data size. These two conditions described above are contradictory and therefore, it is particularly difficult to satisfy them simultaneously with a single ECC. Hence, two error correcting codes are used in concatenated mode such that high code rate as well as high error correction capacity is achieved.

As far as biometrics based key regeneration systems are concerned, this structure was first used by Hao et al. [54] in a fuzzy commitment construct. They used Hadamard codes to correct bit-level errors in Level-1 whereas, Reed-Solomon (RS) codes were used to correct block-level errors in the Level-2. The selection of these ECC was done related to the errors that occur during different iris acquisitions. Iris data can contain random errors due to camera noise, image distortion, etc., and burst errors caused by eye-lids, eye-lashes, specular reflections, etc. The Hadamard codes correct random errors while the RS codes correct the burst errors. An introduction to the Hadamard codes is already given in Section 5.3.1. The other type of ECC, the Reed-Solomon codes, are briefly described in the following subsection.

Reed-Solomon (RS) Codes

Reed-Solomon codes [79] are *nonbinary cyclic* codes with symbols made up of m-bit sequences, where m is any positive integer having a value greater than 2. $RS(n_s, k_s)$ codes on m-bit symbols exist for all n_s and k_s where,

$$0 < k_s < n_s < 2^m + 2. \tag{6.1}$$

Here, k_s is the number of input data symbols being encoded, and n_s is the number of symbols in the output codeword. For the conventional $RS(n_s, k_s)$ code,

$$(n_s, k_s) = (2^m - 1, 2^m - 1 - 2t_s), \tag{6.2}$$

where $2t_s$ is the number of parity symbols and this code can correct at most t_s symbol errors. The RS-codes are systematic codes. An error correcting code is said to be sys-

tematic if its output contains the input in original form. In other words, the output of the RS codes, i.e., the n_s symbols is composed of the k_s original data symbols appended with the $2t_s$ parity symbols. More details about Reed-Solomon code can be found in [79, 119].

As mentioned earlier, one of the factors affecting the selection of the ECC is the amount of errors that need to be corrected. This amount is different for different biometric modalities and also different for different experimental conditions, and hence, the ECC need specific adaptations.

In the next section, we propose the specific adaptations of the generalized scheme described above for iris biometrics. The experimental results and security analysis of the proposed system are also presented. Later, the adaptations for face biometrics along with experimental results and security analysis are presented in Section 6.3.

6.2 Adaptations of the Proposed Generalized Key Regeneration Scheme for Iris Biometrics

6.2.1 Iris Data and Their Noisiness

As far as the iris data is concerned, the iris codes obtained from the iris images contain two types of errors: random errors caused by camera noise, iris distortion, etc., and burst errors resulting from specular reflections, eyelids occlusions, eyelashes, etc. The random errors are generally distributed over the total size of the iris image. Therefore, when the binary iris code is extracted from such an image, the errors are spread over the whole length of the iris code. On the other hand, burst errors generally occur in a localized manner. Therefore, they are concentrated and result in error burst in iris codes.

Hao et al. [54] proposed a concatenated error correction scheme in which Hadamard codes are used to correct the random errors and RS codes to correct the burst errors. We used the same ECC configuration in our proposed scheme because it suites the nature of errors in the iris data.

The limitation of this ECC configuration is that the Hadamard codes can correct only up to 25% errors. In [54], the authors used a private database having small intra-user Hamming distances. The images in this database are acquired with a camera

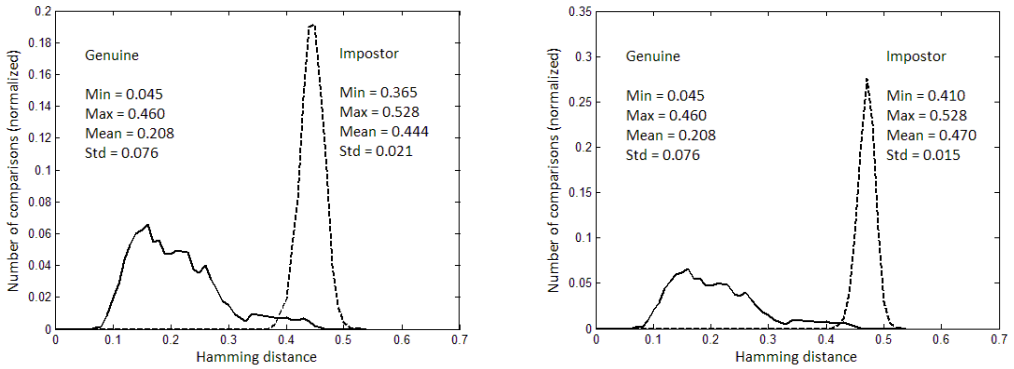
at a fixed measurement distance from the eye. These acquisition conditions may seem to be too restrictive to the user. As a result of these restrictions, the iris data in this database is far less noisy than in some other publicly available iris databases such as the CBS [103] and the NIST-ICE [101] database. The mean of the genuine Hamming distance distribution for the database in [54] is 0.12 while that for the CBS-BiosecureV1 database is 0.21. Clearly, the database we are working with is much more noisy as compared to the one in [54]. We need to correct nearly 35% errors on our databases. The Hadamard codes having up to 25% error correction capacity are not sufficient to correct these high amount of variations. When we applied the Hao et al. [54] scheme on the CBS-database, the false rejection rate is very high (see Table 6.1, page–95). In fact, the minimum FRR that we could obtain is 13.70% at 0% FAR with 42-bit keys. Bringer et al. [26] have also shown that the system in [54], when used on the NIST-ICE database, results in 10% FRR at 0.80% FAR.

Therefore, we propose some adaptations to cope with this high error rate in the following subsection.

6.2.2 Adapting the ECC from Hao et al. [54] to Correct Higher Amount of Errors

A requirement for a biometrics base verification system is that the genuine users are distinguished from impostors. In fuzzy commitment based schemes, this is achieved by adjusting the error correction capacity. During the development of the proposed iris based system, we used the CBS database [103] in order to find out the optimal error correction capacity for iris data.

By observing the Hamming distance distribution plots for genuine and impostor comparisons on the development data sets (CBS-BiosecureV1 and CBS-CasiaV2) shown in Fig. 4.2 and C.1 and reproduced here in Fig. 6.3, we found out that, the error correction capacity should be set to $\approx 35\%$. In the two level error correction scheme shown in Fig. 6.2, majority of error correction is performed by the Level-1 codes (Hadamard codes in case of iris) at the background level. Therefore, the FRR can be decreased by increasing the background error correction capacity of Hadamard codes. Hadamard codes can correct only up to 25% errors, and unfortunately, there is no method in literature to alter the error correction capacity of the Hadamard codes. Therefore, we propose



(a) Baseline iris biometric system on CBS-BiosecureV1 (development) data set

(b) Baseline iris system with shuffling on CBS-BiosecureV1 (development) data set

Figure 6.3: Normalized Hamming distance distributions for genuine and impostor comparisons on the development data set (CBS-BioSecureV1) [103]. These plots are the same as shown in Fig. 4.2.

a zero insertion scheme which makes it possible to correct more than 25% errors using Hadamard codes. It is achieved by uniformly inserting certain amount of zeros in the iris codes. The zeros are inserted in the iris codes at the enrollment as well as verification time. The positions where the zeros are inserted are fixed and the bits in these locations cannot cause any errors. Therefore, when such iris code is divided into blocks for error correction, each block contains certain amount of zeros and some iris code information bits. Since, only the iris code bits can cause errors, the number of errors that can occur in an iris code block decreases. If this amount of errors is less than the error correction capacity of the Hadamard codes, these errors can be corrected.

The zero insertion scheme actually alters the error distribution in the iris codes itself by inserting some similarity in both the iris codes to be compared. Inserting similarity does not change the total amount of errors in the comparison data but the amount of errors present per block decreases. Suppose there are p errors in an n-bit block. If q zeros are uniformly inserted in this block, there will be p errors in $(n+q)$ bits. The error ratio decreases from p/n to $p/(n+q)$. If $p/(n+q)$ is less than 25%, then these p errors can now be corrected by the Hadamard codes. The amount of zeros to be inserted is selected such that the error correction capacity matches the requirement (e.g., 35% error correction in our case). The insertion of zeros into the iris code increases

its length thereby increasing the number of blocks of Hadamard codes. This allows us to obtain keys of higher length.

The length of the key $\mathbf{K}_r$, denoted by $\| \mathbf{K}_r \|$, depends on the length of the biometric data being XOR*ed* with the pseudo code θ_{ps} and the error correction capability t_s. Thus, increasing the iris code length by zero insertion allows to select longer key $\mathbf{K}_r$.

$$n_s = \frac{\text{Length of biometric data being XOR}ed}{2^{m-1}}, \tag{6.3}$$

$$\| \mathbf{K}_r \| = m \times (n_s - 2t_s). \tag{6.4}$$

Here, n_s is an integer. Hence, the length of the biometric data to be XOR*ed* must be an integer multiple of 2^{m-1}. If it is not, some trailing bits are discarded from the biometric data to match the lengths.

Since the Hadamard codes and RS codes are operating in concatenated mode, their dimensions must also be compatible. Hence we set the $m = k + 1$ where m is the number of bits in each RS code symbol and k is the order of the Hadamard code. In other words, the size of Hadamard code input is equal to the number of bits in each RS code symbol.

6.2.3 Experimental Results of the Iris Based Key Regeneration System

Experimental Setup

The experimental setup for performing the experiments on iris modality is the same as employed in previous chapters. The databases and associated protocols are detailed in Section A.2.1. The system parameters, mainly the error correction capacity, are tuned on the development database (CBS). The system is further evaluated on the evaluation database (NIST-ICE).

Experimental Results

Before presenting the results of the proposed system, some preliminary results on the development database are presented. Since the proposed system is based on the Hao et al. [54] system, we re-implemented it and tested it on the development database

(CBS). The error correction capacity in this system is set at 25%. But, as can be seen from the Hamming distance distribution plots in Fig. 6.3, a significant portion of the genuine Hamming distance distribution curve is present above the 25% Hamming distance mark. This leads to a higher number of false rejections which is evident from the Table 6.1. This problem can be solved by allowing a higher number of errors to be corrected.

Table 6.1: Results for the Hao et al. [54] system on the CBS database; $n_s = 37$, $m = 6$, effective iris code length=1,184 bits; Key length is in bits; FAR and FRR values are in %.

t_s	Key Length	CBS-BiosecureV1		CBS-CasiaV2	
		FAR	FRR	FAR	FRR
1	210	0	77.30	0	97.67
2	198	0	68.05	0	94.95
3	186	0	60.27	0	91.20
4	174	0	53.97	0	86.78
5	162	0	48.28	0	80.67
6	150	0	43.72	0	74.22
7	138	0	38.75	0	67.33
8	126	0	34.45	0	60.57
9	114	0	30.50	0	53.32
10	102	0	26.57	0	46.88
11	90	0	23.18	0	41.30
12	78	0	20.13	0	35.68
13	66	0	17.82	0	31.30
14	54	0	15.25	0	27.02
15	42	0	13.70	0	22.88

Table 6.2 shows the results of a modified Hao et al. scheme in which the iris codes are shuffled before using them in the key regeneration framework. A comparison between Table 6.1 and 6.2 shows that applying only the shuffling scheme on the iris codes and using them in the Hao et al. scheme does not improve the performance.

The reason for this no-change can be explained as follows: in classical biometric systems, it is observed that at higher values of the verification threshold, the FRR reduces, but at the expense of increase in FAR. As far as the key regeneration system is concerned, the error correction capacity of the error correcting codes used in the system functions as a threshold. If the amount of errors in the iris codes, which is analogous to the Hamming distance, is less than the error correction capacity, the ECC correct those errors and the key is regenerated successfully which is also an indication of successful

Table 6.2: Results on CBS database: shuffling scheme is applied to the iris codes before using them in the Hao et al. scheme [54]; $n = 37$, $m = 6$; effective iris code length=1,184; Key length is in bits; FAR and FRR values are in %.

t_s	Key Length	BiosecureV1		CasiaV2	
		FAR	FRR	FAR	FRR
1	210	0	72.5	0	96.6
2	198	0	62.82	0	93.2
3	186	0	55.95	0	88.47
4	174	0	50.53	0	83.18
5	162	0	45.25	0	76.58
6	150	0	40.45	0	70.32
7	138	0	36.37	0	63.65
8	126	0	32.55	0	57.33
9	114	0	28.7	0	50.87
10	102	0	25.47	0	45.5
11	90	0	22.78	0	40.18
12	78	0	20.08	0	35.45
13	66	0	17.3	0	30.65
14	54	0	15.32	0	26.9
15	42	0	13.53	0	23.15

user verification.

The difference between the experiments reported in Table 6.1 and 6.2 is only the use of shuffling. The error correction capacity in both these tests is the same. Applying the shuffling scheme increases only the impostor Hamming distances without changing the genuine Hamming distances. Therefore, using a shuffling scheme can decrease the false acceptance rate but the false rejection rate cannot change.

In order to improve the verification performance, the verification threshold, which is the error correction capacity in case of key regeneration system, should be increased. Since, the application of the shuffling scheme only increases the impostor Hamming distance, it gives us an opportunity to increase the threshold without increasing the FAR.

The zero insertion scheme described in Section 6.2 is used in order to increase the background error correction capacity. We experimentally determined the amount of zeros to be inserted in the iris codes by performing a number of tests on the development database. One important aspect of the zero insertion scheme is that the zeros should be inserted uniformly throughout the iris code length. At least, there should be roughly equal number of added zeros in each block (to be decoded by Hadamard codes). Consid-

ering the fact that the 1,188-bit iris code is obtained by using six filters applied at 198 points of the normalized iris image ($198 \times 6 = 1,188$), we added two zeros after every 3-bit block of the iris code. This increases the length of the iris code from 1,188 bits to 1980 bits. The Hadamard code output is a 32-bit block. Since the increased length of the iris code is not an integer multiple of 32, the iris code is truncated to 1,952-bits by discarding the last 28 bits. Considering the added zeros, there can be either 20 or 21 actual iris code bits that can cause errors in a 32-bit Hadamard code block. Inherently, this Hadamard code can correct 7-bit errors in a 32-bit block out of which the added zeros cannot cause errors. Thus, the effective maximum error correction capacity is $7/20 = 35\%$.

Table 6.3 shows the detailed results obtained for the proposed iris based crypto-bio key regeneration scheme on the CBS database. Comparison between the Tables 6.1, 6.2, and 6.3 clearly shows the improvement in verification performance as a result of zero insertion and shuffling. For example, using the Hao et al. scheme [54] on CBS-BiosecureV1 data set (Table 6.1), a 198-bit key can be generated with 0% FAR but at an extremely high (68.05%) value of FRR. Adding the shuffling scheme to the Hao et al. scheme also results in similar outcomes (FRR = 62.82%). Clearly, these values of FRR are too high for practical purposes.

But, when the zero insertion scheme is applied along with the shuffling scheme, the FRR decreases drastically without a significant increase in the FAR. For example, a 198-bit key can be obtained with this scheme at FAR = 0.32% and FRR = 3.25% when tested on the CBS-BiosecureV1 database.

Using the parameters obtained from the tests on the CBS (development) database (number of zeros to be inserted and values of t_s), the proposed system is tested on the evaluation database – NIST-ICE [101]. These results are reported in Table 6.4.

With the proposed iris based key regeneration scheme, we succeed to obtain 198-bit keys at 0.055% FAR and 1.04% FRR on the ICE-Exp1 (right eye experiment). For the ICE-Exp2, keys with the same length are obtained at 0.13% FAR and 1.41% FRR. At low FAR range, e.g., FAR=0.0008%, we succeed to obtain 234-bit keys at 2.48% FRR.

Though the lengths of these keys are quite high, the entropy significantly reduces due to the redundancy added by the ECC. Theoretical analysis of the key entropy

Table 6.3: Results for the proposed iris based key regeneration system on CBS database [103] (development); shuffling is applied on iris codes and 2 zeros added after every 3 bits; $\approx 35\%$ error correction; $n_s = 61$, $m = 6$; effective iris code length=1,952; key length is in bits; FAR and FRR values are in %.

t_s	Key Length	CBS-BiosecureV1		CBS-CasiaV2	
		FAR	FRR	FAR	FRR
1	354	0	30.53	0	49.70
2	342	0	22.12	0	35.78
3	330	0	16.37	0	26.27
4	318	0	12.88	0	19.25
5	306	0	10.65	0	14.82
6	294	0	8.98	0	11.70
7	282	0	8.35	0	9.52
8	270	0	7.27	0	7.32
9	258	0	6.60	0	5.97
10	**246**	**0**	**5.87**	**0**	**4.85**
11	234	0	5.28	0.02	3.77
12	222	0.02	4.57	0.08	3.13
13	210	0.03	3.97	0.12	2.12
14	**198**	**0.32**	**3.25**	**0.52**	**1.57**
15	**186**	**0.70**	**2.67**	**1.15**	**1.07**
16	174	1.38	2.00	2.50	0.63
17	162	2.77	1.43	5.30	0.30
18	150	5.55	1.00	9.68	0.25
19	138	9.57	0.63	17.52	0.15
20	126	16.18	0.42	28.20	0.05
21	114	24.42	0.23	41.32	0.03
22	102	36.22	0.13	56.72	0

is presented in the following subsection.

6.2.4 Security Analysis of the Iris Based Key Regeneration System

In Section 3.3, we have defined two ways of security evaluation of the crypto-biometric systems: (i) theoretical analysis to estimate the entropy of the key, and (ii) experimental security analysis for the stolen biometric and stolen key scenarios. This analysis is presented in the following subsections.

Theoretical Security Analysis of the Iris Based Key Regeneration System

Since the crypto-bio keys obtained using the system described in this section are to be used for cryptographic purposes, it is required to estimate the theoretical

Table 6.4: Results for the proposed iris based key regeneration system on the NIST-ICE database [101]; shuffling is applied on iris codes and 2 zeros added after every 3 bits; ≈ 35% error correction; $n_s = 61$, $m = 6$; effective iris code length=1,952; key length is in bits; FAR and FRR values are in %.

t_s	Key Length	ICE-Exp1		ICE-Exp2	
		FAR	FRR	FAR	FRR
1	354	0	49.39	0	52.99
2	342	0	33.26	0	37.74
3	330	0	24.26	0	25.78
4	318	0	16.50	0	20.10
5	306	0	12.67	0	16.25
6	294	0	10.31	0	11.81
7	282	0	7.29	0	9.42
8	270	0	5.93	0.00009	7.77
9	258	0	4.61	0.0002	6.26
10	246	0.0005	3.63	0.0016	4.54
11	**234**	**0.0008**	**2.48**	**0.0022**	**3.49**
12	222	0.0056	2.13	0.033	3.05
13	210	0.021	1.46	0.018	2.12
14	**198**	**0.055**	**1.04**	**0.13**	**1.41**
15	**186**	**0.096**	**0.76**	**0.21**	**1.09**
16	174	0.33	0.69	0.31	0.94
17	162	0.95	0.47	3.14	0.61
18	150	1.81	0.38	5.62	0.46
19	138	11.37	0.26	7.62	0.39
20	126	11.77	0.15	14.77	0.29
21	114	14.20	0.13	18.38	0.20
22	102	21.99	0.11	30.80	0.13

entropy of such keys. Originally, the key $\mathbf{K}_r$ is randomly generated. But, at the time of regeneration, it is obtained by providing the biometric data along with the shuffling key (or password). Though, ideally, the entropy of the key $\mathbf{K}_r$ is equal to its length, the entropy decreases because of the redundancy added by the error correcting codes to cope with the biometric data variations.

As shown in Hao et al. [54], the sphere packing bound [79] can be used to roughly estimate the number of brut force attempts required for an attacker to guess the key $\mathbf{K}_r$ correctly. If N is the number of information bits being XOR*ed* with the θ_{ps}, and P is the fraction of this information corresponding to the error correction capacity (i.e., $P = N \times$ error correction capacity), the entropy can be estimated using the equation (6.5) as:

$$H \approx \log_2 \frac{2^N}{\binom{N}{P}} \quad \text{bits.} \tag{6.5}$$

Though the iris code used in our system is 1,188-bit long, all these bits are not independent but there are some correlations. Daugman [37] has given a procedure to estimate the degrees-of-freedom of the iris code based on the statistical data obtained from the impostor comparisons. This procedure takes into account the mean (μ) and standard deviation (σ) of the impostor Hamming distance distribution. Following the same procedure, the degrees-of-freedom in the iris codes can be found as:

$$\text{Degrees-of-freedom} \quad N = \mu(1-\mu)/\sigma^2. \tag{6.6}$$

There are two approaches an attacker can follow to obtain the cryptographic key: (1) by providing the un-shuffled biometric data and the shuffling key (or password) separately, or (2) by providing the shuffled biometric data.

For the un-shuffled iris data, the average degrees-of-freedom is found to be $N = 556$. If the attacker selects the first approach (which is to provide the iris data and password separately), then using the equation 6.5, the entropy contribution from the iris code is $H_i = 41$-bits. In our scheme, we suggest that the shuffling key should be randomly generated and then protected by using a password. If an 8-character password is generated randomly, it can have up to 52-bit entropy [31]. Therefore, adding this 52-bit entropy of the password to the 41-bit entropy contributed by the iris, the total entropy of the system in this approach is $H_{total} = 41 + 52 = 93$-bits.

If the attacker chooses the second approach (i.e., by directly providing the shuffled data), he needs to guess the shuffled biometric data within a Hamming distance corresponding to the error correction capacity. In this case, the degrees-of-freedom in the shuffled data needs to be calculated. From the experimental data, the mean is found to be $\mu = 0.47$ and the standard deviation is $\sigma = 0.014$. Applying the equation 6.6, the degrees-of-freedom is calculated to be $N = 1,270$. But, there are only 1,172 information bits in the data being XOR*ed*. Therefore, we consider the degrees-of-freedom to be $N = 1,172$. Then using the equation 6.5, the entropy of the system is calculated which is equal to 83-bits.

Thus, theoretically, in case of iris, it is comparatively easier for an attacker to provide the shuffled data directly instead of obtaining the biometric data and shuffling key (or password) separately. Therefore, in summary, the estimated entropy of the keys obtained using the iris based system is 83 bits.

Experimental Security Analysis of the Iris Based Key Regeneration System

We carried out the experimental security analysis of the proposed iris based key regeneration system by conducting two experiments in two extreme scenarios: (a) stolen biometric scenario and (b) stolen key scenario.

In the stolen biometric scenario, it is considered that the impostor has the biometric data for all the genuine users. Therefore, he provides the stolen biometric data along with a random shuffling key. The verification performance (which is actually the false acceptance rate) in this case is reported in Table 6.5.

Table 6.5: Experimental security analysis in terms of FAR (in %) of the proposed iris based crypto-bio key regeneration scheme. Stolen biometric – when iris images of all the genuine users are stolen; Key length is in bits; Stolen key– when the shuffling keys of all the users are stolen.

t_s	Key Length	ICE-Exp1		ICE-Exp2	
		Stolen biometric	Stolen key	Stolen biometric	Stolen key
10	246	0	1.66	0	1.32
11	234	0	2.11	0.01	1.64
14	198	0.04	14.06	0.14	12.11
15	186	0.05	23.80	0.06	20.56

In the other extreme security scenario, stolen key scenario, it is assumed that the impostor has obtained the shuffling key for all the genuine users. In this case, the impostor provides his biometric data along with the stolen shuffling key of the genuine user. The false acceptance rate in this scenario is reported in Table 6.5.

The values of FAR in both of these cases are higher than that for the random impostor case when neither biometric nor the key is stolen. If the system does not employ shuffling, the FAR is higher than the FAR in the stolen biometric case and is equal to the FAR in the stolen key scenario. Thus, the use of shuffling prohibits an impostor who has stolen the biometric data while in the stolen key scenario, the system performs as good as the one without using shuffling.

6.2.5 Reported Attack on the Iris Based Key Regeneration System and a Proposed Solution

Stoianov et al. [121] have reported an attack on the proposed key regeneration system using the ECC statistics. In the proposed system, we insert certain amount of zeros at fixed locations in order to increase the error correction capacity of the Hadamard codes. Generally, in a 32-bit Hadamard code block, 12 bits are the added zeros.

Stoianov et al. [121] quote:

> "by knowing the locations of only 7 zeros for each 32-bit block, it is possible to reconstruct the entire 198-bit key. For that, the attacker finds the nearest codewords that have the same bits in the known locations (i.e. where the zeros are inserted)."

This attack is possible because the locations of zeros is always fixed and is the same for every user. In order to overcome this attack, we propose a slight modification to the proposed scheme. In the proposed scheme, the iris codes are first shuffled and then zeros are inserted into the shuffled codes. This attack can be avoided if we reverse this process, i.e., if we first add zeros to the iris code and then shuffle it. Since the shuffling key is unique to the user, the locations of the added zeros will be different for each user, thus avoiding the attack.

This attack can also be overcome if the extension of this key regeneration system proposed in Section 6.4 is used. In this extension, the reference biometric data is regenerated in order to obtain constant length keys having higher entropy. This solution involves de-shuffling of the biometric data and can resist the attack described above.

6.3 Adaptations of the Proposed Generalized Key Regeneration Scheme for Face Biometrics

6.3.1 Face Data and Their Noisiness

From the study of the key regeneration system on iris modality described in earlier section, it is clear that the error distribution in the face biometric data should be known, *a priori*, in order to design an error correction scheme for face modality. The amount of variabilities in biometric data changes with the change in experimental conditions and is also different for different modalities.

In Section A.2.2, the face databases and experimental protocols used in this work are explained in detail. We have derived a subset of the FRGCv2 face database [100] for our experiments on the face modality. This subset is composed of 250 subjects each of which has 12 images. Data from the first 125 subjects are used for development and the remaining 125 subjects are used for evaluation.

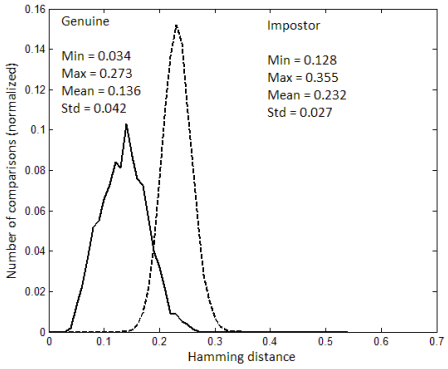
We carried out two separate experiments during development as well as evaluation: FRGC-Exp1* – where the enrollment as well as test images are captured under controlled conditions, and FRGC-Exp4* – in which the enrollment images are from controlled conditions while the test images are from uncontrolled conditions. For the FRGC-Exp1*, 3,500 genuine and 496,000 impostor comparisons are carried out while for FRGC-exp4*, 4,000 genuine and 496,000 impostor comparisons are performed.

These two experiments carried out on the FRGCv2 data set have different characteristics. The FRGC-Exp1*, in which images from controlled set are compared against those in controlled set, the amount of variations in the face codes is less. Our goal is to correct only the intra-user variations in the face data. Observing the Hamming distance distribution curves for the FRGC-Exp1* face data shown in Fig. 6.4(a), we found out that the amount of intra-user variations that need to be corrected is nearly 21%. Clearly, using Hadamard codes in this case cannot work because Hadamard codes correct nearly 25% errors which is much higher than the required capacity. Therefore, we selected BCH codes [79] as Level-1 ECC. For more details about the BCH codes, please refer to [79]. On the contrary, for the FRGC-Exp4* (Fig. 6.4(c)), the intra-user variations are nearly 30%, which is higher than the maximum error correction capacity of BCH as well as Hadamard codes. Therefore, similar to the iris case, we selected Hadamard codes along with zero insertion module to correct those errors.

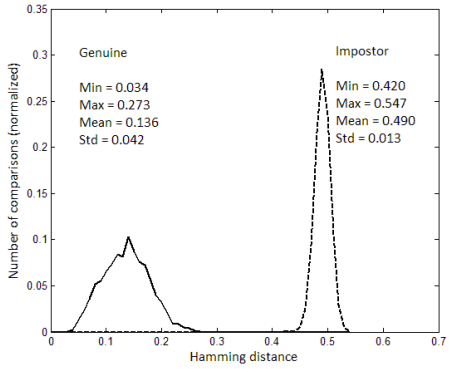
6.3.2 Selecting and Adapting ECC to the Face Data

Observing the Hamming distance distribution plots for the FRGC-Exp1* face data shown in Fig. 6.4(a), we selected BCH(511,28,111) as Level-1 ECC. This code converts an input block of 28 bits into an output block of 511 bits, and this code can correct 111 errors that can occur in the 511 bits. Thus the error correction capacity of these codes is 21.72%. As described in Section 6.1.2, the goal of using concatenated ECC is to increase the code rate which increases the length of the crypto-bio key. Therefore,

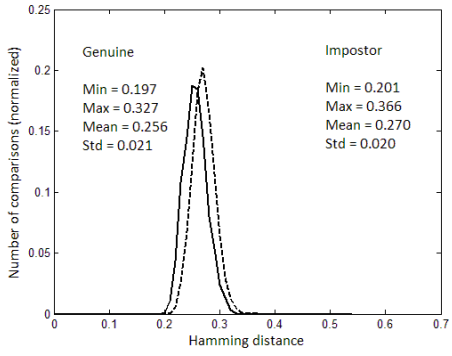
the BCH-codes and RS-codes are employed in concatenated fashion and it is required that their dimensions should be compatible with each other.



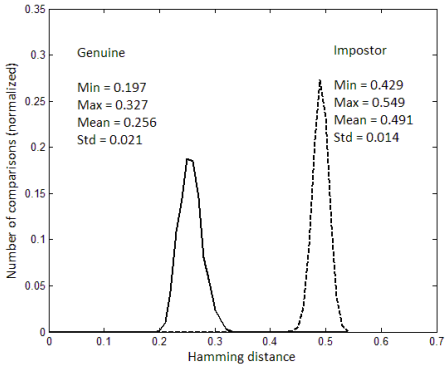
(a) Baseline face biometric system (FRGC-Exp1*)



(b) Baseline face system with shuffling (FRGC-Exp1*)



(c) Baseline face biometric system (FRGC-Exp4*)



(d) Baseline face system with shuffling (FRGC-Exp4*)

Figure 6.4: Normalized Hamming distance distributions for genuine and impostor comparisons on the NIST-FRGCv2 development data set for FRGC-Exp1* and FRGC-Exp4*. These are the same as shown in Fig. 4.7.

Ideally, the output block size of the Level-2 codes should be equal to the input block size of the Level-1 codes. Optionally, the Level-1 codes input block size can be an integer multiple of the output block size of the Level-2 codes so that multiple blocks of Level-2 codes can be combined to obtain a Level-1 codes block. For the FRGC-Exp1*, the RS codes are selected such that each RS-code block has seven bits. Four RS-codes output blocks are concatenated to form an input block of BCH codes.

During decoding, if the BCH code fails to correct the errors, it outputs a decoding failure flag. In such cases, the 28-bit output of the decoder, and therefore, all four RS-decoder input blocks, are treated as erasures. Assuming that there can be α errors and β erasures, the error correction capacity of RS codes is $2\alpha + \beta < d_{min}$, where d_{min} is the minimum distance of the RS-codes. Since we can predict the erasures for RS codes, they are operated in simultaneous error-erasure mode.

For the FRGC-Exp4*, we need 30% error correction. This is achieved by using Hadamard codes along with the zero insertion scheme. In this case, we used 7-bit RS blocks and Hadamard codes of size 64 bits. 14 zeros are uniformly inserted in every 50 bits of the face code thus forming 64-bit blocks. The Hadamard codes can correct 15 error bits in this block. Therefore the effective error correction capacity is $15/50 = 0.3$, because in each 64-bit block, there are only 50 bits that can cause errors. The locations where the zeros are inserted are fixed and hence, the bits in those positions cannot cause errors.

6.3.3 Experimental Results of the Face Based Key Regeneration System

The databases and associated experimental protocols for evaluating the proposed face based system are already introduced in Section 6.3.1. We have carried out two separate experiments on face modality: FRGC-Exp1* (controlled vs controlled) and FRGC-Exp4* (controlled vs non-controlled). The results for these experiments on the development database are reported in Table 6.6 and 6.7. Note that, the amount of variations for FRGC-Exp1* is less as compared to that in FRGC-Exp4*. Hence, the zero insertion is not needed for FRGC-Exp1*. Because of the zero insertion and the different error correcting codes, the effective sizes of the face codes are different for the two experiments. Also, the zero insertion increases the face code length which enables to select higher length keys in case of FRGC-Exp4*. For these two experiments, the shuffling scheme is employed which increases the impostor Hamming distances such that there is no impostor comparison with Hamming distance less than the error correction capacity. For instance, in FRGC-Exp1* (on development data set), the minimum Hamming distance for impostor comparisons is 0.42 whereas the error correction capacity is 21.72%. Therefore, none of the key regeneration attempt for the impostor can be

successful. This is reflected from the zero FAR in Table 6.6. Similar outcome can be observed for the FRGC-Exp4* where the minimum impostor Hamming distance is 0.429 whereas the error correction capacity is 30% resulting in zero FAR (Table 6.7).

Table 6.6: Results for the proposed system on NIST-FRGCv2 development data set, for FRGC-Exp1*; shuffling is applied on the face codes; $\approx 21.72\%$ error correction; $n_s = 24$, $m = 7$; effective face code length=3,066; key length is in bits; FAR and FRR values are in %.

t_s	**Key Length**	**FRGC-Exp1* Development**	
		FAR	**FRR**
1	**154**	**0**	**12.11**
2	140	0	6.11
3	**126**	**0**	**6.11**
4	112	0	3.51
5	**98**	**0**	**3.51**
6	84	0	2.20
7	70	0	2.20
8	56	0	0.97

As shown in Table 6.6 and 6.8, the FRR for FRGC-Exp1 is not zero. As discussed earlier, increasing the error correction capacity is equivalent to increasing the verification threshold. Therefore, the FRR can be further decreased by increasing the error correction capacity. However, increasing the error correction capacity has one drawback. As shown in Section 4.2.2, if the shuffling key is compromised, the impostor Hamming distance overlaps with that of the baseline biometric system. It means that advantage gained from using the shuffling scheme is lost and the original Hamming distance distributions of the baseline biometric system come into effect. As can be seen in Fig. 6.4, the baseline system's Hamming distance curves for genuine and impostor comparisons have a large overlap which results in high recognition errors (FAR, FRR, EER, etc.). Therefore, if the shuffling key is compromised in a system having higher error correction capacity, the FAR will increase. Therefore, we decided not to increase the error correction capacity.

Table 6.7: Results for the proposed system on NIST-FRGCv2 development data set, for FRGC-Exp4*; shuffling is applied on the face codes and zeros are inserted; ≈ 30% error correction; $n_s = 64$, $m = 7$; effective face code length=4,096; key length is in bits; FAR and FRR values are in %.

t_s	Key Length	FRGC-Exp4* Development	
		FAR	FRR
1	434	0	79.90
2	**420**	**0**	**61.00**
3	**406**	**0**	**41.65**
4	**392**	**0**	**27.93**
5	378	0	16.73
6	364	0	9.65
7	350	0	6.15
8	336	0	3.13
9	322	0	1.58
10	308	0	0.93
11	294	0	0.40
12	280	0	0.23
13	266	0	0
14	252	0	0
15	238	0	0
16	224	0	0

6.3.4 Security Analysis of the Face Based Key Regeneration System

Theoretical Security Analysis of the Face Based Key Regeneration System

The crypto-bio keys obtained from the system described in this section are to be used for cryptographic purposes. It is required to estimate the theoretical entropy of such keys. The entropy analysis for the iris based key regeneration system is already provided in Section 6.2.4. It is based on the estimation of the degrees of freedom in the biometric data and the error correction capacity of the ECC.

In case of face data, the two experiments (FRGC-Exp1* and FRGC-Exp4*) have different amount of error correction, and hence, their entropy estimations must be carried out separately. There is no established method to calculate the degrees of freedom in face codes. However, such a method is proposed for iris by Daugman [37]. This method estimates the degrees of freedom from the impostor Hamming distance distribution since this distribution is Gaussian in nature. Since the impostor Hamming distance distribution in case of face modality is also Gaussian in nature, for the sake of completeness, the method used for iris described earlier in Section 6.2.4 is applied to

Table 6.8: Results for the proposed system on NIST-FRGCv2 evaluation data set, for FRGC-Exp1*; shuffling is applied on the face codes; ≈ 21.72% error correction; $n_s = 24$, $m = 7$; effective face code length=3,066; key length is in bits; FAR and FRR values are in %.

t_s	**Key Length**	**FRGC-Exp1* Evaluation**	
		FAR	**FRR**
1	**154**	**0**	**11.05**
2	140	0	5.60
3	**126**	**0**	**5.60**
4	112	0	2.63
5	**98**	**0**	**2.63**
6	84	0	1.14
7	70	0	1.14
8	56	0	0.63

face.

For the FRGC-Exp1*, the number of degrees of freedom calculated with the Equation (6.6) in the un-shuffled face codes are $N = 227$ and that in the shuffled face codes are $N = 1,478$. The error correction capacity for FRGC-Exp1* is 21.72%. If the attacker attempts to guess the face code and password separately, the entropy contributed by the face codes is found to be 60-bits using the equation 6.5. The 52-bit password entropy is added to this resulting in the total entropy to be $60 + 52 = 112$ bits. If the attacker tries to guess the shuffled face data directly, the entropy is found to be 367 bits.

Similarly, for FRGC-Exp4*, there are $N = 447$ degrees of freedom in un-shuffled face codes while those in shuffled face codes are $N = 1,275$. When the attacker provides the face code and password separately, the entropy from face codes is 58-bit and adding the 52-bit entropy from password, the total entropy becomes 110 bits. Providing the shuffled face codes directly result in 157-bit entropy.

The theoretically estimated entropies of the crypto-bio keys obtained with the iris and face based key regeneration systems are summarized in Table 6.10. It is interesting to see that the estimated entropy for face is higher than for iris. The important point highlighted by these results is that if the randomness in the biometric data is less, the increase in entropy because of the shuffling is higher. For example, as shown in Table 6.10, the entropy of the iris based keys increases from 41 bits to 83 bits. But for face (FRGC-Exp1*), it increases from 60 bits to 367 bits.

Table 6.9: Results for the proposed system on NIST-FRGCv2 Evaluation data set, Exp4 (FRGC-Exp4*); shuffling is applied on the face codes and zeros are inserted; $\approx 30\%$ error correction; $n_s = 64$, $m = 7$; effective face code length=4,096; key length is in bits; FAR and FRR values are in %.

t_s	Key Length	FRGC-Exp4* Evaluation	
		FAR	FRR
1	434	0	77.18
2	**420**	**0**	**55.10**
3	**406**	**0**	**38.40**
4	**392**	**0**	**24.07**
5	378	0	15.47
6	364	0	9.35
7	350	0	5.52
8	336	0	3.20
9	322	0	1.55
10	308	0	0.93
11	294	0	0.65
12	280	0	0.25
13	266	0	0.15
14	252	0	0.05
15	238	0	0
16	224	0	0

Table 6.10: Theoretically estimated entropy for the proposed iris and face based key regeneration systems; Approach-1– The attacker provides the biometric data and password separately; Approach-2 – The attacker directly provides the shuffled data.

Approach	Iris	Face: FRGC-Exp1	Face: FRGC-Exp4
Approach-1	93 (41+52)	112 (60+52)	110 (58+52)
Approach-2	83	367	157

Experimental Security Analysis of the Face Based Key Regeneration System

As defined in the experimental methodology in Section 3.3, we carried out the security evaluation of the proposed face based key regeneration system in two extreme scenarios: (a) stolen biometric scenario and (b) stolen key scenario.

In the stolen biometric scenario, an impostor always provides the stolen biometric data of a genuine user. But because of the involvement of the shuffling scheme, the false acceptance rate in this case still remains zero.

In the other security scenario, the stolen key scenario, an impostor always provides the correct shuffling key of the genuine user. In this case, the FAR increases than in the random impostor case. The values of FAR in these two security scenarios

on the FRGC-Exp1* and FRGC-Exp4* experiments are reported in Table 6.11.

Note that, in both these cases, the FAR increases than the FAR for the random impostor case, in which, the biometric as well as the key are assumed secret. We have seen that the shuffling scheme significantly improves the verification performance of the baseline biometric system. Therefore, if this shuffling scheme is not applied, then the FAR of the system increases. This FAR without shuffling is equal to the FAR of the proposed system in stolen key scenario. Thus, the use of shuffling prohibits an impostor who has stolen the biometric data while in the stolen key scenario, the system performs as good as the one without using shuffling.

Table 6.11: Experimental security analysis of the proposed face biometrics based crypto-bio key regeneration scheme in terms of FAR in %; FRGCv2 evaluation data set; Stolen biometric – when face images of all the genuine users are stolen; Stolen key– when the shuffling keys of all the users are stolen.

FRGC-Exp1*			**FRGC-Exp4***		
t_s	**Stolen biometric**	**Stolen key**	t_s	**Stolen biometric**	**Stolen key**
1	0	7.90	1	0	11.92
3	0	16.74	2	0	25.68
5	0	27.08	3	0	41.07
6	0	39.44	4	0	57.49

6.4 Extension of the Proposed Key Regeneration Scheme to Obtain Constant Length Keys with Higher Entropy

The key regeneration scheme described in Section 6.1 can obtain keys from the biometric data. This system protects the biometric data by storing it in locked form. But, as reported in the results for this scheme in Section 6.2.3 and 6.3.3, the length of the key changes with the change in error correction capacity of the RS-codes[1]. However, cryptographic systems require cryptographic keys with fixed lengths (e.g., AES requires keys with 128, 192, or 256 bits). Therefore, in order to be compliant with such cryptosystems, the key regeneration scheme is extended in such a way that it can regenerate the reference biometric data protected in the system. This modified version of the scheme is shown in Fig. 6.5. The hash value of the reference biometric data can

[1]Strictly speaking, the hash value of the regenerated key can be used as a cryptographic key. It can be designed to have the same length irrespective of the length of the regenerated key.

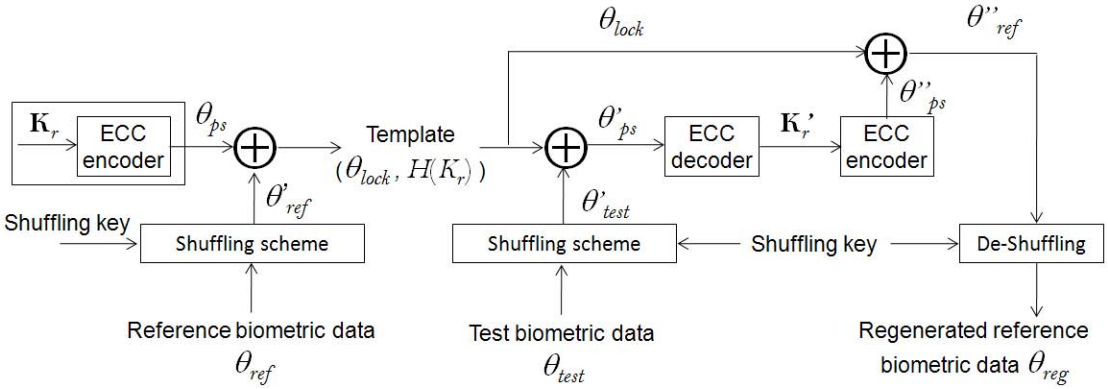


Figure 6.5: Extension of the key regeneration scheme proposed in Section 6.1 to obtain the enrollment biometric data. The hash value of this data will be used as a cryptographic key. This results in constant length keys and has higher entropy.

directly be used as a cryptographic key.

In this scheme, the regenerated key $\mathbf{K}'_r$ is re-encoded using the same error correction scheme as that used in the enrollment phase. When the errors between the enrollment and verification samples of the biometric data are within the error correction capacity of the ECC, the regenerated key $\mathbf{K}'_r$ is the same as the random key $\mathbf{K}_r$. Therefore, when the regenerated key $\mathbf{K}'_r$ is encoded with the ECC, the resultant encoded data θ''_{ps} is the same as the pseudo code θ_{ps}, because,

$$\theta_{ps} = ECC(\mathbf{K}_r), \tag{6.7}$$

and,

$$\mathbf{K}_r = \mathbf{K}'_r. \tag{6.8}$$

Therefore,

$$\theta''_{ps} = ECC(\mathbf{K}'_r) = ECC(\mathbf{K}_r) = \theta_{ps}. \tag{6.9}$$

The locked code θ_{lock} is obtained during the enrollment phase by XOR*ing* the modified biometric data (shuffled and in some cases zero padded) with the pseudo code θ_{ps}. Therefore, the modified biometric data can be regained from the locked code by XOR*ing* the θ''_{ps} with the locked code as:

$$\theta''_{ref} = \theta_{lock} \oplus \theta''_{ps} = \theta_{lock} \oplus \theta_{ps} = \theta'_{ref}. \tag{6.10}$$

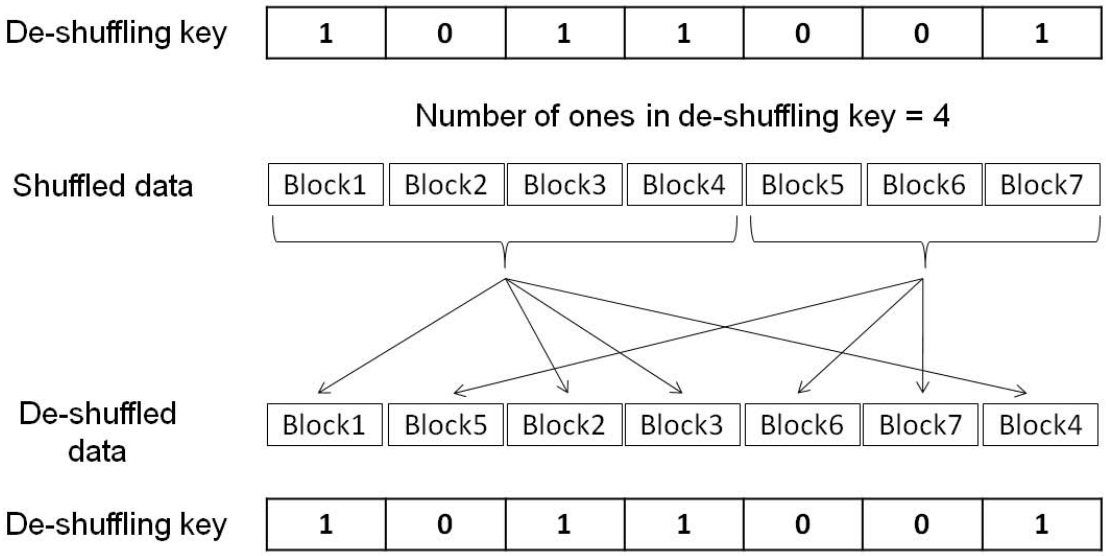


Figure 6.6: Schematic diagram of the de-shuffling process.

If the scheme involves zero insertion, those zeros are present in θ''_{ref} and their locations are known. Therefore, these zeros can easily be removed. Since the reference biometric data is shuffled during enrollment, in order to obtain it back, de-shuffling is required. The schematic diagram of the de-shuffling process is shown in Fig. 6.6.

The de-shuffling process is exactly reverse of the shuffling scheme. During shuffling, the shuffling key acts as a reference for moving the data blocks. In the same way, the de-shuffling key, denoted as $\mathbf{K}_{ds}$, functions as a reference for de-shuffling. In the shuffling process, the blocks, where the shuffling key-bit is one are sorted out in the beginning and the remaining blocks are placed in the end in the shuffled data. Hence, in order to de-shuffle that data, the shuffled data is first divided into two parts, the contents of which are dependent on the de-shuffling key. We consider the length of the de-shuffling key (which is the same as that of the shuffling key) to be L_{sh} and each block of data is L_b-bits. The number of ones in the de-shuffling key are counted and that many blocks from the shuffled data are taken into Part-1, starting from the first. The remaining blocks are taken into Part-2.

An empty array $\mathbf{D}$ of size $L_{sh} \times L_b$ is initialized. The de-shuffling key is read bit-by-bit. If a bit at position i is one, a block from Part-1 is placed at the i^{th} position in the array $\mathbf{D}$. If the bit is zero, a block from Part-2 is used instead. This process is carried

```
N = total number of 1's in the shuffling key;
count1 = 0;
count2 = 0;
for i = 1 to length of the shuffling key,
    if shuffling key(i) = 1,
        count1 = count1 + 1;
        deshuffled_data(i) = shuffled_data(count1);
    else
        count2 = count2 + 1;
        deshuffled_data(i) = shuffled_data(N + count2);
    end if
end for
```

Figure 6.7: Pseudo code for the de-shuffling algorithm.

out for all the bits in the de-shuffling key. At the end, the array **D** will contain blocks from the Part-1 and Part-2 distributed according to the de-shuffling key bit values. A pseudo code for the de-shuffling process is shown in Fig. 6.7.

As far as the shuffling and de-shuffling processes are concerned, if the de-shuffling key is same as the shuffling key used during enrollment, the de-shuffled data is exactly same as the original data. Hence, considering the equation 6.10, and assuming the shuffling key provided is correct, the regenerated reference biometric data θ_{reg} is exactly same as the reference biometric data θ_{ref}. The hash value of this regenerated reference biometric data can be used as a cryptographic key. Irrespective of the error correction capacity t_s, the regenerated biometric data has a constant length.

This extended scheme completely relies on the fact that the regenerated key $\mathbf{K}'_r$ is same as the original random key $\mathbf{K}_r$. If these two keys are not the same, the system cannot recover the correct reference biometric data. Therefore, the result tables presented in Section 6.2.3 and 6.3.3 also apply to this extended scheme.

The security analysis provided in Section 6.2.4 and 6.3.4 can partially apply to this modified system. In that analysis, two approaches are reported which an attacker can chose: (a) to provide the biometric data and the shuffling key separately, and (b) by providing the shuffled data directly. The second approach (to provide the shuffled data directly) is not sufficient because even if the attacker regenerates the key $\mathbf{K}'_r$, he still

needs to provide the de-shuffling key (which is same as the shuffling key) to regenerate the biometric data θ_{reg}. Therefore, an attacker must obtain the biometric data and the password separately in order to regenerate the key. Thus, the entropy of the keys in this scheme is equal to that obtained by the first approach. As shown in Table 6.10, the entropy is 93 bits in case of iris. Whereas the entropy in case of the FRGC-Exp1 experiment is 112 bits and in FRGC-Exp4, it is 110 bits.

6.5 Conclusions and Perspectives

Obtaining high entropy keys using biometrics is a challenging problem. In this chapter, we propose a generalized scheme for biometrics based cryptographic key regeneration. This scheme combines the ideas from cancelable biometrics and fuzzy commitment based key regeneration. A shuffling based cancelable transformation is applied on the biometric data and then this data is used in the fuzzy commitment based scheme. The proposed key regeneration system is useful for generating high entropy keys strongly linked to the user's identity.

The generalized scheme is then adapted for two biometric modalities: iris and face. The most critical aspects of this adaptation are the selection of error correcting codes and tuning their error correction capacities according to the concerned biometric data.

Both, the iris and face based key regeneration systems, satisfy the desired requirements of a crypto-biometric system: non-repudiation, revocability, template diversity, privacy protection, and high entropy keys. Since biometric data is required for regenerating the crypto-bio keys, it is difficult for someone to repudiate. The involvement of the shuffling scheme adds revocability, and template diversity to the system. The privacy of the user is protected in this system.

The estimated entropy of the crypto-bio keys obtained from iris is 83 to 93 bits. For the face based system, the estimated entropy is 110 to 112 bits. The performance of the proposed systems is always better than the underlying baseline biometric system.

There are some issues that need to be solved regarding such key regeneration systems. For the intended operation of the proposed scheme, specific protocols need to be designed. The crypto-bio keys need to be shared between all the parties requiring

secure communication link. In Chapter 8, we propose some protocols which can share the crypto-bio keys obtained with the proposed schemes.

Another issue is the security analysis which should be extended to consider additional attacks. More importantly, an attacker may take advantage of different decoding mechanisms of the error correcting codes used in the system.

The findings of this chapter are used for designing multi-biometrics based key regeneration system in the next chapter.

Chapter 7

Obtaining Cryptographic Keys Using Multi-Biometrics

7.1 Introduction

The crypto-biometric systems described in previous chapters are all uni-biometrics based systems. They are based on information originating from a single biometric trait which puts limitations on the length and entropy of the keys that can be obtained using those systems. In order to overcome this limitation, multi-biometric techniques can be used in a key regeneration framework. Moreover, the advantages of multi-biometric systems over uni-biometric systems can also be inherited by such systems. In this chapter, we explore the possibilities of using multi-biometrics in a fuzzy commitment [64] based key regeneration scheme. Note that, the main objective of using multi-biometrics for key regeneration is to increase the key length and entropy.

In order to overcome some of the limitations of the uni-biometrics based systems, a new class of systems, called multi-biometric systems, is proposed in literature, summarized in [111]. These systems generally improve the verification performance [56]. The improvement in recognition performance of multi-biometric systems can be attributed to the increased user specific information because when more than one biometric information sources are involved, one biometric source can compensate for the limitations of other. In general, when more than one biometric traits are involved, it is more difficult to spoof the biometric system. Therefore, it can increase the security.

We have introduced the concept of multi-biometric systems in Section 1.1.2 (page–3). Depending on the sources of information combined in it, the multi-biometric system can be called as, multi-sensor, multi-sample, multi-algorithm, multi-unit (or multi-instance), and multi-modal. The information fusion can be carried out at different levels in a biometric system, such as, sensor, feature, score, decision, or rank level [111].

Multi-biometrics have many advantages over uni-biometrics, such as:

- Better matching accuracy,
- Increased feature space which can accommodate more number of individuals in a system,
- Multi-biometrics may address the problem of non-universality, e.g., in a voice recognition system, the individuals who cannot speak cannot be enrolled. But, inclusion of another biometric such as iris may enable that person to enroll.
- When multiple biometric traits are involved, it becomes more difficult for an impostor to spoof the system.

Following our classification of crypto-biometric systems in Chapter 2, we have proposed cancelable biometric systems and key regeneration systems. Using multi-biometrics in cancelable biometric systems is straightforward. The cancelable biometric system proposed in Chapter 4 creates modified feature vectors which are compared using the classical Hamming distance matcher. This is similar to a classical biometric system in which the reference and test feature vectors are compared with a distance metric. Therefore, the information fusion techniques available in literature for classical biometric systems can be directly applied to the cancelable biometric system.

As far as cryptographic key regeneration systems are concerned, these systems deliver constant bit-strings. Therefore, specific methods need to be developed in order to integrate multi-biometrics into such key regeneration systems. In this chapter, we propose a novel technique, called *FeaLingECc* (*Fea*ture *L*evel Fus*ion* through Wei*g*hted *E*rror *C*orre*c*tion). With this technique, we combine biometric information obtained from different cues into a fuzzy commitment based key regeneration system.

As described earlier in chapter 6 (page–87), the fuzzy commitment based key regeneration system treats biometric data matching as an error correction issue by con-

sidering it as a problem of data transmission through a noisy communication channel. First, a randomly generated key $\mathbf{K}_r$ is encoded using Error Correcting Codes (ECC) and the variations in the biometric data are transferred onto the encoded key. These variations, treated as errors, are corrected by the ECC to regenerate the random key $\mathbf{K}'_r$. This system does not store the biometric features or templates as in classical biometric systems. The biometric features are stored in a protected form in the crypto-biometric template (which is the locked code θ_{lock} in our case). Since there is no stored biometric template or features, classical biometric comparison cannot be performed in this system and no match score can be obtained. In fact, such systems directly output the regenerated key along with the verification result. Therefore, score level fusion cannot be applied for information fusion in key regeneration systems.

The decision level fusion is possible, but the increase in the key entropy can be a maximum of one bit. In a decision level fusion, depending on the verification results of two individual biometric systems, a combined key can be released. If the length and entropy of each of these keys is N and H bits respectively, the combined key will have a length equal to $2N$ bits but the entropy will increase by only one bit to $H + 1$. The reason behind this is the entropy is measured on logarithmic scale. If an attacker needs 2^H attempts to guess the key, then the entropy is H bits. When two such keys are present, the number of attempts increases to 2×2^H resulting in an entropy of $H + 1$ bits. Thus, the entropy increase in such case is only by one bit.

The rank level fusion works only in identification mode when the output of individual biometric systems is a subset of possible matches sorted in decreasing order of confidence. Our proposed system can work only in verification mode in order to protect the privacy, and therefore, rank level fusion cannot be used in it.

Because of these reasons, in order to achieve the goals of having high entropy and key length, the best solution is to carry out the information fusion prior to the score level, i.e., at sensor level, or feature level.

In this chapter, we explore the possibilities of using multi-biometrics in a fuzzy commitment [64] based key regeneration scheme using two different methodologies:

1. multi-unit (also called as multi-instance) type system combining information from left and right irises of a person, and

2. multi-modal type system which combines information from iris and face biometrics.

For both these systems, the information fusion is carried out at feature level. A novel technique abbreviated as *FeaLingECc* (*Fea*ture *L*evel Fus*i*o*n* through Wei*g*hted *E*rror *C*orre*c*tion) is proposed which allows to apply different weights to different modalities (or different information sources). A general description of this proposed scheme is given in the next section. The multi-unit type system is described in Section 7.3 along with its experimental evaluations and security analysis. In Section 7.4, the proposed multi-modal biometrics based key regeneration system is described. Finally, the conclusions and perspectives are given in Section 7.5.

7.2 Multi-biometrics Based Key Regeneration

The basic structure of our proposed scheme is shown in Fig. 7.1. It is based on the fuzzy commitment scheme [64]. As it is done in our uni-biometrics based key regeneration systems described in Chapter 6, there are two levels of error correction: Level-1 also called as inner level and Level-2 which is the outer level. A randomly generated key $\mathbf{K}$ is assigned to a user and is then encoded using Level-2 encoder. The output of the Level-2 encoder is then randomized with a shuffling key by applying the shuffling scheme described in Section 4.1 (page–48). The shuffled output is further encoded by Level-1 encoder. The output of the encoder is called *pseudo code* θ_{ps}. In fuzzy commitment based systems, the reference biometric data is XOR*ed* with this *pseudo code*.

In the proposed scheme, the biometric data is a combined data from two biometric cues. The biometric information fusion is carried out in the feature domain. The proposed system is based on the fuzzy commitment scheme and therefore requires the feature vectors in binary form. Assuming that the binary feature vector corresponding to the first biometric source is denoted as θ_1 and that to the second biometric source as θ_2, the reference feature code is obtained by concatenating these two feature vectors as, $\theta_{ref} = \theta_1 \| \theta_2$. This reference feature code θ_{ref} is XOR*ed* with the pseudo code θ_{ps} to obtain a locked code θ_{lock},

$$\theta_{lock} = \theta_{ps} \oplus \theta_{ref}. \tag{7.1}$$

This locked code along with the hash value $H(\mathbf{K})$ of the key $\mathbf{K}$ is the crypto-biometric template. The locked code is required for regeneration of the key $\mathbf{K}$, whereas the hash value is required to check the correctness of the regenerated key.

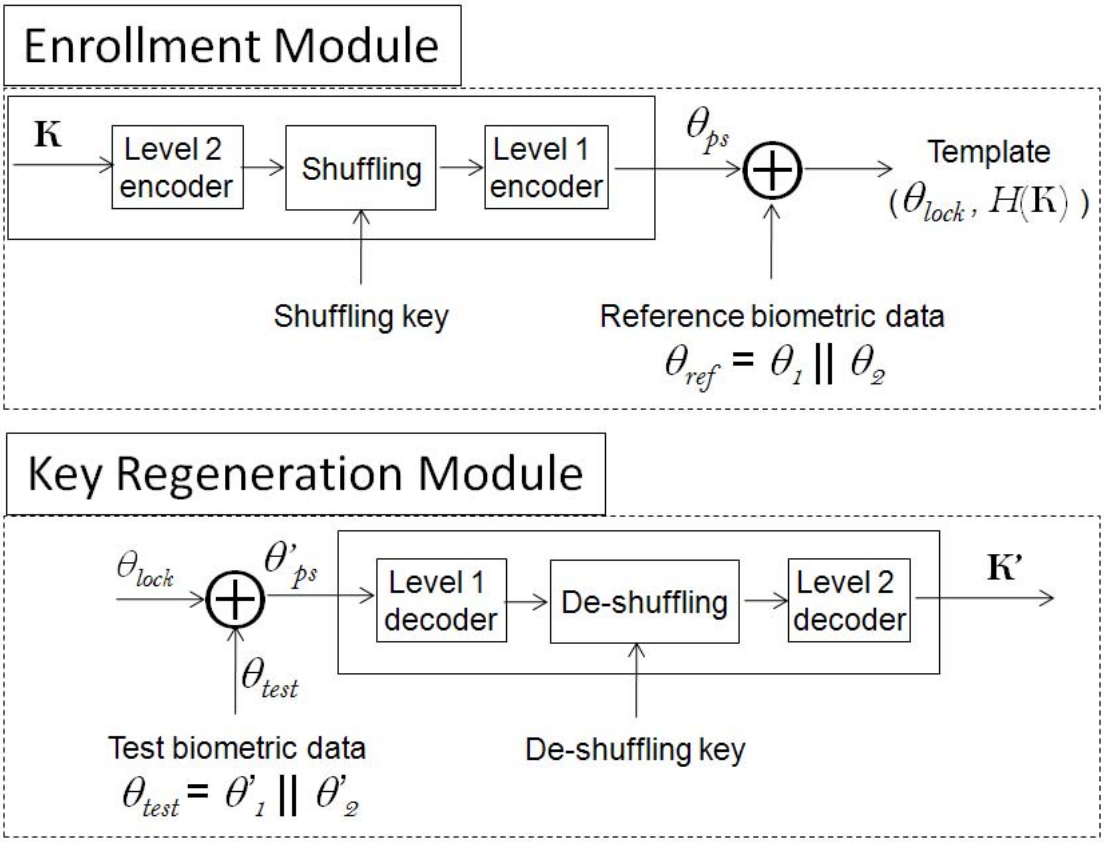


Figure 7.1: Schematic diagram showing the structure of the proposed multi-biometrics based cryptographic key regeneration scheme.

At the time of key regeneration, a multi-biometric test feature vector θ_{test} is obtained by following the procedure similar to that at the enrollment step. This test feature vector is XOR*ed* with the locked code θ_{lock} to obtain a modified version θ'_{ps} of the pseudo code. This modified version consists of the pseudo code θ_{ps} contaminated with the errors e between reference and test biometric vectors. The Error Correcting Codes (ECC) decoding scheme corrects these errors and retrieves a trial value $\mathbf{K}'$ of the random key $\mathbf{K}$. A comparison between the hash values of the original and the regenerated key

is carried out and a positive result indicates key regeneration success.

$$\begin{aligned} \theta'_{ps} &= \theta_{lock} \oplus \theta_{test}, \\ &= \theta_{ps} \oplus \theta_{ref} \oplus \theta_{test}, \\ &= \theta_{ps} \oplus e. \end{aligned} \tag{7.2}$$

$$\mathbf{K}' = ECC^{-1}(\theta'_{ps}). \tag{7.3}$$

The Level-1 error correcting codes perform majority of the error correction. These ECC correct bit-level errors occurring in blocks. If the number of errors in a block is more than the error correction capacity of the Level-1 ECC, that block is decoded incorrectly. Such incorrectly decoded blocks are further treated by the Level-2 codes. Thus, the Level-2 ECC work on block level. In order to cope with the cascading structure of the two ECC, the number of bits in each symbol of the Level-2 ECC must be the same as (or possibly an integer multiple of) the number of bits in Level-1 ECC input block.

7.2.1 *FeaLingECc* (*Fea*ture *L*evel Fus*ion* through Wei*g*hted *E*rror *C*orrection)

When feature vectors corresponding to two biometric sources are combined, it is required that the two vectors have a common representation which is not always the case. For example, fingerprint minutiae set consists of minutiae locations and orientation information, while the iris feature vector is a binary string. The minutiae set is an unordered set while the iris code is an ordered set. Therefore, the two feature vectors must be converted into a common representation. Moreover, the dimensions of the feature vectors can also be different and simply concatenating the two feature vectors may not be beneficial. The difference in the dimensionality of the two feature vectors can cause an adverse effect on the system performance. This problem is called as curse of dimensionality [111]. Therefore, in order to deal with this problem, the feature level fusion module is generally followed by a feature selection module in classical multi-biometric systems.

Moreover, one biometric trait may be performing better than the other in terms of verification performance (e.g., in general, iris performs better than face). This

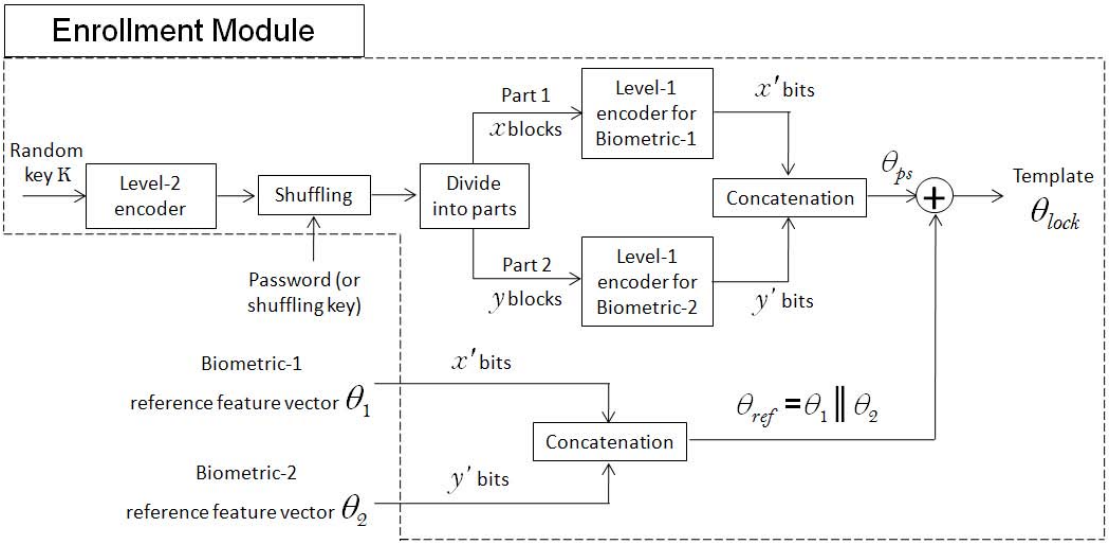
knowledge can be exploited in score level fusion systems by applying different weights to the individual biometric traits. In such systems, higher weight is given to the better performing biometric trait in the verification decision process. This kind of weighting can significantly improve the performance of multi-biometric system.

Since the match scores cannot be computed in key regeneration systems, classical score level fusion techniques cannot be used. Therefore, we propose a novel method in which the features are combined in feature domain and the error correction scheme is designed so that different weights can be applied to the individual biometric traits. This scheme also deals with the problem of curse of dimensionality. It can cope with the differences in the dimensions of individual feature vectors by carefully selecting the dimensions of the Level-1 ECC for the individual biometrics and minimize the effect of dimensions mismatch on the verification performance.

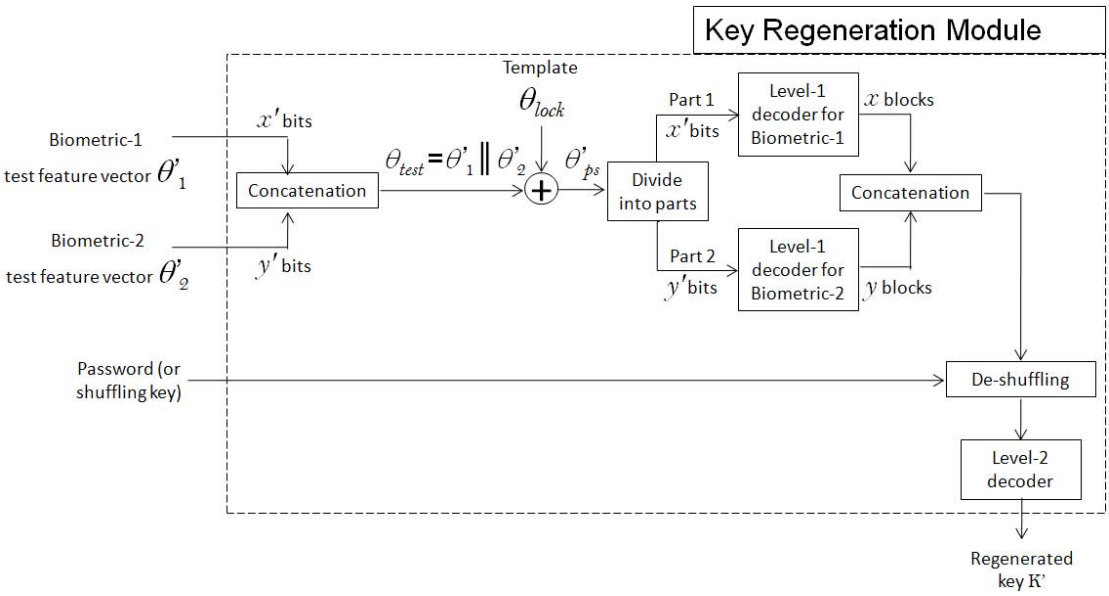
The enrollment and key regeneration modules of the proposed system are shown in Fig. 7.2(a) and 7.2(b). The error correction scheme in the proposed system consists of two levels. The Level-1 work on bit-errors occurring in blocks while the Level-2 ECC correct the block errors which left after the Level-1 ECC action. Since the amount and nature of variations in biometric data are different for different modalities, and they also depend on the acquisition conditions, we need to select different Level-1 ECC for different modalities. As in case of uni-modal biometrics based key regeneration systems described in Chapter 6, the Level-1 ECC and their error correction capacity is selected by observing the Hamming distance distributions for genuine and impostor comparisons for the corresponding trait.

The application of different weights is carried out by assigning different number of blocks of the Level-2 ECC for different biometrics. As shown in Fig. 7.2(a), the output of the Level-2 codes (which is in form of n_s blocks) is split into two parts: Part-1 which consists of x blocks and Part-2 consisting of $y = (n_s - x)$ blocks. Higher weight can be applied to the Biometric-1 by having $x > y$ and vice versa.

The x blocks of Part-1 are further encoded and combined into x' bits by the Level-1 encoder for the first biometric (Biometric-1). The y blocks of Part-2 are encoded and combined into y' bits by the Level-1 encoder of the second biometric (Biometric-2). Here, x' and y' are equal to the number of effective bits in the feature vectors of Biometric-1 and Biometric-2, respectively. The number of bits in each input block of the



(a) User enrollment module



(b) Cryptographic key regeneration module

Figure 7.2: Schematic diagram of the proposed multi-biometric based cryptographic key regeneration scheme using *FeaLingECc* (*Fea*ture *L*evel Fus*i*on through Wei*g*hted *E*rror *C*orrec*c*tion).

Level-1 encoder should be equal to the number of bits in each output block of the Level-2 encoder. Alternatively, the input block size of the Level-1 encoder can be an integer multiple of the output block size of the Level-2 encoder. Concatenation of the outputs

of the two Level-1 encoders yields the pseudo code θ_{ps}. This pseudo code is XOR*ed* with the multi-biometric reference feature vector θ_{ref} (which is obtained by concatenation of two individual feature vectors θ_1 and θ_2) to obtain the locked code θ_{lock}.

The weights are applied by changing the sizes of Part-1 and Part-2. In order to understand the concept, let's take a closer look into the error correction mechanism that takes place during the key regeneration step (see Fig. 7.3). When a multi-biometric test feature vector θ_{test} (which is obtained by concatenation of two individual test feature vectors θ'_1 and θ'_2) is XOR*ed* with the locked code θ_{lock}, the errors between the reference and test feature codes are transferred onto the pseudo code θ_{ps}. Figure 7.3 shows the process of error correction that follows.

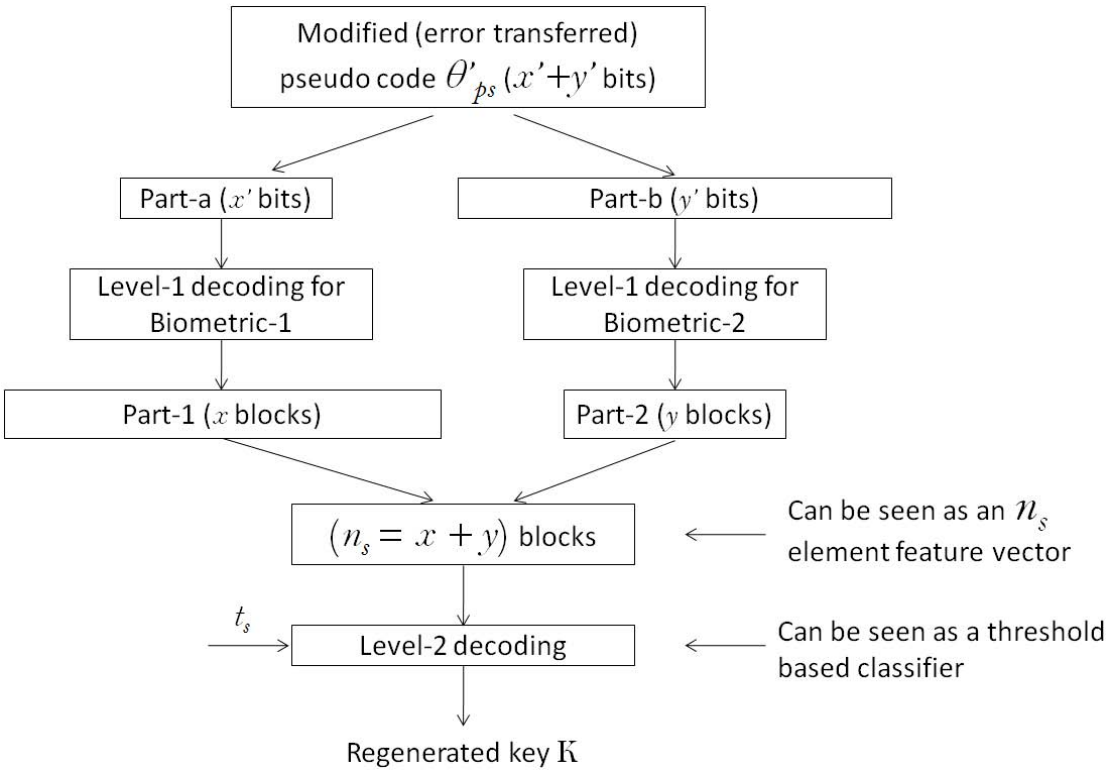


Figure 7.3: Schematic diagram showing the proposed weighted error correction process. Note that Part-b is bigger than Part-a. When Level-1 ECC are applied, this relationship changes. Part-1 becomes bigger than Part-2 which means that higher weight is applied to the Biometric-1 than Biometric-2.

The modified (error transferred) pseudo code θ'_{ps} is divided into two parts: Part-a consists of the first x' bits while the Part-b consists of the remaining y' bits. The

Level-1 decoder corresponding to Biometric-1 is applied on the x' bits to correct the bit errors caused by the Biometric-1. This process yields x blocks. Similarly, y blocks are obtained from the y' bits corresponding to the Biometric-2. These two parts are concatenated to form a single codeword which contains $n_s = (x+y)$ blocks. The Level-2 decoder corrects the erroneous blocks present in this codeword to obtain a trial value $\mathbf{K}'$ of the random key $\mathbf{K}$. The Level-2 decoder can correct up to t_s erroneous blocks where t_s is its error correction capacity. This Level-2 decoder can be seen as a threshold based classifier which operates on an n_s element vector where t_s acts as a threshold. If the number of erroneous blocks are less than or equal to t_s, the key is successfully generated and the verification result is positive. Therefore, if we set $x > y$, a higher weight will be given to the Biometric-1 than the Biometric-2 in the decision process. The condition $x > y$ (or $x < y$ if required) is achieved by properly selecting the dimensions of the Level-1 ECC. However, this selection needs to take care of the error correction capacity which depends on the Hamming distance distribution of the biometric data.

7.2.2 Adding Revocability

As discussed in earlier chapters, biometrics lack the property of revocability and can compromise user's privacy. In order to overcome these drawbacks, we proposed a shuffling based cancelable biometric system in Chapter 4 (page–47). This shuffling scheme is further employed in the key regeneration systems proposed in Chapter 6 (page–85). In a similar way, some cancelable mechanism should be used in the multi-biometrics based system. One simple option is to apply the same shuffling scheme on the two individual biometric feature vectors. In this way, revocability and privacy protection can be added to the multi-biometrics based system.

But, there is a loophole in this design. This loophole appears if the Level-2 error correcting codes used in the system (e.g., we use Reed-Solomon codes as Level-2 codes in our proposal) are of systematic nature. An error correcting code is said to be systematic in nature if the input to the code is present in its original form in the output. The output of such codes comprises of the input data appended by the parity symbols, and thus, the locations of the original data and the parity symbols is known to an attacker. In this case, the attacker can attack the biometric information corresponding only to the data blocks. Clearly, this kind of attack can suppress the

advantage gained by using multiple biometrics. The attacker may need only one set of biometric information to crack the multi-biometric system.

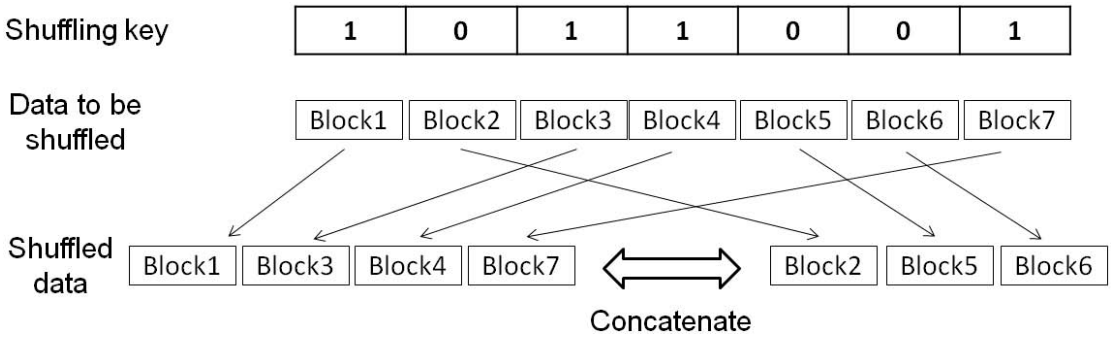
In order to overcome this drawback, we propose to apply the shuffling scheme after the Level-2 encoding instead of applying it on the biometric data. In this case, even if the Level-2 ECC are systematic, the shuffling process breaks the systematic nature of its output. The shuffled output of Level-2 ECC is further encoded with the Level-1 ECC. At the time of key regeneration, the original order of the Level-2 encoder output must be restored in order for the Level-2 decoder to function correctly. This is done by applying the de-shuffling process. For better understanding, the shuffling and de-shuffling processes are shown together in Fig. 7.4. The pseudo program code of the de-shuffling algorithm is given in Fig. 7.5.

The shuffling and de-shuffling processes are the same as described in Section 4.1.1 and 6.4 respectively. The only difference is in the data on which these processes are applied. In these earlier sections, the shuffling and de-shuffling is applied on the biometric data. Whereas, in the multi-biometrics based system, the shuffling and de-shuffling processes are applied on the output of the Level-2 encoder and the input of the Level-2 decoder, respectively.

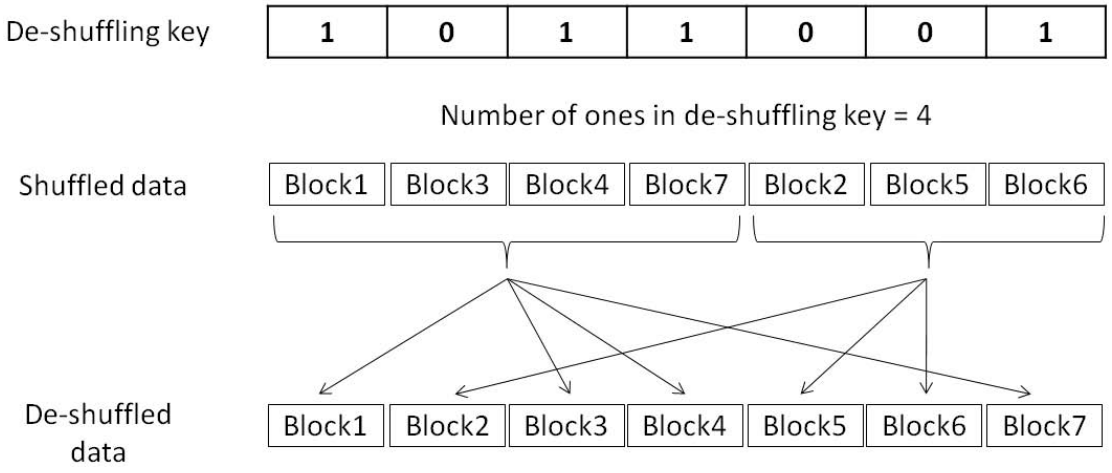
The general multi-biometrics based key regeneration scheme described in this section can be applied to a combination of two sets of biometric information. The pre-requisite for this system is that both the biometric data must be in form of binary vectors. We developed two systems based on this scheme:

- multi-unit type system that combines information from the left and the right irises of a person, and
- multi-modal type system that combines information from an iris with that from the face.

These systems are described in subsequent sections.



(a) Shuffling process



(b) De-shuffling process

Figure 7.4: A schematic diagram showing the shuffling and de-shuffling process. Note that the shuffling and de-shuffling key must be the same to recover the correct data.

7.3 Multi-unit Type Multi-biometrics Based Cryptographic Key Regeneration Scheme

7.3.1 Algorithm for Multi-unit Biometrics Based Key Regeneration

We developed a multi-unit type multi-biometrics system to obtain cryptographic keys. Feature level fusion in multi-unit type systems is comparatively less complicated than in the multi-modal type systems. The reason is that the feature sets obtained from different sources in a multi-unit system are generally similar in nature and dimensions. Our system incorporates information from left and right irises of a

```
N = total number of 1's in the shuffling key;
count1 = 0;
count2 = 0;
for i = 1 to length of the shuffling key,
    if shuffling key(i) = 1,
        count1 = count1 + 1;
        deshuffled_data(i) = shuffled_data(count1);
    else
        count2 = count2 + 1;
        deshuffled_data(i) = shuffled_data(N + count2);
    end if
end for
```

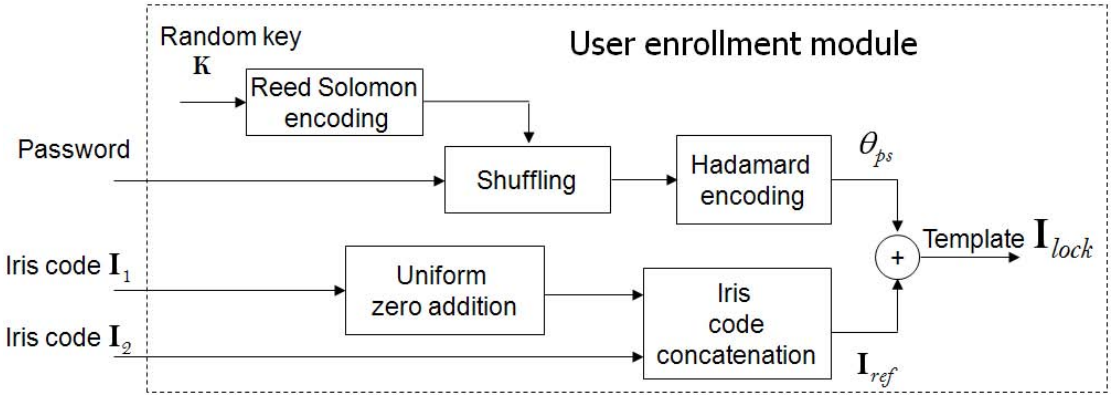
Figure 7.5: Pseudo code for the de-shuffling algorithm.

person in a fuzzy commitment based key regeneration scheme. The information fusion is carried out in feature domain using the weighted error correction approach described in previous section.

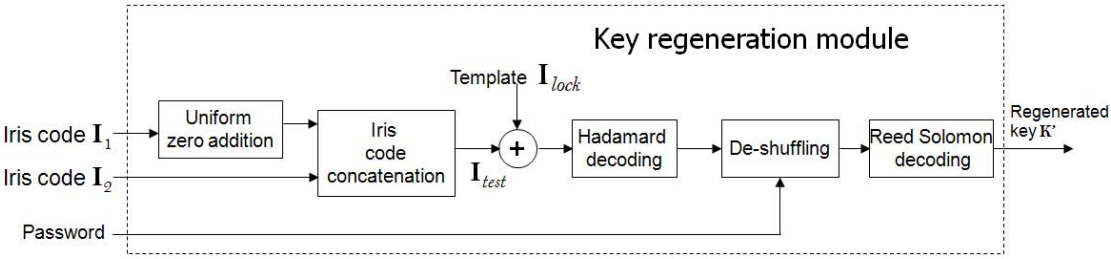
The iris codes obtained from different iris images of the same user contain variabilities which are treated as errors. As described in Section 6.2, there are two types of errors in iris codes: (1) background errors caused by the camera noise, image capture effects, etc., and (2) burst errors which are a result of specular reflections, occlusions, etc. Both these types of errors are corrected using the two level error correction scheme shown in Fig. 7.1.

The enrollment and key regeneration phases of the proposed multi-unit type system are shown in Fig. 7.6. We used Hadamard codes as Level-1 ECC and Reed-Solomon (RS) codes as Level-2 ECC for our two-iris based system. A random bit string $\mathbf{K}$ is generated and assigned to a user and is then encoded using Reed-Solomon (RS) codes, the output of which is further encoded by the Hadamard codes. The Hadamard codes correct the background errors and RS codes correct burst errors. Details about these ECC can be found in [79]. The output of the encoder is called *pseudo code* θ_{ps}. In order to cope with the cascading structure of the two ECC, the number of bits in each symbol of RS and that in the input words of Hadamard codes is set to be equal ($m = 7$). Iris codes $\mathbf{I}_1$ and $\mathbf{I}_2$ from the right and left iris images, respectively, are

concatenated to form a reference (multi-) iris code $\mathbf{I}_{ref}$. This $\mathbf{I}_{ref}$ is XOR*ed* with θ_{ps} to obtain the locked iris code template $\mathbf{I}_{lock}$. In the key regeneration phase, a test (multi-) iris code $\mathbf{I}_{test}$ is obtained similarly and XOR*ed* with $\mathbf{I}_{lock}$. These XOR*ing* operations transfer the errors in the iris codes onto the pseudo code. If the amount of errors is within the error correction capacity of the ECC, the errors are corrected by the ECC decoder part and a key $\mathbf{K}'$ is regenerated which is the same as $\mathbf{K}$. If the amount of errors is more than the error correction capacity of the ECC, $\mathbf{K}' \neq \mathbf{K}$.



(a) User enrollment phase



(b) Key regeneration phase

Figure 7.6: Schematic diagram of the proposed multi-unit type multi-biometric based cryptographic key regeneration scheme using feature level fusion, weighted error correction, and password – (a) User enrollment phase; (b) Cryptographic key regeneration phase.

The Hadamard codes can correct (up to) $2^{(k-2)} - 1$ errors in a 2^k-bit block. If a block has more than $2^{(k-2)} - 1$ errors, that block is not decoded correctly and results in an error. The second level of ECC consists of the RS codes. The output of the Hadamard decoding stage acts as the input to the RS decoder stage. The RS

codes correct the errors caused due to the wrong decoding by the Hadamard codes and generate the key $\mathbf{K}'$.

In the proposed scheme, we apply higher weights to one iris than the other by employing the weighted error correction method described in Section 7.2.1. We use a bigger number of RS blocks for one iris than the other to apply these weights. As described in Section 6.2, inserting certain amount of zeros in the biometric data can increase the error correction capacity of the Hadamard codes. Using this property, we applied the zero insertion scheme to one iris code in order to increase the error correction for it. As shown in Section 6.2.3, using the Hadamard codes without zero insertion scheme results in high false rejections but zero false acceptances (Table 6.1). Thus, the increased error correction for the first iris code helps to increase acceptances while the low error correction for the second iris code increases rejections. The combined effect of the two is the improvement in the verification performance of the key regeneration system. The most important advantage of this scheme is that the feature vector is longer than in uni-biometric based system, and therefore, we can obtain longer keys. The biometric information is also significantly more as compared to the uni-biometric systems which result in higher entropy. Additionally, it experimentally validates our proposal of weighted error correction. The experimental results of this system are reported in the next subsection.

7.3.2 Results and Security Analysis of the Multi-unit (Two-iris) Type System

In this section, we briefly describe the experimental setup, and then present the results and security analysis of the proposed multi-unit type system.

Experimental Setup

The experimental evaluation protocols for this system are described in Section A.2.3. We used the CBS database [103] for development to find out the ECC and error correction capacities. The system is then evaluated on the NIST-ICE database [101]. In the NIST-ICE database, there are 132 subjects out of which, only 112 subjects have recorded images of their both eyes. We select images of these 112 subjects for carrying out our tests. The right iris images are coupled with the left iris images for the multi-iris

tests. The first such image pair of a person is considered for enrollment and a template is registered for that person. The genuine comparisons are carried out by comparing the remaining image pairs of that subject with the enrollment template leading to 1,099 genuine comparisons. For impostor comparisons, one image pair from each of the remaining subjects is randomly selected and these image pairs are compared with the enrollment template. Thus, for each person, we carry out 111 impostor comparisons. In summary, 1,099 genuine and 12,432 impostor comparisons are carried out on the NIST-ICE database for the two-iris experiment.

Experimental Results of the Multi-unit (Two-iris) Type System

Since the proposed systems is based on an iris recognition system, it is worthwhile to report the performance of the baseline biometric system for fair comparison. Hence, such performance results are reported in Table 7.1. Note that the baseline iris system is based on OSIRISv1 with a re-implemented matching module. Classical multi-iris based biometric system is also tested in which the iris codes are simply concatenated and compared. Note that, as expected, the combination of left and right irises results in reduction in the Equal Error Rate (EER).

Table 7.1: Baseline biometric system's (which is based on OSIRISv1, see Section A.1 for details) verification performance in terms of EER in %. Single as well as two-iris tests.

CBS-BiosecureV1 (development)			**NIST-ICE (evaluation)**		
Left	**Right**	**Both irises**	**Left**	**Right**	**Both irises**
3.23	2.90	2.54	2.44	4.81	1.18

For the cryptographic key regeneration system, we first report the results for the simple feature level fusion scheme in Table 7.2. The feature level fusion in this case is by simple concatenation of two feature vectors. For the sake of comparison, the key regeneration results (for CBS database) using single irises are also reported in the same table. The shuffling scheme is not used in any of these tests. It can be observed that the minimum FRR using single iris is 7.37% with a key length of 6-bits. The combination of two irises helps reduce the FRR and also have longer keys such as 35-bit keys at 4.93% FRR. In spite of the improvement, the FRR is still much high and hence we did not carry out these tests on the ICE database.

Table 7.2: Key regeneration system results on the CBS-BiosecureV1 data set when two iris codes are combined using only feature level fusion; no weighting, no shuffling; FRR values are in %; length of key **K** is in bits; FAR is always zero for all these tests.

t_s	**Left iris**		**Right iris**		**Both irises**	
	FRR	**length(K)**	**FRR**	**length(K)**	**FRR**	**length(K)**
16	9.80	30	14.13	30	4.93	35
17	8.60	18	13.10	18	4.57	21
18	**7.37**	**6**	**12.03**	**6**	**4.27**	**7**

When the proposed *FeaLingECc* approach is used, a significant improvement is achieved that can be seen in Table 7.3. As it is done in the uni-biometrics based system, we added certain amount of zeros to the right iris code to correct higher amount of errors in it whereas no zeros are added to left iris code. The Hadamard codes operate on 64-bit blocks and there are 49 such blocks resulting in a total amount of error correction equal to 735 bits. It also allows us to obtain much longer keys with low error rates, e.g., we can have 175-bit keys at 0.38% False Acceptance Rate (FAR) and 1.64% FRR for ICE database.

Finally, the results for the key regeneration scheme with shuffling are presented in Table 7.3. These results are better than any previously published results in literature, e.g., we can generate 147-bit keys at 0.18% FRR and 0% FAR for ICE database. In our experiments, the number of blocks at the output of the RS encoder is 49. Hence we use a 49-bit shuffling key to shuffle those blocks. The shuffling key is protected by a password of eight characters. Note that, there is not much decrease in FRR due to the use of shuffling. The main improvement is in the FAR; the FAR becomes zero which means that the systems becomes more secure by using the shuffling.

Security Analysis of the Multi-unit (Two-iris) Type System

Since the proposed system is for generating cryptographic keys, it is required to analyze the security of the system in terms of key entropy. Though the key is generated randomly at enrollment time, a lot of redundancy is added by the ECC and hence its entropy is bound to decrease. We use the same approach as used by Hao et al. [54] to estimate the entropy. They used the sphere packing bound [79] to roughly estimate the number of brute force attempts required for an attacker to guess the key **K** correctly.

Table 7.3: Key regeneration system results when two iris codes are combined using the proposed *FeaLingECc* method; FAR and FRR values are in %.

t_s	Key length (in bits)	Without shuffling				With shuffling			
		CBS-Bio		NIST-ICE		CBS-Bio		NIST-ICE	
		FAR	FRR	FAR	FRR	FAR	FRR	FAR	FRR
6	259	0	8.37	0	13.28	0	8.50	0	13.74
9	217	0	5.37	0	5.19	0	5.63	0	5.46
10	203	0	4.50	0.016	3.37	0	4.60	0	3.28
11	189	0	4.10	0.06	2.09	0	4.10	0	2.09
12	**175**	**0**	**3.63**	**0.38**	**1.64**	**0**	**3.67**	**0**	**1.36**
13	161	0.10	3.40	1.49	0.55	0	3.50	0	1.00
14	**147**	**0.70**	**3.30**	**2.98**	**0.27**	**0**	**3.30**	**0**	**0.18**
15	133	1.87	3.13	10.46	0.18	0	3.03	0	0.18
16	119	6.40	2.80	15.86	0.09	0	2.37	0	0.09
21	49	84.47	0.23	91.37	0	0	0.30	0	0

Let N be the number of degrees of freedom in the data being XOR*ed* with the pseudo code θ_{ps}, and P is the fraction of this information corresponding to the error correction capacity (i.e., $P = N \times$ error correction capacity). Then the number of brute force attacks an attacker needs to carry out is estimated by the Equation (7.4) as:

$$BF \approx \frac{2^N}{\binom{N}{P}}. \tag{7.4}$$

The number of degrees of freedom can be estimated by the procedure given by Daugman [37]. The iris codes used in our experiments are 1,188 bits long. Following the procedure given in [37], we estimate the degrees of freedom in the iris codes to be 561. Collectively, in two iris codes, we have 1,122 degrees of freedom. In the weighted error correction configuration in the two-iris system, the total amount of error correction is $\approx 30\%$. If, $N = 1,122$ and $P = 0.3 \times N \approx 336$, applying the Equation (7.4), an impostor needs approximately,

$$BF \approx \frac{2^N}{\binom{N}{P}} \approx \frac{2^{1122}}{\binom{1122}{336}} \approx 2^{140}, \tag{7.5}$$

brute force calculations to successfully get the cryptographic key. Thus the entropy of the key is 140 bits, which is much higher than any other reported system.

The shuffling scheme applied in the two-iris system uses 49-bit shuffling key. This key is randomly generated and is protected by a password. We propose to use

a randomly generated 8-character password which can have 52-bit entropy [31]. The shuffling process is embedded into the error correction process and hence the individual entropies add up together resulting in a total key entropy of $140 + 49 = 189$ bits. Thus the minimum entropy of the key is:

$$\text{Entropy} = min(Length(\mathbf{K}), 189)\text{bits}. \tag{7.6}$$

Experimental security evaluation as described in Section 3.3 is also carried out. We have defined two extreme scenarios for security evaluation: (1) stolen biometric scenario – where an impostor always provides a stolen biometric sample of the genuine user, and (2) stolen key scenario – in which the impostor always provides a stolen shuffling key of the genuine user.

In the stolen biometric scenario, the system performance remains unchanged. The shuffling process prevents the impostor from being accepted when he provides the correct biometric data but a wrong shuffling key. Thus, use of shuffling completely eliminates the threat caused by compromised biometric data.

In the other security scenario, stolen key scenario, the system still has two iris codes which provide the security. The performance in this situation degrades but it is equivalent to that of the system without shuffling. Moreover, the performance degradation is only in terms of increase in FAR. The FRR remains unchanged even if the shuffling key is stolen. This is a distinct advantage of the proposed system.

The attack reported in [121] on the iris based uni-biometric key regeneration system described in the previous chapter is ineffective on the multi-iris system because there is no zero insertion for one iris. Moreover, the shuffling (and de-shuffling) process protects the system against such attack.

7.4 Multi-Modal Type Multi-biometrics Based Cryptographic Key Regeneration Scheme

7.4.1 Algorithm for Multi-modal Biometrics Based Key Regeneration

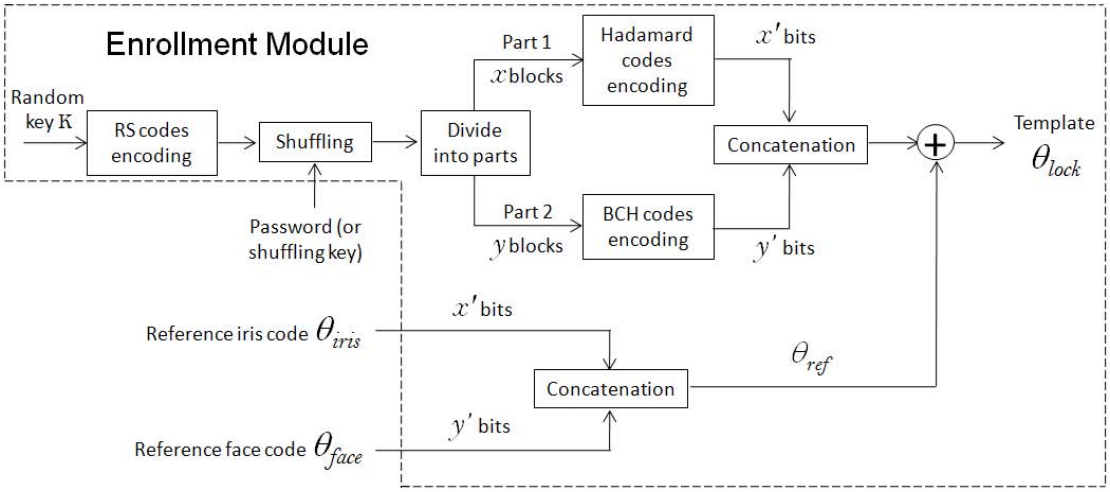
Multi-modal biometric systems, in which the information from multiple biometric traits is combined, can significantly increase the security of the system because of the added efforts required to break such systems. Combination of information from

two biometric traits in the feature domain results in increase of the length of the feature vector. Additionally, the entropy of the crypto-bio keys also increases. We adapt the FeaLingECc scheme proposed in Section 7.2 in order to combine the information from an iris and a face image of a person. The baseline biometric systems for iris and face are discussed in Section A.1. The length of the iris feature vector is 1,188 bits while that of the face feature vector is 3,200 bits. Following the notations of the general scheme described in Section 7.2, we consider iris as Biometric-1 and face as Biometric-2. In the uni-biometrics based systems described in the previous chapter, Hadamard codes were used as Level-1 ECC for iris while the BCH codes were used for face. These ECC were selected according to the Hamming distance distributions of the corresponding biometric data. In the proposed multi-modal system, we employ the same ECC. Reed-Solomon (RS) codes are used as Level-2 ECC, which are common for iris and face. The schematic diagrams of the enrollment and key regeneration phase of the proposed multi-modal biometrics based system are shown in Fig. 7.7(a) and 7.7(b), respectively.

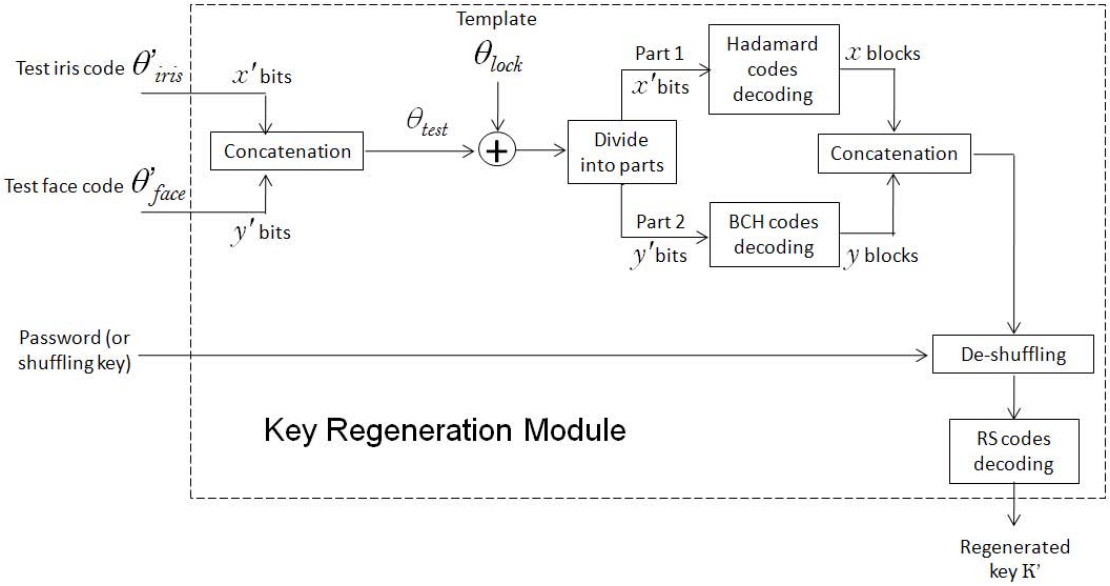
The basic functioning of this scheme is the same as described in Section 7.2. But the involvement of two different types of biometric data raise many design complications. The two biometric data (iris and face) being combined are different in nature. The amount of variabilities, which is treated as errors, is different in iris than in face. In key regeneration systems, the goal is to correct only the intra-user variabilities. The amount of such errors to be corrected is highly dependent on the concerned biometric data set. The error correction capacities for each of the biometric traits need to be set according to their respective Hamming distance distributions for genuine and impostor comparisons.

7.4.2 Experimental Setup

The database and experimental protocol used for this evaluation is detailed in Section A.2.4. We used a virtual database created from two publicly available databases: the NIST-ICE database [101] for iris, and the NIST-FRGCv2 database [100] for face. In this selected data set, there are 175 subjects having five samples each of iris and face images. For each subject, data pairs are formed containing one iris image and one face image corresponding to that subject. Thus, we have five such pairs per subject for 175 subjects. For genuine comparisons, each data pair is compared with every other data



(a) Enrollment phase



(b) Verification phase

Figure 7.7: Schematic diagram of the proposed multi-modal biometrics based cryptographic key regeneration scheme using *FeaLingECc*: (a) Enrollment phase, (b) Key regeneration phase.

pair corresponding to the same subject. Similarly, each data pair is compared with every other data pair of every other subject. This protocol results in 1,750 genuine comparisons and 380,625 impostor comparisons. For the sake of fair comparison with uni-biometric systems, similar protocol is followed to test the uni-biometrics based systems in this

chapter. Note that this experimental protocol is different than the one used in earlier chapters.

From Chapter 6, we know that the iris data need nearly 35% error correction. For face, we used only the controlled subset of the FRGCv2 data set. The error correction required on this subset is nearly 21%. Note that these quantities of error correction requirements are specific to the data set concerned and will change according to the modality and acquisition conditions. Also the amount of error correction required for iris is higher than for face. However, this does not impact the verification performance. The verification performance depends on the separation between genuine and impostor Hamming distance distributions which is better for iris than for face. Therefore, verification performance for iris is better for face.

As shown in Section 6.2.2 (page–92), Hadamard codes along with the zero insertion scheme can achieve the 35% error correction requirement for iris. For face, BCH codes can be applied for correcting the 21% errors. Therefore, we use Hadamard codes as Level-1 ECC for iris and BCH codes as Level-1 ECC for face. The Level-2 ECC are Reed-Solomon (RS) codes which is a common Level for iris and face. But, the error correction scheme in the proposed system is a cascaded structure where the dimensions of the Level-1 and Level-2 codes must be compatible. Each of the three ECC used in the system (RS, BCH, and Hadamard codes) has its own dimensional restrictions.

The Hadamard codes (which are used for iris) have a fixed relation between input and out size: a block of m bits is converted into a block of 2^{m-1} bits. The Reed-Solomon codes of block size m bits can have a maximum of 2^{m-1} blocks. The BCH codes having $\approx$ 21% error correction capacity are: BCH(127,15,27), BCH(255,21,55), BCH(511,28,111), BCH(1023,36,223), BCH(2047,56,443), etc. The suitable ECC sizes also depend on the dimensions of the individual biometric feature vectors. For example, the face code is 3,200-bit. It has to be truncated such that its length is a integer multiple of the BCH code output size. Similarly, the effective iris code length must be an integer multiple of the Hadamard code output size (32 or 64 bits). Moreover, from our experiments (Chapter 4), we know that the iris system performs better (from biometric recognition point of view) than the face system and hence, we need to apply higher weights to iris than to face. This means that more blocks of RS-codes output should be used for iris than for face.

Taking all these requirements into consideration, we fixed the size of the RS-codes block to be equal to $m = 7$. The output of the RS codes encoder is also in form of blocks each of which is 7-bit. Hadamard codes of input size $m = 7$ should be used for compatibility. The output of these Hadamard codes is 64-bits. The length of the iris code after zero insertion is 1984 bits and thus there can be 31 blocks of Hadamard codes. This also means that 31 blocks of RS-codes output are used for iris. The BCH codes should be selected such that the input size of BCH codes is an integer multiple of seven but also keeping in mind that the total number of RS-codes blocks required for face remains less than 31. BCH(127,15,27) and BCH(255,21,55) will require 50 and 36 RS-codes blocks which is more than that required for iris. Therefore, these codes cannot be employed in the system. Hence we applied the next two possible BCH codes: BCH(511,28,111) and BCH(1023,36,223).

In case of BCH(511,28,111), four RS-codes output blocks are concatenated to form a single input block. The 3,200-bit face code is truncated to 3,066 bits which is an integer multiple of 511. There are six such BCH codes blocks which require 24 RS-codes output blocks. Thus, the total number of output blocks required in the RS-codes is $31 + 24 = 55$. The iris has $31/55 = 56\%$ weight in the final verification decision while the face has 44% weight.

For the other possible BCH codes, BCH(1023,36,223), five RS-codes output blocks are concatenated and a zero is appended to it in order to obtain the required 36-bit input block. There can be three such BCH blocks requiring 15 RS-codes blocks. Thus the total number of RS-codes blocks is $31 + 15 = 46$. The iris is given 67% weight in this scenario while the face is given 33% weight.

Experimental performance evaluation along with security analysis of this proposed system is presented in the following subsection.

7.4.3 Results and Security Analysis for the Multi-modal (Iris and Face) Type System

The experimental results and theoretical as well as experimental security analysis is presented in this section.

Experimental Results of the Multi-modal (Iris and Face) Type System

The experimental data set used for this evaluation is different than that used for the uni-biometric systems in earlier chapters. Therefore, for comparison purposes, the baseline biometric systems' verification performances are presented in Table 7.4. The biosecure tool for performance evaluation [5] is used to calculate the EER and confidence intervals. The high improvement in the face verification system after shuffling is due to the high impact of shuffling on impostor face distribution. Shuffling makes the impostor distribution random. The randomness in un-shuffled iris data is higher than that of the face data, and hence, the impact of shuffling on face data is higher than that on iris data.

Table 7.4: Baseline biometric systems' (see Section A.1) user verification performances in terms of EER in % on subsets of NIST-ICE and NIST-FRGCv2 databases; values in bracket indicate the error margins for 90% confidence interval; Baseline – corresponds to baseline biometric system; Shuffled – the shuffling scheme is applied.

Exp.	Iris	Face	Iris+face
Baseline	1.29 [±0.23]	6.53 [±0.52]	1.06 [±0.22]
Shuffled	0.35 [±0.12]	0	0

The results for the iris and face based uni-biometric key regeneration systems are first presented in Table 7.5. These systems have the same settings as in the previous chapter (Table 6.4, page–99 for iris and Table 6.8, page–108 for face).

As said earlier, we evaluated the multi-modal system with two sets of experiments by applying different weights. In Set-1, RS codes having 55 blocks at the output are used. 31 out of these 55 (i.e., $\approx$ 56%) are used for iris and remaining 24 (i.e., $\approx$ 44%) are used for face. BCH(511,28,111) codes are used for face. Since it requires 28-bit input, four RS-code blocks are combined to form that block resulting in a total of 24 RS-code blocks for face.

In a different setting, Set-2, RS codes with 46-block output are selected, and 31 of them are used for iris (i.e., $\approx$ 67%) and remaining 15 blocks for face (i.e., 33%). BCH codes of higher output size are used so that the number of blocks coming from BCH codes will reduce. We selected BCH(1023,36,223) for which the error correction capacity is nearly the same. The 36-bit input required for these BCH codes is obtained

by concatenating five RS-codes blocks appended with a zero. Thus, at the time of decoding, the last bit of the decoded value is discarded. The results for the Set-1 and Set-2 are reported in Table 7.6 and Table 7.7, respectively. For both of these settings, we also carried out experiments without using shuffling which are also reported.

Table 7.5: Results for uni-biometrics based key regeneration systems proposed in Chapter 6; FRR and FAR values are in %. $\|K\|$ indicates length of key K in bits; t_s denotes the error correction capacity of RS-codes.

t_s	**Iris (NIST-ICE subset)**			t_s	**Face (NIST-FRGCv2 subset)**		
	$\|K\|$	**FRR**	**FAR**		$\|K\|$	**FRR**	**FAR**
9	258	3.77	0	2	140	7.08	0
11	234	2.17	0.001	3	126	5.60	0
13	210	1.26	0.027	5	98	3.66	0
15	186	0.86	0.21	6	84	3.14	0

Table 7.6: Results for the proposed multi-modal biometrics based key regeneration system – Set-1 (iris weight = 56%, face weight = 44%). Other symbols have the same meanings as in Table 7.5.

t_s	$\|K\|$	**Without shuffling**		**With shuffling**	
		FRR	**FAR**	**FRR**	**FAR**
3	343	7.54	2.93	7.54	0
9	259	1.94	20.80	1.94	0
12	217	0.91	36.43	0.91	0
16	161	0.17	62.93	0.17	0

Table 7.7: Results for the proposed multi-modal biometrics based key regeneration system – Set-2 (iris weight = 67%, face weight = 33%). Other symbols have the same meanings as in Table 7.5.

t_s	$\|K\|$	**Without shuffling**		**With shuffling**	
		FRR	**FAR**	**FRR**	**FAR**
1	308	8.23	1.31	8.23	0
2	294	5.48	3.80	5.48	0
8	210	0.91	29.80	0.91	0
11	168	0.11	49.33	0.11	0

The improvement in performance over uni-biometrics based systems can be seen by comparing the results in Table 7.5 with those in Table 7.6 and Table 7.7, e.g., at an FRR of 0.91%, we can obtain 217-bit keys with 0% FAR with the multi-biometrics based system. Whereas, at the similar value of FRR (FRR=0.86%), the iris based

system can generate 186-bit keys. Keys obtained from the face based system are even smaller.

Security Analysis of the Multi-Modal (Iris and Face) Type System

Theoretical as well as experimental security evaluation of the proposed system is presented in this section. Using the procedure of Daugman [37], the number of degrees of freedom in the iris and face codes are estimated to be equal to 556 and 243, respectively. Note that, this estimation depends on the impostor Hamming distance distribution and can change with the data set being used for evaluation. The total number of degrees of freedom in the fused feature vector is $N = 556 + 243 = 799$. In total, the system can correct 27% errors in this code (i.e., $P = N{*}0.27 \approx 216$). Applying the Equation (7.4), an impostor needs,

$$BF \approx \frac{2^N}{\binom{N}{P}} \approx \frac{2^{799}}{\binom{799}{216}} \approx 2^{131}, \tag{7.7}$$

brute force calculations to obtain the key. Thus the entropy contributed by the biometric information is 131 bits. The shuffling scheme, which employs a shuffling key obtained with a password can add up to 52 bits of entropy to this estimate resulting in $131 + 52 = 183$ bits entropy. Therefore, the total entropy estimate for the multi-modal type key regeneration system can be given as:

$$\text{Entropy} = min(Length(\mathbf{K}), 183)\text{bits}. \tag{7.8}$$

Experimental security evaluation of the multi-modal type key regeneration system is carried out in a way similar to that performed for the two-iris system. In the stolen biometric scenario, the performance of the system remains unchanged. None of the impostors who provide stolen biometric data along with a wrong shuffling key is accepted by the system. Whereas, in the stolen key scenario, the FAR is equal to that of the system without shuffling. However, applying higher weight to iris proves beneficial in this case. The FAR is much less in stolen key scenario when higher weight is applied to iris.

7.5 Conclusions and Perspectives

In this chapter, we propose to combine multi-biometrics with cryptography for obtaining high entropy keys. Using multi-biometrics has several advantages over uni-biometrics such as: better verification accuracy, larger feature space to accommodate more subjects, and higher security against spoofing. When multi-biometrics are combined with cryptography, in addition to the advantages listed above, we can obtain longer keys with higher entropy.

In order to have long keys, we combine the biometric information in feature domain. We propose a novel method of *Fea*ture *L*evel Fus*ion* through Wei*g*hted *E*rror *C*orre*c*tion (FeaLingECc). With this method, different weights can be applied to different biometric data. The shuffling scheme, which we applied earlier to the biometric data, is used in this system to randomize the error correcting codes data which helps make the system more secure. Additionally, the shuffling scheme induces revocability, template diversity, and privacy protection in the system.

Two systems are proposed: (1) a multi-unit type system, and (2) a multi-modal type system. Information from the left and right iris of a person is combined in the multi-unit type system to obtain long and high entropy crypto-bio keys. The second scheme is a multi-modal biometrics based system in which information from iris and face is combined.

The parameters of the systems are first tuned on development databases and the systems are evaluated on the evaluation databases. For the two-iris tests, we used the NIST-ICE database. On this database, we can obtain 147-bit keys having 147-bit entropy with 0% FAR and 0.18% FRR.

The multi-modal system (iris+face) is evaluated on a virtual database created by combining images from the NIST-ICE and NIST-FRGCv2 databases. We succeed to obtain 210-bit keys having 183-bit entropy at 0.91% FRR and 0% FAR.

The proposed scheme can be adopted to other biometric modalities. The feature level fusion combined with weighted error correction method allows the fusion of different biometric modalities having variation in performances (e.g., face+iris). This opens up new directions for combining biometric information from different sources and dimensions. The difficulty is to find appropriate ECC for that modality and the bina-

rization of the feature vector.

Chapter 8

Biometrics Based Secure Authentication Protocols

In previous chapters, we proposed a number of crypto-biometric systems. The cancelable biometric systems proposed in Chapter 4 and 5 induce the important properties such as revocability, template diversity, and privacy protection into biometric systems. The uni- as well as multi-biometrics based key regeneration systems proposed in Chapter 6 and 7 can deliver long keys with high entropy. These crypto-bio keys, intended to be used in cryptographic applications, are obtained from biometrics and therefore are strongly linked to the user identity. The problem not addressed yet in these chapters, and largely ignored in most of the biometrics based key (re)generation systems found in literature (see Section 2.2), is how to share these keys to be used in cryptography.

Cryptography is generally divided into two main categories: symmetric-key cryptography and public-key cryptography. Symmetric-key cryptography is better suited for real time data transfer because of its speed. In symmetric-key cryptography, the same key is used for encryption and decryption, and therefore, it needs to be shared between all the parties. According to the Kerckhoffs' principle [70], the security of a cryptographic system lies in the key, and therefore, the cryptographic key needs to be protected. In order to share the key only with the intended users, different authentication mechanisms are employed.

In this chapter, we consider that there are two parties, a client and a server,

who need to establish a secure communication link between them. As we defined in the glossary, authentication is a process in which one party (client or server) verifies the authenticity of the other, and then establishes a secure channel between them. In some systems, a process denoted as mutual authentication, is carried out in which both the parties authenticate each other. But, if a large amount of data encrypted with a single symmetric-key is available, several cryptanalytic attacks are made easier.

In public-key cryptography, the two parties, client and server, have their own pairs of public and private keys. The client encrypts a secret with the server's public key. This encrypted data can only be decrypted with the corresponding private key. Since only the server has access to this private key, only the server can perform the decryption and recover the secret. Therefore, technically speaking, public-key cryptography does not need to verify the authenticity of the other party. But, as described in Section 1.2.1 (page–7), the public-key cryptosystems are vulnerable to the man-in-the-middle attack [84]. An example of man-in-the-middle attack is shown in Fig. 1.4 (page–10). Therefore, the public-key cryptosystems also require authentication which is achieved by employing trusted third party certificates.

In order to overcome these shortcomings, many widely used cryptographic protocols, such as the Transport Layer Security (TLS) protocol [42][1], are hybrid systems that use public-key cryptography to exchange a symmetric key. This symmetric key is temporary, and is valid only for the current communication session. Such symmetric key is denoted as session key. Having a temporary session key limits the amount of data encrypted with a single key. Moreover, the server authentication is carried out with the help of certificates issued by a certification authority trusted by both the parties. Additionally, the client can also be authenticated in a mutual authentication mode if he also obtains a certificate from the trusted certification authority.

The crypto-biometric systems summarized in Chapter 2 and those proposed in this thesis can be used for secure authentication in cryptographic communication. In fact, the crypto-bio key generation/regeneration systems do produce a key (as opposed to a single bit verification result of the biometric and cancelable biometric systems). This crypto-bio key can be used for secure cryptographic communication.

[1]TLS is a widely used protocol, e.g., HTTPS (HyperText Transmission Protocol–Secure) uses TLS to secure World Wide Web traffic carried by HTTP. HTTPS is used for secure e-commerce applications such as online payments through internet, online banking applications, etc.

There are very few proposals found in literature for biometrics based secure cryptographic protocols discussed in Section 2.3. Most of those systems do not satisfy all the goals we set for crypto-biometric systems. Many of them require storage of biometric templates. Some others need a pre-existing secure channel for the biometric information exchange.

In this chapter, we propose two protocols which enable us to share the crypto-bio keys securely. We make following assumptions for the proposed protocols:

- There is no trust between the client and the server. Therefore, the client will not pass the authenticators (e.g., biometric data, passwords, etc.) to the server. The server will also not share the stored information with the client.
- The communication link between the client and the server is unprotected. Therefore, the data being transferred through this link should not leak information.
- Biometric data of the user should not be stored in the server or database to protect the user's privacy. The stored data should be revocable.
- The protocol should achieve mutual authentication between the client and the server because none of them trust each other.

The first protocol is proposed in order to securely share the crypto-bio keys obtained with the key regeneration system described in Chapter 6. These keys are always the same for a particular user and can be used in symmetric cryptography. The second protocol is proposed for generation and sharing of biometrics based (crypto-bio) session keys. Both of these protocols achieve mutual authentication without the need of trusted third party certificates. Since biometrics is involved in the crypto-bio key regeneration process, a strong link is established between the user's identity and his cryptographic keys. Additionally, the session key generation and sharing protocol allows easy online update of templates.

In the end, we also perform a case study of how the crypto-bio key regeneration scheme proposed in Chapter 6 can be integrated into a state-of-the-art cryptographic authentication protocol. We integrated this key regeneration scheme in the theoretical proposal of Abid and Afifi [8] in a collaborative work [9] with the authors of [8]. The system in [8] is an ePassport authentication protocol using elliptic curve cryptog-

raphy [88, 72]. They proposed to use fingerprints for generation of the elliptic curve parameters. Based on this scheme, we used our iris based key regeneration scheme in order to obtain a stable input from iris data which is then processed to generate elliptic curve parameters.

The rest of this chapter is organized as follows: in Section 8.1.2 and 8.2, the novel protocols for sharing crypto-bio keys and for generation and sharing of biometrics based session keys are proposed, respectively. The case study of integrating crypto-bio key regeneration system into an existing cryptographic protocol is reported in Section 8.3. This case study is carried out in collaboration with Abid and Afifi, the authors of [8]. It is part of the French Agence Nationale de la Recherche (ANR) project BIOTYFUL [6]. Finally, conclusions and perspectives are given in Section 8.4.

8.1 Biometrics Based Cryptographic Key Regeneration and Sharing

In order to facilitate the understanding of the proposed protocols, the key regeneration system is revisited in the following subsection.

8.1.1 A Recap of the Biometrics Based Key Regeneration Scheme

The biometrics based key regeneration scheme shown in Fig. 8.1 is a hybrid system that combines a transformation based cancelable biometric system with fuzzy commitment based key regeneration scheme. In this scheme, a key $\mathbf{K}_r$ is randomly generated and then encoded into a pseudo code θ_{ps} using Error Correcting Codes (ECC). A cancelable transformation is applied on the reference biometric data θ_{ref} of a user. This transformed data θ_{canc} is then XOR*ed* with the pseudo code θ_{ps} to obtain a locked code template θ_{lock}. At the time of key regeneration, a similar transformation is applied on the test biometric data θ_{test} and then the cancelable data θ'_{canc} is XOR*ed* with the stored template θ_{lock} to obtain θ'_{ps}. The two XOR operations transfer the errors between the reference and test biometric data onto the pseudo code ($\theta'_{ps} = \theta_{lock} \oplus \theta'_{canc} = \theta_{ps} \oplus \theta_{canc} \oplus \theta'_{canc} = \theta_{ps} \oplus e$). If the amount of errors e is less than the error correction capacity of the ECC, all these errors can be corrected after decoding. On successful error correction, a trial value of the random key $\mathbf{K}_r$, denoted as $\mathbf{K}'_r$ is obtained. A

comparison of the hash values of these two keys is carried out, and if they are same, verification success is declared along with releasing the key. If the hash values are different, verification failure is declared.

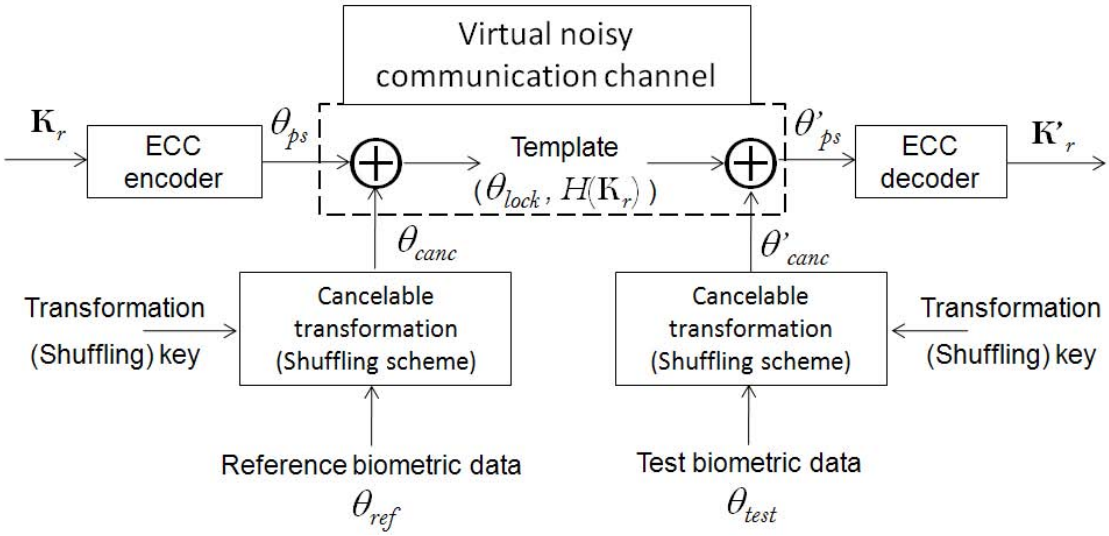


Figure 8.1: Biometrics Based Key Regeneration Scheme proposed in Chapter 6.

The cancelable transformation used in our system is the shuffling scheme (Section 4.1). A randomly generated shuffling key $\mathbf{K}_{sh}$ is assigned to each user and this key is used to randomize the biometric data of that user. The biometric data is divided into blocks and these blocks are rearranged according to the shuffling key. Since the shuffling key is long, it needs to be stored on a smart card or should be generated using a password. The advantage of this shuffling scheme is that it increases only the impostor Hamming distances leaving the genuine Hamming distances intact. Hence, in addition to the properties of cancelability, revocability, template diversity, and privacy protection, the shuffling scheme also improves the verification performance of the system.

8.1.2 Secure Crypto-bio Key Sharing Protocol

Long keys having high entropy can be obtained with the key regeneration system proposed in Chapter 6 which is summarized in previous subsection. Now, we propose a simple and effective protocol to securely share the crypto-biometric keys obtained with this system.

A schematic diagram of the proposed protocol for crypto-biometric key sharing

is shown in Fig. 8.2. The enrollment process (not shown in the figure) is carried out off-line at a secure location. It is basically the same as described in Section 8.1.1. A secure, locked code template θ_{lock} is created using a random key $\mathbf{K}_r$, shuffling key $\mathbf{K}_{sh}$, and the reference biometric data θ_{ref}. This θ_{lock} along with the hash of the key $\mathbf{K}_r$, i.e., $H(\mathbf{K}_r)$ is stored in a database. The system can also employ a smart card to store the shuffling key $\mathbf{K}_{sh}$ in encrypted form. Otherwise, the shuffling key can be directly generated from a password.

At a later time, when the client needs a secure cryptographic key for communication, following steps are carried out:

1. The client sends the authentication request to the server.

2. The server responds with the request accept signal.

3. At the client side, fresh biometric data θ_{test} of the user is captured and shuffled using the shuffling key $\mathbf{K}_{sh}$ to obtain shuffled test code θ'_{canc}. Only the user ID is sent to the server.

4. The server sends the locked code θ_{lock} along with the hash value $H(H(\mathbf{K}_r))$ of the stored hash (i.e., hash of $H(\mathbf{K}_r)$) of the user corresponding to the requested ID to the client.

5. At the client side, a key $\mathbf{K}'_r$ is obtained from θ_{lock} and θ'_{canc} as, $\mathbf{K}'_r = E^{-1}(\theta'_{canc}, \theta_{lock})$ where $E^{-1}(\cdot)$ indicates the decoding function.

6. The client computes $H(H(\mathbf{K}'_r))$ and compares it with the received $H(H(\mathbf{K}_r))$ and if the two values are equal, the shuffled biometric data θ'_{canc} is encrypted using $H(\mathbf{K}'_r)$ and the encrypted data is sent to the server.

7. The server decrypts the received data with $H(\mathbf{K}_r)$ (which is stored in the database) to obtain θ'_{canc} and then regenerates the key $\mathbf{K}'_r$ from θ_{lock} and θ'_{canc}.

8. The server checks the hash values of the original and regenerated keys ($H(\mathbf{K}_r)$ and $H(\mathbf{K}'_r)$, respectively). If they are equal, it sends a start communication signal to the client.

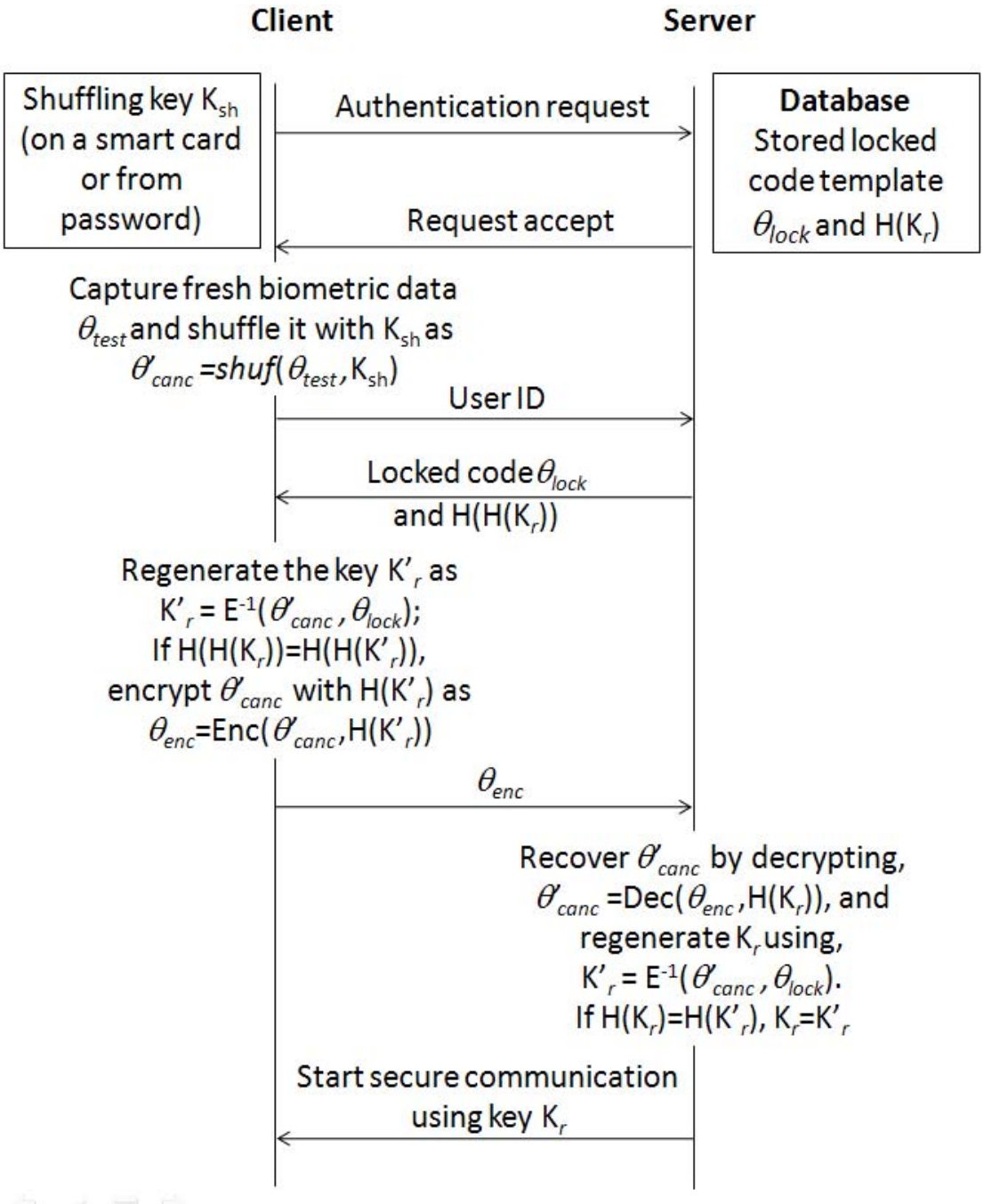


Figure 8.2: The proposed protocol for biometrics based secure key sharing.

Thus a secure channel is established between the client and the server through which secure communication can be carried out. Moreover, the protocol achieves biometric based secure authentication over an unsecured channel. The templates stored in the database are cancelable and the system possesses the properties of revocability,

template diversity, and privacy protection.

8.2 Biometrics Based Session-Key Generation and Sharing Protocol

8.2.1 Session Key Generation and Sharing

The protocol described in the previous section is for sharing crypto-bio keys which can be used in symmetric cryptographic systems. Having a single symmetric key for encrypting a large amount of data is not good for security. Therefore, it is essential to have a scheme which can generate and share session keys based on biometrics for higher security.

In this section, we propose a novel protocol to generate and share session keys based on biometrics. It makes use of the biometrics based key regeneration system described in Section 8.1.1, but it can be generalized to accommodate any other key regeneration scheme. The enrollment is securely carried out off-line during which a cancelable template is generated from the enrollment biometric data of the user and is stored in the database at the server. In our case, the cancelable template is the shuffled biometric data θ_{canc} which is obtained by shuffling the enrollment biometric data θ_{ref} with a shuffling key $\mathbf{K}_{sh}$. The shuffling key $\mathbf{K}_{sh}$ is either stored on a smart card or can be generated from a password.

Figure 8.3 shows a schematic diagram of the proposed session key generation and sharing protocol. The channel between the client and the server is not secure, and hence, no private or sensitive information should be sent over the network unless the channel is secured. When a client desires to securely communicate with the server, following steps are carried out:

1. The client sends authentication request to the server.

2. The server sends acknowledgement to the client.

3. Fresh biometric data θ_{test} of the user is captured and shuffled using the shuffling key $\mathbf{K}_{sh}$ to obtain shuffled test biometric data θ'_{canc} at the client side.

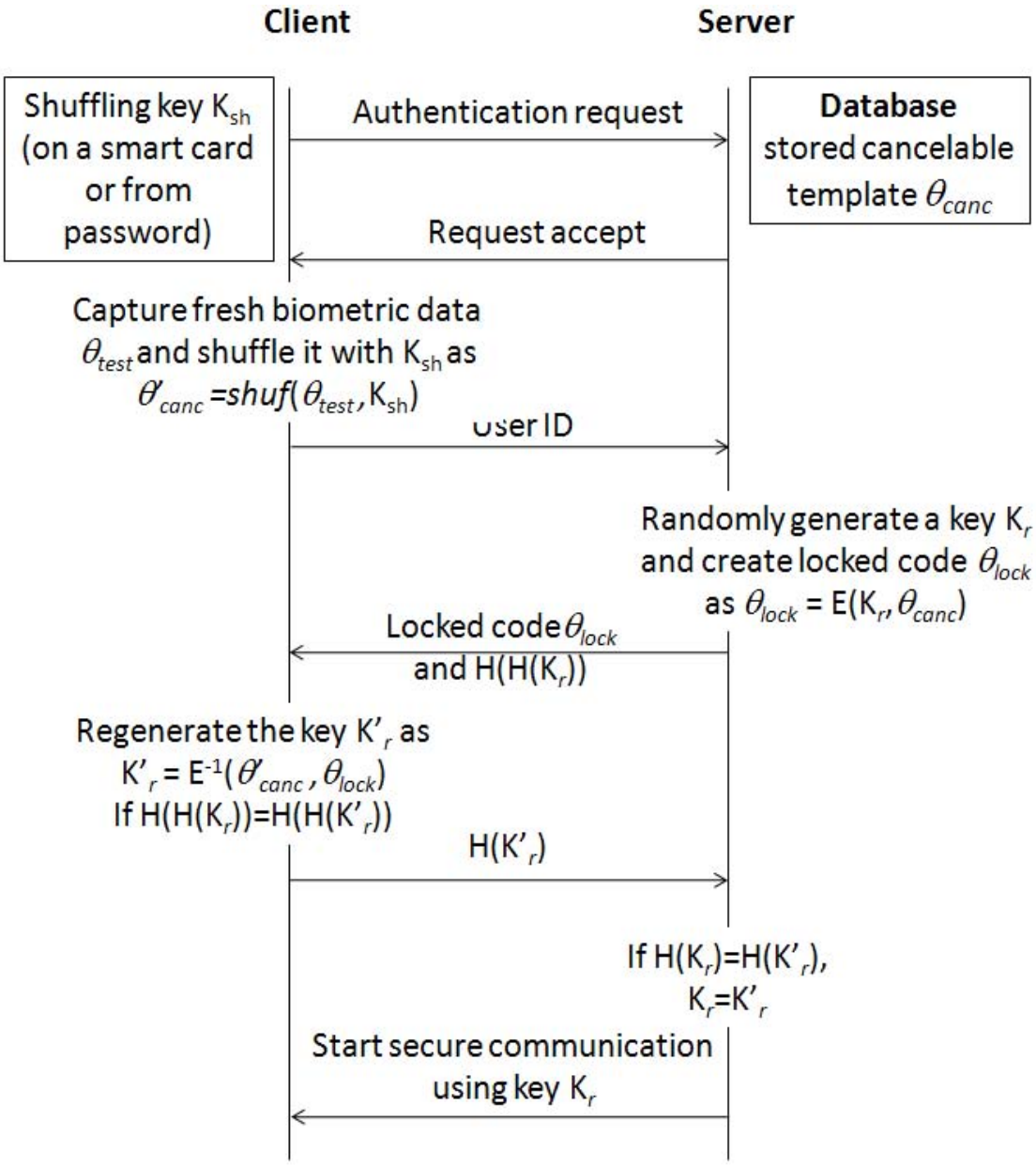


Figure 8.3: The proposed protocol for generating and sharing biometrics based session keys.

4. User ID of the user is sent to the server. Note that the biometric data is not sent to the server.

5. The server generates a random key $\mathbf{K}_r$ and a locked code θ_{lock} is created from $\mathbf{K}_r$ and the stored cancelable template θ_{canc}. This process of obtaining θ_{lock} is the same as shown in Fig. 8.1. It can be summarized as $\theta_{lock} = E(\mathbf{K}_r, \theta_{canc})$ where $E(\cdot)$ indicates the encoding function.

6. The locked code θ_{lock} is sent to the client. A double hashed version of the random key, i.e., $H(H(\mathbf{K}_r))$ is also sent to the client.

7. The client regenerates a trial value $\mathbf{K}'_r$ of the random key using the locked code θ_{lock}, and the shuffled test biometric data θ'_{canc}. This can be summarized as $\mathbf{K}'_r = E^{-1}(\theta'_{canc}, \theta_{lock})$, where $E^{-1}(\cdot)$ indicates the decoding function. The regenerated key $\mathbf{K}'_r$ is hashed twice to obtain $H(\mathbf{K}'_r)$ and $H(H(\mathbf{K}'_r))$.

8. The client compares $H(H(\mathbf{K}'_r))$ with the received $H(H(\mathbf{K}_r))$ and if the two values are equal (which also confirms the server's authenticity), it sends the $H(\mathbf{K}'_r)$ to the server.

9. Server compares the received hash value $H(\mathbf{K}'_r)$ with the hash value of the random key $\mathbf{K}_r$, i.e., with $H(\mathbf{K}_r)$ to check the authenticity of the user. If the two hash values are the same, it means that the user is authentic and has correctly received the randomly generated key $\mathbf{K}_r$. Thus, both the parties have the same key $\mathbf{K}_r$.

10. The key $\mathbf{K}_r$ is then treated as a session key and the server sends the signal to start secure communication using the key $\mathbf{K}_r$.

Thus, at the end of this protocol, the client as well as the server share the same key which can be used for symmetric-key cryptography. Note that, the key is temporary and is destroyed at the end of the communication session. In the next communication session, a new key $\mathbf{K}_r$ will be randomly generated and shared to be used as a session key.

The data being transferred through the channel during the protocol are request, user ID, locked code θ_{lock}, and the hash values $H(H(\mathbf{K}_r))$ and $H(\mathbf{K}'_r)$, none of which reveal the biometric information. Moreover, the template stored in the database is cancelable which itself prevents cross-linking between biometric databases and protects user privacy.

As opposed to the popular and widely used cryptographic protocols such as TLS, the proposed protocol does not need a third party trusted certification authority. In TLS, the third party certification is used to confirm the server authenticity by using digital certificates. In our proposed protocol, client can confirm the authenticity of the server by comparing the double hashed values $H(H(\mathbf{K}'_r))$ and $H(H(\mathbf{K}_r))$. This

comparison can yield positive result only if the server has generated a locked code θ_{lock} from the stored template θ_{canc} of the same user. On the other hand, the server authenticates the client by comparing the hash values $H(\mathbf{K}_r)$ and $H(\mathbf{K}'_r)$. Thus, our protocol achieves mutual authentication without the need of third party certificates. The system described here employs strong authentication by combining biometrics with password (or smart card). Since the user is required to provide specific information in addition to biometric data, the system can resist replay attacks.

The error correction coding is applied at the time of authentication/key regeneration. Hence, it is possible to accommodate different error correcting codes (compatible with the biometric data) in the protocol. As it is done in the TLS, the client and server can negotiate on the choice of ECC and the error correction capacity to be used during authentication.

8.2.2 Online Template Update

Many systems (such as online banking services) require that the user authentication credentials be updated periodically. In password based systems, this means that the user is asked to change his password periodically. On the other hand, the user may also wish to change his credentials.

The distributed nature of our proposed protocol allows the user and/or the system to update the template online. The template update procedure involves changing the cancelable template θ_{canc} by changing the reference biometric data θ_{ref} and the shuffling key $\mathbf{K}_{sh}$. The procedure for this template update is shown in Fig. 8.4.

The steps followed during the template update procedure are:

1. A secure communication channel is created between the client and the server by using the session key generation and sharing protocol described in the previous subsection (shown in Fig. 8.3).

2. A new shuffling key $\mathbf{K}'_{sh}$ is randomly generated at the client side and a cancelable template θ'_{canc} is obtained from the fresh test biometric data θ_{test} and $\mathbf{K}'_{sh}$.

3. The new cancelable template θ'_{canc} is sent to the server through the encrypted channel.

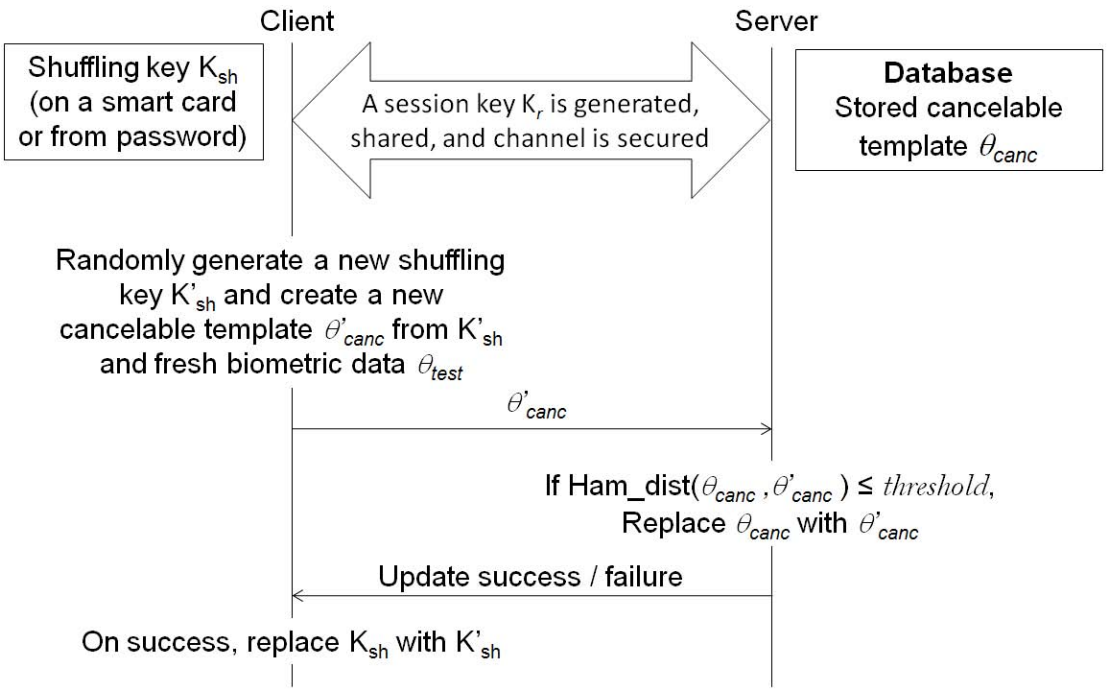


Figure 8.4: Protocol showing online template update. In the beginning of this protocol, the mutual authentication between the client and the server is carried out with the protocol shown in Fig. 8.3. Ham_dist means Hamming distance.

4. The server compares the old template stored in the database θ_{canc} with the received cancelable template θ'_{canc}. If the Hamming distance between the two is less than a threshold, the old template θ_{canc} is replaced with the new one θ'_{canc}. Update success/failure message is sent to the client.

5. If the received message is success, the old shuffling key $\mathbf{K}_{sh}$ stored on the smart card is replaced with the new one $\mathbf{K}'_{sh}$.

Note that, the template update process can be initialized by either the client or the server. Also the mutual authentication between client and server is carried out before initiating the template update procedure during session key generation and sharing. We recommend that template update should be carried after every session for higher security.

The proposed protocol can also be integrated inside classical cryptographic protocols such as TLS. Such classical protocols can first be used to establish a secure connection between the client and the server. Then the protocol shown in Fig. 8.3 can be employed for biometrics based secure mutual authentication between the client and the server.

8.3 Iris Based Authentication Mechanism for ePassports – A Case Study

The crypto-bio key regeneration system described in Chapter 6 can also be integrated into other state-of-the-art cryptographic security protocols. In this section, we present an illustration of such system. We consider biometrics ePassports as a case study.

Biometrics has been included in ePassports in many countries across the world. There are various protocols specifying the use of biometrics in ePassports such as: the International Civil Aviation Organization (ICAO) guidelines for ePassports (Doc 9303 standard [57]), the Extended Access Control (EAC) issued by European Union [46], and the Online Secure ePassport Protocol (OSEP) [102]. Biometric data can be used in these protocols for identity verification. The biometric comparison is carried out in a classical way and hence requires classical biometric template storage. Therefore, the drawbacks of biometric systems described earlier are inherited in these systems.

Abid and Afifi [8] proposed to employ elliptic curve cryptography [88, 72] for the ePassport authentication. They proposed to use fingerprints for generating security parameters for the elliptic curve. In this way, the biometric data is used in an intricate manner for strengthening of the cryptographic protocol. The drawback of this scheme is that it requires a stable bit-string to be extracted from the fingerprints. Extraction of stable bit-string from fingerprint (and in general, any biometric data) is a difficult problem as discussed earlier. In [8], the authors did not address this problem and provided no experimental evaluation for biometric verification performance.

We have addressed the problem of obtaining stable bit-string from biometric data in Chapter 6. In collaboration with the authors of [8], we applied our iris based key regeneration scheme to the scheme in [8]. In the proposed solution, we obtain the elliptic curve parameters from the crypto-bio key regenerated using the subject's iris data.

The proposed solution has three phases: (1) Initialization, (2) Inspection System (IS) authentication, and (3) ePassport bearer's authentication. These three phases are described in following subsections.

8.3.1 Initialization Phase

During the initialization phase, an elliptic curve over Galois Field $GF(p)$ is generated, with p being a prime number. The parameters needed for the Elliptic Curve Diffie-Hellman (ECDH) key agreement algorithm [14] are saved in the chip in the ePassport. This phase is carried out securely at the site of the ePassport issuing authority. Figure 8.5 shows all the entities participating in this phase.

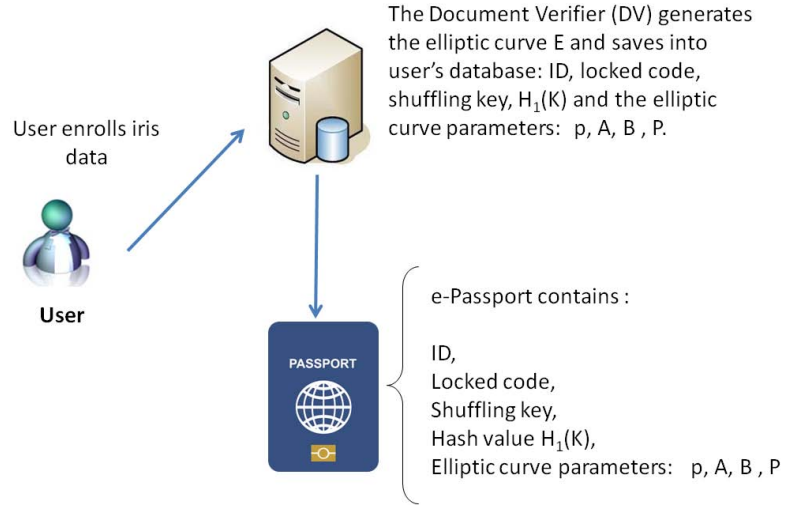


Figure 8.5: Initialization phase: delivering an ePassport to a subject at the issuing authority (Fig. from [9].

The person who needs a new ePassport is asked to provide his iris biometrics. To generate the elliptic curve, the system takes, as input, the enrollment iris. The locked code is generated from a randomly generated key K and a shuffling key. The generated elliptic curve E ($y^2 = x^3 + Ax + B \bmod p$) needs to be an ideal elliptic curve for cryptographic use. This elliptic curve will be used by the chip and the IS to define a session key using the Elliptic Curve Diffie-Hellman (ECDH) key agreement protocol [14] as it is done in [8].

The elliptic curve E is generated and used in the ECDH protocol. We hash the key K with Secure Hash Algorithm (SHA-256) [7] to create the hash value $h_2(K)$. This hash value $h_2(K)$ is used with a prime number p coded on 128 bits and a big number A also coded on 128 bits to choose the suitable elliptic curve **E**. This procedure is shown in Figure 8.6.

First of all, a $P_0(X_0, Y_0)$ will be generated from the hash value $h_2(K)$. Since

length of $h_2(K)$ is 256 bits, it is divided into two parts: X_0 and Y_0, each of which is coded on 128 bits.

The coefficient $A \in GF(p)$ is chosen by the Document verifier (DV). Then, the DV sets $B = Y_0^2 - X_0^3 - AX_0$, and checks that $4A^3 + 27B^2 \neq 0$. If this condition is satisfied, $N = Card(E)$ is computed, where N is the cardinality of the curve. If N is prime, a certificate of primality is generated.

At the end, the DV gets a curve suitable for cryptographic use. It chooses a point $P \in E$ which will be the public point of the chip. Then, the ePassport is ready to be delivered to the traveler. The parameters stored in the user database and the ePassport chip are:

- ID: identifier;
- the locked code;
- the shuffling key;
- $h_1(K)$: the hash value;
- p: the prime number;
- the parameters generated using iris biometric:
  - P - the public point;
  - A and B - the coefficients of the elliptic curve.

Other conventional parameters for the ePassport like name, country, age, gender, etc., can also be added.

The parameters A, B, p, and P are certified by the document verifier. At the end, the ePassport is delivered to the traveler. When the validity time of the ePassport is finished, a new ePassport will be issued to the bearer where the system will generate a new elliptic curve different from the previous one.

8.3.2 Inspection System (IS) Authentication

The second phase is the Inspection System (IS) authentication which is carried out when a traveler presents his ePassport to the border control authorities. In this

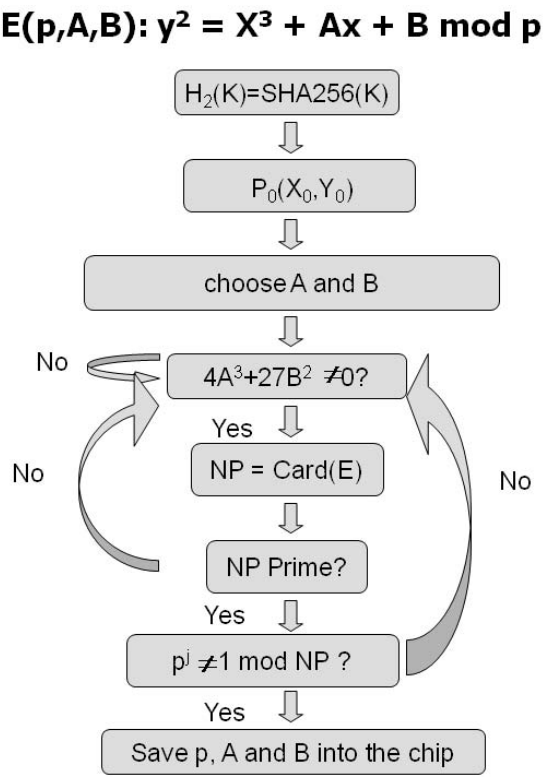


Figure 8.6: Elliptic curve generation (Fig. from [9].

phase, authenticity of the ePassport is verified and then the IS and the chip agree on a session key to be used to secure the communication between them.

Figure 8.7 shows all the entities participating in this phase: the ePassport chip, the IS, and the DV. The chip has public and private keys certified by the DV. The IS and the DV have public and private keys certified by the Certification Authority of the Visiting Country.

To agree on a session key shared between the chip and the IS, the Elliptic Curve Diffie Hellman Key Agreement protocol is used. It involves an exchange of two points in $\mathbf{E}$ ($Q_C = N_C * P$ from the chip and $Q_{IS} = N_{IS} * P$ from the IS) where N_C and N_{IS} are two random numbers. The IS verifies the data sent by the chip by asking to the DV. The latter also checks the authenticity of the IS since the certificates are signed by the certification authority of the visiting country. When the IS is convinced about the authenticity of the ePassport, IS sends his point Q_{IS}. At the end of the exchange, the point, $Q = N_{IS} * Q_C = N_C * Q_{IS}$ is shared between the chip and the IS. The session key can be the abscissa X or the ordinate Y of the point Q, or their concatenation $X||Y$

(X and Y are coded on 128 bits).

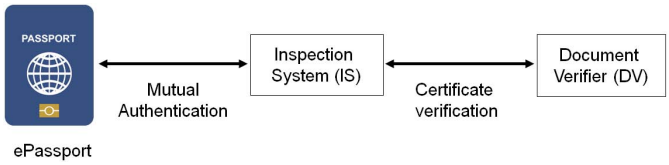


Figure 8.7: Entities involved in the Mutual Authentication (Fig. from [9].

8.3.3 ePassport Bearer's Authentication

Once the ePassport is authenticated, the bearer of the ePassport is asked to prove his identity at the border control. The procedure of ePassport bearer's authentication is shown in Figure 8.8. The traveler provides fresh iris biometrics data. The chip sends the data needed by the Inspection System (IS) to retrieve the elliptic curve for the authentication of the bearer. These data are, the locked code, the shuffling key, the hash value $h_1(K)$ and the parameters of the elliptic curve p, A, B.

First of all, the IS generate K' using the fresh Iris biometric data, the Shuffling Key and the Locked code. The IS can check if $h_1(K')$ is equal to $h_1(K)$.

The IS hashes K' using SHA-256 to get hash value $h_2(K')$. A point $P'_0(X'_0, Y'_0)$ is created using $h_2(K')$. The point P'_0, p and A are used to generate the elliptic curve $\mathbf{E}'$ ($y^2 = x^3 + Ax + B' \bmod p$). If the value B' is equal to B, it means that the bearer of the ePassport is genuine.

In the end, the IS and the chip agree upon a session key extracted from K. The chip can release its data to the IS in a secure way.

8.3.4 Experimental Evaluation of the Iris Based ePassport Authentication Protocol

Biometric verification performance of the proposed iris based ePassport authentication protocol is evaluated on the NIST-ICE database [101]. We used the Multi-precision Integer and Rational Arithmetic C/C++ Library (MIRACL) [2]. This library implements the primitives necessary to design big number cryptography including cryptographic algorithms such as the SHA-256, AES, and the elliptic curve functions. The iris based key regeneration system is implemented in MATLAB.

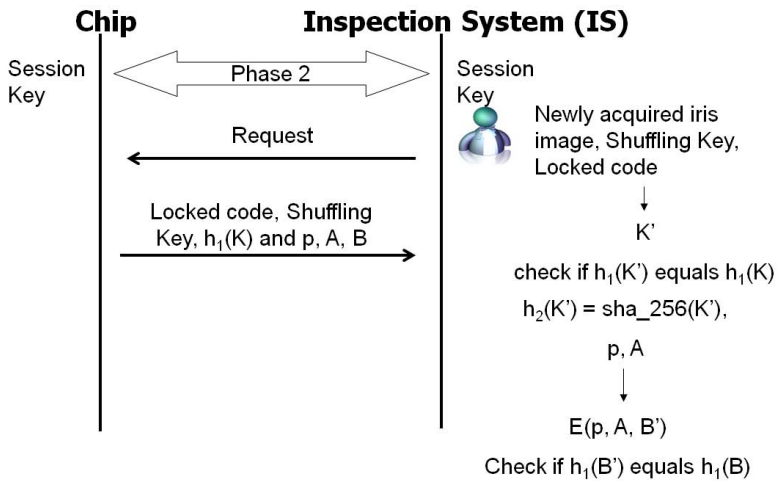


Figure 8.8: Procedure of ePassport bearer's authentication at the border control using fresh iris data (Fig. from [9].

On a computer with Intel Xeon E5430 CPU @ 2,66 GHz and 16 GB RAM, the initialization phase of the protocol takes $\sim$ 3 minutes while the ePassport bearer's authentication phase requires nearly seven seconds. Considering these time constraints, we used a subset of the NIST-ICE database [101] for our evaluations. This data set is the same as that used for evaluation of the multi-modal biometrics based key regeneration system in Chapter 7. This data set contains 875 images of 175 different irises (175×5). A total of 1750 genuine and 308,625 impostor comparisons are carried out.

The iris based key regeneration system is already tested on the whole NIST-ICE database in Chapter 6 (see Table 6.4, page 99) for different values of error correction capacity t_s (of the Reed-Solomon codes). We selected two of these values for our case study – $t_s = 10$ and $t_s = 15$. The results are shown in Table 8.1. The EER of the baseline iris system (see Section A.1) on this data set is 1.29%.

Table 8.1: Experimental results of the iris based ePassport authentication protocol on a subset of the NIST-ICE database.

t_s	**FAR (in %)**	**FRR (in %)**
10	0.01	5.26
15	0.2	2.60

The low FAR (e.g., 0.01% for $t_s = 10$; in Table 6.4, it is 0.0005%) indicates higher security while the low FRR (e.g., 2.60% for $t_s = 15$; in Table 6.4, it is 0.76%)

indicates better user friendliness. The performance is slightly degraded as compared to the performance of the iris based key regeneration system reported in Table 6.4. The reason behind this degradation is the fact that the elliptic curve algorithm requires the points on the curve to be prime for security reasons. Therefore, if the generated point is not prime, the algorithm performs a predefined number of iterations until it is prime. If it still fails to obtain a prime number, the algorithm declares curve generation failure. Because of such situations, the performance gets degraded.

8.4 Conclusions and Perspectives

In this chapter, the important issue of application of the crypto-bio keys obtained from the biometrics based key regeneration systems is addressed. To begin with, we proposed a novel protocol which enables sharing of the crypto-bio keys between two parties (a client and a server) over a completely un-secure communication channel. The limitation of this protocol is that the crypto-bio keys obtained during its execution are always the same. In order to have better security, we proposed a new protocol which enables generation and sharing of biometrics based session keys. The session keys are precisely valid for a single communication session and are destroyed afterwards. Both of these two protocols achieve mutual authentication between the client and the server in a zero trust environment (client does not trust the server and vice-versa). The underlying key regeneration scheme, and in turn, these protocols possess the properties of revocability, template diversity, and privacy protection.

In addition, as a case study, we also developed an iris based ePassport authentication protocol (a collaborative work). In this protocol, the biometrics based key regeneration scheme proposed in earlier chapters is integrated in a state-of-the-art cryptographic protocol for ePassport authentication. This protocol is based on elliptic curve cryptography. The crypto-bio keys obtained with the key regeneration scheme are used to generate the elliptic curve parameters. This proposal is validated by carrying our experimental evaluation for biometric verification performance on a subset of the publicly available NIST-ICE database where we obtain FRR=2.60% at FAR=0.20%.

Chapter 9

Conclusions, Perspectives, and Future Directions

9.1 Conclusions and Perspectives

Biometrics and cryptography are two techniques which have high potential for providing information security. Unfortunately, both of these have certain limitations. Cryptography requires keys, but these keys are not strongly bound to the user's identity, whereas, biometrics suffer from nonrevocability, non-template diversity, and possibility of privacy compromise. A combination of biometrics and cryptography is a good solution for eliminating these limitations. We call such systems, in which biometrics and cryptography are combined, crypto-biometric systems.

This is an emerging field which started from late 90's, and lacks a uniform terminology. Therefore, first of all, we presented a systematic classification of the crypto-biometric systems found in literature based on their principal goals and working methodologies. We proposed two major classes: (i) protection of biometric data and (ii) obtaining biometrics based cryptographic keys. The systems in these two categories were further classified based on their working methodology.

Secondly, we proposed a cancelable biometric system to protect the biometric data. A user specific shuffling technique was proposed which adds revocability and template diversity to the biometric systems. Additionally, it protects the biometric data privacy, information privacy, and identity privacy of the user. The most important

advantage of this shuffling technique is that it improves the verification performance of the baseline biometric system by more than 80%. The reason for this improvement is that the shuffling scheme increases the impostor Hamming distances without changing the genuine Hamming distances. A distinctive feature of this scheme is that, if the shuffling key for all the users is compromised, its performance is equal to that of the baseline biometric system. This is a distinct advantage of this system with respect to other systems found in literature.

It is a well known fact among the biometrics research community that the biometric data coming from a single source contain variability. In order to reduce such variability, we proposed a novel approach of using Error Correcting Codes (ECC). By careful selection of the ECC, this scheme can correct more errors (variability) in genuine cases than in impostors. When combined with the shuffling scheme, the performance improves by more than 90%.

Further, we combined the shuffling scheme with a fuzzy commitment based system to develop a hybrid scheme for biometrics based cryptographic key regeneration. The shuffling scheme is applied on the biometric features to make them cancelable. The cancelable features are then used in a fuzzy commitment based key regeneration system. This scheme possesses the important properties of revocability and template diversity, and it also protects user's privacy. This generic scheme was adapted for iris and face modalities. The estimated entropy of the crypto-bio keys obtained from iris is 83 to 93 bits. For the face based system, the estimated entropy is 110 to 112 bits.

In order to have longer keys with higher entropy, we proposed to use multi-biometrics for key regeneration. A novel idea denoted as *FeaLingECc* (*Fea*ture *L*evel Fus*ion* through Wei*g*hted *E*rror *C*orre*c*tion) was proposed. The FeaLingECc allows fusion of different biometric modalities having variation in performances (e.g., face+iris). Using this scheme, we proposed a multi-unit system based on two-irises and a multi-modal system using a combination of iris and face. We succeeded to obtain crypto-bio keys having 189-bit entropy with the two-iris system while the entropy of keys obtained with the iris-face system is 183 bits.

Finally, we addressed the problem of securely sharing the crypto-bio keys. We designed protocols which can share the crypto-bio keys while achieving mutual authentication between a client and a server. The protocol also allows sharing of session specific

(one-time) crypto-bio keys with the help of biometrics. This protocol operates in a zero trust environment and achieves mutual authentication without the need of trusted third party certificates. Session keys have high practical importance, and therefore, the proposed biometrics based session key generation and sharing protocol can find a large number of applications. This protocol has a potential to replace the existing key sharing protocols. On the other hand, it can be seamlessly integrated into classical key sharing protocols to provide additional level of security.

9.2 Future Research Directions

The cryptographic key regeneration systems proposed in this thesis use Error Correcting Codes (ECC). A shortcoming of our work is that we have not presented an analysis of the systems against the attacks based on the ECC statistics. An attacker can run the ECC in soft decoding or erasure mode which can reduce the entropy of the keys. Such analysis will be carried out through theoretical as well as experimental evaluations.

The proposed crypto-biometric systems can be further tested with different biometric modalities. Open source biometric reference systems can be very useful for this purpose. Moreover, the systems can further be tested on larger databases. Especially, the proposed multi-biometric system is evaluated on a virtual multi-modal biometric database. It can further be tested on a real multi-modal biometric database.

The biometrics based session key generation and sharing protocol can further be extended to integrate multi-biometrics. Using multi-biometrics has many advantages over uni-biometrics such as better verification accuracy, larger feature space to accommodate more subjects, and higher security against spoofing. Additionally, it can provide longer keys having higher entropy.

It has been observed that the noninvertibility and performance improvement properties of the cancelable biometric systems are contradictory to each other. Indeed, if a cancelable transformation is noninvertible, it generally degrades the verification performance. The transformations that improve the verification performance are generally invertible. A problem for future research could be to design a noninvertible cancelable transformation which improves verification performance.

Bibliography

[1] Online: http://svnext.it-sudparis.eu/svnview2-eph/ref_syst/.

[2] Multiprecision Integer and Rational Arithmetic C/C++ Library (MIRACL). Online: http://www.shamus.ie.

[3] United States Visitor and Immigrant Status Indicator Technology (US-VISIT). Ofiicial website: http://www.dhs.gov/files/programs/usv.shtm.

[4] Advanced Encryption Standard (AES), November 2001.

[5] Biosecure Tool: Performance Evaluation of A Biometric Verification System. Online: http://svnext.it-sudparis.eu/svnview2-eph/ref_syst/Tools/PerformanceEvaluation/doc/, 2007.

[6] BIOmetrics and crypTographY for Fair aUthentication Licensing (BIOTYFUL). Agence Nationale de la Recherche (ANR), 2007–2010. ANR-06-TCOM-018.

[7] Secure Hash Standard (SHS), October 2008.

[8] Mohamed Abid and Hossam Afifi. Towards a Secure e-Passport Protocol Based on Biometrics. *Journal of Information Assurance and Security (JIAS) (Special Issue on Access Control and Protocols)*, 4(4):338–345, 2009.

[9] Mohamed Abid, Sanjay Kanade, Dijana Petrovska-Delacrétaz, Bernadette Dorizzi, and Hossam Afifi. Iris Based Authentication Mechanism for e-Passports. In *2nd International Workshop on Security and Communication Networks (IWSCN)*, 2010.

[10] Andy Adler. Sample Images Can be Independently Restored from Face Recognition Templates. In *Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2003.

[11] Gaurav Aggarwal, Nalini K. Ratha, Jonathan H. Connell, and Ruud M. Bolle. Physics Based Revocable Face Recognition. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2008.

[12] Savvas Argyropoulos, Dimitrios Tzovaras, Dimosthenis Ioannidis, and Michael G. Strintzis. A Channel Coding Approach for Human Authentication From Gait Sequences. *IEEE Transactions on Information Forensics and Security*, 4(3):428 – 440, 2009.

[13] Lucas Ballard, Seny Kanmara, and Michael K. Reiter. The Practical Subtleties of Biometric Key generation. In *17th USENIX Security Symposium*, 2008.

[14] Elaine Barker, Don Johnson, and Miles Smid. NIST Special Publication 800-56A, Recommendation for Pair-Wise Key Establishment Schemes Using Discrete Logarithm Cryptography. Technical report, National Institute of Standards and Technology (NIST), March 2007.

[15] Mauro Barni, Tiziano Bianchi, Dario Catalano, Mario Di Raimondo, Ruggero Donida Labati, Pierluigi Failla, Dario Fiore, Riccardo Lazzeretti, Vincenzo Piuri, Fabio Scotti, and Alessandro Piva. Privacy-Preserving Fingercode Authentication. In *The 12th ACM Workshop on Multimedia and Security (MM&Sec10)*, Rome, Italy, Sept 2010.

[16] Claude Barral. *Biometrics & Security: Combining Fingerprints, Smart Cards and Cryptography.* PhD thesis, Ecole Polytechnique Fédérale de Lausannel, June 2010.

[17] Rima Belguechi and Christophe Rosenberger. Study on the Convergence of FingerHashing and a Secured Biometric System. In *Proceedings of the Confrence Internationale sur l'Informatique et ses Applications*, 2009.

[18] J. O. Berger. *A Stastical Decision Theory.* Springer-Verlag, 1980.

[19] J. R. Beveridge, D. Bolme, B. A. Raper, and M. Teixeira. The CSU Face Identification Evaluation System. *Machine Vision and Applications*, 16(2):128–138, 2005.

[20] Vishnu Naresh Boddeti, Fei Su, and B.V.K. Vijaya Kumar. A Biometric Key-Binding and template Protection Framework Using Correlation Filters. In

M. Tistarelli and M. Nixon, editors, *International Conference on Biometrics (ICB)*, pages 919–929, 2009.

[21] Ruud M. Bolle, Nalini K. Ratha, and Sharath Pankanti. Error Analysis of Pattern Recognition Systems – the Subsets Bootstrap. *Computer Vision and Image Understanding*, 93(1):1–33, January 2004.

[22] T. Boult. Robust Distance Measures for Face-Recognition Supporting Revocable Biometric Tokens. In *IEEE Conference on Automatic Face and Gesture Recognition (FG)*, 2006.

[23] T. E. Boult, W. J. Scheirer, and R. Woodworth. Revocable fingerprint biotokens: Accuracy and security analysis. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.

[24] Xavier Boyen. Reusable cryptographic fuzzy extractors. In *11th ACM Conference on Computer and Communications Security (CCS)*, 2004.

[25] Xavier Boyen, Yevgeniy Dodis, Jonathan Katz, Rafail Ostrovsky, and Adam Smith. Secure Remote Authentication Using Biometric Data. In *Eurocrypt*, 2005.

[26] J. Bringer, H. Chabanne, G. Cohen, B. Kindarji, and G. Zémor. Optimal Iris Fuzzy Sketches. In *IEEE Conference on Biometrics: Theory, Applications and Systems*, 2007.

[27] Julien Bringer, Hervé Chabanne, Malika Izabachène, David Pointcheval, Qiang Tang, and Sébastien Zimmer. An Application of the Goldwasser-Micali Cryptosystem to Biometric Authentication. In *The 12th Australasian Conference on Information Security and Privacy (ACISP '07)*, 2007.

[28] Julien Bringer, Hervé Chabanne, and Bruno Kindarji. The best of both worlds: Applying secure sketches to cancelable biometrics. *Science of Computer Programming*, 74(1–2):43–51, December 2008.

[29] Ileana Buhan. *Cryptographic Keys from Noisy Data*. PhD thesis, University of Twente, Netherlands, 2008.

[30] Ileana Buhan, Jeroen Doumen, Pieter Hartel, and Raymond Veldhuis. Secure Ad-hoc Pairing with Biometrics: SAfE. Technical report, University of Twente, 2007.

[31] William E. Burr, Donna F. Dodson, and W. Timothy Polk. Electronic authentication guideline: Recommendations of the National Institute of Standards and Technology, April 2006.

[32] R. Cappelli, A. Lumini, D. Maio, and D. Maltoni. Can Fingerprints be Reconstructed from ISO Templates? In *9th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, 2006.

[33] R. Cappelli, D. Maio, A. Lumini, and D. Maltoni. Fingerprint Image Reconstruction from Standard Templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(9):1489–1503, September 2007.

[34] Ann Cavoukian and Alex Stoianov. Biometric encryption: A positive-sum technology that achieves strong authentication, security and privacy. White paper, Information and privacy commissioner of Ontario, March 2007.

[35] C. Chen, R.N.J. Veldhuis, T.A.M. Kevenaar, and A.H.M. Akkermans. Biometric Binary String Generation with Detection Rate Optimized Bit Allocation. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2008.

[36] T. Charles Clancy, Negar Kiyavash, and Dennis J. Lin. Secure smartcard-based fingerprint authentication. In *Proceeding of ACM SIGMM Workshop on Biometrics Methods and Applications*, pages 45–52, November 2003.

[37] John Daugman. The importance of being random: Statistical principles of iris recognition. *Pattern Recognition*, 36(2):279–291, February 2003.

[38] John Daugman. How Iris Recognition Works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14:21–30, January 2004.

[39] G. I. Davida, Y. Frankel, and B. J. Matt. On enabling secure applications through off-line biometric identification. In *Proceedings of the IEEE Symposium on Privacy and Security*, pages 148–157, 1998.

[40] G. I. Davida, Y. Frankel, B. J. Matt, and R. Peralta. On the relation of error correction and cryptography to an offline biometric based identification scheme. In *Proc. Workshop on Coding and Cryptography*, pages 129–138, 1999.

[41] T. Dierks and C. Allen. The TLS Protocol, Version 1.0. Request for Comments: 2246, Internet Engineering Task Force (IETF), January 1999.

[42] T. Dierks and E. Rescorla. The Transport Layer Security (TLS) Protocol Version 1.2. Request for Comments: 5246, Internet Engineering Task Force (IETF), August 2008.

[43] Whitfield Diffie and Martin Hellman. New Directions in Cryptography. *IEEE Transactions on Information Theory*, 22(6):644–654, 1976.

[44] Y. Dodis, L. Reyzin, and A. Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In *Proceedings of the Eurocrypt 2004*, pages 523–540, 2004.

[45] Taher Elgamal. A public key cryptosystem and a siggnature scheme based on discrete logarithms. In *Crypto*, pages 10–18, 1984.

[46] EU. Advanced Security Mechanisms for Machine Readable Travel Documents. Technical Guideline TR-03110, 2010. Version 2.03.

[47] Faisal Farooq, Ruud M. Bolle, Tsai-Yang Jea, and Nalini Ratha. Anonymous and Revocable Fingerprint Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[48] Yi C. Feng, Pong C. Yuen, and Anil K. Jain. A Hybrid Approach for Generating Secure and Discriminating Face Template. *IEEE Transactions on Information Forensics and Security*, 5(1):103 –117, March 2010.

[49] Manuel R. Freire, Julian Fierrez, and Javier Ortega-Garcia. Dynamic Signature Verification with Template Protection Using Helper Data. In *International Conference on Acaustics, Speech, and Signal Processing (ICASSP)*, 2008.

[50] Bo Fu, Simon X. Yang, Jianping Li, and Dekun Hu. Multibiometric Cryptosystem:

Model Structure and Performance Analysis. *IEEE Transactions on Information Forensics and Security*, 4(4):867–882, 2009.

[51] Alwyn Goh and David C.L. Ngo. Computation of cryptographic keys from face biometrics. In A. Lioy and D. Mazzocchi, editors, *Proceedings of the International Federation for Information Processing*, volume 2828 of *Lecture Notes in Computer Science*, pages 1–13. Springer Berlin / Heidelberg, 2003.

[52] Shafi Goldwasser and Silvio Micali. Probabilistic Encryption and How to Play Mental Poker Keeping Secret All Partial Information. In *Proceedings of the Fourteenth Annual ACM Symposium on Theory of Computing*, 1982.

[53] Feng Hao. *On using fuzzy data in security mechanisms.* Phd thesis, Queens College, Cambridge, April 2007.

[54] Feng Hao, Ross Anderson, and John Daugman. Combining crypto with biometrics effectively. *IEEE Transactions on Computers*, 55(9):1081–1088, 2006.

[55] Feng Hao and Choong Wah Chan. Private key generation from on-line handwritten signatures. *Information Management and Computer Security*, 10(4):159–164, 2002.

[56] Lin Hong, Anil K. Jain, and Sharath Pankanti. Can Multibiometrics Improve Performance? In *Proceedings of IEEE Workshop on Automatic Identification Advanced Technologies (AutoID)*, pages 59–64, October 1999.

[57] International Civil Aviation Organization (ICAO). Part 1 - Machine Readable Passport - Volume 2: Specifications for Electronically Enabled Passports with Biometric Identification Capabilities. ICAO Doc 9303, 2006.

[58] ISO/IEC CD 2382.37. Information processing systems Vocabulary Part 37 : Harmonized Biometric Vocabulary, 2010.

[59] Y. Itakura and S. Tsujii. Proposal on a multifactor biometric authentication method based on cryptosystem keys containing biometric signatures. *International Journal of Information Security*, pages 288–296, 2005.

[60] Anil K. Jain, Karthik Nandakumar, and Abhishek Nagar. Biometric template security. *EURASIP Journal on Advances in Signal Processing*, 2008(Article ID 579416):17 pages.

[61] Andrew Teoh Beng Jin and Tee Connie. Remarks on biohashing based cancelable biometrics in verification system. *Neurocomputing*, 69(16-18):2461–2464, October 2006. Brain Inspired Cognitive Systems - Selected papers from the 1st International Conference on Brain Inspired Cognitive Systems (BICS 2004).

[62] Andrew Teoh Beng Jin, David Ngo, Chek Ling, and Alwyn Goh. Biohashing: two factor authentication featuring fingerprint data and tokenised random number. *Pattern Recognition*, 37(11):2245–2255, November 2004.

[63] A. Juels and M. Sudan. A fuzzy vault scheme. In A. Lapidoth and E. Teletar, editors, *Proc. IEEE Int. Symp. Information Theory*, page 408. IEEE Press, 2002.

[64] A. Juels and M. Wattenberg. A fuzzy commitment scheme. In *Proceedings of the Sixth ACM Conference on Computer and communication Security (CCCS)*, pages 28–36, 1999.

[65] Sanjay Kanade, Danielle Camara, Emine Krichen, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi. Three Factor Scheme for Biometric-Based Cryptographic Key Regeneration Using Iris. In *The 6th Biometrics Symposium (BSYM)*, September 2008.

[66] Sanjay Kanade, Emine Krichen, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi. Three Factor Scheme for Biometric-based Cryptographic Key Regeneration Using Iris. Technical report, Institut TELECOM: TELECOM SudParis, 2008.

[67] Sanjay Kanade, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi. Cancelable Iris Biometrics and Using Error Correcting Codes to Reduce Variability in Biometric Data. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2009.

[68] Sanjay Kanade, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi. Obtaining

Cryptographic Keys Using Feature Level Fusion of Iris and Face Biometrics for Secure User Authentication. In *IEEE CVPR Workshop on Biometrics*, June 2010.

[69] E. J. C. Kelkboom, B. Gökberk, T. A. M. Kevenaar, A. H. M. Akkermans, and M. van der Veen. "3D Face": Biometric Template Protection for 3D Face Recognition. In *International Conference on Biometrics*, 2007.

[70] Auguste Kerckhoffs. La Cryptographie Militaire. *Journal des Sciences Militaires*, 9:5–38, January, 161–191, February, 1883.

[71] T.A.M. Kevenaar, G.J. Schrijen, M. van der Veen, A.H.M. Akkermans, and F. Zuo. Face recognition with renewable and privacy preserving binary templates. In *Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, October 2005.

[72] Neal Koblitz. Elliptic Curve Cryptosystems 48, 1987, pp. 203209. *Mathematics of Computation*, 48:203–209, 1987.

[73] Adams Kong, King-Hong Cheung, David Zhang, Mohamed Kamel, and JaneYou. An analysis of biohashing and its variants. *Pattern Recognition*, 39(7):1359–1368, July 2006.

[74] Peter Kovesi. Matlab and octave functions for computer vision and image processing. Online: http://www.csse.uwa.edu.au/∼pk/Research/MatlabFns/, 2005.

[75] Emine Krichen, Bernadette Dorizzi, Zhenan Sun, Sonia Garcia-Salicetti, and Tieniu Tan. Iris Recognition. In Dijana Petrovska-Delacrétaz, Gérard Chollet, and Bernadette Dorizzi, editors, *Guide to Biometric Reference Systems and Performance Evaluation*, pages 25–50. Springer-Verlag, 2009.

[76] Martin Lades, Jan C. Vorbrüuggen, Joachim Buhmann, Jöorg Lange, Christoph v.d. Malsburg, Rolf P. Wüurtz, and Wolfgang Konen. Distortion Invariant Object Recognition in the Dynamic Link Architecture. *IEEE Transactions on Computers*, 42(3):300–311, March 1993.

[77] Qiming Li, Muchuan Guo, and Ee-Chien Chang. Fuzzy Extractors for Asymmetric Biometric Representations. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, (CVPRW '08)*, pages 1–6, June 2008.

[78] Alessandra Lumini and Loris Nanni. An improved biohashing for human authentication. *Pattern Recognition*, 40(3):1057–1065, March 2007.

[79] F. J. MacWilliams and N. J. A. Sloane. *Theory of Error-Correcting Codes*. North Holland, 1991.

[80] E. Maiorana, M. Martinez-Diaz, P. Campisi, J. Ortega-Garcia, and A. Neri. Template protection for hmm-based on-line signature authentication. In *IEEE CVPR Workshop on Biometrics*, 2008.

[81] Emanuele Maiorana, Patrizio Campisi, Julian Fierrez, Javier Ortega-Garcia, and Alessandro Neri. Cancelable Templates for Sequence Based Biometrics with Application to On-line Signature Recognition. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 40(3):525 – 538, 2010.

[82] Emanuele Maiorana, Patrizio Campisi, and Alessandro Neri. User adaptive fuzzy commitment for signature template protection and renewability. *Jounal of Electronic Imaging*, 17(1), 2008.

[83] Emanuele Maiorana, Patrizio Campisi, Javier Ortega-Garcia, and Alessandro Neri. Cancelable Biometrics for HMM-based Signature Recognition. In *IEEE Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2008.

[84] Wenbo Mao. *Modern Cryptography: Theory and Practice*. Prentice Hall, August 2003.

[85] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki. The DET Curve in Assessment of Detection Task Performance. In *Proceedings of the Eurospeech*, 1997.

[86] Microsoft Corporation. Windows Biometric Framework – Guidelines for IHV, ISVs and OEMs. Online, March 19 2009. http://www.microsoft.com/whdc/Device/biometric/WBFIntro.mspx.

[87] Preda Mihăilescu. The Fuzzy Vault for Fingerprints is Vulnerable to Brute Force Attack. Online, 2007. http://arxiv.org/abs/0708.2974v1.

[88] Victor Miller. Use of Elliptic Curves in Cryptography. In *Advances in Cryptology (CRYPTO '85)*, 1986.

[89] F. Monrose, M.K. Reiter, Qi Li, and S. Wetzel. Cryptographic key generation from voice. In *Proceedings of the IEEE Symposium on Security and Privacy*, pages 202–213, May 2001.

[90] F. Monrose, M.K. Reiter, and R. Wetzel. Password hardening based on keystroke dynamics. In *Proceedings of the Sixth ACM Conference on Computer and communication Security (CCCS)*, pages 73–82, 1999.

[91] Abhishek Nagar and Santanu Chaudhury. Biometrics based Asymmetric Cryptosystem Design Using. Modified Fuzzy Vault Scheme. In *18th International Conference on Pattern Recognition (ICPR)*, 2006.

[92] Abhishek Nagar and Anil K. Jain. On the Security of Non-Invertible Fingerprint Template Transforms. In *IEEE International Workshop on Information Forensics and Security*, December 2009.

[93] Abhishek Nagar, Karthik Nandakumar, and Anil K. Jain. Securing fingerprint template: Fuzzy vault with minutiae descriptors. In *International Conference on Pattern Recognition (ICPR)*, pages 1–4, 2008.

[94] Abhishek Nagar, Karthik Nandakumar, and Anil K. Jain. A Hybrid Biometric Cryptosystem for Securing Fingerprint Minutiae Templates. *Pattern Recognition Letters*, 31(8):733–741, 2009.

[95] Karthik Nandakumar. *Multibiometric Systems: Fusion Strategies and Template Security*. Phd thesis, Department of Computer Science and Engineering, Michigan State University, 2008.

[96] Karthik Nandakumar and Anil K. Jain. Multibiometric template security using fuzzy vault. In *IEEE Second International Conference on Biometrics: Theory, Applications and Systems*, 2008.

[97] Karthik Nandakumar, Anil K. Jain, and Sharath Pankanti. Fingerprint-based fuzzy vault: Implementation and performance. *IEEE Transactions of Information Forensics and Security*, 2(4):744–757, December 2007.

[98] Karthik Nandakumar, Abhishek Nagar, and Anil K. Jain. Hardening Fingerprint Fuzzy Vault Using Password. In *Proceedings of the International Conference on Biometrics*, 2007.

[99] Loris Nanni and Alessandra Lumini. Random subspace for an improved BioHashing for face authentication. *Pattern Recognition Letters*, 29:295–300, 2008.

[100] National Institute of Science and Technology (NIST). Face Recognition Grand Challenge, 2005. http://www.frvt.org/FRGC/.

[101] National Institute of Science and Technology (NIST). Iris Challenge Evaluation, 2005. http://iris.nist.gov/ice.

[102] Vijayakrishnan Pasupathinathan, Josef Pieprzyk, and Huaxiong Wang. An Online Secure e-Passport Protocol. In *Proceedings of the 4th International Conference on Information Security Practice and Experience*, pages 14–28, 2008.

[103] Dijana Petrovska-Delacrétaz, Gérard Chollet, and Bernadette Dorizzi, editors. *Guide to Biometric Reference Systems and Performance Evaluation.* Springer-Verlag, 2009.

[104] N. K. Ratha, J. H. Connell, and R. M. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM Systems Journal*, 40(3):614–634, 2001.

[105] Nalini K. Ratha, Sharat Chikkerur, Jonathan H. Connell, and Ruud M. Bolle. Generating cancelable fingerprint templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):561–572, April 2007.

[106] Ronald Rivest, Adi Shamir, and Len Adleman. A Method for Obtaining Digital Signatures and Public-key Cryptosystems. *Communications of the ACM*, 21(2):120–126, 1978.

[107] A. Ross and A. K. Jain. Multimodal biometrics: an overview. In *Proceedings of the 12th European Signal Processing Conference (EUSIPCO)*, pages 1221–1224, September 2004.

[108] Arun Ross. An Introduction to Multibiometrics. In *Proceedings of the 15th Eurpoean Signal Processing Conference (EUSIPCO)*, 2007.

[109] Arun Ross, Jidnya Shah, and Anil K. Jain. From template to image: Reconstructing fingerprints from minutiae points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):544–560, April 2007.

[110] Arun Ross, Samir Shah, and Jidnya Shah. Image Versus Feature Mosaicing: A Case Study in Fingerprints. In *SPIE Conference on Biometric Technology for Human Identification*, 2006.

[111] Arun A. Ross, Karthik Nandakumar, and Anil K. Jain. *Handbook of Multibiometrics*. International Series on Biometrics. Springer, 2006.

[112] A. Sahai and B. Waters. Fuzzy Identity-Based Encryption. In *EUROCRYPT*, pages 457–473, 2005.

[113] Marios Savvides, B.V.K. Vijaya Kumar, and P.K. Khosla. Cancelable biometric filters for face recognition. In *Proceedings of the 17th International Conference on Pattern Recognition (ICPR04)*, volume 3, pages 922–925, August 2004.

[114] Tobias Scheidat, Claus Vielhauer, and Jana Dittmann. Advanced Studies on Reproducibility of Biometric Hashes. In B. Schouten et al., editor, *Biometrics and Identity Management (BIOID)*, 2008.

[115] W. J. Scheirer and T. E. Boult. Bio-Cryptographic Protocols with Bipartite Biotokens. In *Biometric Symposium*, 2008.

[116] W. J. Scheirer and T. E. Boult. Bipartite Biotokens: Definitions, Implementation, and Analysis. In *International Conference on Biometrics (ICB)*, 2009.

[117] Walter J. Scheirer and Terrance E. Boult. Cracking Fuzzy Vaults and Biometric Encryption. In *Biometrics Symposium*, 2007.

[118] Weiguo Sheng, Gareth Howells, Michael Fairhurst, and Farzin Deravi. Template-free biometric-key generation by means of fuzzy genetic clustering. *IEEE Transactions on Information Forensics and Security*, 3(2):183–191, June 2008.

[119] Bernard Sklar. Reed-solomon codes. Available online (16 June 2010). www.phptr.com/content/images/art_sklar7_reed-solomon/elementLinks/art_sklar7_reed-solomon.pdf.

[120] C. Soutar, D. Roberge, A. Stoianov, R. Gilroy, and B.V.K. Vijaya Kumar. Biometric encryption. In *ICSA guide to Cryptography*. McGraw-Hill, 1999.

[121] A. Stoianov, T. Kevenaar, and M. van der Veen. Security Issues of Biometric Encryption. In *IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH)*, 2009.

[122] Alex Stoianov. Security of Error Correcting Code for Biometric Encryption (critical note). In *Eighth Annual International Conference on Privacy, Security and Trust*, 2010.

[123] Y. Sutcu, Qiming Li, and N. Memon. Secure biometric templates from fingerprint-face features. In *IEEE Conference on Computer Vision and Pattern Recognition, 2007*, pages 1–6, June 2007.

[124] Yagiz Sutcu, Qiming Li, and Nasir Memon. Protecting biometric templates with sketch: Theory and practice. *IEEE Transactions on Information Forensics and Security*, 2(3):503–512, September 2007.

[125] Qiang Tang, Julien Bringer, Hervé Chabanne, and David Pointcheval. A Formal Study of the Privacy Concerns in Biometric-Based Remote Authentication Schemes. In *Information Security Practice and Experience Conference (ISPEC)*, 2008.

[126] Andrew Beng Jin Teoh and Chong Tze Yuang. Cancelable biometrics realization with multispace random projections. *IEEE Transactions on Systems, Man, and Cybernetics, Part B-Cybernetics*, 37(5):1096–1106, October 2007.

[127] Andrew B.J. Teoh, Alwyn Goh, and David C.L. Ngo. Random Multispace Quantization as an Analytic Mechanism for BioHashing of Biometric and Random Identity Inputs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):1892 – 1901, December 2006.

[128] Andrew B.J. Teoh, Yip Wai Kuan, and Sangyoun Lee. Cancellable Biometrics and Annotations on BioHash. *Pattern Recognition*, 41(6):20342044, June 2008.

[129] Andrew B.J. Teoh and David C.L. Ngo. Cancellable Biometerics Featuring with Tokenised Random Number. *Pattern Recognition Letters*, 26(10):1454–1460, July 2005.

[130] Andrew B.J. Teoh, David C.L. Ngo, and Alwyn Goh. Personalised cryptographic key generation based on facehashing. *Computers & Security*, 23:606–614, 2004.

[131] Achint O. Thomas, Nalini K. Ratha, Jonathan H. Connell, and Ruud M. Bolle. Comparative analysis of registration based and registration free methods for cancelable fingerprint biometrics. In *International Conference on Pattern Recognition (ICPR)*, 2008.

[132] Valrie Viet Triem Tong, Herv Sibert, Jrmy Lecoeur, and Marc Girault. Biometric fuzzy extractors made practical: A proposal based on fingercodes. In Seong-Whan Lee and Stan Z. Li, editors, *Proceedings of ICB*, pages 604–613, 2007.

[133] Pim Tuyls, Anton H.M. Akkermans, Tom A.M. Kevenaar, Geert-Jan Schrijen, Asker M. Bazen, and Raymond N.J. Veldhuis. Practical Biometric Authentication with Template Protection. In *Audio- and Video-Based Biometric Person Authentication (AVBPA)*, 2005.

[134] Pim Tuyls and Jasper Goseling. Capacity and Examples of Template-Protecting Biometric Authentication Systems. In D. Maltoni and Anil K. Jain, editors, *Biometric Authentication Workshop*, 2004.

[135] Pim Tuyls, Boris Škorić, and Tom Kevenaar, editors. *Security with Noisy Data*. Springer, 2007.

[136] Yoshifumi Ueshige and Kouichi Sakurai. A Proposal of One-Time Biometric Authentication. In H. R. Arabnia and S. Aissi, editors, *Security and Management*, 2006.

[137] U. Uludag and A. Jain. Securing fingerprint template: Fuzzy vault with helper data. In *Proc. of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*, pages 163–170, June 2006.

[138] U. Uludag, S. Pankanti, S. Prabhakar, and A.K. Jain. Biometric cryptosystems: issues and challenges. *Proceedings of the IEEE*, 92(6):948–960, June 2004.

[139] Umut Uludag. *Secure Biometric Systems.* Phd thesis, Michigan State University, 2006.

[140] Umut Uludag and Anil K. Jain. Fuzzy fingerprint vault. In *Proc. Workshop: Biometrics: Challenges Arising from Theory to Practice*, pages 13–16, August 2004.

[141] Michiel van der Veen, Tom Kevenaar, Geert-Jan Schrijen, Ton H. Akkermans, and Fei Zuo. Face biometrics with renewable templates. In *Security, Steganography, and Watermarking of Multimedia Contents VIII*, January 2006.

[142] Claus Vielhauer, Ralf Steinmetz, and Astrid Mayerhöfer. Biometric hash based on statistical features of online signatures. In *Proceedings of the 16th International Conference on Pattern Recognition*, 2002.

[143] Zhifang Wang, Qi Han, Xiamu Niu, and Christoph Busch. A Novel Template Protection Algorithm for Iris Recognition. In *International Conference on Intelligent Systems Design and Applications (ISDA)*, 2008.

[144] Shenglin Yang and Ingrid Verbauwhede. Automatic Secure Fingerprint Verification System Based on Fuzzy Vault Scheme. In *International Conference on Acoustics, Speech, and Signal Processing*, 2005.

[145] Shenglin Yang and Ingrid Verbauwhede. Secure iris verification. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-2007)*, volume 2, pages II–133 – II–136, April 2007.

[146] Shenglin Yang and Ingrid M. Verbauwhede. Secure fuzzy vault based fingerprint verification system. In *Thirty-Eighth Asilomar Conference on Signals, Systems and Computers*, 2004.

[147] Gang Zheng, Wanqing Li, and Ce Zhan. Cryptographic key generation from biometric data using lattice mapping. In *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, pages 513–516, Washington, DC, USA, 2006. IEEE Computer Society.

[148] Xuebing Zhou. Template protection and its implementation in 3d face recognition systems. *Proceedings of the SPIE*, 6539:65390L, 2007.

[149] Xuebing Zhou and Christoph Busch. A novel privacy enhancing algorithm for biometric system. In *Proceedings of the Special Interest Group on Biometrics and Electronic Signatures*, September 2008.

[150] Jinyu Zuo, Nalini K. Ratha, and Jonathan H. Connell. Cancelable Iris Biometric. In *International Conference on Pattern Recognition (ICPR)*, 2008.

Appendix A

Baseline Biometric Systems, Databases, and experimental Protocols

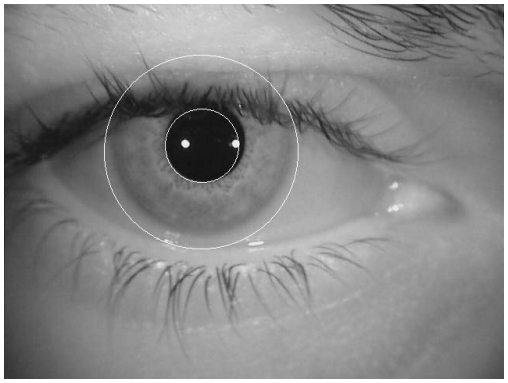
A.1 Baseline Systems Used for Extracting Features from Biometric Data

A.1.1 Baseline Open Source Iris System – OSIRISv1

The Open Source Iris Recognition System – OSIRISv1 described in [75] (available online at [1]) – is used for extracting binary iris code features from iris images. The circular iris region first needs to be detected correctly using a process called iris segmentation. In order that our system is not influenced by segmentation errors, we manually adjusted the segmentation if necessary. The segmented circular iris region is then converted into a rectangular image of fixed size using Daugman's rubber sheet model [38]. In our experiments, the size of the normalized images is 512×80 pixels. The normalized image is decomposed using Gabor filters and phase information from the decomposed images is binarized and concatenated to form a one-dimensional binary iris code. In all our experiments, the extracted iris codes are 1,188 bits long. Figure A.1 shows an iris image at different processing levels.

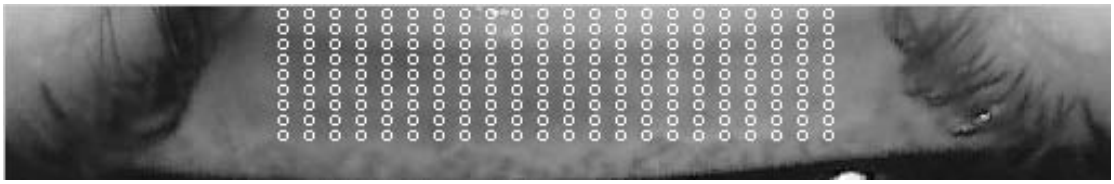
The iris codes obtained from the reference and test iris images are compared

(a) Raw iris image



(b) Segmented image showing the location of the iris

(c) Normalized iris image



(d) Normalized iris image showing the locations at which the binary features are calculated

Figure A.1: Illustration of processing of an iris image: (a) raw iris image, (b) segmented iris image, (c) normalized iris image, and (d) normalized iris image with the locations where Gabor filters are applied for binary feature extraction.

using the Hamming distance. Generally, in such systems, noise masks are employed to eliminate noisy portions of the iris image from comparison. Equation (A.1) gives the formula for calculating the Hamming distance.

$$\text{Hamming distance} = \frac{\|(codeA \oplus codeB) \cap maskA \cap maskB\|}{\|maskA \cap maskB\|}. \tag{A.1}$$

In this equation, *codeA* and *codeB* are the iris codes from imageA and imageB whereas maskA and maskB are their respective noise masks. From this equation, it is clear that the reference noise mask is required to be logically AND*ed* with the test mask

for Hamming distance calculation and therefore, needs to be stored. In our crypto-biometric systems, this may leak information about the reference biometric template. Therefore we do not take the noise masks into consideration. This makes the crypto-biometric system design more challenging. Instead, we use a static mask which eliminates the probable noisy regions such as the possible locations of eye-lids and eye-lashes. An example of a normalized image along with the fixed mask is shown in Fig. A.1(d).

During acquisition, there is a possibility of rotation of iris due to head tilt. In order to cope with such iris rotations, the OSIRISv1 shifts the normalized test iris image horizontally in both the directions. The iris codes extracted from such rotated images are compared with the reference iris code using the Hamming distance metric and the minimum of these Hamming distances is considered for comparison with a threshold. In OSIRISv1, the normalized test iris image is shifted 10 times in both directions thus resulting in 21 comparisons. For our crypto-biometric systems, we need only the binary feature vectors from OSIRISv1. Therefore, the rotation adjustment module is re-implemented and validated experimentally.

A.1.2 Baseline Face System

The key regeneration systems proposed in this thesis are based on the fuzzy commitment scheme [64]. It requires that the biometric data are represented as an ordered set of binary values. In order to extract a binary face code from the face images, first a Gabor filter based approach is employed [76]. The face image is geometrically normalized using program from the CSU Face Recognition Evaluation System [19], and then processed using log-Gabor filters having four scales and eight orientations using the MATLAB source code available at [74]. Magnitude of the filtered output is calculated, downsampled, and concatenated to form a 3,200-element feature vector. The values of this vector are then binarized to obtain a 3,200-bit string called face code. The binarization process used is very simple. The median of the values of each feature vector is taken as a threshold. The elements having higher value than the threshold are converted to one while the remaining are converted to zeros.

The reason for using such a basic approach for a baseline face recognition system is that, in order to obtain long keys with the fuzzy commitment scheme, the length of the feature vector should be high. There are many other baseline face systems

such as the SudParis Face Recognition system (SudFROG) [103] which perform better than the one used in this thesis. But, in SudFROG, after applying the Gabor filters to the normalized face images, dimensionality reduction technique called DLDA are used. This results in the feature vector with very small size. Therefore, it cannot be used for our key regeneration systems in its present status.

Some other possible binarization mechanisms (which are applied to fingerprints) are summarized in [135]. These techniques are: simple binarization, reliable component selection [71, 69], and Detection Rate Optimized Bit Allocation (DROBA) [35]. These binarization mechanisms take multiple samples as input and use statistical characteristics such as mean and standard deviation of the genuine distribution for determining the quantization threshold. Using these binarization methods can improve the verification performance of the system. However, these methods require storage of user specific auxiliary information needed for the binarization step during verification. This auxiliary information is highly sensitive and may leak information about the biometric data and user identity. Indeed, the stored auxiliary information from different systems can be used for cross-matching between databases thus compromising privacy, one of the main goals of the proposed systems.

When compared to these approaches, the binarization method used in our system has an advantage that it does not require storage of any user specific information for binarization. In this way, the problem of information leakage is eliminated. Moreover, the median of the feature vector is used as binarization threshold. This ensures that there are nearly equal number of zeros and ones in the binary feature vector.

A.2 Databases and Experimental Protocols

Evaluation of the systems developed in this thesis is carried out by testing the system on publicly available databases. Moreover, the development data set is different than the test data set without overlap. Various parameters of the crypto-biometric systems, such as, length of the shuffling key, choice of ECC and error correction capacity, are tuned on the development database. The systems are later tested on the evaluation data sets using these parameters. The databases used for iris as well as face, along with their associated experimental protocols, are described in the following subsections.

A.2.1 Iris Databases and Experimental Protocols

For iris, two databases are used namely Casia-BioSecure (CBS) database [75] (OKI device subset) for development and NIST-ICE database [101] for performance evaluation.

The CBS database is composed of two parts: CBS-BiosecureV1 and CBS-CasiaV2. Each of these two parts contains 1,200 images from 60 eyes of 30 persons with 20 images from each eye. According to the protocol described in [75], each of these data sets is divided into two parts as:

1. Enrollment data set consisting of the first 10 images of each eye, and
2. Test data set composed of the remaining 10 images.

For intra-class comparisons, the 10 images from enrollment data set are compared to the 10 images of the same eye from the test data set. For inter-class comparisons, the enrollment data set images are compared with 10 randomly selected images from other eyes from the test data set. This results in comparisons between images obtained in different sessions and different illumination conditions, and between images of eyes with and without glasses. In total, there are 6,000 genuine and 6,000 impostor comparisons for the two data sets – CBS-BiosecureV1 and CBS-CasiaV2.

Once the system parameters are tuned on the CBS database, the system is evaluated on the NIST-ICE database [101]. This database consists of 2,953 images obtained from 244 eyes of 132 users. 124 users have recorded right eye images while 120 have recorded left eye images with 112 users being present in both these sets. As described in the ICE protocol [101], two different experiments are carried out for this database. Experiment-1 (ICE-exp1) consists of comparison of right eye images while Experiment-2 (ICE-Exp2) consists of left eye comparisons. All possible comparisons between image pairs are carried out for each of these experiments. In total, 12,214 genuine and 1,002,386 impostor comparisons were carried out in ICE-Exp1, whereas in ICE-exp2, 14,653 genuine, and 1,151,975 impostor comparisons were performed.

A.2.2 Face Database and Experimental Protocols

For our experiments on face modality, a subset of the FRGCv2 (Face Recognition Grand Challenge version 2) face database [100] is selected. The full FRGCv2

(a) An example image from the controlled set (b) An example image from the non-controlled set

Figure A.2: Examples of images from the FRGCv2 database: (a) an image from the controlled set, and (b) an image from the non-controlled set of the same subject.

database contains images from 466 subjects and is composed of 16,028 controlled still images captured under controlled conditions and 8,024 non-controlled still images captured under uncontrolled lighting conditions. There are two types of expressions: smiling and neutral and a large time variability exists. There are many experiments defined in order to evaluate the performance of algorithms for different sets of parameters.

The crypto-biometric systems developed in this thesis have much more computational complexity than the baseline biometric systems. Therefore, a significantly more amount of time is required to carry out the performance evaluation. Moreover, there are multiple test that need to be carried out for different parameters of the crypto-biometric systems. This poses practical difficulties for running the full experimental protocols defined for the FRGCv2 database. In order to reduce the number of comparisons, we selected a subset of the FRGCv2 database for our experiments.

Our data set consists of 250 subjects each of which has 12 facial images. Out of these 250 users, first 125 users are used for development and the remaining 125 are

used for evaluation purposes, thus there is no overlap between the development and evaluation data sets. The images in FRGCv2 database are captured under two different acquisition conditions: controlled and uncontrolled. The controlled images were taken in a studio setting, are full frontal facial images taken under two lighting conditions and with two facial expressions (smiling and neutral). The uncontrolled images were taken in varying illumination conditions; e.g., hallways, atriums, or outside. Each set of uncontrolled images contains two expressions, smiling and neutral. Examples of face images from controlled and uncontrolled sets are shown in Fig. A.2. Among the 12 images of a person, first eight are from the controlled set and remaining four are from the uncontrolled set. Two separate experiments are carried out: FRGC-exp1* – in which, the images from the controlled set are taken for enrollment as well as test; and FRGC-exp4* – where the enrollment images are taken from the controlled set and those from uncontrolled set are used as tests (query). We put the star (*) in order to stress the fact that these experiments are not exactly the same as described in the full FRGCv2 protocols. For the FRGC-exp1*, 3,500 genuine and 496,000 impostor comparisons are carried out while for FRGC-exp4*, 4,000 genuine and 496,000 impostor comparisons are done.

A.2.3 Two Iris Protocol

In this thesis, a multi-unit type multi-biometric based cryptographic key regeneration system is proposed (Section 7.3, page–127). It combines information from left and right irises of a person for key regeneration. As it is done for single iris based systems, the multi-iris based system is also developed on the CBS database [75] and then evaluated on the NIST-ICE database [101]. The development set is required mainly for setting the error correction capacities of the error correcting codes used in the system.

The CBS database protocol described earlier in Section A.2.1 is for single iris comparisons. For the multi-iris comparisons, the same structure is followed thus resulting in exactly half the number of tests than for the single eye protocol.

The CBS-BiosecureV1 OKI device subset is used for the experiments. In this subset, there 30 subjects each having 20 images of each eye. Image pairs, comprising of images of right and left irises of a person, are created. The first 10 image pairs are taken for enrollment and the remaining are used during verification. For genuine

comparisons, the 10 enrollment image pairs are compared with the 10 verification image pairs of the same person. For impostor tests, the 10 enrollment image pairs are compared with 10 randomly selected image pairs of other persons. Thus there are 3,000 genuine comparisons and 3,000 impostor comparisons.

In the NIST-ICE database [101], there are 132 subjects out of which, only 112 subjects have recorded images of their both eyes. We select images of these 112 subjects for carrying out our tests. The right iris images are coupled with the left iris images for the multi-iris tests. The first such image pair of a person is considered for enrollment and a template is registered for that person. The genuine comparisons are carried out by comparing the remaining images pairs of that subject with the enrollment template which results in 1,099 genuine comparisons. For impostor comparisons, one image pair from each of the remaining subjects is randomly selected and these image pairs are compared with the enrollment template. Thus, for each person, we carry out 111 impostor comparisons. In summary, 1,099 genuine and 12,432 impostor comparisons are carried out on the NIST-ICE database for the two-iris experiment.

A.2.4 Iris-Face Protocol

In order to evaluate the multi-modal biometric based key regeneration system described in Section 7.4 (page–134), we used a virtual database created from two publicly available databases: the NIST-ICE database [101] for iris, and the NIST-FRGCv2 database [100] for face described in previous sections. These two databases have different number of users and different number of images per subject. In order to evaluate the multi-biometric system, we needed a data set that has equal number of users for both the modalities and equal number of samples per modality. Daugman [38] has shown that the right and left eyes of a person are uncorrelated. Following this assumption, we treated them as different identities in order to have a higher number of subjects. The NIST-ICE database has images from 244 different eyes but a number of those eyes have only one image. For our experiments, we selected only those eyes for which we have at least five images. There are 175 such different eyes resulting in a total of $175 \times 5 = 875$ iris images. Similarly we selected 875 face images corresponding to 175 users. Then we created 175 virtual users having iris images taken from the NIST-ICE database and face images taken from the NIST-FRGCv2 database. In the NIST-FRGCv2 database, there

are face images captured in two scenarios: controlled and uncontrolled. We selected images only from the controlled data-set for our experiments.

For each subject, data pairs are formed containing one iris image and one face image corresponding to that subject. Thus, we have five such pairs per subject for 175 subjects. For genuine comparisons, each data pair is compared with every other data pair corresponding to the same subject. Similarly, each data pair is compared with every other data pair of every other subject. This protocol results in 1,750 genuine comparisons and 380,625 impostor comparisons. For the sake of fair comparison with uni-biometric systems, similar protocol is followed to test the uni-biometrics based systems in Chapter 7.

Appendix B

Biosecure Tool for Performance Evaluation

The performance of the biometric systems (in terms of FAR, FRR, and EER along with their error margins for 90% confidence intervals is calculated with the help of the Biosecure performance evaluation tool [103]. The method for estimating the confidence intervals in this tool is described below.

B.1 Parametric Confidence Interval Estimation

It is clear that, in a biometric system performance evaluation experiment, it is not possible to cover the whole human population. The score distribution that we obtain is a subset of the complete population. Therefore, the verification error rates (e.g., EER, OP) obtained during the evaluation can be erroneous. In order to predict the possible error margins, a 90% interval of confidence is calculated. The method described here is the one proposed by Bolle et al. [21].

Suppose we have M client scores and N impostor scores. We denote these sets of scores by $\mathbf{X} = \{X_1, ..., X_M\}$ and $\mathbf{Y} = \{Y_1, ..., Y_N\}$ respectively. In the following, we assume that available scores are similarity measures.

Let $\mathbf{S}$ be the set of thresholds used to calculate the score distributions.

For the set of client scores, $\mathbf{X}$, assume that this is a sample of M numbers drawn from a population with distribution F, that is, $F(x) = Prob(X \leq x), x \in \mathbf{S}$.

Let the impostor scores $\mathbf{Y}$ be a sample of N numbers drawn from a population

with distribution $G(y) = Prob(Y \leq y), y \in S$.

In this way, $FRR(x) = F(x)$ and $FAR(y) = 1 - G(y)$, x and $y \in \mathbf{S}$. From now, we have to find an estimate of these distributions at some threshold $t_0 \in \mathbf{S}$ and then, we have to estimate the confidence interval for these estimations.

- The estimate of $F(t_0)$ using data $\mathbf{X}$ is the unbiased statistic:

$$\hat{F}(t_0) = \frac{1}{M} \sum_{i=1}^{M} 1(X_i \leq t_0). \tag{B.1}$$

Thus, $\hat{F}(t_0)$ is obtained by simply counting the $X_i \in \mathbf{X}$ that are smaller than t_0 and dividing by M.

In the same way, the estimate $G(t_0)$ using the data $\mathbf{Y}$ is the unbiased statistic:

$$\hat{G}(t_0) = \frac{1}{N} \sum_{i=1}^{N} 1(Y_i \leq t_0). \tag{B.2}$$

- In the following, let us concentrate on the distribution F. For the moment, let us keep $x = t_0$ and let us determine the confidence interval for $\hat{F}(t_0)$.

First define Z as a binomial random variable, the number of successes, where success means $(X \leq t_0)$ is true, in M trials with probability of success $F(t_0) = Prob(X \leq t_0)$. This random variable Z has binomial probability mass distribution:

$$P(Z = z) = \binom{M}{z} F(t_0)^z (1 - F(t_0))^{M-z}, \quad z = 0, ..., M. \tag{B.3}$$

The expectation of Z is $E(Z) = MF(t_0)$ and the variance is $\sigma^2(Z) = MF(t_0)(1 - F(t_0))$.

From this, it follows that the random variable Z/M has expectation $F(t_0)$ and variance $F(t_0)(1 - F(t_0))/M$. When M is large enough, using the law of large numbers, Z/M is distributed according to a normal distribution, i.e., $Z/M \sim \mathcal{N}(F(t_0), F(t_0)(1 - F(t_0))/M)$.

Now it can be seen that, $\hat{Z}/M = \hat{F}(t_0)$. Hence, for large M, $\hat{F}(t_0)$ is normally distributed with an estimate of the standard deviation given by:

$$\hat{\sigma}(t_0) = \sqrt{\frac{(\hat{F}(t_0))(1 - \hat{F}(t_0))}{M}}. \tag{B.4}$$

The confidence intervals can be determined from $\hat{\sigma}(t_0)$. For example, a 90% interval of confidence is:

$$F(t_0) \in [\hat{F}(t_0) - 1.645\hat{\sigma}(t_0), \hat{F}(t_0) + 1.645\hat{\sigma}(t_0)]. \tag{B.5}$$

Estimate $\hat{G}(t_0)$ for the probability distribution $G(t_0)$ using a set of impostor scores $\mathbf{Y}$ can be obtained in a similar fashion. Parametric confidence intervals for this estimate can be calculated by replacing $\hat{F}(t_0)$ with $\hat{G}(t_0)$ and M with N in equations (B.4) and (B.5).

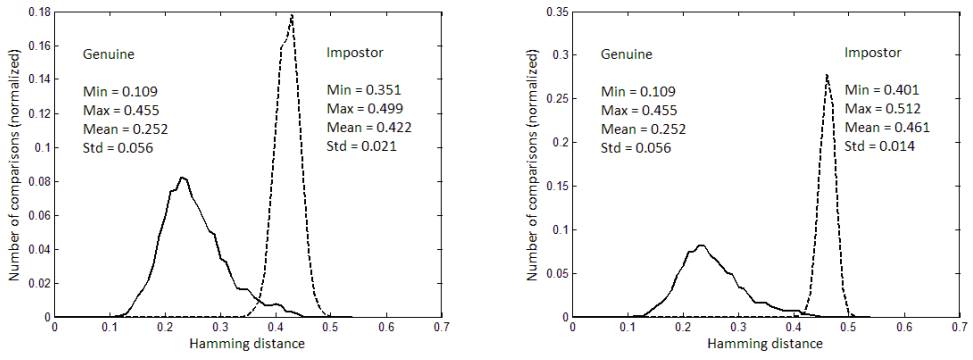
Appendix C

Additional Results

Some additional results and plots are reported in this section.

C.1 Additional Results from Chapter 4

The Hamming distance curves for the baseline biometric system and the shuffling based cancelable biometric system proposed in Chapter 4 on the CBS-CasiaV2 development data set are shown in Fig. C.1. As in the case of CBS-BiosecureV1, the mean of impostor Hamming distance distribution increases (from 0.42 to 0.46) as a result of the shuffling.

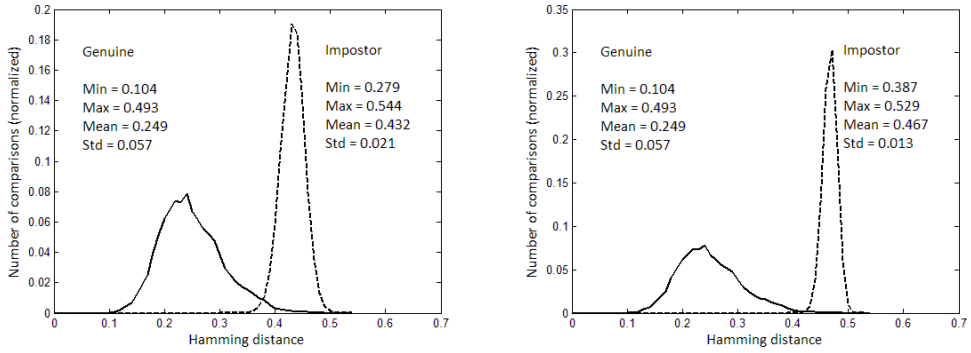


(a) Baseline iris biometric system (b) Baseline iris system with shuffling

Figure C.1: Normalized Hamming distance distributions for genuine and impostor comparisons on the CBS-CasiaV2 development data set [103].

Hamming distance distributions for the baseline biometric system and the shuf-

fling based cancelable biometric system proposed in Chapter 4 on the NIST-ICE evaluation data set, ICE-Exp2 (left-eye experiment) are shown in Fig. C.2.



(a) Baseline iris biometric system (ICE-Exp2) (b) Baseline iris system with shuffling (ICE-Exp2)

Figure C.2: Normalized Hamming distance distributions for genuine and impostor comparisons on the NIST-ICE [101] evaluation database for ICE-Exp1.

Detection Error Tradeoff (DET) curves for the proposed shuffling based cancelable biometric system along with the security threats are shown in Fig. 4.4 for the iris modality; evaluation data set (NIST-ICE, ICE-Exp2 experiment).

C.2 Additional Results from Chapter 5

The Hamming distance distribution plots for these three tests on the CBS-CasiaV2 (development) data sets are shown in Fig. C.6.

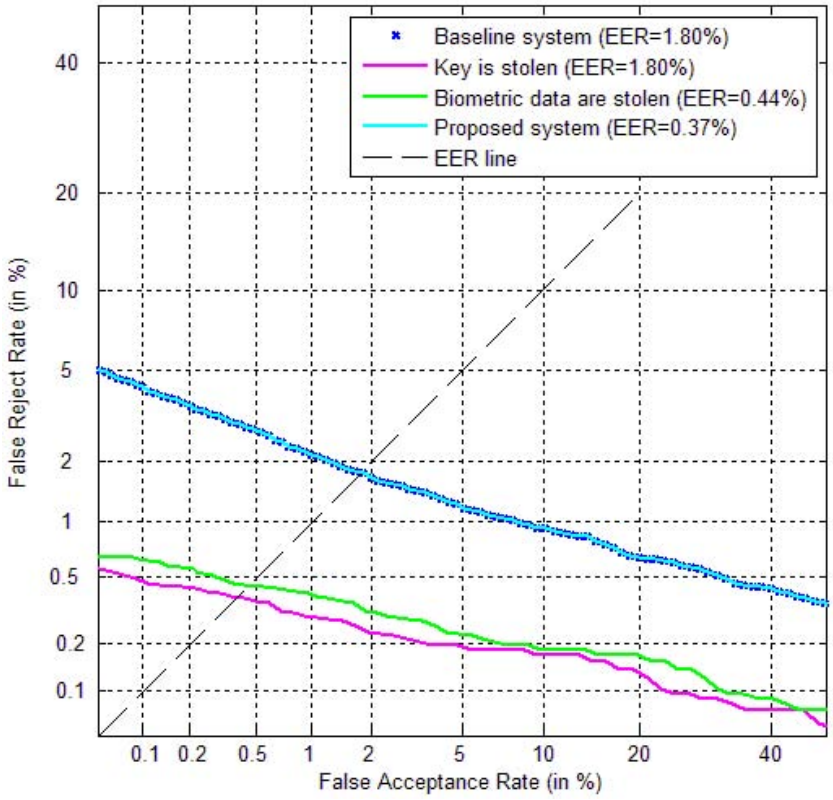


Figure C.3: DET curves for the proposed system performance along with the possible security threats for iris modality on the NIST-ICE database [101].

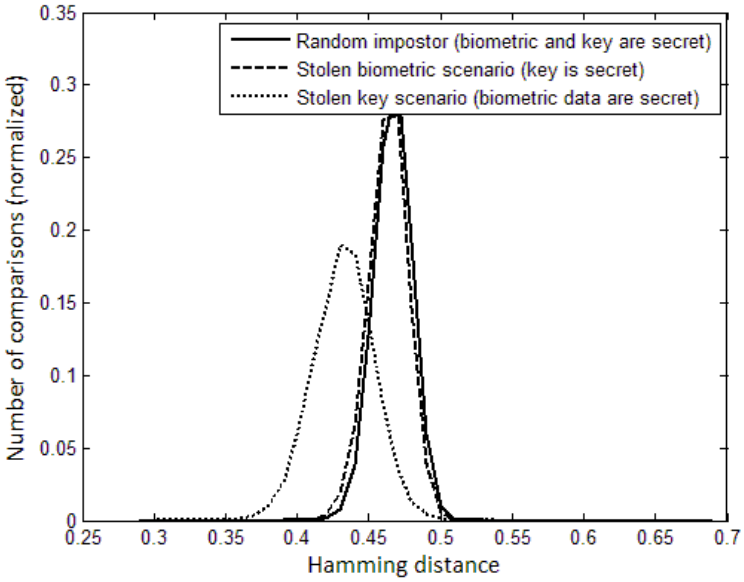


Figure C.4: Impostor Hamming distance distributions for the proposed system along with the possible security threats for iris modality on the NIST-ICE database [101] (ICE-Exp2).

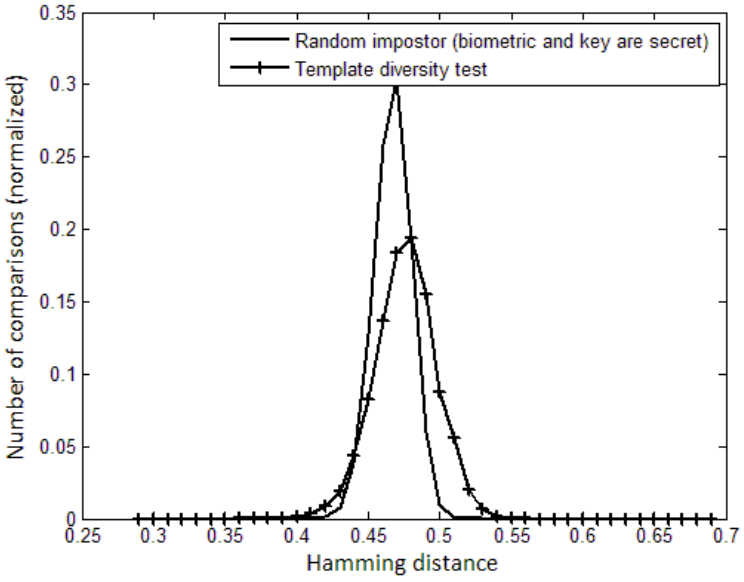
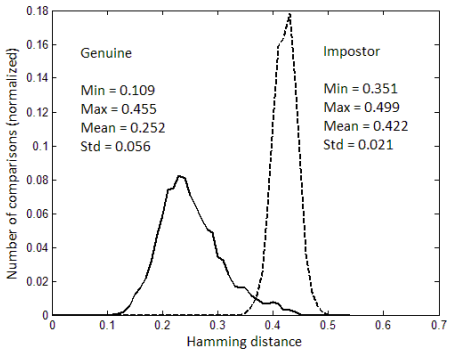
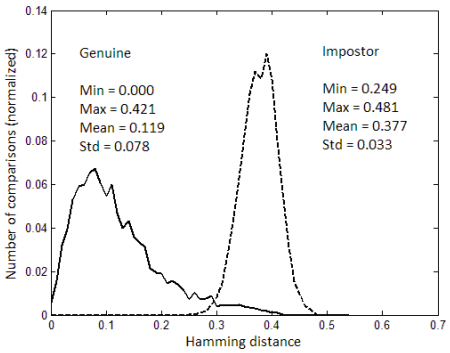


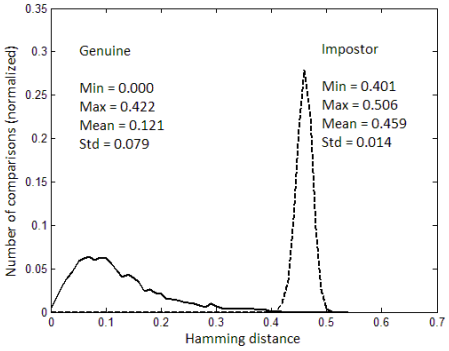
Figure C.5: Impostor Hamming distance distributions for the proposed system along with the Hamming distance distributions for the template diversity test on iris modality on the NIST-ICE database [101] (ICE-Exp2).



(a) Baseline iris biometric system



(b) Baseline iris system with error correction



(c) The proposed system (baseline iris biometric system with error correction and shuffling

Figure C.6: Normalized Hamming distance distributions for genuine and impostor comparisons on the development data set (CBS-CasiaV2 database [103]).

Appendix D

Majority Coding approach for feature variation reduction

Davida et al. [39, 40] and Soutar et al. [120] have suggested majority coding to reduce feature variations. Majority coding extracts a stable template from multiple input samples. For a particular bit position, majority voting is applied and the bit is set to one or zero depending on the bit values in that position. If there are more than half ones at a particular position, the bit at that position is set to one in the reference template otherwise it is reset to zero. We employed this coding scheme on our iris codes to extract a reference iris template and later tested it in the complete algorithm. For evaluating the system with majority coding approach, a reference iris template is created from first five iris codes of a person. This reference template is compared against five other iris codes from the same eye for genuine comparisons. For impostor comparisons, each reference template is compared with 150 iris codes of other users. There are 300 genuine comparisons and 9000 impostor comparisons. But we found that majority coding does not work at all for iris images. The system completely failed to regenerate the cryptographic key for all the irises. We obtained 100% FRR for all levels of error correction. The possible reason for this failure can be the iris rotation. For multiple iris images, we have not applied the rotation adjustment. The results are reported in Table D.1.

Table D.1: Results for majority coding approach,no rotation adjustment, no password adjustment, 4 zeros added in 12 bits $\approx$ 33% error correction. $n = 50$, $m = 6$, effective code length=1600

t_s	**Key Length**	**Biosecure V1**		**Casia V2**	
		FAR	**FRR**	**FAR**	**FRR**
1	288	0	100	0	100
2	276	0	100	0	100
3	264	0	100	0	100
4	252	0	100	0	100
5	240	0	100	0	100
6	228	0	100	0	100
7	216	0	100	0	100
8	204	0	100	0	100
9	192	0	100	0	100
10	180	0	100	0	100
11	168	0	100	0	100
12	156	0	100	0	100
13	144	0	100	0	100
14	132	0	100	0	100
15	120	0	100	0	100
16	108	0	100	0	100
17	96	0	100	0	100
18	84	0	100	0	100
19	72	0	100	0	100
20	60	0.01	100	0	100
21	48	0.01	100	0	100
22	36	0.09	100	0	100
23	24	0.15	100	0	100
24	12	0.38	100	0.12	100

Appendix E

List of Publications

E.1 Conference Publications

1. Sanjay Kanade, Danielle Camara, Emine Krichen, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi. Three Factor Scheme for Biometric-Based Cryptographic Key Regeneration Using Iris. In *The 6th Biometrics Symposium (BSYM)*, September 2008.

2. Sanjay Kanade, Danielle Camara, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi. Application of Biometrics to Obtain High Entropy Cryptographic Keys. In *Proceedings of World Academy on Science, Engineering, and Technology*, Hong Kong, March 2009.

3. Sanjay Kanade, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi. Cancelable Iris Biometrics and Using Error Correcting Codes to Reduce Variability in Biometric Data. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2009.

4. Sanjay Kanade, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi. Multi-Biometrics Based Cryptographic Key Regeneration Scheme. In *IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, September 2009.

5. Sanjay Kanade, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi. Obtaining Cryptographic Keys Using Feature Level Fusion of Iris and Face Biometrics for

Secure User Authentication. In *IEEE CVPR Workshop on Biometrics*, June 2010.

6. Mohamed Abid, Sanjay Kanade, Dijana Petrovska-Delacrétaz, Bernadette Dorizzi, and Hossam Afifi. Iris Based Authentication Mechanism for e-Passports. In *2nd International Workshop on Security and Communication Networks (IWSCN)*, 2010.

7. Sanjay Kanade, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi. Generating and Sharing Biometrics Based Session Keys for Secure Cryptographic Applications. *IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, September, 2010.

8. Dijana Petrovska-Delacrétaz, Sanjay Kanade, Rima Belguechi, Christophe Rosenberger, and Bernadette Dorizzi. Can an Algorithmic Solution be Proposed That Helps the CNIL to Guarantee the Privacy of our Biometric Data? In *Workshop Interdisciplinaire sur la Sécurité Globale (WISG)*, 2010

E.2 Patents

1. Procédé de Vérification de l'Identité d'un Individu Avec des Codes Correcteurs d'Erreurs, *Brevet PR88357 Français*, déposé en Juin 2009.

E.3 Presentations, Talks, & Others

1. Sanjay Kanade, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi, Secure Cryptographic Key Generation Using Iris-Biometrics, Poster presentation and demonstration at CeBIT, Hanover, March 2009.

2. Sanjay Kanade, Dijana Petrovska-Delacrétaz, Bernadette Dorizzi, Obtaining Cryptographic Keys by Feature Level Fusion of Iris and Face Biometrics, Presented at the *COST 2101 Workshop on Privacy and Security for Biometrics*, Las Palmas de Gran Canaria, Canary Islands, May 10–11, 2010.

3. Sanjay Kanade, Emine Krichen, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi. Three Factor Scheme for Biometric-based Cryptographic Key Regen-

eration Using Iris. Technical report, Institut TELECOM: TELECOM SudParis, 2008.