ON THE PERFORMANCE OF HELPER DATA TEMPLATE PROTECTION SCHEMES

Emile J. C. Kelkboom

The work described in this thesis has been carried out
at the Philips Research Laboratories Eindhoven and
at the University of Twente, the Netherlands.

CIP data Koninklijke Bibliotheek, The Hague, The Netherlands
Kelkboom, Emile
On the Performance of Helper Data Template Protection Schemes
Thesis University of Twente,
ISBN 978-90-365-3074-3
Keywords.: biometrics, template protection, privacy enhancing technologies (PET), helper data systems (HDS), bit extraction.

Cover: The color combinations, the stripes, and the stars are inspired by the Arubian flag. The design contains an artistic impression of the graphical illustration of the Helper Data System template protection scheme in Figure 5.2.
Design: Henny Herps

ON THE PERFORMANCE OF HELPER DATA TEMPLATE PROTECTION SCHEMES

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit van Twente,
op gezag van de rector magnificus,
prof.dr. H. Brinksma,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op vrijdag 1 oktober 2010 om 13.15 uur

door

Emile Josephus Carlos Kelkboom

geboren op 16 juni 1980
te Aruba

Dit proefschrift is goedgekeurd door de promotor:
prof.dr. W. Jonker

Assistent Promotor:
dr.ir. R.N.J. Veldhuis

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
prof.dr. W. Jonker	Universiteit van Twente, promotor
dr.ir. R.N.J. Veldhuis	Universiteit van Twente, assistent promotor
prof.dr.ir. C.H. Slump	Universiteit van Twente
prof.dr. P.H. Hartel	Universiteit van Twente
prof.dr.ir. A.J. Han Vinck	University of Duisburg-Essen, Germany
prof.dr. C. Busch	Hochschule Darmstadt, Germany
dr.ir. F.M.J. Willems	Technische Universiteit Eindhoven
dr.ir. J. Breebaart	Philips Research

Na mi mayor- y rumanan
To my parents and sisters

Samenvatting

Biometrie biedt de mogelijkheid om de identiteit van een persoon vast te stellen op basis van fysieke of gedragseigenschappen. Enkele voorbeelden van fysieke eigenschappen zijn vingerafdrukken, gezichtskenmerken of een irispatroon; voorbeelden van gedragseigenschappen zijn loopbeweging, een handtekening, of spraakkarakteristieken. Omdat deze biometrische eigenschappen toebehoren aan de persoon zelf, bestaat er een sterke link tussen de persoon en het identificatiemiddel. Door deze sterke band verhoogt biometrie de veiligheid van toegangs- en grenscontrole systemen of het authenticeren van een persoon op afstand in een netwerk. Daarnaast kan biometrie het authenticatieproces gebruiksvriendelijker maken, omdat bijvoorbeeld een paswoord onthouden of het dragen van een badge niet meer noodzakelijk is.

Het gebruik van biometrie blijft toenemen en wordt tegenwoordig al wereldwijd toegepast in het electronisch paspoort, het zogenaamde ePassport. Om een persoon te authenticeren dienen de biometrische informatie in de vorm van een referentietemplate opgeslagen te worden tijdens de registratiefase, zoals het in het electronisch paspoort gebeurd. Het grootschalig gebruik van biometrische systemen en het opslaan van referentietemplates brengt nieuwe privacy- en veiligheidsrisico's met zich mee. Voorbeelden van zulke risico's zijn (i) identiteitsfraude, (ii) traceerbaarheid, (iii) onvervangbaarheid van de referentietemplate, en (iv) het achterhalen van gevoelige medische informatie. Het reduceren van deze risico's is essentieel bij grootschalig gebruik van biometrie.

Deze risico's kunnen beperkt worden door templateprotectie technieken toe te passen. De vereiste protectie-eigenschappen zijn (i) onomkeerbaarheid, (ii) vernieuwbaarheid, (iii) en ontraceerbaarheid. Gedurende het laatste decennium werden diverse methoden gepubliceerd om biometrische gegevens te beschermen, waaronder het helper data systeem (HDS). Het fundamentele principe achter het HDS is het binden van een binaire vector met de biometrische gegevens, door gebruik te maken van helper data en cryptografische technieken. Die binding gebeurt zodanig dat de binaire vector reproduceerbaar is gegeven nieuwe biometrische gegevens van hetzelfde individu. Hierbij wordt de binaire vector gebruikt als een cryptografische sleutel. De identiteitscontrole wordt veilig uitgevoerd door middel van het vergelijken van de hash van de sleutel die zowel tijdens de registratie- als de authenticatiefase zijn afgeleid. De lengte van de sleutel bepaalt de mate van protectie.

Dit proefschrift beschrijft een uitgebreid onderzoek van het HDS, namelijk het (i) bepalen van het theoretische classificatievermogen, (ii) afleiden van de bovengrens

van de sleutellengte, (iii) analyseren van de onomkeerbaarheids- en ontraceerbaarheids-eigenschappen voor verschillende bitextractiemethoden, en (iv) het bepalen van de optimale fusie methode.

Het theoretische classificatievermogen wordt bepaald door aan te nemen dat de features, die zijn afgeleid uit de biometrische gegevens, een Gaussische verdeling hebben. De resultaten tonen aan dat een simpel model, waarbij uniformiteit van intra-klasse featurevariantie en onafhankelijke featurecomponenten verondersteld worden, niet toereikend is om het classificatievermogen goed te schatten. Complexere modellen worden geïntroduceerd, waarbij de variabiliteit van de featurevariantie en de afhankelijkheid tussen featurecomponenten in acht worden genomen. Op basis van het theoretisch model wordt de invloed van de bitextractiemethode op het classificatievermogen van het biometrisch systeem onderzocht. Het gebruik van een bitextractiemethode die een enkel bit per featurecomponent met behulp van een vaste kwantisatiedrempel extraheert, leidt tot een verlies in classificatievermogen.

Met behulp van het theoretisch model wordt de bovengrens afgeleid voor de lengte van de cryptografische sleutel, op basis van de aanname dat de foutcorrigerende code op de zogenaamde Shannonlimiet staat ingesteld. Het onderzoek toont de relatie aan tussen het classificatievermogen van het biometrisch system en de lengte van de sleutel.

Verschillende kwetsbaarheden die de onomkeerbaarheids- en ontraceerbaarheids-eigenschappen negatief beïnvloeden worden aangetoond met een bijbehorende oplossing. De eerste kwetsbaarheid betreft de bitextractiemethode DROBA, welke meerdere bits per component kan extraheren. Voor deze bitextractiemethode wordt aangetoond dat de onomkeerbaarheidseigenschap is aangetast. Een mogelijke oplossing hiervoor is het beperken van het bitextractiemethode zonder verlies in classificatievermogen. De tweede kwetsbaarheid betreft het gebruik van een lineaire foutcorrigerende code die negatieve consequenties heeft voor de ontraceerbaarheidseigenschap. Een mogelijke oplossing is het introduceren van een specifiek randomisatieproces op de binaire vector die is afgeleid van de biometrie. Als laatste wordt het verband geanalyseerd tussen het systeem classificatie- en traceerbaarheidsvermogen voor verschillende bitextractiemethoden. In dit onderzoek varieert de mate van gebruik van persoonsgebonden informatie dat wordt opgeslagen tijdens de registratie fase. Het traceerbaarheidsvermogen stijgt naarmate de persoonsgebonden informatie toeneemt. Verder tonen de resultaten aan dat in het geval dat het aantal registratiewaarnemingen toeneemt het traceerbaarheidsvermogen het classificatievermogen van het biometrisch system kan overtreffen.

De optimale fusiemethode, toepasbaar in het HDS, wordt bestudeerd voor meerdere waarnemingen van een biometrische karakteristiek of meerdere feature-extractiealgoritmen. Neemt men het gemiddelde van de features uit de gemaakte waarnemingen, dan leidt dit tot de meest compacte referentietemplate zonder verlies in classificatievermogen. Wanneer meerdere feature-extractiealgorithmen dienen te worden gecombineerd, blijkt fusie van scores tot het beste classificatievermogen leidt.

Summary

Biometrics enables the establishment of a person's identity by means of the person's physiological or behavioral traits. Examples of the physical traits include fingerprints, face, or iris and examples of behavioral properties include gait, signature, or voice. Biometrics creates a strong link between the person and its credentials because the properties belong to the person. Because of this strong link, biometrics can improve the security in access- or border-control systems, or in case of a remote personal authentication in a networked system. Furthermore, biometrics can make the personal authentication process more convenient by replacing the burden of remembering passwords or carrying a badge or token.

The use of biometrics looks promising as it is already being applied in electronic passports, ePassports, on a global scale. Because the biometric data has to be stored as a reference template on either a central or personal storage device, its wide-spread use introduces new security and privacy risks such as (i) identity fraud, (ii) cross-matching, (iii) irrevocability and (iv) leaking sensitive medical information. Mitigating these risks is essential to obtain the acceptance from the subjects of the biometric systems and therefore facilitating the successfully implementation on a large-scale basis.

A solution to mitigate these risks is to use template protection techniques, also known as privacy enhancing technologies (PET). The required protection properties are (i) irreversibility, (ii) renewability and (iii) unlinkability. In the last decade, different approaches have been introduced in the literature, including the one known as the helper data system (HDS). The fundamental principle of the HDS is to bind a binary vector with the biometric sample with use of helper data and cryptography, as such that the binary vector can be reproduced or released given another biometric sample. The binary vector is then used as a cryptographic key. The identity check is then performed in a secure way by comparing the hash of the key. Hence, the size of the key determines the amount of protection.

This thesis extensively investigates the HDS system, namely (i) the theoretical classification performance, (ii) the maximum key size, (iii) the irreversibility and unlinkability properties, and (iv) the optimal fusion method.

The theoretical classification performance of the biometric system is determined by assuming that the features extracted from the biometric sample are Gaussian distributed. The results show that a simple model, which assumes independent feature components and homogeneous within-class variance across all subjects, is not sufficient to estimate the classification performance of the biometric system. More complex models are intro-

duced incorporating the within-class variability and the dependencies between the features. With the simple model, the influence of the bit extraction scheme on the classification performance is investigated. Given a bit extraction scheme that extracts a single bit per feature based on a fixed quantization threshold, the results indicate that the classification performance before the bit extraction scheme is better than the performance after the bit extraction.

With use of the theoretical framework, the maximum size of the key is determined by assuming the error-correcting code to operate on Shannon's bound. The study indicates the relationship between the system classification performance and the maximum key size.

Multiple vulnerabilities are analyzed and a solution is proposed. The first vulnerability concerns the bit extraction scheme named DROBA, which can extract multiple bits per component, where the original algorithm has a negative impact on the irreversibility property. A solution is proposed to restrict the DROBA algorithm such that no loss of classification performance is observed. The second vulnerability concerns the use of linear error-correcting codes, which has a negative impact on the unlinkability property. A solution is the use of a specific randomization process on the extracted binary vector. Furthermore we analyze the relationship between the system and cross-matching classification performance for different bit extraction schemes varying in the degree of subject-specific information that is used. The results also show that when increasing the number of enrolment samples the cross-matching performance can outperform the system performance.

The optimal way of applying multi-sample and multi-algorithm fusion with the HDS is studied. Taking the average of features of the multiple enrollment samples has the advantage of a single protected template while having a similar classification performance. In case of multi-algorithm fusion, applying fusion at score-level leads to the best classification performance.

Compilacion

Biometria ta ofrece e posibilidad pa determina identidad di un persona, basa riba caracteristicanan fisico of di comportacion. Algun ehempel di caracteristicanan fysico ta imprenta di dede, caracteristica di cara of un patronchi di iris; caracteristicanan adecuado di comportacion ta e manera di cana, un firma of e manera di papia. Pa motibo cu tur esaki ta pertenece na e persona mes, ta surgi un relacion fuerte entre e persona y e manera di identificacion. Pa motibo di e laso fuerte, biometria ta mehora siguridad di systema di entrada y control na frontera of autenticidad di un persona riba distancia den un systema di red. Ademas biometria por haci e proceso di autenticidad di persona mas complaciente, pa motibo cu no ta necesario mas pa corda un codigo di entrada of cana cu badge.

E usamento di biometria ta munstra prometedor ya cu ta us'e caba den mundo electronico: por ehempel den e paspoort electronico, ePassport. Pa motibo cu mester warda datonan biometrico como base (template) di referencia, manera den ePasport, e usamento na scala grandi di biometria ta lanta risiconan nobo di privacidad y siguridad. Ehempelnan di risiconan asina ta (i) fraude di identidad, (ii) autenticidad, (iii) base di referencia irevocabel, y (iv) pone man riba informacion medico sensibel. Reduci e risiconan aki ta esencial pa e usamento na scala grandi di biometria.

E solucion pa limita e risiconan ta tuma luga cu implementacion di tecnicanan di proteccion di e base (template), conosi como e tecnologia di proteccion di privacidad, Privacy Enhancing Technologies (PET). E caracteristicanan exigi di proteccion ta (i) irevocabel, (ii) renobabel, y (iii) bo no por localisa nan. Durante e ultimo decada, nan a publica diferente metodo pa proteha e datonan biometrico, entre nan e "Helper Data System" (HDS). E principio fundamental tras di e HDS ta pa acopla un vector cu datonan biometrico cu ayudo di "helper data" y cryptografia, di tal forma cu por reproduci of publica e vector binaire cu datonan biometrico nobo di e mesun individuo. Por usa anto e vector binair como un yabi cryptografico. E control di indentificacion ta ehecuta na manera sigur cu comparacion di e mexcla (hash) di e yabi. Largura di e yabi ta determina grandura di proteccion.

E tesis doctoral aki ta describi un investigacion amplio di e HDS, sea (i) determinacion di e poder teoretico di clasificacion, (ii). determina e nivel maximo di e largura di e yabi, (iii) analisis di e caracteristica irevocabel y imposibel pa localisa cu diferente metodo "bit extractie" y (iv) e metodo obtimal di fusion.

E poder teoretico di clasificacion ta determina door di asumi, cu e caracteristicanan, saca for di e datonan biometrico ta distribui segun e systema Gaussis. E resultadonan

ta munstra cu un modelo simpel, den cual ta supone uniformidad di variante intra-clase di e caracteristicanan y cu e componentenan di e caracteristicanan ta independiente, no ta suficiente pa calcula exactamente e forsa di clasificacion. Nos a introduci modelonan mas compleho cu ta carga cu nan e variabilidad y dependencia entre e componentenan di e caracteristicanan. Basa riba e cuadra teoretico, nan ta investiga e influencia di e metodo di extracto di cada bit riba e poder di clasificacion di e systema biometrico. E usamento di un metodo di extracto di cada bit, cu ta localisa un solo bit pa componente di e caracteristicanan cu ayudo di un barera di cuantisacion fiho, ta mustra cu e comportacion di clasificacion prome cu e proceso di extracto di bit ta miho compara cu despues di e extracto di bit. Cu ayudo di e modelo teoretico, ta determina e nivel maximo pa largura di e yabi, basa riba acceptacion cu e codigo di coreccion di fayo ta traha segun e systema di Shannon. E investigacion ta munstra e relacion entre e poder di clasificacion di e systema y largura di e yabi.

Nos ta analisa diferente asunto vulnerabel y ta propone e solucion corespondiente. E prome asunto vulnerabel ta trata e metodo di extracto di bit DROBA, cu por aisla mas cu un bit pa componente, cu ta munstra cu e caracter irevocabel ta atacha. Un posibel solucion pa esaki ta limitacion di un algoritmo di extracto di bit sin ta perde e poder di clasificacion. E di dos caso vulnerabel ta trata e usamento di un codigo linear di coreccion di fayo, cu tin consecuentia negativo pa e caracter di no por localis'e. Un posibel solucion ta introduccion di un proceso di arbitrahe specifico riba e vector binair extradita. Como ultimo nos ta analisa e relacion entre e systema di poder di clasificacion y localisa pa diferente metodo di extracto di bit, cu ta varia den e grandura di usamento di informacion cu ta mara na persona. Mas cu nan ta usa e informacion mara na persona, mas miho e poder di localisa ta bira. Ademas e resultadonan ta munstra, cu den caso cu e cantidad di observacionnan di registracion ta aumenta, e poder di localisacion por ta hasta mas miho cu e poder di clasificacion di e systema biometrico.

Nos a studia e manera optimal di fusion cu e HDS pa mas observacion di un caracteristica biometrico of mas cu un extracto algoritmico di e caracteristicanan. E promedio di e caracteristicanan for di diferente observacion ta hiba pa e base di referencia mas compacto, sin perdemento di e poder di clasificacion. Na momento cu mester combina mas extracto algoritmico di e caracteristicanan, ta resulta cu e fusion riba e nivel di resultado ta genera e miho forsa di clasificacion.

Translated by Emile Kelkboom Sr.

Acknowledgements

The last four years flew by, which means that I truly had a great time with my Ph.D. project. Although there is a single author written on the cover of this thesis, this work could never have existed without the contribution and support of my promotor prof.dr. Willem Jonker, assistant promotor dr.ir. Raymond Veldhuis, my daily supervisors dr.ir Tom Kevenaar and dr.ir Jeroen Breebaart, my colleagues at Philips Research, the University of Twente and the European project 3DFace, and of course my family and friends.

Willem opened my eyes on conveying the essential information that managers are looking for. He taught me his helicopter-view approach as I tended to get lost into the fine details of problems. With Raymond I enjoyed the many talks and discussions and the various trips we made together. He still amazes me with his extensive knowledge and creativity. I had the opportunity to have had two daily supervisors. Tom was my first supervisor for roughly the first two years, before he joined the successful spin-off of priv-ID. Many thanks for his help and extensive knowledge of the field, and I also enjoyed listening to him playing guitar. I wish him and the priv-ID team, including Michiel van der Veen, lots of business opportunities. My supervisor for the final two years of my Ph.D. was Jeroen. Although new to the field, he amazed me with the speed he became an expert. Many thanks for the countless suggestions and corrections for improving my thesis. I will also miss the music produced by his keyboard while typing.

I would like to thank all my colleagues within the Information and System Security (ISS) department, led by Bart. I really appreciate the four years of commitment and freedom that I received from Bart. A special thanks would go to Ileana, Asim, Koen, Sabri, Jeroen, Fons, Ton, Milan, Jorge, Sandeep, and Sye Loong for the great discussions either during the coffee breaks, lunch, wok, and walks. Odette, our secretary, was always ready to help me and she is also a great group event organizer. In short, I will miss the ISS group. I would like to thank all the participants within the European project 3DFace, especially Christoph and Xuebing, for the wonderful three years of collaboration and successfully integrating the template protection technology into the prototype. Since August, I joined the Brain, Body, & Behavior (BBB) department and I would like to thank Ans for giving me this opportunity. Also, I would like to thank Gary, who helped me a lot with my first journal publication when he was a colleague within the ISS department. I am looking forward to continue our collaboration within the BBB department. Ludo and Stan demystified the field of error-correcting codes for me.

Furthermore, I would also like to thank the Philips PhD and PostDoc Community (PPC) committee members, Marjolein, Janneke, Greg, Nele, Maarten, Jos, Tommi, Aaron,

Jurgen, and Alberto and the active members. In the last 18 months we successfully initiated this community and organized many social events. I wish them all the best with keeping up the good work and to have an exciting first Symposium in November.

From the University of Twente, I would like to thank Berk & Pinar, Luuk, Chun, Haiyun, Sanja, and Anne for showing me a glimpse of the life of a Ph.D. student at the university, which is different than the one within a company.

From the Technical University of Eindhoven, I would like to thank Frans, Berry, Boris, and Tanya for the many fruitful discussions.

On a more sportive note, the last four years I picked up playing basketball again thanks to the international/multi-cultural campus basketball club whose leader is Bob. The players that join on a consistent basis can be divided into four groups, namely the Serbian Gangsters, Bob, professor Milan, Alex, Milos, Zoran, and Vojkan, the Italian Mafiosi, Danilo, Alessio, Pietro, and Giovanni, the Greek Mob, Evangelos, Pavlos, Nektarios, Emmanuel, and the United World Domination Force, Konrad, Qing, Geert-Jan, Marek, Andrei, Carlos, Anne, Vadim, Xiaojun, Ignacio, Roger, and Jan. I hope we will have many more nights of great fun of playing basketball and arguing about fouls and rules. The last two years were even more sportive since I also joined the Almonte basketball club. With help of the trust and confidence from the coaches, Koen, Daan and Stephan, and teammates, Ronald, Carlos, Thijs, Kai, Niels, Sven, Mikke, Niel, Siem, Klaas, and Raul, I improved as such that my nickname has changed from "De Lompe" to a more graceful one, namely "Èhmile". Let's go for the championship and play "eerste divisie" nationally next season!!

Besides my friends from basketball and work, I would also like to show my token of appreciation to my other friends Bel & Raffy, Angela, Andres, Charisa & Marlon, Kristel & Sergio, Ivan, Theo, Chee, Alberto, Bala, Robin, Quintin and Nestor for the many drinks, talks, birthday parties, movie nights, and BBQ's we celebrated together. I shouldn't forget Nancy from whom, in recent months, I have learned a lot about life. A special thanks goes to Wendy for supporting me in the last seven years and most of my Ph.D. work. I am fully confident that you will become one of the best dermatologists.

Standing as strong as a pyramid in my life, that would be my family. The most important ones are my parents. Because of the great dedication, nurture and guidance from my mother Marij and father Emile Sr, I am standing here. My two beloved sisters, Esther and Sandra, and their family, from whom I know that they will always be there for me. I am also grateful to all my aunts, uncles, cousins, nieces, and nephews.

For the ones who I may have missed, my apologies, but your help and support or simply your presence is being appreciated.

I would like to end with the following: *Take some time away from your busy life or take a look outside your self-created invisible wall and contemplate on the following questions; "Who am I?", "What do I want?", and "Am I happy with my life as it is?", because each day that passes in which you haven't smiled is a day you haven't lived to its fullest. Be Happy and Smile! Sea Contento y Cana cu Sonrisa! :-)*

Emile Kelkboom Jr. 26 August 2010
Eindhoven, the Netherlands

Contents

Chapter 1

Introduction

The size and complexity of our society and world calls for the ability to accurately and automatically identify people, also referred to as *personal identification* [1]. Personal identification can be established by means of verification or recognition [1]. In a verification setup, also referred to as *(user-)authentication* [2–4], the system tries to verify whether the identity claim provided by a subject is correct, namely "Am I who I claim I am?". In a recognition setup, your identity is automatically established from a set of known identities, e.g. a database of identities. In the literature, recognition is also referred to as identification. This thesis is mainly focused on the verification/authentication setup as validating a claimed identity is the most common method for identity management in commercial and governmental applications.

The most common approaches for user-authentication being used today are based on (i) personal possessions (what you have), (ii) knowledge (what you know), and (iii) biometrics (who you are), from which the latter is becoming more popular. Examples of personal possessions include passports, national identity cards, driver's license, bank cards, company badges, and the old fashioned tangible keys. Examples of knowledge-based authentication include the use of passwords, personal identification numbers (PIN), and answers to a set of questions to which the answers have been recorded in an earlier phase. Biometrics is the field of uniquely and automatically recognizing humans based upon one or more intrinsic physiological or behavioral traits. Examples of physiological traits include fingerprint, face, iris, retina, hand geometry, and palm, while examples of behavioral traits include voice, signature, keystroke dynamics, and gait. Hence, biometrics creates a strong connection between an individual's identity and body. There are also systems that combine two or more factors of authentication, referred to as *multi-factor authentication*, such as payment systems where both the ATM card and its corresponding PIN have to be provided, or the passport that includes a face image and other personal information.

The drawback of possession-based authentication is that the corresponding object has to be presented, while it can be forgotten, lost or stolen. Similarly, passwords used in knowledge-based authentication are often forgotten. The studies [5, 6] analyzed the num-

ber of passwords people have to remember. Both studies report that roughly 20% of the participants have to remember 15 or more passwords for their job, while 35% and 57%, respectively, have between six and 15 passwords to remember. A more recent study [7] reports that 66% of the participants have 11 or more password-protected accounts, where 47% use different passwords for each or almost all accounts. Besides these convenience drawbacks, possession- and knowledge-based authentication are also sensitive to the *repudiation attack*. For example, a person could legitimately gain access to a building by using his own badge and still claim it wasn't him because he assumably lost his badge. Similarly, this attack also exists when using passwords.

These drawbacks can be overcome by using biometrics. It is very difficult to "forget" or "lose" your biometric trait. Therefore, biometrics can make the authentication procedure more convenient by replacing the burden of remembering long passwords or carrying a badge. The incorporation of physiological and/or behavioral traits as evidence for authentication also helps to prevent a repudiation attack.

Because of its advantages, the interest in biometric systems has significantly increased in recent years. Examples are the planned introduction of the United Kingdom National Identity Card based on biometrics required by the Identity Cards Act 2006 [8], the recommendation by the International Civil Aviation Organization (ICAO) [9] to adopt the ePassport that also includes biometric data, the implementation of the iris-based Privium border control system in Schiphol Airport in the Netherlands [10], and the many implementations in the financial sector such as in ATMs in Japan [11, 12] and payment systems in Singapore [13], US [13], and Mexico [14].

The use of biometrics looks promising. Its wide-spread use, however, introduces new security and privacy risks as will be discussed in Section 1.2. Mitigating these risks is essential for obtaining the acceptance from the subjects of the biometric systems and therefore facilitating the successful implementation on a large-scale base. *Methods to address and mitigate these risks are the main topics of this thesis.*

In the remainder of this chapter we first describe a general biometric verification system and its performance measures in more detail in Section 1.1, and follow with the security and privacy risks in Section 1.2. Furthermore, in Section 1.3 we discuss the guidelines and countermeasures to mitigate these risks. This thesis focuses on the countermeasure known as *template protection*. We introduce the template protection scheme of interest that is used throughout this thesis, namely the *Helper Data System* (HDS). We present the research questions and discuss the corresponding contributions within this thesis in Section 1.4. We conclude the chapter with the outline of this thesis in Section 1.5.

1.1 Biometric Verification Systems

As mentioned previously, biometrics is the field of uniquely and automatically recognizing or verifying humans based upon one or more intrinsic physiological or behavioral traits. Desired properties of the biometric traits are [1, 15]:

* **Universality**, which implies that the trait should be existing for each subject,

* **Uniqueness**, which means that the trait should be different for each subject within the population,
* **Permanence**, which indicates that the trait remains constant with time,
* **Collectability**, which means that the trait can be measured quantitatively.
* **Performance**, which implies that a certain verification accuracy can be achieved with specific resource requirements, and within working and environmental factors.
* **Acceptability**, which suggests the willingness for people to accept the biometric system. Note that any privacy or security risk of the biometric system left untreated can affect its acceptability.
* **Circumvention**, which indicates the difficulty to spoof the system. Spoofing is the act of fooling the system and obtaining unauthorized access by means of fraudulent techniques. Researchers have shown successful attacks on fingerprint recognition systems by using fake fingerprints, for example by creating "gummy" fingerprints or a wafer thin silicon dummy that can be glued on the finger [16, 17]. Results at that time showed that these methods worked effectively on multiple fingerprint sensors both for the scenarios where the fake fingerprint is created (i) with full co-operation from the subject being impersonated, and (ii) from a latent fingerprint without cooperation.

A biometric verification system consists of an enrolment and verification phase as portrayed in Figure 1.1. In the enrolment phase, the individual is presented to the biometric system for the first time. One or more biometric samples are captured by a sensor. In Figure 1.1 we show an example of a camera that captures depth information of the individual's face, namely a 3D face image, as the biometric sample. Usually, the capturing process is followed by the *Feature Extraction* module, where either a real-valued *feature vector* (e.g. Gabor filter responses), a *binary vector* (e.g. iris code), or an unordered *set of values* (e.g. minutiae set) is extracted from the biometric sample and stored as the *reference template* on a storage device. Examples of storage devices include tokens, smart cards, and a central database. In the verification phase, a probe biometric sample is captured from the same biometric trait. The biometric sample is passed through the same feature extraction process and compared with the stored reference template corresponding to the individual's claimed identity. The *Comparator* module returns a match if the features extracted in the verification phase are similar to the reference template. In some cases, the biometric sensor data are stored as the reference template, for example in the form of a JPEG image. In that case, the comparison process incorporates the feature extraction process for both the reference as well as the probe sample.

There are two types of comparisons, namely a comparison between biometric samples of the same individual, which is referred to as a genuine comparison, and a comparison between biometric samples of different individuals, which is referred to as an imposter comparison. In general, the comparison process entails first the computation of a *score* followed by a decision based on the score. There are two types of scores, namely a *similarity score* and *dissimilarity score*, which tells you how similar and different the two biometric samples are, respectively. The decision is made by means of a threshold T. In case of a similarity (dissimilarity) score, a match is returned when the score is larger (smaller) than the threshold T. A match implies that the biometric samples from

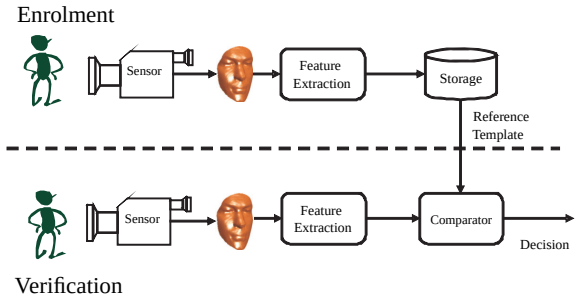
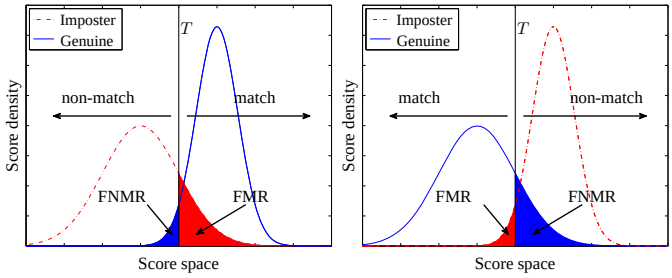


Figure 1.1: General biometric verification system where as an example a video camera captures a 3D face image as the biometric sample.



(a) Similarity score (b) Dissimilarity score

Figure 1.2: Illustration for the case of (a) similarity and (b) dissimilarity score with the corresponding match and non-match region, and FMR and FNMR.

the enrolment and verification phase are believed to have been acquired from the same individual, hence the claimed identity is considered to be genuine. On the other hand, a non-match results in a reject of the claimed identity. An illustration of the similarity and dissimilarity score with its corresponding match and non-match region is portrayed in Figure 1.2(a) and Figure 1.2(b), respectively. The dashed-line density corresponds to the scores obtained from imposter comparisons, while the solid-line density corresponds to the scores obtained from genuine comparisons.

As classification performance indicators we use the *false match rate* (FMR, α) and the *false non-match rate* (FNMR, β). The FMR is the rate of obtaining a match at imposter comparisons, while the FNMR is the rate of obtaining a non-match at genuine comparisons.[1] In Figure 1.2, the FMR and FNMR are indicated by the red and blue

[1]The FMR and FNMR are performance measurements of the recognition algorithm specifically and are related to the false-acceptance rate (FAR) and false-rejection rate (FRR) at system level by combining the FNMR and FMR with the failure to enrol (FTE) and failure to acquire (FTA) rates. The FTE is the rate of not being able to create a reference template of sufficient quality in the enrolment phase, while the FTA is the rate of not acquiring a biometric sample and feature vector of sufficient quality in the verification phase.

shaded areas, respectively. Note that both the FMR and FNMR depend on the threshold T. Therefore, the threshold is also referred to as the operating point of the biometric system. The relationship between the FMR and FNMR at different operating points can be illustrated by means of a detection error tradeoff (DET) or a receiver operating characteristics (ROC) curve. Note that when changing the operating point, either the FMR or FNMR decreases while the other increases, thus both the FMR and FNMR cannot be decreased or increased simultaneously. Single number performance indicators that are often used are the equal-error rate (EER), which is achieved at the operating point T_{EER} where both FNMR(T_{EER}) and FMR(T_{EER}) are equal, the FNMR achieved at a target FMR or the FMR achieved at a target FNMR.

1.1.1 Fusion

As stated in [18], the basic principle of fusion is the reconciliation of evidence presented by multiple sources of biometric information in order to enhance the classification performance. Multiple sources of biometric information can be extracted from the same biometric modality by (see Figure 1.3 for the case of fingerprints): (i) capturing a sample of multiple instances (left and right index fingerprint or iris) with the same sensor, (ii) using different sensors to acquire a different type of biometric samples from the same instance, (iii) capturing multiple samples using the same sensor and instance, and (iv) extracting multiple feature representations of the same biometric sample using different algorithms. These cases are referred to as the multi-instance, multi-sensor, multi-sample [2], and multi-algorithm systems, respectively. Further more, the fifth type is the multi-modal system, which is the fusion of sources of biometric information from multiple modalities, for example fingerprint, face, iris, voice, palm or retina. To complete the summary from [18], the sixth type is referred to as the hybrid system, which consists of a combination of the aforementioned fusion types. Each multi-biometric fusion type can be implemented at feature-level, score-level, or decision-level of the biometric system.

1.2 Security and Privacy Risks

The storage and processing of biometric data, and the widespread use of biometric systems introduce various security and privacy risks. We would define a security risk as a vulnerability of the system that facilitates an adversary to attack the system or increases the adversary's success rate of attacking the system. Privacy risks are related to vulnerabilities in which the adversary extracts valuable information about the individuals that use the biometric system and may not directly be related to increasing an adversary's attacking success rate. Mitigating these risks is essential to obtain the acceptance from the subjects of the biometric systems and therefore facilitating the successfully implemented on a large-scale base. The security and privacy risks are:

i **Identity fraud**, where for example an adversary steals the stored reference template and impersonates the genuine subject of the system by some spoofing mechanism.

[2] Within ISO [19], multi-sample fusion is referred to as multi-presentation.

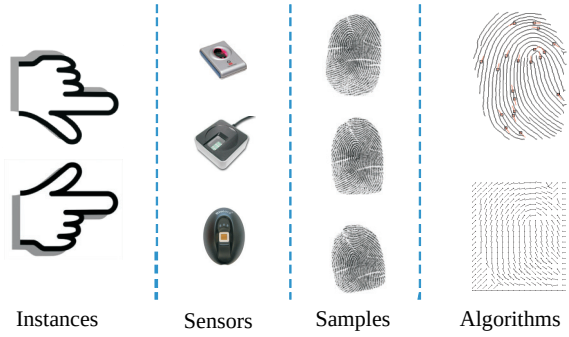


Figure 1.3: Multiple sources of biometric information using fingerprints as the single modality.

ii **Limited-renewability**, implying the limited capability to renew a compromised reference template due to the limited number of biometric instances, for example we only have ten fingers, two irises or retinas, and a single face.

iii **Cross-matching**, linking reference templates of the same individual across databases of different applications. With cross-matching it is possible to track the presence of an individual across multiple applications based on biometrics.

iv **Leaking sensitive personal information**, where it is known that biometric data may reveal the gender, ethnicity, or medical information such as the presence of certain diseases [20–22].

For fingerprints, real-life examples exist of spoofing a biometric system based on unauthorized use of fingerprints [16, 17], thus allowing identity fraud. It was thought that storing the set of minutiae points extracted from the fingerprint image instead would solve this problem, because the transformation was considered to be one-way. However, it has been shown in [23, 24] that from the set of minutiae points an artificial fingerprint can be created to spoof a minutiae-based fingerprint recognition system. Retrieving information about the original biometric sample may therefore lead to the leakage of sensitive personal information as indicated by the fourth risk, which is thus of a privacy nature.

The limited number of biometric instances makes it impossible to 'endlessly' renew a compromised reference template. If one revokes a compromised template, the corresponding biometric instance of the individual cannot be used within the biometric system anymore. Hence, this creates a security risk, because a compromised template cannot be revoked without disturbing the operational use of the system. This is a significant drawback compared to possession- or password-based authentication, where for example a new credit card with a new serial number can be issued or a new password can be created once they are compromised.

The limited number of biometric instances combined with the desired property of permanency leads to the cross-matching risk, which is a privacy risk. Using the same biometric instance of the same trait in multiple applications allows for verification whether

an individual is enrolled in different application assuming the application databases to be accessible. Again this is a drawback compared to possession- or password-based authentication, where for example different cards/tokens or usernames/passwords can be used for each application, however with some convenience drawback of needing to carry or remembering multiple cards/tokens or usernames/passwords, respectively. The cross-matching possibility consequently introduces the undesired threat of *function creep*. An example of function creep is the case that a database of biometric data is collected for a specific purpose, for example independent performance testing of biometric recognition systems, but is also used for another purpose without the consent of the participants, for example cross-matching the collected database with the criminal justice database containing biometric data related to unsolved crimes.

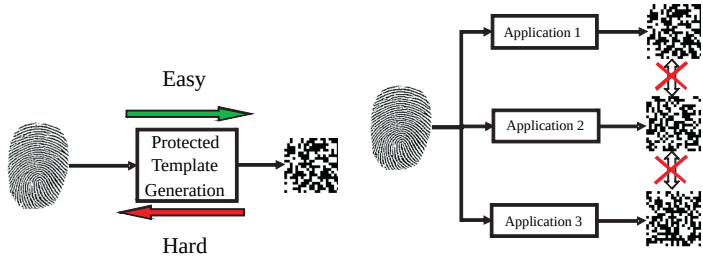
1.3 Protecting the Reference Template

Mitigating the privacy and security risks discussed in Section 1.2 is essential for biometric systems, in order to be accepted by the subjects and, therefore, a prime condition to successful large scale deployment.

According to several laws and directives, biometric data is considered to be personally identifiable information (PII) and requires proper protection in terms of procedures for handling the data and methods to prevent unauthorized use. ISO guidelines [25] for the proper protection of biometric data include the following requirements for stored biometric data:

i **Data minimization**, referring to only collecting the necessary data for the biometric verification as the reference template.
ii **Confidentiality**, ensuring that the reference template is accessible only to those authorized to have access.
iii **Integrity**, meaning that the reference template cannot be modified without authorization.
iv **Irreversibility**, implying that it is impossible or at least very difficult to retrieve the original biometric sample from the reference template.
v **Renewability**, where it is possible to create different reference templates when one gets compromised.
vi **Unlinkability**, which guarantees that different and unlinkable reference templates can be created for different applications in order to prevent cross-matching.

Reducing the stored reference template to information that is strictly required for verification, for example by storing extracted features rather than the biometric sample, reduces the risk of unauthorized use. The confidentiality guideline ensures that non-authorized persons do not gain access to the reference template, thus limiting the privacy risks of leaking personal information. Ensuring the integrity guarantees that an adversary is not able to modify the reference template in order to improve its success rate of attacking the biometric system. An illustration of the irreversibility property is shown in Figure 1.4(a).



(a) Irreversible (b) Renewability/unlinkable

Figure 1.4: (a) Irreversible and (b) renewability/unlinkable property.

Given a biometric sample it is easy to create the protected template, but given the protected template it is impossible or at least difficult to retrieve the biometric sample. People may claim that biometrics are not secret [26], as your face or fingerprint can easily be captured covertly, and therefore protecting them with an irreversibility property may not make sense. However, new biometric traits such as hand vein or palm vein are much more difficult to obtain covertly and their classification performance look very promising. Therefore, it is essential to protect the reference templates derived from these traits. The renewability and unlinkability properties are illustrated in Figure 1.4(b). The subtle difference between the renewability and unlinkability property is that for the renewability property different reference templates need to be derived from the same biometric sample, while the unlinkability property requires that these different templates cannot be linked back to the same data subject. Fulfilling the unlinkability property inherently fulfills the renewability property.

Some known countermeasures to safeguard the privacy and security by enforcing some of the ISO guidelines are

i The practice of **data separation** where the most privacy sensitive information is stored on an individual smart card or token. This reduces the risk of security breaches of centralized databases, and provides more control to the subject of the biometric data and the processing thereof.

ii The use of **data minimization** principles, such as feature extraction techniques. For example, store only the extracted minutiae set instead of the complete fingerprint image.

iii The use of **classical encryption** techniques such as DES, AES, or RSA, to provide confidentiality or integrity during storage and transmission of biometric data.

iv The implementation of **template protection techniques**, to provide irreversibility, renewability, and unlinkability.

Separating the privacy sensitive data across different storage devices increases the effort for the adversary to collect all data. Furthermore, by storing the privacy sensitive data on a storage device under the supervision of the subjects of the biometric system themselves, the subjects have more control of the use and processing of their biometric data

and protecting their privacy therefore also includes their own responsibility.

Instead of separating the data, the risks could be mitigated by storing only the data required for verification. For example, the use of feature extraction algorithms that extract only the essential information for verification from the captured biometric sample, namely storing the minutiae set instead of the fingerprint image.

With classical encryption schemes, the reference template would be encrypted before being stored in the database and in the verification phase it would be decrypted prior to the comparison process. Hence, the database consists of encrypted reference templates and is protected as long as the encryption key is kept secret. Confidentiality is achieved because only the key holder has access to the content of the reference template. By using digital signature schemes, the integrity of the reference template can be guaranteed. By using a different key for each application the protected templates are renewable and unlinkable when the keys are not compromised and therefore neutralizing the cross-matching risk. However, the drawback of this encryption method is that the encrypted reference templates have to be decrypted and are in the clear in the verification phase prior to the comparison. Furthermore, if the encryption key gets compromised the whole database could be decrypted, therefore the key has to be kept secret and requires a secure key infrastructure. Alternatively, comparison is performed on the encrypted domain [27–29]. However, these techniques are currently not sufficiently mature for wide-spread use in applications.

Template protection techniques inherently protect the reference template without the use of a single encryption key or having the reference template decrypted and in the clear. Template protection techniques mainly focus on implementing the irreversibility, renewability and unlinkability properties[3]. In the context of this thesis, a biometric reference template that has the aforementioned properties is referred to as "protected template". Note that these properties have to be met while maintaining a similar classification performance as for the case of the unprotected reference templates. The field of template protection is relatively young, however there is a significant interest to successfully develop and implement these techniques as shown by their prominent position within the European projects 3DFace [30] and TURBINE (TrUsted Revocable Biometric IdeNtitiEs) [31] from the 6th and 7th Framework Programme, respectively, the great interest from privacy offices such as the Office of the Information and Privacy Commissioner of Ontario [32], and the current ISO standardization activities [25]. This thesis focuses only on the template protection countermeasure.

1.3.1 Helper Data System (HDS)

In this section we briefly present the template protection scheme being used in the remainder of this thesis, which is known as the *Helper Data System* (HDS). A more detailed description of the HDS is provided in Section 2.3.1. An abstract overview of the HDS scheme as used in [33–35] is portrayed in Figure 1.5 and consists of two main parts: (i) *Bit Extraction* and (ii) *Bit Protection* part.

[3]The integrity and confidentiality property can easily be achieved by combining template protection techniques with cryptographic techniques, and are therefore considered not to be part of template protection and out of the scope of this thesis.

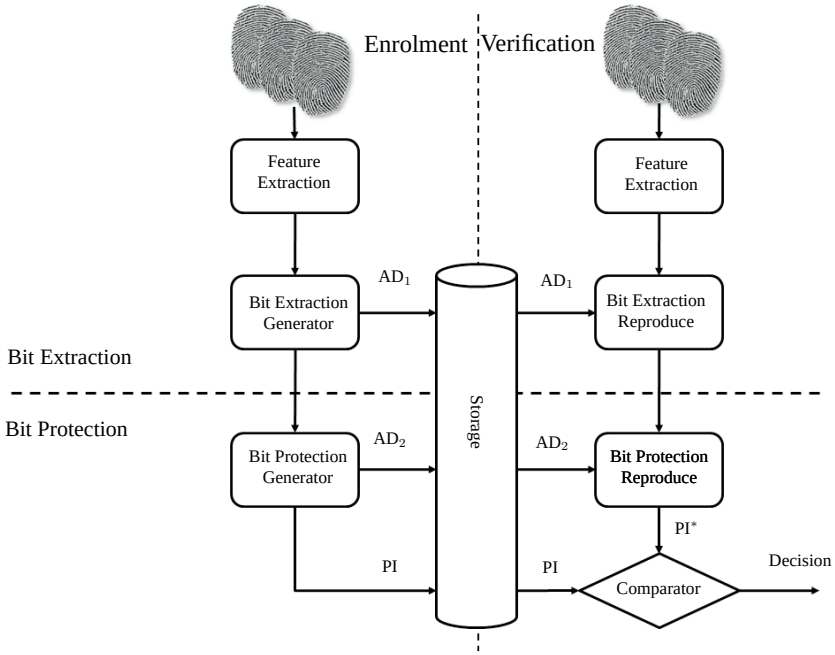


Figure 1.5: Template protection scheme including a *Bit Extraction* module.

In the enrolment phase, first a real-valued feature vector is extracted from each acquired enrolment sample by the *Feature Extraction* module. Hereafter, a single binary vector is created from the multiple feature vectors, within the *Bit Extraction Generator* module. The bit extraction scheme could be subject-specific in order to extract more robust bits, therefore some auxiliary data AD_1 containing the subject-specific information has to be stored as part of the protected template for use in the verification phase. The final step in the enrolment phase is the protection of the binary vector by the *Bit Protection Generator* module. The HDS is based on the key binding principle known as the fuzzy commitment scheme (FCS) from Juels and Wattenberg (1998) [36]. It randomly generates a key and binds it to the binary vector. The binding output is referred to as the *code-offset auxiliary data* AD_2. Furthermore, a *pseudonymous identifier* (PI) is derived from the random key using cryptographic primitives and is considered as part of the protected template. Concluding, the protected template is the triplet $\{\mathrm{AD}_1, \mathrm{AD}_2, \mathrm{PI}\}$.

In the verification phase, the *Feature extraction* module extracts a real-valued feature vector from each of the multiple acquired verification samples. Hereafter, the *Bit Extraction Reproduce* module derives a single probe binary vector from the multiple feature vectors with help of the stored auxiliary data AD_1 from the enrolment phase. The *Bit Protection Reproduce* module extracts a candidate pseudonymous identifier PI^* from the probe binary vector and the code-offset auxiliary data AD_2. The *Comparator* module compares both PI and PI^* and returns a decision. A match is returned if PI and PI^* are

equal, which occurs only if the probe binary vector is similar to the enrolment binary vector, otherwise a non-match is returned. In order to be robust against bit differences between the enrolment and probe binary vector, error-correcting codes (ECC) are being used.

1.3.2 Irreversibility, Renewability and Unlinkability Properties

In order to achieve the irreversibility property, given the protected template $\{\mathrm{AD}_1, \mathrm{AD}_2, \mathrm{PI}\}$ it should be difficult to retrieve information about the enrolment biometric data, its extracted real-valued feature vector, or the extracted binary feature vector. Therefore, (i) the bit extraction auxiliary data AD_1 should ideally not leak information about either the input real-valued feature vector or the biometric data, (ii) the code-offset auxiliary data AD_2 should preferably not leak information about the extracted binary vector or the random key, and (iii) the pseudonymous identifier PI should ideally not leak information about the randomly generated key, where the key size determines the difficulty of reversing PI. As the type of leakage we consider whether the leaked information is about the biometric samples, its extracted real-valued feature vector, or the extracted binary feature vector. We express the amount of the information leakage by the degree the adversary is able to increase the FMR at an impersonation attack. A greater increase of the FMR would imply a greater information leakage.

The renewability property is based on the possibility of creating many different protected templates given a biometric sample. The number of different keys that can be used in the binding procedures determines the renewability property, hence the key size plays another essential role.

The unlinkability property is stricter than the renewability property as it also requires that the protected template of the first $\{\mathrm{AD}_{1,1}, \mathrm{AD}_{2,1}, \mathrm{PI}_1\}$ and second $\{\mathrm{AD}_{1,2}, \mathrm{AD}_{2,2}, \mathrm{PI}_2\}$ application should not be linkable. The protected template may leak information that could be used for cross-matching. Hence, we express the amount of the information leakage by the cross-matching performance between two protected templates, which should be kept at a minimum in order to optimize the unlinkability property.

Concluding, there are two important attributes to study, namely (i) the key size, and (ii) the type and amount of information leakage from the protected template affecting the irreversibility or unlinkability property.

1.4 Research Questions and Contributions

As the title of the thesis suggests, the main research question is

What is the performance of the helper data template protection scheme (HDS)?

The term “performance” in this context is broad and includes the classification performance and the effectiveness of the privacy and security protection of the HDS. The main research question can be subdivided into four smaller and more specific questions.

Given the helper data template protection scheme:

1 What is the **theoretical classification performance**?
   - i How can we model the classification performance?
   - ii How do the system parameters influence it?
   - iii How does it compare with the classification performance without template protection?
2 What is the **maximum key size** at a given target classification performance and system parameters?
3 How does the **information leakage from the auxiliary data** affect the irreversibility and unlinkability property?
4 How can one realize **fusion with protected templates** and to what extent can it improve the classification performance?

In the following sections we discuss the related work and our contributions for each research question separately.

1.4.1 Theoretical Classification Performance

The irreversibility, renewability, and unlinkability properties of the template protection technique, as discussed in Section 1.3, have to be achieved while maintaining a similar classification performance as in the case of the unprotected reference templates. As will be explained in Chapter 3, it can be shown that the classification performance of the HDS is, given some limitations, identical to a Hamming-distance classifier operating on the binary feature vector. Furthermore, the classification performance for the unprotected case is assumed to be of the real-valued feature vector prior to the bit extraction. Hence, it is of importance to investigate the classification performance of the binary vectors, i.e. the classification performance on the binary level, and compare it with the optimal classification performance of the real-valued feature vectors, i.e. the classification performance on the continuous level.

To enable the analysis, we model the extracted real-valued features as a source with within-class and between-class Gaussian probability densities. The within-class density models the biometric variability and measurement noise, while the between-class models the diversity of a feature across the whole population. The bit extraction scheme we consider extracts a single bit per component using the mean of the between-class density as the binarization threshold. We also include the case where multiple enrolment and verification samples are taken and we analyze their effect on the classification performance.

With the Gaussian source model and bit extraction scheme we analytically estimate the theoretical classification performance of the template protection system in Chapter 3. As the naive model, we assume the within-class variance of a component to be homogeneous across all subjects, i.e. equal for each subject of the population, and each feature component to be independent. We validate the naive Gaussian analytical framework using biometric data. The naive model does not fully describe the performance curve and thus we adapt the model in order to incorporate the properties of non-homogeneous within-class variances and dependent feature components.

We conclude Chapter 3 by comparing the theoretical classification performance on binary level with the classification performance on continuous level, i.e. a binary classifier versus continuous classifier performance comparison. As the continuous classifier we considered the optimal likelihood ratio adapted from Veldhuis and Bazen (2005) [37] by including the number of verification samples. With the comparison performance we can judge the effect of the template protection scheme, mainly due to the bit extraction part, on the classification performance.

1.4.2 Maximum Key Size

In Section 1.3.2 we outlined the influence of the size of the key on the irreversibility and renewability property of the template protection system. By assuming the bits of the key to be uniformly random and independent, the size of the key is indicative for its entropy. Hence, the irreversibility and renewability property can be optimized by maximizing the key size.

In Chapter 4 we analytically determine the maximum key size based on the naive Gaussian framework presented in Chapter 3. Similar to the published work of Willems and Ignatenko (2009) [38], we model the real-valued feature vectors as a Gaussian continuous source, which has a discriminating power equal to its Gaussian channel capacity. The discriminating power is referred to as the input capacity. However, our approach differs because we fix the input capacity and distribute the capacity among the feature components. Furthermore, we assume the error-correcting capability of the ECC to be equal to Shannon's bound[4].

With the analytical classification performance, determined in Chapter 3, we have the relationship between the classification performance and the number of bits that have to be corrected T, namely FNMR(T) and FMR(T). In Chapter 4 we combine this relationship with Shannon's theory, which stipulates the relationship between the key size and the error-correcting capability, and therefore we obtain the relationship between the performance and the key size. Furthermore, we also investigate the influence of the system parameters, which are the input capacity and the number of feature components, the number of enrolment and verification samples, and the target FNMR or FMR, on the key size. We extend the analysis by investigating the effect of distributing the input capacity uniformly or non-uniformly among the feature components and we also include the case where feature components are dependent.

1.4.3 Information Leakage of the Auxiliary Data

Our goal is to determine the information leakage of the auxiliary data $\{\mathrm{AD}_1, \mathrm{AD}_2\}$ about the key, the enrolment real-valued feature vector or binary vector affecting the irreversibility property, and to which extent can the auxiliary data be used for cross-matching, which will affect the desired unlinkability property. We perform this analysis on the bit protection part (Chapter 5) and the bit extraction part (Chapter 6) of the HDS in Figure 1.5 separately.

[4] In practice, ECCs cannot realize this bound and the results are therefore theoretical upper limits.

Bit Protection Part (AD_2)

Recent publications showed that AD_2 could be used for cross-matching due to the linear property of the ECC, known as the decodability attack in the literature [39, 40]. They determined the theoretical FMR when comparing AD_2 of arbitrary protected templates from different applications. In Chapter 5 we extend the analysis and also determine the theoretical FNMR. We show that as long as the HDS is balanced, i.e. there are equal number of enrolment and verification samples, the cross-matching classification performance is worse than the classification performance of the HDS. Besides the extended analysis, we also provide a solution for the decodability attack based on randomization in order to mitigate the cross-matching performance close to random.

Bit Extraction Part (AD_1)

Firstly, we analyze the information leakage of the bit extraction auxiliary data AD_1 of a specific bit extraction scheme affecting the irreversibility property. Secondly, for several bit extraction schemes we study the cross-matching performance of AD_1 affecting the unlinkability property.

With respect to the irreversibility analysis, it has been shown in Ballard et al. (2008) [41] that the bit extraction auxiliary data from certain schemes do indeed leak information that could be exploited by an adversary to improve its impersonation success rate by increasing the FMR. This information leakage affects the irreversibility property, because it is easier to guess the feature representation of the biometric data due to the increase of the FMR. We analyze the information leakage for the case of the Detection Rate Optimized Bit Allocation (DROBA) bit extraction scheme proposed in Chen et al. (2009) [42], which extracts multiple bits per feature component. We show with biometric data that AD_1 allows an adversary to increase the FMR by two orders of magnitude compared to the FMR obtained without access to AD_1. Furthermore, we analyze the cause of the information leakage and provide a remedy which essentially requires the restriction of the allocation freedom of the DROBA algorithm.

With respect to the unlinkability analysis, we study the cross-matching performance of AD_1 affecting the unlinkability property for several bit extraction schemes. In the literature, numerous bit extraction schemes have been proposed using subject-specific information stored in AD_1 in order to extract more robust bits, i.e. bits with a smaller bit-error probability [33–35, 42–45]. We limit the scope of our analysis to the simple binarization scheme, the reliable component selection (RCS) scheme [33–35], and the DROBA scheme [42].

Firstly, we demonstrate that the use of subject-specific information can improve the system classification performance. Secondly, we determine the cross-matching performance of the bit extraction auxiliary data and illustrate the difference between the system and cross-matching performance with respect to the number of enrolment and verification samples. The results show that the more subject-specific information the bit extraction uses, the greater its cross-matching performance will be. Having an unbalanced sys-

tem where the number of enrolment samples is greater than the number of verification samples can also cause the cross-matching performance to be better than the system performance. Thirdly, we show that reconstructing the bit allocation strategy from the verification samples, in order to prevent cross-matching, significantly deteriorates the system performance. Fourthly, we investigate whether the system performance can be improved by fusion of the system and the cross-matching performance.

1.4.4 Fusion

Fusion is the art of combining multiple sources of biometric information in order to improve the classification performance. The HDS system only outputs a decision which protects it against hill-climbing attacks, which are based on the availability of the score. However, the drawback of not having a score is that it is not possible to apply fusion at score-level. Therefore, published work on fusion with template protection are mainly focussed on fusion at feature-level or at decision-level [33, 34, 46–48]. However, we show in Chapter 7 that by extending the PI reconstruction process with the derivation of a dissimilarity score, it is possible to apply fusion at score-level, given some limitations on the match and non-match regions that can be created. Furthermore, we compare the fusion classification performance at score-level with the one obtained at feature-level and decision-level fusion. We will do this comparison for multi-sample and multi-algorithm fusion in Section 7.2 and Section 7.3, respectively. From our results we observe that, despite the aforementioned limitations of fusion at score-level, its classification performance outperforms fusion at feature-level or decision-level for multi-algorithm fusion, while no significant differences was found for multi-sample fusion.

1.5 Outline of the Thesis

Chapter 2 provides an overview of proposed template protection schemes known in the literature. We provide the advantages and disadvantages of the different types of template protection schemes and compare them with the scheme of interest in this thesis, namely the HDS scheme.

Chapter 3 answers the first research question of "Given the helper data template protection scheme, what is the **theoretical classification performance**?". We determine the theoretical classification performance of the HDS system assuming a Gaussian modeled biometric source and a single bit extraction scheme. We conclude the chapter with the comparison of the theoretical classification performance of the binary classifier, i.e. on the binary vector level, and the continuous classifier, i.e. on the real-valued feature level.

Chapter 4 answers the second research question of "Given the helper data template protection scheme, what is the **maximum key size** at a given target classification performance and system parameters?". With the theoretical classification performance of the binary classifier determined in Chapter 3 and the assumption that the ECC operates on Shannon's bound, we compute the maximum key size and analyze the influence of the system parameters, such as the discriminating power of the input Gaussian source and its

number of feature components, the number of enrolment and verification samples, and the target FNMR or the target FMR.

Chapter 5 and Chapter 6 combined answer the third research question of "Given the helper data template protection scheme, how does the **information leakage from the auxiliary data** affect the irreversibility and unlinkability property?". The information leakage from the auxiliary data of the HDS is performed in two parts, firstly the analysis of the bit protection part (AD_2) in Chapter 5 and secondly the analysis of the bit extraction part (AD_1) in Chapter 6. Chapter 5 investigates the cross-matching vulnerability, known as the decodability attack, affecting the unlinkability property of the bit protection part. Besides the analysis of the cross-matching performance we also propose a remedy based on randomization. On the other hand, Chapter 6 discusses the information leakage of the bit extraction part affecting both the irreversibility and unlinkability properties. We identify and solve the information leakage problem of the DROBA bit extraction scheme and we also investigate the relationship between the classification and cross-matching performances of different bit extraction schemes.

Furthermore, in Chapter 7 we answer the fourth research question of "Given the helper data template protection scheme, how can one realize **fusion with protected templates** and to what extent can it improve the classification performance?" We show that it is possible to improve the classification performance by applying multi-samples and multi-algorithm fusion at feature-, score-, and decision-level with the HDS template protection system.

We conclude the thesis with Chapter 8, where we outline the contributions and the answers to the research question of this thesis. We also propose recommendations and possible future directions.

It is noted to the reader that, as a consequence of integrating the full versions of the published papers in this thesis, some parts of the chapters may contain some overlapping content.

Chapter 2

Overview of Template Protection Schemes

2.1 Introduction

As described in Jain et al. (2008) [49], the template protection schemes proposed in the literature can be divided into two categories, namely (i) *feature transformation* and (ii) *biometric cryptosystems*.

Template protection schemes based on feature transformation essentially transform the enrolment biometric data using a transformation function in order to create the reference template. In order to protect the biometric data, the transformation should either be non-invertible, difficult to invert, or the transformation function or its parameters should be kept secret. In the verification phase, the same transformation is applied on the new biometric data before comparison. It is the intention to use the same classifier for the comparison of the transformed biometric data as would have been on the original biometric data.

Biometric cryptosystems on the other hand protect the biometric sample by either extracting a key from it or binding a key to it. The same key has to be extracted from the verification biometric sample or released from the reference template using the verification sample, respectively. The entropy of the key determines the amount of protection of the biometric data. Using the name cryptosystems may impose false expectations of the use of keys with an entropy common in the field of cryptography, which is currently advised by NIST to be at least 80 bits and will increase in 2011 to 112 bits [50]. As is known in the literature and as we will study in Chapter 4, the upperbound for the key size expressed in bits equal to $-\log_2(\text{FMR})$. Because the range of the operating FMR of published biometric performances are mainly in the order of $10^{-3} - 10^{-6}$ which correspond to the range of 10-20 bits, we cannot expect effective key sizes close to NIST requirements of at least 80 bits. Therefore, we suggest to use the label *key based protection* instead of *biometric cryptosystems*.

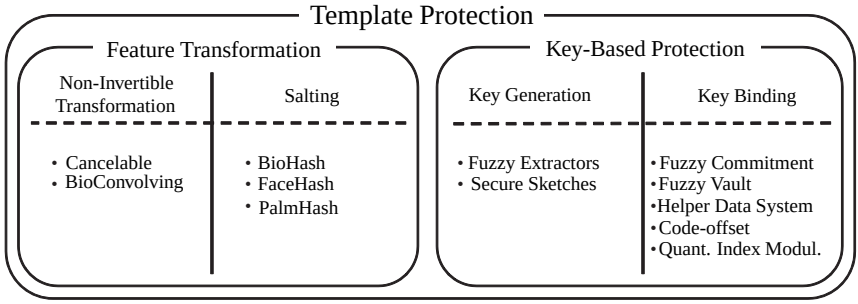


Figure 2.1: Overview of template protection techniques (adapted from Jain et al. (2008) [49]).

An overview of the template protection techniques adapted from Jain et al. (2008) [49] is portrayed in Figure 2.1. In the following sections we discuss the feature transformation and key-based template protection schemes in more detail.

2.2 Feature Transformation

As mentioned above, feature transformation schemes transform both the enrolment and verification biometric data using a transformation function and it is the intention to use the same classifier for the comparison of the transformed biometric data as would have been on the original biometric data. The transformation can be non-invertible, difficult to invert, or easy to invert. If the transformation is easy to invert, the transformation parameters have to be kept secret or depend on an external input such as a key, password, or PIN in order to protect the reference templates. The transformation schemes where the transformation parameters have to be kept secret are referred to as *salting* schemes and the schemes where the transformation is non-invertible or difficult to invert are labelled as the *non-invertible* transformation schemes.

2.2.1 Salting

For feature transformation schemes based on salting (see Figure 2.2), the transformation parameters have to be kept secret because the transformation itself is reversible to a certain extend. Examples of such schemes are the following. The work of Teoh and Ngo (2005) [51] proposes FaceHash in which the features extracted from the face are transformed depending on a random number from a token. Teoh et al. (2006) [52] introduces BioHash, which employs a random multispace quantization based on external input on the features extracted from face images, with similar implementation on iris [53] and palmprints named PalmHashing [54], or the work from Ong et al. (2008) [55] using dynamic quantization transformation on features extracted from fingerprints. Both the work of Farooq et al. (2007) [56] and of Lee and Kim (2010) [57] use a key or PIN from the subject

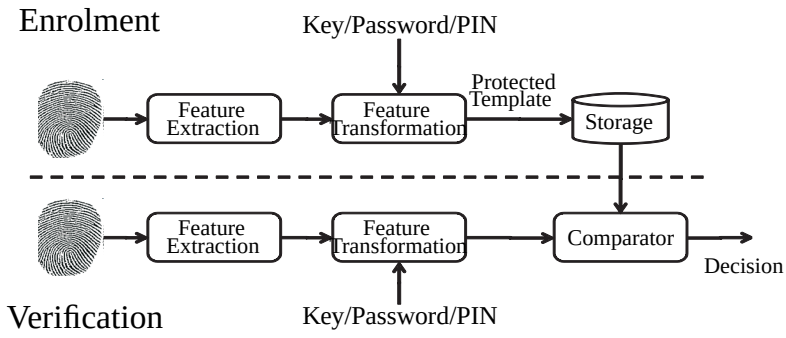


Figure 2.2: General depiction of a template protection scheme based on salting. Note that the randomness and protection of the feature transformation depends on the external input of either a key, password, or PIN from the subjects.

in order to randomize the binary string extracted from minutiae points from fingerprints.

The renewability property is based on the existence of a great number of different transformations. The irreversibility and unlinkability property are based on the secrecy of the transformation parameters.

Because the transformation parameters derived from a password, key or PIN provided by the subjects and are considered to be secret, the classification performance results of most of the published papers from above show that the performance of the protected templates are significantly better than the performance of the unprotected templates. Note that the performance improvements is mainly due to the fact that the classification performance is actually based on a multi-factor authentication setup of biometrics combined with either possession or knowledge entities. Therefore, most of the mentioned work also provide the performance for the scenario in which the transformation parameters are no longer considered to be secret. For this scenario the performance will depend on the biometric instance only. Hence, due to the multi-factor authentication approach, caution has to be taken when comparing the classification performance obtained with template protection schemes based on salting with the other types of template protection schemes.

2.2.2 Non-Invertible Transformation Schemes

A general depiction of the non-invertible transformation is shown in Figure 2.3. The most common technique based on non-invertible transformations is known as *Cancelable Biometrics* and was first introduced by Ratha et al. (2001) [58]. The main difference between cancelable biometrics and salting is the use of a non-invertible transformation that does not necessarily need an external input and due to the non-invertible property it is impossible to obtain the original biometric sample from the reference template. Note that the non-invertible transformation can also be applied on the biometric sample itself, such as a face or fingerprint image, without the need to extract a feature vector first. Some non-invertible transformations adapted from Ratha et al. [59] are portrayed in Figure 2.4. In

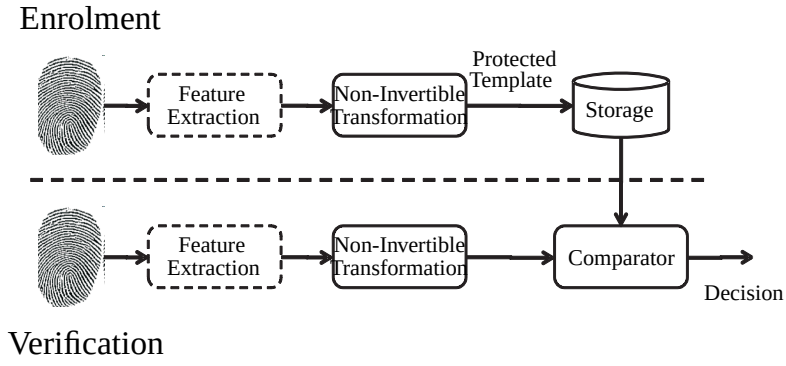


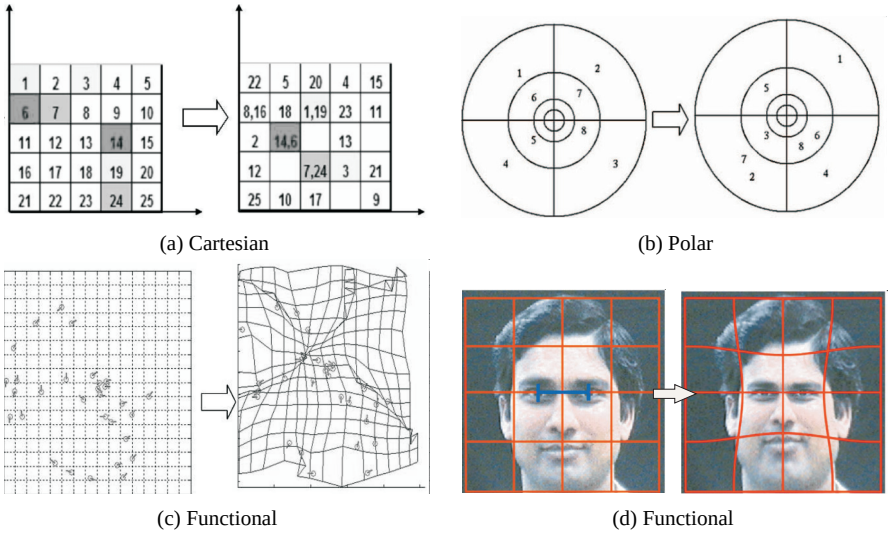
Figure 2.3: General depiction of a template protection scheme based on non-invertible transformation without use of an external input as for the case of salting in Figure 2.2.

case of the Cartesian transformation of Figure 2.4(a) the 2D feature space of for example a minutiae point cloud is divided into smaller squares which are shuffled by the transformation. Non-invertibility is achieved by merging several squares into a single one. For the polar transformation Figure 2.4(b), radial and angular sectors are created instead of squares and shuffled by the transformation. Similarly, non-invertibility is achieved by merging multiple sectors. The third transformation type is the functional transformation such as the folding transformation shown in Figure 2.4(c) or the morphing transformation shown in Figure 2.4(d). Another cancelable approach specifically for biometrics whose template can be represented by a set of sequences is known as BioConvolving proposed by Maiorana et al. (2010) [47]. The sequence is chopped into multiple parts of equal length, from which the convolution is taken of a number of randomly selected parts. Some other work focussed on obtaining registration free cancelable template from the minutiae set are Chikkerur et al. 2009 [60] and Yang et al. (2010) [61]. The work of Bringer et al. (2009) [62] propose a method to create cancelable templates that are time-dependent. Cancelable transformation can also be applied on the binary vector extracted from the biometric sample as has been shown in Zuo et al. (2008) [63] for iris images. The binary vector is divided into several smaller binary vectors and the XOR and XNOR operation is taken on randomly selected pairs creating a new and protected binary vector.

The renewability property is based on the existence of a great number of different transformations. The irreversibility and unlinkability property are based on the non-invertibility of transformation. The main drawback of the cancelable approach is the fact that the classification performance is reduced as can be seen in [47, 59, 62].

2.3 Key-Based Protection

As previously mentioned, there are two types of key-based template protection schemes, namely (i) schemes that bind a key to the biometric data in the enrolment phase and subsequently releases the same key from the reference template with use of the biometric



(a) Cartesian (b) Polar

(c) Functional (d) Functional

Figure 2.4: Examples of cancelable transformations adapted from Ratha et al. (2007) [59], namely (a) Carthesian, (b) polar and (c-d) functional transformations. Printed with the permission of the publisher, © 2007 IEEE.

data in the verification phase, and (ii) schemes that extract a key from the biometric data in the enrolment phase that has to be reproduced from the biometric data in the verification phase.

2.3.1 Key Binding

A general description of the key binding and release template protection technique is depicted in Figure 2.5. The principle idea is to bind or embed an arbitrary key with the biometric in the enrolment phase such that the protected template in the ideal case does not reveal any information from the enrolment biometric data. In the verification phase, the key can be released by combining the protected template with a newly captured probe biometric sample.

Examples of known key binding and release implementations are the *Code-Offset Construction* [40, 64, 65], *Fuzzy Commitment Scheme* (FCS) [36, 66–71], the *Helper Data System* (HDS) [33–35, 72, 73], *Quantization Index Modulation* (QIM) [48, 74, 75], and the *Fuzzy Vault* [46, 76–86]. The first three schemes are related to each other as portrayed in Figure 2.6.

The code-offset construction is common in both the FCS and HDS. In the enrolment phase, the code-offset construction consists of the random generation of the key $\mathbf{K} \in \{0,1\}^{k_c}$ in the *Random Number Generator* module, the encoding of the key to a codeword $\mathbf{C} \in \{0,1\}^{n_c}$ from the codebook $\mathcal{C}$ by the *ECC Encoder* module, and the XOR

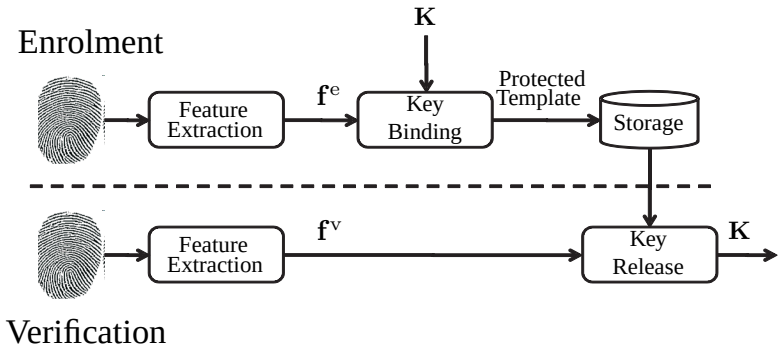


Figure 2.5: General depiction of the key binding and release template protection technique.

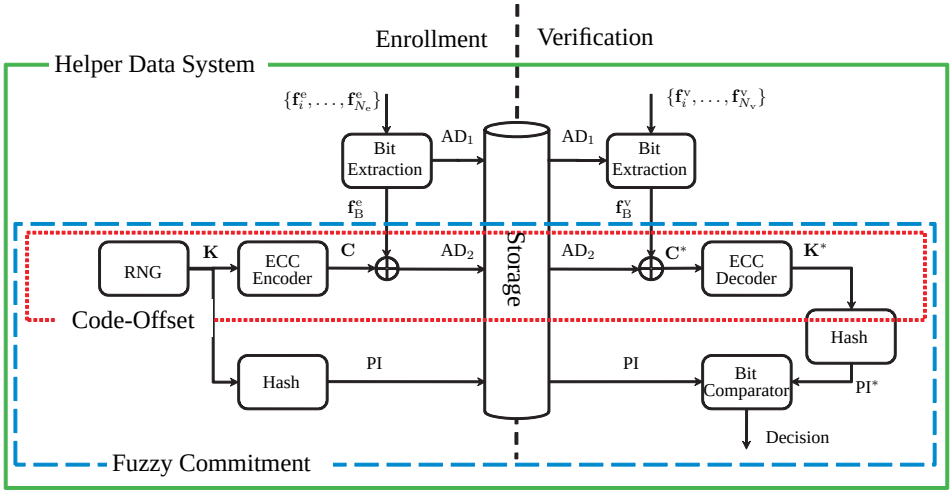


Figure 2.6: Key binding schemes.

operation of the codeword with the biometric binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ creating the code-offset auxiliary data AD_2. The XOR operation can be considered to be similar to a one-time-pad encryption algorithm. In the verification phase, the new biometric binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$ is XORed with AD_2 resulting into the possibly corrupted codeword $\mathbf{C}^* = \mathrm{AD}_2 \oplus \mathbf{f}_{\mathrm{B}}^{\mathrm{v}} = \mathbf{C} \oplus (\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \oplus \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}) = \mathbf{C} \oplus \mathbf{e}$, where the Hamming distance $\epsilon = d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}) = ||\mathbf{e}||$ indicates the number of errors corrupting the codeword $\mathbf{C}$. Decoding $\mathbf{C}^*$ by the *ECC Decoder* module leads to the candidate key $\mathbf{K}^*$. Note that the ECC enables the scheme to be robust against bit differences between binary vector from the enrolment and verification phase induced by measurement noise or biometric variability. The candidate key $\mathbf{K}^*$ is equal to

the enrolment key only if the number of bit errors ϵ is smaller than the error correcting capability of the ECC indicated by t_{c}. The irreversibility is based on the fact that given AD_2 there are $2^{k_{\mathrm{c}}}$ possibilities of $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$. Similarly, the renewability property is based on the fact that given a $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ there are $2^{k_{\mathrm{c}}}$ different AD_2. The unlinkability property is based on the fact that the $2^{k_{\mathrm{c}}}$ different AD_2 cannot be linked. Hence, the size of the key is indicative for the irreversibility, renewability and unlinkability properties of the template protection scheme.

With the code-offset construction only it is not possible to determine whether the candidate key $\mathbf{K}^*$ from the verification phase is equal to the enrolment key. This is made possible with the FCS, which is the extension of the code-offset construction by hashing the key into the pseudonymous identifier (PI) and storing both PI and AD_2 as the protected template. In the verification phase the candidate key $\mathbf{K}^*$ is hashed into PI^*. The candidate and enrolment key are equal if the two hashes PI and PI^* are equal and the decision of the *Bit Comparator* will be a match, otherwise a non-match. Storing the hash of the key adds an additional requirement for the irreversibility property, namely that it should be impossible or at least difficult to derive the key from its hashed version. As described in Juels and Wattenberg (1998) [36], $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ is equivalent to the *witness* that is used to commit the codeword $\mathbf{C}$ by means of the XOR operation. The outcome of the commitment is the AD_2 and PI pair, which together is also known as the *blob*. To successfully decommit the blob, a new witness $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$ has to be provided that is within t_{c} bit differences from the original witness $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$.

The HDS extends the FCS with a *Bit Extraction* module in order to convert the real-valued feature vectors into a binary string that can be used within the FCS. The bit extraction module can use subject-specific information, which is stored as the bit extraction auxiliary data AD_1. An additional requirement for the HDS is that AD_1 should not leak much information about the biometric data possibly affecting the irreversibility, renewability, and unlinkability properties.

A common requirement of the code-offset construction, HDS, and FCS schemes is that the input feature vector has to be of fixed length and ordered. The main differences between these schemes observed in the literature is the use of different ECC codes or different bit extraction schemes. Examples of different ECCs include the use of BCH codes in Tuyls et al. (2005) [35], concatenated codes such as a Reed-Solomon and Hadamard code in Hao et al. (2006) [66], maximum likelihood decoder in Chang and Roy (2007) [65], or product codes in Bringer et al. (2008) [71]. One of the first papers using ECC is Davida et al. (1998) [87]. Examples of different bit extraction schemes include the reliable component selection (RCS) method in Tuyls et al. (2005) [35], or the multi-bits extraction schemes such as the subject specific quantizer presented in Chang et al. (2004) [44] or the detection rate optimized bit allocation (DROBA) in Chen et al. (2009) [42].

QIM schemes protect the biometric data by introducing ambiguity into the bit extraction scheme cf. [48, 74, 75]. In the case when a single bit is extracted from a component, the single dimension feature space is divided into equidistant quantization bins with alternating bit values of either '0' or '1'. The ambiguity is introduced by having multiple bins with the same bit value. Increasing the number of bins with the same bit value increases

the ambiguity and the protection capability. The auxiliary data indicates the distance of the biometric sample with respect to the middle of the closest quantization bin with a bit value corresponding to the key. The same shift is applied on the verification sample before quantization. By increasing the quantization bins, the bit error probability will decrease, but on the other hand the ambiguity will also decrease. If the ambiguity decreases the protection capability also decreases.

Fuzzy vault schemes lock the key within the vault which can only be opened with another biometric sample that is similar to the enrolment sample. A common implementation of the fuzzy vault is shown in Figure 2.7. In the enrolment phase, first a polynomial function is created using the key. The biometric data, for example the x or y coordinates of minutiae points, are then projected on the polynomial as indicated by circles. In order to protect the biometric data, many random points referred to as chaff points (squares) are added in the space. The fuzzy vault or the protected template is thus the set of minutiae coordinates, their polynomial projection and the chaff points. In the verification phase, using the newly acquired probe biometric data, the closest points from the fuzzy vault to the biometric data are selected using a filtering mechanism. From these selected points, the same polynomial function is attempted to be reconstructed, leading to the same key. The vault is said to be unlocked if the same enrolment key has been recovered. Note that due to this approach there is no strict requirement to have a fixed length or ordered feature vector. It suffices to select enough points from the vault that would recover the same polynomial function and therefore the same key if there is a sufficient match. Therefore, the fuzzy vault is quite popular when extracting minutiae points from fingerprints, because the number of extracted minutiae points can vary significantly. Because the biometric data is actually stored in the clear but combined with chaff points, the protection of the fuzzy vault is thus based on the obfuscation of the biometric data by the chaff points. There is a tradeoff between the number of chaff points, the protection capability and the key recovery rate. For more detail about the fuzzy vault and its implementation we would refer the reader to the many published papers [46, 76–86].

For the key binding schemes, the irreversibility property is based on the difficulty of determining the key or biometric data from the output of the binding process. The renewability is based on the number of different keys that can be used in the binding process, while the unlinkability property requires that the binding output with different keys are not linkable.

Multiple key binding schemes can also be merged as shown in Nagar et al. (2008) [88] where they combined the fuzzy vault scheme with the fuzzy commitment scheme. The combination of the fuzzy commitment scheme and cancelable biometrics is shown in Bringer er al. (2008) [89].

2.3.2 Key Generation

The two commonly known key generation methods are the *Secure Sketch* and *Fuzzy Extractor*.

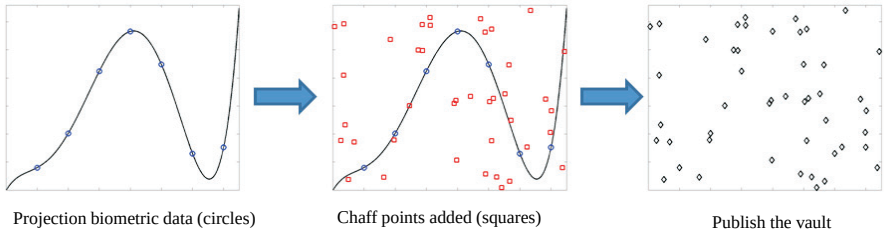


Figure 2.7: An example of the construction of a fuzzy vault.

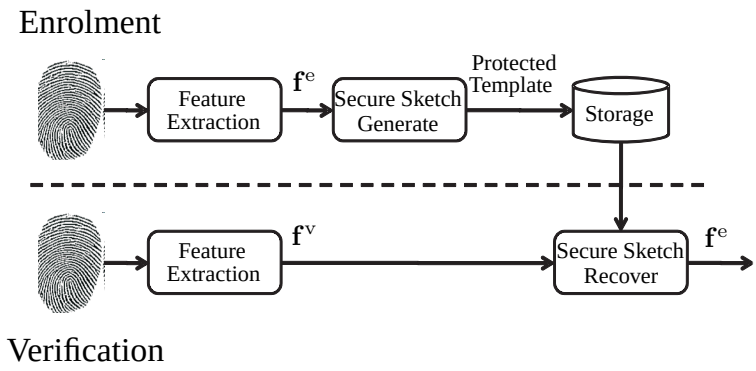


Figure 2.8: Depiction of the secure sketch.

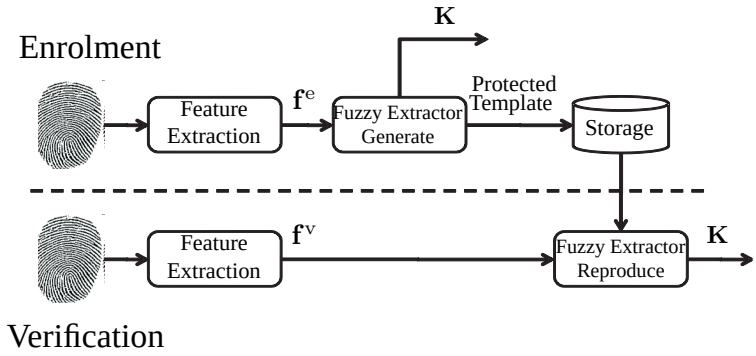


Figure 2.9: Depiction of the fuzzy extractor.

As described in Dodis et al. (2004) [64], *Secure Sketch* is defined as follows. There is a pair of procedures, namely (i) the *sketch* procedure that receives the enrolment biometric data $\mathbf{f}^e$ as input and outputs the public data P as the protected template, which does not

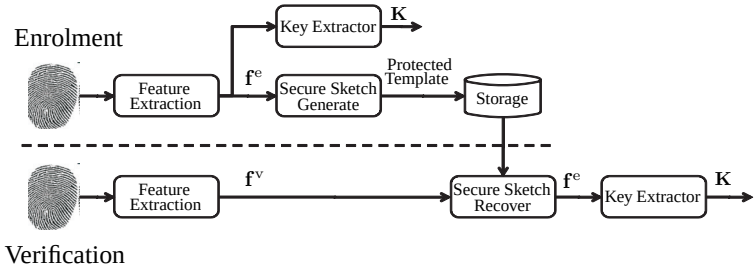


Figure 2.10: Depiction of a fuzzy extractor based on a secure sketch.

reveal too much information about $\mathbf{f}^{e}$, and (ii) the *recover* procedure which reconstructs $\mathbf{f}^{e}$ when the verification biometric sample $\mathbf{f}^{v}$ is similar enough with respect to $\mathbf{f}^{e}$. Because of the ability of reconstructing $\mathbf{f}^{e}$ it can be used as the key. A depiction of the secure sketch is shown in Figure 2.8. Examples of published papers based on secure sketch are [65,90,91]. Note that a secure sketch can be created by using the code-offset construction presented in Section 2.3.1. By taking the XOR of the candidate codeword $\mathbf{C}^{*}$ and the auxiliary data AD_2 we obtain the enrolment binary vector $\mathbf{f}_{\mathrm{B}}^{e}$ if and only if there is a match, namely when $\mathbf{C} = \mathbf{C}^{*}$.

A *Fuzzy Extractor* consists of a *generate* and *reproduce* procedure. Given the enrolment biometric data, the generate procedure outputs the public data P as the protected template and a key $\mathbf{K}$. Given the verification biometric data and the protected template P, the reproduce procedure outputs the same key $\mathbf{K}$ if the enrolment and verification samples are similar. A depiction of the fuzzy extractor is shown in Figure 2.9. Essentially, a fuzzy extractor can be created by combining a secure sketch with a key extractor that extracts a key with bits being close to uniformly random and independent. Examples of fuzzy extractors are given in [92–94]. The main difference between the fuzzy extractor and the secure sketch is the emphasis of the fuzzy extractor to extract a key with bits being close to uniformly random and independent, while the key properties within the secure sketch depend on the properties of the enrolment feature vector $\mathbf{f}^{e}$.

The irreversibility property is based on the difficulty of extracting information about the biometric data from the public data P. The renewability and unlinkability property are based on the capability of creating many different public data P that cannot be linked to the corresponding subject.

Chapter 3

Theoretical Classification Performance

3.1 Chapter Introduction

In this chapter the first research question will be addressed, namely

> **Given the HDS template protection scheme: What is the theoretical classification performance, and how do the system parameters influence it?**

In the first part, Section 3.2, we analytically determine the theoretical classification performance of the HDS by assuming the extracted feature vectors to be modelled by a Gaussian source and considering a single bit extraction scheme based on a single quantization threshold. The effect of the system parameters such as the number enrolment and verification samples is included in the analysis. The main results are published in Kelkboom et al. (2010) [95][1], which also includes the validation of the analysis using fingerprint and 3D face images.

In the second part, Section 3.3, we compare the classification performance of the HDS with the performance of the unprotected case. The HDS performance is equivalent to the classification performance on the binary level, while the performance for the unprotected case is equivalent to the performance on the continuous level. We consider the optimal likelihood ratio classifier as the continuous classifier. The main results are published in Kelkboom et al. (2010) [96][2].

[1] E. J. C. Kelkboom, G. Garcia Molina, J. Breebaart, R. N. J. Veldhuis, T. A. M. Kevenaar, and W. Jonker, "Binary biometrics: An analytic framework to estimate the performance curves under gaussian assumption," in *IEEE Transactions on Systems, Man and Cybernetics Part A, Special Issue on Advances in Biometrics: Theory, Applications and Systems*, vol. 40, no. 3, pp. 555-571, May 2010.

[2] E. J. C. Kelkboom, R. N. J. Veldhuis, and J. Breebaart, "Classification performance comparison of a continuous and binary classifier under gaussian assumption," *in The 31st Symposium on Information Theory in the Benelux*, 2010, pp. 129 - 136.

3.2 Binary Biometrics: An Analytic Framework to Estimate the Performance Curves under Gaussian Assumption

3.2.1 Abstract

In recent years the protection of biometric data has gained increased interest from the scientific community. Methods such as the fuzzy commitment scheme, helper data system, fuzzy extractors, fuzzy vault and cancelable biometrics have been proposed for protecting biometric data. Most of these methods use cryptographic primitives or error-correcting codes (ECC) and use a binary representation of the real-valued biometric data. Hence, the difference between two biometric samples is given by the Hamming distance or bit errors between the binary vectors obtained from the enrollment and verification phases respectively. If the Hamming distance is smaller (larger) than the decision threshold, then the subject is accepted (rejected) as genuine. Because of the use of ECCs, this decision threshold is limited to the maximum error-correcting capacity of the code, consequently limiting the false rejection rate (FRR) and false acceptance rate (FAR) trade-off. A method to improve the FRR consists in using multiple biometric samples in either the enrollment or verification phase. The noise is suppressed, hence reducing the number of bit errors and decreasing the Hamming distance. In practice, the number of samples is empirically chosen without fully considering its fundamental impact. In this work, we present a Gaussian analytical framework for estimating the performance of a binary biometric system given the number of samples being used in the enrollment and the verification phase. The error detection trade-off (DET) curve that combines the false acceptance and false rejection rates is estimated to assess the system performance. The analytic expressions are validated using the FRGC v2 and FVC2000 biometric databases.

3.2.2 Introduction

With the increased popularity of biometrics and its application in society, privacy concerns are being raised by privacy protection watchdogs. This has stimulated research into methods for protecting the biometric data in order to mitigate these privacy concerns. Numerous methods such as the *fuzzy commitment scheme* [36], *helper data system* [33, 34, 48], *fuzzy extractors* [64,65], *fuzzy vault* [80,84] and *cancelable biometrics* [58] have been proposed for transforming the biometric data in such a way that the privacy is safeguarded. Several of these privacy or template protection techniques use some cryptographic primitives (e.g. hash functions) or error-correcting codes (ECC). Therefore they use a binary representation of the biometric data, referred to as the *binary vector*. The transition from real-valued to binary representation of the biometric allows the difference between two biometric samples to be quantified by the Hamming distance (HD), i.e. the number of different bits (bit errors) between two binary vectors.

Eventually the biometric system has to verify the claimed identity of a subject. If verified, this identity is considered as genuine. The decision of either rejecting or accepting the subject as genuine depends on whether the Hamming distance is larger than a prede-

termined decision threshold (T). In template protection systems that use an ECC, T is usually determined by its error-correcting capacity. Hence, the false rejection rate (FRR) depends on the number of genuine matches that produce a Hamming distance larger than the decision threshold.

Attackers may attempt to gain access by impersonating a genuine subject. The associated comparisons are referred to as the imposter comparisons and will be accepted if the Hamming distance is smaller or equal to T, thus leading to a false accept. The success rate of impersonation attacks is quantified by the false acceptance rate (FAR).

Therefore, the performance of a biometric system can be expressed by its FAR and FRR, which depends on the genuine (ϕ_{ge}) and imposter (ϕ_{im}) Hamming distance probability mass functions (pmf) and the decision threshold T. A graphical representation is given in Figure 3.1.

One of the problems with template protection systems based on ECCs is that the FRR is lower bounded by the error-correcting capacity of the ECC. A large FRR makes the biometric system inconvenient, because many genuine subjects will be wrongly rejected. In some practical cases [33, 34] high FRR values were obtained because it was impossible to further increase the decision boundary, since the used ECC was unable to correct more bits. The method they used to improve the FRR consists in using multiple biometric samples in order to suppress the noise and thus reducing the number of bit errors resulting in a smaller Hamming distance.

The main objective of this study is to analytically estimate, under the Gaussian assumption, the performance of a biometric system based on binary vectors under Hamming distance comparison and considering the use of multiple biometric samples. We present a framework for analytically estimating both the genuine and imposter Hamming distance pmfs from the analytically estimated bit-error probability presented in [97] under the assumption that both the wihin- and between-class of the real-valued features are Gaussian distributed. Firstly, due to the central limit theorem we can assume that the real-valued features will tend to approximate a Gaussian distribution when they result from a linear combinations of many components, e.g. feature extraction techniques based on the principle component analysis (PCA) or linear discriminant analysis (LDA). PCA or LDA techniques are often being used to perform dimension reduction in order to prevent overfitting or to simplify the classifier [98], and in the field of template protection PCA is also used to decorrelate the features in order to guarantee uniformly distributed keys extracted from the biometric sample [65]. Secondly, the Gaussian assumption makes it possible to obtain an analytical closed-form expression for the Hamming distance pmf.

This paper is organized as follows. In Section 3.2.3 we present a general description of a biometric system with template protection and model each processing component. We present the Gaussian model assumption describing the probability density function (pdf) of the real-valued biometric features extracted from the biometric sample, the binarization method under consideration, and the interpretation of the template protection block. Then, we present the analytic expression for estimating the genuine and imposter Hamming distance pmfs, and the FRR and FAR curves in Section 3.2.4. In Section 3.2.5 we validate these analytic expressions with two different real biometric databases namely, the FRGC v2 3D face images [99] and the FVC2000 fingerprint images [100]. We further extend the framework in Section 3.2.6 and 3.2.7 in order to relax the assumptions made in

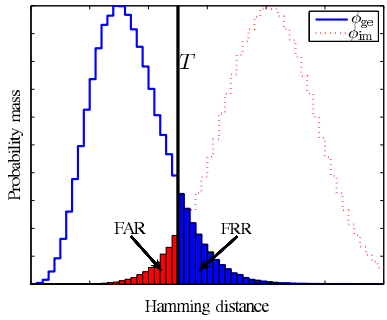


Figure 3.1: FRR and FAR from the genuine and imposter Hamming distance pmfs, ϕ_{ge} and ϕ_{im}, respectively.

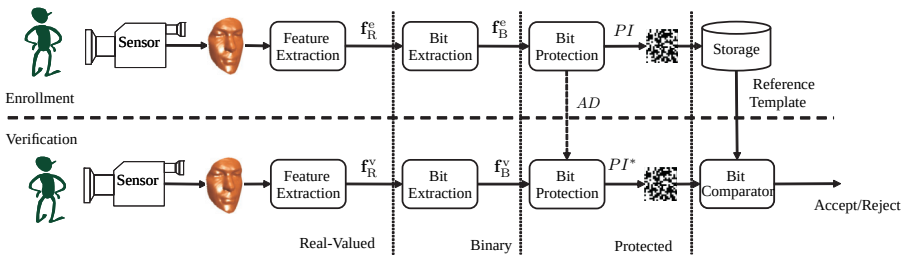


Figure 3.2: A general scheme of a biometric system with template protection based on helper data.

Section 3.2.3. Furthermore, some practical considerations are discussed in Section 3.2.8. Section 3.2.9 concludes this paper and outlines the future work.

3.2.3 Modeling of a Biometric System with Template Protection

A general scheme of a biometric system with template protection based on helper data is shown in Figure 3.2. In the enrollment phase a biometric sample, for example a 3D shape image of the face of the subject, is obtained by the acquisition system and presented to the Feature-Extraction module. The biometric sample is preprocessed (enhancement, alignment, etc.) and a real-valued *feature vector* $\mathbf{f}_{\mathrm{R}}^{\mathrm{e}} \in \mathbb{R}^{N_{\mathrm{F}}}$ is extracted, where N_{F} is the number of feature components or dimension of the feature vector. In the Bit-Extraction module, a binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \in \{0,1\}^{N_{\mathrm{B}}}$ is extracted from the real-valued feature vector, where N_{B} is the number of bits and in general does not need to be equal to N_{F}. Quantization schemes range from simple, extracting a single bit out of each feature component [33, 34] to more complex, extracting multiple bits per feature component [44, 101]. Hereafter, the binary vector is protected within the Bit-Protection module. The Bit-Protection module safeguards the privacy of the subjects of the biometric system by enabling accurate

comparisons without the need to store the original biometric data $\mathbf{f}_{\mathrm{R}}^{\mathrm{e}}$ or $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$. We focus on the helper data system that is based on ECCs and cryptographic primitives, for example hash functions. A unique but renewable key is generated for each subject and kept secret by using a hash function. Robustness to measurement noise and biometric variability is achieved by effectively using error-correcting codes. The output is a pseudonymous identifier (PI), represented as a binary vector, accompanied by some auxiliary data also known as helper data (AD) [102]. Finally, PI and AD have to be stored for use in the verification phase.

In the verification phase, another live biometric measurement is acquired from which its real-valued feature vector $\mathbf{f}_{\mathrm{R}}^{\mathrm{v}}$ is extracted followed by the quantization process, which produces the binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$. In the Bit-Protection module a candidate pseudonymous identifier PI* is created using AD and the binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$. There is an exact match between PI and PI* when the same AD is presented together with a biometric sample with similar characteristics as the one presented in the enrollment phase. In a classical biometric system, the comparator bases its decision on the similarity or distance between the feature vectors $\mathbf{f}_{\mathrm{R}}^{\mathrm{e}}$ and $\mathbf{f}_{\mathrm{R}}^{\mathrm{v}}$. For a binary biometric system, the decision is based on the difference between $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ and $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$, which can be quantified using the Hamming distance. For a template protection system, there is an acceptance only when PI and PI* are identical.

In summary, the biometric system incorporating template protection can be divided into three blocks; (i) the Acquisition and Feature-Extraction modules where the input is the subject's biometric and the output is a real-valued feature vector $\mathbf{f}_{\mathrm{R}} \in \mathbb{R}^{N_{\mathrm{F}}}$, (ii) the Bit-Extraction module that extracts a binary vector $\mathbf{f}_{\mathrm{B}}$ out of $\mathbf{f}_{\mathrm{R}}$, and (iii) the Bit-Protection and Bit-Matching modules which protects the binary vector and performs the matching and decision making based on PI and PI*. *To build an analytical framework, we have to model each block.* In this Section we present a simple model for each block. However, the simple model incorporating the Acquisition and Feature-Extraction block is built under strong assumptions and will be relaxed later in the paper.

Acquisition and Feature-Extraction Block

The input of the Acquisition and Feature-Extraction block is a captured biometric sample of the subject and the output is a real-valued feature vector $\mathbf{f}_{\mathrm{R}} = [f_{\mathrm{R}}[1], f_{\mathrm{R}}[2], \ldots, f_{\mathrm{R}}[N_{\mathrm{F}}]]'$ of dimension N_{F}, where ' ' is the transpose operator. The feature vector $\mathbf{f}_{\mathrm{R}}$ is likely to be different between two measurements, even if they are acquired immediately after each other. Causes for this difference include sensor noise, environment conditions (e.g. illumination) and biometric variabilities (e.g. pose or expression).

To model these variabilities, we consider Parallel Gaussian Channels (PGC) as portrayed in Figure 3.3. We assume an ideal Acquisition and Feature-Extraction module which always produces the same feature vector $\boldsymbol{\mu}_i$ for subject i. Such ideal module is thus robust against all aforementioned variabilities. However, the variability of component j is modeled as an additive zero-mean Gaussian noise $w[j]$ with its pdf $p_{\mathrm{w[j],}i} \sim \mathcal{N}(0, \sigma^2_{\mathrm{w},i}[j])$. Adding the noise $w[j]$ with the mean $\mu_i[j]$ results into the noisy feature component $f_{\mathrm{R}}[j]$, in vector notation $\mathbf{f}_{\mathrm{R}} = \boldsymbol{\mu}_i + \mathbf{w}$. The observed variability within one subject is characterized by the variance of the *within-class* pdf and is referred to as within-class variability.

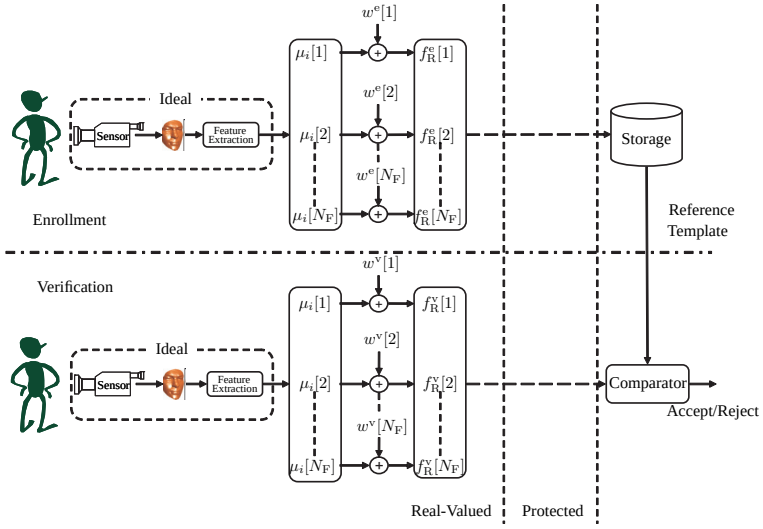


Figure 3.3: The Parallel Gaussian Channel for both the enrollment and verification phase.

We assume that each subject has the same within-class variance, i.e. homogeneous within-class variance $\sigma^2_{\mathrm{w},i}[j] = \sigma^2_{\mathrm{w}}[j], \forall i$. For each component, the within-class variance can be different and we assume the noise to be independent.

On the other hand, each subject should have a unique mean in order to be distinguishable. Across the population we assume $\mu_i[j]$ to be another Gaussian random variable with density $p_{\mathrm{b}[j]} \sim \mathcal{N}(\mu_{\mathrm{b}}[j], \sigma^2_{\mathrm{b}}[j])$. The variability of $\mu_i[j]$ across the population is referred to as the *between-class* variability. Figure 3.4 shows an example of the within-class and between-class pdfs for a specific component and a given subject. The *total* pdf describes the observed real-valued feature value $f_{\mathrm{R}}[j]$ across the whole population and is also Gaussian with $p_{\mathrm{t}[j]} \sim \mathcal{N}(\mu_{\mathrm{t}}[j], \sigma^2_{\mathrm{t}}[j])$, where $\mu_{\mathrm{t}}[j] = \mu_{\mathrm{b}}[j]$ and $\sigma^2_{\mathrm{t}}[j] = \sigma^2_{\mathrm{w}}[j] + \sigma^2_{\mathrm{b}}[j]$. For simplicity but without loss of generality we consider $\mu_{\mathrm{t}}[j] = \mu_{\mathrm{b}}[j] = 0$.

As depicted in Figure 3.3, in both the enrollment and verification phase the PGC adds random noise $\mathbf{w}^{\mathrm{e}}$ and $\mathbf{w}^{\mathrm{v}}$ with the same probability density to $\boldsymbol{\mu}_i$, resulting in $\mathbf{f}^{\mathrm{e}}_{\mathrm{R}}$ and $\mathbf{f}^{\mathrm{v}}_{\mathrm{R}}$, respectively. Thus $\boldsymbol{\mu}_i$ is sent twice over the same Gaussian channel.

Bit-Extraction Block

The function of the Bit-Extraction block is to extract a binary representation from the real-valued representation of the biometric sample. As the bit extraction method, we use the thresholding version used in [33, 34], where a single bit is extracted from each feature component. Hence, the obtained binary vector $\mathbf{f}_{\mathrm{B}} \in \{0,1\}^{N_{\mathrm{F}}}$ has the same dimension as $\mathbf{f}_{\mathrm{R}}$. Furthermore, the binarization threshold for each component $\delta[j]$ is set equal to the mean of the between-class pdf $\mu_{\mathrm{b}}[j]$; if the value of $f_{\mathrm{R}}[j]$ is smaller than $\delta[j]$ then it is set to "0" otherwise it is set to "1", see Figure 3.4. More complex binarization schemes

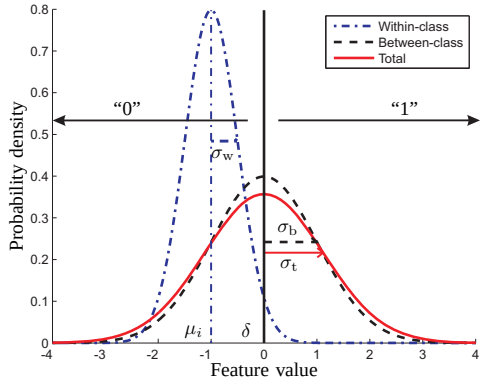


Figure 3.4: Modeling of a single feature component of the real-valued biometric.

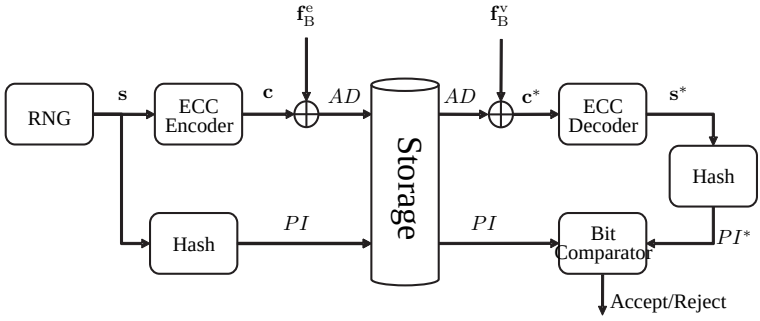


Figure 3.5: Fuzzy commitment scheme.

could be used [44, 101], but the simple binarization is used more frequently. Therefore, we only focus on the single bit binarization method. Note that the binarization method is similar in both the enrollment and verification phase. In the case where multiple biometric samples are used in either the enrollment (N_e) or verification (N_v) phase, the average of all the corresponding $\mathbf{f}_R$ is taken prior to the binarization process.

Bit-Protection and Bit-Comparator Block

Many bit protection or template protection schemes are based on the capability of generating a robust binary vector or key out of different biometric measurements of the same subject. However, the binary input vector $\mathbf{f}_B$ itself cannot be used as the key because it is most likely not exactly the same in both the enrollment and verification phase ($\mathbf{f}_B^e \neq \mathbf{f}_B^v$), due to measurement noise and biometric variability that lead to *bit errors*. The number of bit errors is also referred to as the Hamming distance $d_H(\mathbf{f}_B^e, \mathbf{f}_B^v)$. Therefore, error-correcting codes are used to deal with these bit errors. A possible way of integrating an

ECC is shown in Figure 3.5, which is also known as the fuzzy commitment scheme [36].

In the enrollment phase, a binary secret or message vector $\mathbf{K}$ is randomly generated by the *Random-Number-Generator* (*RNG*) module. The security level of the system is higher at larger secret lengths. A codeword $\mathbf{C}$ of an error-correcting code is obtained by encoding $\mathbf{K}$ in the *ECC-Encoder* module. The codeword is XOR-ed with $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ in order to obtain the auxiliary data AD. Furthermore, the hash of $\mathbf{K}$ is taken in order to obtain the pseudonymous identifier PI. For the sake of coherence we use the terminology proposed in [102, 103].

In the verification phase, the possibly corrupted codeword $\mathbf{C}^*$ is created by XOR-ing $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$ with AD. The candidate secret $\mathbf{K}^*$ is obtained by decoding $\mathbf{C}^*$ in the *ECC-Decoder* module. We compute the candidate pseudonymous identifier PI* by hashing $\mathbf{K}^*$. The decision in the Bit-Comparator block is based on whether PI and PI* are bitwise identical.

In order to illustrate our framework with practical parameter values, we choose the linear block type Bose, Ray- Chaudhuri, Hocquenghem (BCH) encoder/decoder as an example ECC. While more sophisticated ECCs can be used, the BCH accommodates our framework due to its Hamming distance classifier property. For example if we would consider the binary symbol based Reed-Solomon code, the number of bits it can correct depends on the error pattern. Hence, their probabilistic decoding behavior also needs to be modelled which is out of the scope of the framework described in this paper. The ECC is specified by the codeword length (n_{c}), message length (k_{c}), and the corresponding number of bits that can be corrected (t_{c}), in short $[n_{\mathrm{c}}, k_{\mathrm{c}}, t_{\mathrm{c}}]$. Because the BCH ECC can correct random bit errors, the Bit-Protection module yields equivalent PI and PI* when the number of bit errors between the binary vectors $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ and $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$ is smaller or equal to the error-correcting capability t_{c}. Thus, there is a match when the Hamming distance is smaller than t_{c}, $d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}) = ||\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \oplus \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}||_1 \leq t_{\mathrm{c}}$, and the Bit-Protection module can be modeled as a Hamming distance classifier with threshold t_{c}. Some $[n_{\mathrm{c}}, k_{\mathrm{c}}, t_{\mathrm{c}}]$ settings of the BCH code are given in Table 3.1. Note, that the maximum number of bits that can be corrected lies between 20-25% of the binary vector.

Table 3.1: Some examples of the BCH code given by the codeword (n_{c} and message (k_{c}) length, the corresponding number of correctable bits (t_{c}), and the bit error rate $t_{\mathrm{c}}/n_{\mathrm{c}}$.

n_{c}	k_{c}	t_{c}	BER $= t_{\mathrm{c}}/n_{\mathrm{c}}$
15	5	3	20.0%
	11	1	6.7%
31	6	7	22.6%
	16	3	9.7%
63	7	15	23.8%
	16	11	17.5%

Modeling Summary

Here follows a summary of the modeling choices and assumptions that we have made:

- **Acquisition and Feature-Extraction Block $\mathbf{f}_{\mathrm{R}}$**
  - Modeled as a Parallel Gaussian Channel, where each feature component is defined by:
    * Within-class pdf $\sim \mathcal{N}(0, \sigma^2_{\mathrm{w}}[j])$
      - Describes the genuine biometric variability and measurement noise
      - Homogeneous variance across subjects $\sigma^2_{\mathrm{w},i}[j] = \sigma^2_{\mathrm{w}}[j], \forall i$
      - Noise is independent across channels, measurements, and subjects
    * Between-class pdf $\sim \mathcal{N}(0, \sigma^2_{\mathrm{b}}[j])$
      - Characterizes the $\mu_i[j]$ variability across the population
      - Feature components are independent
    * Total pdf $\sim \mathcal{N}(0, \sigma^2_{\mathrm{t}}[j])$
      - Defines $f_{\mathrm{R}}[j]$ across the population
- **Bit-Extraction Block $\mathbf{f}_{\mathrm{B}}$**
  - Single bit extraction method, with binarization threshold $\delta[j] = \mu_{\mathrm{b}}[j]$
- **Bit-Protection and Bit-Comparator Block**
  - Hamming distance classifier with the ECC settings defining its decision boundary.

3.2.4 Analytical Estimation of Bit-Error Probabilities, FRR and FAR.

The goal of this study is to analytically estimate the performance of the presented general template protection system. In Section 3.2.3, we have presented a comprehensive description of such a system including the modeling approach or properties of each block that forms the basis of our analytic framework. In case of a Hamming distance classifier, the goal is to analytically estimate the expected genuine and imposter Hamming distance pmfs ϕ_{ge} and ϕ_{im}, respectively (see Figure 3.1). With these pmfs we can compute the false rejection rate β (FRR) and the false acceptance rate α (FAR), where β is the probability that a genuine subject is incorrectly rejected and α is the probability that an imposter is incorrectly accepted by the biometric system.

The Hamming distance between two binary vectors is the number of bit errors between them. Knowing the bit-error probability for each bit $P_{\mathrm{e}}[j]$, the expected Hamming distance $\bar{d}_H$ between $\mathbf{f}^{\mathrm{e}}_{\mathrm{B}}$ and $\mathbf{f}^{\mathrm{v}}_{\mathrm{B}}$ is

$$\bar{d}_H(\mathbf{f}^{\mathrm{e}}_{\mathrm{B}}, \mathbf{f}^{\mathrm{v}}_{\mathrm{B}}) = \sum_{j=1}^{N_{\mathrm{F}}} P_{\mathrm{e}}[j]. \tag{3.1}$$

Further, we define the pmf of the number of bit errors of component j as $P_j = [1 - P_{\mathrm{e}}[j], P_{\mathrm{e}}[j]]$, where $P_j(0)$ is the probability of no bit error ($d_{\mathrm{H}} = 0$) and $P_j(1)$

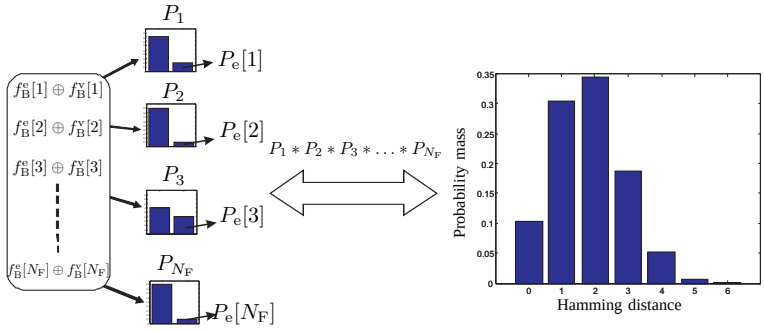


Figure 3.6: A toy example of the convolution method given by (3.2).

is the probability of a single bit error ($d_{\mathrm{H}} = 1$). Under the assumption that the bit-error probabilities are independent, the pmf of $d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}})$ is defined as

$$\begin{aligned} \phi(k) &\stackrel{\mathrm{def}}{=} \mathcal{P}\{d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}) = k\} \\ &= (P_1 * P_2 * \ldots * P_{N_{\mathrm{F}}})(k), \end{aligned} \tag{3.2}$$

where the convolution is taken of the pmf of the number of bit errors per component. A toy example is shown in Figure 3.6. For the two extreme cases of (3.2) we have

$$\phi(0) = \prod_{j=1}^{N_{\mathrm{F}}} P_j(0) = \prod_{j=1}^{N_{\mathrm{F}}} (1 - P_{\mathrm{e}}[j]), \tag{3.3}$$

$$\phi(N_{\mathrm{F}}) = \prod_{j=1}^{N_{\mathrm{F}}} P_j(1) = \prod_{j=1}^{N_{\mathrm{F}}} P_{\mathrm{e}}[j], \tag{3.4}$$

which are the probabilities of having zero or N_{F} errors, respectively. The FRR corresponding to a Hamming distance threshold T, $\beta(T)$, is the probability that the Hamming distance for a genuine comparison is greater than T, therefore

$$\begin{aligned} \beta(T) &= \mathcal{P}\{d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B},i}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B},i}^{\mathrm{v}}) > T\} \\ &= \sum_{k=T+1}^{N_{\mathrm{F}}} \phi_{\mathrm{ge}}(k). \end{aligned} \tag{3.5}$$

Furthermore, $\alpha(T)$ is the probability that the Hamming distance for an imposter comparison is smaller or equal to the threshold T, hence we have

$$\begin{aligned} \alpha(T) &= \mathcal{P}\{d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B},i}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B},j}^{\mathrm{v}}) \leq T, \forall i \neq j\} \\ &= \sum_{k=0}^{T} \phi_{\mathrm{im}}(k). \end{aligned} \tag{3.6}$$

In other words, if we want to estimate $\beta(T)$ and $\alpha(T)$ analytically we have to obtain an analytic closed-form expression of the average bit-error probability $P_{\mathrm{e}}[j]$ across the population for both the genuine and imposter case, $P_{\mathrm{e}}^{\mathrm{ge}}[j]$ and $P_{\mathrm{e}}^{\mathrm{im}}[j]$ respectively. Because

of the PGC modeling approach, $P_{\mathrm{e}}^{\mathrm{ge}}[j]$ will depend on the within-class and between-class variances $\sigma_{\mathrm{w}}^2[j]$ and $\sigma_{\mathrm{b}}^2[j]$, respectively. Furthermore, we also want to find the relationship between $P_{\mathrm{e}}^{\mathrm{ge}}[j]$ and the number of enrollment N_{e} and verification N_{v} samples. As mentioned in Section 3.2.3, in case of multiple samples the average of the extracted $\mathbf{f}_{\mathrm{R}}$ of each samples is taken prior to the binarization process.

P_{e} Estimation for the Imposter Case: $P_{\mathrm{e}}^{\mathrm{im}}$

For the imposter case, we are considering the comparison between binary vectors of two different subjects, $d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B},i}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B},j}^{\mathrm{v}}), \forall i \neq j$. As mentioned in Section 3.2.3, we focus on the binarization method based on thresholding with $\delta = \mu_{\mathrm{b}} = \mu_{\mathrm{t}}$ (see Figure 3.4). Because the total pdf is assumed to be Gaussian with mean μ_{t}, we have equiprobable bit values. This implies that the bit-error probability of randomly guessing a bit is 1/2, $P_{\mathrm{e}}^{\mathrm{im}}[j] = 1/2, \forall j$. Thus, under the assumption that the feature components are independent, imposter comparisons are similar to matching $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ with a random binary vector.

Since $P_{\mathrm{e}}^{\mathrm{im}}[j] = 1/2, \forall j$, we can simplify $\phi_{\mathrm{im}}(k)$ as the binomial pmf

$$\begin{aligned} \phi_{\mathrm{im}}(k) &= (P_1 * P_2 * \ldots * P_{N_{\mathrm{F}}})(k) && (3.7)\\ &= \binom{N_{\mathrm{F}}}{k} (P_{\mathrm{e}}^{\mathrm{im}}[j])^k (1 - P_{\mathrm{e}}^{\mathrm{im}}[j])^{N_{\mathrm{F}}-k} && (3.8)\\ &= \binom{N_{\mathrm{F}}}{k} 2^{-N_{\mathrm{F}}}, && (3.9) \end{aligned}$$

where the simplification step from (3.7) to (3.8) holds because of $P_{\mathrm{e}}^{\mathrm{im}}[i] = P_{\mathrm{e}}^{\mathrm{im}}[j],\ \forall\, i \neq j$. Furthermore, $\alpha(T)$ turns into

$$\alpha(T) = \sum_{k=0}^{T} \phi_{\mathrm{im}}(k) = 2^{-N_{\mathrm{F}}} \sum_{k=0}^{T} \binom{N_{\mathrm{F}}}{k}, \tag{3.10}$$

which corresponds to what is used in [104].

P_{e} Estimation for the Genuine Case: $P_{\mathrm{e}}^{\mathrm{ge}}$

We focus on estimating the bit-error probability for each component $P_{\mathrm{e}}^{\mathrm{ge}}[j]$, and for convenience purposes we omit the component index j. Using the Gaussian model approach as defined in Section 3.2.3 and depicted in Figure 3.7, the expected bit-error probability $P_{\mathrm{e}}^{\mathrm{ge}}$ over the whole population is defined by

$$\begin{aligned} P_{\mathrm{e}}^{\mathrm{ge}} &= E\left[P_{\mathrm{e}}^{\mathrm{ge}}(\mu)\right] \\ &= \int_{-\infty}^{\infty} p_{\mathrm{b}}(\mu) P_{\mathrm{e}}^{\mathrm{ge}}(\mu)\, \mathrm{d}\mu, \end{aligned} \tag{3.11}$$

where $P_{\mathrm{e}}^{\mathrm{ge}}(\mu)$ is the bit-error probability given μ and p_{b} is the between-class pdf. With the binarization threshold $\delta = \mu_{\mathrm{b}} = 0$, this problem becomes symmetric with respect to

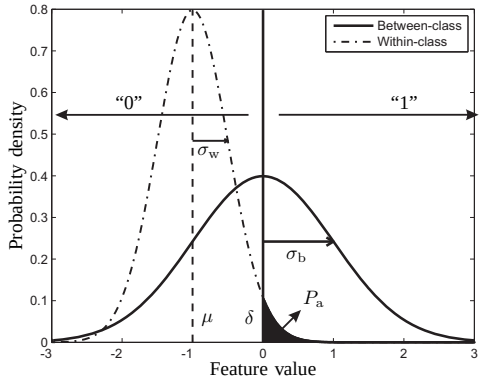


Figure 3.7: Measurement error P_a.

δ. Consequently, (3.11) becomes

$$\begin{aligned} P_e^{ge} &= 2 \int_{-\infty}^{0} p_b(\mu) P_e^{ge}(\mu)\, d\mu \\ &= 2 \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}\sigma_b} e^{-\left(\frac{\mu}{\sqrt{2}\sigma_b}\right)^2} P_e^{ge}(\mu)\, d\mu \\ &= \frac{2\lambda}{\sqrt{\pi}} \int_{-\infty}^{0} e^{-(\lambda\mu)^2} P_e^{ge}(\mu)\, d\mu, \end{aligned} \tag{3.12}$$

where $\lambda = \frac{1}{\sqrt{2}\sigma_b}$.

We define the measurement or acquisition error probability P_a, depicted by the shaded area in Figure 3.7, as the probability that the measured bit is different than the bit defined by the mean μ of the feature value. P_a becomes smaller at either a larger distance between μ and the binarization threshold δ or a smaller within-class variance. Since multiple enrollment (N_e) and verification (N_v) samples are considered, P_a also depends on the number of samples N, given as

$$P_a(\mu; N) = \int_{0}^{\infty} \frac{\sqrt{N}}{\sqrt{2\pi}\sigma_w} e^{-\left(\frac{\sqrt{N}(x-\mu)}{\sqrt{2}\sigma_w}\right)^2} dx, \tag{3.13}$$

where we used the fact that when averaging N samples the within-class variance decreases as

$$\sigma_{w,N}^2 = \frac{\sigma_w^2}{N} \Rightarrow \sigma_{w,N} = \frac{\sigma_w}{\sqrt{N}}. \tag{3.14}$$

With use of the error function

$$\mathtt{erf}(z) = \frac{2}{\sqrt{\pi}} \int_{0}^{z} e^{-t^2}\, dt. \tag{3.15}$$

and by defining $\eta = \frac{\sqrt{N}}{\sqrt{2}\sigma_{\mathrm{w}}}$, $P_{\mathrm{a}}(\mu; N)$ can be rewritten as

$$\begin{aligned} P_{\mathrm{a}}(\mu; N) &= \tfrac{\eta}{\sqrt{\pi}} \int_{0}^{\infty} \mathrm{e}^{-(\eta(x-\mu))^2} \mathrm{d}x \\ &= \tfrac{1}{\sqrt{\pi}} \int_{-\eta\mu}^{\infty} \mathrm{e}^{-z^2} \mathrm{d}z, \text{ with } z = \eta(x-\mu) \\ &= \tfrac{1}{\sqrt{\pi}} \left[\int_{0}^{\infty} \mathrm{e}^{-z^2} \mathrm{d}z - \int_{0}^{-\eta\mu} \mathrm{e}^{-z^2} \mathrm{d}z \right], \text{for } \mu \leq 0 \\ &= \tfrac{1}{\sqrt{\pi}} \left[\tfrac{\sqrt{\pi}}{2} - \tfrac{\sqrt{\pi}}{2} \mathtt{erf}(-\eta\mu) \right] \\ &= \tfrac{1}{2} \left[1 - \mathtt{erf}(-\eta\mu) \right], \end{aligned} \tag{3.16}$$

where we used the well known result $\int_0^\infty \lambda \mathrm{e}^{-(\lambda\mu)^2} \mathrm{d}\mu = \frac{\sqrt{\pi}}{2}$. There is a bit-error probability only when there is a measurement error at either the enrollment or the verification phase. If there is a measurement error in both phases then the measured bits still have the same bit value, thus no bit error. Hence, $P_{\mathrm{e}}(\mu)$ of (3.12) becomes

$$\begin{aligned} P_{\mathrm{e}}^{\mathrm{ge}}(\mu; N_{\mathrm{e}}, N_{\mathrm{v}}) = & (1 - P_{\mathrm{a}}(\mu; N_{\mathrm{e}})) P_{\mathrm{a}}(\mu; N_{\mathrm{v}}) \\ & + P_{\mathrm{a}}(\mu; N_{\mathrm{e}})(1 - P_{\mathrm{a}}(\mu; N_{\mathrm{v}})) \\ = & \tfrac{1}{4} \left[(1 + \mathtt{erf}(-\eta_{\mathrm{e}}\mu))(1 - \mathtt{erf}(-\eta_{\mathrm{v}}\mu)) \right. \\ & \left. + (1 - \mathtt{erf}(-\eta_{\mathrm{e}}\mu))(1 + \mathtt{erf}(-\eta_{\mathrm{v}}\mu)) \right] \\ = & \tfrac{1}{2} \left[1 - \mathtt{erf}(-\eta_{\mathrm{e}}\mu) \mathtt{erf}(-\eta_{\mathrm{v}}\mu) \right], \end{aligned} \tag{3.17}$$

where $\eta_{\mathrm{e}} = \frac{\sqrt{N_{\mathrm{e}}}}{\sqrt{2}\sigma_{\mathrm{w}}}$ and $\eta_{\mathrm{v}} = \frac{\sqrt{N_{\mathrm{v}}}}{\sqrt{2}\sigma_{\mathrm{w}}}$. By substituting (3.17) into (3.12) we obtain

$$\begin{aligned} P_{\mathrm{e}}^{\mathrm{ge}}(N_{\mathrm{e}}, N_{\mathrm{v}}) = & \frac{\lambda}{\sqrt{\pi}} \int_{-\infty}^{0} \mathrm{e}^{-(\lambda\mu)^2} \left[1 - \mathtt{erf}(-\eta_{\mathrm{e}}\mu) \mathtt{erf}(-\eta_{\mathrm{v}}\mu) \right] \mathrm{d}\mu \\ = & \frac{\lambda}{\sqrt{\pi}} \int_{0}^{\infty} \mathrm{e}^{-(\lambda\mu)^2} \left[1 - \mathtt{erf}(\eta_{\mathrm{e}}\mu) \mathtt{erf}(\eta_{\mathrm{v}}\mu) \right] \mathrm{d}\mu \\ = & \frac{1}{2} - \frac{\lambda}{\sqrt{\pi}} \int_{0}^{\infty} \mathrm{e}^{-\lambda^2\mu^2} \mathtt{erf}(\eta_{\mathrm{e}}\mu) \mathtt{erf}(\eta_{\mathrm{v}}\mu) \, \mathrm{d}\mu. \end{aligned} \tag{3.18}$$

The integral of the `erf` function can be solved using the general solution of `erf` integrals [105] given as

$$\int_{0}^{\infty} \mathrm{e}^{-\gamma x^2} \mathtt{erf}(ax) \mathtt{erf}(bx) \, \mathrm{d}x = \frac{\arctan\left(\frac{ab}{\sqrt{\gamma(a^2+b^2+\gamma)}} \right)}{\sqrt{\gamma\pi}}. \tag{3.19}$$

Thus, (3.18) can be solved by using (3.19) with $\gamma = \lambda^2$, $a = \eta_{\mathrm{e}}$, and $b = \eta_{\mathrm{v}}$ as

$$
\begin{aligned}
P_{\mathrm{e}}^{\mathrm{ge}}(N_{\mathrm{e}}, N_{\mathrm{v}}, \sigma_{\mathrm{w}}, \sigma_{\mathrm{b}}) &= \tfrac{1}{2} - \frac{\lambda}{\sqrt{\pi}} \frac{\arctan\left(\frac{\eta_{\mathrm{e}}\eta_{\mathrm{v}}}{\sqrt{\lambda^2\left(\eta_{\mathrm{e}}^2+\eta_{\mathrm{v}}^2+\lambda^2\right)}}\right)}{\lambda\sqrt{\pi}} \\
&= \tfrac{1}{2} - \tfrac{1}{\pi}\arctan\left(\frac{\eta\sqrt{N_{\mathrm{e}}N_{\mathrm{v}}}}{\lambda\sqrt{N_{\mathrm{e}}+N_{\mathrm{v}}+\left(\frac{\lambda}{\eta}\right)^2}}\right) \\
&= \tfrac{1}{2} - \tfrac{1}{\pi}\arctan\left(\frac{\sigma_{\mathrm{b}}\sqrt{N_{\mathrm{e}}N_{\mathrm{v}}}}{\sigma_{\mathrm{w}}\sqrt{N_{\mathrm{e}}+N_{\mathrm{v}}+\left(\frac{\sigma_{\mathrm{b}}}{\sigma_{\mathrm{w}}}\right)^{-2}}}\right),
\end{aligned}
\tag{3.20}
$$

where we also included σ_{w} and σ_{b} as an argument of the estimation function. As can be observed, $P_{\mathrm{e}}^{\mathrm{ge}}$ is dependent on the $\sigma_{\mathrm{b}}/\sigma_{\mathrm{w}}$ ratio, N_{e}, and N_{v}.

Summary

We have presented the analytic expressions of the genuine ($\phi_{\,\mathrm{ge}}$) and imposter (ϕ_{im}) Hamming distance pmfs and the corresponding FRR ($\beta(T)$) and FAR ($\alpha(T)$) curves. Because of the choice of the binarization scheme the imposter bit-error probability $P_{\mathrm{e}}^{\,\mathrm{im}}[j]$ does not need to be estimated and can be assumed to be equal to 1/2 for each feature component. However, the genuine bit-error probability $P_{\mathrm{e}}^{\mathrm{ge}}[j]$ has to be estimated using the analytic expression in (3.20). Therefore, in the remainder of this study we only need to estimate $P_{\mathrm{e}}^{\mathrm{ge}}[j]$ and for convenience reason we frequently omit the ge superscript.

3.2.5 Experimental Evaluation with Biometric Databases

In this section, the analytic expressions and the effect of the Gaussian assumption are validated using two real biometric databases, which are discussed in Section 3.2.5. To estimate $P_{\mathrm{e}}[j]$ using (3.20), we need to estimate the within- and between-class variances $\sigma_{\mathrm{w}}^2[j]$ and $\sigma_{\mathrm{b}}^2[j]$, respectively. In Section 3.2.5 we show that the within-class variance influences the between-class variance estimation and we present a corrected estimator. Due to the limited size of the databases, estimation errors do occur when estimating $P_{\mathrm{e}}[j]$ even in the case when the underlying model is correct. We account for these errors by estimating the 95 percentile boundaries in Section 3.2.5. We then present the results of estimating $P_{\mathrm{e}}[j]$ in Section 3.2.5, and the effect of using PCA as a mean to generate uncorrelated features in Section 3.2.5. We conclude by portraying the experimental $\phi_{\,\mathrm{ge}}(k)$, $\phi_{\mathrm{im}}(k)$, $\beta(T)$, $\alpha(T)$, and DET curves in Section 3.2.5.

Biometric Databases and Feature Extraction

The first database (db1) consists of 3D face images from the FRGC v2 dataset [99], where we used the shape-based 3D face recognizer of [106] to extract feature vectors of dimension $N_{\mathrm{orig}} = 696$. Subjects with at least 8 samples were selected resulting in $N_{\mathrm{s}} = 230$ subjects with a total of $N_{\mathrm{t}} = 3147$ samples. The number of samples per subject varies between 8 and 22 with an approximate average of $\bar{N}_{\mathrm{i}} = 14$ samples per subject. The second database (db2) consists of fingerprint images from Database 2 of FVC2000 [100], and uses a feature extraction algorithm based on Gabor filters and directional fields [107]

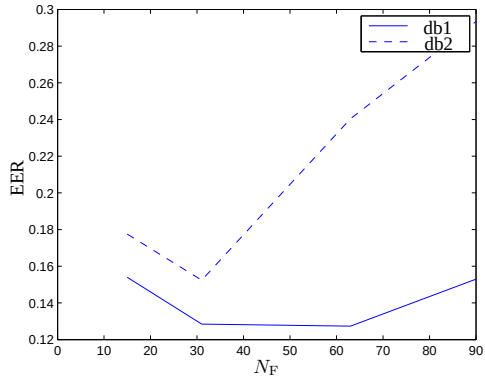


Figure 3.8: EER of the training set after applying PCA for different reduced number of features N_{F}.

resulting in 1536 features ($N_{\mathrm{orig}} = 1536$). There are $N_{\mathrm{s}} = 110$ subjects with $N_i = 8$ samples each. An overview is given in Table 3.2.

The components of the original feature vectors are dependent. Therefore, we applied the principle component analysis (PCA) technique to decorrelate the features and reduce the dimension of the feature space if necessary. Furthermore, we partitioned both databases into a training and testing set containing 25% and 75% of the number of subjects, respectively. The size of the test set is a very important factor in this analytic framework, thus we traded off the size of the training set and limited it to 25 % of the number of subjects. We applied PCA on the training set and reduced the dimensionality (N_{F}) of the feature vectors to the codeword lengths presented in Table 3.1 and computed the equal error rate (EER) (see Figure 3.8), which is defined as the point where FAR equals FRR. The optimal performance is computed using the bit-extraction method in Section 3.2.3 and a Hamming distance classifier. The optimal number of features for both db1 and db2 are in the range of 15, 31, and 63. Note that the best EER of 12.7% for db1 and 15.2% for db2 is higher than the reported performance of template protection systems based on these databases in the literature ($\approx 8\%$ for db1 in [33] and $\approx 5\%$ for db2 in [35])[3]. However, our proposed analytic framework is not focused on optimizing the performance but on analytically estimating the performance. The effect of the PCA transformation on the feature value distribution and the error probability estimation is discussed in Section 3.2.5. Unless stated otherwise, the remainder of this analysis is based on the PCA transformed test set using the PCA matrix obtained from the training set. For convenience, the remainder of this work is mainly focussed on the optimal setting of $N_{\mathrm{F}} = 31$.

Table 3.2: Overview of the biometric databases

Database	N_{orig}	N_{s}	N_{t}	$\bar{N}_{\text{i}} = N_{\text{t}}/N_{\text{s}}$
FRGC v2 (db1)	696	230	3147	≈ 14
FVC2000 (db2)	1536	110	880	8

Table 3.3: Variance estimation table as defined in [108].

Source of variation	*Sum of squares*	*d.f.*	Auxiliary
Within	$\sum_{i=1}^{N_{\text{s}}} \sum_{j=1}^{N_i} (f_{i,j} - \hat{\mu}_i)^2$	$N_{\text{t}} - N_{\text{s}}$	$\hat{\mu}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} f_{i,j}$
Between	$\sum_{i=1}^{N_{\text{s}}} N_i (\hat{\mu}_i - \hat{\mu})^2$	$N_{\text{s}} - 1$	$\hat{\mu} = \frac{1}{N_{\text{t}}} \sum_{i=1}^{N_{\text{s}}} \sum_{j=1}^{N_i} f_{i,j}$
Total	$\sum_{i=1}^{N_{\text{s}}} \sum_{j=1}^{N_i} (f_{i,j} - \hat{\mu})^2$	$N_{\text{t}} - 1$	

Variance Estimation of σ_{w}^2 and σ_{b}^2

The analytic expression $P_{\text{e}}^{\text{ge}}(N_{\text{e}}, N_{\text{v}}, \sigma_{\text{w}}, \sigma_{\text{b}})$ in (3.20) requires the standard deviations σ_{w} and σ_{b}. The estimated values $\hat{\sigma}_{\text{w}}$ and $\hat{\sigma}_{\text{b}}$ are obtained from the test set of the database under consideration. The variances $\hat{\sigma}_{\text{w}}^2$ and $\hat{\sigma}_{\text{b}}^2$ are estimated according to the *variance estimation* table given in Table 3.3 from [108], where $f_{i,j}$ is the jth real-valued feature vector of subject i, N_{s} is the number of subjects, N_i is the number of samples or feature vectors of subject i and N_{t} is the total number of samples $N_{\text{t}} = \sum_{i=1}^{N_{\text{s}}} N_i$. This table is also used in ANOVA (analysis of variance) models and describes the method for computing the *sum of squares* of the source of the within-class (SSW), between-class (SSB), and the total (SST) variation. Two important facts deriving from this table are that (i) the total sum of squares is equal to sum of the within-class and between-class sum of squares $\text{SST} = \text{SSW} + \text{SSB}$, and (ii) the total number of *degrees of freedom* (d.f.) is equal to the sum of the between-class and the within-class degrees of freedom. The details are in [108]. With the use of the table, the variance estimation is given as the sum of squares divided by the d.f., thus

$$\hat{\sigma}_{\text{w}}^2 = \frac{1}{N_{\text{t}} - N_{\text{s}}} \sum_{i=1}^{N_{\text{s}}} \sum_{j=1}^{N_i} (f_{i,j} - \hat{\mu}_i)^2, \tag{3.21}$$

$$\hat{\sigma}_{\text{b}}^2 = \frac{1}{\bar{N}_{\text{i}}(N_{\text{s}} - 1)} \sum_{i=1}^{N_{\text{s}}} N_i (\hat{\mu}_i - \hat{\mu})^2 \text{ with } \bar{N}_{\text{i}} = \frac{N_{\text{t}}}{N_{\text{s}}}, \tag{3.22}$$

$$\hat{\sigma}_{\text{t}}^2 = \frac{1}{N_{\text{t}} - 1} \sum_{i=1}^{N_{\text{s}}} \sum_{j=1}^{N_i} (f_{i,j} - \hat{\mu})^2, \tag{3.23}$$

[3] In [33] the most reliable feature components were selected and in [35] six enrollment samples were used.

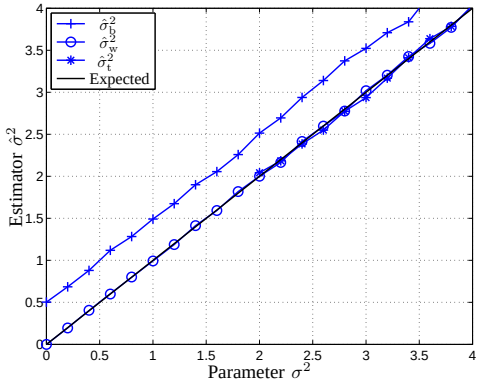


Figure 3.9: The within-class, between-class, and total variance estimation for different settings of $\{\sigma^2_{\rm w}, \sigma^2_{\rm b}\}$.

with the exception of $\hat{\sigma}^2_{\rm b}$, which is also divided by the average number of samples per subject $\bar{N}_{\rm i}$. Notice that $\hat{\sigma}^2_{\rm w}$ is calculated as the variance of the aggregated zero-mean samples of subjects, while taking into account that $N_{\rm s}$ degrees of freedom are lost because of the need to estimate the mean of each subject $\hat{\mu}_i$. Furthermore, $\hat{\sigma}^2_{\rm w}$ is also equal to the weighted average of the variance of each subject, because (3.21) can also be written as

$$\begin{aligned}
\hat{\sigma}^2_{\rm w} &= \tfrac{1}{N_{\rm t}-N_{\rm s}} \sum_{i=1}^{N_{\rm s}} (N_i - 1)\hat{\sigma}^2_{{\rm w},i} \\
&= \tfrac{1}{N_{\rm s}(\bar{N}_{\rm i}-1)} \sum_{i=1}^{N_{\rm s}} (N_i - 1)\hat{\sigma}^2_{{\rm w},i} \\
&= \tfrac{1}{\frac{1}{N_{\rm s}}\sum_{i=1}^{N_{\rm s}}(N_i-1)} \tfrac{1}{N_{\rm s}} \sum_{i=1}^{N_{\rm s}} (N_i - 1)\hat{\sigma}^2_{{\rm w},i}, \text{ with} \\
\hat{\sigma}^2_{{\rm w},i} &= \tfrac{1}{N_i-1} \sum_{j=1}^{N_i} \left(f_{i,j} - \hat{\mu}_i\right)^2,
\end{aligned} \tag{3.24}$$

which turns into $\hat{\sigma}^2_{\rm w} = \frac{1}{N_{\rm s}} \sum_{i=1}^{N_{\rm s}} \hat{\sigma}^2_{{\rm w},i}$ when N_i is equal for each subject.

The variance estimators are validated using a synthetically generated database of $N_{\rm s} = 1000$ subjects with $N_i = 4$ samples each. The parameters $\{\sigma^2_{\rm w}, \sigma^2_{\rm b}\}$ are used during the synthesis and we estimated $\{\hat{\sigma}^2_{\rm w}, \hat{\sigma}^2_{\rm b}, \hat{\sigma}^2_{\rm t}\}$ using (3.21), (3.22) and (3.23), respectively. The synthesis and estimation processes are performed ten times (10-fold) and the average of the result is taken. Figure 3.9 shows the estimation results of $\hat{\sigma}^2_{\rm w}$ for different values of $\sigma^2_{\rm w}$ with $\sigma^2_{\rm b} = 2$, and both $\hat{\sigma}^2_{\rm b}$ and $\hat{\sigma}^2_{\rm t}$ for different values of $\sigma^2_{\rm b}$ with $\sigma^2_{\rm w} = 2$. We can conclude that the $\hat{\sigma}^2_{\rm w}$ and $\hat{\sigma}^2_{\rm t}$ estimators give values that closely resemble the underlying model parameters $\sigma^2_{\rm w}$ and $\sigma^2_{\rm t}$, but we observe a constant estimation error for the $\hat{\sigma}^2_{\rm b}$ estimator . This estimation error is examined for different values of $\sigma^2_{\rm w}$ and N_i, as shown in Figure 3.10(a) and (b), respectively. The figures show that the estimation error increases when $\sigma_{\rm w}$ increases or when N_i decreases.

The constant estimation error of $\hat{\sigma}^2_{\rm b}$ is caused by the estimation error of the sample

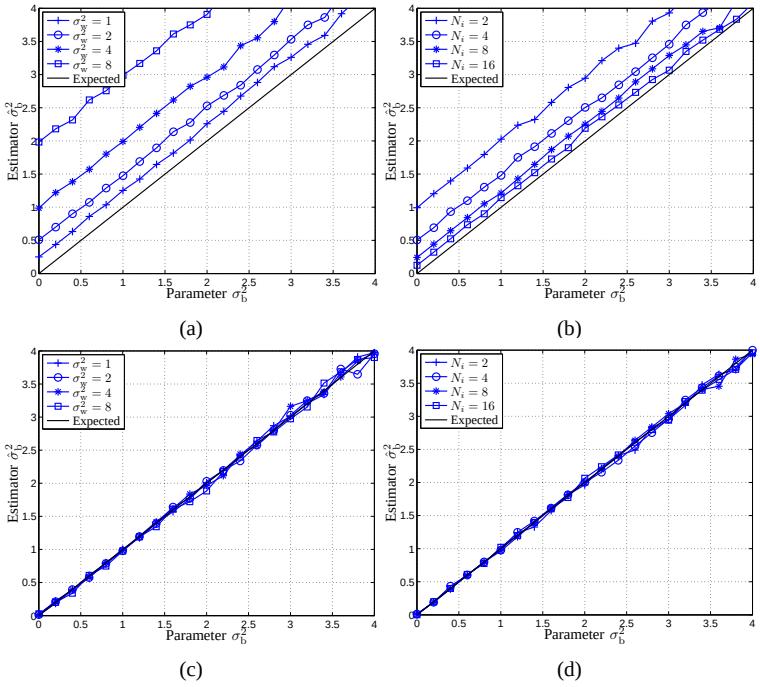


Figure 3.10: The between-class estimation of (3.22) at (a) different values of σ_{w}^2 with $N_i = 2$ and (b) different values of N_i with $\sigma_{\mathrm{w}}^2 = 2$, with its corrected version (3.27) in (c) and (d), respectively.

mean of each subject $\hat{\mu}_i$. From [108], we know that the variance of the sampling distribution of the sample mean $\hat{\mu}_i$ is given by

$$\sigma_{\hat{\mu}_i}^2 = \frac{\sigma_{\mathrm{w},i}^2}{N_i}. \tag{3.25}$$

If more samples are taken to estimate the sample mean, the estimation variance decreases. This implies that the estimation $\hat{\sigma}_{\mathrm{b}}^2$ of (3.22) is in fact

$$\hat{\sigma}_{\mathrm{b}}^2 = EST(\sigma_{\mathrm{b}}^2 + \sigma_{\hat{\mu}}^2) = EST(\sigma_{\mathrm{b}}^2 + \frac{\sigma_{\mathrm{w}}^2}{N_{\mathrm{i}}}), \tag{3.26}$$

where $EST(\tau) \triangleq \hat{\tau}$ is the estimation of parameter τ. The corrected version of the between-class estimation $\check{\sigma}_{\mathrm{b}}^2$ thus becomes

$$\check{\sigma}_{\mathrm{b}}^2 = \hat{\sigma}_{\mathrm{b}}^2 - \frac{\hat{\sigma}_{\mathrm{w}}^2}{N_{\mathrm{i}}}. \tag{3.27}$$

Figure 3.10(c)(d) shows the results of applying this correction on the results of Figure 3.10(a)(b) and the estimation has clearly improved.

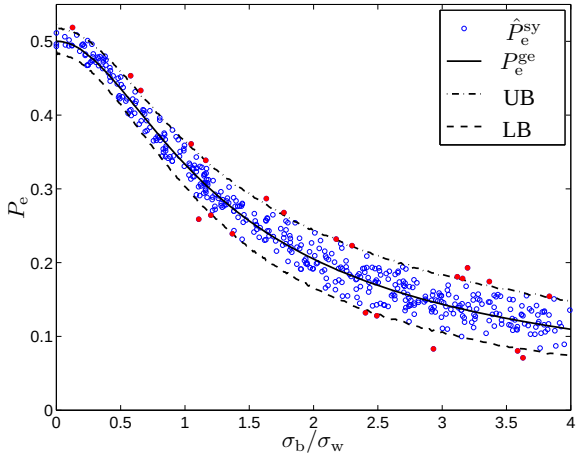


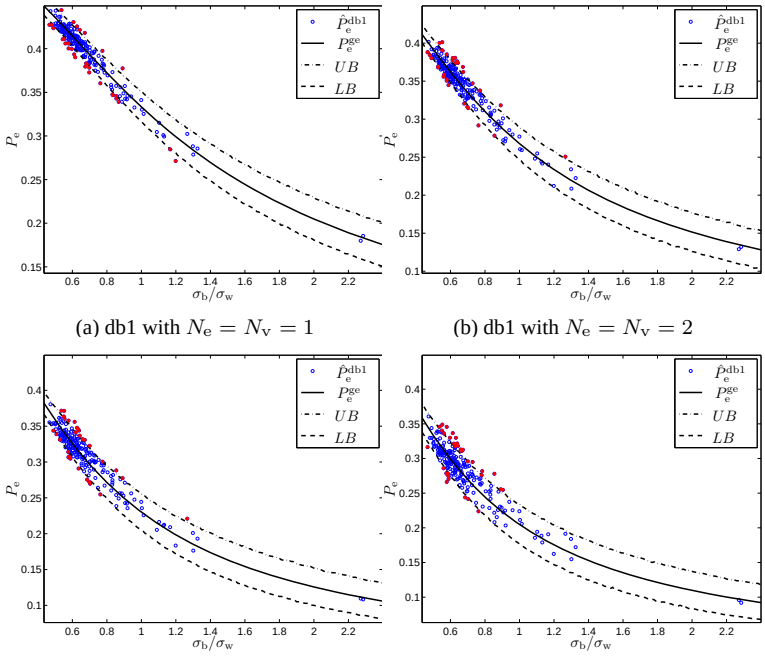
Figure 3.11: Random estimation errors due to the random nature and the upper (UB) and lower (LB) boundaries.

Boundaries of Tolerated Estimation Errors

When estimating $P_{\mathrm{e}}[j]$ of a given biometric database, there are always estimation errors because of its random nature. Even if we randomly generate a synthetic database that fully complies with the Gaussian modeling assumption, there are still estimation errors. These estimation errors are caused by the random nature of the problem and should be tolerated. Hence, we compute the upper (UB) and lower (LB) tolerance bounds for the estimation errors. Such an example is depicted in Figure 3.11 for a synthetic dataset of similar size as db2 ($N_{\mathrm{s}} = 110$ and $N_i = 8$) but with $N_{\mathrm{F}} = 500$ and $\sigma_{\mathrm{w}}^2[j] = 1$ with $\sigma_{\mathrm{b}}^2[j]$ randomly drawn from the uniform distribution $U(0, 16)$ with minimum and maximum value of 0 and 16, respectively. Figure 3.11 compares the estimated bit-error probability of the synthetic dataset $\hat{P}_{\mathrm{e}}^{\mathrm{sy}}[j]$ with the corresponding analytically obtained $P_{\mathrm{e}}^{\mathrm{ge}}[j]$, which stands for $P_{\mathrm{e}}^{\mathrm{ge}}(N_{\mathrm{e}}, N_{\mathrm{v}}, \hat{\sigma}_{\mathrm{w}}[j], \check{\sigma}_{\mathrm{b}}[j])$ of (3.20), where $\hat{\sigma}_{\mathrm{w}}[j]$ and $\check{\sigma}_{\mathrm{b}}[j]$ are estimated using (3.21) and (3.27), respectively. $\hat{P}_{\mathrm{e}}^{\mathrm{sy}}[j]$ is reported by a circle ('o') at its estimated $\hat{\sigma}_{\mathrm{b}}[j]/\hat{\sigma}_{\mathrm{w}}[j]$ ratio and its analytic estimation is the value of the solid line at the same $\hat{\sigma}_{\mathrm{b}}[j]/\hat{\sigma}_{\mathrm{w}}[j]$ ratio. A greater vertical distance implies a greater analytical estimation error.

The test protocol for calculating $\hat{P}_{\mathrm{e}}^{\mathrm{sy}}[j]$ is as follows: for each feature component, $\hat{P}_{\mathrm{e}}^{\mathrm{sy}}[j]$ is calculated as the average across the bit-error probability of each subject $\hat{P}_{\mathrm{e},i}^{\mathrm{sy}}[j]$. The subject bit-error probability $\hat{P}_{\mathrm{e},i}^{\mathrm{sy}}[j]$ results from performing 200 matches and determining the relative number of errors. For each match, N_{e} distinct feature vectors are randomly selected, averaged and binarized (enrollment phase). The obtained bit is compared to the bit obtained from averaging and binarizing N_{v} different randomly selected feature vectors of the same subject (verification phase).

We empirically estimate the upper (UB) and lower (LB) boundaries by clustering the



(a) db1 with $N_{\rm e} = N_{\rm v} = 1$

(b) db1 with $N_{\rm e} = N_{\rm v} = 2$

(c) db1 with $N_{\rm e} = N_{\rm v} = 3$

(d) db1 with $N_{\rm e} = N_{\rm v} = 4$

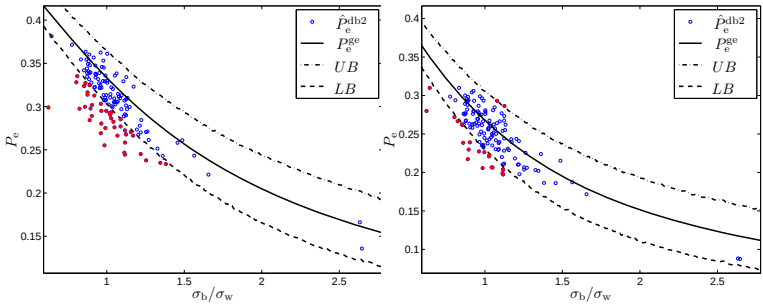
Figure 3.12: Comparison between $P_{\rm e}^{\rm ge}[j]$ and $\hat{P}_{\rm e}^{\rm db1}[j]$ for different settings (a) $N_{\rm e} = N_{\rm v} = 1$, (b) $N_{\rm e} = N_{\rm v} = 2$, (c) $N_{\rm e} = N_{\rm v} = 3$, and (d) $N_{\rm e} = N_{\rm v} = 4$. The circles (discs) correspond to cases where $\hat{P}_{\rm e}^{\rm db1}[j]$ falls within (outside) the boundaries.

points into equidistant intervals on the x-axis and compute the 95 percentile range of the $\hat{P}_{\rm e}^{\rm sy}[j]$ values in each interval. The circles (discs) correspond to cases where $\hat{P}_{\rm e}^{\rm sy}[j]$ is within (outside) the 95 percentile boundaries.

Validation of the Analytic Expression $P_{\rm e}^{\rm ge}$

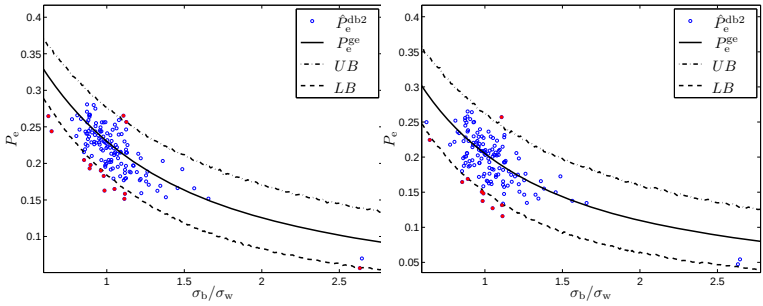
In this section we experimentally validate the analytic expression of the bit-error probability $P_{\rm e}^{\rm ge}$. In the previous section, we have discussed the use of PCA for decorrelating the feature components and for reducing the dimension to $N_{\rm F} = 31$. In order to have more components for the validation we apply PCA but without reducing the number of features. Hence, we consider the original number of features (696) for database db1. However, for database db2 we only consider 223 components since 25% of the total number of subjects (i.e. 28 subjects) with a total of 224 feature vectors were used to derive the PCA projection. Thus, to avoid singularities we have reduced the number of features to 223.

To assess the model assumptions, we compared the estimated bit-error probability of the biometric database $\hat{P}_{\rm e}^{\rm db}[j]$ with the corresponding analytically obtained $P_{\rm e}^{\rm ge}[j]$. The same test protocol is used as discussed in Section 3.2.5. The experimental results for db1



(a) db2 with $N_{\rm e} = N_{\rm v} = 1$

(b) db2 with $N_{\rm e} = N_{\rm v} = 2$



(c) db2 with $N_{\rm e} = N_{\rm v} = 3$

(d) db2 with $N_{\rm e} = N_{\rm v} = 4$

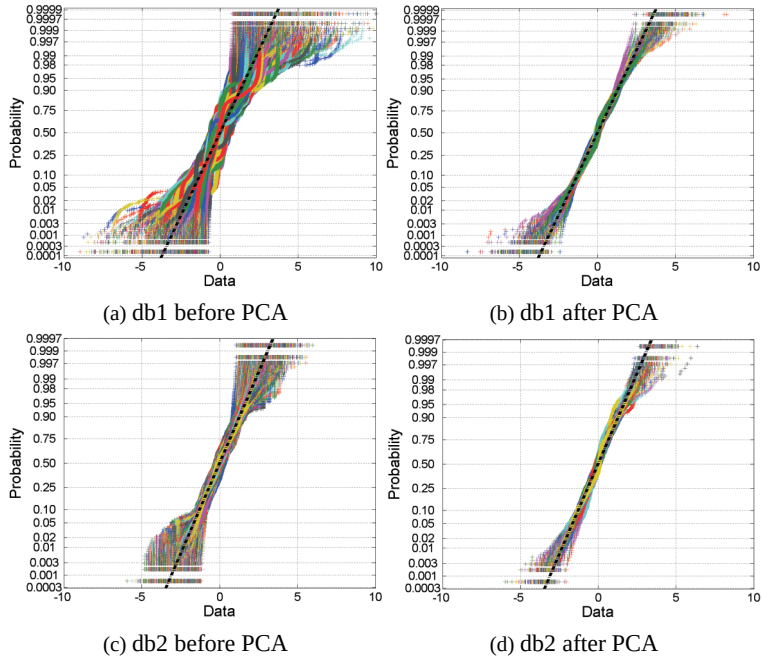
Figure 3.13: Comparison between $P_{\rm e}^{\rm ge}[j]$ and $\hat{P}_{\rm e}^{\rm db2}[j]$ for different settings (a) $N_{\rm e} = N_{\rm v} = 1$, (b) $N_{\rm e} = N_{\rm v} = 2$, (c) $N_{\rm e} = N_{\rm v} = 3$, and (d) $N_{\rm e} = N_{\rm v} = 4$. The circles (discs) correspond to cases where $\hat{P}_{\rm e}^{\rm db2}[j]$ falls within (outside) the boundaries.

and db2 for different values of $N_{\rm e}$ and $N_{\rm v}$ are shown in Figure 3.12 and Figure 3.13, respectively. The circles (discs) correspond to cases where $\hat{P}_{\rm e}^{\rm db}[j]$ is within (outside) the 95 percentile boundaries. We refer to the number of discs as the estimation error $\epsilon_{P_{\rm e}}$. If all the assumptions hold then we expect the relative $\epsilon_{P_{\rm e}}$ to be around 5%. Table 3.4 reports the absolute and relative $\epsilon_{P_{\rm e}}$. Because $\epsilon_{P_{\rm e}}$ is noisy due to the random selection of $N_{\rm e}$ and $N_{\rm v}$ samples within the test protocol, we repeat the estimation 20 times and report its mean. For db1, $\epsilon_{P_{\rm e}}$ is 16.7% for $N_{\rm e} = N_{\rm v} = 1$ and decreases to 13% for $N_{\rm e} = N_{\rm v} = 4$. In the case of db2, $\epsilon_{P_{\rm e}}$ is very large; 27.3% for $N_{\rm e} = N_{\rm v} = 1$ but decreases significantly when both $N_{\rm e}$ and $N_{\rm v}$ are increased, reaching 6.3% when $N_{\rm e} = N_{\rm v} = 4$. Thus, for both databases there is a clear improvement when increasing the number of samples. We conjecture that the improved bit-error probability estimation performance is due to the fact that the feature value distribution becomes more Gaussian when averaging multiple samples as stated by the central limit theorem [109]. Also note that many $\hat{P}_{\rm e}^{\rm db1}[j]$ estimations of db1 are very close to the 95 percentile boundaries, hence small estimation errors can lead to large variation in $\epsilon_{P_{\rm e}}$ that could explain the bit-error probability estimation performance differences between db1 and db2 observed in the table.

Table 3.4: The number of cases ϵ_{P_e} where $\hat{P}_e^{db}[j]$ is outside the 95% percentile boundaries per database and $\{N_e, N_v\}$ setting.

	db1		db2	
Setting	Abs. ϵ_{P_e}	Rel. ϵ_{P_e}.	Abs. ϵ_{P_e}	Rel. ϵ_{P_e}
$N_e = N_v = 1$	116	16.7 %	61	27.3%
$N_e = N_v = 2$	103	14.8 %	33	14.8%
$N_e = N_v = 3$	91	13.1 %	18	8.1%
$N_e = N_v = 4$	92	13.2 %	14	6.3%

The effect of PCA on the Gaussian Assumption

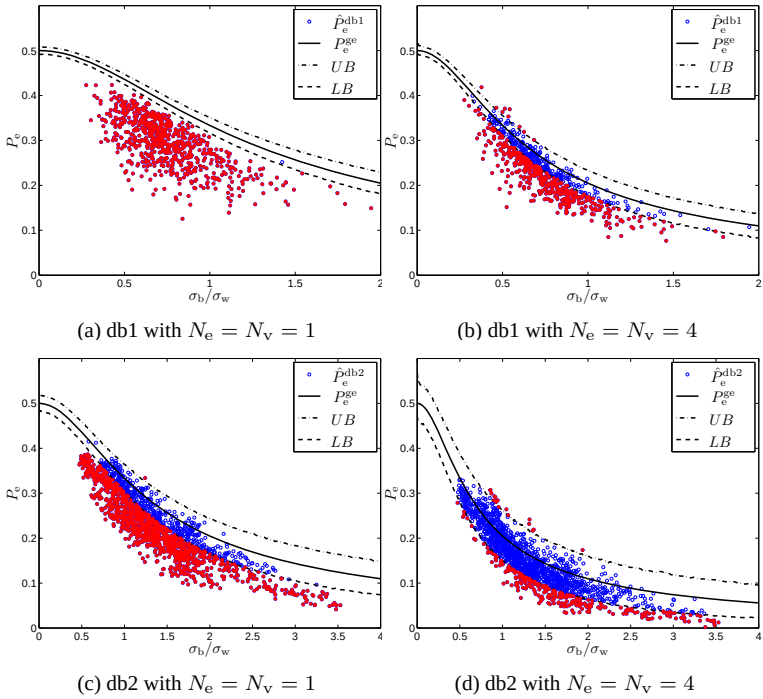


(a) db1 before PCA (b) db1 after PCA

(c) db2 before PCA (d) db2 after PCA

Figure 3.14: Normal probability plot of each feature vector component of db1 and db2 before and after applying PCA.

As described in Section 3.2.3, the analytic framework is based on the Gaussian model assumption. Figures 3.14(a)(c) show the normal probability plot for each component of the feature vectors of db1 and db2 respectively, before applying the PCA transformation. The normal probability plot is a graphical technique for assessing the degree to which a dataset approximates a Gaussian distribution. If the curve of the data closely follows the



(a) db1 with $N_{\rm e} = N_{\rm v} = 1$

(b) db1 with $N_{\rm e} = N_{\rm v} = 4$

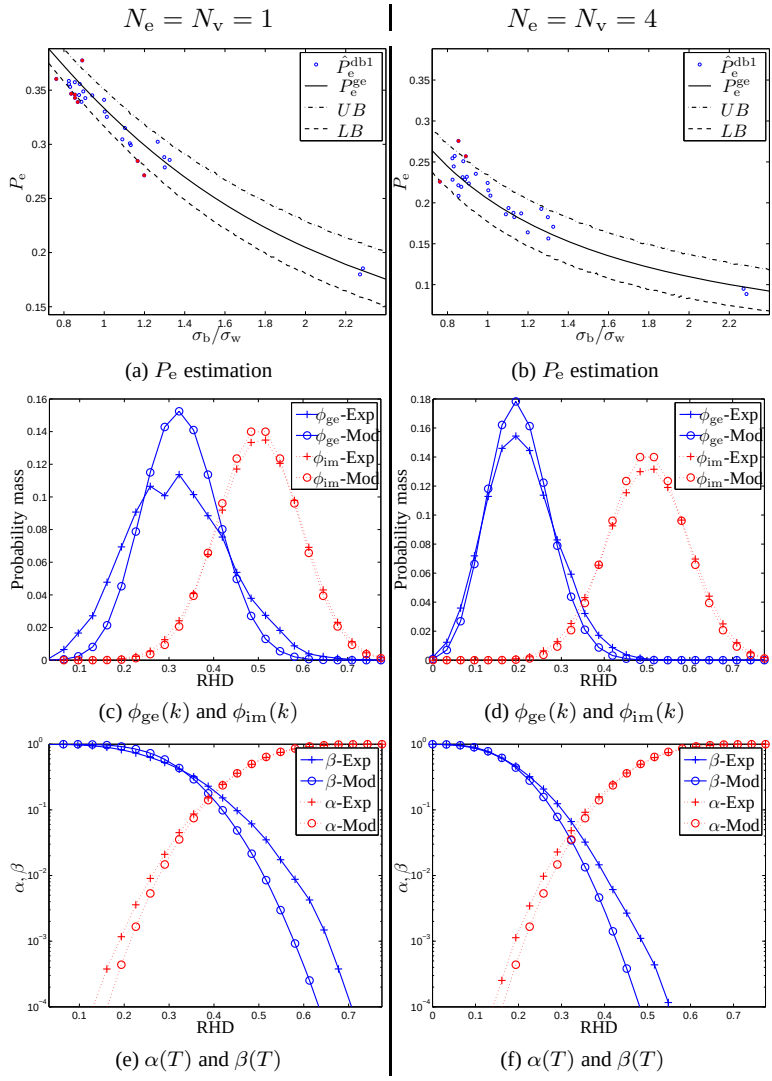
(c) db2 with $N_{\rm e} = N_{\rm v} = 1$

(d) db2 with $N_{\rm e} = N_{\rm v} = 4$

Figure 3.15: $\hat{P}_{\rm e}^{\rm dbx}[j]$ at different settings of $N_{\rm e}$ and $N_{\rm v}$ for both db1 and db2 before applying the PCA transform.

dashed-thick line then the data can be assumed to be approximately Gaussian distributed. Prior to comparing, we normalized each feature so that it has zero-mean and unit-variance. For both databases it is evident that the distributions before applying PCA are not Gaussian, because they significantly deviate from the dashed-thick line that represents a perfect Gaussian distribution. Figures 3.14 (b)(d) depict the normal probability plot for each of the 696 components of db1 and the 223 components of db2 respectively, after applying PCA. For both databases the figures show that after applying PCA the features tend to behave more like Gaussians. Yet, the tails deviate the most from being Gaussian where for the most cases the empirical distribution is wider.

Figure 3.15 shows the $P_{\rm e}$ estimations before applying PCA for both databases in two cases: $N_{\rm e} = N_{\rm v} = 1$ and $N_{\rm e} = N_{\rm v} = 4$. Note that before PCA db1 and db2 have 696 and 1536 components, respectively. For db1 $\epsilon_{P_{\rm e}}$ is equal to 99.8% for the $N_{\rm e} = N_{\rm v} = 1$ and 61.2% for the $N_{\rm e} = N_{\rm v} = 4$ case, while for db2 $\epsilon_{P_{\rm e}}$ is 71% and 18%, respectively. Comparing these results with the $\epsilon_{P_{\rm e}}$ values when applying PCA, see Table 3.4, we can also conclude that applying PCA makes the features significantly more Gaussian.



(a) P_e estimation (b) P_e estimation

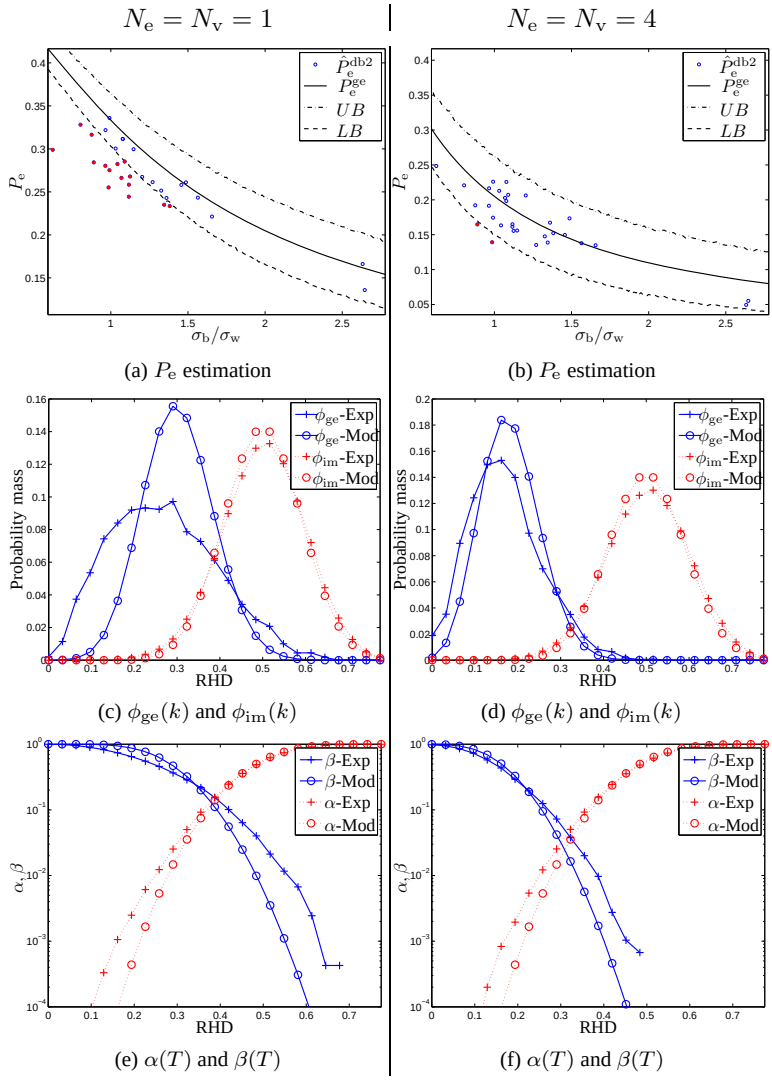
(c) $\phi_{ge}(k)$ and $\phi_{im}(k)$ (d) $\phi_{ge}(k)$ and $\phi_{im}(k)$

(e) $\alpha(T)$ and $\beta(T)$ (f) $\alpha(T)$ and $\beta(T)$

Figure 3.16: Results for db1 with $N_F = 31$, (a)(b) $\hat{P}_e^{db1}$ and the analytical estimation of P_e^{ge}, (c)(d) $\phi_{ge}(k)$ and $\phi_{im}(k)$ pmfs, and (e)(f) the $\alpha(T)$ and $\beta(T)$ curves. The graphs on the left (right) correspond to $N_e = N_v = 1$ ($N_e = N_v = 4$).

Validation of the Analytic Expression of FRR and FAR

For both db1 and db2, we analytically estimate the genuine $\phi_{ge}(k)$ and imposter $\phi_{im}(k)$ Hamming distance pmfs, and the $\beta(T)$ and $\alpha(T)$ curves. The results are presented in Figure 3.16 and Figure 3.17 for db1 and db2, respectively. The experimentally calcu-



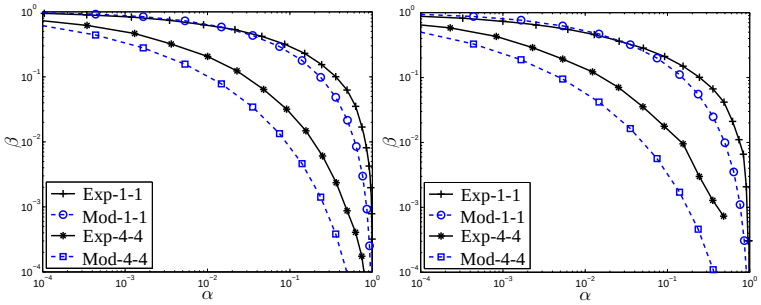
(a) P_e estimation (b) P_e estimation

(c) $\phi_{ge}(k)$ and $\phi_{im}(k)$ (d) $\phi_{ge}(k)$ and $\phi_{im}(k)$

(e) $\alpha(T)$ and $\beta(T)$ (f) $\alpha(T)$ and $\beta(T)$

Figure 3.17: Results for db2 with $N_F = 31$, (a)(b) $\hat{P}_e^{db2}$ and the analytical estimation of P_e^{ge}, (c)(d) $\phi_{ge}(k)$ and $\phi_{im}(k)$ pmfs, and (e)(f) the $\alpha(T)$ and $\beta(T)$ curves. The graphs on the left (right) correspond to $N_e = N_v = 1$ ($N_e = N_v = 4$).

lated pmfs are indicated by 'Exp' while the ones obtained using the analytical model are indicated by 'Mod'. The experimental results are obtained using the same protocol as the one discussed in Section 3.2.5, but storing the Hamming distance pmfs of each subject instead. We focus on the cases corresponding to $N_F = 31$ with $N_e = N_v = 1$ and $N_e = N_v = 4$.



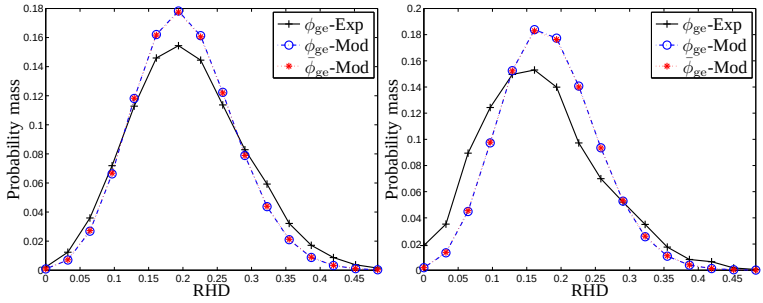
(a) db1 with $N_{\mathrm{F}} = 31$ (b) db2 with $N_{\mathrm{F}} = 31$

Figure 3.18: DET curves for both db1 and db2 for $N_{\mathrm{F}} = 31$ with different values of N_{e}, and N_{v}. The values N_{e} and N_{v} are indicated in the legend in the subsequent order. The experimentally obtained curves are denoted by 'Exp' while the analytical by 'Mod'.

Both Figure 3.16 and Figure 3.17 indicate that there is a good agreement between $\phi_{\mathrm{im}}(k)$-Exp and $\phi_{\mathrm{im}}(k)$-Mod. Large differences are observed between $\phi_{\mathrm{ge}}(k)$-Exp and $\phi_{\mathrm{ge}}(k)$-Mod. However, the differences decrease when both N_{e} and N_{v} are increased. Averaging multiple independent samples leads to a higher Gaussianity degree in accordance with the central limit theorem. This effect was also observed for the P_{e} estimation results in previous section. It is interesting to note the differences between the estimation errors of $\phi_{\mathrm{ge}}(k)$ of db1 and db2. For db1 the center of gravity of $\phi_{\mathrm{ge}}(k)$-Exp and $\phi_{\mathrm{ge}}(k)$-Mod practically coincide. The only difference is the width of the pmfs, since the experimentally obtained pmf is wider than the theoretical one. In case of db2, we see that there is both an alignment and a width error, $\phi_{\mathrm{ge}}(k)$-Exp is skewed to the left.

Eventually, we are interested in estimating the DET curves. Because the DET curves combine both β and α, they are thus prone to estimation errors associated with β or α. The DET curves for db1 and db2 for $N_{\mathrm{F}} = 31$ with different values of N_{e} and N_{v} are shown in Figure 3.18. From these figures we can conclude that increasing N_{e} and N_{v} leads to greater estimation errors of the DET curve, which contradicts the previous finding that increasing N_{e} and N_{v} leads to better estimations of P_{e} and $\phi_{\mathrm{ge}}(k)$. This can be explained by the fact that in the $N_{\mathrm{e}} = N_{\mathrm{v}} = 4$ case, the area of interest with $\beta(T) \in [0.01, 0.1]$ occurs for smaller values of $\alpha(T)$, because the number of bit errors decreases when N_{e} and N_{v} increase, i.e. the performance improves. As shown by the $\alpha(T)$ curves in Figure 3.16 and Figure 3.17, there is a greater estimation error at smaller values of $\alpha(T)$ thus amplifying the estimation error of the DET curve.

A summary of the probable causes for the observed differences, starting from the most probable, are (i) the non-homogeneous within-class variance (ii) the dependency between features, and (iii) the dependency between bit errors. Database db2 seems to be clearly not adhering to the homogeneous within-class variance assumption, resulting into a skewed $\phi_{\mathrm{ge}}(k)$ with a large tail. Such a tail is caused by subjects that have on average a worse performance than the other subjects. These subjects have many feature components with a larger within-class variance leading to larger $P_{\mathrm{e}}[j]$ values and thus



(a) db1 (b) db2

Figure 3.19: The approximation of the genuine Hamming distance pmf as binomial with $\bar{P}_{\mathrm{e}}$ ((3.28)) for the $N_{\mathrm{e}} = N_{\mathrm{v}} = 4$ case with $N_{\mathrm{F}} = 31$.

greater Hamming distances. In the literature these subjects are referred to as goats [110, 111]. If the features are dependent, then the Hamming distance pmf becomes wider while keeping its original mean. This effect is visible for both $\phi_{\mathrm{ge}}(k)$ and $\phi_{\mathrm{im}}(k)$ for both databases. On the other hand, certain disturbances such as occluded biometric images or strong biometric variabilities can cause multiple errors to occur simultaneously. Thus, the bit errors are dependent causing the tails on the right side of the genuine Hamming distance pmf. A right tail is slightly visible for db1, but is clearly present for db2 as illustrated in Figures 3.16(c)(d) and Figures 3.17(c)(d), respectively.

In Section 3.2.6 we propose a modified model that incorporates the non-homogeneous within-class variance property, while in Section 3.2.7 we further extend the model to include dependencies.

3.2.6 Relaxing the Homogeneous Within-Class Variance Assumption

In this section we propose a modified model that takes the non-homogeneous property into account, while still assuming independent feature components. The proposed method makes use of the approximation of the convolution of (3.2) with the binomial pmf. For the genuine case, this would be

$$\bar{\phi}_{\mathrm{ge}}(k) \quad = \binom{N_{\mathrm{F}}}{k} (\bar{P}_{\mathrm{e}}^{\mathrm{ge}})^k (1 - \bar{P}_{\mathrm{e}}^{\mathrm{ge}})^{N_{\mathrm{F}}-k}, \tag{3.28}$$

where $\bar{P}_{\mathrm{e}}^{\mathrm{ge}}$ is the average bit-error probability across the feature components $\bar{P}_{\mathrm{e}}^{\mathrm{ge}} = 1/N_{\mathrm{F}} \sum_{j=1}^{N_{\mathrm{F}}} P_{\mathrm{e}}^{\mathrm{ge}}[j]$. The approximate pmfs $\bar{\phi}_{\mathrm{ge}}(k)$ are depicted in Figure 3.19(a) for db1 and Figure 3.19(b) for db2 for the $N_{\mathrm{e}} = N_{\mathrm{v}} = 4$ case with $N_{\mathrm{F}} = 31$. For both databases, the approximation is reasonably accurate.

Thus we can model the non-homogeneous effect by assuming that $\bar{P}_{\mathrm{e},i}^{\mathrm{ge}}$ is not equal for each subject and is distributed according to a probability density $p_{\bar{P}_{\mathrm{e}}^{\mathrm{ge}}}$. The following step consists in determining the pdf $p_{\bar{P}_{\mathrm{e}}^{\mathrm{ge}}}$ across the population and computing the average

genuine Hamming distance pmf defined as

$$\bar{\Phi}_{\mathrm{ge}}(k) \quad = \int_0^{1/2} p_{\bar{P}_{\mathrm{e}}^{\mathrm{ge}}}(\tau)\bar{\phi}_{\mathrm{ge}}(k|\tau)\mathrm{d}\tau, \tag{3.29}$$

where the integral limits are due to the fact that $P_{\mathrm{e}} \in [0, 1/2]$ and $\bar{\phi}_{\mathrm{ge}}(k|\tau)$ is the generic case of (3.28) as

$$\bar{\phi}_{\mathrm{ge}}(k|\tau) \quad = \binom{N_{\mathrm{F}}}{k}(\tau)^k(1-\tau)^{N_{\mathrm{F}}-k}. \tag{3.30}$$

We propose a method for estimating $p_{\bar{P}_{\mathrm{e}}^{\mathrm{ge}}}$ using only the estimated within-class variance of each subject $\hat{\sigma}^2_{\mathrm{w},i}[j]$. Because of the limited number of samples N_i, we know from [108] that the estimation ratio $((N_i-1)\hat{\sigma}^2_{\mathrm{w},i}[j])/\sigma^2_{\mathrm{w}}[j]$ follows the χ^2 distribution with $N_i - 1$ degrees of freedom, where $\sigma^2_{\mathrm{w}}[j]$ is the underlying within-class variance that has to be estimated and is assumed to be homogeneous. However, in practice $\sigma^2_{\mathrm{w}}[j]$ is unknown, therefore we have to replace it by its estimate $\hat{\sigma}^2_{\mathrm{w}}[j]$. It is well known that the mean associated with a χ^2 distribution is equal to its number of degrees of freedom, thus by omitting the $(N_i - 1)$ multiplications it becomes unit mean.

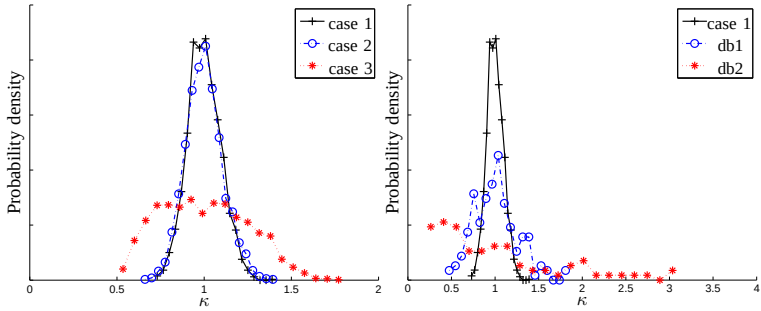
The next step is to take the average ratio over all feature components as

$$\kappa_i \quad = \tfrac{1}{N_{\mathrm{F}}} \sum_{j=1}^{N_{\mathrm{F}}} \hat{\sigma}^2_{\mathrm{w},i}[j]/\hat{\sigma}^2_{\mathrm{w}}[j]. \tag{3.31}$$

We can model the non-homogeneous property by assuming that for all components of subject i the within-class variance is $\sigma^2_{\mathrm{w},i}[j] = \kappa_i\sigma^2_{\mathrm{w}}[j]$. If the homogeneous assumption holds and the number of features is large, then the pdf of κ_i across the whole population becomes Gaussian with unit mean and a variance that decreases when N_{F} increases. The variance decreases at larger values of N_{F} because this would be similar to having N_{F} times more samples and therefore a better estimation of its mean. When there are "goat-like" subjects, the homogeneous assumption does not hold, then the variance of the pdf of κ_i increases.

Figure 3.20(a) shows the empirically estimated pdf of κ_i for a synthetically generated databases containing 2000 subjects with $N_{\mathrm{F}} = 31$, $N_i = 8$, and $\sigma^2_{\mathrm{b}}[j] = 1$, where for 'case 1' every subject has the same $\sigma^2_{\mathrm{w},i}[j] = 1$, in 'case 2' $\sigma^2_{\mathrm{w},i}[j] = 1 + \nu_i[j]$, and for 'case 3' $\sigma^2_{\mathrm{w},i}[j] = 1 + \nu_i$ where ν_i is drawn from U(-0.4,0.4) and is redrawn for each feature component separately in 'case 2'. The results imply that the variance of the κ_i pdf increases when $\sigma^2_{\mathrm{w},i}[j]$ is different for each subject ('case 2') and increases significantly when there is a positive correlation with the variance offset, for example when subjects have all their $\sigma^2_{\mathrm{w},i}[j]$ larger or smaller than the average value ('case 3'). Hence, in 'case 3' there is a clear existence of goats or doves, where the latter are the subjects that have a small number of bit errors when matched against themselves [112].

Figure 3.20(b) compares the κ_i pdf of 'case 1', db1, and db2. The results show that both db1 and db2 do not adhere to the homogeneous property. The κ_i pdf found for db1 looks similar to 'case 3'. However, the pdf found for db2 significantly deviates from the synthetic cases, which confirms the existence of goats and doves. This may also explain the significant discrepancy found when estimating the genuine Hamming distance pmfs of db2 as shown in Figure 3.17.



(a) Synthetic cases (b) db1 & db2

Figure 3.20: Empirical estimated probability density p_{κ_i} using synthetic databases (a) of 2000 subjects with $N_{\rm F} = 31, N_i = 8, \sigma_{\rm b}^2[j] = 1$, where for 'case 1' every subject has the same $\sigma_{{\rm w},i}^2[j] = 1$, in 'case 2' $\sigma_{{\rm w},i}^2[j] = 1 + \nu_i[j]$, and for 'case 3' $\sigma_{{\rm w},i}^2[j] = 1 + \nu_i$ where ν_i is drawn from U(-0.4,0.4) and is redrawn for each feature component separately in 'case 2'. In (b) the comparison between 'case 1', db1, and db2 is shown.

Now we can empirically estimate the probability density $p_{\bar{P}_{\rm e}^{\rm ge}}$ using p_{κ_i}. The relationship between κ_i and $\bar{P}_{{\rm e},i}^{\rm ge}$ is given by

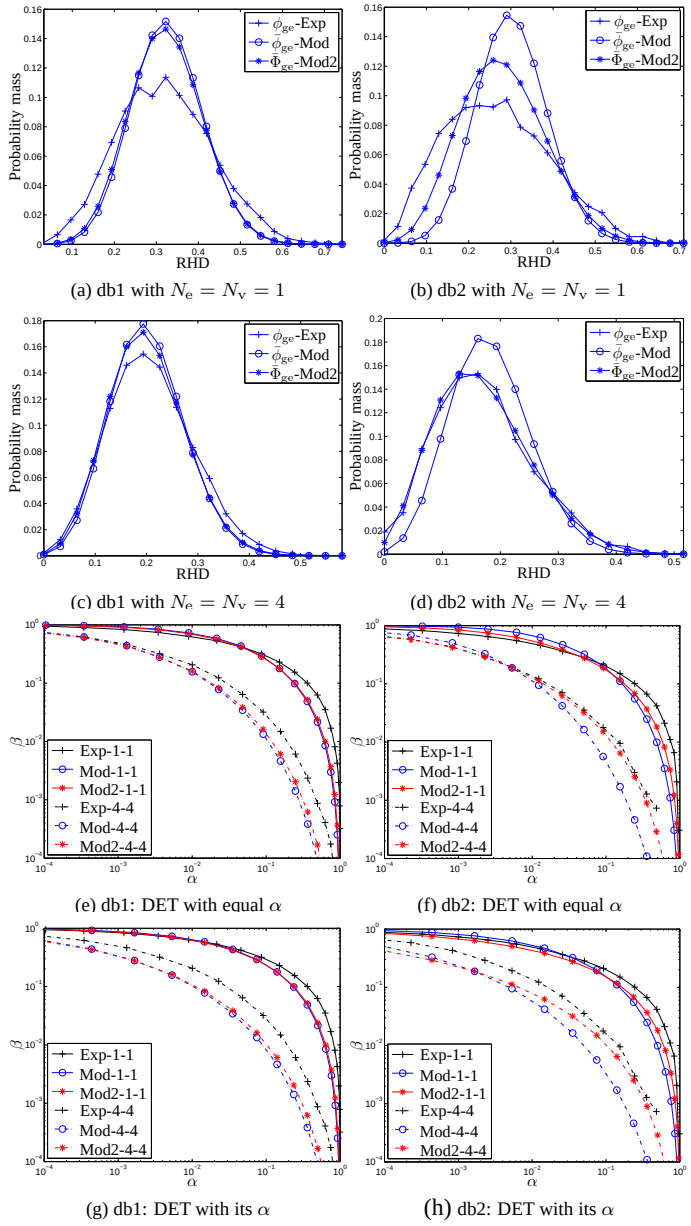
$$\bar{P}_{{\rm e},i}^{\rm ge} \quad = \frac{1}{N_{\rm F}} \sum_{j=1}^{N_{\rm F}} P_{\rm e}^{\rm ge}(N_{\rm e}, N_{\rm v}, \sqrt{\kappa_i \hat{\sigma}_{\rm w}^2[j]}, \hat{\sigma}_{\rm b}[j]), \tag{3.32}$$

where we take the average of $P_{\rm e}^{\rm ge}[j]$ across all features, while using $\hat{\sigma}_{\rm b}[j]$ and the modified within-class variance estimation $\sqrt{\kappa_i \hat{\sigma}_{\rm w}^2[j]}$. Because of the nonlinear relationship between $P_{\rm e}^{\rm ge}[j]$ and $\hat{\sigma}_{\rm w}[j]$ we take the average over $P_{\rm e}^{\rm ge}[j]$ instead of estimating $P_{\rm e}^{\rm ge}$ using the average of $\hat{\sigma}_{\rm w}[j]$.

In practice, we can rewrite (3.29) as:

$$\bar{\Phi}_{\rm ge}(k) \quad = \frac{1}{N_{\rm s}} \sum_{i=1}^{N_{\rm s}} \bar{\phi}_{\rm ge}(k|\bar{P}_{{\rm e},i}^{\rm ge}). \tag{3.33}$$

We applied this new method for estimating $\phi_{\rm ge}(k)$ of db1 and db2 and the results are shown in Figures 3.21(a-d) for the $N_{\rm e} = N_{\rm v} = 1$ and $N_{\rm e} = N_{\rm v} = 4$ cases with $N_{\rm F} = 31$, where $\phi_{\rm ge}(k)$-Exp is the experimentally obtained pmf, $\phi_{\rm ge}(k)$-Mod is obtained using (3.2), and $\bar{\Phi}_{\rm ge}(k)$-Mod2 with (3.33). The results show that $\phi_{\rm ge}$-Exp is better approximated when using the new method $\bar{\Phi}_{\rm ge}(k)$-Mod2. In case of db1 there is a small improvement, but for db2 there is a significant improvement and even a better estimation is obtained when $N_{\rm e} = N_{\rm v} = 4$. Furthermore, Figures 3.21(e-h) show the DET curve results. In Figures 3.21(e)(f) the same α is used for each DET curve in order to isolate the estimation errors of $\phi_{\rm ge}(k)$, while in Figures 3.21(g)(h) α-Exp is used for the 'Exp' curves and α-Mod is used for both the 'Mod' and 'Mod2' curves. With the new method the DET curve estimation has improved, most significantly for db2. However, the differences between Figures 3.21(e)(f) and Figures 3.21(g)(h) clearly indicate



(a) db1 with $N_e = N_v = 1$ (b) db2 with $N_e = N_v = 1$

(c) db1 with $N_e = N_v = 4$ (d) db2 with $N_e = N_v = 4$

(e) db1: DET with equal α (f) db2: DET with equal α

(g) db1: DET with its α (h) db2: DET with its α

Figure 3.21: Results of the proposed method incorporating the non-homogeneous property of db1 and db2 for the cases $N_e = N_v = 1$ and $N_e = N_v = 4$ with $N_F = 31$. Figures (a-d) show the Hamming distance pmf estimations while figures (e-h) show the DET curves estimation, where 'Mod' and 'Mod2' indicate the modeling method without and with the non-homogeneous property, respectively. In (e) and (f) all the DET curves are plotted using the experimentally obtained α-Exp, while in (g) and (h) we use the α-Exp for the 'Exp' curves and α-Mod for both the 'Mod' and 'Mod2' curves.

that the remaining estimation errors are caused by the estimation of α. As shown in Figures 3.16(c)(d) and Figures 3.17(c)(d) there is an estimation error of ϕ_{im}, which we consider to be caused by the fact that the feature components are dependent.

3.2.7 Incorporating Feature Component Dependencies

In previous section we observed that a significant part of the remaining DET estimation errors is related to the estimation errors of the ϕ_{im}-Exp pmf. In this section we propose a further extension of the analytical framework in order to incorporate dependencies between feature components. We propose to estimate the dependency from the ϕ_{im} pmf and apply it to the ϕ_{ge} pmf estimation. Hence, we assume that both pmfs are influenced by the dependency to the same extent.

We estimate the dependency from ϕ_{im}-Exp by fitting it with a Gaussian approximation of the binomial pmf of (3.9) with the variance as the fitting parameter. For large values of N_{F}, the binomial pmf with probability P_{e} and dimension N_{F} can be approximated by the Gaussian density $\mathcal{N}(N_{\mathrm{F}}P_{\mathrm{e}}, N_{\mathrm{F}}P_{\mathrm{e}}(1-P_{\mathrm{e}}))$, with mean $N_{\mathrm{F}}P_{\mathrm{e}}$ and variance $N_{\mathrm{F}}P_{\mathrm{e}}(1-P_{\mathrm{e}})$. For the imposter case we know that $P_{\mathrm{e}} = 1/2$, from which its mean and variance become $N_{\mathrm{F}}/2$ and $N_{\mathrm{F}}/4$, respectively. Hence, the Gaussian approximation of the ϕ_{im}-Exp pmf with the variance parameter ϑ used for fitting becomes

$$\begin{aligned} \phi_{\mathrm{im}}(k)\text{-Mod-}\vartheta &= \tfrac{1}{\sqrt{2\pi\vartheta\sigma^2}}\mathrm{e}^{-\frac{(k-\mu)^2}{2\vartheta\sigma^2}} \\ &= \tfrac{1}{\sqrt{2\pi\vartheta N_{\mathrm{F}}P_{\mathrm{e}}(1-P_{\mathrm{e}})}}\mathrm{e}^{-\frac{(k-N_{\mathrm{F}}P_{\mathrm{e}})^2}{2\vartheta N_{\mathrm{F}}P_{\mathrm{e}}(1-P_{\mathrm{e}})}} \\ &= \tfrac{2}{\sqrt{2\pi\vartheta N_{\mathrm{F}}}}\mathrm{e}^{-\frac{(2k-N_{\mathrm{F}})^2}{2\vartheta N_{\mathrm{F}}}}, \end{aligned} \tag{3.34}$$

where the optimal ϑ is computed by minimizing the mean-square error (MMSE) as

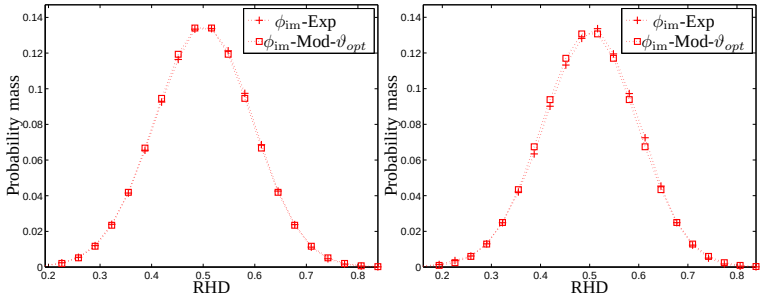
$$\vartheta_{opt} = \arg\min_{\vartheta} \sum_{k=0}^{N_{\mathrm{F}}} \Big(\phi_{\mathrm{im}}(k)\text{-Exp} - \phi_{\mathrm{im}}(k)\text{-Mod-}\vartheta\Big)^2. \tag{3.35}$$

The estimation results of ϑ_{opt} for the $N_{\mathrm{e}} = N_{\mathrm{v}} = 1$ case are shown in Figure 3.22 for both databases. The optimal value of ϑ_{opt} is 1.11 for db1 and 1.17 for db2. For both databases ϑ_{opt} is very similar, which may indicate that the amount of dependencies between the feature components is relative similar for both databases. Furthermore, the ϕ_{im}-Exp pmf is better estimated when compared to its first estimation disregarding the feature component dependencies as depicted in Figure 3.16(c) and Figure 3.17(c) for db1 and db2, respectively.

With the Gaussian approximation including the variance correction with ϑ_{opt} we have a better estimation of the ϕ_{ge} pmf by rewriting (3.33) as

$$\bar{\Phi}_{\mathrm{ge}}(k) = \tfrac{1}{N_{\mathrm{s}}} \sum_{i=1}^{N_{\mathrm{s}}} \tfrac{1}{\sqrt{2\pi\sigma_{cor}^2}} \mathrm{e}^{\frac{-\left(k-\bar{P}_{\mathrm{e},i}^{\mathrm{ge}}N_{\mathrm{F}}\right)^2}{2\sigma_{cor}^2}}, \tag{3.36}$$

with $\sigma_{cor}^2 = \vartheta_{opt}N_{\mathrm{F}}\bar{P}_{\mathrm{e},i}^{\mathrm{ge}}(1-\bar{P}_{\mathrm{e},i}^{\mathrm{ge}})$. Because of the Gaussian approximation errors it does not hold that the sum of the probability mass equals to one, therefore we normalize it



(a) db1 $\vartheta_{opt} = 1.11$ (b) db2 $\vartheta_{opt} = 1.17$

Figure 3.22: Results of estimating ϑ_{opt} from ϕ_{im}-Exp using (3.35) for the $N_{\mathrm{e}} = N_{\mathrm{v}} = 1$ case for both databases. The variance corrected Gaussian approximated curve as described by (3.34) is depicted as ϕ_{im}-Mod-ϑ.

according to

$$\bar{\Phi}'_{\mathrm{ge}}(k) = \frac{1}{\sum_{k=0}^{N_{\mathrm{F}}} \bar{\Phi}_{\mathrm{ge}}(k)} \bar{\Phi}_{\mathrm{ge}}(k). \tag{3.37}$$

The estimation results using (3.37) for the cases of $\vartheta = 1$ and $\vartheta = \vartheta_{opt}$ are depicted in Figure 3.23. For the $\vartheta = 1$ case the Gaussian approximation is used without the variance correction. Figures 3.23(a-d) show that the $\phi_{\mathrm{ge}}(k)$ pmf estimation has slightly improved. The $\bar{\Phi}'_{\mathrm{ge}}$-Mod-$\vartheta_{opt}$ curve is closer to $\phi_{\mathrm{ge}}(k)$-Exp than $\bar{\Phi}'_{\mathrm{ge}}$-Mod-$\vartheta_1$. This holds across the whole curve for the $N_{\mathrm{e}} = N_{\mathrm{v}} = 1$ case and mainly for the right tail for the $N_{\mathrm{e}} = N_{\mathrm{v}} = 4$ case. The same conclusions are also portrayed by the DET curves of Figures 3.23(e-f), where each DET curve uses the same α curve, namely the experimentally obtained α-Exp in order to isolate the ϕ_{ge}(k) pmf estimation errors. The DET curves in Figures 3.23(g-h) use the actual α curves, thus α-Mod-ϑ_1 for the DET-Mod-ϑ_1 curves and α-Mod-ϑ_{opt} for the DET-Mod-ϑ_{opt} curves, respectively. The curves show that the DET-Mod-ϑ_{opt} curve is clearly closer to DET-Exp curve, because α-Mod-ϑ_{opt} is a better approximation of α-Exp as we have shown earlier.

3.2.8 Practical Considerations

In previous sections we have presented several analytical models for estimating the DET performance curve. However, as stated previously, because of the use of an ECC the FRR is lower bounded because of the limited number of bits the ECC can correct. For the setting of $N_{\mathrm{F}} = 31$, which equals the codeword length n_{c}, the BCH ECC can correct up to 7 bits as shown in Table 3.1. The experimentally achieved performance and its analytical estimates at this operating point are given in Table 3.5. The results indicate that at this operating point there is not a significant difference between the estimations using the 'Mod' and 'Mod2' models, while the 'Mod-ϑ_{opt}' estimator leads to the best estimation where its significant improvement is of the α.

Although we have presented an analytical framework for analysis, it could also be used in practical cases. For example, consider the scenario where a database has been collected with a maximum of five samples per subject. Hence, the performance could only be calculated for cases where $N_{\mathrm{e}} + N_{\mathrm{v}} \leq 5$. However, this restriction does not hold for our proposed analytical framework. By estimating σ_{w}^2, σ_{b}^2, κ_i, and ϑ_{opt} from the given database, the performance could be estimated for the cases where $N_{\mathrm{e}} + N_{\mathrm{v}} \geq 5$. Either the performance could be estimated for a specific N_{e} and N_{v} setting or the lower bounds of the N_{e} and N_{v} setting could be estimated in order to obtain a certain performance or better. Given the same scenario as for Table 3.5 where the performance is estimated at the maximum error capability of the ECC for both databases, db1 is expected to reach $\beta \leq 0.1$ when $N_{\mathrm{e}} = N_{\mathrm{v}} \geq 8$, while $N_{\mathrm{e}} = N_{\mathrm{v}} \geq 7$ for db2.

3.2.9 Conclusions

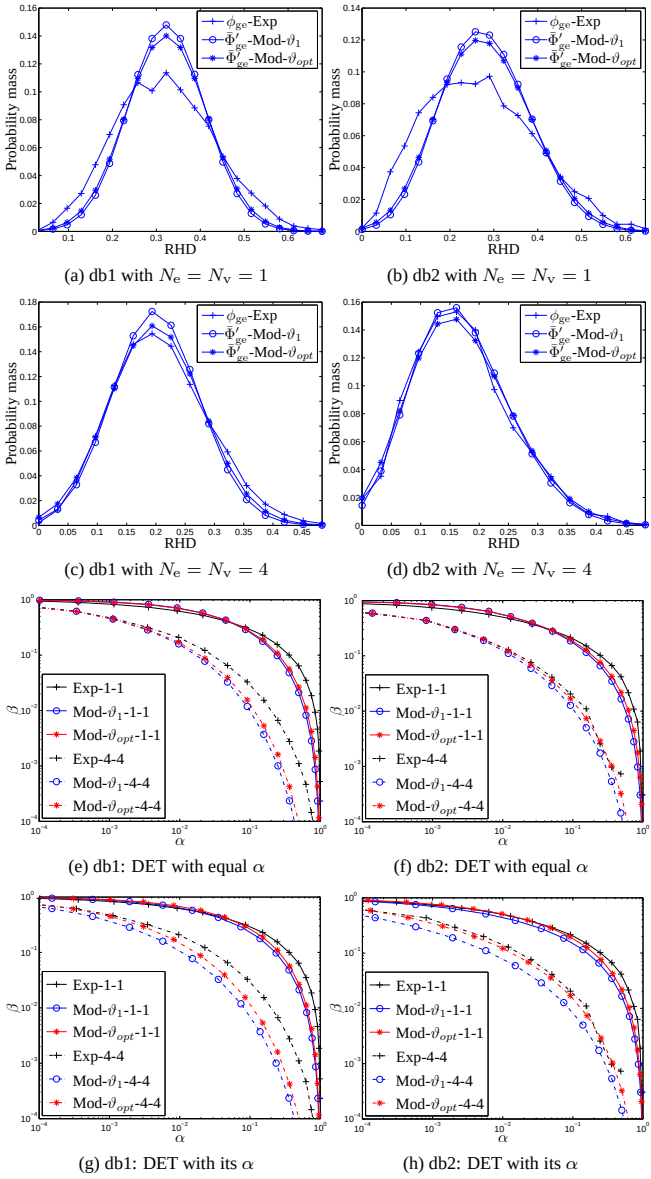
We have proposed an analytical framework for estimating the DET performance curve of a biometric system, based on binary feature vectors, for different settings of N_{e} and N_{v}.

The first proposed estimation method used a simple Parallel Gaussian Channel framework for modeling the pdf of the real-valued features. Each component has its own channel with the corresponding additive Gaussian noise representing the biometric variability and measurement noise, called the within-class variability. The results showed significant estimation errors and were far from optimal, mainly because of the homogeneous within-class variance assumption. Consequently we proposed a modified framework to incorporate the non-homogeneous property, which in fact assumes that the within-class variance is different for each subject. The estimation improved significantly and the remaining estimation error is thought to be caused by the estimation errors of the false acceptance curve due to dependency between feature components and corresponding bits.

Table 3.5: The experimentally ('Exp') achieved α and β and its analytical estimates using the simplistic model ('Mod'), the model relaxing the homogeneous property ('Mod2'), and the model also incorporating the feature component dependencies ('Mod-ϑ_{opt}').

	db1			
	$N_{\mathrm{e}} = N_{\mathrm{v}} = 1$		$N_{\mathrm{e}} = N_{\mathrm{v}} = 4$	
	α	β	α	β
Exp	$3.59 \cdot 10^{-3}$	$7.33 \cdot 10^{-1}$	$3.73 \cdot 10^{-3}$	$3.17 \cdot 10^{-1}$
Mod	$1.66 \cdot 10^{-3}$	$8.43 \cdot 10^{-1}$	$1.66 \cdot 10^{-3}$	$2.79 \cdot 10^{-1}$
Mod2	$1.66 \cdot 10^{-3}$	$8.25 \cdot 10^{-1}$	$1.66 \cdot 10^{-3}$	$2.77 \cdot 10^{-1}$
Mod-ϑ_{opt}	$3.06 \cdot 10^{-3}$	$8.15 \cdot 10^{-1}$	$3.06 \cdot 10^{-3}$	$2.94 \cdot 10^{-1}$
	db2			
	$N_{\mathrm{e}} = N_{\mathrm{v}} = 1$		$N_{\mathrm{e}} = N_{\mathrm{v}} = 4$	
	α	β	α	β
Exp	$6.35 \cdot 10^{-3}$	$5.58 \cdot 10^{-1}$	$5.28 \cdot 10^{-3}$	$1.96 \cdot 10^{-1}$
Mod	$1.66 \cdot 10^{-3}$	$7.66 \cdot 10^{-1}$	$1.66 \cdot 10^{-3}$	$1.88 \cdot 10^{-1}$
Mod2	$1.66 \cdot 10^{-3}$	$6.31 \cdot 10^{-1}$	$1.66 \cdot 10^{-3}$	$1.88 \cdot 10^{-1}$
Mod-ϑ_{opt}	$3.80 \cdot 10^{-3}$	$6.31 \cdot 10^{-1}$	$3.94 \cdot 10^{-3}$	$2.01 \cdot 10^{-1}$

The final proposed framework also incorporated feature component dependency, whose value was derived from the calculated imposter Hamming distance pmf of the database. This method resulted in the most optimum estimation of the DET performance curves.



(a) db1 with $N_{\mathrm{e}} = N_{\mathrm{v}} = 1$ (b) db2 with $N_{\mathrm{e}} = N_{\mathrm{v}} = 1$

(c) db1 with $N_{\mathrm{e}} = N_{\mathrm{v}} = 4$ (d) db2 with $N_{\mathrm{e}} = N_{\mathrm{v}} = 4$

(e) db1: DET with equal α (f) db2: DET with equal α

(g) db1: DET with its α (h) db2: DET with its α

Figure 3.23: Results of the proposed method incorporating both the dependency and non-homogeneous property of db1 and db2 for the cases $N_{\mathrm{e}} = N_{\mathrm{v}} = 1$ and $N_{\mathrm{e}} = N_{\mathrm{v}} = 4$ with $N_{\mathrm{F}} = 31$. Figures (a-d) show the ϕ_{ge} estimations, while (e-h) show the DET curves estimation. The label 'Mod-ϑ_1' indicates the new modeling method but with $\vartheta = 1$, hence using only the Gaussian approximation of the binomial pmf including the non-homogeneous property. The label 'Mod-ϑ_{opt}' indicates the cases where $\vartheta = \vartheta_{opt}$. In (e) and (f) all the DET curves are plotted using the experimentally obtained α-Exp, while in (g) and (h) we use the α-Exp for the 'Exp' curves, α-Mod-ϑ_1 for the 'Mod-ϑ_1' curves and α-Mod-ϑ_{opt} for the 'Mod-ϑ_{opt}' curves.

3.3 Classification Performance Comparison of a Continuous and Binary Classifier under Gaussian Assumption

3.3.1 Abstract

Template protection techniques are privacy and security enhancing techniques of biometric reference data within a biometric system. Several of the template protection schemes known in the literature require the extraction of a binary representation from the real-valued biometric sample, which raises the question whether the bit extraction method reduces the classification performance. In this work we provide the theoretical performance of the optimal log likelihood ratio continuous classifier and compare it with the theoretical performance of a binary Hamming distance classifier with a single bit extraction scheme as known from the literature. We assume biometric data modeled by a Gaussian between-class and within-class probability density with independent feature components and we also include the effect of averaging multiple enrolment and verification samples.

3.3.2 Introduction

The introduction of the ePassport with fingerprint raised some question marks on the privacy of the subjects and the security of the stored biometric data, especially when the Dutch government decided to store the fingerprint samples in a centralized database [113]. The security and privacy risks related to the storage of biometric data are (i) *identity fraud* where an adversary steals the stored reference template and impersonates the genuine subject of the system by some spoofing mechanism, (ii) *limited-renewability* implying the limited capability to renew a compromised reference template due to the limited number of biometric instances (for example we only have ten fingers, two irises or retinas, and a single face), (iii) *cross-matching* or linking reference templates of the same subject across databases of different applications, and (iv) derivation of *sensitive medical information* where it is known that biometric data may reveal the presence of certain diseases.

The field of template protection aims at mitigating these privacy and security risks by developing techniques that provide (i) *irreversibility* implying that it is impossible or at least very difficult to retrieve the original biometric sample from the reference template, (ii) *renewability* where it is possible to renew the reference template when necessary, and (iii) *unlinkability* which prevents cross-matching. In the literature, numerous template protection methods such as the *Fuzzy Commitment Scheme* (FCS) [36], *Helper Data System* (HDS) [33, 34, 48], *Fuzzy Extractors* [64, 65], *Fuzzy Vault* [80, 84] and *Cancelable Biometrics* [59] have been proposed.

In general, the extracted feature vector from the biometric sample is real-valued, while several of the proposed template protection schemes depend on the extraction of a binary representation from the biometric sample. The classification performance of the template protection scheme thus depends on the combination of the bit extraction process and the binary classifier. Yet, an unanswered question is what the difference is between the theoretical classification performance at binary level (after the bit extraction) and the per-

formance at the continuous level (before the bit extraction). A potential performance loss after the bit extraction process may represent the penalty for the requirement to extract a binary representation from the biometric sample. In [95], the performance of a single bit extraction process with a Hamming distance classifier has been theoretically determined under the assumption that the biometric data is Gaussian distributed. In this work we first discuss the theoretical performance of the optimal likelihood-ratio continuous classifier, under the assumption that the biometric data is Gaussian distributed. In [37], the theoretical performance has been derived where the reference template is the average of N_{e} enrolment samples with a single verification sample. We extend this analysis by including the averaging of N_{v} verification samples. Lastly, we compare the theoretical performance difference between the continuous and binary classifier and study the influence of the number of feature components and the number of enrolment and verification samples.

The outline of this paper is as follows. In Section 3.3.3 we briefly describe the model of the biometric data under Gaussian assumption including the averaging of multiple enrolment and verification samples. The theoretical performance estimation for the continuous classifier is derived in Section 3.3.4 and Section 3.3.5 briefly describes the theoretical performance for the binary classifier known from the literature. The theoretical performance comparison between the two classifiers and the effect of averaging multiple enrolment and verification samples is studied in Section 3.3.6. We conclude with our final remarks in Section 3.3.7.

3.3.3 Preliminaries

Random variables are underlined. Let $\underline{x}_i \simeq N(\underline{\mu}_{\mathrm{e}}, \sigma_{\mathrm{w}}^2)$, $i = 1, \ldots, N_{\mathrm{e}}$ denote the enrolment samples (features, in fact) and $\underline{y}_i \simeq N(\underline{\mu}_{\mathrm{v}}, \sigma_{\mathrm{w}}^2)$, $i = 1, \ldots, N_{\mathrm{v}}$ the verification samples with σ_{w}^2 being the within-class variance. We assume that for a given class mean μ the samples drawn from that class are i.i.d. The enrolment and verification class means are also Gaussian random variables, in particular $\underline{\mu}_{\mathrm{e}}, \underline{\mu}_{\mathrm{v}} \simeq N(0, \sigma_{\mathrm{b}}^2)$ with σ_{b}^2 being the between-class variance. The reference template $\underline{r}$ and the verification template $\underline{v}$ are sample means, i.e.

$$\underline{r} = \frac{1}{N_{\mathrm{e}}} \sum_{i=1}^{N_{\mathrm{e}}} \underline{x}_i \tag{3.38}$$

$$\underline{v} = \frac{1}{N_{\mathrm{v}}} \sum_{i=1}^{N_{\mathrm{v}}} \underline{y}_i. \tag{3.39}$$

Because the samples are assumed to be independent we obtain $\underline{r} \simeq N(\underline{\mu}_{\mathrm{e}}, \frac{\sigma_{\mathrm{w}}^2}{N_{\mathrm{e}}})$ and $\underline{v} \simeq N(\underline{\mu}_{\mathrm{v}}, \frac{\sigma_{\mathrm{w}}^2}{N_{\mathrm{v}}})$.

In the genuine case, the features originate from the same, unknown, mean, i.e. $\underline{\mu}_{\mathrm{e}} = \underline{\mu}_{\mathrm{v}} = \underline{\mu}$. In the impostor case the features originate form arbitrary means drawn from the between-class density. The purpose of the classifier is to discriminate between genuine and impostor comparisons.

3.3.4 Continuous Classifier Performance

The Log Likelihood Ratio Comparison Score

Let $p_{\underline{r},\underline{v}}(r,v|\mathrm{gen})$, $p_{\underline{r},\underline{v}}(r,v|\mathrm{imp})$ denote the joint probability densities of $\underline{r}$ and $\underline{v}$ in the genuine and impostor cases, respectively. The likelihood ratio in this case is defined by

$$l(r,v) = \frac{p_{\underline{r},\underline{v}}(r,v|\mathrm{gen})}{p_{\underline{r},\underline{v}}(r,v|\mathrm{imp})}. \tag{3.40}$$

We conveniently arrange $\underline{r}$ and $\underline{v}$ in a column vector $\mathbf{z} = (\underline{r},\underline{v})^{\mathrm{T}}$. We write

$$p_{\underline{r},\underline{v}|\mathrm{gen}}(r,v|\mathrm{gen}) = \frac{1}{2\pi\sqrt{|C_{\mathrm{gen}}|}}\mathrm{e}^{-\frac{\mathbf{z}^{\mathrm{T}}C_{\mathrm{gen}}^{-1}\mathbf{z}}{2}} \tag{3.41}$$

$$p_{\underline{r},\underline{v}|\mathrm{imp}}(r,v|\mathrm{imp}) = \frac{1}{2\pi\sqrt{|C_{\mathrm{imp}}|}}\mathrm{e}^{-\frac{\mathbf{z}^{\mathrm{T}}C_{\mathrm{imp}}^{-1}\mathbf{z}}{2}}, \tag{3.42}$$

where C_{gen} and C_{imp} are the co-variance matrices for the genuine and imposter comparisons, respectively. For $p_{\underline{r},\underline{v}|\mathrm{gen}}(r,v|\mathrm{gen})$, we can write

$$p_{\underline{r},\underline{v}|\mathrm{gen}}(r,v|\mathrm{gen}) = \int_{-\infty}^{\infty} p_{\underline{r}|\underline{\mu}}(r|\mu)p_{\underline{v}|\underline{\mu}}(v|\mu)p_{\underline{\mu}}(\mu)\mathrm{d}\mu. \tag{3.43}$$

Using this we obtain $E\{\underline{r}|\mathrm{gen}\} = E\{\underline{v}|\mathrm{gen}\} = 0$, $E\{\underline{r}^2|\mathrm{gen}\} = \sigma_{\mathrm{b}}^2 + \frac{1}{N_{\mathrm{e}}}\sigma_{\mathrm{w}}^2$, $E\{\underline{v}^2|\mathrm{gen}\} = \sigma_{\mathrm{b}}^2 + \frac{1}{N_{\mathrm{v}}}\sigma_{\mathrm{w}}^2$, and $E\{\underline{r}\underline{v}|\mathrm{gen}\} = \sigma_{\mathrm{b}}^2$, therefore,

$$C_{\mathrm{gen}} = \begin{pmatrix} \sigma_{\mathrm{b}}^2 + \frac{1}{N_{\mathrm{e}}}\sigma_{\mathrm{w}}^2 & \sigma_{\mathrm{b}}^2 \\ \sigma_{\mathrm{b}}^2 & \sigma_{\mathrm{b}}^2 + \frac{1}{N_{\mathrm{v}}}\sigma_{\mathrm{w}}^2 \end{pmatrix}. \tag{3.44}$$

In the impostor case, $\underline{r}$ and $\underline{v}$ are independent and

$$C_{\mathrm{imp}} = \begin{pmatrix} \sigma_{\mathrm{b}}^2 + \frac{1}{N_{\mathrm{e}}}\sigma_{\mathrm{w}}^2 & 0 \\ 0 & \sigma_{\mathrm{b}}^2 + \frac{1}{N_{\mathrm{v}}}\sigma_{\mathrm{w}}^2 \end{pmatrix}. \tag{3.45}$$

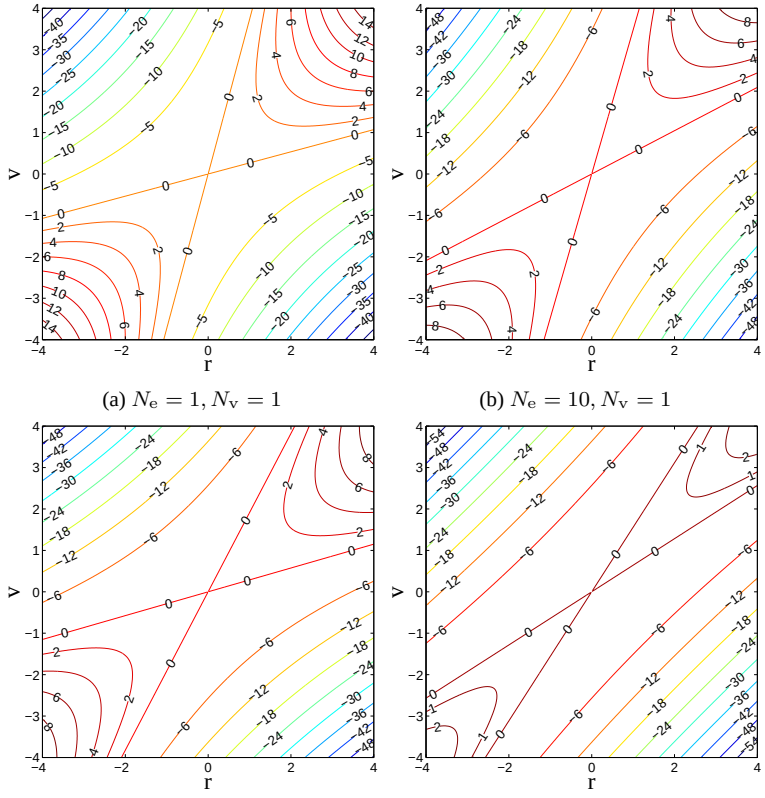
Instead of the likelihood ratio we compute a comparison score based on the log likelihood ratio, from which constant terms and factors have been removed:

$$s(r,v;N_{\mathrm{e}},N_{\mathrm{v}}) = -\mathbf{z}^{\mathrm{T}}C_{\mathrm{gen}}^{-1}\mathbf{z} + \mathbf{z}^{\mathrm{T}}C_{\mathrm{imp}}^{-1}\mathbf{z}. \tag{3.46}$$

On substitution of (3.44) and (3.45) into (3.46) and after simplification and elimination of constants we obtain the following expression for the comparison score

$$s(r,v;N_{\mathrm{e}},N_{\mathrm{v}}) = -\frac{r^2}{\sigma_{\mathrm{b}}^2 + \frac{1}{N_{\mathrm{e}}}\sigma_{\mathrm{w}}^2} - \frac{v^2}{\sigma_{\mathrm{b}}^2 + \frac{1}{N_{\mathrm{v}}}\sigma_{\mathrm{w}}^2} + 2\frac{rv}{\sigma_{\mathrm{b}}^2}, \tag{3.47}$$

in which we included the number of enrolment N_{e} and verification N_{v} samples as parameters. Examples of $s(r,v;N_{\mathrm{e}},N_{\mathrm{v}})$ are portrayed by contour plots in Figure 3.24 for



(a) $N_{\mathrm{e}} = 1, N_{\mathrm{v}} = 1$ (b) $N_{\mathrm{e}} = 10, N_{\mathrm{v}} = 1$

(c) $N_{\mathrm{e}} = 1, N_{\mathrm{v}} = 10$ (d) $N_{\mathrm{e}} = 10, N_{\mathrm{v}} = 10$

Figure 3.24: Contour plot of the log likelihood ratio comparison score $s(r, v; N_{\mathrm{e}}, N_{\mathrm{v}})$ from (3.47) with within-class and between class variance $\sigma_{\mathrm{w}}^2 = \sigma_{\mathrm{b}}^2 = 1$ for different number of enrolment N_{e} or verification N_{v} samples.

different number of enrolment N_{e} or verification N_{v} samples with within-class and between class variance $\sigma_{\mathrm{w}}^2 = \sigma_{\mathrm{b}}^2 = 1$. Positive comparisons scores are obtained when the $\{r, v\}$-pair is close the $r = v$-axis (the positive diagonal line) and being further away from the origin increases the comparison score. Negative comparisons scores are obtained when the $\{r, v\}$-pair is closer the $-r = v$-axis (the negative diagonal line) and increases when further away from the origin. Increasing both the number of enrolment and verification samples shifts the zero-contour lines closer to the $r = v$-axis, because the expected uncertainty has decreased due to the reduction of the within-class variance by averaging multiple samples. Hence, a similar behavior can be expected when decreasing the within-class variance directly. Increasing only the number of enrolment (verification) samples mainly shifts the horizontal (vertical) zero-contour line closer to the $r = v$-axis.

Comparison Score Density and the Classification Performance

In order to estimate the performance, first we to have to derive the density of the log likelihood comparison score $s(r, v; N_{\mathrm{e}}, N_{\mathrm{v}})$ from (3.47), denoted as $p_{\underline{s_j}|\mathrm{gen}}(s|\mathrm{gen})$ for the genuine case and $p_{\underline{s_j}|\mathrm{imp}}(s|\mathrm{imp})$ for the imposter case. By combining (3.47) with the joint probability density $p_{\underline{r},\underline{v}|\mathrm{gen}}(r, v|\mathrm{gen})$ from (3.41) for the genuine and $p_{\underline{r},\underline{v}|\mathrm{imp}}(r, v|\mathrm{imp})$ from (3.42) for the imposter case, respectively, we approximate the score density by means of numerical integration of the joint probability density along the score contour. Because $s(r, v; N_{\mathrm{e}}, N_{\mathrm{v}})$ from (3.47) is derived for the univariate case, thus the score densities $p_{\underline{s_j}|\mathrm{gen}}(s|\mathrm{gen})$ and $p_{\underline{s_j}|\mathrm{imp}}(s|\mathrm{imp})$ are for the univariate case as denoted by the j subscript.

For the multivariate case, when there are n independent feature components, the likelihood ratio equals the product of the likelihood ratio of each component. Because we use the log likelihood ratio as the comparison score, the multivariate comparison score equals the sum of the n univariate scores defined in (3.47). Hence, the multivariate comparison score density for the genuine $p_{\underline{s}|\mathrm{gen}}(s|\mathrm{gen})$ and imposter case $p_{\underline{s}|\mathrm{imp}}(s|\mathrm{imp})$ becomes the convolution of the univariate score density $p_{\underline{s_j}|\mathrm{gen}}(s|\mathrm{gen})$ and $p_{\underline{s_j}|\mathrm{imp}}(s|\mathrm{imp})$, respectively, namely

$$p_{\underline{s}}(s) \stackrel{\mathrm{def}}{=} (p_{\underline{s_1}} * p_{\underline{s_2}} * \ldots * p_{\underline{s_n}})(s). \tag{3.48}$$

Because the log likelihood comparison score is a similarity score, a match is returned only when the comparison score is larger than or equal to the operating point T. The two error types are a match obtained at an imposter comparison known as a false match and a non-match at a genuine comparison known as a false non-match. As the performance measures, we use the false non-match rate (FNMR) $\beta(T)$ and the false match rate (FMR) $\alpha(T)$ at the operating point T. With the multivariate score density we can compute the FNMR and FMR as

$$\beta(T) = \int_{\infty}^{T} p_{\underline{s}|\mathrm{gen}}(s|\mathrm{gen})\mathrm{d}s, \tag{3.49}$$

$$\alpha(T) = \int_{T}^{\infty} p_{\underline{s}|\mathrm{imp}}(s|\mathrm{imp})\mathrm{d}s. \tag{3.50}$$

Results

Figure 3.25 illustrates several examples of the approximated score density at (a) genuine and (b) imposter comparisons for the univariate case for different number number of enrolment and verification samples with $\sigma_{\mathrm{b}}^2 = \sigma_{\mathrm{w}}^2 = 1$, and (c) their corresponding receiver operating characteristics (ROC) curves. Similarly for the multivariate case in (d), (e) and (f), respectively, but for different dimensions n with $\sigma_{\mathrm{b}}^2 = \sigma_{\mathrm{w}}^2 = N_{\mathrm{e}} = N_{\mathrm{v}} = 1$. Note that the genuine score density is symmetric at a score of zero, while the imposter density is skewed towards the negative scores. Averaging multiple enrolment and verification samples has the effect of concentrating the genuine score density closer to zero, while skewing the imposter score density further towards the negative values. Both effects improve the performance as observed by the ROC curves. For the multivariate case, when

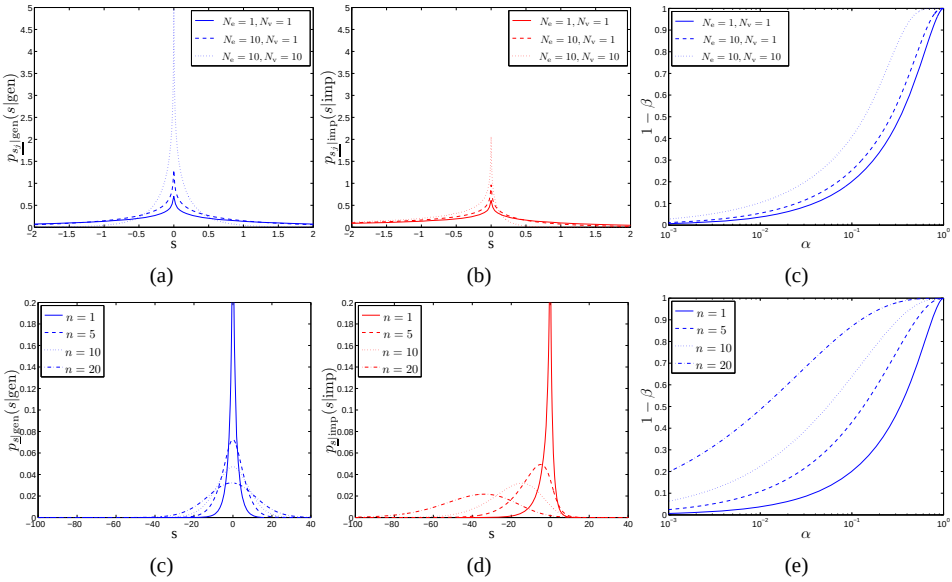


Figure 3.25: The approximated comparison score density for the univariate case with within-class and between class variance $\sigma_{\mathrm{w}}^2 = \sigma_{\mathrm{b}}^2 = 1$ for different number of enrolment N_{e} or verification N_{v} samples is shown (a) for the genuine $p_{\underline{s_j}|\mathrm{gen}}(s|\mathrm{gen})$ and (b) imposter $p_{\underline{s_j}|\mathrm{imp}}(s|\mathrm{imp})$ case, and (c) portrays the corresponding ROC curves. Furthermore, for the multivariate case is shown (d) $p_{\underline{s_j}|\mathrm{gen}}(s|\mathrm{gen})$, (e) $p_{\underline{s_j}|\mathrm{imp}}(s|\mathrm{imp})$, and (f) the ROC curves for different number of components n with $\sigma_{\mathrm{w}}^2 = \sigma_{\mathrm{b}}^2 = N_{\mathrm{e}} = N_{\mathrm{v}} = 1$.

increasing the number of components n the imposter score density significantly skews and shifts to the negative values while the genuine density becomes broader but remains symmetric. Overall, both effects combined improve the performance as illustrated by the ROC curves.

3.3.5 Binary Classifier Performance

The theoretical performance of a binary classifier when using a bit extraction method based on a single threshold at the background mean has been studied in [95]. For the genuine comparisons, the average bit-error probability of component j is analytically determined to be equal to

$$P_{\mathrm{e}}^{\mathrm{ge}}[j] = \tfrac{1}{2} - \tfrac{1}{\pi} \arctan\left(\frac{\sigma_{\mathrm{b}}[j]}{\sigma_{\mathrm{w}}[j]} \frac{\sqrt{N_{\mathrm{e}} N_{\mathrm{v}}}}{\sqrt{N_{\mathrm{e}} + N_{\mathrm{v}} + \left(\frac{\sigma_{\mathrm{b}}[j]}{\sigma_{\mathrm{w}}[j]}\right)^{-2}}}\right). \tag{3.51}$$

The bit-error probability determines the number of bit errors or Hamming distance ϵ between the binary vectors extracted in the enrolment and verification phase. Under the

assumption of having independent components, the probability mass function (pmf) of ϵ is the following convolution

$$p_{\underline{\epsilon}}(\epsilon) \stackrel{\text{def}}{=} (P_1 * P_2 * \ldots * P_{n_\text{c}})(\epsilon), \tag{3.52}$$

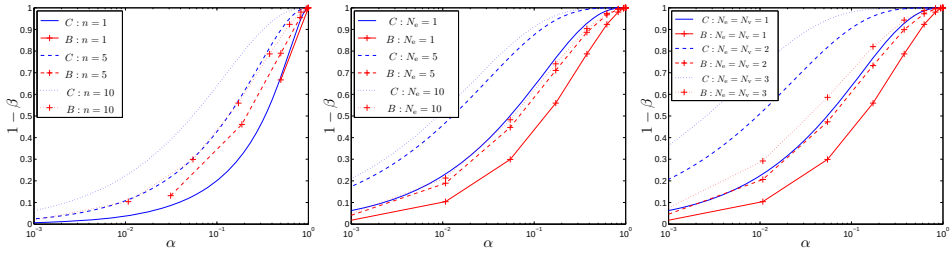
where $P_j = [1 - P_\text{e}[j], P_\text{e}[j]]$ is the marginal pmf of the single bit extracted from component j. Note that the number of bit errors ϵ is a distance score and a match is obtained when ϵ is smaller or equal to the operating point T. Thus, the FNMR $\beta(T)$ and FMR $\alpha(T)$ at the operating point T are defined as

$$\begin{aligned} \beta(T) &= \sum_{\epsilon=T+1}^{n} p_{\underline{\epsilon|\text{gen}}}(\epsilon|\text{gen}), \\ \alpha(T) &= \sum_{\epsilon=0}^{T} p_{\underline{\epsilon|\text{imp}}}(\epsilon|\text{imp}), \end{aligned} \tag{3.53}$$

where the bit-error probability P_e^ge from (3.51) is used for the genuine case and $P_\text{e}^\text{im} = 0.5$ for the imposter case.

3.3.6 Performance Comparison

A comparison of the theoretical performances determined in Section 3.3.4 for the continuous classifier and Section 3.3.5 for the binary classifier is portrayed by the ROC curves in Figure 3.26(a) for different feature dimensions n with $\sigma_\text{w}^2 = \sigma_\text{b}^2 = N_\text{e} = N_\text{v} = 1$, for different number of enrolment samples N_e with $n = 10$ and $\sigma_\text{w}^2 = \sigma_\text{b}^2 = N_\text{v} = 1$ in Figure 3.26(b), and in Figure 3.26(c) for different number of enrolment and verification samples $N_\text{e} = N_\text{v}$ with $n = 10$ and $\sigma_\text{w}^2 = \sigma_\text{b}^2 = 1$. The continuous classifier is denoted by the prefix C, while the binary classifier is denoted by the prefix B. In all three cases the results clearly show that the continuous classifier outperforms the binary classifier and changing either the dimension n or the number of enrolment or verification samples has a greater improvement for the continuous classifier. A drawback of the binary classifier is that the binarization process under consideration extracts a single bit by coarsely dividing the feature space of a component in two regions only and therefore discarding essential information. This loss is clearly shown by the 'n=1' ROC curve in Figure 3.26(a), where the continuous classifier ROC curve has an infinite number of operating points and can reach any FMR of FNMR value, while the binary classifier has only two operating points where the smallest FMR is 50%. As observed in Figure 3.26(a), this information loss has a snowball effect when increasing the dimension n, because the performance of the continuous classifier has a greater improvement with increasing n than the binary classifier performance. Extracting a single bit becomes more disadvantageous when the within-class variance is suppressed by increasing the number of enrolment or verification samples, or similarly having better feature components, i.e. feature components with a larger feature quality ratio $\frac{\sigma_\text{b}}{\sigma_\text{w}}$. When having better feature components it may be better to extract more bits instead of one.



(a) (b) (c)

Figure 3.26: The ROC performance comparison between the continuous (denoted by C) and binary classifier (denoted by B) for (a) different feature dimensions n with $\sigma_{\mathrm{w}}^2 = \sigma_{\mathrm{b}}^2 = N_{\mathrm{e}} = N_{\mathrm{v}} = 1$, (b) different number of enrolment samples N_{e} with $n = 10$ and $\sigma_{\mathrm{w}}^2 = \sigma_{\mathrm{b}}^2 = N_{\mathrm{v}} = 1$, and (c) different number of enrolment and verification samples $N_{\mathrm{e}} = N_{\mathrm{v}}$ with $n = 10$ and $\sigma_{\mathrm{w}}^2 = \sigma_{\mathrm{b}}^2 = 1$.

3.3.7 Conclusions

The requirement to extract a binary representation from the real-valued biometric sample for several template protection schemes known in the literature raises the question whether the bit extraction method reduces the classification performance. In this work we compared the theoretical performance of the optimal log likelihood ratio continuous classifier with the binary Hamming distance classifier under the assumption of Gaussian biometric data modeled by the between-class and within-class densities with independent feature components and including the averaging of multiple enrolment and verification samples.

In the literature, the theoretical performance for the binary classifier consisting of a single bit extraction method based on thresholding has been studied. Similarly, the theoretical performance of a continuous classifier based on the log likelihood ratio comparison scores has been analyzed, but was limited to the averaging of multiple enrolment samples only. Hence, in this work we extended the analysis by including the averaging of multiple verification samples. We approximated the density of the comparison score for the univariate and multivariate case, from which we computed the corresponding performance curve.

Consequently, we compared the theoretical performance of the continuous and binary classifier and studied the effect of the number of the feature dimension and the number of enrolment and verification samples. In all cases the continuous classifier outperforms the binary classifier, which is expected as the likelihood ratio is the optimal classifier if the class-conditional probability is well-known. In this work we assumed the class-conditional probability to be well defined. In practice, however, the performance advantage of the continuous classifier will be less because it is known to be difficult to have a perfect estimation of the class-conditional probability, especially at high feature dimensions or correlated feature components. A drawback of the binary classifier under consideration is that the bit extraction method coarsely divides the feature space of a

component in only two regions in order to extract a single bit and therefore discarding essential information. This drawback is amplified when the within-class noise is suppressed by increasing the number of enrolment or verification samples, where it may be more advantageous to extract more than one bit from each feature component.

As future work, it would be of great interest to derive the theoretical performance of more advanced bit extraction methods that can extract more robust bits or multiple bits from each component in order to close the gap between the continuous and binary classifier. Furthermore, it is important to investigate the sensitivity of both classifiers with respect to correlated feature components and estimation errors of the class-conditional probability.

3.4 Chapter Conclusions

With the first part, Section 3.2, we have shown that it is possible to theoretically determine the classification performance of the HDS based on a single bit extraction scheme employing a single quantization threshold. This was primarily accomplished by deriving a closed-form analytical expression of the average bit-error probability of extracted bit from a component. We experimentally validated the performance estimation using fingerprint and 3D face data. The naive model assuming independent feature components with a homogeneous within-class variance has a large deviation, which can be reduced by incorporating the dependent and non-homogeneous feature components. Increasing system parameters, such as the number of enrolment and verification samples, improve the classification performance by reducing the within-class variance, and also improve the performance estimation. The performance estimation becomes more accurate because when averaging multiple acquired biometric samples the within-class distribution becomes more Gaussian as dictated by the central limit theorem, hence the averaged samples will fit the Gaussian model more closely.

With the second part, Section 3.3, we have shown that the classification performance of the unprotected templates (on continuous level) using the optimal likelihood ratio classifier is better than the performance of the protected templates using the HDS with a single bit extraction scheme based on a single quantization threshold. The performance difference increases with the feature vector dimension and the number of enrolment or verification samples. The results are however optimistic, because we considered the naive model of independent components with a homogeneous within-class variance, which does not hold in practice as we have shown in Section 3.2. In order for the likelihood ratio classifier to be optimal the class-conditional probability of a feature vector has to be well defined, thus an accurate estimation of the feature dependency and non-homogeneous property is required. We conjecture that the observed difference between the protected and unprotected performance will decrease when there are inaccuracies in the estimation of the class-conditional probability.

Chapter 4

Maximum Key Size

4.1 Chapter Introduction

In this chapter the second research question will be addressed, namely

> **Given the HDS template protection scheme: What is the maximum key size at a given target classification performance and system parameters?**

Using the naive model introduced in Chapter 3 we determine the maximum key size by assuming the ECC to be operating on Shannon's bound and given the system parameters such as the input capacity, the number of feature components, the number of enrolment and verification samples, and the target performance. Section 4.2.3 and a great part of Section 4.2.4 overlap with the modeling work in Section 3.2.3, and can be skipped at first reading. We also investigate the maximum key size for the case where the model includes fully dependent feature components and components with dissimilar feature quality. The main results are published in Kelkboom et al. (2010) [114] [1].

[1] E. J. C. Kelkboom, J. Breebaart, I. R. Buhan, and R. N. J. Veldhuis, "Analytical template protection performance and maximum key size given a Gaussian modeled biometric source," Submitted to IEEE Transactions on Information Forensics and Security, 2010.

4.2 Analytical Template Protection Performance and Maximum Key Size given a Gaussian Modeled Biometric Source: A trade-off between privacy, security and convenience

4.2.1 Abstract

Template protection techniques are used within biometric systems in order to protect the stored biometric template against privacy and security threats. A great portion of template protection techniques are based on extracting a key from, or binding a key to a biometric sample. The achieved privacy and security depend on the entropy of the key. We focus on the key binding method known as the Fuzzy Commitment Scheme. In the literature it can be observed that there is a large variation on the reported key lengths at similar classification performance of the same template protection system, even when based on the same biometric modality and database. In this work we determine the analytical relationship between the classification performance of the Fuzzy Commitment Scheme and the theoretical maximum key size given as input a Gaussian modeled biometric source. We show the effect of the system parameters such as the biometric source capacity, the number of feature components, the number of enrolment and verification samples, and the target performance on the maximum key size. We also show that a trade-off exists between the privacy and security of the Fuzzy Commitment Scheme and its convenience for its subjects. Furthermore, we provide an analysis of the effect of feature interdependencies and differences in their quality. Finally, we analyze these findings on the MCYT fingerprint database using two feature extraction algorithms.

4.2.2 Introduction

A biometric system consist of an enrolment and verification phase. In the enrolment phase, a biometric sample is captured from which a reference template is created and stored. In the verification phase, a new biometric sample is captured and compared with the stored reference template. The subject is considered as being genuine if the new biometric sample is sufficiently similar to the stored reference template. In recent years, the interest in biometric systems has significantly increased. Examples are the planned introduction of the United Kingdom National Identity Card based on biometrics required by the Identity Cards Act 2006 [8] or the recommendation by the International Civil Aviation Organization (ICAO) [9] to adopt the ePassport that also includes biometric data. The widespread use of biometrics and its necessity of storing a reference template introduces new security and privacy risks such as (i) *identity fraud* where an adversary steals the stored reference template and impersonates the genuine subject of the system by some spoofing mechanism, (ii) *limited-renewability* implying the limited capability to renew a compromised reference template due to the limited number of biometric instances (for example we only have ten fingers, two irises or retinas, and a single face), (iii) *cross-matching* linking reference templates of the same subject across databases of different

applications, and (iv) *(sensitive) medical information leakage* where it is known that biometric data may reveal the gender, ethnicity, or the presence of certain diseases.

The field of template protection is focused on mitigating these privacy risks by developing template protection techniques that provide (i) *irreversibility* implying that it is impossible or at least very difficult to retrieve the original biometric sample from the reference template, (ii) *renewability* where it is possible to renew the reference template when necessary, and (iii) *unlinkability* which prevents cross-matching. The field of template protection is relatively young, however there is a significant interest to successfully develop and implement these techniques as shown by their prominent position within the European projects 3DFace [30] and TURBINE (TrUsted Revocable Biometric IdeNtitiEs) [31] from the 6th and 7th Framework Programme, respectively, and the great interest from privacy offices such as the Office of the Information and Privacy Commissioner of Ontario [32].

Overview of the Template Protection Field

As described in Jain et al. (2008) [49], the template protection techniques proposed in the literature can be divided into two categories, namely (i) *feature transformation* and (ii) *biometric cryptosystems*. The most common technique based on feature transformation is known as *Cancelable Biometrics* [58, 61]. With cancelable biometrics, the reference template is generated by applying a non-invertible transformation on the enrolment sample. Due to the non-invertible property of the transformation it is impossible to obtain the original biometric sample from the reference template. In the verification phase, the same non-invertible transformation is applied on the verification sample, and the matching is thus performed on the transformed version of both the enrolment and verification sample. Biometric cryptosystem techniques can be sub-divided into (1) *key binding* and (2) *key generation* methods. In the enrolment phase, the key binding techniques combines the key with a biometric sample into auxiliary data as such that the same key can be successfully released in the verification phase. The key release process in the verification phase uses a new biometric sample and the stored auxiliary data. Examples of the key binding techniques are the *Fuzzy Commitment Scheme* (FCS) [36], the *Helper Data System* (HDS) [48], the *Fuzzy Vault* [80]. Key generation techniques extract a robust key from the biometric sample in the enrolment phase, with auxiliary data if necessary. In the verification phase the same key has to be extracted using a new biometric sample and, when available, the auxiliary data. *Fuzzy Extractors* are the most common key generation techniques, which can be created using *Secure Sketches* [115].

Privacy and Security, and Convenience

It is known from the key binding technique that given the protected template, an adversary could retrieve the binary vector extracted from the biometric sample by randomly guessing the key and inverting the key binding process. Compromising the binary vector facilitates a possible replay or cross-matching attack and is therefore clearly a security and privacy breach, respectively. Besides the cross-matching privacy breach, the binary vector could also reveal sensitive or medical information of the subject. Therefore, the

achieved privacy and security protection depends on the entropy of the key. Considering the key to consist out of independent and uniform bits, its entropy is then determined by its size. Having a key of k_c bits on average will take 2^{k_c-1} guesses in order to obtain the correct one, hence adding a single bit to the key doubles the adversary's effort.

On the other hand, the classification performance of the template protection system also determines the effort of inverting the key-binding process. In the remainder of this work we refer to the classification performance of the template protection system as the system performance. The system performance can be expressed by the *false match rate* (FMR) and the *false non-match rate* (FNMR). The FMR is the probability of incorrectly classifying the biometric samples from two different subjects as similar and genuine, hence leading to a false match. Thus, the FMR also indicates the likelihood of finding a random biometric sample from an existing database that will lead to a match and therefore a security breach, which is also known as the FMR attack. The work of Korte and Plaga (2007) [116] describes a relationship between the FMR and the key size, namely $k_c \leq -\log_2(FMR)$. Furthermore, the FNMR is the probability of incorrectly classifying two biometric samples from the same subject as different or imposter, thus leading to a false non-match. We consider the FNMR as part of the convenience factor of the biometric system, because it determines the probability that subjects have to repeat the verification process which is considered as an unpleasant experience. It is also known that increasing the FNMR usually results into an decrease of the FMR, and therefore a possible increase the key size. Furthermore, acquiring multiple biometric samples will improve the performance as shown in Kittler et. al (1997) [117], Faltemier et al. (2008) [118], and Kelkboom et al. [97], but is likely to be considered as inconvenient by the subjects. Hence, both the FNMR and the number of samples do influence the key size, hence showing a possible trade-off between the privacy and security and the convenience of the template protection system.

Reported Performances with Corresponding Key Size

In the literature, there is a significant variability in the reported key size with respect to the system performance. Table 4.1 shows an overview of the reported system performance and key size for different template protection techniques, databases and feature extraction methods. It is difficult to find a relationship between the system performance and the key size. For example, consider the cases 6 and 11 that use the same template protection technique and modality, and a similar database. While having similar reported performance, the key size in case 11 is almost three times larger than in case 6. Likewise, when comparing the cases 2c and 10a with similar template protection technique, modality, and database, the key size reported in case 10a is almost double of the one of case 2c. As last example, the separate cases 7 and 10 show that using exactly the same template protection technique on the same modality but different database may lead to a different performance at an equal key size as in case 7 or different key sizes at similar performance as in case 10. *Hence, in practice there seems no clear relationship between the system performance and the key size*.

Table 4.1: Overview of the key size and the classification performance of different biometric cryptosystems techniques, modalities and databases found in the literature. The biometric cryptosystems under considerations are the fuzzy extractors, the fuzzy commitment schemes (FCS), the helper data systems (HDS), the fuzzy vault, and the code-offset construction.

Work	Case	Method	Modality	Database (# samples / instance)	FMR	FNMR	Key size [bits]
Arakala et al. [93]	1	fuzzy extractors	fingerprint	FVC2000 DB1a (800/100)	0.15	0.15	34
Bringer et al. [71]	2a	FCS	iris	ICE 2005 (2953/244)	$< 10^{-5}$	0.0562	42
	2b	FCS	iris	CASIA (756/108)	≈ 0	0.0665	42
	2c	FCS	fingerprint	FVC2000 DB2a (800/100)	0.0553	0.0273	42
Chang et al. [65]	3	FCS	fingerprint	NIST 4 (4000/2000)	≈ 0.001	≈ 0.10	10
Clancy et al. [79]	4	fuzzy vault	fingerprint	-	-	0.20-0.30	69
Hao et al. [66]	5	code-offset	iris	private (700/70)	≈ 0	0.0047	140
					0.02%	0.15%	112
Kelkboom et al. [33]	6	HDS	3D face	FRGC v2 subset (2347/145)	0.0019	0.16	35
Kevenaar et al. [34]	7a	HDS	2D face	FERET ($> 948/237$)	≈ 0	0.35	58
	7b	HDS	2D face	Caltech ($> 209/19$)	≈ 0	0.035	58
Nandakumar et al. [119]	8a	fuzzy vault	fingerprint	FVC2000 DB2a (800/100)	$\approx 10^{-4}$	0.09	≈ 40
	8b	fuzzy vault	fingerprint	MSU-DBI (640/160)	$\approx 2 \cdot 10^{-4}$	0.175	≈ 40
Sutcu et al. [120]	9	FCS	fingerprint	Mitsubishi (1035/69)	$1.19 \cdot 10^{-4}$	0.11	30
Tuyls et al. [35]	10a	HDS	fingerprint	FVC2000 DB2a&b (880/110)	0.052	0.054	76
	10b	HDS	fingerprint	Univ. Twente (2500/500)	0.035	0.054	40
Zhou et al. [121]	11	HDS	3D face	FRGC v1 subset ($> 396/99$)	0.004	0.12	107

Related Work and Contributions

We are interested in determining the relationship between the maximum key size and the system performance and investigate the influence of the system parameters such as the input capacity, the number of feature components, and the number of enrolment and verification samples.

An analysis about the maximum key size given a discrete biometric source is done in Ignatenko and Willems (2009) [122] (which is an extended version of Ignatenko and Willems (2008) [123]) and a similar work of Lai et al. (2008) [124], where they estimated the secret-key rate. The work of Willems and Ignatenko (2009) [38] analyzed the secret-key rate for a Gaussian distributed continuous biometric source. The framework of these works assumes that if the number of feature components goes to infinity, the discriminating power of each component remains constant. Assuming independent feature components, this would imply that the biometric source has an infinite discriminating power. This would not hold for a biometric system, where the discriminating power of a biometric trait is limited due to its practical nature, namely measurement noise or biometric variability.

In our work we fix the discriminating power of our Gaussian modeled continuous biometric source, referred to as the input capacity, and distribute its discriminating power among the feature components. We present five contributions. Firstly, we analytically determine the classification performance of the Fuzzy Commitment Scheme where the input is a Gaussian modeled biometric source. We also include the number of enrolment and verification samples. Secondly, from the estimated performance we analytically determine the theoretical maximum key size at the operating point determined by the target FNMR, assuming an ECC with decoding capabilities at Shannon's bounded. We also verify the known relationship between the maximum key size and the FMR and illustrate the gap when errors have to be corrected. Thirdly, we investigate by means of numerical analysis the effect of the parameters such as the Gaussian capacity of the biometric source, the number of enrolment and verification samples, and the target FNMR on the maximum key size. Fourthly, we provide an analysis of the effect of feature interdependencies and differences in their quality. Finally, we analyze these findings on the MCYT fingerprint database using two feature extraction algorithms.

Outline

The outline of this paper is as follows. We briefly describe the FCS construction in Section 4.2.3. In Section 4.2.4 we present the analytical framework that models the biometric source as parallel Gaussian channels. Furthermore, we derive the analytical system performance and the theoretical maximum key size. Section 4.2.5 illustrates by means of numerical analysis the effect of the system parameters, feature interdependencies and differences in their quality on the maximum key size. The experimental setup using the MCYT database and the obtained results are discussed in Section 4.2.6. Our final remarks and conclusions are given in Section 4.2.7.

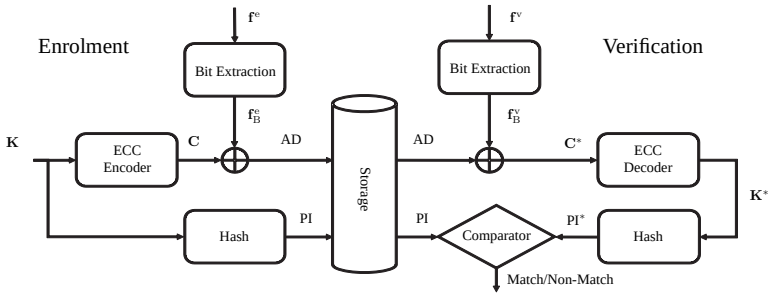


Figure 4.1: The FCS construction combined with a *Bit Extraction* module.

4.2.3 Fuzzy Commitment Scheme

The FCS construction combined with a *Bit Extraction* module is depicted in Figure 4.1. Note that the FCS is considered to be a key-binding technique. In the enrolment phase or the key-binding process, the real-valued column *feature vector* $\mathbf{f}^{\,\mathrm{e}} \in \mathbb{R}^{N_\mathrm{F}}$ is extracted from each of the N_e biometric enrolment samples by the feature extraction algorithm. A single binary column vector $\mathbf{f}_\mathrm{B}^\mathrm{e} \in \{0,1\}^{N_\mathrm{F}}$ is created from the mean of the N_e feature vectors within the *Bit Extraction* module, which we will discuss in Section 4.2.4. In context of template protection, the work of Kelkboom et al. (2009) [125] shows that multi-sample fusion at feature level, i.e. taking the mean of the multiple samples, has the greatest advantage in terms of security and privacy, and also storage requirements, while the performance is close to the optimal found at score-level fusion. Furthermore, a random key $\mathbf{K} \in \{0,1\}^{k_\mathrm{c}}$ is created and encoded by the *ECC Encoder* module into a codeword $\mathbf{C} \in \mathcal{C}$ of size $\{0,1\}^{n_\mathrm{c}}$, where $\mathcal{C}$ is the ECC codebook (the set of codewords). As the key-binding method, the codeword is XOR-ed with the binary vector $\mathbf{f}_\mathrm{B}^\mathrm{e}$, creating the helper data AD also referred to as the *Auxiliary Data* in Breebaart et al. (2008) [102], which is in line with standardization activities in ISO [25]. AD is stored as part of the protected template together with the hash of $\mathbf{K}$, which is referred to as the *Pseudonymous Identifier* (PI). Because of the XOR operation and the fact that a single bit is extracted from each feature component, it implies that the size of the extracted real-valued and binary vector are equal to the codeword size, namely $n_\mathrm{c} = N_\mathrm{F}$, and in the remainder of this work we will only use n_c.

In the verification phase or the key-release process, the binary vector $\mathbf{f}_\mathrm{B}^\mathrm{v}$ is created by quantizing the mean of the N_v verification feature vectors $\mathbf{f}^\mathrm{v}$. Hereafter, the auxiliary data AD is XOR-ed with $\mathbf{f}_\mathrm{B}^\mathrm{v}$ resulting into the possibly corrupted codeword $\mathbf{C}^*$. Decoding $\mathbf{C}^*$ by the *ECC Decoder* module leads to the candidate secret $\mathbf{K}^*$. The candidate pseudonymous identifier PI* is obtained by hashing $\mathbf{K}^*$. A match is returned by the *Comparator* module if PI and PI* are equal, which occurs only when $\mathbf{K}$ and $\mathbf{K}^*$ are equal, i.e. the key-released process was successful.

The key-binding and key-release process can be modeled by a binary symmetric channel (BSC) as portrayed in Figure 4.2, where the error pattern $\mathbf{e} = \mathbf{f}_\mathrm{B}^\mathrm{e} \oplus \mathbf{f}_\mathrm{B}^\mathrm{v}$ of weight

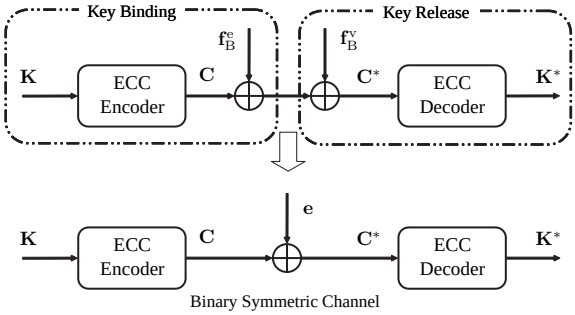


Figure 4.2: Modeling the key binding and release process by a Binary Symmetric Channel.

$\epsilon = ||\mathbf{e}|| = d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}})$ corrupts the original codeword used in the key-binding process. The bit-error probability P_{e}, which is the probability that a bit of $\mathbf{e}$ is '1', determines the number of bit-errors that have to be corrected by the *ECC Decoder* and therefore also the system performance. The bit-error probability depends on the quantization method being used, the quality of the features, and the number of samples (see Section 4.2.4) and is different for imposter and genuine comparisons.

4.2.4 The Analytical Framework

In this section we present the analytical framework for modeling the biometric source, quantization method, system performance, and the maximum key size that can be extracted. An overview of this framework is depicted in Figure 4.3. The *Source Modeling* module models the biometric source from which the enrolment and verification feature vectors $\mathbf{f}$ are derived. Given the input capacity C_{in} and the number of feature components n_{c} as it parameters the *Source Modeling* module outputs the quality of feature component j defined by the within-class and between-class standard deviation ratio $\frac{\sigma_{\mathrm{b}}[j]}{\sigma_{\mathrm{w}}[j]}$, referred to as the feature quality. With the quantization method under consideration, the number of enrolment N_{e} and verification N_{v} samples, and the feature quality $\frac{\sigma_{\mathrm{b}}[j]}{\sigma_{\mathrm{w}}[j]}$, the *Quantization* module estimates the bit-error probability of the extracted bit from feature component j at genuine $P_{\mathrm{e}}^{\mathrm{ge}}[j]$ and imposter $P_{\mathrm{e}}^{\mathrm{im}}[j]$ comparisons. Knowing the bit-error probabilities the *Performance Estimation* module estimates the analytical system performance defined by the false match rate (FMR) $\alpha(T)$ and the false non-match rate (FNMR) $\beta(T)$ at all possible operating points T. Given the system performance and the target FNMR β_{tar}, the maximum extracted key size k_{c}^{*} is determined in the *Maximum Key Size* module. In the remainder of this section we discuss each module in more detail.

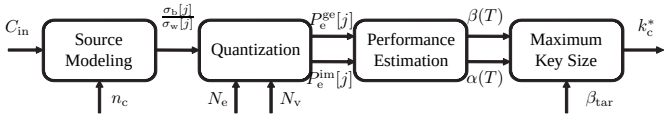


Figure 4.3: An overview of the framework used to model the biometric source defined by the feature quality $\frac{\sigma_{\text{b}}[j]}{\sigma_{\text{w}}[j]}$ of the j-th component, the resulting bit-error probabilities $P_{\text{e}}^{\text{ge}}[j]$ and $P_{\text{e}}^{\text{im}}[j]$, the corresponding performance defined by the FMR $\alpha(T)$ and the FNMR $\beta(T)$ at the operating point T, and the maximum key size k_{c}^* that can be extracted.

Biometric Source Modeling with Parallel Gaussian Channels

The input of the FCS template protection system is a real-valued column feature vector $\mathbf{f} = [f[1], f[2], \ldots, f[n_{\text{c}}]]'$ of dimension n_{c}, where ' $'$ ' is the transpose operator. The feature vector $\mathbf{f}$ is extracted from a biometric sample by the feature extractor and is likely to be different between two measurements, even if they are acquired immediately after each other. Causes for this difference include sensor noise, environmental conditions and biometric variabilities. To model these variabilities, we use the Parallel Gaussian Channels (PGC) as portrayed in Figure 4.4(a). This approach has been successfully used on estimating the performance of two biometric databases in Kelkboom et al. (2010) [95]. We assume an ideal Acquisition and Feature-Extraction module which always produces the same feature vector $\boldsymbol{\mu}_i$ for subject i. Such ideal module is thus robust against all aforementioned variabilities. However, the variability of component j is modeled as an additive zero-mean Gaussian noise $w[j]$ with its pdf $p_{\text{w[j],}i} \sim \mathcal{N}(0, \sigma_{\text{w},i}^2[j])$. Adding the noise $w[j]$ with the mean $\mu_i[j]$ results into the noisy feature component $f[j]$, in vector notation $\mathbf{f} = \boldsymbol{\mu}_i + \mathbf{w}$. The observed variability within one subject is characterized by the variance of the *within-class* pdf and is referred to as within-class variability. We assume that each subject has the same within-class variance, i.e. homogeneous within-class variance $\sigma_{\text{w},i}^2[j] = \sigma_{\text{w}}^2[j], \forall i$. We also assume the noise to be independent across components j, subjects i, and across measurements. Hence, the feature vector extracted from each biometric sample is equivalent to retransmitting $\boldsymbol{\mu}_i$ over the same PGC channels.

Each subject should have a unique mean in order to be distinguishable. Across the population we assume $\mu_i[j]$ to be another Gaussian random variable with density $p_{\text{b}[j]} \sim \mathcal{N}(\mu_{\text{b}}[j], \sigma_{\text{b}}^2[j])$. The variability of $\mu_i[j]$ across the population is referred to as the *between-class* variability. Figure 4.4(b) shows an example of the within-class and between-class pdfs for a specific component and a given subject. The *total* pdf describes the observed real-valued feature value $f[j]$ across the whole population and is also Gaussian with $p_{\text{t}[j]} \sim \mathcal{N}(\mu_{\text{t}}[j], \sigma_{\text{t}}^2[j])$, where $\mu_{\text{t}}[j] = \mu_{\text{b}}[j]$ and $\sigma_{\text{t}}^2[j] = \sigma_{\text{w}}^2[j] + \sigma_{\text{b}}^2[j]$. For simplicity but without loss of generality we consider $\mu_{\text{t}}[j] = \mu_{\text{b}}[j] = 0$.

The capacity of each channel is given by the Gaussian channel capacity $C_{\text{G}}[j]$ as defined in Cover and Thomas (1991) [126]

$$C_{\text{G}}[j] = \tfrac{1}{2} \log_2 \left(1 + \left(\frac{\sigma_{\text{b}}[j]}{\sigma_{\text{w}}[j]} \right)^2 \right), \tag{4.1}$$

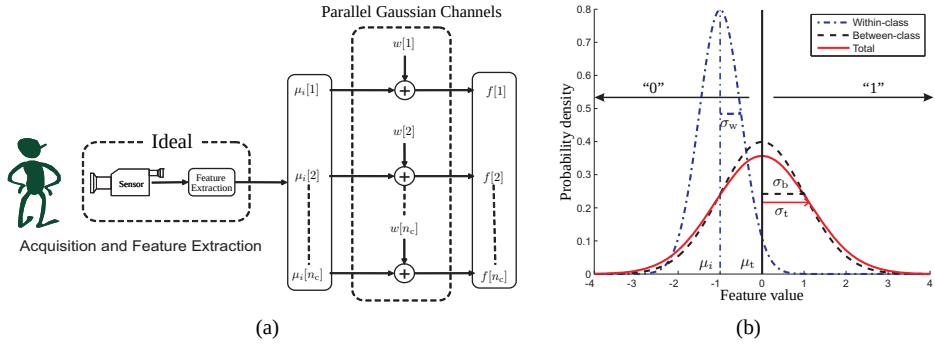


Figure 4.4: (a) The Parallel Gaussian Channels modeling the real-valued features and (b) the within-class, between-class and the total density and the quantization method based on thresholding.

which in fact states that a maximum of $C_{\mathrm{G}}[j]$ bits could be send per transmission. Note that the Gaussian channel capacity only depends on the ratio $\frac{\sigma_{\mathrm{b}}[j]}{\sigma_{\mathrm{w}}[j]}$ and in Section 4.2.4 we will also show that the bit-error probability P_{e} depends on this ratio. Therefore, we can define the ratio $\frac{\sigma_{\mathrm{b}}[j]}{\sigma_{\mathrm{w}}[j]}$ as the feature quality of component j and taking its inverse of (4.1) we obtain

$$\frac{\sigma_{\mathrm{b}}[j]}{\sigma_{\mathrm{w}}[j]} = \sqrt{2^{2C_{\mathrm{G}}[j]} - 1}, \tag{4.2}$$

where the relationship is graphically represented in Figure 4.5(a).

With the capacity of feature component j to be equal to the Gaussian channel capacity $C_{\mathrm{G}}[j]$, we can define the total capacity of the input biometric source C_{in} as the following sum

$$C_{\mathrm{in}} = \sum_{j=1}^{n_{\mathrm{c}}} C_{\mathrm{G}}[j]. \tag{4.3}$$

The input capacity C_{in} thus represents the amount of discriminating information in a biometric sample across the whole population and is distributed among the n_{c} components. In this work, we will analyze the difference between a uniform and a non-uniform distribution of C_{in}. If the input capacity is uniformly distributed among the n_{c} components, consequently given the input capacity C_{in} the Gaussian capacity of each component $C_{\mathrm{G}}[j]$ is equal to $\frac{C_{\mathrm{in}}}{n_{\mathrm{c}}}$. By substituting $C_{\mathrm{G}}[j] = \frac{C_{\mathrm{in}}}{n_{\mathrm{c}}}$ in (4.2) the feature quality parameter $\frac{\sigma_{\mathrm{b}}}{\sigma_{\mathrm{w}}}$ is defined as

$$\frac{\sigma_{\mathrm{b}}}{\sigma_{\mathrm{w}}} = \sqrt{2^{\frac{2C_{\mathrm{in}}}{n_{\mathrm{c}}}} - 1}, \tag{4.4}$$

and is also equal for each component. Thus, for this special case of a uniformly distributed input capacity, (4.4) gives the relationship between the input and output parameters of the *Source Modeling* module. In all other cases, (4.2) and (4.3) determine these properties.

Quantization Module based on Thresholding

Figure 4.4(b) depicts the quantization method under consideration, which is a binarization method based on thresholding, where the mean of the total density μ_{t} is taken as the threshold [33–35]. If the real-valued feature is larger than the threshold, then a bit of value '1' is allocated, otherwise '0'. To estimate the analytical system performance we need to estimate the bit-error probability $P_{\mathrm{e}}[j]$ for each component j at imposter and genuine comparisons. In this section we analytically estimate $P_{\mathrm{e}}[j]$ given the quantization scheme, the feature quality $\frac{\sigma_{\mathrm{b}}[j]}{\sigma_{\mathrm{w}}[j]}$, and the number of enrolment N_{e} and verification N_{v} samples.

Imposter Bit-Error Probability $P_{\mathrm{e}}^{\mathrm{im}}[j]$ At imposter comparisons, each bit is compared with the bit extracted from a randomly selected feature value from the total density. Because μ_{t} is the binarization threshold, there is a 50% probability that a randomly selected bit from the whole population will be equal, hence $P_{\mathrm{e}}^{\mathrm{im}}[j] = \frac{1}{2}$. Note that both the number of enrolment and verification samples do not have an influence on $P_{\mathrm{e}}^{\mathrm{im}}[j]$, and $P_{\mathrm{e}}^{\mathrm{im}}[j]$ is equal for each component.

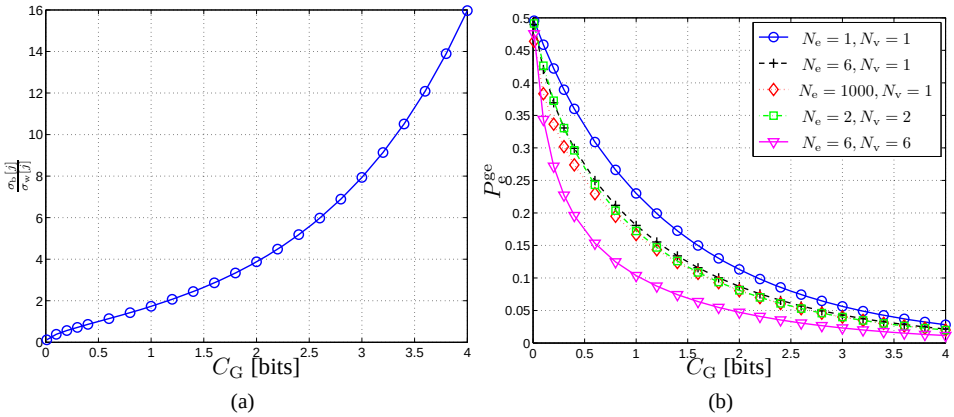
Genuine Bit-Error Probability $P_{\mathrm{e}}^{\mathrm{ge}}[j]$ At genuine comparisons, the analytical bit-error probability $P_{\mathrm{e}}^{\mathrm{ge}}[j]$ has been derived in Kelkboom et al. (2008) [97], namely

$$P_{\mathrm{e}}^{\mathrm{ge}}[j] = \tfrac{1}{2} - \tfrac{1}{\pi}\arctan\left(\frac{\frac{\sigma_{\mathrm{b}}[j]}{\sigma_{\mathrm{w}}[j]}\sqrt{N_{\mathrm{e}}N_{\mathrm{v}}}}{\sqrt{N_{\mathrm{e}}+N_{\mathrm{v}}+\left(\frac{\sigma_{\mathrm{b}}[j]}{\sigma_{\mathrm{w}}[j]}\right)^{-2}}}\right), \tag{4.5}$$

where it can be seen that the standard deviation ratio $\frac{\sigma_{\mathrm{b}}[j]}{\sigma_{\mathrm{w}}[j]}$ (the feature quality) and the number of enrolment N_{e} and verification N_{v} samples determine $P_{\mathrm{e}}^{\mathrm{ge}}[j]$. Note that $P_{\mathrm{e}}^{\mathrm{ge}}[j]$ is the average bit-error probability across the population. Some subjects have a larger bit-error probability because their mean $\mu_i[j]$ is closer to the quantization threshold $\mu_{\mathrm{t}}[j]$, while others have a smaller bit-error probability because their mean is further away. However, for estimating the analytical system performance across an infinite number of subjects, it is only necessary to compute the average bit-error probability as shown in Kelkboom et al. (2010) [95]. Substituting (4.2) into (4.5) we obtain

$$P_{\mathrm{e}}^{\mathrm{ge}}[j] = \tfrac{1}{2} - \tfrac{1}{\pi}\arctan\left(\frac{\sqrt{\left(2^{2C_{\mathrm{G}}[j]}-1\right)N_{\mathrm{e}}N_{\mathrm{v}}}}{\sqrt{N_{\mathrm{e}}+N_{\mathrm{v}}+\left(2^{2C_{\mathrm{G}}[j]}-1\right)^{-1}}}\right). \tag{4.6}$$

With (4.6) it is easy to show that $P_{\mathrm{e}}^{\mathrm{ge}}$ for the $N_{\mathrm{e}} = N_{\mathrm{v}} = 2X$ case converges to the $\{N_{\mathrm{e}} = \infty, N_{\mathrm{v}} = X\}$ case when $C_{\mathrm{G}}[j]$ and thus the feature quality becomes larger as such that $\left(2^{2C_{\mathrm{G}}[j]}-1\right)^{-1} \ll X$. Figure 4.5(b) depicts the bit-error probability $P_{\mathrm{e}}^{\mathrm{ge}}$ as a function of C_{G} for different settings of N_{e} and N_{v} as defined by (4.6). By increasing N_{e}, $P_{\mathrm{e}}^{\mathrm{ge}}$ decreases because the bits extracted in the enrolment phase are more stable. However, when increasing N_{e} further to infinity, $P_{\mathrm{e}}^{\mathrm{ge}}$ stays close to the $N_{\mathrm{e}} = N_{\mathrm{v}} = 2$ case and converges when C_{G} increases. To further decrease $P_{\mathrm{e}}^{\mathrm{ge}}$ it is thus necessary to also increase N_{v}.



(a) (b)

Figure 4.5: The (a) feature quality $\frac{\sigma_b[j]}{\sigma_w[j]}$ as a function of the Gaussian capacity per channel C_G and (b) the genuine bit-error probability P_e^{ge} as a function of C_G for different values of the number of enrolment N_e and verification N_v samples.

For the special case where the input capacity is uniformly distributed among the n_c components the feature quality $\frac{\sigma_b[j]}{\sigma_w[j]}$ is equal for each component and therefore $P_e^{ge}[j]$ is equal for each extracted bit. By substituting (4.4) into (4.5) we obtain

$$P_e^{ge} = \frac{1}{2} - \frac{1}{\pi} \arctan \left(\frac{\sqrt{\left(2^{\frac{2C_{in}}{n_c}} - 1\right) N_e N_v}}{\sqrt{N_e + N_v + \left(2^{\frac{2C_{in}}{n_c}} - 1\right)^{-1}}} \right). \tag{4.7}$$

System Performance

In Section 4.2.3 we have modeled the FCS template protection system as a binary symmetric channel with bit-error probability $P_e[j]$. The bit-error probability determines the probability mass function (pmf) of the number of bit errors or Hamming distance $\epsilon = d_H(\mathbf{f}_B^e, \mathbf{f}_B^v)$. As presented in Kelkboom et al. (2010) [95], the pmf is defined by the convolution

$$\begin{aligned} \phi(\epsilon) &\stackrel{\text{def}}{=} \mathcal{P}\{d_H(\mathbf{f}_B^e, \mathbf{f}_B^v) = \epsilon\} \\ &= (P_1 * P_2 * \ldots * P_{n_c})(\epsilon), \end{aligned} \tag{4.8}$$

where $P_j = [1 - P_e[j], P_e[j]]$ is the marginal pmf of the single bit extracted from component j. A toy example is depicted in Figure 4.6. The toy example shows the marginal pmf at comparisons between the enrolment and verification bits $f_B^e[1]$ and $f_B^v[1]$, respectively. Taking the convolution of all marginal pmf leads to the pmf of the Hamming distance ϵ.

For the special case where the input capacity is uniformly distributed across the n_c components and therefore $P_e[j]$ is equal for each component, the convolution in (4.8)

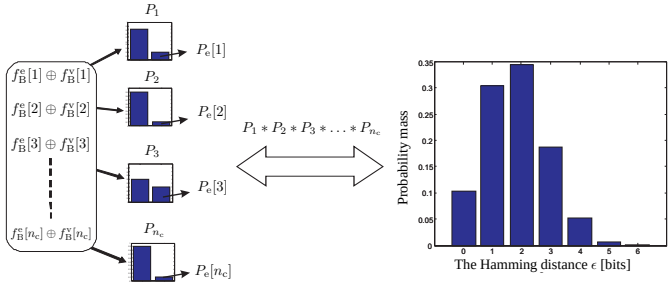


Figure 4.6: A toy example of the convolution method given by (4.8). (From Kelkboom et al. (2010) [95])

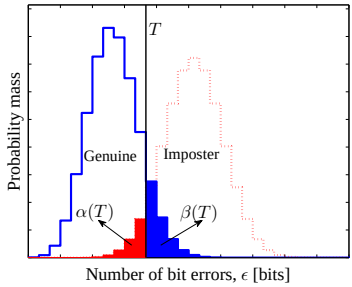


Figure 4.7: The false match rate (FMR) and the false non-match rate (FNMR) given the probability mass function of the number of errors ϵ at imposter and genuine comparisons.

becomes a binomial pmf $P_{\mathrm{b}}(\epsilon; N, p)$ as discussed in Daugman (2003) [104]

$$P_{\mathrm{b}}(\epsilon; N, p) = \binom{N}{\epsilon} p^{\epsilon}(1-p)^{(N-\epsilon)}, \tag{4.9}$$

with dimension $N = n_{\mathrm{c}}$ and probability $p = P_{\mathrm{e}}$.

False Match Rate The false match rate (FMR) depends on the pmf of the Hamming distance ϵ at imposter comparisons, where we have the bit-error probability $P_{\mathrm{e}}^{\mathrm{im}}$ that is equal for each extracted bit. Therefore, the pmf of the Hamming distance ϵ is the binomial pmf with p equal to $P_{\mathrm{e}}^{\mathrm{im}}$. Hence, the FMR at the operating point T, $\alpha(T)$, is the probability that ϵ is smaller or equal to T (see Figure 4.7), namely

$$\begin{aligned} \alpha(T) &\stackrel{\mathrm{def}}{=} \mathcal{P}\{\epsilon \leq T \mid \text{imposter comparisons}\} \\ &= \sum_{i=0}^{T} P_{\mathrm{b}}(i; n_{\mathrm{c}}, P_{\mathrm{e}}^{\mathrm{im}}) \\ &= 2^{-n_{\mathrm{c}}} \sum_{i=0}^{T} \binom{n_{\mathrm{c}}}{T}. \end{aligned} \tag{4.10}$$

False Non-Match Rate In general, $P_{\mathrm{e}}^{\mathrm{ge}}$ is not equal for each bit and therefore the pmf of the Hamming distance ϵ at genuine comparisons is defined by the convolution of (4.8) with marginal pmf's $P_j^{\mathrm{ge}} = [1 - P_{\mathrm{e}}^{\mathrm{ge}}[j], P_{\mathrm{e}}^{\mathrm{ge}}[j]]$. Hence, the false non-match rate at the operating point T, $\beta(T)$, is the probability that ϵ is larger than T (see Figure 4.7), namely

$$\begin{aligned}\beta(T) &\stackrel{\mathrm{def}}{=} \mathcal{P}\{\epsilon > T \mid \text{genuine comparisons}\} \\ &= \sum_{i=T+1}^{n_{\mathrm{c}}} (P_1^{\mathrm{ge}} * P_2^{\mathrm{ge}} * \ldots * P_{n_{\mathrm{c}}}^{\mathrm{ge}})(i).\end{aligned} \tag{4.11}$$

For the special case where the input capacity is distributed uniformly among the n_{c} components, the pmf of ϵ is defined by the binomial pmf with probability $p = P_{\mathrm{e}}^{\mathrm{ge}}$, namely

$$\beta(T) = \sum_{i=T+1}^{n_{\mathrm{c}}} P_{\mathrm{b}}(i; n_{\mathrm{c}}, P_{\mathrm{e}}^{\mathrm{ge}}). \tag{4.12}$$

Maximum Key Size

In this section, we determine the theoretical maximum key size or message length that can be transmitted given the BSC depicted in Figure 4.2 with bit-error probability P_{e} and assuming an optimal binary ECC that corrects up to t_{c} errors with decoding properties defined by Shannon's bound. We determine the maximum key size at the operating point determined by Shannon's theorem, at the operating point where the equal-error rate (EER) is achieved, and at the operating point determined by the target FNMR, β_{tar}. The EER is the performance achieved at the operating point where both the FMR and the FNMR are equal. Note that we assume an ECC code that considers the bit-error probability to be equal for each bit, hence we have a Hamming distance classifier with equal weights.

Operating Point from Shannon's Theorem With the code rate R equal to the ratio of the key size and the codeword size, $\frac{k_{\mathrm{c}}}{n_{\mathrm{c}}}$, Shannon's theorem shows that there exists a decoding technique that can decode the corrupted codeword with a bit-error rate p with an arbitrary small probability of a decoding error when

$$R < C(p) \tag{4.13}$$

for a sufficiently large value of n_{c}, where $C(p)$ is the channel capacity defined as

$$C(p) = 1 - h(p), \tag{4.14}$$

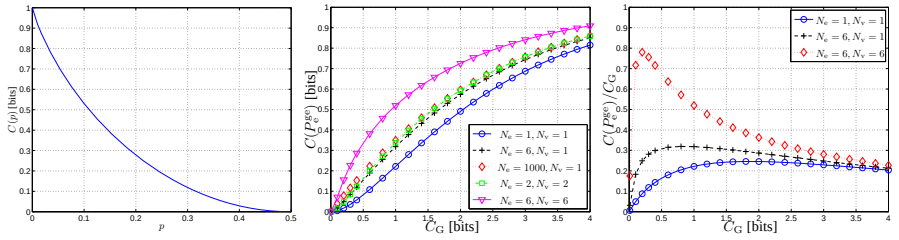
with $h(p)$ being the binary entropy function (see Figure 4.8(a))

$$h(p) = -p \log_2 p - (1-p) \log_2 (1-p). \tag{4.15}$$

Hence, the key size k_{c} has an upper limit given by Shannon's bound with $p = P_{\mathrm{e}}^{\mathrm{ge}}$ as

$$k_{\mathrm{c}} = n_{\mathrm{c}} R < n_{\mathrm{c}} C(P_{\mathrm{e}}^{\mathrm{ge}}). \tag{4.16}$$

Note that this bound only holds under the assumption that n_{c} is sufficiently large. With use of (4.6) we have the relationship between the Gaussian channel capacity C_{G} and



(a) (b) (c)

Figure 4.8: The (a) binary symmetric channel (BSC) capacity as a function of the bit-error probability p, (b) the BSC capacity $C(P_{\mathrm{e}}^{\mathrm{ge}})$ as a function of the Gaussian capacity per channel C_{G} at different values of the number of enrolment N_{e} and verification N_{v} samples, and (c) the key extraction efficiency defined as the ratio $\frac{C(P_{\mathrm{e}}^{\mathrm{ge}})}{C_{\mathrm{G}}}$.

the BSC channel capacity $C(P_{\mathrm{e}}^{\mathrm{ge}})$ as illustrated in Figure 4.8(b) for different number of enrolment N_{e} and verification N_{v} samples settings. Increasing the number of samples improves the BSC channel capacity $C(P_{\mathrm{e}}^{\mathrm{ge}})$ because of the decrease of the genuine bit-error probability $P_{\mathrm{e}}^{\mathrm{ge}}$. We also portray the key extraction efficiency defined as the ratio $\frac{C(P_{\mathrm{e}}^{\mathrm{ge}})}{C_{\mathrm{G}}}$ in Figure 4.8(c). Thus, with a single enrolment and verification samples, it is only possible to extract $\approx 24\%$ of the Gaussian channel capacity C_{G} and thus also from the input capacity C_{in}. The efficiency can be significantly improved by increasing the number of enrolment or verification samples, consequently shifting the optimum at a smaller C_{G}.

In practice, however, n_{c} is not large. As described in Daugman (2003) [104], the intrinsic degrees of freedom of the binary iris code is 249, which has been derived by fitting the imposter Hamming distance pmf with a binomial pmf with probability $p = 0.5$ and dimension $N = 249$. A toy example of the achieved FNMR when choosing the operating, $\frac{T}{n_{\mathrm{c}}} = 0.2$ close to $P_{\mathrm{e}}^{\mathrm{ge}} = 0.19$ as stipulated by Shannon's theorem for different values of n_{c} is depicted in Figure 4.9. At a large codeword size of $n_{\mathrm{c}} = 10000$ bits the achieved FNMR is 0.6%, which is acceptable. Note however that the FNMR significantly increases once n_{c} decreases, namely 22.1% and 43.9% at $n_{\mathrm{c}} = 1000$ and $n_{\mathrm{c}} = 100$ bits, respectively. Hence, with iris having 249 independent bits and is known as one of the best biometrics modality, we can conclude that the codeword size is expected to be too small and not fulfilling the requirement of Shannon's theorem. To lower the FNMR we have to correct more bits. In the following section we describe two alternative operating points, namely at the EER operating point or at the target FNMR β_{tar}.

The EER Operating Point with Gaussian Approximation In order to find an analytical expression of the EER operating point, T_{EER}, we approximate the binomial density used for modelling the pmf of the Hamming distance ϵ by a Gaussian density. The EER oper-

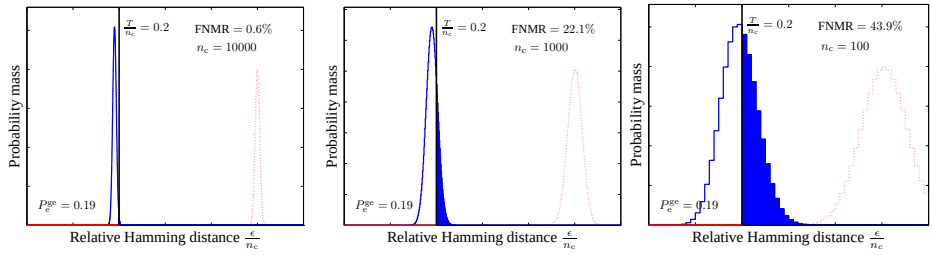


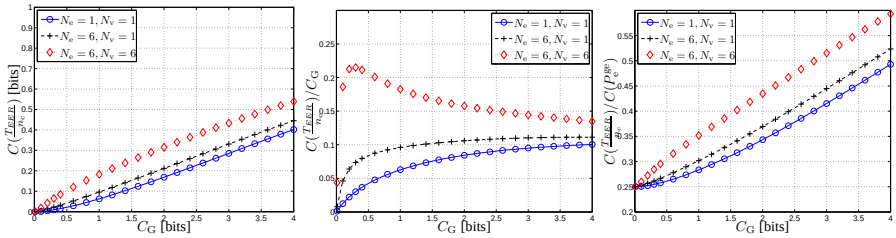
Figure 4.9: A toy example of the achieved FNMR when choosing the operating, $\frac{T}{n_c} = 0.2$ close to $P_e^{ge} = 0.19$ as stipulated by Shannon's theorem for different values of n_c. The solid (blue) curve portrays the pmf of the Hamming distance ϵ at genuine comparisons, while the dotted (red) curves depicts the pmf at imposter comparisons.

ating point in terms of P_e^{ge} becomes

$$\frac{T_{EER}}{n_c} = \frac{\sqrt{P_e^{ge}(1-P_e^{ge})}+P_e^{ge}}{2\sqrt{P_e^{ge}(1-P_e^{ge})}+1}. \tag{4.17}$$

where the complete derivation is presented in Section 4.A. Note that the relative operating point $\frac{T_{EER}}{n_c}$ and thus the BSC channel capacity at the EER operating point $C(\frac{T_{EER}}{n_c})$ is fully determined by P_e^{ge}. By combing (4.6) with (4.17) we obtain the relationship between the Gaussian channel capacity C_G and $C(\frac{T_{EER}}{n_c})$ for different number of enrolment N_e and verification N_v samples as depicted in Figure 4.10(a). Using more samples improves the key extraction efficiency as portrayed by Figure 4.10(b). However, the key extraction efficiency at the EER operating point is much smaller than the efficiency achieved at the operating point stipulated by Shannon's theorem. This difference is portrayed in Figure 4.10(c) by the ratio $C(\frac{T_{EER}}{n_c})/C(P_e^{ge})$, where it shows that within this range of C_G, N_e and N_v, the key extraction efficiency at the EER operating point is between 25-55% of the efficiency at Shannon's operating point. This reduction is caused by the fact that we force the FNMR to be equal to the FMR, hence significant more bits have to be corrected, consequently limiting the channel capacity. Therefore, in practice, the EER operating point may not be ideal in terms of performance. As a result, we discuss in the following section an alternative method where the operating point is determined by a target FNMR β_{tar}.

Operating Point at the Target FNMR β_{tar} We have shown that Shannon's theory leads to an optimistic upper bound with a high FNMR, while the EER operating point may not be the ideal operating point of a biometric system in terms of FMR, which consequently leads to a smaller maximum key size. In this section we present a different operating point determined by the target performance, namely the target FNMR, β_{tar}. Hence, instead of correcting $t_c = n_c P_e^{ge}$ or T_{EER} bits, we will correct $t_c = T_{tar}$ bits, where T_{tar} is the



(a) (b) (b)

Figure 4.10: The (a) BSC channel capacity at the EER operating point $C(\frac{T_{EER}}{n_c})$ as a function of the Gaussian channel capacity C_G at different values of N_e and N_v with its key extraction efficiency defined as $C(\frac{T_{EER}}{n_c})/C_G$ depicted in (b). Furthermore, (c) illustrates the ratio between the key extraction efficiency obtained when using the operating point from Shannon's theory and at the EER point, namely the ratio between $C(\frac{T_{EER}}{n_c})$ and $C(P_e^{ge})$.

operating point in order to reach β_{tar}, namely

$$T_{tar} = \arg\min_{T}(|\beta(T) - \beta_{tar}|). \tag{4.18}$$

Hence, the theoretical maximum key size assuming an ECC at Shannon's bound with $p = \frac{T_{tar}}{n_c}$ is then equal to

$$k_c^* \stackrel{\text{def}}{=} n_c C\left(\frac{T_{tar}}{n_c}\right) = n_c\left(1 - H\left(\frac{T_{tar}}{n_c}\right)\right). \tag{4.19}$$

Because $\frac{T_{tar}}{n_c}$ is larger than P_e^{ge} and will not exceed $\frac{1}{2}$, we know that k_c^* will be smaller than the upper bound $n_c C(P_e^{ge})$ from (4.16). However, if β_{tar} is larger than the EER then k_c^* will be larger than $C(\frac{T_{EER}}{n_c})$. Thus, the key extraction efficiency depends on the target FNMR, however if $\beta_{tar} \leq 50\%$ the key extraction efficiency should be between the one from Shannon's theory of Figure 4.8(c) and from the EER operating point of Figure 4.10(b).

We have defined the maximum key size k_c^*, which we will use in the remainder of this work. In the following section, we study the effect of the system parameters of the framework shown in Figure 4.3 on k_c^*.

Relationship between the Maximum Key Size k_c^* and the Target FMR α_{tar} The work of Korte and Plaga (2007) [116] showed the relationship between the key size k_c and the FMR to be $k_c \leq -\log_2(\alpha(T))$ by using the Hamming bound theorem. Namely, from theorem 6 on Page 19 in MacWilliams and Sloane (1977) [127] (The sphere packing or Hamming bound) states: *A t_c-error binary code of length n_c containing M codewords must satisfy*

$$M\left(1 + \binom{n_c}{1} + \binom{n_c}{2} + \ldots + \binom{n_c}{t}\right) \leq 2^{n_c}. \tag{4.20}$$

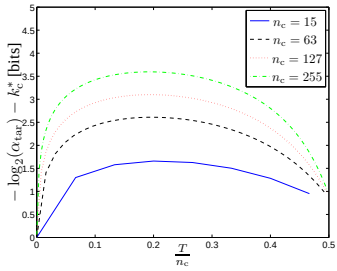


Figure 4.11: The difference $(-\log_2(\alpha_{\mathrm{tar}}) - k_{\mathrm{c}}^*)$ as a function of relative operating point $\frac{T_{tar}}{n_{\mathrm{c}}}$ with n_{c} fixed at different n_{c} settings.

With the FMR defined in (4.10) as $\alpha(T) = 2^{-n_{\mathrm{c}}} \sum_{i=0}^{T} \binom{n_{\mathrm{c}}}{T}$ with $t_{\mathrm{c}} = T$ and $M = 2^{k_{\mathrm{c}}}$, we obtain

$$\begin{aligned} k_{\mathrm{c}} &\leq -\log_2(\alpha(T)) \\ &\leq -\log_2(\alpha_{\mathrm{tar}}), \text{ with } T = T_{tar}, \end{aligned} \tag{4.21}$$

where we define the FMR at the target operating point T_{tar} as α_{tar}. Thus, we have two upper bounds for the key size, namely $\log_2(\alpha_{\mathrm{tar}})$ from the Hamming bound theorem from (4.21) and k_{c}^* from Shannon's theorem from (4.19). We compare the difference between the two bounds $(-\log_2(\alpha_{\mathrm{tar}}) - k_{\mathrm{c}}^*)$ as a function of the relative operating point $\frac{T}{n_{\mathrm{c}}}$ at a fixed number of components n_{c}, as illustrated in Figure 4.11 for different n_{c} settings. We observe that if no errors have to be corrected, $T = 0$, then there is no difference because $(-\log_2(\alpha_{\mathrm{tar}}) - k_{\mathrm{c}}^*) = 0$. However, if errors have to be corrected we observed a difference, where its maximum is around $\frac{T_{tar}}{n_{\mathrm{c}}} = 0.2$. A larger maximum is observed for larger n_{c} values. A rule of thumb we could use is that the maximum difference increases with 0.5 bits when doubling n_{c}.

Hence, $-\log_2(\alpha_{\mathrm{tar}})$ is an upper bound of the key size k_{c} at the target operating point, however it is larger than the maximum key size k_{c}^* upper bound derived from Shannon's bound when errors have to be corrected. Furthermore, the difference between the two bounds increases when there are more components.

4.2.5 Numerical Analysis of the System Performance and the Maximum Key Size

By means of a numerical analysis we illustrate the effect of the system parameters on both the system performance and the theoretical maximum key size k_{c}^*. As the system parameters we have the input capacity C_{in}, the number of components n_{c}, the number of enrolment N_{e} and verification N_{v} samples, the target FNMR β_{tar}, and the target FMR α_{tar}. In Section 4.2.5 we analyze the case where the input capacity C_{in} is uniformly distributed among the n_{c} components leading to feature components with an equal feature quality. On the other hand, we consider the case where the components have unequal

feature qualities due to the non-uniformly distribution of C_{in} in Section 4.2.5. In Section 4.2.5 we conclude with the analysis of dependencies between the feature components.

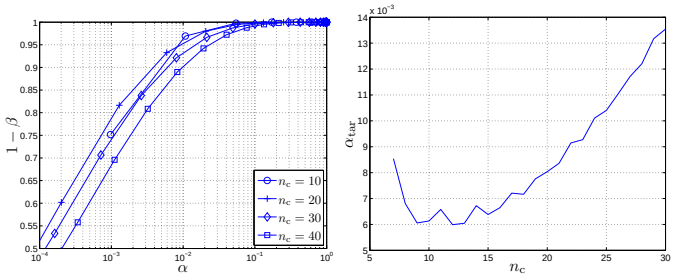
Biometric Source with Equal Feature Quality

System Performance First, we illustrate the effect of equally distributing a fixed input capacity C_{in} among a different number of components n_{c} on the system performance, hence it holds that the Gaussian channel capacity for equal component is equal to $C_{\mathrm{G}}[j] = \frac{C_{\mathrm{in}}}{n_{\mathrm{c}}}$. We present the system performance with the receiver operating characteristics (ROC) curve, where the detection rate 1-β is displayed as a function of the FMR α. For different settings of n_{c}, Figure 4.12(a) portrays the ROC curves obtained for the case when $C_{\mathrm{in}} = 50$ bits and Figure 4.12(b) depicts the obtained FMR α_{tar} at the target FNMR $\beta_{\mathrm{tar}} = 5\%$. The target FMR is computed by first estimating the operating point T_{tar} defined by (4.18), then determining the FMR on that operating point. Note that the operating point determines the number of errors that have to be corrected and can only be integer numbers. The target FNMR will most likely not be achieved precisely on an integer number, and some interpolation method has to be used in order to obtain a more precise (fractional) operating point. A simple linear interpolation between the closed two integer values can be used, however we have noticed that the obtained results such as in Figure 4.12(b) may have a high degree of fluctuations. Therefore, we take the linear interpolation of the logarithm of both the FMR and FNMR in order to have a more accurate estimate.

Figure 4.12 shows that the system performance depends on n_{c}. If n_{c} is too large or too small the performance deteriorates and hence is not optimal. At smaller n_{c} values, the genuine bit-error probability $P_{\mathrm{e}}^{\mathrm{ge}}$ will be smaller, because the capacity per component increases due to the fixed input capacity assumption. However, the number of subjects that can be distinguished reduces as well. In a perfect system where $P_{\mathrm{e}}^{\mathrm{ge}} = 0$, it is only possible to distinguish $2^{n_{\mathrm{c}}}$ subjects without any errors. As a consequence, significantly decreasing n_{c} will degrade the system performance. On the other hand, at larger n_{c} values it is possible to distinguish more subjects, but $P_{\mathrm{e}}^{\mathrm{ge}}$ increases due to the overall constant input capacity C_{in} leading to a system performance deterioration. Consequently, for each $\{C_{\mathrm{in}}, N_{\mathrm{e}}, N_{\mathrm{v}}\}$ setting we determine the optimal number of components n_{c}^* leading to the minimum α_{tar} and use the corresponding ROC curve for comparison between different settings.

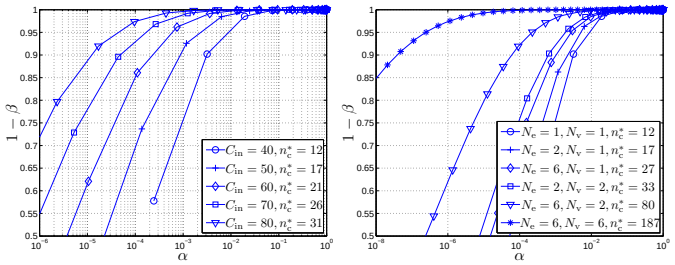
With the optimal number of components n_{c}^* determined, Figure 4.13(a) depicts the ROC curve at different input capacity C_{in} settings, while Figure 4.13(b) shows the ROC curve at different number of enrolment N_{e} and verification N_{v} settings with $C_{\mathrm{in}} = 40$ bits. The figures show that the ROC improves when either increasing C_{in}, N_{e}, or N_{v}. The most significant performance improvement is obtained when increasing both N_{e} and N_{v}.

Maximum Key Size with Fixed Number of Components Having shown the effect of the $\{C_{\mathrm{in}}, n_{\mathrm{c}}, N_{\mathrm{e}}, N_{\mathrm{v}}\}$ parameters on the ROC performance curve, we will now illustrate the effect of the $\{\beta_{\mathrm{tar}}, N_{\mathrm{e}}, N_{\mathrm{v}}\}$ parameters on the maximum key size k_{c}^* and the relative operating point $\frac{T_{tar}}{n_{\mathrm{c}}}$ given a fixed input capacity $C_{\mathrm{in}} = 40$ bits uniformly distributed among $n_{\mathrm{c}} = 12$ components. The influence of the target FNMR β_{tar} is shown



(a) (b)

Figure 4.12: For different number of components n_c, (a) the obtained ROC curves with an input capacity of $C_{in} = 40$ bits and a single of enrolment and verification sample, and (b) the obtained FMR α_{tar} at the target FNMR $\beta_{tar} = 5\%$.



(a) (b)

Figure 4.13: (a) ROC curve for the input capacity at $C_{in} \in \{40, 50, 60, 70, 80\}$ and (b) the ROC curve for different number of enrolment N_e and verification N_v samples with input capacity $C_{in} = 40$ bits. In both cases the optimal number of components n_c^* is determined and used for each settings, and the target FNMR is at $\beta_{tar} = 5\%$.

in Figure 4.14(a) for the maximum key size and in Figure 4.14(c) for the relative operating point. Increasing β_{tar} decreases the number of components to be corrected, hence decreasing the relative operating point and therefore increasing the maximum key size. The maximum key size converges towards the number of components n_c. Increasing the number of samples reduces the bit-error probability and therefore decreasing the relative operating point and consequently increasing the maximum key size. Increasing both the number of enrolment and verification samples has a greater increase on the maximum key size which converges towards n_c.

Maximum Key Size with Optimal Number of Components In Section 4.2.5 we have shown the effect of the $\{\beta_{tar}, N_e, N_v\}$ parameters on the maximum key size k_c^* where we considered the input capacity C_{in} and the number of components n_c to be fixed. We

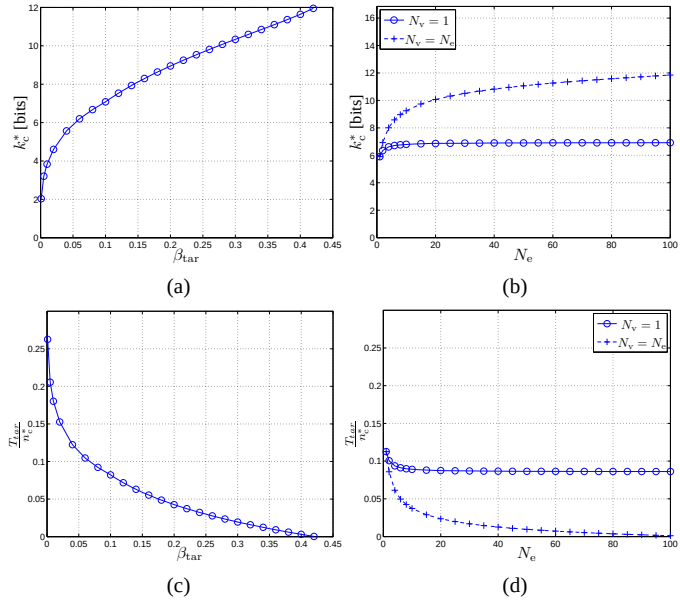


Figure 4.14: The influence of the target FNMR β_{tar} on the (a) maximum key size k_{c}^* and (c) the relative operating point $\frac{T_{tar}}{n_{\text{c}}}$, and similarly the influence of the number of enrolment N_{e} or verification N_{v} samples in (b) and (d), respectively.

will now illustrate the effect of the $\{C_{\text{in}}, \beta_{\text{tar}} N_{\text{e}}, N_{\text{v}}\}$ parameters on the maximum key size k_{c}^* considering the case where the optimal number of components n_{c}^* is determined given the parameter setting as discussed in Section 4.2.5. First, we discus the effect of $\{C_{\text{in}}, \beta_{\text{tar}}\}$ followed by the effect of $\{C_{\text{in}}, N_{\text{e}}, N_{\text{v}}\}$.

Figure 4.15(a)(b) portray the effect of the target FNMR β_{tar} and the input capacity C_{in} on the maximum key size k_{c}^* with a single enrolment and verification sample $N_{\text{e}} = N_{\text{v}} = 1$, where Figure 4.15(a) depicts k_{c}^* as a function of C_{in} with different β_{tar} settings and Figure 4.15(b) shows k_{c}^* as a function of β_{tar} with different C_{in} settings. Similarly, the effect of β_{tar} and C_{in} on the relative operating point $\frac{T_{tar}}{n_{\text{c}}^*}$ and the optimal number of components n_{c}^* are illustrated in Figure 4.15(c)(d) and Figure 4.15(e)(f), respectively. The results show that increasing either the input capacity C_{in} or the target FNMR β_{tar} increases the maximum key size k_{c}^* and the optimal number of components n_{c}^*, but decreases the relative operating point $\frac{T_{tar}}{n_{\text{c}}^*}$. Both the increase of n_{c}^* and the decrease of $\frac{T_{tar}}{n_{\text{c}}^*}$ have a positive effect on the maximum key size k_{c}^*. Doubling β_{tar} from 10% to 20% on average adds around 2 bits to k_{c}^*, but from 2.5% to 5% on average adds 1 bit. Furthermore, doubling C_{in} roughly doubles k_{c}^* for the case when $\beta_{\text{tar}} = 20\%$ and almost triples for the case when $\beta_{\text{tar}} = 2.5\%$. Also, Figure 4.15(b) shows that if β_{tar} is small, namely $\leq 5\%$, there is a significant drop of k_{c}^* when β_{tar} decreases further. At smaller β_{tar} it is required to correct more bits (as shown in Figure 4.15(c) by the increase

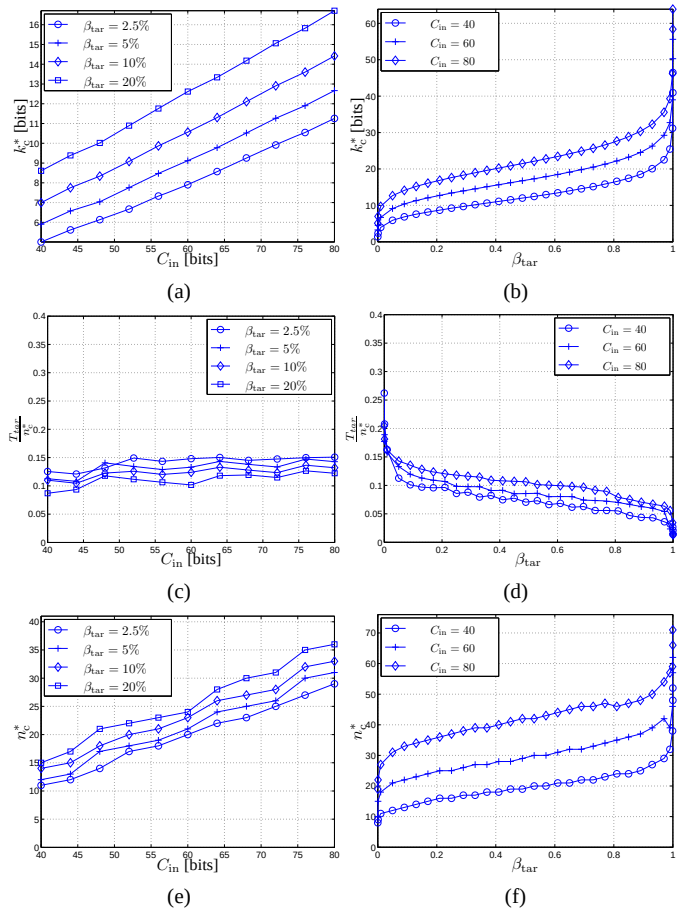


Figure 4.15: Sub-figures (a)(c)(e) depict the maximum key size $k_{\rm c}^*$, the relative targeted operating point $\frac{T_{tar}}{n_{\rm c}}$, and the optimal number of components $n_{\rm c}^*$ as a function of the input capacity $C_{\rm in}$ at different target FNMR $\beta_{\rm tar}$ settings, respectively. Similarly, (b)(d)(f) depict them as function of $\beta_{\rm tar}$ with different $C_{\rm in}$ settings.

in $\frac{T_{tar}}{n_{\rm c}}$), hence it is important to extract bits with smaller bit-error probabilities $P_{\rm e}^{\rm ge}[j]$. Therefore, at a fixed $C_{\rm in}$, there have to be less components in order for each component to have a better feature quality $\frac{\sigma_{\rm b}}{\sigma_{\rm w}}$ or Gaussian capacity $C_{\rm G}[j]$ leading to a smaller $P_{\rm e}^{\rm ge}[j]$. On the contrary, when $\beta_{\rm tar}$ is close to 1, there is a significant increase in $k_{\rm c}^*$. If $\beta_{\rm tar}$ converges to 1, $k_{\rm c}^*$ goes to infinity. In this case, because of the large target FNMR it is not necessary to correct many bits with its extreme case where no bits at all have to be corrected. Hence, many components (see Figure 4.15(f)) can be extracted with a worse feature quality or a smaller $C_{\rm G}[j]$.

Figure 4.16 depicts the effect of the $\{N_{\rm e}, N_{\rm v}, C_{\rm in}\}$ parameters on the maximum key

size k_{c}^*, the relative operating point $\frac{T_{tar}}{n_{\mathrm{c}}}$, and the optimal number of components n_{c}^*. The effect of the input capacity C_{in} is similar as illustrated in Figure 4.15(a). Furthermore, increasing either the number of enrolment N_{e} or verification N_{v} samples leads to an increase of k_{c}^*. However, keeping either N_{e} or N_{v} fixed while increasing the other shows that k_{c}^* increases asymptotically and is limited (see Figure 4.16(b)). Changing both N_{e} and N_{v} significantly increase k_{c}^*. In general, increasing the number of samples enables the use of components with a worse feature quality, hence increasing the optimal number of components n_{c}^* when the input capacity C_{in} is fixed. Consequently, the relative operating point $\frac{T_{tar}}{n_{\mathrm{c}}^*}$ increases because of the lower quality leading to a larger bit-error probability. A larger $\frac{T_{tar}}{n_{\mathrm{c}}^*}$ leads to a smaller channel capacity and therefore a smaller possible key size. However, the optimal number of components increases stronger leading to a net increase of the maximum key size k_{c}^*.

Some examples of the maximum key size increase are as follows. Within the specific range of target FNMR $2.5\% \leq \beta_{\mathrm{tar}} \leq 20\%$ and the input capacity $40 \leq C_{\mathrm{in}} \leq 80$, doubling the target FMR adds 1 to 2 bits to the maximum keys size k_{c}^*. Doubling the input capacity C_{in} doubles the maximum key size k_{c}^* when $\beta_{\mathrm{tar}} = 20\%$ and almost triples when $\beta_{\mathrm{tar}} = 2.5\%$. Furthermore, for the case where the target FNMR is at $\beta_{\mathrm{tar}} = 5\%$, increasing the number of enrolment samples N_{e} from one to six samples increases the maximum key size k_{c}^* with 0.6 bits (from 5.9 to 6.5) at $C_{\mathrm{in}} = 40$ bits and 2.9 bits (from 12.7 to 15.6) bits at $C_{\mathrm{in}} = 80$ bits. Keeping $N_{\mathrm{e}} = 6$ and increasing the number of verification samples N_{v} from one to two samples increases k_{c}^* with 3.0 bits at $C_{\mathrm{in}} = 40$ and 7.6 bits at $C_{\mathrm{in}} = 80$ bits. A further increase of N_{v} from two to six samples increases k_{c}^* with 9.3 bits at $C_{\mathrm{in}} = 40$ and 20.8 bits at $C_{\mathrm{in}} = 80$ bits.

Determining the Optimal Number of Components In Section 4.2.5 we determined the optimal number of components n_{c}^* based on the performance, namely the smallest FMR α_{tar} at the target FNMR β_{tar}. However, the actual goal is to determine the optimal number of components that leads to the maximum key size k_{c}^*. In this section we analyze the difference between the two optimization methods.

In order to investigate the differences between the two optimization methods, Figure 4.17 depicts k_{c}^* as function of $-\log_2(\alpha_{\mathrm{tar}})$ at different number of components n_{c} and the number of enrolment N_{e} and verification N_{v} samples. Note that a smaller α_{tar} leads to a larger $-\log_2(\alpha_{\mathrm{tar}})$. We vary the number of components n_{c} and compute $-\log_2(\alpha_{\mathrm{tar}})$ and k_{c}^* at $\beta_{\mathrm{tar}} = 5\%$ with the input capacity $C_{\mathrm{in}} = 50$ bits. Small n_{c} values correspond to the upper-left of the curves. When increasing n_{c}, the curve follows the upper-right direction and after reaching the upper-right corner it goes down to the lower-left. Figure 4.17 shows that both $-\log_2(\alpha_{\mathrm{tar}})$ and k_{c}^* are close to their largest value at roughly the same n_{c} value. Thus, we can conclude that the optimal number of components n_{c}^* determined by either the smallest α_{tar} or largest k_{c}^* are similar with some margin of error, however the margin of error decreases at larger number of components. Note the difference between $-\log_2(\alpha_{\mathrm{tar}})$ and k_{c}^* at n_{c}^* in Figure 4.17. This difference has been discussed in Section 4.2.4

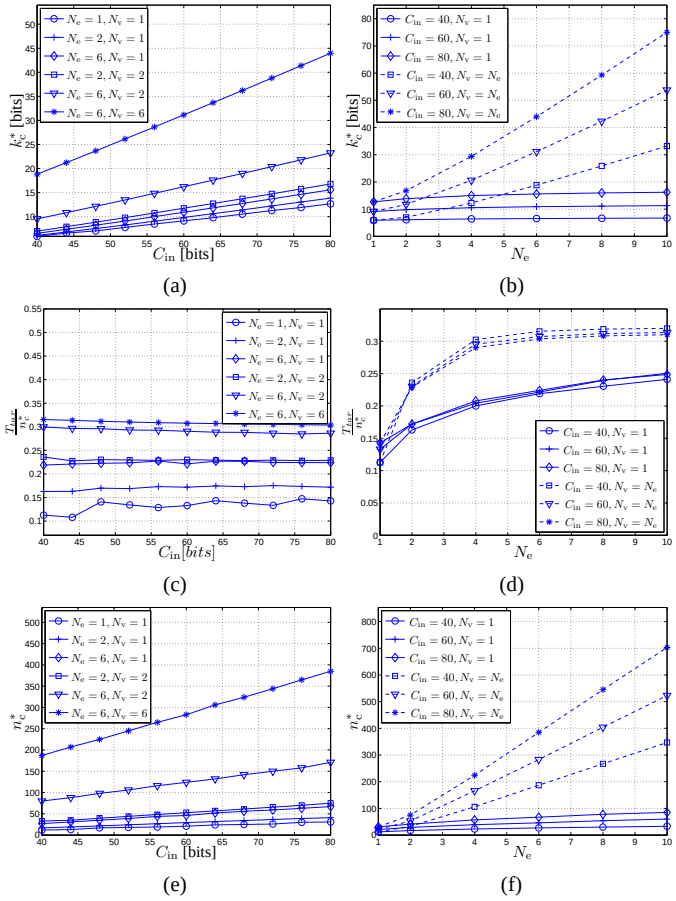


Figure 4.16: Sub-figures (a)(c)(e) depict the maximum key size k_{c}^*, the relative targeted operating point $\frac{T_{tar}}{n_{\mathrm{c}}^*}$, and the number of components n_{c}^* as a function of input capacity C_{in} at different $\{N_{\mathrm{e}}, N_{\mathrm{v}}\}$ settings, respectively. Similarly, (b)(d)(f) depict them as a function of $\{N_{\mathrm{e}}, N_{\mathrm{v}}\}$ with different C_{in} settings. In all cases we have $\beta_{\mathrm{tar}} = 5\%$.

Optimal Parameter Settings and Key Extraction Efficiency at the Target Performance In the previous sections we considered the input capacity C_{in} to be given, while we determined the optimal number of components that leads to the best performance. We are now interested in the different possible settings of the number of components n_{c} and the input capacity C_{in} that can reach the target performance defined by both the target FMR α_{tar} and target FNMR β_{tar}. The results are shown in Figure 4.18 for the case of $\beta_{\mathrm{tar}} = 5 \times 10^{-2}$ and $\alpha_{\mathrm{tar}} = 1 \times 10^{-5}$. Figure 4.18(a) illustrates the optimal Gaussian channel capacity C_{G} at the target performance given the number of components n_{c}. We can conclude that if n_{c} increases, smaller Gaussian channel capacities C_{G}, i.e. lower

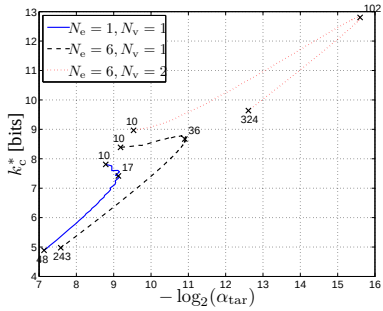


Figure 4.17: The FMR $-\log_2(\alpha_{\mathrm{tar}})$ and the maximum key size k_{c}^* estimated at the target FNMR $\beta_{\mathrm{tar}} = 5\%$ at different number of feature components n_{c} and number of enrolment samples $\{N_{\mathrm{e}}, N_{\mathrm{v}}\}$ with input capacity $C_{\mathrm{in}} = 50$ bits. For each curve we indicate with $\times$ the cases where we have the optimal number of components at the smallest α_{tar}, and where we have the smallest or largest number of components under consideration.

quality feature components, are necessary in order to reach the same target performance. Similarly, Figure 4.18(c) shows that the maximum key size k_{c}^* decreases at larger n_{c} with a difference of 2.5 bits between the maximum k_{c}^* achieved at $n_{\mathrm{c}} = \lceil -\log_2(\alpha_{\mathrm{tar}}) \rceil = 17$ and $n_{\mathrm{c}} = 100$. When $n_{\mathrm{c}} \leq \lceil -\log_2(\alpha_{\mathrm{tar}}) \rceil$ it is not possible to reach the target FMR. Thus, increasing n_{c} larger than $\lceil -\log_2(\alpha_{\mathrm{tar}}) \rceil$ only decreases k_{c}^*. Furthermore, from Figure 4.18(b) we observe that the input capacity has a minimum at $n_{\mathrm{c}} = 34$. Hence, when analyzing the key extraction efficiency defined by the ratio $\frac{k_{\mathrm{c}}^*}{C_{\mathrm{in}}}$ and portrayed in Figure 4.18(d) we observe a maximum at $n_{\mathrm{c}} = 30$, where 16.2% of the input capacity is extracted as the key.

Similar to the results of Figure 4.18, we analyze the influence of the target FMR α_{tar} in Figure 4.19(a)(b)(c)(d), the target FNMR β_{tar} in Figure 4.19(e)(f)(g)(h), and the number of enrolment N_{e} and verification N_{v} samples in Figure 4.19(i)(j)(k)(l). Because of the relationship $n_{\mathrm{c}} \geq \lceil -\log_2(\alpha_{\mathrm{tar}}) \rceil$, smaller α_{tar} values increase the minimum number of components and also the maximum key size k_{c}^* due to a similar relationship as shown in Section 4.2.4. Furthermore, the Gaussian channel capacity C_{G} per component increases, consequently also the input capacity C_{in}. Finally, the key extraction efficiency also increases at smaller α_{tar} values while its optimum shifts towards larger n_{c} values. At the smallest setting of $\alpha_{\mathrm{tar}} = 10^{-8}$ the key extraction efficiency is around 17.2% with the optimum at $n_{\mathrm{c}} = 54$. Decreasing the target FNMR β_{tar} also increases the feature quality and input capacity requirement but does not increase k_{c}^*. Consequently, the key extraction efficiency decreases at smaller β_{tar} and its optimum slightly shifts towards larger n_{c} values. Increasing either N_{e} or N_{v} reduces the requirement on the feature quality and the input capacity, while k_{c}^* is kept unchanged. Thus, the key extraction efficiency improves when increasing either N_{e} or N_{v} and the optimum is obtained at larger n_{c} values.

We can conclude that there is a trade-off between the maximum key size and the key extraction efficiency. On the one hand all the optimal key extraction efficiency is obtained

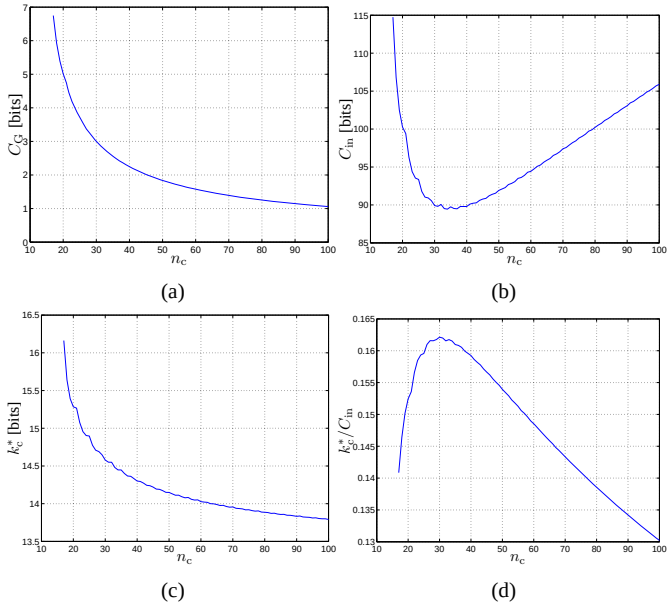


Figure 4.18: As function of the number of feature components, we present the requirement in terms of (a) the Gaussian channel capacity of each component C_{G} and (b) the input capacity C_{in} to reach the target performance given by the target FMR $\alpha_{\mathrm{tar}} = 1 \times 10^{-5}$ and the target FNMR $\beta_{\mathrm{tar}} = 5 \times 10^{-2}$ with a single enrolment and verification sample ($N_{\mathrm{e}} = N_{\mathrm{v}} = 1$). As output we show (c) the maximum key size k_{c}^* and (d) the key extraction efficiency defined by the ratio $\frac{k_{\mathrm{c}}^*}{C_{\mathrm{in}}}$.

only when the optimal number of components is used and can be further increased by decreasing the target FMR or increasing either the target FNMR or the number of enrolment or verification samples. On the other hand, given a target FMR and the number of components leading to the optimal key extraction efficiency, the maximum key size can be further improved by only decreasing the number of components. By decreasing the number of components, the feature quality has to increase as such that the input capacity also increases. The input capacity increase is greater than the maximum key size improvement that therefore reducing the key size efficiency.

Biometric Source with Unequal Feature Quality

In Section 4.2.5 we discussed the special case where the input capacity C_{in} is uniformly distributed among the n_{c} components, hence leading to components with an equal feature quality $\frac{\sigma_{\mathrm{b}}}{\sigma_{\mathrm{w}}}$ and Gaussian channel capacity C_{G}. In practice, however, this scenario would be unlikely to occur. Therefore, in this section we consider the case where C_{in} is non-uniformly distributed and thus leading to components with unequal feature qualities. The

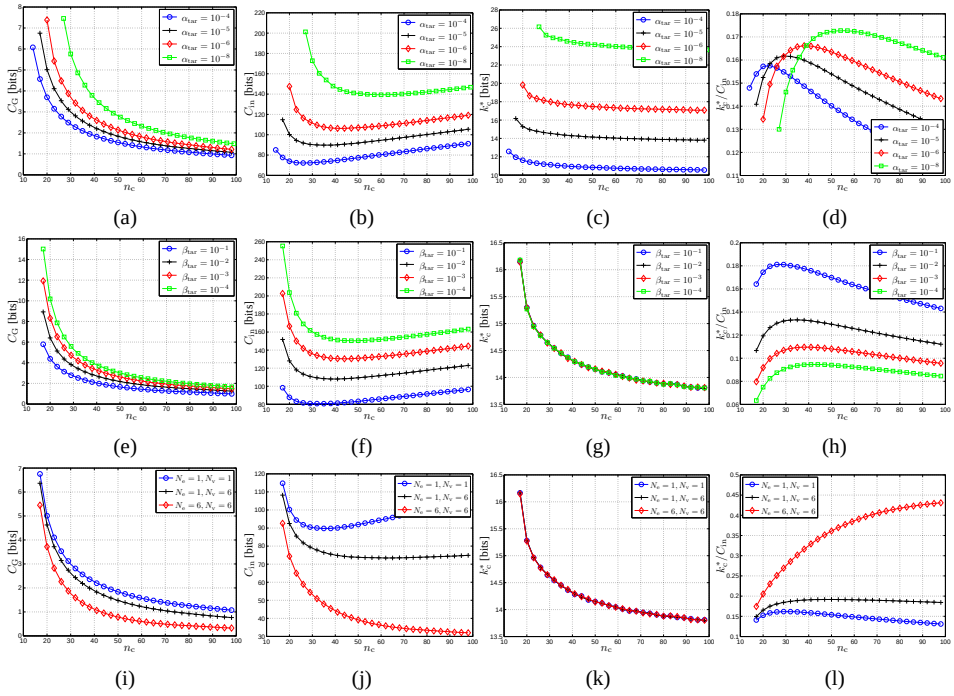
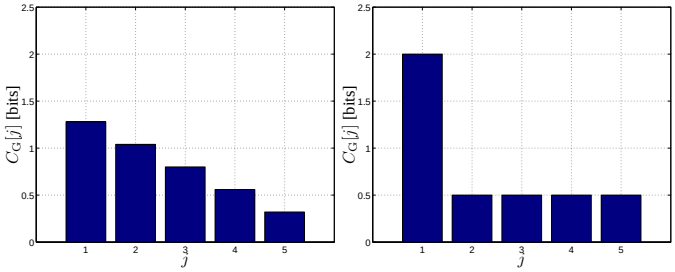


Figure 4.19: As function of the number of feature components, we present the requirement to reach the target performance in terms of (a)(e)(i) the Gaussian channel capacity of each component C_{G} and (b)(f)(j) the input capacity C_{in} with the output (c)(g)(k) the maximum key size k_{c}^{*}, and (d)(h)(l) the key extraction efficiency defined by the ratio $\frac{k_{\mathrm{c}}^{*}}{C_{\mathrm{in}}}$. This analysis is presented for different target FMR α_{tar} in (a)(b)(c)(d) with $\beta_{\mathrm{tar}} = 5 \times 10^{-2}$ and a single enrolment and verification sample ($N_{\mathrm{e}} = N_{\mathrm{v}} = 1$), different target FNMR β_{tar} in (e)(f)(g)(h) with $\alpha_{\mathrm{tar}} = 1 \times 10^{-5}$ and $N_{\mathrm{e}} = N_{\mathrm{v}} = 1$, and different number of enrolment N_{e} and verification N_{v} samples in (i)(j)(k)(l) with $\alpha_{\mathrm{tar}} = 1 \times 10^{-5}$ and $\beta_{\mathrm{tar}} = 5 \times 10^{-2}$.

main requirement is that the sum of the Gaussian channel capacity among the n_{c} components is equal to the input capacity C_{in}, namely $C_{\mathrm{in}} = \sum_{j=1}^{n_{\mathrm{c}}} C_{\mathrm{G}}[j]$. We consider the two cases of non-uniformly distribution as portrayed in Figure 4.20. In the first case, case 1, the first component has the largest Gaussian channel capacity C_{G}, while the following components have a linearly decreasing capacity. We define the non-uniformity ratio $r = \frac{C_{\mathrm{G}}[1]}{C_{\mathrm{G}}[n_{\mathrm{c}}]} \geq 1$ and together with $C_{\mathrm{in}} = \sum_{j=1}^{n_{\mathrm{c}}} C_{\mathrm{G}}[j]$ fully define the capacity of each



(a) case 1 (b) case 2

Figure 4.20: The two cases we consider where C_{in} is non-uniformly distributed among the n_{c} components leading to components with unequal capacity $C_{\mathrm{G}}[j]$. (a) Depicts the first case , case 1, where the first component has the largest capacity and the following components have a linearly decreasing capacity and (b) illustrates the second case , case 2 where only the first component has an increased capacity. For both cases we define the non-uniformity as the ratio $r = \frac{C_{\mathrm{G}}[1]}{C_{\mathrm{G}}[n_{\mathrm{c}}]}$ which together with $C_{\mathrm{in}} = \sum_{j=1}^{n_{\mathrm{c}}} C_{\mathrm{G}}[j]$ fully define the non-uniformity. For both cases we have $C_{\mathrm{in}} = 4$, $n_{\mathrm{c}} = 5$, and $r = 4$.

component, namely

$$\begin{aligned} C_{\mathrm{G}}[1] &= \frac{r(n_{\mathrm{c}}-1)C_{\mathrm{in}}}{(n_{\mathrm{c}}-1)n_{\mathrm{c}}+(r-1)\sum_{j=1}^{n_{\mathrm{c}}-1} j}, \\ C_{\mathrm{G}}[n_{\mathrm{c}}] &= \frac{C_{\mathrm{G}}[1]}{r}, \\ C_{\mathrm{G}}[j] &= C_{\mathrm{G}}[1] - (j-1)\frac{C_{\mathrm{G}}[1]-C_{\mathrm{G}}[n_{\mathrm{c}}]}{n_{\mathrm{c}}-1}, \text{ for } 2 \leq j \leq n_{\mathrm{c}}-1. \end{aligned} \tag{4.22}$$

For the second case under consideration, case 2, the first component has the largest capacity, while the other $n_{\mathrm{c}} - 1$ components have an equal but r times smaller capacity as depicted in Figure 4.20(b). Again we have the non-uniform ratio $r = \frac{C_{\mathrm{G}}[1]}{C_{\mathrm{G}}[n_{\mathrm{c}}]}$ and together with $C_{\mathrm{in}} = \sum_{j=1}^{n_{\mathrm{c}}} C_{\mathrm{G}}[j]$ we obtain

$$\begin{aligned} C_{\mathrm{G}}[1] &= \frac{rC_{\mathrm{in}}}{n_{\mathrm{c}}-1+r}, \\ C_{\mathrm{G}}[j] &= \frac{C_{\mathrm{G}}[1]}{r}, \text{ for } 2 \leq j \leq n_{\mathrm{c}}. \end{aligned} \tag{4.23}$$

Note that for both non-uniform cases we can obtain the uniform case by setting $r = 1$.

Numerical analysis of the two non-uniform cases are portrayed in Figure 4.21 showing the maximum key size k_{c}^* and the relative operating point $\frac{T_{tar}}{n_{\mathrm{c}}}$ as a function of non-uniformity ratio r for different settings of the input capacity C_{in} and the number of enrolment N_{e} and verification N_{v} samples. We keep the number of components fixed instead of using the optimal n_{c}^*. The results show that both non-uniform cases have a smaller k_{c}^* than the uniform case with $r = 1$, because increasing r decreases k_{c}^* for both non-uniform cases. At small r values, the maximum key size k_{c}^* is larger for the case 2 cases, however at larger r values the decrease of k_{c}^* continues for case 2, while k_{c}^* for case 1 stabilizes.

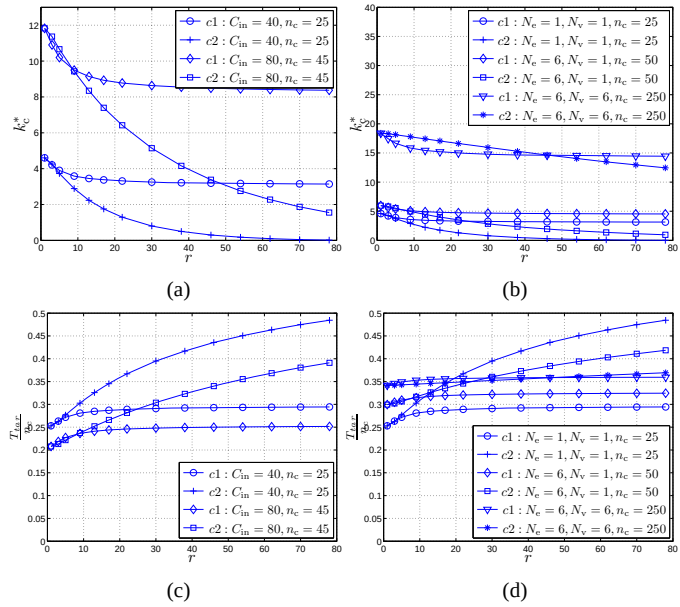


Figure 4.21: The influence of the input capacity $C_{\rm in}$ being non-uniformly distributed among the $n_{\rm c}$ components for the non-uniformity cases, where case 1 and case 2 are indicated by 'c1' and 'c2', respectively. For the sub-figures (a)(c) we have the settings $C_{\rm in} = \{40, 80\}$ bits with a fixed number of components $n_{\rm c} = \{25, 45\}$, respectively, and for sub-figures (b)(d) we change the number of enrolment $N_{\rm e}$ and verification $N_{\rm v}$ samples with $C_{\rm in} = 40$ bits and a fixed $n_{\rm c}$ for each $\{N_{\rm e}, N_{\rm v}\}$ setting.

Consequently, the case 1 has the largest $k_{\rm c}^*$ at larger non-uniformity r. This crossing point where $k_{\rm c}^*$ is equal for both cases increases when increasing either $C_{\rm in}$, $N_{\rm e}$, or $N_{\rm v}$.

The main conclusion of these results is the that the maximum key size $k_{\rm c}^*$ is larger when the feature components have an equal feature quality than when they are unequal. Thus, any deviation from uniformity is suboptimal. Note that we considered an ECC that does not exploit the prior knowledge of bits having different bit-error probabilities.

Biometric Source with Dependent Feature Components

Until now we have assumed the extracted feature vector components and the channel noise to be independent across components and measurements. However, in practice the components may be dependent and therefore may influence the performance and the maximum key size $k_{\rm c}^*$. We only consider the extreme cases where a number of components are fully dependent, because a detailed analysis of the dependencies is beyond the scope of this work. Consider a feature vector with $N_{\rm F}$ components. We assume that the first n_ρ components have in addition κ_ρ components that are fully dependent

(duplicate or identical components), while the remaining $n_{\overline{\rho}}$ components have no duplicates. Hence, it holds that $N_{\mathrm{F}} = n_\rho + n_{\overline{\rho}}$ and the total number of components n_{c} is equal to $n_{\mathrm{c}} = n_\rho(\kappa_\rho + 1) + n_{\overline{\rho}}$. Furthermore, we define the array with n zeros as $\mathbb{O}_n = [0_1, 0_2, \ldots, 0_n]$. With the assumed dependency model, the pmf of the number of bit errors ϵ as defined by (4.8) becomes

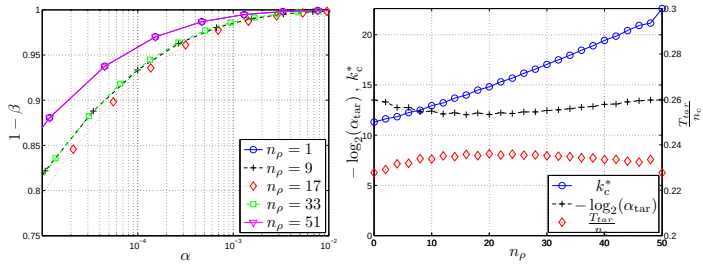
$$\begin{aligned}\phi(\epsilon) &\stackrel{\text{def}}{=} \mathcal{P}\{d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}) = \epsilon\} \\ &= (P_{\rho,1} * P_{\rho,2} * \ldots * P_{\rho,n_\rho} * P_{\overline{\rho},n_\rho+1} * \ldots * P_{\overline{\rho},n_{\mathrm{c}}})(\epsilon),\end{aligned} \tag{4.24}$$

where $P_{\rho,j} = [1 - P_{\mathrm{e}}[j], \mathbb{O}_{\kappa_\rho}, P_{\mathrm{e}}[j]]$ is the marginal pmf of the Hamming distance from the extracted bits from the set of $\kappa_\rho + 1$ identical components for the first n_ρ components and $P_{\overline{\rho},j} = [1 - P_{\mathrm{e}}[j], P_{\mathrm{e}}[j]]$ is the pmf for the extracted bit from the last $n_{\overline{\rho}}$ components without duplicates. For the set of $\kappa_\rho + 1$ identical bits it is only possible to have zero or $\kappa_\rho + 1$ bit errors with probability $1 - P_{\mathrm{e}}$ and P_{e}, respectively. As in the previous sections, we can use the same equations for estimating the performance and the maximum key size at the target FNMR.

The results for the case where $N_{\mathrm{F}} = 50$ with input capacity $C_{\mathrm{in}} = 80$ bits and target FNMR at $\beta_{\mathrm{tar}} = 5\%$ is portrayed in Figure 4.22, where the first n_ρ components have a single duplicate $\kappa_\rho = 1$. The ROC performance curve deteriorates once duplicate components are added as shown in Figure 4.22(a). In other words, the FMR α_{tar} at the target FNMR β_{tar} increases, as illustrated by the decrease of $-\log_2(\alpha_{\mathrm{tar}})$ in Figure 4.22(b). Furthermore, the relative operating point $\frac{T_{tar}}{n_{\mathrm{c}}}$ also increases. Although the increase of $\frac{T_{tar}}{n_{\mathrm{c}}}$ reduces the capacity $C(\frac{T_{tar}}{n_{\mathrm{c}}})$, we observe that the maximum key size k_{c}^* increases due to the increase of n_{c}. However, further increasing n_ρ until each component has κ_ρ duplicates ($n_\rho = N_{\mathrm{F}}$) leads to the same α_{tar} and $\frac{T_{tar}}{n_{\mathrm{c}}}$ as for the case where no components have a duplicate ($n_\rho = 0$). Although the performance is similar, the maximum key size k_{c}^* has doubled.

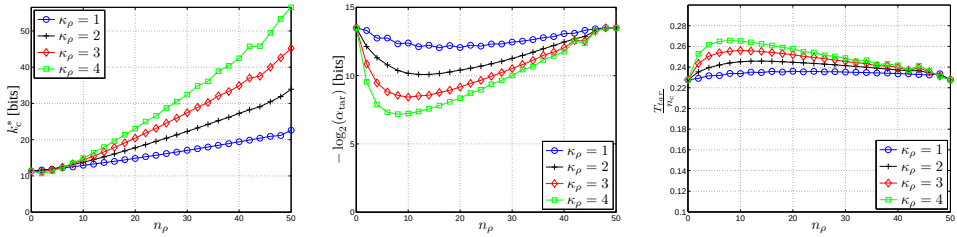
The effects of changing κ_ρ are shown in Figure 4.23. When all feature components have a duplicate, $n_\rho = N_{\mathrm{F}}$, we can see from Figure 4.23(a) that the maximum key size k_{c}^* increases by $(\kappa_\rho + 1)$ when compared to the case where no feature components have a duplicate $n_{\overline{\rho}} = 0$. Furthermore, Figure 4.23(b) shows that the FMR deviation increases when increasing the number of duplicates κ_ρ. Note that the largest FMR, hence the smallest $-\log_2(\alpha_{\mathrm{tar}})$, is achieved at the point where the average Hamming distance from the dependent and independent bits are equal, namely $(\kappa_\rho+1)n_\rho = n_{\overline{\rho}}$. With $n_{\overline{\rho}} = N_{\mathrm{F}} - n_\rho$, we obtain the point $n_\rho = \frac{N_{\mathrm{F}}}{\kappa_\rho+2}$. Not only does κ_ρ influence the FMR at the target FNMR and therefore also the maximum key size k_{c}^*, it also influences the relative operating point $\frac{T_{tar}}{n_{\mathrm{c}}}$, which increases with κ_ρ.

Hence, it seems that the maximum key size k_{c}^* could be increased by adding identical components. However, we argue that the protection actually does not increase because the FMR α_{tar} at the target FNMR β_{tar} is either kept unchanged or even decreases. We also observed in Section 4.2.4 that another upper bound for the key size is $-\log_2(\alpha_{\mathrm{tar}})$, which will no longer hold once identical bits are added by either increasing n_ρ or κ_ρ. This discrepancy between the FMR and the maximum key size is caused by the fact that the ECC is modeled as a Hamming distance classifier that considers each bit to be independent. Hence, the space $\{0,1\}^{n_{\mathrm{c}}}$ is assumed to be fully used and only under this



(a) (b)

Figure 4.22: The (a) performance ROC curve for different n_ρ settings and (b) the maximum key size $k_{\rm c}^*$, the log of the FMR at the operating point $-\log_2(\alpha_{\rm tar})$, and the relative operating point $\frac{T_{tar}}{n_{\rm c}}$ as a function of the number of dependent components n_ρ. For both cases the input capacity is $C_{\rm in} = 80$ bits with the targer FNMR at $\beta_{\rm tar} = 5\%$.



(a) (b) (c)

Figure 4.23: The (a) maximum key size $k_{\rm c}^*$, (b) the log of the FMR at the operating point $-\log_2(\alpha_{\rm tar})$, and the relative operating point $\frac{T_{tar}}{n_{\rm c}}$ as a function of the number of component duplicates κ_ρ.

assumption the maximum key size could be achieved. By adding identical components the space $\{0,1\}^{n_{\rm c}}$ is not fully used, but is reduced to $\{0,1\}^{N_{\rm F}}$. Therefore, the actual maximum key size is smaller and a tighter upperbound would be $-\log_2(\alpha_{\rm tar})$ with its known offset depending on $N_{\rm F}$ as discussed in Section 4.2.4. Note that if the upperbound $-\log_2(\alpha_{\rm tar})$ is taken as the actual maximum key size, the key size will decrease when identical components are added due to the fact that the $\alpha_{\rm tar}$ increases.

We can conclude that by adding multiple κ_ρ identical components to the feature vector the maximum key size can be increased artificially, however the actual protection indicated by the FMR will at most stay equal. The FMR is only kept unchanged when all components have exactly κ_ρ identical components, otherwise the FMR will degrade.

4.2.6 Experiments

By means of numerical analysis, previous sections illustrated the effects of the system parameters such as the number of enrolment N_{e} and verification N_{v} samples on the performance and the maximum key size k_{c}^{*}. In this section we will analyze these findings using an actual biometric database and two feature extraction algorithms.

Biometric Modality and Database

The database we use is the MCYT (Ministerio de Ciencia y Tecnología) containing fingerprint images from a capacitive and optical sensor as described in Ortega-Garcia et al. (2003) [128]. It contains 12 images of all 10 fingers from 330 subjects for each sensor. However, we limit our dataset to only the images of the right-index finger from the optical sensor.

Feature Extraction Algorithms

Two types of texture based features are extracted from a fingerprint, namely *directional field* and *Gabor* features. In order to compensate for possible translations between enrolled and verification measurements, a translation-only pre-alignment step is performed during the feature extraction process. Such pre-alignment requires extraction of the core point which is performed according to the algorithm described in Ignatenko et al. (2002) [129]. Around the core point we define a 17×17 grid with eight pixels between each grid point. The following feature extraction algorithms extract a feature value on each grid point. Our feature extraction algorithm failed to extract a feature vector from one subject, so we excluded it from the dataset, hence there are effectively only $N_{\mathrm{s}} = 329$ subjects.

Direction Field Feature The first feature extraction algorithm is based on directional fields. A directional field vector describes the estimated local ridge-valley edge orientation in a fingerprint structure and is based on gradient vectors. The orientation of the ridge-valley edge is orthogonal to the gradient's angle. Therefore a directional field vector that signifies the orientation of the ridge-valley edge is perpendicular positioned to the gradient vector. In order to extract directional field features from a fingerprint the algorithm described in Gerez and Bazen (2002) [130] is applied on each grid point. The direction field features have a dimension of $N_{\mathrm{F}} = 578$ and are referred to as the DF features.

Gabor Filters Feature The second type of extracted features are the Gabor filters (GF) features, described in Bazen and Veldhuis (2004) [107], where each grid point is filtered using a set of four 2D Gabor filters at angles of $\left\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\right\}$, respectively. The feature vector is the concatenation of the modulus of the four complex responses at each grid point, resulting into a feature vector dimension of $N_{\mathrm{F}} = 1156$.

Dimension Reduction To decorrelate and reduce the number of feature components we use the principle component analysis (PCA) and the linear discriminant analysis (LDA)

techniques, where the LDA transformation is also used to obtain more discriminating feature components. The PCA and LDA transformation matrices are computed using the training set. N_{PCA} is the reduced dimension after applying the PCA transformation and N_{LDA} is the reduced dimension after applying the LDA transformation. We limit N_{LDA} to the number of subjects within the training set from which the transformation matrices are determined.

Testing Protocol

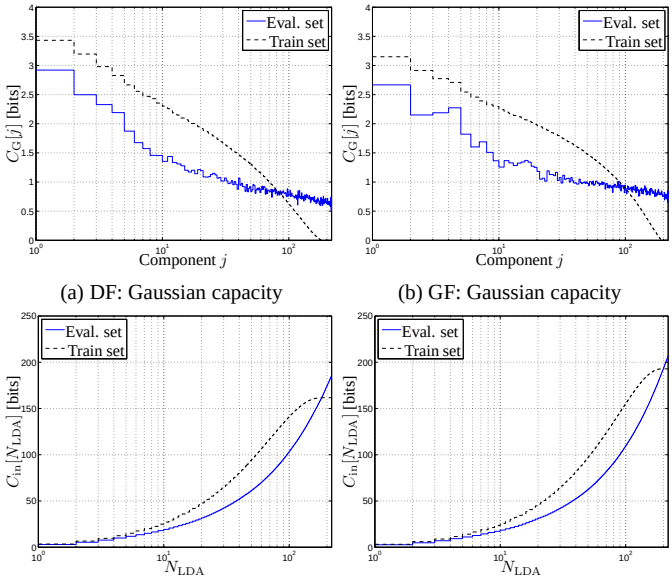
The performance testing protocol consists of randomly selecting 219 out of $N_{\mathrm{s}} = 329$ subjects as the training set and the remaining 110 subjects as the evaluation set, which is referred to as the training-evaluation-set split. The template protection system parameters such as the quantization thresholds used within the *Quantization* module of Figure 4.1 and the PCA and LDA transformation matrices are estimated using the training set.

From the evaluation set, N_{e} samples of each subject are randomly selected as the enrolment samples while the remaining samples are considered as the verification samples. This split is referred to as the enrolment-verification split. The protected template is generated using all the enrolment samples and compared with the average of N_{v} verification samples. When the verification sample is from the same subject as of the protected template, it is referred to as a genuine comparison, otherwise it is an imposter comparison. Note that the number of genuine and imposter comparisons depends on the number of enrolment and verification samples. For the genuine case we have 30250 comparisons for the $N_{\mathrm{e}} = N_{\mathrm{v}} = 1$ case, 16500 for the case of $N_{\mathrm{e}} = 6$ and 2750 comparisons for $N_{\mathrm{e}} = N_{\mathrm{v}} = 6$ case. For the imposter case we have 3297250, 1798500, and 299750 comparisons, respectively.

The training-evaluation-set split is performed five times, while for each of these splits the enrolment-verification split is also performed five times. From each enrolment-verification split we estimate the operating point T_{tar} at the target FNMR β_{tar} and the corresponding FMR α_{tar}. Note that the splits are performed randomly, however the seed at the start of the protocol is always the same, hence all the splits are equal for the performance tests at different settings. Hence, the splitting process does not contribute to any performance differences.

Results

First we determine the Gaussian channel capacity $C_{\mathrm{G}}[j]$, which is indicative for the feature quality, of component j of the feature vector obtained after applying the PCA/LDA transformation. We consider both on the training set and the evaluation set. The capacity for the 218 components are illustrated in Figure 4.24 for both the directional field DF and the Gabor filters GF features indicating that the capacity is not equal for each component. Note that the capacity is greater for the transformed training set than the transformed evaluation set, because the PCA/LDA transformation matrix is determined on the same set and can thus be perfectly trained and the training and evaluation sets are disjunct. This perfect training is also confirmed by the fact that the last components of the training set have a capacity $C_{\mathrm{G}}[j]$ close or equal to zero, while they are larger than zero for the evaluation



(a) DF: Gaussian capacity (b) GF: Gaussian capacity

(c) DF: Input Capacity (d) GF: Input Capacity

Figure 4.24: For both the GF and DF features, (a)(b) illustrate the Gaussian channel capacity $C_{\mathrm{G}}[j]$ of each component from the training set and evaluation set, and (c)(d) the input capacity C_{in} taken as the cumulative sum of $C_{\mathrm{G}}[j]$ of all N_{LDA} components, namely $C_{\mathrm{in}} = \sum_{m=1}^{N_{\mathrm{LDA}}} C_{\mathrm{G}}[m]$.

set. By assuming all components to be independent, we observe that the DF feature has an input capacity $C_{\mathrm{in}} = 162$ bits on the training set and $C_{\mathrm{in}} = 186$ bits on the evaluation set, while $C_{\mathrm{in}} = 193$ and $C_{\mathrm{in}} = 207$ bits for the GF features. Because the capacities are not equally divided, we already know that the achieved performance and maximum key size will be suboptimal.

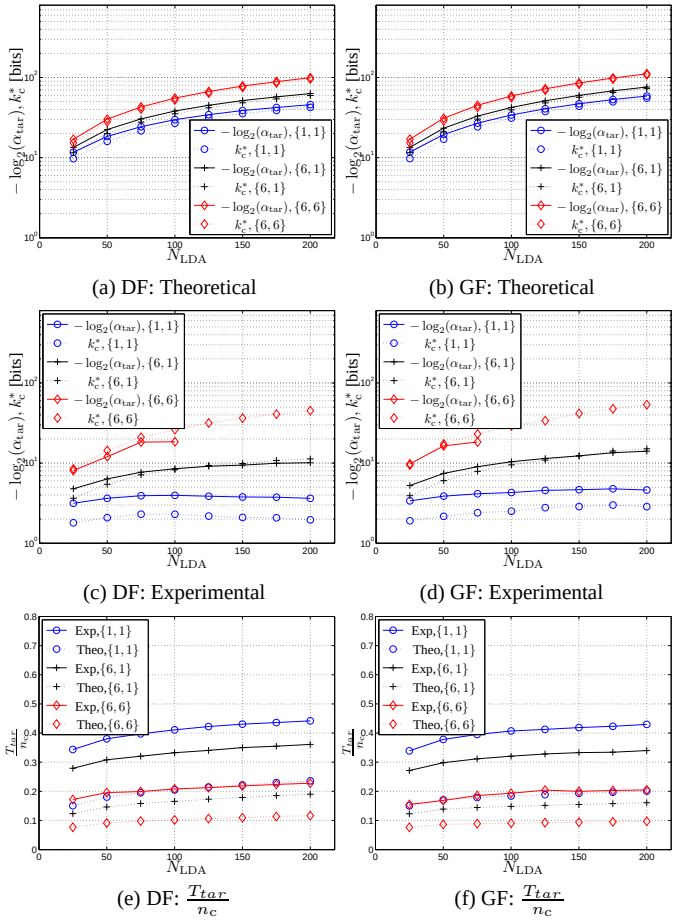
With the known capacity of each component, we can thus compare the maximum key size k_{c}^{*} and the log of the FMR at the target FNMR $-\log_2(\alpha_{\mathrm{tar}})$ from the theoretical performance and the experimental performance. The theoretical performance is obtained using the analytical framework. These results are shown in Figure 4.25 for different number of enrolment N_{e} and verification N_{v} samples for both the DF and GF features. Note that due to the limited number of imposter comparisons, it is not possible to obtain a α_{tar} smaller than $\frac{1}{299750} = 3.3 \times 10^{-6}$ except zero for the experimental case with $N_{\mathrm{e}} = N_{\mathrm{v}} = 6$. From the results we observe four effects. First of all, both the experimental and theoretical results confirm the finding in Section 4.2.5 that the components with a smaller capacity have a greater improvement when more samples are used. For the single enrolment and verification sample case, the experimental results even show that the last components with a much smaller capacity deteriorates the performance and therefore also the maximum key size. However, an improvement is observed when we increase the

number of enrolment samples to $N_{\mathrm{e}} = 6$, and a greater improvement is observed for when we also increase the number of verification samples to $N_{\mathrm{v}} = 6$. Secondly, the results also indicate that the estimated k_{c}^{*} and $-\log_2(\alpha_{\mathrm{tar}})$ are much greater for the theoretical case than for the experimental one. The results in Figure 4.25(e)(f) portray the significant difference between the obtained relative operating point $\frac{T_{tar}}{n_{\mathrm{c}}}$ between the theoretical and experimental cases. This clearly indicates that the FNMR curve is not correctly estimated, namely the target FNMR for the experimental case is at a larger relative operating point than for the theoretical case. As discussed in Kelkboom et al. (2010) [95], estimation errors are introduced by deviations from the made assumptions such as the Gaussian distribution, an equal and independent within-class for each subject, and independent feature components. They proposed a modified analytical framework for relaxing these assumption, however this approach is out of the scope of this work. Thirdly, we observe that the relative difference between the theoretical an experimental results is greater for the $N_{\mathrm{e}} = N_{\mathrm{v}} = 1$ case and decreases when increasing N_{e} and N_{v}. It has also been shown in Kelkboom et al. (2010) [95] that by increasing the number of samples the feature vectors distribution tend to be more Gaussian. Hence, a better Gaussian approximation due to the increase of the number of samples may be the cause behind the improvement of the estimation error. The forth and last difference we observed between the theoretical and experimental results in Figure 4.25(a)(b) and Figure 4.25(c)(d) is the relationship between $-\log_2(\alpha_{\mathrm{tar}})$ and the maximum key size k_{c}^{*}. We have shown in Section 4.2.4 that they are related to each other, namely $k_{\mathrm{c}}^{*} < -\log_2(\alpha_{\mathrm{tar}})$, and this relationship is confirmed by the theoretical case in Figure 4.25(a)(b). However, the results in Figure 4.25(c)(d) show that for the experimental cases $-\log_2(\alpha_{\mathrm{tar}})$ is not always larger than k_{c}^{*}. These deviations are caused by the estimation errors of the FMR curve, leading to an optimistically smaller FMR and thus a larger $-\log_2(\alpha_{\mathrm{tar}})$ at the same operating point.

As discussed in Kelkboom et al. (2010) [95], having dependent feature components has a great influence on the FMR curve estimation. Due to the dependencies, the variance of the relative Hamming distance (the Hamming distance relative to n_{c}) distribution at imposter comparisons is larger than the expected variance of the binomial distribution. Because the variance of the relative Hamming distance is inverse proportional to the dimension, namely $\sigma^2 = \frac{p(1-p)}{N}$, the intrinsic dimension decreases when there is a stronger dependency. Similar as in the work of Daugman (2003) [104], we will estimate the intrinsic dimension by fitting the imposter Hamming distance distribution with a binomial distribution with a dimension smaller than n_{c} and a bit-error probability smaller than $\frac{1}{2}$. Given the relative Hamming distances at each comparison we estimate its variance $\hat{\sigma}^2_{\mathrm{im}}$ and mean $\hat{\mu}_{\mathrm{im}}$, from which we can estimate the new binomial dimensions $\hat{n}_{\mathrm{c}}$ with bit-error probability $\hat{P}_{\mathrm{e}}^{\mathrm{im}}$ as

$$\begin{aligned} \hat{P}_{\mathrm{e}}^{\mathrm{im}} &= \hat{\mu}_{\mathrm{im}}, \\ \hat{n}_{\mathrm{c}} &= \left\lfloor \frac{\hat{P}_{\mathrm{e}}^{\mathrm{im}}(1-\hat{P}_{\mathrm{e}}^{\mathrm{im}})}{\hat{\sigma}^2_{\mathrm{im}}} \right\rfloor . \end{aligned} \tag{4.25}$$

An example of this approximation is shown in Figure 4.26(a) for the pmf of the relative Hamming distances and in Figure 4.26(b) for the FNMR curve. The experimentally obtained curves are indicated with 'Exp', while the original theoretical model curve is indicated with 'Theo', and its corrected version for the intrinsic dimension by 'Theo-cor'. Note that we multiplied the pmf for the 'Theo-cor' case with $\frac{\hat{n}_{\mathrm{c}}}{n_{\mathrm{c}}}$ in order for its area under



(a) DF: Theoretical (b) GF: Theoretical

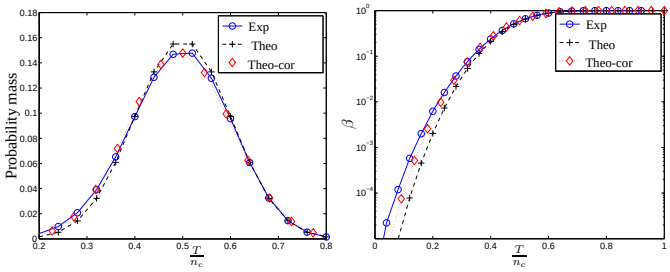
(c) DF: Experimental (d) GF: Experimental

(e) DF: $\frac{T_{tar}}{n_{\rm c}}$ (f) GF: $\frac{T_{tar}}{n_{\rm c}}$

Figure 4.25: The maximum key size $k_{\rm c}^*$, the log of the FMR at the target FNMR $-\log_2(\alpha_{\rm tar})$, and the relative operating point $\frac{T_{tar}}{n_{\rm c}}$ as a function of the LDA dimension $N_{\rm LDA}$ at different $N_{\rm e}$ and $N_{\rm v}$ settings indicated as $\{N_{\rm e}, N_{\rm v}\}$ in the legend. Sub-figures (a)(b) are for the theoretical case for the DF and GF features, respectively, similarly sub-figures (c)(d) are for the experimental case, and (e)(f) are the theoretical and experimental case combined.

the curve to be as large for the other two cases for a fair comparison. From these results we observe that the corrected pmf 'Theo-cor' approximates the experimentally obtained results much better, however the estimation errors are now mainly at the tails of the pmf and thus at the smallest values of the FNMR.

The estimated bit-error probability $\hat{P}_{\rm e}^{\rm im}$ and the intrinsic dimension $\hat{n}_{\rm c}$ at imposter comparisons for different LDA dimensions $N_{\rm LDA}$ and number of enrolment $N_{\rm e}$ or verification $N_{\rm v}$ samples are depicted in Figure 4.27 for both the DF and GF features. Instead



(a) (b)

Figure 4.26: (a) The Hamming distance pmf at imposter comparisons from the experimental case ('Exp'), from the theoretical case ('Theo') and the corrected theoretical case ('Theo-cor') where the experimental data is fitted with a binomial distribution with dimension $\hat{n}_{\mathrm{c}}$ and bit-error probability $\hat{P}_{\mathrm{e}}^{\mathrm{im}}$. Furthermore, (b) shows the corresponding FMR β curve for the three cases in (a).

of the actual estimated intrinsic dimension $\hat{n}_{\mathrm{c}}$ we show the ratio $\frac{\hat{n}_{\mathrm{c}}}{n_{\mathrm{c}}}$. The results from Figure 4.27(a)(b) indicate that when adding more components by increasing N_{LDA}, the relative intrinsic dimension decreases while the bit-error probability converges towards $\frac{1}{2}$. Note that the relative intrinsic dimension also decreases when more samples are used, hence taking the average of N_{e} or N_{v} samples increases the dependencies between the bit errors at imposter comparisons.

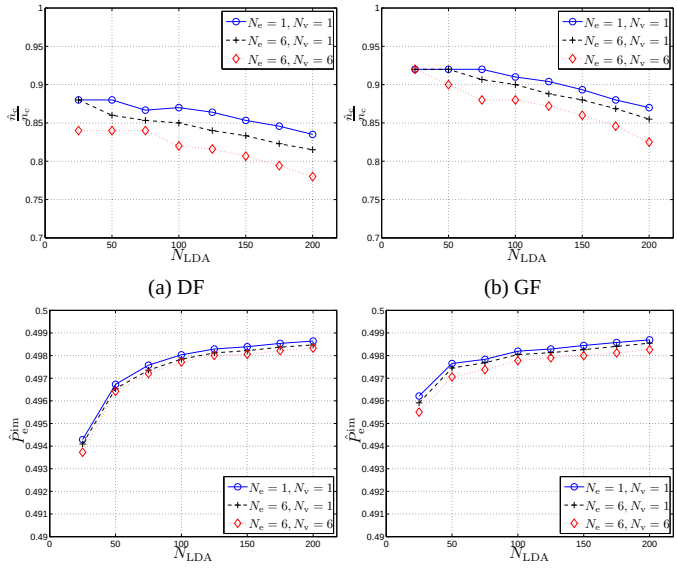
The maximum key size estimation can be improved by incorporating the intrinsic dimension as

$$\begin{aligned} k_{\mathrm{c}}^{*}\text{-cor} &\stackrel{\mathrm{def}}{=} \hat{n}_{\mathrm{c}} C(\tfrac{T_{tar}}{n_{\mathrm{c}}}) \\ &= \tfrac{\hat{n}_{\mathrm{c}}}{n_{\mathrm{c}}} k_{\mathrm{c}}^{*}, \end{aligned} \tag{4.26}$$

where the corrected maximum key size k_{c}^{*}-cor is the relative intrinsic dimension $\frac{\hat{n}_{\mathrm{c}}}{n_{\mathrm{c}}}$ times the original maximum key size k_{c}^{*}. The improved results are illustrated in Figure 4.28. Now also for the $N_{\mathrm{e}} = 6$, $N_{\mathrm{v}} = 1$ case, the corrected maximum key size is always smaller than $-\log_2(\alpha_{\mathrm{tar}})$. The estimation has also improved for the $N_{\mathrm{e}} = 6, N_{\mathrm{v}} = 6$ case, however there are still some deviations, which may be caused by the limited database.

4.2.7 Discussion and Conclusions

The Fuzzy Commitment Scheme is a well known template protection scheme in the literature and is based on a key-binding and key-release mechanism, where the entropy of the key is indicative for the amount of privacy and security. Considering the key to consist out of independent and uniform bits, its entropy is then mainly determined by its size. We have analytically determined the classification performance and the maximum key size of the Fuzzy Commitment Scheme given a Gaussian modeled biometric source, a single bit extraction quantization scheme, the number of enrolment and verification samples, an ECC with decoding capabilities at Shannon's bound, and the target FNMR. Further-



(a) DF (b) GF

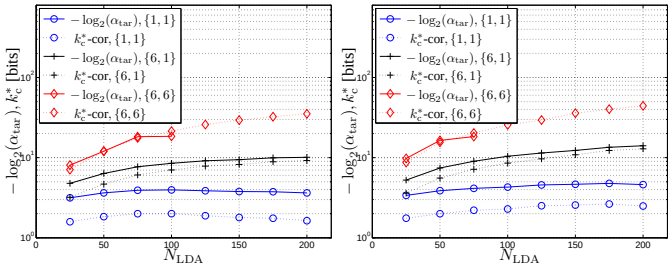
(c) DF (d) GF

Figure 4.27: (a)(b) The estimated relative intrinsic degrees of freedom or dimension $\frac{\hat{n}_c}{n_c}$ of the Hamming distance pmf at imposter comparisons for different LDA settings N_{LDA} and number of enrolment N_e or verification N_v samples, and (c)(d) the corresponding estimate bit-error probability $\hat{P}_e^{im}$ for both the DF and GF features.

more, we modeled the Fuzzy Commitment Scheme as a binary symmetric channel with its corresponding bit-error probability.

The biometric source is modeled by parallel Gaussian channels. Each Gaussian channel models the real-valued behavior of a feature component by means of a Gaussian density for the between-class variance and an additive zero-mean Gaussian density for the within-class variance. Furthermore, we considered the within-class noise to be independent across components and measurements, and homogeneous, i.e. given a component all subjects have an equal variance. However, the within-class variance can be different for each component. The ratio between the between-class and within-class standard deviation is used as the feature quality. Because of the Gaussian assumption we used the Gaussian channel capacity as the discriminant power for each component. Consequently, the discriminant power of the biometric source, referred to as the input capacity, is defined by the sum of the Gaussian channel capacity across all components.

As the quantization scheme, we used a known method where a single bit is extracted per component using a binarization scheme based on thresholding. As the threshold we used the mean of the between-class density. With this setup we estimated the bit-error probability disturbing the binary symmetric channel, where we also included the effect of the number of enrolment and verification samples on the bit-error probability. We have



(a) DF (b) GF

Figure 4.28: The corrected maximum key size k_c^*-cor, the log of the FMR at the target FNMR $-\log_2(\alpha_{\text{tar}})$ as a function of the LDA dimension N_{LDA} at different N_e and N_v settings for the DF and GF features. The number of samples is indicated in the legend with $\{N_e, N_v\}$.

shown that having an infinite enrolment samples with X verification samples approximates the performance when both are equal to $2X$, if the feature quality is large enough.

We estimated the maximum key size assuming an optimal binary ECC that corrects up to t_c random bit errors with decoding capabilities at Shannon's bound. We determined the maximum key size at the operating point determined by Shannon's theorem, at the operating point where the EER is achieved, and at the operating point determined by the target FNMR. We showed that the maximum key size obtained at the operating point dictated by Shannon's theory is optimistic and has a high FNMR, namely close to 50%. The high FNMR is due to the assumption from Shannon's theory that the codeword should be sufficiently large, while it is not large enough even for the best performance biometric, namely iris, which has a degree of freedom of 249 bits. We proposed two other operating points, namely the analytical operating point at the EER and the operating point given the target FNMR. The maximum key size at the EER is always smaller, down to 25%, than at the operating point from Shannon's theory. At the EER more bits have to be corrected due the smaller FNMR requirement, consequently the operating point is larger leading to a smaller key size. The operating point at the target FNMR is a compromise between the two aforementioned methods, and leads to the maximum key size with the desired FNMR. We also discussed the relationship between the maximum key and the target FMR at the target FNMR. We showed that the upperbound from literature, namely that the maximum key size is smaller than $-\log_2(\text{FMR})$, is not so tight when errors have to be corrected. The difference increases when using larger codewords, and could be around 3 bits when the codeword is 127 bits long.

With the analytical framework we studied, by means of numerical analysis, the effect of the system parameters such as the source capacity, the number of feature components, the number of enrolment and verification samples, and the target FNMR on the classification performance and the maximum key size. There are three main scenario, namely (i) the scenario where the input capacity is uniformly distributed among the components,

(ii) the scenario where it is non-uniformly distributed, and (iii) the scenario where a set of components are fully dependent.

For the first scenario, we extensively studied the effect of the system parameters, namely (i) the case where both the input capacity and the number of components are given, (ii) the case where the optimal number op components is determined given the input capacity, and (iii) the case where the input capacity and the number of components are determined in order to reach the target FMR and target FNMR. In the first case we showed that increasing the number of enrolment and verification samples and the target FNMR increases the maximum key size, however the maximum key size is limited to the number of components. The greatest improvement is obtained by increasing both the number of enrolment and verification samples. Similar results were obtained for the second case, however due to the variable number of components, the maximum key size had a greater increase, especially when increasing both the enrolment and verification samples. The main finding was the fact that components with worse discriminant power could be used when increasing the system parameters, which was also confirmed by the experiment results with the fingerprint database. With the range of the input capacity between 40 bits and 80 bits, we found the following numerical analysis results. Doubling the input capacity roughly tripled the key size at a target FNMR of 2.5%, while doubling the target FNMR from 2.5% to 5% on average added around 1 bit. Increasing the number of enrolment samples from one to six added 2.9 bits. With six enrolment samples and increasing the number of verification samples from one to two added 7.6 bits, while increasing from two to six samples added 20.8 bits. Thus, if the subjects of the biometric system have no issue with a less convenient system where the target FNMR has increased or more biometric samples have to be acquired, we could create a protected template that is more difficult to break by an adversary. Namely, doubling the target FNMR also doubles the search space of the key. Moreover, switching from a single to six enrolment and verification samples increases the search space by almost 2^{32}. Supplying six samples during enrolment seems acceptable, because it only needs to be done once. Although capturing six samples during verification may be considered inconvenient, it still gives a good insight in what can be achieved by such a system. In both the first and second case we observed that the maximum key size significantly reduces if the target FNMR is smaller than 5%. In the third case we showed the trade-off between the key extraction efficiency and the optimal maximum key size given a target FMR. If the maximum key size has to be increase the input capacity has to increase unproportionately, hence reducing the key extraction efficiency.

Comparing the results of the first two scenarios, we can conclude that given a certain input capacity, any deviation from a uniformly distribution is sub-optimal with respect to the maximum key size. Note that we considered an ECC that does not exploit the prior knowledge of bits having different bit-error probabilities. Furthermore, in the third scenario we showed that adding fully dependent bits does not improve the performance, the FMR can even increase at the same target FNMR, while the maximum key size can be artificially increased. We conjecture that the discrepancies between the reported key size and system performance shown in Table 4.1 is mainly caused by this artificially increase of the key size due to dependencies between feature components. Hence, both the reported key size and FMR have to be taken into account when analyzing the actual privacy

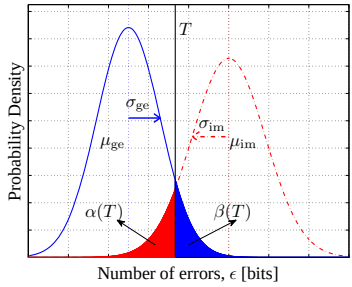


Figure 4.29: The Gaussian approximation of the pmf of the number of errors ϵ at genuine (the solid blue curve) and imposter (the dashed-dotted red curve) comparisons from Figure 4.7.

and security of a template protection system.

We can conclude that we analytically obtained the relationship between the system performance and the maximum key size given the system parameters. Having independent feature components of equal quality is necessary in order to be optimal in terms of performance and key extraction. Furthermore, we revealed the trade-off between the convenient use of the biometric system, determined by the target FNMR and the number of samples to be acquired, and the privacy and security protection indicated by the maximum key size. Essentially, if desired, more protection can be achieved by sacrificing some convenience.

4.A The EER Operating Point with Gaussian Approximation

In order to find an analytical expression of the EER operating point, T_{EER}, we approximate the binomial density used for modelling the pmf of the Hamming distance ϵ by a Gaussian density. Hence, instead of (4.9) we use

$$\begin{aligned} P_{\mathrm{G}}(\epsilon; N, p) &= \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\left(\frac{\epsilon-\mu}{\sigma\sqrt{2}}\right)^2}, \\ &= \frac{1}{\sqrt{n_{\mathrm{c}}(1-p)(p)}\sqrt{2\pi}} \mathrm{e}^{-\left(\frac{\epsilon-n_{\mathrm{c}}p}{\sqrt{2n_{\mathrm{c}}(1-p)p}}\right)^2}, \end{aligned} \tag{4.27}$$

where we use the mean and variance of the binomial density, namely the mean $\mu = n_{\mathrm{c}}p$ and standard deviation $\sigma = \sqrt{n_{\mathrm{c}}(1-p)p}$. The resulting approximated probability density as a function of the Hamming distance ϵ is shown in Figure 4.29.

Thus given the operating point T, the FNMR from (4.12) can be rewritten as

$$\begin{aligned}\beta(T) &= \int_{i=T}^{\infty} P_{\mathrm{G}}(i; n_{\mathrm{c}}, P_{\mathrm{e}}^{\mathrm{ge}})\,\mathrm{d}i \\ &= \int_{i=T}^{\infty} \tfrac{1}{\sigma_{\mathrm{ge}}\sqrt{2\pi}} \mathrm{e}^{-\left(\frac{i-\mu_{\mathrm{ge}}}{\sigma_{\mathrm{ge}}\sqrt{2}}\right)^2}\,\mathrm{d}i,\end{aligned} \tag{4.28}$$

with $\mu_{\mathrm{ge}} = n_{\mathrm{c}} P_{\mathrm{e}}^{\mathrm{ge}}$ and $\sigma_{\mathrm{ge}} = \sqrt{n_{\mathrm{c}}(1-P_{\mathrm{e}}^{\mathrm{ge}})P_{\mathrm{e}}^{\mathrm{ge}}}$. By applying the following change of variable $\tau = \frac{i-\mu_{\mathrm{ge}}}{\sigma_{\mathrm{ge}}}$ with $\mathrm{d}i = \sigma_{\mathrm{ge}}\mathrm{d}\tau$ we obtain

$$\beta(T) = \int_{\tau=z_{\mathrm{ge}}(T)}^{\infty} \tfrac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\tau^2}\,\mathrm{d}\tau, \tag{4.29}$$

where we have the genuine z-score $z_{\mathrm{ge}}(T) = \frac{T-\mu_{\mathrm{ge}}}{\sigma_{\mathrm{ge}}}$ that fully determines the FNMR. Similarly, for the FMR we have

$$\begin{aligned}\alpha(T) &= \int_{i=-\infty}^{T} P_{\mathrm{G}}(i; n_{\mathrm{c}}, P_{\mathrm{e}}^{\mathrm{im}})\,\mathrm{d}i \\ &= \int_{\tau=-\infty}^{z_{\mathrm{im}}} \tfrac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\tau^2}\,\mathrm{d}\tau \\ &= \int_{\tau=-z_{\mathrm{im}}(T)}^{\infty} \tfrac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\tau^2}\,\mathrm{d}\tau,\end{aligned} \tag{4.30}$$

where we applied the same variable change, defined the imposter z-score $z_{\mathrm{im}}(T) = \frac{T-\mu_{\mathrm{im}}}{\sigma_{\mathrm{im}}}$ and used the property that the integral is symmetric. Because $P_{\mathrm{e}}^{\mathrm{im}} = \frac{1}{2}$, we have $\mu_{\mathrm{im}} = \frac{n_{\mathrm{c}}}{2}$ and $\sigma_{\mathrm{im}} = \frac{\sqrt{n_{\mathrm{c}}}}{2}$ Being at the EER operating point T_{EER} implies that $\alpha(T_{EER}) = \beta(T_{EER})$. Hence, equation (4.29) and (4.30) have to be equal. Both equations are equal when $z_{\mathrm{ge}}(T_{EER}) = -z_{\mathrm{im}}(T_{EER})$, thus T_{EER} becomes

$$\begin{aligned} z_{\mathrm{ge}}(T_{EER}) &= -z_{\mathrm{im}}(T_{EER}), \\ \tfrac{T_{EER}-\mu_{\mathrm{ge}}}{\sigma_{\mathrm{ge}}} &= -\tfrac{T_{EER}-\mu_{\mathrm{im}}}{\sigma_{\mathrm{im}}}, \\ T_{EER} &= \tfrac{\mu_{\mathrm{im}}\sigma_{\mathrm{ge}}+\mu_{\mathrm{ge}}\sigma_{\mathrm{im}}}{\sigma_{\mathrm{im}}+\sigma_{\mathrm{ge}}}. \end{aligned} \tag{4.31}$$

Substituting the genuine parameters $\mu_{\mathrm{ge}} = n_{\mathrm{c}} P_{\mathrm{e}}^{\mathrm{ge}}$ and $\sigma_{\mathrm{ge}} = \sqrt{n_{\mathrm{c}}(1-P_{\mathrm{e}}^{\mathrm{ge}})P_{\mathrm{e}}^{\mathrm{ge}}}$, and the imposter parameters $\mu_{\mathrm{im}} = \frac{n_{\mathrm{c}}}{2}$ and $\sigma_{\mathrm{im}} = \frac{\sqrt{n_{\mathrm{c}}}}{2}$, we obtain

$$\begin{aligned} T_{EER} &= \tfrac{n_{\mathrm{c}}\left(\sqrt{P_{\mathrm{e}}^{\mathrm{ge}}(1-P_{\mathrm{e}}^{\mathrm{ge}})}+P_{\mathrm{e}}^{\mathrm{ge}}\right)}{2\sqrt{P_{\mathrm{e}}^{\mathrm{ge}}(1-P_{\mathrm{e}}^{\mathrm{ge}})}+1}, \text{ or} \\ \tfrac{T_{EER}}{n_{\mathrm{c}}} &= \tfrac{\sqrt{P_{\mathrm{e}}^{\mathrm{ge}}(1-P_{\mathrm{e}}^{\mathrm{ge}})}+P_{\mathrm{e}}^{\mathrm{ge}}}{2\sqrt{P_{\mathrm{e}}^{\mathrm{ge}}(1-P_{\mathrm{e}}^{\mathrm{ge}})}+1}. \end{aligned} \tag{4.32}$$

Note that the relative operating point $\frac{T_{EER}}{n_{\mathrm{c}}}$ and thus the BSC channel capacity at the EER operating point $C(\frac{T_{EER}}{n_{\mathrm{c}}})$ is fully determined by $P_{\mathrm{e}}^{\mathrm{ge}}$.

4.2 Chapter Conclusions

The maximum key size that can be extracted depends on the performance of the underlying biometric recognition system. Note that the FMR and FNMR depend on the operating point T, which is equal to the number of bits the ECC has to correct. With the relative number of bits that have to be corrected we can determine the maximum key size by assuming the ECC to be operating Shannon's bound. An important finding of this work is the fact that the upperbound of the key size for the HDS known in the literature, namely $-\log_2(\alpha_{\mathrm{tar}})$ of (4.21) where α_{tar} is the target FMR, is not tight when compared to the maximum key derived assuming an ECC operating on Shannon's bound. The difference can be a couple of bits and increases with the number of feature components. When the FMR is taken as the target performance, the key size depends on the operating point determined by the target FMR and has an upperbound given by $-\log_2(\alpha_{\mathrm{tar}})$. However, when taking the FNMR as the target performance, the operating point and therefore the maximum key size depend on

i The target FNMR: increasing the FNMR increases the maximum key size.

ii The input capacity: increasing the input capacity increases the maximum key size and having feature components of equal quality, where the quality is defined by the ratio of the between-class and within-class variance, optimizes the key size that can be achieved.

iii The number of feature components: deviating from the optimal number of components reduces the key size.

iv The number of enrolment and verification samples: increasing the number of samples increases the maximum key size. Furthermore, increasing the number of samples increases the optimal number of components, hence the feature extraction algorithm can output larger feature vectors.

Considering the fact that having a larger target FNMR and more enrolment and verification samples do influence the convenience of the biometric system, we have shown a trade-off between the protection capability of the HDS in terms of key size with respect to its convenience.

With respect to the number of enrolment and verification samples, we have shown that the classification performance for the $N_{\mathrm{e}} = N_{\mathrm{v}} = 2X$ case converges to the $\{N_{\mathrm{e}} = \infty, N_{\mathrm{v}} = X\}$ case when the feature quality increases. In other words increasing the number of enrolment samples to infinity leads to a similar performance when doubling both the number of enrolment and verification samples.

With Table 4.1 we have provided the discrepancy between the reported FMR and key size published in the literature. With the adapted model incorporating fully dependent feature components, we may be able to explain one possible cause of this discrepancy, namely the dependency between the feature vector components. Because of this dependency, the reported key size can be larger than the upper bound depending on the FMR, namely $-\log_2(\alpha_{\mathrm{tar}})$. Hence, in this case the key size is not indicative for the remaining uncertainty about the biometric data given the protected template.

Chapter 5

Information Leakage Analysis of the Bit Protection Part

5.1 Chapter Introduction

In this chapter the first part of the third research question will be addressed, namely

> **Given the HDS template protection scheme: How does the information leakage from the auxiliary data affect the irreversibility and unlinkability property?**

More specifically, this chapter answers this research question only for the bit protection part of the HDS of Figure 1.5, also known as the fuzzy commitment scheme. Chapter 6 will answer the question for the bit extraction part. Recent publications [39, 40] have shown a vulnerability affecting the unlinkability property, namely the cross-matching possibility based on the code-offset auxiliary data AD_2 only, due to the linear property of the ECC. Simoens et al. (2009) [40] determined the theoretical FMR when comparing AD_2 of arbitrary protected templates from different application. In Section 5.2 we extend this analysis and also determine the theoretical FNMR. We also show that as long as the HDS is balanced, i.e. there are equal number of enrolment and verification samples ($N_\mathrm{e} = N_\mathrm{v}$), the cross-matching classification performance is worse than the classification performance of the HDS. Besides this extended analysis, we also provide a solution based on randomization in order to mitigate the cross-matching performance close to random. The main results are published in Kelkboom et al. (2010) [131][1].

[1] E. J. C. Kelkboom, J. Breebaart, T. A. M. Kevenaar, I. R. Buhan, and R. N. J. Veldhuis, "Preventing the decodability attack based cross-matching in a fuzzy commitment scheme," *Submitted to IEEE Transactions on Information Forensics and Security*, 2010.

5.2 Preventing the Decodability Attack based Cross-matching in a Fuzzy Commitment Scheme

5.2.1 Abstract

Template protection techniques are used within biometric systems in order to safeguard the privacy of the system's subjects. This protection also includes unlinkability, i.e. preventing cross-matching between two or more reference templates from the same subject across different applications. In the literature, the template protection techniques based on fuzzy commitment, also known as the code-offset construction, have recently been under scrutiny. Recent work presented the decodability attack vulnerability facilitating cross-matching based on the protected templates and its theoretical analysis. Firstly, we extend the theoretical analysis and include the comparison between the system and cross-matching performance. We validate the presented analysis using real biometric data from the MCYT fingerprint database. Secondly, we show that applying a random bit-permutation process secures the fuzzy commitment scheme from cross-matching based on the decodability attack.

5.2.2 Introduction

When using an application based on biometrics, first a *reference template* is generated from the biometric sample provided in the enrolment phase for later use. In the authentication phase, a new biometric sample is acquired and compared with the reference template. Hence, the application requires this reference template for a successful authentication and therefore it needs to be stored. Basically, there are two options of storage, namely on a token carried by the subjects themselves or in a centralized database. The latter case is considered to be more convenient for the subjects. However storing unprotected biometric reference templates in centralized databases for each application increases the privacy risk. For example, if these databases are compromised, an adversary could check the types of applications or services a specific subject has subscribed to. In the literature, this is known as *cross-matching*.

Therefore it is not a surprise that the ISO guidelines [25] dictate the avoidance of centralized databases if possible. Some known countermeasures to safeguard the privacy and security by enforcing some of the ISO guidelines are (i) the practice of *data separation* where the most privacy sensitive information is stored on an individual smartcard or token, (ii) the use of *data minimization* principles, (iii) the use of *classical encryption* techniques such as DES, AES, RSA to augment the confidentiality or integrity of the reference template and (iv) the implementation of *template protection* which creates irreversible, renewable and unlinkable reference templates, i.e. protected reference templates. In our work we focus only on the template protection method.

In the literature, numerous template protection methods such as the *Fuzzy Commitment Scheme* (FCS) [36], *Helper Data System* (HDS) [33, 34, 48], *Fuzzy Extractors* [64, 65], *Fuzzy Vault* [80, 84] and *Cancelable Biometrics* [59] have been proposed, with the claim of preventing cross-matching. However, recently it was presented in [132] that fuzzy vaults were susceptible to cross-matching and [119] solved this issue by hard-

ening the protected reference template using a secret key or password provided by the subject. The requirement of keeping the key or password secret however, has a serious impact on the convenience of the biometric system.

In the FCS construction, also known as the code-offset construction, the binary vector extracted from the biometric sample is XOR-ed with a randomly selected codeword resulting into auxiliary data that is stored as part of the protected template. Certain implementations of the Helper Data System, Fuzzy Extractors are based on this FCS construction. Possible cross-matching vulnerabilities for template protection systems based on the FCS construction are briefly discussed in [89] and are based on attack methods using exhaustive search. More recently, a new vulnerability known as the *decodability attack* has been published for the case when the FCS is based on a linear error-correcting codes (ECC). To the best of our knowledge, the cross-matching vulnerability of the FCS construction is first published by the presentation of Dr. Stoianov at the European Biometrics Forum (EBF) Biometric Encryption Seminar [39]. Cross-matching is made possible by simply checking whether decoding the XOR of two auxiliary data elements stored in different databases leads to a valid codeword. If it leads to a valid codeword the two auxiliary data most likely belong to the same subject and is labeled as genuine. Therefore, this vulnerability is also known as the decodability attack. More recently, a theoretical analysis is presented in [40] where the authors determine the probability that the decodability attack incorrectly labels two auxiliary data from different subjects as genuine under the assumption that across the whole population the bits of the binary vector are independent and uniform.

Contributions: As our first contribution, we extend the theoretical analysis from [40] and show the relationship between the cross-matching performance with the template protection system performance itself. Furthermore, we empirically evaluate the theoretical analysis using real biometric data from the MCYT fingerprint database and show that if no care is taken cross-matching based on the decodability attack is indeed possible. However, as our second contribution we will show that this vulnerability can be prevented by implementing a bit-permutation or shuffling randomization process on the binary vector. Consequently, the cross-matching performance is close to random.

The outline of this paper is as follows. In Section 5.2.3 we briefly describe the FCS construction, present the properties of a linear error-correcting code (ECC), and discuss a probability estimation case extensively used in the remainder of this work. In Section 5.2.4 we discuss the possible cross-matching attacks including the newly published decodability attack [39, 40]. In Section 5.2.5 we theoretically analyze both the cross-matching and template protection system performance and show their relationship. Validation of the theoretical performances are conducted in Section 5.2.6 using the MCYT fingerprint database. In Section 5.2.7 we show that a bit-permutation randomization process reduces the effectiveness of the decodability attack. Conclusions are given in Section 5.2.8.

5.2.3 Preliminaries

The template protection scheme under consideration is known as the Fuzzy Commitment Scheme (FCS) from [36] and is based on an error-correcting code (ECC). We first discuss

the notations related to the ECC and thereafter we present the FCS. Furthermore, we discuss the estimation of the probability mass function (pmf) of the number of bit errors when XOR-ing two random binary vectors, which is extensively used in the remainder of this work.

Linear Error-Correcting Code

We denote a t_{c}-error linear binary error-correcting code as $[n_{\mathrm{c}}, k_{\mathrm{c}}, t_{\mathrm{c}}]$, where n_{c} is the length of the codeword $\mathbf{C}$, k_{c} the length of the message or key $\mathbf{K}$, and t_{c} the error-correcting capability.

The *ECC Encoder* (*Enc*) function converts the key $\mathbf{K} \in \{0,1\}^{k_{\mathrm{c}}}$ into its corresponding codeword $\mathbf{C} \in \{0,1\}^{n_{\mathrm{c}}}$. The codebook $\mathcal{C}$ is the set of all valid codewords of the ECC with cardinality $|\mathcal{C}| = 2^{k_{\mathrm{c}}}$. As the distance function we use the Hamming distance denoted as $d_{\mathrm{H}}\{.,.\}$ and the Hamming weight denoted as $||.||$. The minimum distance of the codebook $\mathcal{C}$ is $d = 2t_{\mathrm{c}} + 1$, therefore it can correct up to t_{c} bit errors. Because of the linearity property of the ECC it holds that the XOR operation between any pair of codewords leads to another codeword from the same codebook $\mathcal{C}$, namely $\forall\ \mathbf{C}_i, \mathbf{C}_j \in \mathcal{C} : \mathbf{C}_i \oplus \mathbf{C}_j = \mathbf{C}_k$, with $\mathbf{C}_k \in \mathcal{C}$. Furthermore, we define $W_{\mathcal{C}}$ to be the set of possible weights w of the codewords from $\mathcal{C}$, while the function $N_{\mathcal{C}}(w)$ returns the number of codewords n_w with weight w, with $\sum_{w \in W_{\mathcal{C}}} N_{\mathcal{C}}(w) = |\mathcal{C}|$.

Given a word $\mathbf{w} \in \{0,1\}^{n_{\mathrm{c}}}$ and the smallest distance to any codeword defined as $d_{\mathrm{c}}(\mathbf{w}, \mathcal{C}) \stackrel{\mathrm{def}}{=} \min_{\mathbf{C} \in \mathcal{C}} d_{\mathrm{H}}(\mathbf{w}, \mathbf{C})$, the *ECC Decoder* (*Dec*) function returns the key corresponding to the closest codeword from the codebook $\mathcal{C}$ if the smallest distance $d_{\mathrm{c}}(\mathbf{w}, \mathcal{C})$ is smaller than or equal to the error-correcting capability t_{c}, i.e. $d_{\mathrm{c}}(\mathbf{w}, \mathcal{C}) \leq t_{\mathrm{c}}$. When the smallest distance is larger than the error-correcting capability, $d_{\mathrm{c}}(\mathbf{w}, \mathcal{C}) > t_{\mathrm{c}}$, then the word is not decodable and the *ECC Decoder* function either returns a decoding error or randomly selects a key.

In our experiments we use the linear block type ECC "Bose, Ray-Chaudhuri, Hocquenghem" (BCH), with some $[n_{\mathrm{c}}, k_{\mathrm{c}}, t_{\mathrm{c}}]$ settings given in Table 5.1. For the BCH ECC we use the maximum error-correcting capability t_{c}^{*} is limited to around 25% of the codeword size n_{c} (see Table 5.1), and if the word is not decodable it outputs the first k_{c} bits of the word as the key.

Fuzzy Commitment Scheme

The fuzzy commitment scheme (FCS) from [36] is one of the first template protection techniques and is based on the *bit commitment technique* known within the field of cryptography. The FCS works on discrete biometric data, while in practice most biometric data are continuous. Figure 5.1 portrays the FCS construction combined with a bit extraction module.

In the enrolment phase the real-valued column *feature vector* $\mathbf{f}^{\mathrm{e}} \in \mathbb{R}^{N_{\mathrm{F}}}$ is extracted from each N_{e} biometric enrolment sample by the feature extraction algorithm. From the N_{e} feature vectors, a single binary column vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \in \{0,1\}^{N_{\mathrm{F}}}$ is created. For each component, we extract a single bit using a bit extraction scheme based on thresholding,

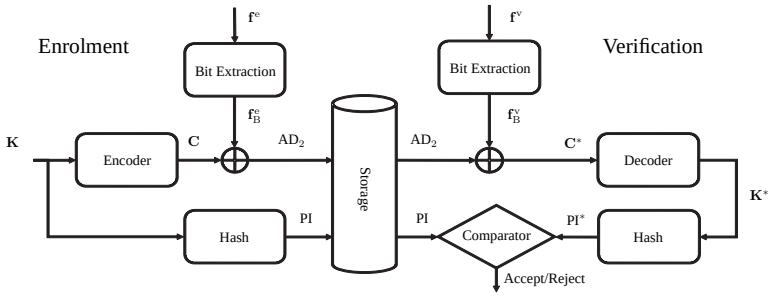


Figure 5.1: The fuzzy commitment scheme (FCS) combined with a bit extraction module.

where the mean of the background density is chosen as the threshold and estimated from a disjoint training set [33–35]. Prior to thresholding the mean of the N_{e} feature vectors is taken. Furthermore, a random key $\mathbf{K} \in \{0,1\}^{k_{\mathrm{c}}}$ is created and encoded by the *ECC Encoder* module into a codeword $\mathbf{C} \in \{0,1\}^{n_{\mathrm{c}}}$ from $\mathcal{C}$. The fundamental property of the FCS is the XOR operation of the codeword $\mathbf{C}$ and the binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ creating the offset AD_2 as helper data, $\mathrm{AD}_2 = \mathbf{C} \oplus \mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$. The helper data AD_2 is also referred to as the *Auxiliary Data* in [102], in line with standardization activities in ISO [25]. Together with the hash of $\mathbf{K}$, also referred to as the *Pseudonymous Identifier* (PI), we obtain the protected template. As described in [36], $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ is equivalent to the *witness* with which we commit the codeword $\mathbf{C}$ using the XOR operation considered to be similar to the one-time-pad encryption algorithm. The outcome of the commitment is the AD_2 and PI pair, which together is also known as the *blob*.

In the verification phase, the binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$ is created by quantizing the mean of the N_{v} verification feature vectors $\mathbf{f}^{\mathrm{v}}$. Hereafter, the auxiliary data AD_2 is XOR-ed with

Table 5.1: Examples of the BCH ECC given by the codeword (n_{c}) and key (k_{c}) length, the corresponding correctable bits (t_{c}), and the relative error correcting capability $t_{\mathrm{c}}/n_{\mathrm{c}}$.

n_{c} [bits]	k_{c} [bits]	t_{c} [bits]	$t_{\mathrm{c}}/n_{\mathrm{c}}$
31	6	7	22.6%
	11	5	16.1%
	16	3	9.7%
63	7	15	23.8%
	16	11	17.5%
	24	7	11.1%
127	8	31	24.4%
	22	23	18.1%
	36	15	11.8%

$\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$ resulting into the possibly corrupted codeword $\mathbf{C}^* = \mathrm{AD}_2 \oplus \mathbf{f}_{\mathrm{B}}^{\mathrm{v}} = \mathbf{C} \oplus (\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \oplus \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}) = \mathbf{C} \oplus \mathbf{e}$, where the Hamming distance $\epsilon = d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}) = ||\mathbf{e}||$ indicates the number of errors corrupting the codeword $\mathbf{C}$. Decoding $\mathbf{C}^*$ by the *ECC Decoder* module leads to the candidate key $\mathbf{K}^*$. The candidate pseudonymous identifier PI^* is obtained by hashing $\mathbf{K}^*$. A match is returned by the *Comparator* module if both PI and PI^* are equal, which occurs only when $\mathbf{K}$ and $\mathbf{K}^*$ are equal. Both secrets are equal when the Hamming distance between the binary vectors $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ and $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$ is smaller or equal to the error-correcting capability of the ECC, $\epsilon = d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}) \leq t_{\mathrm{c}}$. Hence, to successfully decommit the blob, a new witness $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$ has to be provided that is within t_{c} bit differences with the original witness $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$.

An illustration of the code-offset is presented in Figure 5.2, where the n_{c} dimensional problem is simplified into a 2D problem. The grid of small dots represent the word space $\{0,1\}^{n_{\mathrm{c}}}$, while the bigger dots represents the codewords from $\mathcal{C}$ with the error-correcting capability represented by the circles with radius t_{c}. The auxiliary data AD_2 shifts the enrolment binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ to the codeword $\mathbf{C}$. In the verification phase, the same shift is applied to $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$ and will lead to a match only if it is within the radius t_{c} of codeword $\mathbf{C}$. Hence, all binary vectors $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$ within the dashed circle with radius t_{c} and center point $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ will lead to a match.

In this work we consider two cases of the FCS, namely the *unbalanced* and *balanced system*. For the unbalanced system there are $N_{\mathrm{e}} \neq N_{\mathrm{v}}$ enrolment samples with N_{v} verification samples, while for the balanced case the number of verification samples is equal to the number of enrolment samples, $N_{\mathrm{v}} = N_{\mathrm{e}}$.

Hamming Weight after XOR-ing two Random Binary Vectors

In many derivations in the remainder of this work we need a solution to the following problem. Consider the case of having two words $\mathbf{w}_1$ and $\mathbf{w}_2$ randomly selected from $\{0,1\}^{n_{\mathrm{c}}}$ with weights w_1 and w_2, respectively. Defining the number of bit errors or differences ϵ between $\mathbf{w}_1$ and $\mathbf{w}_2$, namely $\epsilon = d_{\mathrm{H}}(\mathbf{w}_1, \mathbf{w}_2)$, we are interested in the probability mass function (pmf) of ϵ.

Lemma 5.2.1 (Hamming Weight after the XOR of two Binary Vectors)**.** *Given two random binary vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ with Hamming weight w_1 and w_2, respectively and defining $w_{\min} = \min(w_1, w_2)$, and $w_{\max} = \max(w_1, w_2)$, the number of possible bit errors $\epsilon = d_{\mathrm{H}}(\mathbf{w}_1, \mathbf{w}_2)$ is given by the set $E = \{\epsilon_{\min}, \epsilon_{\min}+2, \ldots, \epsilon_{\max}-2, \epsilon_{\max}\}$ with probability $P_{w\times w}(\epsilon; w_1, w_2, n_{\mathrm{c}})$ defined as*

$$P_{w\times w}(\epsilon; w_1, w_2, n_{\mathrm{c}}) \stackrel{\mathrm{def}}{=} \begin{cases} 0 & \text{if } \epsilon \notin E \\ \frac{1}{\binom{n_{\mathrm{c}}}{w_{\min}}} \binom{w_{\max}}{w_{\min}-(\epsilon-\epsilon_{\min})/2} \binom{n_{\mathrm{c}}-w_{\max}}{(\epsilon-\epsilon_{\min})/2} & \text{if } \epsilon \in E \end{cases}, \tag{5.1}$$

where $\epsilon_{\min} = |w_1 - w_2|$, and $\epsilon_{\max} = n_{\mathrm{c}} - |w_1 + w_2 - n_{\mathrm{c}}|$.

Proof. Because $\mathbf{w}_1$ and $\mathbf{w}_2$ have w_1 and w_2 bits of value 1, respectively, the minimum number of possible errors equals the difference $\epsilon_{\min} = |w_1 - w_2|$. For example, let $w_1 > w_2$, i.e. $w_{\max} = w_1$ and $w_{\min} = w_2$, and the first w_1 bits of $\mathbf{w}_1$ have a value 1 while the remaining $n_{\mathrm{c}} - w_1$ bits have a value 0. The case with $\epsilon_{\min}$ errors can be obtained

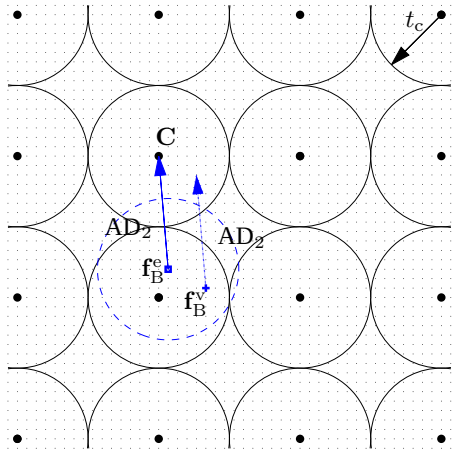


Figure 5.2: An illustration of the FCS construction principles. The grid of small dots represent the word space $\{0,1\}^{n_c}$, while the bigger dots represents the codewords from $\mathcal{C}$ with the error-correcting capability represented by the circles with radius t_c. AD_2 shifts the enrolment binary vector $\mathbf{f}_B^e$ to the codeword $\mathbf{C}$. In the verification phase, the same shift is applied to $\mathbf{f}_B^v$ and will lead to a match if it is within the radius t_c of codeword $\mathbf{C}$. Hence, all binary vectors $\mathbf{f}_B^v$ within the dashed circle with radius t_c and center point $\mathbf{f}_B^e$ will lead a match.

by allocating the w_2 bits of value 1 as the first bits of $\mathbf{w}_2$. Overall, there are $\binom{w_1}{w_2}$ possible combinations of having w_2 bits of value 1 of $\mathbf{w}_2$ at locations where the bits of $\mathbf{w}_1$ have a value of 1. Thus, the probability of having ϵ_{min} errors is equal to the ratio of the number of possibilities with respect to the number of binary vectors of length n_c with weight w_2, namely $\binom{w_1}{w_2}/\binom{n_c}{w_2}$.

Note that two bit errors are introduced if one bit of value 1 of $\mathbf{w}_2$ is allocated where $\mathbf{w}_1$ has a bit value of 0 instead of value 1. Hence, there are $\binom{w_1}{w_2-1}\binom{n_c-w_1}{1}$ possible combinations of introducing 2 bit errors. The first binomial coefficient $\binom{w_1}{w_2-1}$ is the number of possibilities of locating w_2-1 bits of value 1 of $\mathbf{w}_2$ at the w_1 locations where $\mathbf{w}_1$ has bits of value 1. The second binomial coefficient $\binom{n_c-w_1}{1}$ is the number of possibilities of allocating a single bit of value 1 of $\mathbf{w}_2$ at the n_c-1 locations where $\mathbf{w}_1$ has a bit of value 0. Similarly, four bit errors are introduced when two bits of value 1 of $\mathbf{w}_2$ are allocated where $\mathbf{w}_1$ has a bit value of 0 with $\binom{w_1}{w_2-2}\binom{n_c-w_1}{2}$ possible combinations.

The maximum number of bit errors ϵ_{max} is introduced by allocating all w_2 bits of value 1 of $\mathbf{w}_2$ at locations where the bits of $\mathbf{w}_1$ have a value 0. When $w_1+w_2>n_c$, the number of bits of $\mathbf{w}_1$ of value 0 is smaller than the number of bits of $\mathbf{w}_2$ of value 1, namely $n_c-w_1<w_2$, because of the $w_1>w_2$ assumption. Consequently, the maximum number of bit errors is limited to $\epsilon_{max}=n_c-|w_1+w_2-n_c|$. □

5.2.4 Cross-Matching Attacks

The setup of the cross-matching analysis is depicted in Figure 5.3. We consider the scenario where there are two applications using the same biometric trait and identical template protection algorithms. Each application creates a protected template from independent enrolment samples of its subjects and stores it into its centralized database. We consider both centralized databases to be accessible by the adversary. Furthermore, we consider two cases differing on what is stored in the centralized database. In the first case, Case 1, both the auxiliary data AD_2 and the pseudonymous identifier PI are stored. Hence, the protected template for the first and second application is the pair $\{\mathrm{PI}_1, \mathrm{AD}_{2,1}\}$ and $\{\mathrm{PI}_2, \mathrm{AD}_{2,2}\}$, respectively. In the second case, Case 2, we consider only AD_2 to be stored in the centralized databases that are accessible, while PI may be stored within a personal storage device such as a smart-card which is not compromised. The adversary has access to all protected templates in both databases and tries to find subjects that are enrolled in both applications. Two protected templates, each taken from a different database, are compared by a *cross-matching classifier* in the *Comparator* module in order to determine whether they were derived from the same subject. The cross-matching classifier computes a cross-matching distance score s_{CM} on which to base its decision whether the two protected templates belong to the same subject (genuine) or not (imposter). The comparison between the protected templates of the same subject is referred to as a genuine comparison and between different subjects as an imposter comparison. In the ideal case, it would be impossible to link the protected samples from the same subjects across the two databases.

In this section, we discuss several cross-matching classifier methods. We discuss the exhaustive search approach for Case 1 and Case 2. We omitted the third possible case where only PI is stored in the centralized databases that are accessible by the adversary, because it can be easily shown that cross-matching is not possible. If the key could be derived from PI, they could still not be used for cross-matching because the keys where generate randomly within each application. Furthermore, we discuss the recently published method known as the decodability attack [39] [40], which is not based on an exhaustive search and only consists of an XOR and decoding operation by exploiting the linearity property of the ECC.

Exhaustive Search Attack

Given two protected templates, the exhaustive search type of the cross-matching attack relies on searching the complete codebook $\mathcal{C}$ in order to determine whether the two protected templates belong to the same subject.

Case 1: PI and AD_2. Recall that the pseudonymous identifier PI is the hash of the randomly selected key $\mathbf{K}$. Because the PI is part of the protected template, it is more effective to search the key from the PI. Assuming that the probability of a collision is small, i.e. the probability that two different keys have the same hash value, the key leading to the hash value equal to PI can be found by searching the key space of $\{0,1\}^{k_{\mathrm{c}}}$ and taking its hash value. The enrolled binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ can be obtained by computing

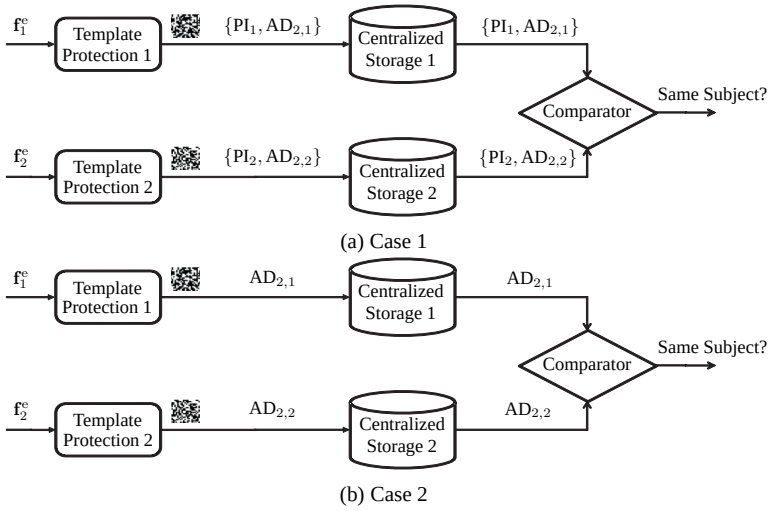


Figure 5.3: Two cases of the cross-matching attack scenario between two application databases that are accessible by the adversary. The first case (Case 1) both PI and AD_2 are stored in the centralized database. In the second case (Case 2) only AD_2 is stored in the centralized database accessible by the adversary, while AD_2 is assumed to be stored in a secure way and is not accessible by the adversary.

the XOR of auxiliary data AD_2 and the codeword $\mathbf{C}$ corresponding to the obtained key $\mathbf{K}$, namely $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} = \mathrm{AD}_2 \oplus \mathbf{C}$. By performing this exhaustive search on each protected template we obtain the binary vector $\mathbf{f}_{\mathrm{B},1}^{\mathrm{e}}$ and $\mathbf{f}_{\mathrm{B},2}^{\mathrm{e}}$ for the first and second application, respectively. As the cross-matching distance score s_{CM} we use the Hamming distance $s_{\mathrm{CM}} = \epsilon_{CM} = d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B},1}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B},2}^{\mathrm{e}})$. On average only half of the key space has to be searched, hence the average effort of finding the key corresponding to PI is $2^{k_{\mathrm{c}}-1}$. Consequently, finding both keys separately only takes twice the effort, namely $2^{k_{\mathrm{c}}}$.

Case 2: Only $\mathbf{AD}_2$. Because PI is not available, the distance measure has to be obtained from AD_2 only. By defining the XOR operation of the two auxiliary data as $\mathrm{AD}_{\oplus} \stackrel{\mathrm{def}}{=} \mathrm{AD}_{2,1} \oplus \mathrm{AD}_{2,2}$, we can rewrite $\mathrm{AD}_{\oplus}$ as

$$\begin{aligned} \mathrm{AD}_{\oplus} &= \mathrm{AD}_{2,1} \oplus \mathrm{AD}_{2,2} \\ &= (\mathbf{f}_{\mathrm{B},1}^{\mathrm{e}} \oplus \mathbf{C}_1) \oplus (\mathbf{f}_{\mathrm{B},2}^{\mathrm{e}} \oplus \mathbf{C}_2) \\ &= (\mathbf{f}_{\mathrm{B},1}^{\mathrm{e}} \oplus \mathbf{f}_{\mathrm{B},2}^{\mathrm{e}}) \oplus (\mathbf{C}_1 \oplus \mathbf{C}_2) \\ &= \mathbf{e} \oplus \mathbf{C}_3, \end{aligned} \tag{5.2}$$

where $\mathbf{f}_{\mathrm{B},1}^{\mathrm{e}}$ ($\mathbf{C}_1$) and $\mathbf{f}_{\mathrm{B},2}^{\mathrm{e}}$ ($\mathbf{C}_2$) are the binary vectors (codewords) in the enrolment phase for application 1 and 2 respectively, $\mathbf{e}$ is the error pattern between the enrolment binary vectors, and we used the property of linear codes where the XOR of two codewords leads to another codeword from the same codebook. A graphical representation of the XOR

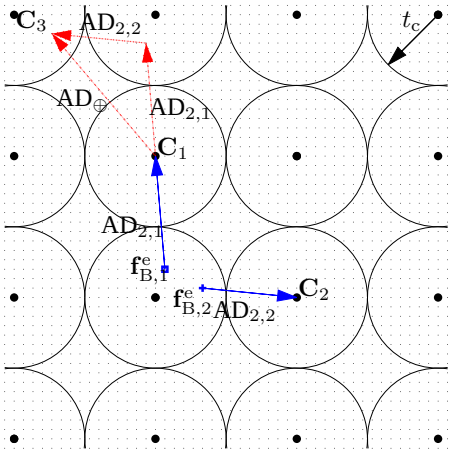


Figure 5.4: An illustration of the XOR of $\mathrm{AD}_{2,1}$ and $\mathrm{AD}_{2,2}$ obtained from the enrolled binary samples $\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1}$ and $\mathbf{f}^{\mathrm{e}}_{\mathrm{B},2}$ from the same subject.

operation is presented in Figure 5.4. Hence, all possible error patterns can be computed by exhaustively taking the XOR of $\mathrm{AD}_{\oplus}$ with any codeword from $\mathcal{C}$, which is an effort of 2^{k_c}. As the cross-matching distance score s_{CM} we take the error pattern with the smallest Hamming weight, namely $s_{\mathrm{CM}} = \min_{\mathbf{C}\in\mathcal{C}}||\mathrm{AD}_{\oplus} \oplus \mathbf{C}||$. Note that it holds that $s_{\mathrm{CM}} = \epsilon_{CM} = d_{\mathrm{H}}(\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1}, \mathbf{f}^{\mathrm{e}}_{\mathrm{B},2})$ only when $\epsilon_{CM} = d_{\mathrm{H}}(\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1}, \mathbf{f}^{\mathrm{e}}_{\mathrm{B},2}) \leq t_{\mathrm{c}}$, because in this case $\mathbf{C}_3$ will lead to the smallest distance. For the case when $d_{\mathrm{H}}(\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1}, \mathbf{f}^{\mathrm{e}}_{\mathrm{B},2}) > t_{\mathrm{c}}$ there is a probability that we obtain $s_{\mathrm{CM}} \leq t_{\mathrm{c}}$. Because the distance of $\mathrm{AD}_{\oplus}$ to $\mathbf{C}_3$ is larger than t_{c} there is a probability that another neighboring codeword is closer, due to the existence of multiple codewords at the minimum distance $d = 2t_{\mathrm{c}} + 1$. The obtained cross-matching score is equal to $s_{\mathrm{CM}} = ||\mathbf{e}^*|| \leq t_{\mathrm{c}}$ only if the error pattern can be rewritten as $\mathbf{e} = \mathbf{e}^* \oplus \mathbf{C}_i$ with $||\mathbf{e}^*|| \leq t_{\mathrm{c}}$ and $\mathbf{C}_i \in \mathcal{C}$.

Note that when the codeword $\mathbf{C}_3$ is known, it is not possible to derive the binary vectors $\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1}$ and $\mathbf{f}^{\mathrm{e}}_{\mathrm{B},2}$, because the codewords $\mathbf{C}_1$ and $\mathbf{C}_2$ are not known. Because of the linear property of the ECC there are 2^{k_c} possible combinations of $\mathbf{C}_1$ and $\mathbf{C}_2$ that lead to $\mathbf{C}_3$. Hence, with this cross-matching attack we obtain only a distance measure between the two enrolment binary vectors $\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1}$ and $\mathbf{f}^{\mathrm{e}}_{\mathrm{B},2}$ but not their actual value.

The effort of determining the cross-matching distance score s_{CM} is case-dependent. If we obtained a cross-matching score s_{CM} smaller than t_{c}, the average effort of the corresponding cross-matching attack equals 2^{k_c-1}, because the search can be stopped once a score smaller than t_{c} has been obtained. When $s_{\mathrm{CM}} > t_{\mathrm{c}}$ then the complete codebook had to be searched and the effort is then 2^{k_c}.

Decodability Attack

The decodability attack method presented in both [39, 40] is based on cross-matching with only AD_2. For linear ECCs, they show that when $\mathrm{AD}_{\oplus}$ is decodable, the two presented auxiliary data are most probable derived from the same subject. More formally if $Dec(\mathrm{AD}_{\oplus})$ is successful the cross-matching classifier outputs a match. From (5.2) we can derive that $\mathrm{AD}_{\oplus}$ is decodable when $||\mathbf{e}|| \leq t_{\mathrm{c}}$ or $\mathbf{e} = \mathbf{e}^* \oplus \mathbf{C}_i$ with $||\mathbf{e}^*|| \leq t_{\mathrm{c}}$ and $\mathbf{C}_i \in \mathcal{C}$. Hence, the decodability attack exploits the same underlying mechanism as shown in Case 2 in Section 5.2.4 and has therefore the same performance. However, the effort is significantly reduced towards a single decoding operation by using the decoding function of the ECC. Similarly, only a distance measure between the binary vectors $\mathbf{f}_{\mathrm{B},1}^{\mathrm{e}}$ and $\mathbf{f}_{\mathrm{B},2}^{\mathrm{e}}$ can be obtained but not their actual value.

5.2.5 Relating the Cross-matching and System Performance

In Section 5.2.4 we presented several cross-matching attack methods from which the decodability attack is the most serious one because of its reduced effort towards a single decoding operation of the ECC. In this section we will determine the cross-matching classification performance in terms of the *false match rate* (FMR) and *false non-match rate* (FNMR) under the assumption that the subjects are in both databases. Furthermore, we compare the cross-matching performance with the system performance of the fuzzy commitment scheme. We assume the extracted bits to be independent with equal bit-error probability.

False Match Rate Relationship

Lemma 5.2.2 (FMR Relationship)**.** *Under the assumption that the bits of* $\mathbf{f}_{\mathrm{B}} \in \{0.1\}^{n_{\mathrm{c}}}$ *across the population are independent and uniform and given a* t_{c}*-error binary linear ECC, the cross-matching and system FMR,* α_{CM} *and* α_{TP} *respectively, at the error correcting threshold* t_{c} *are related according to* $\alpha_{\mathrm{CM}}(t_{\mathrm{c}}, n_{\mathrm{c}}) = 2^{k_{\mathrm{c}}}\alpha_{\mathrm{TP}}(t_{\mathrm{c}}, n_{\mathrm{c}})$.

Proof. The false-acceptance rate for the template protection system α_{TP} depends on the probability mass function (pmf) of the Hamming distance $\epsilon = d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}})$ at imposter comparisons. As presented in [104] under the assumption that the bits of $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ across the population are independent and uniform, the imposter Hamming distance pmf can be modeled by the binomial density

$$P_{\mathrm{b}}(d; N, p) \quad \stackrel{\mathrm{def}}{=} \binom{N}{d} p^d (1-p)^{(N-d)} \tag{5.3}$$

with dimension $N = n_{\mathrm{c}}$ and bit-error probability $p = P_{\mathrm{e}}^{\mathrm{im}} = \frac{1}{2}$, where $P_{\mathrm{e}}^{\mathrm{im}}$ is the bit-error probability at imposter comparisons. Due to the single-bit extraction scheme employing a quantization threshold that is equal to the background mean, the bit-error probability $P_{\mathrm{e}}^{\mathrm{im}}$ does not depend on either the number of enrolment N_{e} or verification N_{v} samples. Hence, the false-acceptance α_{TP} rate at threshold t_{c} is the following sum of the

binomial pmf

$$\begin{aligned}\alpha_{\text{TP}}(t_{\text{c}}, n_{\text{c}}) &\stackrel{\text{def}}{=} \sum_{i=0}^{t_{\text{c}}} P_{\text{b}}(i; n_{\text{c}}, P_{\text{e}}^{\text{im}}) \\ &= \tfrac{1}{2^{n_{\text{c}}}} \sum_{i=0}^{t_{\text{c}}} \tbinom{n_{\text{c}}}{i} \\ &= \tfrac{1}{2^{n_{\text{c}}}} V_2(n_{\text{c}}, t_{\text{c}}),\end{aligned} \tag{5.4}$$

where $V_2(n, r) = \sum_{i=0}^{r} \binom{n}{i}$ is the number of vectors in a sphere with radius r in $\{0,1\}^n$. An illustration of the binary vectors that will lead to a match at the verification phase is depicted in Figure 5.2. Examples of $\alpha_{\text{TP}}(t_{\text{c}}, n_{\text{c}})$ at several BCH ECC settings are given in Table 5.2. Increasing the codeword size n_{c} decreases the FMR. Increasing the key size k_{c} and therefore decreasing the error-correcting capability t_{c}, also decreases the FMR.

As shown in Section 5.2.4, the FMR of the cross-matching classifier α_{CM} is the probability that the XOR of the auxiliary data from two different subjects is decodable. As defined in [40], under the assumption that the bits of $\mathbf{f}_{\text{B}}$ are independent and uniform with $P_{\text{e}}^{\text{im}} = \frac{1}{2}$, the α_{CM} is equal to the probability of randomly selecting a word $\mathbf{w} \in_R \{0,1\}^{n_{\text{c}}}$ that is decodable, i.e. within t_{c} bits of any codeword from $\mathcal{C}$, namely

$$\alpha_{\text{CM}}(t_{\text{c}}, n_{\text{c}}) \stackrel{\text{def}}{=} \mathcal{P}\{d_{\text{c}}(\mathbf{w}, \mathcal{C}) \leq t_{\text{c}}\} = \frac{2^{k_{\text{c}}} V_2(n_{\text{c}}, t_{\text{c}})}{2^{n_{\text{c}}}}. \tag{5.5}$$

An illustration of the binary vectors that will lead to a match is shown in Figure 5.5. The $\alpha_{\text{CM}}(t_{\text{c}}, n_{\text{c}})$ is equal to the ratio of all possible vectors within the dashed circles with respect to all possible vectors in the $\{0,1\}^n$ space. Examples of $\alpha_{\text{CM}}(t_{\text{c}}, n_{\text{c}})$ at some BCH ECC settings are given in Table 5.2. Increasing the codeword size n_{c} decrease $\alpha_{\text{CM}}(t_{\text{c}}, n_{\text{c}})$, however increasing the key size k_{c} does not always decrease $\alpha_{\text{CM}}(t_{\text{c}}, n_{\text{c}})$. Note the special case of $n_{\text{c}} = 31$ with $[k_{\text{c}}, t_{\text{c}}] = [26, 1]$, where $\alpha_{\text{CM}} = 1$ because the full $\{0,1\}^{n_{\text{c}}}$ space is decodable. Thus, this $[n_{\text{c}}, k_{\text{c}}, t_{\text{c}}]$ setting of the BCH ECC leads to a perfect code.

Table 5.2: Examples of α_{TP} and α_{CM} for different $n_{\text{c}} \in \{127, 63, 31\}$ and $[k_{\text{c}}, t_{\text{c}}]$ settings.

	$\mathbf{n_c = 127}$			
$[k_{\text{c}}, t_{\text{c}}]$	$[8, 31]$	$[22, 23]$	$[36, 15]$	$[78, 7]$
α_{TP}	$3.16 \cdot 10^{-9}$	$8.48 \cdot 10^{-14}$	$7.89 \cdot 10^{-20}$	$5.57 \cdot 10^{-28}$
α_{CM}	$8.10 \cdot 10^{-7}$	$3.56 \cdot 10^{-7}$	$5.42 \cdot 10^{-9}$	$1.68 \cdot 10^{-4}$
	$\mathbf{n_c = 63}$			
$[k_{\text{c}}, t_{\text{c}}]$	$[7, 15]$	$[16, 11]$	$[24, 7]$	$[45, 3]$
α_{TP}	$1.88 \cdot 10^{-5}$	$8.37 \cdot 10^{-8}$	$6.82 \cdot 10^{-11}$	$4.52 \cdot 10^{-15}$
α_{CM}	$2.41 \cdot 10^{-3}$	$5.48 \cdot 10^{-3}$	$1.14 \cdot 10^{-3}$	$1.59 \cdot 10^{-1}$
	$\mathbf{n_c = 31}$			
$[k_{\text{c}}, t_{\text{c}}]$	$[6, 7]$	$[11, 5]$	$[16, 3]$	$[26, 1]$
α_{TP}	$1.66 \cdot 10^{-3}$	$9.61 \cdot 10^{-5}$	$2.32 \cdot 10^{-6}$	$1.49 \cdot 10^{-8}$
α_{CM}	$1.06 \cdot 10^{-1}$	$1.97 \cdot 10^{-1}$	$1.52 \cdot 10^{-1}$	1.00

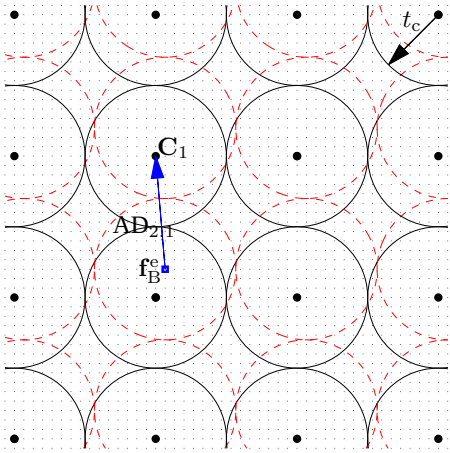


Figure 5.5: An illustration of the binary vectors that would lead to a match.

By combing the system FMR α_{TP} from (5.4) and the cross-matching FMR α_{CM} from (5.5) we obtain

$$\alpha_{\text{CM}}(t_c, n_c) = 2^{k_c}\alpha_{\text{TP}}(t_c, n_c), \tag{5.6}$$

which implies that the cross-matching FMR is 2^{k_c} times larger than the system FMR under the assumption that the bits of $\mathbf{f}_B^e$ across the population are independent and uniform. □

False Non-Match Rate Relationship

Lemma 5.2.3 (FNMR Relationship)**.** *Under the assumption that the bits of* $\mathbf{f}_B \in \{0.1\}^{n_c}$ *are independent with equal bit-error probability* P_e^{ge}*, given a balanced system where* $N_v = N_e$ *and a* t_c*-error binary linear ECC, the cross-matching* β_{CM} *at the error correcting threshold* t_c *is smaller than the system FNMR* β_{TP}*, namely* $\beta_{\text{CM}}(t_c, n_c) < \beta_{\text{TP}}(t_c, n_c)$*.*

Proof. For the template protection system, a false non-match occurs when $\epsilon = d_H(\mathbf{f}_B^e, \mathbf{f}_B^v) > t_c$ at genuine comparisons. Similar as in Section 5.2.5, we model the pmf of ϵ with a binomial density with dimension n_c, however with bit-error probability P_e^{ge}. The theoretical FNMR of the template protection system at threshold t_c, $\beta_{\text{TP}}(t_c, n_c)$, is the following sum of the binomial pmf

$$\beta_{\text{TP}}(t_c, n_c) \overset{\text{def}}{=} \sum_{i=t_c+1}^{n_c} P_b(i; n_c, P_e^{ge}). \tag{5.7}$$

For the cross-matching classifier, β_{CM} is the probability that the XOR of the auxiliary data $\mathrm{AD}_{2,1}$ and $\mathrm{AD}_{2,2}$ from the same subject at different databases is not decodable, hence an

non-match at a genuine comparison. As discussed in Section 5.2.4, the decodability probability is determined by the Hamming distance between the binary vectors at enrolment, namely $\epsilon_{CM} = d_{\mathrm{H}}(\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1}, \mathbf{f}^{\mathrm{e}}_{\mathrm{B},2})$. Because of the balanced system assumption the bit-error probability is also equal $P^{\mathrm{ge}}_{\mathrm{e}}$, consequently the pmf of $\epsilon_{CM} = d_{\mathrm{H}}(\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1}, \mathbf{f}^{\mathrm{e}}_{\mathrm{B},2})$ is equal to the pmf of $\epsilon = d_{\mathrm{H}}(\mathbf{f}^{\mathrm{e}}_{\mathrm{B}}, \mathbf{f}^{\mathrm{v}}_{\mathrm{B}})$ and for convenience we use ϵ in the remainder of this section. As discussed in Section 5.2.4, there is also a probability that when $\epsilon > t_{\mathrm{c}}$, the XOR of the auxiliary data $\mathrm{AD}_{\oplus}$ will also be decodable and hence correctly labeled as genuine. We define the decodability probability $P_{\mathrm{AD}_{\oplus}}(\epsilon; t_{\mathrm{c}}, \mathcal{C})$ as

$$P_{\mathrm{AD}_{\oplus}}(\epsilon; t_{\mathrm{c}}, \mathcal{C}) \quad \stackrel{\mathrm{def}}{=} \mathcal{P}\{d_{\mathrm{c}}(\mathrm{AD}_{\oplus}, \mathcal{C}) \leq t_{\mathrm{c}} \mid \epsilon\} \tag{5.8}$$

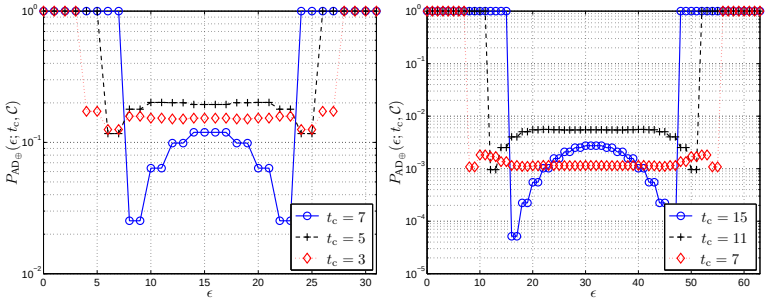
which has to be taken into account when estimating β_{CM} according to

$$\beta_{\mathrm{CM}}(t_{\mathrm{c}}, n_{\mathrm{c}}) \quad \stackrel{\mathrm{def}}{=} \sum_{i=t_{\mathrm{c}}+1}^{n_{\mathrm{c}}} \Big(1 - P_{\mathrm{AD}_{\oplus}}(i; t_{\mathrm{c}}, \mathcal{C})\Big) P_{\mathrm{b}}(i; n_{\mathrm{c}}, P^{\mathrm{ge}}_{\mathrm{e}}). \tag{5.9}$$

Observe that $\beta_{\mathrm{CM}}(t_{\mathrm{c}}, n_{\mathrm{c}})$ from (5.9) is equal to $\beta_{\mathrm{TP}}(t_{\mathrm{c}}, n_{\mathrm{c}})$ from (5.7) when $1 - P_{\mathrm{AD}_{\oplus}}(\epsilon; t_{\mathrm{c}}, \mathcal{C}) = 1$ for $\epsilon > t_{\mathrm{c}}$. In other words $P_{\mathrm{AD}_{\oplus}}(\epsilon; t_{\mathrm{c}}, \mathcal{C}) = 0$ stating that $\mathrm{AD}_{\oplus}$ should not be decodable for any cases of $\mathbf{f}^{\mathrm{e}}_{\mathrm{B}}$ and $\mathbf{f}^{\mathrm{v}}_{\mathrm{B}}$ with error pattern of weight $\epsilon > t_{\mathrm{c}}$. However, $\beta_{\mathrm{CM}}(t_{\mathrm{c}}, n_{\mathrm{c}}) < \beta_{\mathrm{TP}}(t_{\mathrm{c}}, n_{\mathrm{c}})$ if there is at least one case of $\mathbf{f}^{\mathrm{e}}_{\mathrm{B}}$ and $\mathbf{f}^{\mathrm{v}}_{\mathrm{B}}$ with error pattern of weight $\epsilon > t_{\mathrm{c}}$ where $\mathrm{AD}_{\oplus}$ is decodable. Hence, it suffice to prove that there is at least one case of $\mathbf{f}^{\mathrm{e}}_{\mathrm{B}}$ and $\mathbf{f}^{\mathrm{v}}_{\mathrm{B}}$ with error pattern of weight $\epsilon > t_{\mathrm{c}}$ where $\mathrm{AD}_{\oplus}$ is decodable.

Let the codebook be $\mathcal{C} = \{\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3\}$ with minimum distance $d = 2t_{\mathrm{c}} + 1$, where the codewords $\mathbf{C}_1$ and $\mathbf{C}_2$ are used in the enrolment phase of application 1 and 2, respectively, and $\mathbf{C}_3 = \mathbf{C}_1 \oplus \mathbf{C}_2$. Note that the XOR of the auxiliary data can be rewritten as $\mathrm{AD}_{\oplus} = (\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1} \oplus \mathbf{C}_1) \oplus (\mathbf{f}^{\mathrm{e}}_{\mathrm{B},2} \oplus \mathbf{C}_2) = \mathbf{e} \oplus \mathbf{C}_3$ with $\epsilon = ||\mathbf{e}||$ and is decodable for the $\epsilon > t_{\mathrm{c}}$ cases only if the error pattern can be rewritten as $\mathbf{e} = \mathbf{e}^* \oplus \mathbf{C}_i$ with $||\mathbf{e}^*|| \leq t_{\mathrm{c}}$ and $\mathbf{C}_i \in \{\mathbf{C}_1, \mathbf{C}_2\}$. Hence, there are at least two cases where $\mathrm{AD}_{\oplus}$ with $\epsilon > t_{\mathrm{c}}$ is decodable, namely the cases $\mathrm{AD}_{\oplus} = \mathbf{C}_1 \oplus \mathbf{C}_3$ or $\mathrm{AD}_{\oplus} = \mathbf{C}_2 \oplus \mathbf{C}_3$ where $||\mathbf{e}^*|| = 0$. □

Lemma 5.2.3 only states that $\beta_{\mathrm{CM}}(t_{\mathrm{c}}, n_{\mathrm{c}}) < \beta_{\mathrm{TP}}(t_{\mathrm{c}}, n_{\mathrm{c}})$ for any settings of t_{c} and n_{c}. In order to know the actual difference between $\beta_{\mathrm{CM}}(t_{\mathrm{c}}, n_{\mathrm{c}})$ and $\beta_{\mathrm{TP}}(t_{\mathrm{c}}, n_{\mathrm{c}})$ we have to determine $P_{\mathrm{AD}_{\oplus}}(\epsilon; t_{\mathrm{c}}, \mathcal{C})$ given a specific codebook $\mathcal{C}$. Assume we have an ECC with the codebook $\mathcal{C}$ consisting of one codeword of weight 0 ($\mathbf{C}_0$) and n_{c} ($\mathbf{C}_{n_{\mathrm{c}}}$) and n_w codewords $\mathbf{C}_w$ of weight w. Because of the properties of linear codes, each codeword has n_w neighbors at a distance w and one codeword at a distance n_{c}. Consider the case of being at codeword $\mathbf{C}_0$ and having a binary vector $\mathbf{w}_\epsilon$ with ϵ errors with respect to $\mathbf{C}_0$, hence having the weight w_ϵ. There are n_w neighboring codewords at a distance of w bits from $\mathbf{C}_0$, thus they have a weight of w. Furthermore, the error-correcting capability is equal to t_{c}. The fundamental question we want to answer is the decodability probability of the binary vector of $\mathbf{w}_\epsilon$. If its weight w_ϵ is within the error-correcting capability t_{c}, $w_\epsilon \leq t_{\mathrm{c}}$, $\mathbf{w}_\epsilon$ will always be decodable with respect to $\mathbf{C}_0$. However, if $w_\epsilon > t_{\mathrm{c}}$ the binary vector $\mathbf{w}_\epsilon$ will not be decodable with respect to $\mathbf{C}_0$ but there is a probability that $\mathbf{w}_\epsilon$ is decodable with respect to one of the n_w neighboring codewords at distance w. $\mathbf{w}_\epsilon$ will only be decodable if its distance to the neighboring codewords is smaller or equal to t_{c}, i.e. $||\mathbf{w}_\epsilon \oplus \mathbf{C}_w|| \leq t_{\mathrm{c}}$. In Section 5.2.3 we have discussed the probability $P_{w \times w}(\epsilon; w_1, w_2, n_{\mathrm{c}})$ of the weight of



(a) $n_c = 31$ (b) $n_c = 63$

Figure 5.6: $P_{\mathrm{AD}_\oplus}(\epsilon; t_c, \mathcal{C})$ values for different t_c settings at $n_c \in \{31, 63\}$.

the binary vector after XOR-ing two binary vectors of length n_c and weights w_1 and w_2, respectively. Hence, the decodability probability with respect to the n_w neighboring codewords of weight w is equal to $n_w \sum_{i=0}^{t_c} P_{w\times w}(i; w_\epsilon, w, n_c)$. Similarly, the decodability probability with respect to the codeword $\mathbf{C}_{n_c}$ has to be included, which is equal to $\sum_{i=0}^{t_c} P_{w\times w}(i; w_\epsilon, n_c, n_c)$.

For a general codebook $\mathcal{C}$, the decodability probability at ϵ errors, $P_{\mathrm{AD}_\oplus}(\epsilon; t_c, \mathcal{C})$, is given by

$$P_{\mathrm{AD}_\oplus}(\epsilon; t_c, \mathcal{C}) = \sum_{w \in W_\mathcal{C}} N_\mathcal{C}(w) \sum_{i=0}^{t_c} P_{w\times w}(i; \epsilon, w, n_c), \tag{5.10}$$

where $W_\mathcal{C}$ is the set of the unique weights w of the codewords from $\mathcal{C}$ and the function $N_\mathcal{C}(w)$ returns the number of codewords n_w with weight w, with $\sum_{w \in W_\mathcal{C}} N_\mathcal{C}(w) = |\mathcal{C}|$. Some examples of $P_{\mathrm{AD}_\oplus}(\epsilon; t_c, \mathcal{C})$ for the BCH code we consider are portrayed in Figure 5.6 for $n_c \in \{31, 63\}$ and different t_c settings. From these figures we can conclude that when $\epsilon \geq n_c - t_c$, $\mathrm{AD}_\oplus$ will always be decodable, because of the existence of a complementary codeword at distance n_c with respect to each codeword from $\mathcal{C}$. Furthermore, when $n_c = 31$ at most $\approx 20\%$ of the cases where $t_c < \epsilon < n_c - t_c$ are still decodable, which is significantly decreased to $\approx 0.6\%$ when $n_c = 63$. Some examples of β_{TP} and β_{CM} for different $n_c \in \{31, 63\}$ and $P_e^{\mathrm{ge}} \in \{0.20, 0.15\}$ settings are given in Table 5.3. There is no significant difference between β_{TP} and β_{CM} for the $n_c = 63$ case, however there is a clear difference for the $n_c = 31$ case.

Performance Relationship

Conjecture 5.2.1 (Performance Relationship)**.** *Under the assumption that the bits of* $\mathbf{f}_B \in \{0.1\}^{n_c}$ *are independent with equal bit-error probability* P_e^{ge} *and* $P_e^{\mathrm{im}} = \frac{1}{2}$ *at genuine and imposter comparisons respectively, given a balanced system where* $N_v = N_e$ *the cross-matching performance is worse than the system performance.*

With Lemma 5.2.2 we showed that FMR between the cross-matching and system is related according to $\alpha_{\text{CM}}(t_{\text{c}}, n_{\text{c}}) = 2^{k_{\text{c}}}\alpha_{\text{TP}}(t_{\text{c}}, n_{\text{c}})$, where the cross-matching FMR is $2^{k_{\text{c}}}$ worse than the system FMR. However, with Lemma 5.2.3 we showed that the FNMR at cross-matching is better than the system FNMR, however the difference is marginal at larger codeword lengths. In order to compare the overall performance we use the *receiver operating characteristic* (ROC) curves as illustrated in Figure 5.7 for the $n_{\text{c}} \in \{31, 63\}$ and $P_{\text{e}}^{\text{ge}} = 0.15$ settings. The system performance is given by the ROC labeled as TP_{b}, while the cross-matching performance is indicated by the points labeled with different markers representing the different $[k_{\text{c}}, t_{\text{c}}]$ settings of the ECC. Note that a performance is considered as being better when it is closer to the upper-left corner of the graph. Because the system ROC curve is clearly closer to the upper-left corner, we have shown that the system performance is better than the cross-matching performance.

5.2.6 Experiments

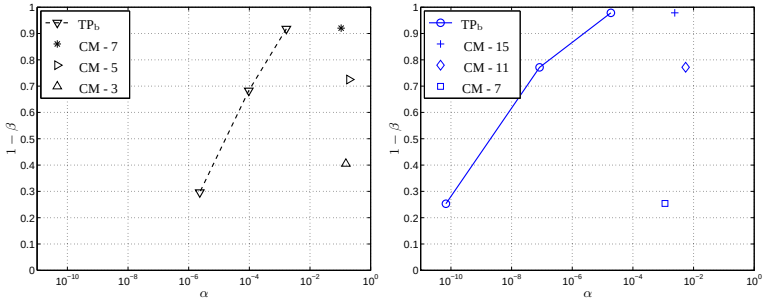
In this section we empirically estimate both the template protection system and cross-matching performance based on a fingerprint database in Section 5.2.6 and Section 5.2.6, respectively. The biometric database, feature extraction and evaluation protocol are described in Section 5.2.6.

Experimental Setup

Biometric Modality and Database The database we use is the MCYT (Ministerio de Ciencia y Tecnología) containing fingerprint images from a capacitive and optical sensor as described in [128]. It contains 12 images of all 10 fingers from $N_{\text{s}} = 330$ subjects for each sensor. However, we limit our dataset to the images of the right-index finger from the optical sensor.

Table 5.3: Comparison between β_{TP} and β_{CM} for different $n_{\text{c}} \in \{31, 63\}$, $[k_{\text{c}}, t_{\text{c}}]$ and $P_{\text{e}}^{\text{ge}} \in \{0.15, 0.20\}$ settings.

$\mathbf{n_c = 31}$				
$[k_{\text{c}}, t_{\text{c}}]$		$[6, 7]$	$[11, 5]$	$[16, 3]$
$P_{\text{e}}^{\text{ge}} = 0.15$	β_{TP}	0.0822	0.3173	0.7039
	β_{CM}	0.0796	0.2749	0.5948
$P_{\text{e}}^{\text{ge}} = 0.20$	β_{TP}	0.2700	0.6069	0.8930
	β_{CM}	0.2598	0.5176	0.7592
$\mathbf{n_c = 63}$				
$[k_{\text{c}}, t_{\text{c}}]$		$[7, 15]$	$[16, 11]$	$[24, 7]$
$P_{\text{e}}^{\text{ge}} = 0.15$	β_{TP}	0.0215	0.2287	0.7471
	β_{CM}	0.0215	0.2283	0.7460
$P_{\text{e}}^{\text{ge}} = 0.20$	β_{TP}	0.1789	0.6246	0.9527
	β_{CM}	0.1789	0.6231	0.9513



(a) $n_{\mathrm{c}} = 31$ (b) $n_{\mathrm{c}} = 63$

Figure 5.7: Performance comparison between the template protection system (TP_{b}) and cross-matching performance (CM) for the (a) $n_{\mathrm{c}} = 31$ and (b) $n_{\mathrm{c}} = 63$ case, under the assumption of independent bits with bit-error probabilities $P_{\mathrm{e}}^{\mathrm{im}} = 0.5$ and $P_{\mathrm{e}}^{\mathrm{ge}} = 0.15$, and a balanced system $N_{\mathrm{e}} = N_{\mathrm{v}}$. The suffix indicates t_{c}.

Feature Extraction Algorithms In order to compensate for possible translations between the enrolment and verification measurements, a translation-only pre-alignment step is performed during the feature extraction process. Such pre-alignment requires extraction of the core point which is performed according to the algorithm described in [129]. Around the core point we define a 17×17 grid with eight pixels between each grid point. The feature extraction algorithm extracts a feature value on each grid point. Our feature extraction algorithm failed to extract a feature vector from a single subject, so we excluded it from the dataset, hence there are effectively $N_{\mathrm{s}} = 329$ subjects.

The feature extraction method is based on the Gabor filter response, described in [107], where each grid point is filtered using a set of four 2D Gabor filters at angles of $\left\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\right\}$, respectively. The feature vector is the concatenation of the modulus of the four complex responses at each grid point, resulting into a feature vector dimension of $N_{\mathrm{F}} = 1156$.

Performance Evaluation Protocol The performance evaluation protocol consists of randomly selecting 219 out of $N_{\mathrm{s}} = 329$ subjects as the training set and the remaining 110 subjects as the evaluation set, which is referred to as the training-evaluation-set split. To decorrelate the feature components we use the principle component analysis (PCA) and the linear discriminant analysis (LDA) techniques. The PCA and LDA transformation matrices are computed using the training set, where N_{PCA} is the reduced dimension after applying the PCA transformation and N_{LDA} is the reduced dimension after applying the LDA transformation. Furthermore, the template protection system parameters such as the quantization thresholds, used within the *Bit Extraction* module, are also estimated on the training set.

From the evaluation set we evaluate both the system and cross-matching classification performance.

- For the *system performance* evaluation, N_{e} samples of each subject are randomly selected as the enrolment samples while the remaining samples are considered as the verification samples. The protected template is generated using all the N_{e} enrolment samples and compared with disjoint groups of N_{v} verification samples where the mean of the feature vectors is taken prior to the bit extraction process.

- For the *cross-matching performance* evaluation, we randomly select N_{e} samples for the enrolment for the first application and another random N_{e} samples for the second application as such that we have distinct samples for each application. For each application we create the protected template and compare all protected templates using the cross-matching classifier.

This split of creating the enrolment and verification set or the enrolment set for application one and two is referred to as the enrolment-verification split. If the verification sample is from the same subject as of the protected template, it is referred to as a genuine comparison, otherwise it is an imposter comparison.

Both the training-evaluation-set and the enrolment-verification splits are performed five times. Note that the splits are performed randomly, however the seed at the start of the protocol is always the same, hence all the splits are equal for the performance tests at different settings. Therefore, the splitting process does not contribute to any performance differences.

Template Protection System Performance

We evaluate the template protection system classification performance using the evaluation protocol in Section 5.2.6 with $N_{\mathrm{e}} = 6$ and $N_{\mathrm{v}} \in \{1, 6\}$. The case where $N_{\mathrm{v}} = N_{\mathrm{e}}$ is referred to as the balanced ($\mathrm{TP_b}$) case and the unbalanced ($\mathrm{TP_u}$) case when $N_{\mathrm{v}} \neq N_{\mathrm{e}}$.

The optimal N_{PCA} setting was found to be around 220 components and we set N_{LDA} equal to n_{c} to evaluate the performance. Note that we assume the FCS construction to act as a Hamming distance classifier as discussed in Section 5.2.3, hence we actually evaluate the scores $s_{\mathrm{TP}} = \epsilon = d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}})$ and limit the ROC curve at the threshold equal to t_{c}. The ROC curves for $n_{\mathrm{c}} \in \{31, 63\}$ settings are portrayed in Figure 5.8. For both n_{c} settings, the balanced case has a better performance because taking the average of N_{v} feature vectors suppresses the noise during verification which significantly improves the performance. Because of the BCH error-correcting limitation the FNMR is lower bounded and the FMR is upper bounded. The performance of the $n_{\mathrm{c}} = 63$ case is better, however the BCH limitation has a greater impact on the FNMR and FMR. Note that the estimated α_{TP} for both the balanced and unbalanced case are very similar, however they deviate from the theoretical expectation presented in Section 5.2.5. At t_{c}, the estimated FMR is ten times larger for the $n_{\mathrm{c}} = 63$ case, while twice larger for the $n_{\mathrm{c}} = 31$ case. The main cause of the deviating is the fact that the bits are still slightly dependent, while the theoretic work assumed independent bits. We omitted the $n_{\mathrm{c}} = 127$ case due to the limited dataset with respect to its small theoretic FMR at the maximum error-correcting capability t_{c}^*, namely $\alpha_{\mathrm{TP}}(t_{\mathrm{c}}^* = 31, n_{\mathrm{c}} = 127) \approx 3.16 \cdot 10^{-9}$.

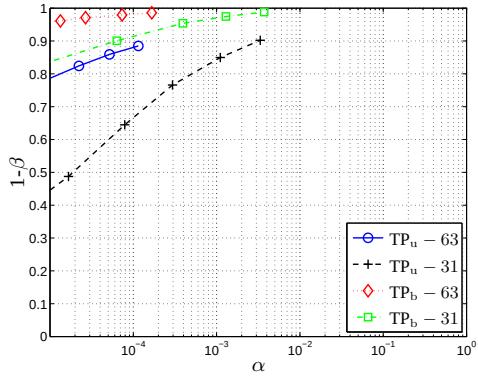


Figure 5.8: The ROC curve of the balanced and unbalanced and (TP $_{\mathrm{u}}$) template protection system derived from the $s_{\mathrm{TP}} = d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}})$ scores for the $n_{\mathrm{c}} \in \{31, 63\}$ settings. For the balanced case we have $N_{\mathrm{e}} = N_{\mathrm{v}} = 6$, while $N_{\mathrm{e}} = 6$ and $N_{\mathrm{v}} = 1$ for the unbalanced case.

Cross-Matching Performance Evaluation

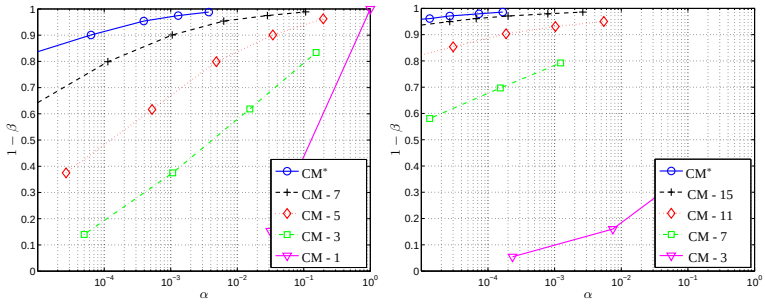
As discussed in Section 5.2.6 for the cross-matching (CM) performance evaluation we create two datasets containing the same subjects with $N_{\mathrm{e}} = 6$ distinct samples of each subject. The two datasets represent the enrolment samples for the two applications. From the each dataset we compute the binary vectors $\mathbf{f}_{\mathrm{B},1}^{\mathrm{e}}$ and $\mathbf{f}_{\mathrm{B},2}^{\mathrm{e}}$, and auxiliary data $\mathrm{AD}_{2,1}$ and $\mathrm{AD}_{2,2}$ from two randomly generated codewords $\mathbf{C}_1$ and $\mathbf{C}_2$, respectively.

The cross-matching classifier from the decodability attack, as presented in Section 5.2.4, is based on the property whether the XOR of the auxiliary data $\mathrm{AD}_{\oplus} = \mathrm{AD}_{2,1} \oplus \mathrm{AD}_{2,1}$ is decodable, i.e. $\mathit{Dec}(\mathrm{AD}_{\oplus})$ is successful, where *Dec* is the ECC decoding function. When successful the classifier outputs a match, otherwise a non-match. The decoding function of the BCH ECC we use does not return an error when it is not decodable, but returns the first k_{c} bits of $\mathrm{AD}_{\oplus}$ as the key instead. Therefore, we compute the cross-matching distance score s_{CM} as

$$\begin{aligned} s_{\mathrm{CM}} &= d_{\mathrm{CM}}(\mathrm{AD}_{2,1}, \mathrm{AD}_{2,2}) \\ &= d_{\mathrm{H}}(\mathrm{AD}_{\oplus}, \mathit{Enc}(\mathit{Dec}(\mathrm{AD}_{\oplus}))), \end{aligned} \tag{5.11}$$

where d_{CM} is the distance measure of the cross-matching classifier, and *Enc* and *Dec* are the encoding and decoding function of the BCH ECC, respectively. Consequently, we can extend the cross-matching classifier beyond the decision of either match or non-match with a score indicating how similar the comparison is.

The cross-matching performance ROC curves (CM) are depicted in Figure 5.9 for $n_{\mathrm{c}} = \{31, 63\}$ and different $[k_{\mathrm{c}}, t_{\mathrm{c}}]$ settings. Because of the availability of a score value instead of a decision, the ROC curves consist of multiple points instead of a single point as in Figure 5.7, where the outmost right-upper point corresponds to the decision-based performance. We also show the ROC curve from the Hamming distance of the enrolled



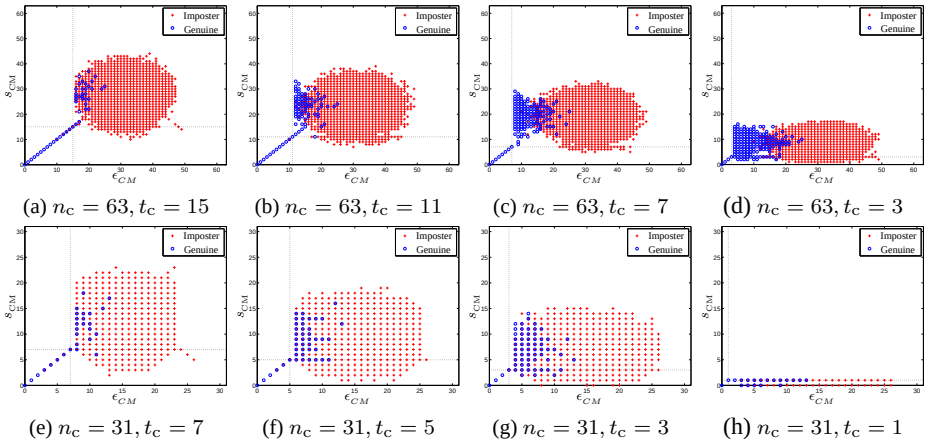
(a) $n_c = 31$ (b) $n_c = 63$

Figure 5.9: The ROC curve of cross-matching using AD 2 (CM) at different n_c and t_c indicated by the suffix. As reference, the ROC curve corresponding to $\epsilon_{CM} = d_H(\mathbf{f}^e_{B,1}, \mathbf{f}^e_{B,2})$ is used and is labeled as CM*.

binary vectors, $\epsilon_{CM} = d_H(\mathbf{f}^e_{B,1}, \mathbf{f}^e_{B,2})$, indicated by CM*. Note that the CM* ROC curve is equal to the balanced system performance ROC curve TP_b from Figure 5.8. Thus confirming the assumption made in Section 5.2.5 that the pmf of $\epsilon = d_H(\mathbf{f}^e_B, \mathbf{f}^v_B)$ is equal to the pmf of $\epsilon_{CM} = d_H(\mathbf{f}^e_{B,1}, \mathbf{f}^e_{B,2})$. With Figure 5.9 we also experimentally validate Lemma 5.2.1 dictating that the cross-matching performance is always worse than the balanced system performance. Also note that the difference significantly increases when t_c is decreased and thus increasing k_c. However, the cross-matching performance can be better than the unbalanced system performance as shown by the comparison of the $TP_u - 31$ and $TP_u - 63$ ROC curves from Figure 5.8 with the CM-7 and CM-15 curves from Figure 5.9(a) and Figure 5.9(b), respectively. *Hence, designing a balanced system with $N_e = N_v$ guarantees that the cross-matching performance is always worse than the system performance itself.*

For further analysis we show the comparison between the cross-matching Hamming distance ϵ_{CM} and distance score s_{CM} in Figure 5.10. These figures illustrate that for both the genuine and imposter comparisons if $\epsilon_{CM} \leq t_c$ than $s_{CM} \leq t_c$. Furthermore, from the imposter comparisons, notably for the $n_c = 31$ case, we also observe that when $\epsilon_{CM} \geq n_c - t_c$ than it holds that $s_{CM} = n_c - \epsilon_{CM}$, because for each codeword there also exists its complementary one with a distance of n_c bits. For the case when $t_c < \epsilon < n_c - t_c$, $AD_\oplus$ is occasionally decodable leading to a score $s_{CM} \leq t_c$ with probability $P_{AD_\oplus}(\epsilon; t_c, \mathcal{C})$ from (5.10) only when we can rewrite $(\mathbf{f}^e_{B,1} \oplus \mathbf{f}^e_{B,2}) = \mathbf{C}_i \oplus \mathbf{e}^*$ with $||\mathbf{e}^*|| \leq t_c$ and $\mathbf{C}_i \in \mathcal{C}$.

Also note that the average of the scores s_{CM}, for the cases when $AD_\oplus$ is not decodable and leading to a score $s_{CM} > t_c$, decreases when t_c decreases. Because of the systematic implementation of the BCH ECC and the fact that the decoding function of the ECC returns the first k_c bits as the key, guarantees that the first k_c bits between the corresponding codeword and $AD_\oplus$ are always equal while the remaining bits will be random. Hence, the expected bit difference is equal to $\frac{n_c - k_c}{2}$.



(a) $n_c = 63, t_c = 15$ (b) $n_c = 63, t_c = 11$ (c) $n_c = 63, t_c = 7$ (d) $n_c = 63, t_c = 3$

(e) $n_c = 31, t_c = 7$ (f) $n_c = 31, t_c = 5$ (g) $n_c = 31, t_c = 3$ (h) $n_c = 31, t_c = 1$

Figure 5.10: Comparison between $\epsilon_{CM} = d_{\mathrm{H}}(\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1}, \mathbf{f}^{\mathrm{e}}_{\mathrm{B},2})$ and $s_{\mathrm{CM}} = d_{\mathrm{CM}}(\mathrm{AD}_{2,1}, \mathrm{AD}_{2,2})$ for $n_{\mathrm{c}} = \{31, 63\}$ and different $[k_{\mathrm{c}}, t_{\mathrm{c}}]$ settings.

5.2.7 Decodability Attack Resilience with Bit-Permutation Randomization

We have shown that cross-matching is possible by using the decodability attack. However, if the system is designed as such that it is balanced, namely $N_{\mathrm{e}} = N_{\mathrm{v}}$, the cross-matching performance is always worse than the system performance, but still having a discriminating power. Ideally, it is preferred that the cross-matching performance is as close as possible to random.

In this section we introduce a randomization module within the FCS construction rendering the cross-matching performance close to random. As illustrated in Figure 5.11, prior to the XOR operation of the binary vector $\mathbf{f}^{\mathrm{e}}_{\mathrm{B}}$ and the codeword, we randomize $\mathbf{f}^{\mathrm{e}}_{\mathrm{B}}$ by multiplying it with a bit-permutation matrix $A_\pi \in \Pi$, obtaining $\mathbf{g}^{\mathrm{e}}_{\mathrm{B}} = A_\pi \mathbf{f}^{\mathrm{e}}_{\mathrm{B}}$, where A_π is a $n_{\mathrm{c}} \times n_{\mathrm{c}}$ matrix derived by randomly permuting the rows of the identity matrix and Π is the set of all possible permutation matrices. Because A_π is an orthogonal matrix its inverse is equal to its transpose, $A_\pi^{-1} = A'_\pi$. At each enrolment a new randomly generated bit-permutation matrix is used and stored as auxiliary data AD_3 and is considered as public. It is important to note that in the current approach the randomization matrix A_π is not considered to be secret, which is in contrast to earlier methods such as [89].

The XOR of the auxiliary data $\mathrm{AD}_{\oplus}$ can now be rewritten as

$$\begin{aligned} \mathrm{AD}_{\oplus} &= (\mathbf{g}^{\mathrm{e}}_{\mathrm{B},1} \oplus \mathbf{C}_1) \oplus (\mathbf{g}^{\mathrm{e}}_{\mathrm{B},2} \oplus \mathbf{C}_2) \\ &= (A_{\pi,1}\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1} \oplus A_{\pi,2}\mathbf{f}^{\mathrm{e}}_{\mathrm{B},2}) \oplus (\mathbf{C}_1 \oplus \mathbf{C}_2) \\ &= \mathbf{e}_\pi \oplus \mathbf{C}_3, \end{aligned} \tag{5.12}$$

with $\epsilon_\pi = ||\mathbf{e}_\pi|| = d_{\mathrm{H}}(A_{\pi,1}\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1}, A_{\pi,2}\mathbf{f}^{\mathrm{e}}_{\mathrm{B},2}) = d_{\mathrm{H}}(\mathbf{g}^{\mathrm{e}}_{\mathrm{B},1}, \mathbf{g}^{\mathrm{e}}_{\mathrm{B},2})$ being the number of errors after permutation instead of $\epsilon_{CM} = d_{\mathrm{H}}(\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1}, \mathbf{f}^{\mathrm{e}}_{\mathrm{B},2})$ when no permutation has been applied. Because of the randomization process it is likely that at genuine comparisons more errors

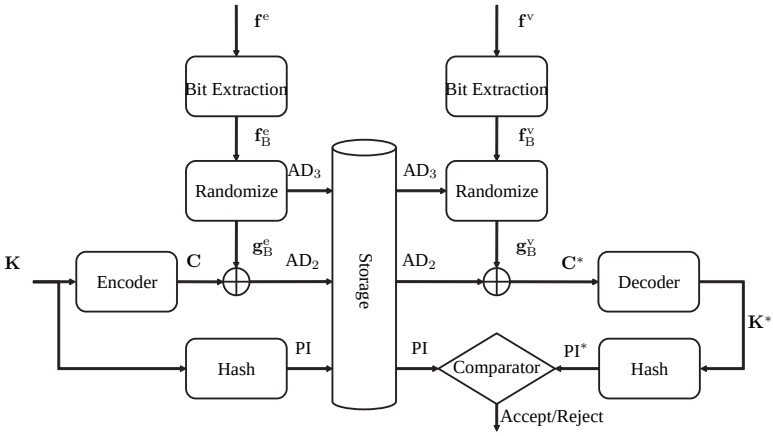


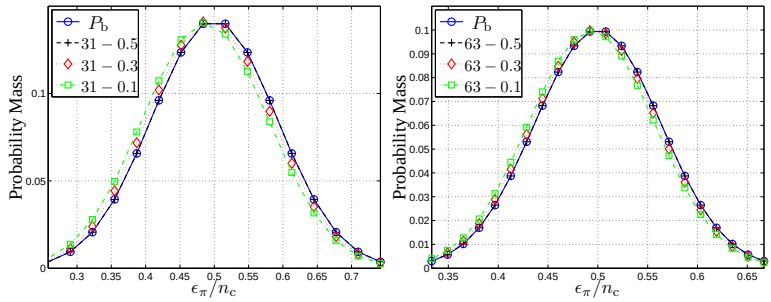
Figure 5.11: The code-offset system with randomization.

are introduced, namely $\epsilon_\pi > \epsilon_{CM}$, hence decreasing the probability that $\mathrm{AD}_\oplus$ is decodable, which significantly decreases when $\epsilon_\pi > t_{\mathrm{c}}$ (see Figure 5.6). As discussed in Section 5.2.4, under the assumption of having independent bits with bit-error probability $P_{\mathrm{e}}^{\mathrm{ge}}$ between genuine comparisons, the pmf of ϵ_π can be modeled by a binomial distribution with dimension n_{c} and $p = P_{\mathrm{e}}^{\mathrm{ge}}$, namely $P_{\mathrm{b}}(\epsilon_{CM}; n_{\mathrm{c}}, P_{\mathrm{e}}^{\mathrm{ge}})$. However, the pmf of ϵ_π will depend on both the pmf of ϵ_π and on the effect of the permutation, which we will analyze further. When the weight of the binary vectors $\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1}$ and $\mathbf{f}^{\mathrm{e}}_{\mathrm{B},2}$ are w_1 and w_2, respectively, the probability of ϵ_π number of errors after randomizing them is thus equal to $P_{w\times w}(\epsilon_\pi; w_1, w_2, n_{\mathrm{c}})$ as discussed in Section 5.2.3. Hence, the expected probability of ϵ_π irrespective of the weights $P_{\epsilon_\pi}(\epsilon_\pi; P_{\mathrm{e}}^{\mathrm{ge}}, n_{\mathrm{c}})$ is the average of $P_{w\times w}(\epsilon_\pi; w_1, w_2, n_{\mathrm{c}})$ across all possible weights. The possible combinations of w_1 and w_2 depend on the number of errors ϵ_{CM} between $\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1}$ and $\mathbf{f}^{\mathrm{e}}_{\mathrm{B},2}$. If w_1 and ϵ_{CM} are known than the probability of w_2 is determined by $P_{w\times w}(w_2; w_1, \epsilon_{CM}, n_{\mathrm{c}})$, because the error pattern can be considered as another binary vector of weight ϵ_{CM}. With the probability of randomly selecting a binary vector of weight w_1 equal to $P_{\mathrm{b}}(w_1; n_{\mathrm{c}}, \frac{1}{2})$, we obtain

$$\begin{aligned} P_{\epsilon_\pi}(\epsilon_\pi; P_{\mathrm{e}}^{\mathrm{ge}}, n_{\mathrm{c}}) \stackrel{\mathrm{def}}{=} \sum_{\epsilon_{CM}=0}^{n_{\mathrm{c}}} \sum_{w_1=0}^{n_{\mathrm{c}}} \sum_{w_2=0}^{n_{\mathrm{c}}} P_{w\times w}(\epsilon_\pi; w_1, w_2, n_{\mathrm{c}}) \times \\ \times P_{w\times w}(w_2; w_1, \epsilon_{CM}, n_{\mathrm{c}}) P_{\mathrm{b}}(w_1; n_{\mathrm{c}}, \tfrac{1}{2}) P_{\mathrm{b}}(\epsilon_{CM}; n_{\mathrm{c}}, P_{\mathrm{e}}^{\mathrm{ge}}), \end{aligned} \tag{5.13}$$

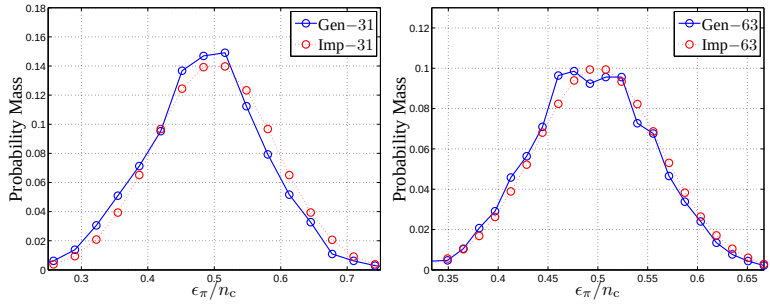
Figure 5.12 portrays the pmf of ϵ_π at genuine comparisons obtained with (5.13) for different settings of $P_{\mathrm{e}}^{\mathrm{ge}} \in \{\frac{1}{10}, \frac{3}{10}, \frac{1}{2}\}$ and $n_{\mathrm{c}} = \{31, 63\}$. As a reference we use the case where $\mathbf{e}_\pi$ is random binary vector with the pmf of its weight defined by the binomial pmf $P_{\mathrm{b}}(\epsilon_\pi; n_{\mathrm{c}}, \frac{1}{2})$. The figures show that the expected pmf of ϵ_π is very close to the case of being random, if either $P_{\mathrm{e}}^{\mathrm{ge}}$ and n_{c} increases the difference becomes smaller. If $P_{\mathrm{e}}^{\mathrm{ge}} = \frac{1}{2}$ the pmf of ϵ_π is equal to the case of being random.

Experimental results of the effects of the permutation randomization process, based



(a) $n_c = 31$ (b) $n_c = 63$

Figure 5.12: The pmf of $\epsilon_\pi = d_H(\mathbf{g}^e_{B,1}, \mathbf{g}^e_{B,2})$ from (5.13) at genuine comparisons for settings of $P^{ge}_e \in \{\frac{1}{10}, \frac{3}{10}, \frac{1}{2}\}$ and $n_c = \{31, 63\}$ compared with a binomial distribution $P_b(\epsilon_\pi; n_c, \frac{1}{2})$.

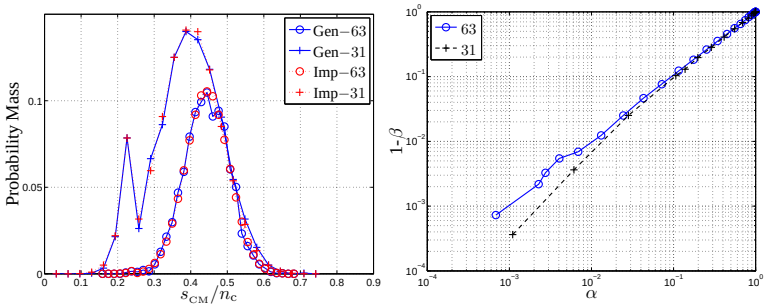


(a) $n_c = 31$ (b) $n_c = 63$

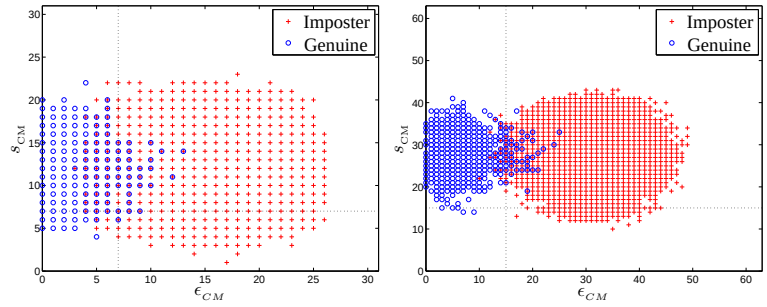
Figure 5.13: The pmf of $\epsilon_\pi = d_H(\mathbf{g}^e_{B,1}, \mathbf{g}^e_{B,2})$ at both the genuine (Gen) and imposter (Imp) comparisons for (a) $n_c = 31$ and (b) $n_c = 63$ settings.

on the same experimental setup from Section 5.2.6, are shown in Figure 5.13. We observe that the pmf of ϵ_π at genuine comparisons is close, however not equal, to the pmf at imposter comparisons, implying that it is difficult to distinguish a genuine comparison from an imposter comparison. These results confirm the theoretical expectations presented in Figure 5.12. Note that due to the fewer number of genuine comparisons than imposter comparisons, the pmf for the genuine case is more noisy.

Finally, the cross-matching performance with the randomization process is estimated based on the score s_{CM} from (5.11) and the results are shown in Figure 5.14. Figure 5.14(a) depicts the pmf of s_{CM} at genuine (Gen) and imposter comparisons (Imp) for the $n_c = \{31, 63\}$ settings. In contrast to the results in Figure 5.9 we also include the scores larger than t_c. Both the genuine and imposter pmfs are very similar, hence no distinguishing performance can be extracted by the adversary. The cross-matching ROC curve for the $n_c = \{31, 63\}$ settings are shown in Figure 5.14(b). As expected,



(a) s_{CM} pmf (b) ROC curves



(c) $n_{\mathrm{c}} = 31$, $t_{\mathrm{c}} = 15$ (d) $n_{\mathrm{c}} = 63$, $t_{\mathrm{c}} = 7$

Figure 5.14: (a) The pmf of s_{CM} and (b) the cross-matching ROC curve on logarithmic axes for $n_{\mathrm{c}} \in \{63, 31\}$, and the comparison of s_{CM} against $\epsilon_{CM} = d_{\mathrm{H}}(\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1}, \mathbf{f}^{\mathrm{e}}_{\mathrm{B},2})$ for (c) $n_{\mathrm{c}} = 31$ and (d) $n_{\mathrm{c}} = 63$.

the ROC curves are close to the one of a random classifier whose ROC curve is defined by $1 - \beta = \alpha$. Because of the limited genuine comparisons, the ROC curve for the $n_{\mathrm{c}} = 63$ case looks to be a bit worse than the random classifier. Furthermore, the comparison between s_{CM} and $\epsilon_{CM} = d_{\mathrm{H}}(\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1}, \mathbf{f}^{\mathrm{e}}_{\mathrm{B},2})$ are portrayed in Figure 5.14(c) and (d) for the $n_{\mathrm{c}} = 31$ and $n_{\mathrm{c}} = 63$ case, respectively. Due to the bit-permutation randomization process, the relationship between s_{CM} and ϵ_{CM}, as observed in Figure 5.10, no longer exists.

Inverting the Randomization Process

The randomization process and the bit-permutation matrix $A_{\pi,1}$ stored as auxiliary data AD_3 are considered as public. Hence, the adversary could apply the inverse on AD_2, namely $A'_{\pi,1}\mathrm{AD}_2$ with $A'_{\pi,1} = A^{-1}_{\pi,1}$, before applying the decodability attack on $\mathrm{AD}_{\oplus}$.

With the inverse process $\mathrm{AD}_{\oplus}$ becomes

$$\begin{aligned}\mathrm{AD}_{\oplus} &= A'_{\pi,1}\mathrm{AD}_{2,1} \oplus A'_{\pi,2}\mathrm{AD}_{2,2} \\ &= A'_{\pi,1}(\mathbf{g}^{\mathrm{e}}_{\mathrm{B},1} \oplus \mathbf{C}_1) \oplus A'_{\pi,2}(\mathbf{g}^{\mathrm{e}}_{\mathrm{B},2} \oplus \mathbf{C}_2) \\ &= (A'_{\pi,1}A_{\pi,1}\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1} \oplus A'_{\pi,2}A_{\pi,2}\mathbf{f}^{\mathrm{e}}_{\mathrm{B},2}) \oplus (A'_{\pi,1}\mathbf{C}_1 \oplus A'_{\pi,2}\mathbf{C}_2) \\ &= (\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1} \oplus \mathbf{f}^{\mathrm{e}}_{\mathrm{B},2}) \oplus (A'_{\pi,1}\mathbf{C}_1 \oplus A'_{\pi,2}\mathbf{C}_2),\end{aligned} \tag{5.14}$$

with $A'_{\pi}A_{\pi} = I$. Note that due to the inverse operation, additional errors may be introduced by the fact that both codewords are permuted by two different bit-permutation matrices, namely $(A'_{\pi,1}\mathbf{C}_1 \oplus A'_{\pi,2}\mathbf{C}_2) \in \mathcal{C}$. The additional errors guarantee that the cross-matching performance will be worse than the system performance. The only case where no errors are introduced is when $(A'_{\pi,1}\mathbf{C}_1 \oplus A'_{\pi,2}\mathbf{C}_2) \in \mathcal{C}$. We will show that this probability is very small, and thus there is a high probability that the cross-matching performance after taken the inverse is still worse than the system performance.

We will analyze this problem in two steps. First, given the codebook $\mathcal{C}$ we estimate the probability of obtaining a binary vector of weight w from $(A'_{\pi,1}\mathbf{C}_1 \oplus A'_{\pi,2}\mathbf{C}_2)$, defined as $P_{\pi^{-1}}(w;\mathcal{C})$. Hereafter, we estimate the probability that this binary vector is indeed a codeword, namely $P_{\pi^{-1}}(\mathcal{C})$.

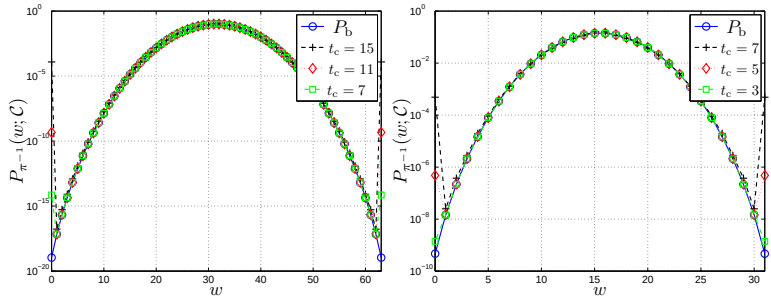
With $W_{\mathcal{C}}$ defined as the set of possible weights w of the codewords from $\mathcal{C}$ and the function $N_{\mathcal{C}}(w)$ returning the number of codewords n_w with weight w with $\sum_{w\in W_{\mathcal{C}}} N_{\mathcal{C}}(w) = |\mathcal{C}| = 2^{k_{\mathrm{c}}}$, the probability $P_{\pi^{-1}}(w;\mathcal{C})$ is equal to

$$\begin{aligned}P_{\pi^{-1}}(w;\mathcal{C}) &\stackrel{\mathrm{def}}{=} \mathcal{P}\big\{w = ||A'_{\pi,1}\mathbf{C}_1 \oplus A'_{\pi,2}\mathbf{C}_2|| \,| \\ &\qquad \forall\, \mathbf{C}_1, \mathbf{C}_2 \in \mathcal{C},\ A_{\pi,1}, A_{\pi,2} \in \Pi\big\} \\ &= \sum_{\substack{w_2\in W_{\mathcal{C}} \\ w_1\in W_{\mathcal{C}}}} \Big\{ P_{w\times w}(w;w_1,w_2,n_{\mathrm{c}})\times \\ &\qquad \times \tfrac{N_{\mathcal{C}}(w_1)N_{\mathcal{C}}(w_2)}{2^{2k_{\mathrm{c}}}} \Big\},\end{aligned} \tag{5.15}$$

where we take the sum, across all possible weights w_1 and w_2 of codewords $\mathbf{C}_1$ and $\mathbf{C}_2$, of the product of $P_{w\times w}(w;w_1,w_2,n_{\mathrm{c}})$ from (5.1) which is the probability that the XOR of two random binary vectors of weights w_1 and w_2 will lead to a binary vector of weight w, and $\frac{N_{\mathcal{C}}(w_1)N_{\mathcal{C}}(w_2)}{2^{2k_{\mathrm{c}}}}$ which is the probability of randomly selecting two codewords of weights w_1 and w_2 from $\mathcal{C}$. Figure 5.15 illustrates $P_{\pi^{-1}}(w;\mathcal{C})$ for different n_{c} and $[k_{\mathrm{c}}, t_{\mathrm{c}}]$ settings of the BCH ECC, compared with a binomial distribution $P_{\mathrm{b}}(w;n_{\mathrm{c}},\frac{1}{2})$. Note that $P_{\pi^{-1}}(w;\mathcal{C})$ is very similar to the binomial probability except at weights zero and n_{c}, where the difference increases when t_{c} increases. The weight, $w = ||A'_{\pi,1}\mathbf{C}_1 \oplus A'_{\pi,2}\mathbf{C}_2||$ is zero when both $||\mathbf{C}_1||$ and $||\mathbf{C}_2||$ are zero or n_{c}, or equal to n_{c} when one of the codewords has weight of zero and the other one n_{c}. Both cases have the probability $P_{\pi^{-1}}(0;\mathcal{C}) = P_{\pi^{-1}}(n_{\mathrm{c}};\mathcal{C}) = \frac{2}{2^{2k_{\mathrm{c}}}}$.

With $P_{\pi^{-1}}(w;\mathcal{C})$ we can estimate the probability $P_{\pi^{-1}}(\mathcal{C})$ of the occurrence where no additional errors are introduced when the adversary applies the inverse, namely

$$\begin{aligned}P_{\pi^{-1}}(\mathcal{C}) &\stackrel{\mathrm{def}}{=} \mathcal{P}\big\{(A'_{\pi,1}\mathbf{C}_1 \oplus A'_{\pi,2}\mathbf{C}_2) \in \mathcal{C} \,| \\ &\qquad \forall\, \mathbf{C}_1, \mathbf{C}_2 \in \mathcal{C}, \forall A_{\pi,1}, A_{\pi,2} \in \Pi\big\} \\ &= \sum_{w\in W_{\mathcal{C}}} P_{\pi^{-1}}(w;\mathcal{C}) \tfrac{N_{\mathcal{C}}(w)}{\binom{n_{\mathrm{c}}}{w}},\end{aligned} \tag{5.16}$$



(a) $n_c = 63$ (b) $n_c = 31$

Figure 5.15: The probability of obtaining a binary vector of weight $w = ||\mathbf{C}_1 \oplus A_\pi \mathbf{C}_2||$ given by $P_{\pi^{-1}}(w;\mathcal{C})$ from (5.15) for different n_c and t_c settings compared to a binomial distribution $P_b(\epsilon_\pi; n_c, 0.5)$.

where $\frac{N_\mathcal{C}(w)}{\binom{n_c}{w}}$ is the probability that the binary vector of weight w is a codeword. Some examples of $P_{\pi^{-1}}(\mathcal{C})$ for different n_c and [k_c,t_c] settings are given in Table 5.4. At smaller k_c settings $P_{\pi^{-1}}(\mathcal{C})$ is close to $\frac{4}{2^{2k_c}}$, which is the probability of only selecting codewords of either weight zero or n_c. For those cases, no additional errors are introduced by $A'_{\pi,1}\mathbf{C}_1 \oplus A'_{\pi,2}\mathbf{C}_2$. The probability $P_{\pi^{-1}}(\mathcal{C})$ can be reduced even further by removing these two codewords from the original codebook, thus obtaining the codebook $\mathcal{C}\backslash\{0, n_c\}$. The probability is then given by $P_{\pi^{-1}}(\mathcal{C}\backslash\{0, n_c\})$ and its value for the same n_c and [k_c,t_c] settings are given in Table 5.4. At smaller k_c values, $P_{\pi^{-1}}(\mathcal{C}\backslash\{0, n_c\})$ is significantly smaller than $P_{\pi^{-1}}(\mathcal{C})$. Hence, in order to be more robust against the inverse of the bit-permutation process prior to the decodability attack, it is recommended not to use the codewords of weight zero or n_c. The drawback is that the key space is reduced to $2^{k_c} - 2$, which becomes negligible for larger k_c values. However at larger k_c values both $P_{\pi^{-1}}(\mathcal{C})$ and $P_{\pi^{-1}}(\mathcal{C}\backslash\{0, n_c\})$ converge to each other. From the results of Figure 5.15, we observe that at larger k_c values it holds that $P_{\pi^{-1}}(w;\mathcal{C}) \approx P_b(w; n_c, \frac{1}{2})$, consequently (5.16) becomes

$$\begin{aligned} P_{\pi^{-1}}(\mathcal{C}) &= \sum_{w\in W_\mathcal{C}} P_{\pi^{-1}}(w;\mathcal{C}) \tfrac{N_\mathcal{C}(w)}{\binom{n_c}{w}} \\ &\approx \sum_{w\in W_\mathcal{C}} P_b(w; n_c, \tfrac{1}{2}) \tfrac{N_\mathcal{C}(w)}{\binom{n_c}{w}} \\ &= \sum_{w\in W_\mathcal{C}} \tfrac{\binom{n_c}{w}}{2^{n_c}} \tfrac{N_\mathcal{C}(w)}{\binom{n_c}{w}} \\ &= \tfrac{1}{2^{n_c}} \sum_{w\in W_\mathcal{C}} N_\mathcal{C}(w) \\ &= 2^{k_c - n_c} \end{aligned} \tag{5.17}$$

which is the probability of randomly guessing a codeword from $\mathcal{C}$. Empirical results shown in Figure 5.16 confirm that inverting the randomization process prior to applying the decodability attack does not give the adversary an advantage when using the decodability attack, because the ROC curve is still close to random.

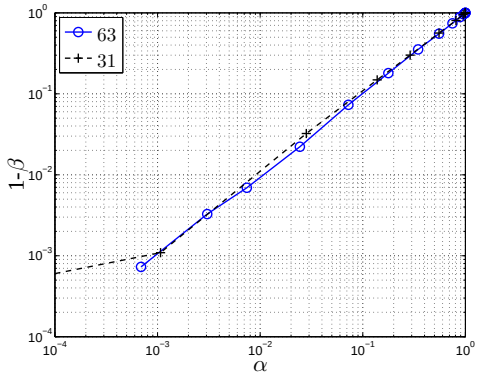


Figure 5.16: The cross-matching ROC curve when applying the decodability attack after inverting the randomization process on logarithmic axes for the $n_{\mathrm{c}} \in \{63, 31\}$ settings.

Ineffectiveness of the Noise-Addition Randomization Method

We will show that not all randomization processes will work. For example, taking the XOR of $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ with a random bit pattern δ, hence obtaining $\mathbf{g}_{\mathrm{B}}^{\mathrm{e}} = \mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \oplus \delta$ does not work, because this randomization process is fully reversible. When taking the XOR between $\mathrm{AD}_{2,1}$ and $\mathrm{AD}_{2,2}$ we obtain

$$\begin{aligned} \mathrm{AD}_{2,1} \oplus \mathrm{AD}_{2,2} &= (\mathbf{g}_{\mathrm{B},1}^{\mathrm{e}} \oplus \mathbf{C}_1) \oplus (\mathbf{g}_{\mathrm{B},2}^{\mathrm{e}} \oplus \mathbf{C}_2) \\ &= ((\mathbf{f}_{\mathrm{B},1}^{\mathrm{e}} \oplus \delta_1) \oplus (\mathbf{f}_{\mathrm{B},2}^{\mathrm{e}} \oplus \delta_2)) \oplus (\mathbf{C}_1 \oplus \mathbf{C}_2) \\ &= (\delta_1 \oplus \delta_2) \oplus (\mathbf{f}_{\mathrm{B},1}^{\mathrm{e}} \oplus \mathbf{f}_{\mathrm{B},2}^{\mathrm{e}}) \oplus (\mathbf{C}_1 \oplus \mathbf{C}_2) \end{aligned} \tag{5.18}$$

Hence, it is sufficient to take the XOR of the auxiliary data AD_2 with the publicly known bit pattern δ prior to applying the decodability attack, namely

$$\begin{aligned} (\delta_1 \oplus \mathrm{AD}_{2,1}) \oplus (\delta_2 \oplus \mathrm{AD}_{2,2}) &= (\delta_1 \oplus \delta_2) \oplus (\mathrm{AD}_{2,1} \oplus \mathrm{AD}_{2,2}) \\ &= (\mathbf{f}_{\mathrm{B},1}^{\mathrm{e}} \oplus \mathbf{f}_{\mathrm{B},2}^{\mathrm{e}}) \oplus (\mathbf{C}_1 \oplus \mathbf{C}_2), \end{aligned} \tag{5.19}$$

Table 5.4: The probability $P_{\pi^{-1}}(\mathcal{C})$ and $P_{\pi^{-1}}(\mathcal{C}\backslash\{0, n_{\mathrm{c}}\})$ for different settings of n_{c} and $[k_{\mathrm{c}}, t_{\mathrm{c}}]$.

$\mathbf{n_c = 31}$			
$[k_{\mathrm{c}}, t_{\mathrm{c}}]$	$[6, 7]$	$[11, 5]$	$[16, 3]$
$P_{\pi^{-1}}(\mathcal{C})$	$9.7660 \cdot 10^{-4}$	$1.9103 \cdot 10^{-6}$	$3.0521 \cdot 10^{-5}$
$P_{\pi^{-1}}(\mathcal{C}\backslash\{0, n_{\mathrm{c}}\})$	$2.8424 \cdot 10^{-8}$	$9.5271 \cdot 10^{-7}$	$3.0517 \cdot 10^{-5}$
$\mathbf{n_c = 63}$			
$[k_{\mathrm{c}}, t_{\mathrm{c}}]$	$[7, 15]$	$[16, 11]$	$[24, 7]$
$P_{\pi^{-1}}(\mathcal{C})$	$2.4414 \cdot 10^{-4}$	$9.3133 \cdot 10^{-10}$	$1.8332 \cdot 10^{-12}$
$P_{\pi^{-1}}(\mathcal{C}\backslash\{0, n_{\mathrm{c}}\})$	$1.3555 \cdot 10^{-17}$	$7.1052 \cdot 10^{-15}$	$1.8190 \cdot 10^{-12}$

because $(\delta_1 \oplus \delta_2) \oplus (\delta_1 \oplus \delta_2)$ cancel each other out. Hence, the adversary obtains the same error pattern $(\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1} \oplus \mathbf{f}^{\mathrm{e}}_{\mathrm{B},2}) \oplus (\mathbf{C}_1 \oplus \mathbf{C}_2)$ with which cross-matching is possible as shown in Section 5.2.4.

Effect on the Exhaustive Search Attack

In Section 5.2.4 we discussed both the decodability attack and the attacks based on exhaustive searches. With the bit-permutation process we reduced the effectiveness of the decodability attack, however both exhaustive attack methods still exist. With the bit-permutation process, the exhaustive search type of Case 1, where both the auxiliary data AD_2 and Pseudonymous Identifier PI are available, remains unchanged. By guessing the codeword from PI, the permuted binary vector $\mathbf{g}^{\mathrm{e}}_{\mathrm{B}}$ can be computed from which we can obtain $\mathbf{f}^{\mathrm{e}}_{\mathrm{B}}$ by inverting the bit-permutation process with A_{π}. However the exhaustive search type of Case 2, where only the auxiliary data is available, changes. The exhaustive search attack without the bit-permutation process as discussed in Section 5.2.4 has to search for a single codeword from the codebook $\mathcal{C}$ leading to the smallest distance score $s_{\mathrm{CM}} = \min_{\mathbf{C} \in \mathcal{C}} ||\mathrm{AD}_{\oplus} \oplus \mathbf{C}||$ with an average effort around $\approx 2^{k_{\mathrm{c}}-1}$. However, once the codeword was found there was still an ambiguity about the binary vector $\mathbf{f}^{\mathrm{e}}_{\mathrm{B}}$ of $2^{k_{\mathrm{c}}}$ possibilities. With the bit-permutation process, the XOR of the inverse of the auxiliary data of (5.2) becomes

$$\mathrm{AD}_{\oplus} \quad = (\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1} \oplus \mathbf{f}^{\mathrm{e}}_{\mathrm{B},2}) \oplus (A'_{\pi,1}\mathbf{C}_1 \oplus A'_{\pi,2}\mathbf{C}_2), \tag{5.20}$$

where the linear property of the ECC no longer holds as in (5.2). Instead of searching the codebook $\mathcal{C}$ only once, all combinations of the permuted codewords $A'_{\pi,1}\mathbf{C}_1 \oplus A'_{\pi,2}\mathbf{C}_2$ with known bit-permutation matrices has to be searched leading to the smallest distance score $s_{\mathrm{CM}} = \min_{\mathbf{C}_1,\mathbf{C}_2 \in \mathcal{C}} ||\mathrm{AD}_{\oplus} \oplus A'_{\pi,1}\mathbf{C}_1 \oplus A'_{\pi,2}\mathbf{C}_2||$. Thus, the effort has significantly increased towards $\approx 2^{2k_{\mathrm{c}}-1}$. However, once the codewords $\mathbf{C}_1$ and $\mathbf{C}_2$ have been found, the binary vector $\mathbf{f}^{\mathrm{e}}_{\mathrm{B}}$ is fully known. Hence, there is a trade-off between the case where cross-matching with the effortless decodability attack is possible with protection of the binary vectors or the case where cross-matching matching is possible with a significant effort of $2^{2k_{\mathrm{c}}-1}$ but revealing the binary vectors at a successful cross-match.

5.2.8 Conclusions

We analyzed the cross-matching performance of the auxiliary data AD_2 of the Fuzzy Commitment Scheme (FCS). We showed two attacks based on an exhaustive search, resulting in a significant attack effort, as well as a recently introduced attack requiring only a single decoding operation of the ECC, known as the decodability attack. Both attacks have the same cross-matching performance. To the best of our knowledge, the decodability attack was first presented in [39] and theoretically analyzed in [40]. We extended this theoretical analysis and showed the relationship between the balanced template protection system where $N_{\mathrm{v}} = N_{\mathrm{e}}$ and the cross-matching performance. The FMR at cross-matching is $2^{k_{\mathrm{c}}}$ larger than the FMR of the system, were k_{c} is the key size of the ECC. On the contrary, the FNMR at cross-matching is smaller than the FNMR of the system. However, the difference significantly decreases for larger n_{c} values. When comparing both

the FMR and FNMR in a ROC curve, we showed that the cross-matching performance is clearly worse than the system performance. We empirically validated the presented theoretical analysis using real biometric data from the MCYT fingerprint database. *Concluding, designing a balance template protection system with $N_{\mathrm{v}} = N_{\mathrm{e}}$ guarantees that the cross-matching performance is always worse than the system performance itself.*

Ideally, the cross-matching performance should be close to random. We provided a solution based on a bit-permutation randomization process that reduces the cross-matching performance of the decodability attack very close to random under the assumption that independent samples are taken for each application. During the enrolment phase, a random bit-permutation matrix is generated and used to permute the binary vector prior to creating the auxiliary data. We can consider the bit-permutation matrix of the randomization process to be publicly known because we have shown that the cross-matching performance is still close to random even when inverting the bit-permutation randomization process.

We showed the following trade-off. Without the proposed bit-permutation randomization process the decodability cross-matching attack is effortless, however without revealing the enrolled binary vectors. With the bit-permutation randomization process, the decodability cross-matching attack is neutralized however cross-matching based on exhaustive search is still possible. The effort of the exhaustive search increased towards $2^{2k_{\mathrm{c}}-1}$, instead of $2^{k_{\mathrm{c}}}$ when the bit-permutation randomization process is not applied. However, the effort increase is obtained with a drawback, namely revealing the enrolled binary vectors at a successful cross-match.

5.3 Chapter Conclusions

The vulnerability of cross-matching based on the decodability attack published in [39,40] indeed holds, however the cross-matching performance is worse than the classification performance of the HDS when it is balanced, i.e. the number of enrolment and verification samples are equal ($N_{\mathrm{e}} = N_{\mathrm{v}}$). The cross-matching FMR is a factor $2^{k_{\mathrm{c}}}$ larger than the FMR of the HDS with k_{c} being the key size, while the FNMR difference becomes negligible within increasing feature dimension. The cross-matching performance based on the decodability attack can be made close to random by introducing a bit-permutation matrix randomization process that permutes the binary vector. Note that the bit-permutation matrix is considered to be public data and has to be different for each application. Hence, the effect of the cross-matching performance on the unlinkability property is negligible. When implemented, what remains is the cross-matching based on exhaustive search with an average effort of $2^{2k_{\mathrm{c}}-1}$, however more protection beyond the exhaustive search cannot be guaranteed.

Chapter 6

Information Leakage Analysis of the Bit Extraction Part

6.1 Chapter Introduction

In this chapter the second part of the third research question will be addressed, namely

> **Given the HDS template protection scheme: How does the information leakage from the auxiliary data affect the irreversibility and unlinkability property?**

More specifically, this chapter answers this research question for the bit extraction part of the HDS of Figure 1.5. Firstly, in Section 6.2 we investigate the information leakage that could be exploited by an adversary to improve its impersonation success rate by increasing the FMR, thus affecting the irreversibility property. We focus only on the Detection Rate Optimized Bit Allocation (DROBA) bit extraction scheme proposed in Chen et al. (2009) [42]. We show with biometric data that the amount of information that AD_1 leaks is enough to increase the adversary's FMR by two orders of magnitude. Furthermore, we analyze the cause of the information leakage and provide a remedy. The main results are published in Kelkboom et al. (2009) [133] [1].

Secondly, we study the cross-matching performance affecting the unlinkability property of AD_1 of several bit extraction schemes that vary in the amount of subject-specific information that is used. We investigate the relationship between the improvement of the HDS performance by using more subject-specific information and the corresponding cross-matching performance. Results are published in Kelkboom et al. (2010) [134] [2].

[1] E. J. C. Kelkboom, K. T. J. de Groot, C. Chen, J. Breebaart, and R. N. J. Veldhuis, Pitfall of the detection rate optimized bit allocation within template protection and a remedy, in *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, 2009. (BTAS 09), 2009, pp. 1-8.

[2] E. J. C. Kelkboom, J. Breebaart, and R. N. J. Veldhuis, "Analysis of the system and cross-matching performance of bit extraction schemes with template protection," Submitted to *EURASIP Journal on Advances in Signal Processing*, 2010.

6.2 Pitfall of the Detection Rate Optimized Bit Allocation within Template Protection and a Remedy

6.2.1 Abstract

One of the requirements of a biometric template protection system is that the protected template ideally should not leak any information about the biometric sample or its derivatives. In the literature, several proposed template protection techniques are based on binary vectors. Hence, they require the extraction of a binary representation from the real-valued biometric sample. In this work we focus on the Detection Rate Optimized Bit Allocation (DROBA) quantization scheme that extracts multiple bits per feature component while maximizing the overall detection rate. The allocation strategy has to be stored as auxiliary data for reuse in the verification phase and is considered as public. This implies that the auxiliary data should not leak any information about the extracted binary representation. Experiments in our work show that the original DROBA algorithm, as known in the literature, creates auxiliary data that leaks a significant amount of information. We show how an adversary is able to exploit this information and significantly increase its success rate on obtaining a false accept. Fortunately, the information leakage can be mitigated by restricting the allocation freedom of the DROBA algorithm. We propose a method based on population statistics and empirically illustrate its effectiveness. All the experiments are based on the MCYT fingerprint database using two different texture based feature extraction algorithms.

6.2.2 Introduction

The widespread use of biometric systems introduces new privacy risks, for example identity fraud or cross-matching. These risks can be mitigated by applying template protection techniques. An overview of the privacy risks and template protection techniques are presented in [49]. A subclass of template protection techniques is based on a transformation of a biometric measurement to a binary vector as initial step. Hence, they require the extraction of a binary representation from the real- valued biometric sample. In the literature, numerous quantization schemes have been proposed. They vary from a simple method of extracting a single bit per feature component [33] [34] to a more complex, multiple bits per feature component, extraction method [44] [42]. If the quantization scheme is *subject-specific* the information has to be stored as *auxiliary data* for further use in the verification phase.

One of the requirements of a template protection system is that the stored auxiliary data ideally should not leak any information about the binary representation or the biometric sample itself. Hence, the subject-specific quantization scheme stored as the auxiliary data should not reveal any information that may facilitate an adversary on increasing its success rate guessing the binary representation of the biometric sample in order to obtain a false accept.

The work of [41] showed that the quantization schemes proposed in [135] and [136] do indeed leak information that could be exploited by an adversary. Their attack model is to guess the secret key in an off-line mode by using the auxiliary data and population

statistics. They use the guessing distance, consisting of the number of attempts required for a correct guess, as the measure of the degree of difficulty. Their results showed that the guessing distance is much smaller than what is expected based on the claimed security in [135] and [136], respectively. We focus on the Detection Rate Optimized Bit Allocation (DROBA) quantization scheme proposed in [42] that extracts multiple bits per feature component. For each enrolled subject the optimization algorithm allocates the optimal number of bits per component while maximizing the overall detection rate. The bit allocation strategy has to be stored as *auxiliary data* for further use during the verification phase.

Contribution: Our contribution is threefold. Firstly, we show that if the DROBA quantization scheme is not correctly implemented it will leak information about the binary representation of the biometric sample. Secondly, we illustrate an attack method an adversary could use in order to increase its success rate on reproducing a binary representation that leads to a false accept. Instead of using the guessing distance, we use the *false-acceptance* rate (FAR, α) as the degree of difficulty. We consider the template protection technique known as the helper-data system [33] [34] [35]. However, *any template protection technique incorporating the DROBA quantization scheme is susceptible to this vulnerability*. Thirdly, we outline a solution and propose an implementation guideline as a remedy. The remedy significantly mitigates the information leakage and guarantees a more private template.

The outline of this paper is as follows. In Section 6.2.3 we briefly discuss the considered template protection system with the DROBA quantization scheme. In Section 6.2.4 we describe our experimental setup concerning a fingerprint database, two feature extraction algorithms, and a testing protocol followed by the analysis of the information leakage due to the improper implementation of the DROBA quantization scheme. With use of the information leakage we demonstrate an attack method in Section 6.2.5 that significantly increases the false accept probability. As a remedy, we propose an implementation guideline in Section 6.2.6 and show that it significantly mitigates the information leakage. We finish with the conclusions in Section 6.2.7.

6.2.3 Template Protection Scheme with DROBA

The template protection technique under consideration is known as the helper-data system [33] [34] [35] and is portrayed in Figure 6.1. As input we have the real-valued feature vector of dimension N_{F}, $\mathbf{f} \in \mathbb{R}^{N_{\mathrm{F}}}$, which is extracted from the biometric sample by the feature extraction algorithm. Subsequently, a binary vector $\mathbf{f}_{\mathrm{B}} \in \{0,1\}^{N_{\mathrm{B}}}$ is extracted by the DROBA quantization module and outputs the first auxiliary data AD_1 containing the allocation strategy. Many template protection schemes are based on the capability of generating a robust binary vector or key out of different biometric measurements of the same subject. However, the binary vector $\mathbf{f}_{\mathrm{B}}$ itself cannot be used as the key because it is most likely not exactly the same in both the enrollment and verification phase ($\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \neq \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$), due to measurement noise and biometric variability that lead to *bit errors*. The number of bit errors between two binary vectors is also referred to as the Hamming distance

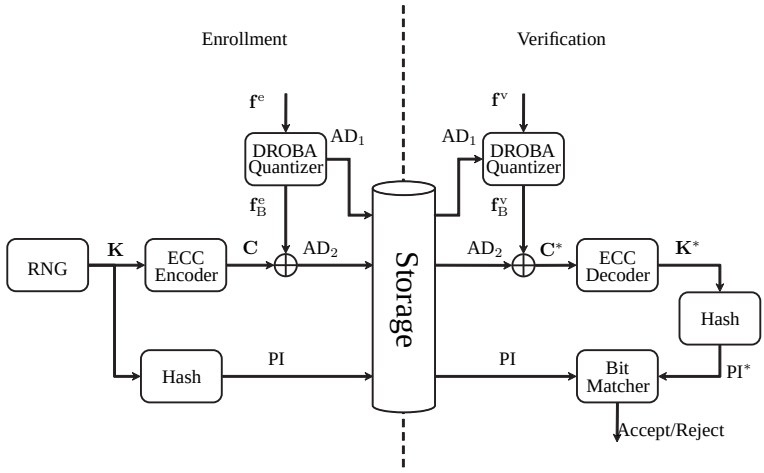


Figure 6.1: Template protection scheme with DROBA implementation.

(HD) $d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}})$. Therefore, ECCs are used to deal with these bit errors. As shown in Figure 6.1, the ECC and hash function are integrated using the well-known Fuzzy Commitment scheme [36]. For the sake of coherence we use the terminology proposed in [102].

Within the fuzzy commitment scheme we use the linear block type ECC "Bose, Ray-Chaudhuri, Hocquenghem" (BCH) that corrects random errors. The codeword $\mathbf{C}$ corresponding to a randomly generated secret $\mathbf{K}$ is XOR-ed with the $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ in order to obtain the auxiliary data AD_2. Furthermore, the hash of $\mathbf{K}$ is taken in order to obtain the pseudo identity PI. In the verification phase this process is reversed with help of the auxiliary data resulting into a candidate pseudonymous identifier PI^*. Only when $d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}) \leq t_{\mathrm{c}}$ then PI and PI^* are equal, thus resulting into an accept. Hence, the Fuzzy Commitment scheme can be considered as a HD-classifier. More details about the template protection system can be found in [33] [34].

As mentioned previously, the binary vector $\mathbf{f}_{\mathrm{B}}$ is extracted from the real-valued input vector $\mathbf{f}$ by the DROBA quantization scheme and algorithm proposed in [42]. The DROBA algorithm has the flexibility to extract multiple bits from a single component. The number of bits extracted from component j is given by b_j. The quantization schemes for the $b_j \in \{1, 2, 3\}$ cases are shown in Figure 6.2(a), (b), and (c), respectively. For convenience we refer the $b_j = 1$ case as b_1^*, and b_2^* and b_3^* for the $b_j = 2$ and $b_j = 3$ cases, respectively. The 2^{b_j} quantization intervals are defined as such that the occurrence of each interval is equiprobable with respect to the *total* density, which we assume to be Gaussian distributed $p_{\mathrm{t}} \sim \mathcal{N}(\mu_{\mathrm{t}}, \sigma_{\mathrm{t}}^2)$ with mean μ_{t} and variance σ_{t}^2. The total density defines the observed variability of that component across the whole population. Each quantization interval is assigned a unique b_j bits Gray code [137]. Furthermore, we model the observed biometric variability and measurement errors of the feature vector component of

a specific subject with the *within-class* density, which for simplicity is assumed to be another Gaussian density $p_{\mathrm{w}} \sim \mathcal{N}(\mu_{\mathrm{w}}, \sigma_{\mathrm{w}}^2)$. Note that μ_{w} and σ_{w}^2 can be different for each component or subject. From [42] the detection rate γ is defined as the probability that the next measurement of the feature component will be in the same quantization interval. For component j the detection rate is computed as

$$\gamma_j(b_j) = \int_{Q_{\mu_{\mathrm{w}}}(b_j)} p_{\mathrm{w}}(v)\mathrm{d}v, \tag{6.1}$$

where $Q_{\mu_{\mathrm{w}}}(b_j)$ is the quantization interval corresponding to μ_{w} and also depends on the number of bits b_j to be extracted. Thus, the detection rate is the part of the within-class density within the quantization interval corresponding to μ_{w}, portrayed by the shaded area in Figure 6.2. For the case where no bits are extracted ($b_j = 0$) the detection rate is defined as $\gamma_j(0) = 1$. Note that the detection rate decreases when b_j increases. Under the assumption that the N_{F} feature components are independent, the overall detection rate is defined as

$$\gamma_t = \prod_{j=1}^{N_{\mathrm{F}}} \gamma_j(b_j). \tag{6.2}$$

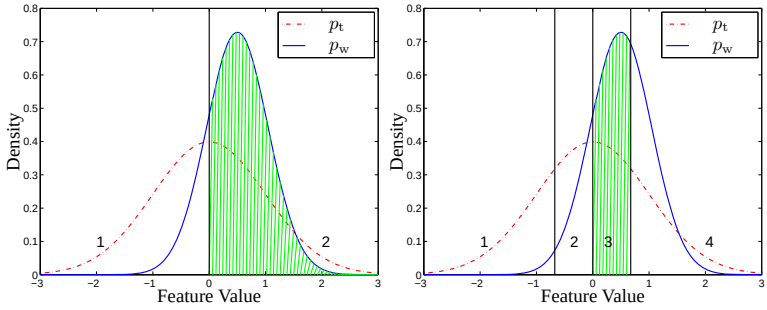
The DROBA algorithm has to create a binary vector of length N_{B}, hence it has to allocate N_{B} bits across all components. We also refer to N_{B} as the bit-budget. With use of the multiple (N_{e}) enrollment samples, the DROBA algorithm analyzes the subject-dependent feature statistics (μ_{w} and σ_{w}^2) of each component and allocates the optimal number of bits b_j to component i with the constrains of maximizing the overall detection rate γ_t and allocating the bit-budget $\sum_{j=1}^{N_{\mathrm{F}}} b_j = N_{\mathrm{B}}$. The optimal allocation strategy is stored as auxiliary data $\mathrm{AD}_1 = [b_1, b_2, \ldots, b_{N_{\mathrm{F}}}]$ for reuse at the verification phase. The optimization is implemented using the dynamic programming approach presented in [42].

6.2.4 Experiments

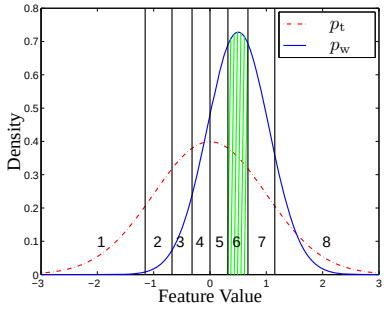
If the DROBA implementation is correct, auxiliary data AD_1 should not leak any information about the enrolled binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$. We will empirically analyze whether there is any information leakage by means of a fingerprint database and two feature extraction algorithms. We first discuss the experiment setup including the testing protocol followed by the information leakage analysis.

Experiment Setup

Biometric Modality and Database The database we use is the MCYT (Ministerio de Ciencia y Tecnología) containing fingerprint images [128]. It contains 12 images of all 10 fingers from $N_{\mathrm{s}} = 330$ subjects. However, we limit our dataset to the images of the right-index finger only.



(a) b_1^* case with 2 intervals (b) b_2^* case with 4 intervals



(c) b_3^* case with 8 intervals

Figure 6.2: The total density p_t with an example of a within-class density p_w and the corresponding detection rate γ_j at different quantization scheme where (a) $b_j = 1$ (b_1^*), (b) $b_j = 2$ (b_2^*), (c) $b_j = 3$ (b_3^*) bits are extracted.

Feature Extraction Algorithms Two types of texture based features are extracted from a fingerprint, namely *directional field* and *Gabor* features. In order to compensate for possible translations between enrolled and verification measurements, a translation-only pre-alignment step is performed during the feature extraction process. Such pre-alignment requires extraction of the core point which is performed according to the algorithm described in [129]. Around the core point we define a 17×17 grid with eight pixels between each grid point. The following feature extraction algorithms extract a feature value on each grid point.

The first feature extraction algorithm is based on directional fields. A directional field vector describes the estimated local ridge-valley edge orientation in a fingerprint structure and is based on gradient vectors. The orientation of the ridge-valley edge is orthogonal to the gradient's angle. Therefore a directional field vector that signifies the orientation of the ridge-valley edge is perpendicular positioned to the gradient vector. In order to extract directional field features from a fingerprint the algorithm described in [130] is applied on each grid point. The direction field features have a dimension of $N_F = 578$ and are referred to as the DF features.

The second type of extracted features are the Gabor (GF) features, described in [107], where each grid point is filtered using a set of four 2D Gabor filters at angles of $\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$. The feature vector is the concatenation of the modulus of the four complex responses at each grid point, resulting into a feature vector dimension of $N_{\mathrm{F}} = 1156$.

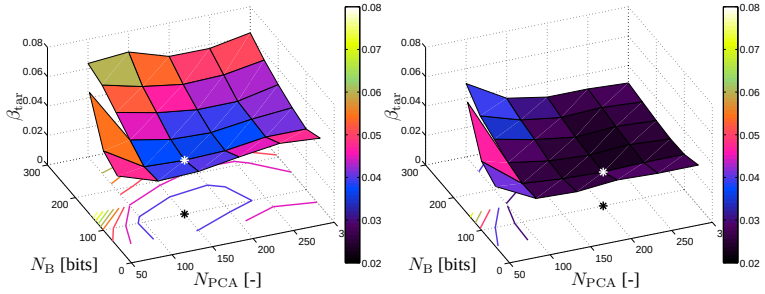
Testing Protocol The performance testing protocol consists of randomly selecting 220 out of N_{s} subjects as the training set and the remaining 110 subjects as the evaluation set, which is referred to as the training-evaluation-set split. To decorrelate the feature components we use the principle component analysis (PCA) and the linear discriminant analysis (LDA) techniques, where the LDA transformation is also used to obtain more discriminating feature components from which we expect to extract more bits from. The PCA and LDA transformation matrices are computed using this training set, where N_{PCA} is the reduced dimension after applying the PCA transformation and N_{LDA} is the reduced dimension after applying the LDA transformation. To avoid singularities we ensure that $N_{\mathrm{LDA}} \leq 220$. Furthermore, the template protection system parameters such as the quantization thresholds, used within the *Bit Extraction* module, are also estimated on the training set. From the evaluation set, 6 samples of each subject are randomly selected as the enrollment samples while the remaining samples are considered as the verification samples. This split is referred to as the enrollment-verification split. The protected template is generated using all the enrollment samples and compared with each individual verification sample. When the verification sample is from the same subject as of the protected template, it is referred to as a genuine comparison, otherwise it is an imposter comparison.

The training-evaluation-set split is performed five times, while for each of these splits the enrollment-verification split is performed 3 times. From each enrollment-verification split we estimate the β_{tar} (the *false-rejection* rate (FRR, β) at the targeted FAR of $\alpha_{\mathrm{tar}} = 0.1\%$) and the *equal-error* rate (EER) where the FAR is equal to the FRR. Note, that the splits are performed randomly, however the seed at the start of the protocol is always the same, hence all the splits are equal for the performance tests at different settings. Hence, the splitting process does not contribute to any performance differences.

Analysis of the Information Leakage

First of all we empirically derive the $\{N_{\mathrm{PCA}}, N_{\mathrm{LDA}}, N_{\mathrm{B}}\}$ setting leading to the optimal performance in terms of β_{tar}. We evaluate the performance for the settings of $N_{\mathrm{PCA}} \in \{50, 100, \ldots, 300\}$ and $N_{\mathrm{B}} \in \{50, 100, \ldots, \min(N_{\mathrm{PCA}} \cdot b_{\max}, 300)\}$, while the N_{LDA} parameter is set to $N_{\mathrm{LDA}} = \min(N_{\mathrm{PCA}}, 220)$ as discussed in Section 6.2.4. The achieved β_{tar} performance for the different $\{N_{\mathrm{PCA}}, N_{\mathrm{LDA}}, N_{\mathrm{B}}\}$ settings are depicted in Figures 6.3(a) and (b) for the DF and GF features, respectively.

For the DF features the optimal setting is achieved at $\{150, 150, 100\}$, while at $\{200, 200, 100\}$ for the GF features. At the optimal performance settings, the error-rate (α and β) curves with respect to the relative Hamming distance (RHD) between $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ and $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$ is portrayed in Figure 6.4(a) and (b) for the DF and GF features, respectively. The β_{tar} is 3.66% for the DF features and 2.30% for the GF features, while the EER is 1.49% and 1.29%, respectively.

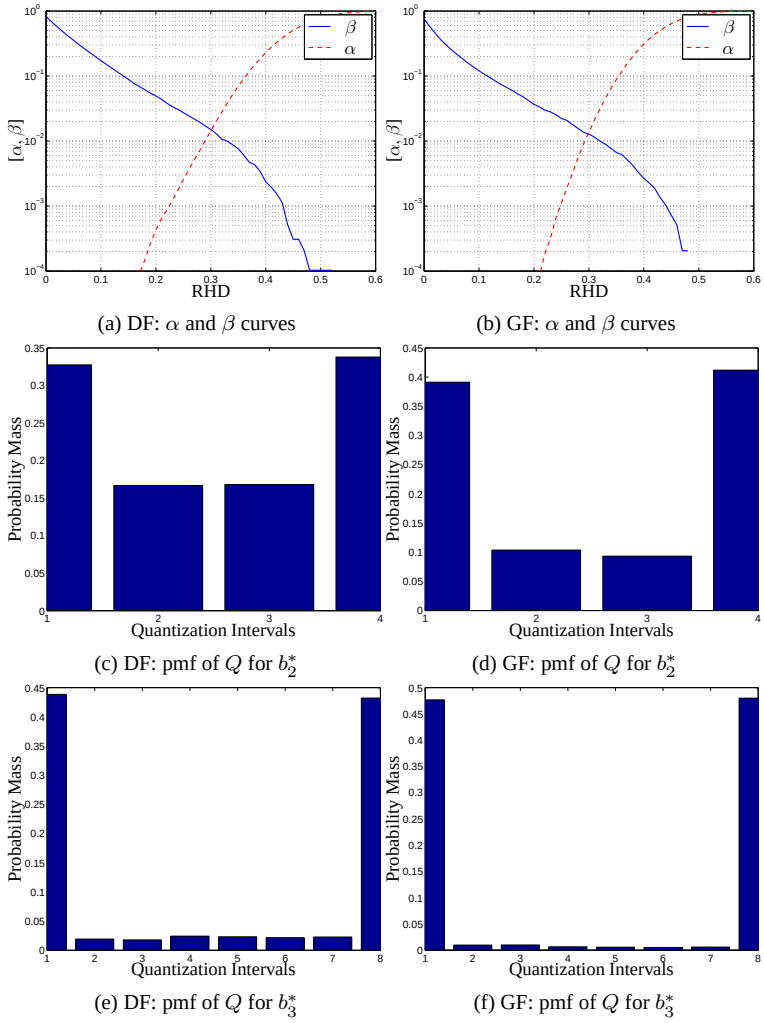


(a) DF: β_{tar} (b) GF: β_{tar}

Figure 6.3: The β_{tar} for different $\{N_{\mathrm{PCA}}, N_{\mathrm{LDA}}, N_{\mathrm{B}}\}$ settings for the DF and GF features. The optimal performance for each case is indicated by both the black and white star.

If the DROBA implementation is correct, AD_1 should not leak any information about the enrolled binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$. We know that AD_1 is a concatenation of b_i of each feature component, hence knowing b_i should not leak any information about the actual b_i allocated bits. The allocated bits are equal to the Gray code assigned to the quantization interval in which the sample mean μ_{w} of the subject is measured. This implies that the probability of each quantization interval across the population should be equal irrespective of b_i. Hence, we analyze the probability of each quantization interval, referred to as the probability mass function (pmf) of Q, where we represent the quantization intervals by a discrete random variable Q. For the b_1^* case the pmf is uniform, however for the b_2^* and b_3^* cases a significantly non-uniform pmf is observed, see Figure 6.4(c-f). For the b_2^* case roughly 66% of the cases μ_{w} is found to be in the outer quantization intervals for the DF features, while 80% for the GF features. For the b_3^* case it is around 87% for the DF feature and around 96% for the GF features. Due to the cyclic nature of Gray codes, the binary codes assigned to the outer quantization intervals differ in only a single bit. Hence, if multiple bits are extracted it is an advantage for the adversary to randomly select the binary code corresponding to one of the outer quantization intervals when guessing the binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$.

In order to illustrate at which $\{N_{\mathrm{PCA}}, N_{\mathrm{LDA}}, N_{\mathrm{B}}\}$ settings the most non-uniform pmf of Q is obtained, we define δ as the difference between the average probability of the two outer quantization intervals and the average probability of the remaining inner intervals. Hence, the closer δ is to zero the more the pmf is uniform and its maximum value is $\frac{1}{2}$. Furthermore, δ_2 is defined for the b_2^* case and δ_3 is for the b_3^* case. The δ values for the different settings are depicted in Figure 6.5. From the figures we can observe that the non-uniformity is stronger when N_{B} decreases or N_{PCA} increases, which corresponds to the cases where the DROBA algorithm has more freedom to allocate the N_{B} bits. The maximum observed values are $\delta_2 = 0.256$ and $\delta_3 = 0.458$ of the DF features and $\delta_2 = 0.360$ and $\delta_3 = 0.485$ for the GF features. The pmf is close to uniform when $N_{\mathrm{B}} \approx b_{\max} N_{\mathrm{PCA}}$, which is the case where the maximum number of bits is mostly extracted



(a) DF: α and β curves

(b) GF: α and β curves

(c) DF: pmf of Q for b_2^*

(d) GF: pmf of Q for b_2^*

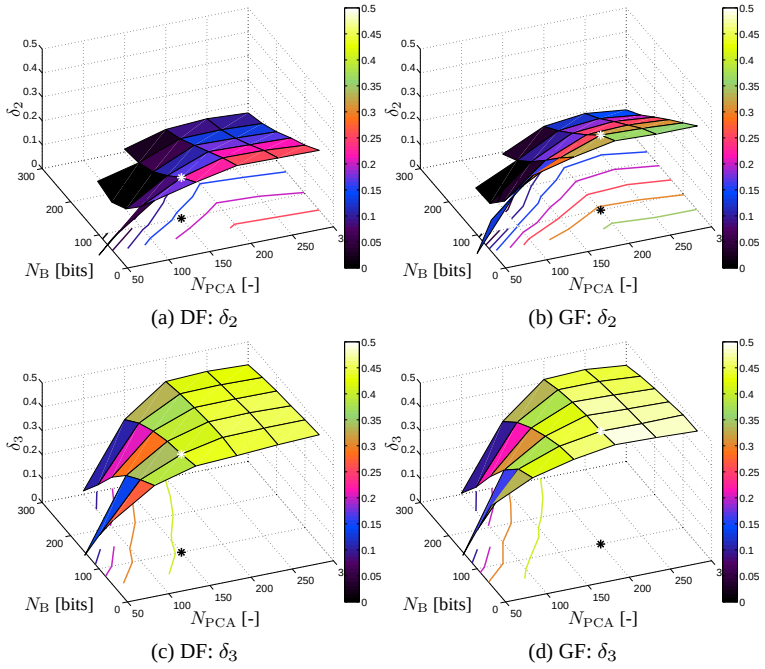
(e) DF: pmf of Q for b_3^*

(f) GF: pmf of Q for b_3^*

Figure 6.4: The error-rate curves for and the pmf of Q for the b_2^* and b_3^* cases, for the DF and GF features.

from each component. Note that at the optimal setting (indicated by the black and white star) the non-uniformity is close to its strongest.

Furthermore, we define $p(b_x^*)$ to be the average probability that a bit is derived from a b_x^* case. The $p(b_x^*)$ probabilities are different for each $\{N_{\mathrm{PCA}}, N_{\mathrm{LDA}}, N_{\mathrm{B}}\}$ setting as shown in Figures 6.6 for the $p(b_2^*)$ and $p(b_3^*)$ cases for the DF and GF features. Because the sum of the probabilities is one, the probability $p(b_1^*)$ can be derived from $p(b_2^*)$ and $p(b_3^*)$. The figures show that if N_{B} increases, more bits are extracted from the b_3^* case and less from the b_1^* case. The number of bits extracted from the b_2^* case stays relatively stable.



(a) DF: δ_2 (b) GF: δ_2

(c) DF: δ_3 (d) GF: δ_3

Figure 6.5: The δ_2 and δ_3 for different settings of N_{PCA} and N_{B} for the DF and GF features. The optimal performance setting is indicated with both the black and white star.

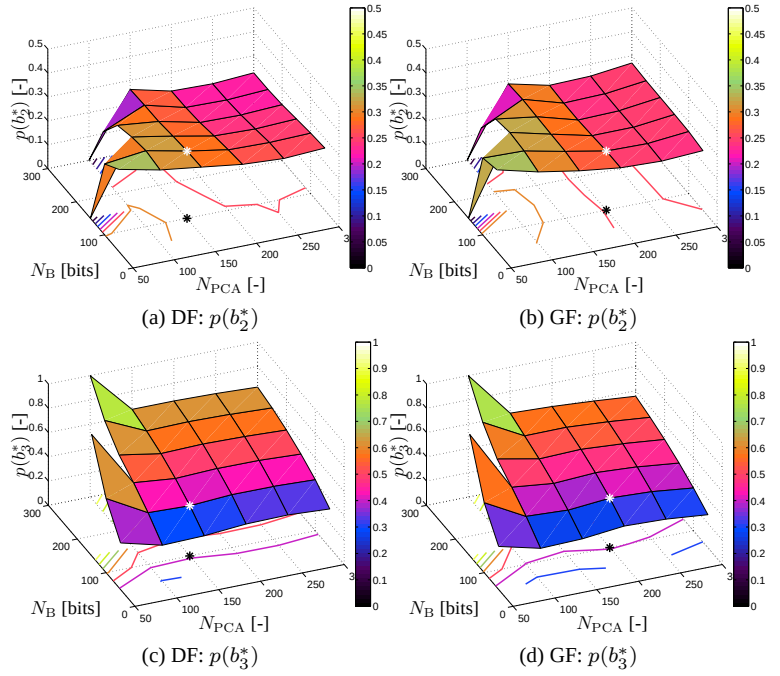
For the optimal setting we have the probabilities $p(b_1^*) = 0.345$, $p(b_2^*) = 0.247$, and $p(b_3^*) = 0.408$ for the DF features, and $p(b_1^*) = 0.304$, $p(b_2^*) = 0.282$, and $p(b_3^*) = 0.414$ for the GF features, respectively. Note that the majority of the bits are extracted from a multiple-bits extraction case, from which we know that information is leaked as shown in Figure 6.5. More precisely, the largest portion of bits are extracted from the b_3^* case, which leaks the most information.

6.2.5 Exploitation of the Leakage

In the previous section we have shown that the information leakage from the auxiliary data AD_1 about the enrolled binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ is significant even at the optimal perfor-

Table 6.1: The $p(b_1^*)$, $p(b_2^*)$, $p(b_3^*)$, δ_2, δ_3 values for the DF and GF features.

Features	EER [%]	β_{tar} [%]	$p(b_1^*)$	$p(b_2^*)$	$p(b_3^*)$	δ_2	δ_3
DF	1.49	3.66	0.345	0.247	0.408	0.1706	0.4106
GF	1.29	2.30	0.304	0.282	0.414	0.3136	0.4727



(a) DF: $p(b_2^*)$ (b) GF: $p(b_2^*)$

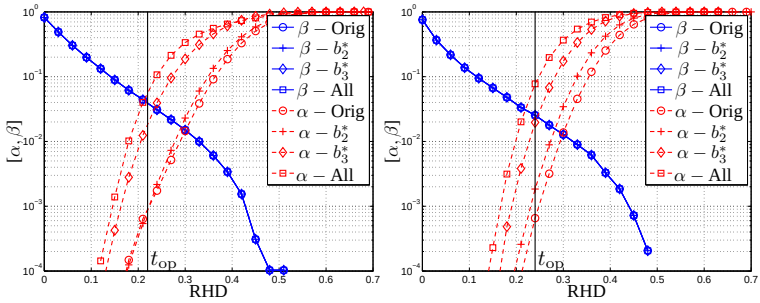
(c) DF: $p(b_3^*)$ (d) GF: $p(b_3^*)$

Figure 6.6: The $p(b_2^*)$, $p(b_3^*)$ for different settings of N_{PCA} and N_{B} for the DF and GF features. The optimal performance setting is indicated with both the black and white star.

mance setting. However, it does not show what the actual practical advantage is for the adversary. In this section we propose a simple method the adversary could use in order to take advantage of the leaked information.

We consider the attack scenario where the adversary has the protected template, which is the collection of public auxiliary data AD_1, AD_2 and PI, of an unknown subject and tries to obtain a false accept by the biometric system. As defined in [138] we focus on the attack level of "*overriding the feature extraction process*". A possible attack method would be a dictionary attack, where a random image sample from a publicly available fingerprint database is selected, its feature vector $\mathbf{f}$ is extracted and send to the next modules as if it is authentic. The probability of an accept is equal to the FAR of the template protection system, because the imposter comparisons in fact do represent a dictionary attack. In our work, the targeted FAR is $\alpha_{\mathrm{tar}} = 0.1\%$, thus on average $\frac{1}{\alpha_{\mathrm{tar}}} = 1000$ attempts are expected in order to obtain a successful accept.

In our proposed attack method we also consider the *DROBA Quantizer* module to be compromised. Hence, the binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ is generated and send to the next module. The leaked information can be exploited in the following way. We change the *DROBA Quantizer* module as such that if multiple bits are extracted (the b_2^* and b_3^* cases indicated by AD_1), we randomly select one of the two outer quantization intervals and return the corre-



(a) DF: α and β curves (b) GF: α and β curves

Figure 6.7: The error-rate curves pmfs for the (a) DF and (b) GF features when using the proposed attack at the imposter comparisons.

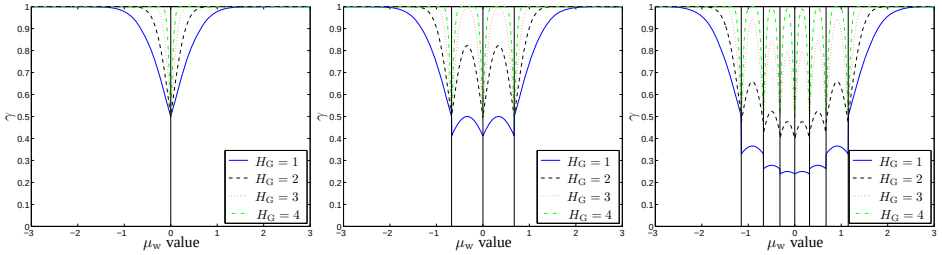
sponding Gray code. Hence, if AD_1 indicates that it is a b_2^* case, then either quantization intervals 1 or 4 are selected with 50% probability and when it is a b_3^* case the quantization intervals 1 or 8 are selected at random.

The attack results are given by the error-rate curves in Figure 6.7(a) and (b) for the DF and GF features, respectively. Note that the attack is only carried out on the imposter comparisons and hence only the FAR curves are influenced. The original FAR is indicated with the "Orig" suffix, which is previously shown in Figure 6.4 and represents the case where the attacker plainly selects a random sample from the database for the verification comparison without using any available knowledge and is the common FAR reported in the literature. For the attacks including the knowledge of the information leakage, we first study the method where only the information leakage from the b_2^* cases are exploited, hereafter we consider the method where only the b_3^* cases are exploited, and as the last method both the b_2^* and b_3^* cases are exploited. These attack methods are indicated with the suffix "b_2^*", "b_3^*", and "All", respectively.

The operating point of a biometric system is determined using the α-Orig curve. The closest operating point t_{op} where the FAR reaches the targeted $\alpha_{\text{tar}} = 0.1\%$ without exceeding it, is portrayed with the solid vertical line. The operating point is at a RHD $= 0.22$ with $\alpha = 8.71 \cdot 10^{-2}\%$ for the DF features and RHD $= 0.23$ with $\alpha = 6.56 \cdot 10^{-2}\%$ for the GF features. The FAR obtained at the operating point for the different attack

Table 6.2: The operating point t_{op} at α_{tar} of the original case and the FAR obtained at the different attack scenario.

	Orig case		FAR at t_{op} at attack scenario		
Features	t_{op} [RHD]	$\approx \alpha_{\text{tar}}$ [%]	b_2^* [%]	b_3^* [%]	All [%]
DF	0.22	$8.71 \cdot 10^{-2}$	$8.23 \cdot 10^{-2}$	1.89	5.78
GF	0.23	$6.56 \cdot 10^{-2}$	$1.84 \cdot 10^{-1}$	1.97	7.75



(a) b_1^* case with 2 intervals (b) b_2^* case with 4 intervals (c) b_3^* case with 8 intervals

Figure 6.8: The detection rate γ for different values of μ_{w} for the (a) b_1^*, (b) b_2^*, (c) b_3^* case with different feature qualities $H_{\mathrm{G}} \in \{1, 2, 3, 4\}$.

methods are given in Table 6.2. The results show that α-b_3^* is larger than α-b_2^*, which confirms the fact that the information leakage of the b_3^* cases is significantly larger than of the b_2^* cases. Furthermore, the advantage of the adversary is further increased by using the information leakage of both cases, because α-All is even larger. Hence, the largest achieved α is 5.78% for the DF features and 7.75% for the GF features. For the DF features the FAR has increased with a gain factor $G_\alpha = 66$, while for the GF features $G_\alpha = 118$. Thus, for both features the adversary gain is around two orders of magnitude. The necessary effort for the adversary to obtain an accept has significantly decreased from on average 1148 attempts to 17 attempts for the DF features and from 1524 to 13 for the GF features. Hence, the gain factor G_α can be seen as the gain of the adversary by exploiting the information leakage.

6.2.6 An Implementation Guideline as Remedy

In the previous section we have shown that if no precaution is taken, an adversary with knowledge of the DROBA implementation could significantly increase its false-acceptance rate with two orders of magnitude by exploiting the information leakage embedded in the auxiliary data AD_1 of the protected template. In this section we will address the cause of the information leakage and propose an implementation guideline for mitigating the leakage.

The Cause

Recall the fact that the DROBA algorithm is allowed to extract multiple bits from all feature components of $\mathbf{f}$, irrespective of its discriminating power or quality. Using the Gaussian model for describing the feature distribution of $\mathbf{f}$ (see Section 6.2.3), we can analyze the detection rate at different subject's mean μ_{w} for the b_1^*, b_2^*, and b_3^* cases and at different qualities of the feature components. As a measurement of the feature quality we use the Gaussian channel capacity or entropy H_{G} as defined in [126]

$$H_{\mathrm{G}} = \tfrac{1}{2} \log_2 \left(1 + \tfrac{\sigma_{\mathrm{b}}^2}{\sigma_{\mathrm{w}}^2}\right), \tag{6.3}$$

which only depends on the ratio $\frac{\sigma_{\mathrm{b}}^2}{\sigma_{\mathrm{w}}^2}$ and where σ_{b}^2 is the variance of the between-class Gaussian density p_{b} describing the variability of the mean μ_{w} across the population and σ_{w}^2 is the variance of the within-class Gaussian density p_{w}.

Assuming the total density p_{t} to have a unit variance and using $\sigma_{\mathrm{t}}^2 = \sigma_{\mathrm{w}}^2 + \sigma_{\mathrm{b}}^2$ we can rewrite H_{G} as

$$\begin{aligned} H_{\mathrm{G}} &= \tfrac{1}{2}\log_2\left(1 + \tfrac{\sigma_{\mathrm{t}}^2 - \sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{w}}^2}\right) \\ &= \tfrac{1}{2}\log_2\left(\tfrac{1}{\sigma_{\mathrm{w}}^2}\right) \\ &= -\log_2(\sigma_{\mathrm{w}}). \end{aligned} \tag{6.4}$$

Hence, feature components with $H_{\mathrm{G}} = 1$ have a within-class standard deviation of $\sigma_{\mathrm{w}} = \frac{1}{2^{H_{\mathrm{G}}}} = \frac{1}{2}$, similarly for the cases $H_{\mathrm{G}} = [2, 3, 4]$ we have $\sigma_{\mathrm{w}} = \left[\frac{1}{4}, \frac{1}{8}, \frac{1}{16}\right]$, respectively.

Using (6.1) the detection rate γ for different values of μ_{w} for different b_x^* cases and feature qualities $H_{\mathrm{G}} \in \{1, 2, 3, 4\}$ are shown in Figure 6.8. Note that the quantization intervals are fixed because of the unit variance assumption of p_{t}. The figures show that for the b_2^* and b_3^* cases the maximum detection rate γ for the inner quantization intervals are much lower than for the outer intervals, because the width of the inner quantization intervals are much smaller in order to be equiprobable with respect to the total density. The detection rate difference between the inner and outer quantization bins depend on the feature quality H_{G} and on the b_x^* case. A larger γ difference is observed for smaller H_{G} values and when more bits are extracted.

As discussed in Section 6.2.3, the DROBA algorithm maximizes the overall detection rate γ_t as given by (6.2). Due to the optimization criteria, the DROBA algorithm tends to allocate multiple bits mostly for the cases where the subject's mean μ_{w} is in the outer quantization intervals due to the larger γ values. This behavior is stronger for the lower quality feature components because γ is significantly larger for the outer quantization intervals as shown in Figure 6.8.

We illustrate the non-uniformity effect introduced by the DROBA algorithm with the following simplified case. Consider the case where there are three feature components of equal quality of $H_{\mathrm{G}} = 2$ from which four bits ($N_{\mathrm{B}} = 4$) have to be extracted and only two bits are allowed to be extracted from each component (b_2^* case). Assume, the first component analyzed has a detection rate of $\gamma_1 = 0.8$. The probability that the next component has a detection rate γ_2 larger than threshold $\gamma_{\mathrm{thr}} = \gamma_1$ is portrayed by the shaded area of the p_{t} density shown in Figure 6.9 which is $\Pr(\gamma_2 > \gamma_{\mathrm{thr}}) \approx 0.5$. Note that the probability of each quantization interval is not equiprobable. For the outer quantization intervals we obtain $p(q_1) = p(q_4) = 0.38$, while for the inner quantization intervals $p(q_2) = p(q_3) = 0.12$. Hence the difference is $\delta_2 = 0.26$. If it turns out that $\gamma_2 > \gamma_{\mathrm{thr}}$, then when analyzing the third component the threshold becomes $\gamma_{\mathrm{thr}} = \gamma_2$. Because of the larger γ_{thr} for the third component, the probability of obtaining a higher γ_2 in one of the quantization intervals becomes more uniform and δ_2 is thus larger. Note that this effect is stronger for lower quality feature components with a smaller H_{G} or when more bits are extracted.

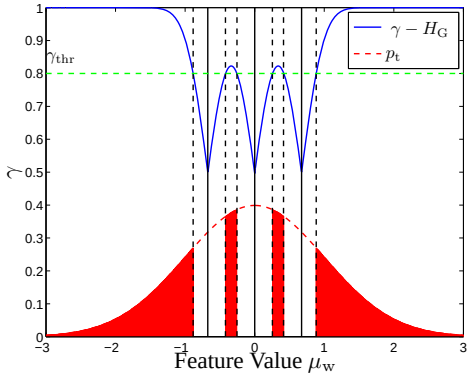


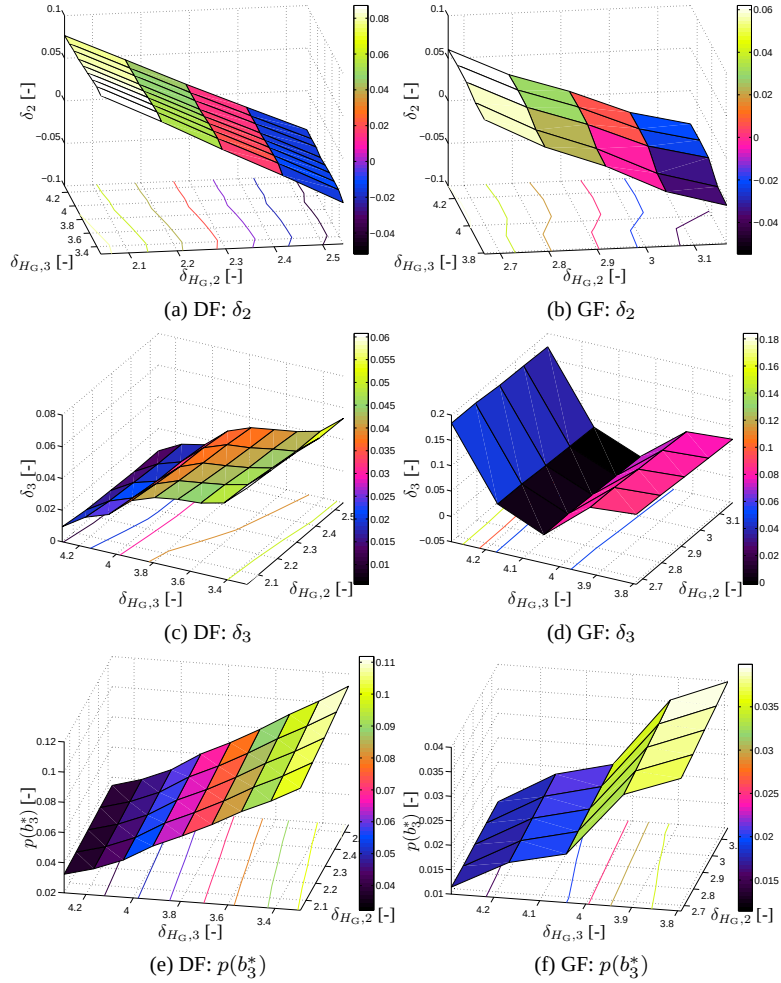
Figure 6.9: The probability of selecting each quantization interval leading to a detection rate γ larger than a threshold γ_{thr}.

The Remedy: Restricting DROBA

As remedy we propose to restrict the DROBA algorithm. The maximum number of bits $b_{\max}$ that the DROBA algorithm is allowed to extract from a component should depend on the overall feature quality of the corresponding component. For each component, we compute the overall feature quality using (6.3) where we take the average of the subject dependent within-class variance across the population. We introduce the thresholds $\delta_{H_G,2}$ and $\delta_{H_G,3}$, where $\delta_{H_G,2}$ defines the minimum overall feature quality requirement of the component for extracting two bits and similarly $\delta_{H_G,3}$ for the case of extracting three bits. We empirically estimate the optimal threshold settings that minimize the information leakage, i.e. induce δ_2 and δ_3 to be close to zero. The δ_2 and δ_3 values for different $\delta_{H_G,2}$ and $\delta_{H_G,3}$ settings are shown in Figure 6.10 for both features. For the δ_2 case we obtain $\delta_2 \approx 0$ by setting $\delta_{H_G,2} = 2.35$ for the DF features and $\delta_{H_G,2} = 2.95$ for the GF features. However, for the δ_3 case it does not reach zero. By increasing $\delta_{H_G,3}$ even further has the consequence that there are only a few b_3^* cases, even less than one case per subject for the GF features as shown by Figure 6.10(f). Eventually we select $\delta_{H_G,3}$ with the biggest drop in δ_3, which is at $\delta_{H_G,3} = 4.05$ for the DF features and $\delta_{H_G,3} = 4.15$ for the GF features.

We implement the proposed remedy to the DROBA algorithm and evaluate the performance and information leakage on the optimal performance setting obtained in Section 6.2.4 of $\{150, 150, 100\}$ and $\{200, 200, 100\}$ for the DF and GF features, respectively. The pmf of Q for the b_2^* and b_3^* cases, and the error-rate curves are shown in Figure 6.11. The pmf of Q for the b_2^* case for both the DF and GF features are very close to uniform, while for the b_3^* case they tend to become more uniform. Because the threshold $\delta_{H_G,3}$ was limited, otherwise no bits would have been extracted from a b_3^* case, the pmf of Q is not uniform.

Comparing the error-rate curves, we observe that the β-Remedy curve has shifted to the right compared to the original curve, β-Orig. However, the α-Remedy curve has also



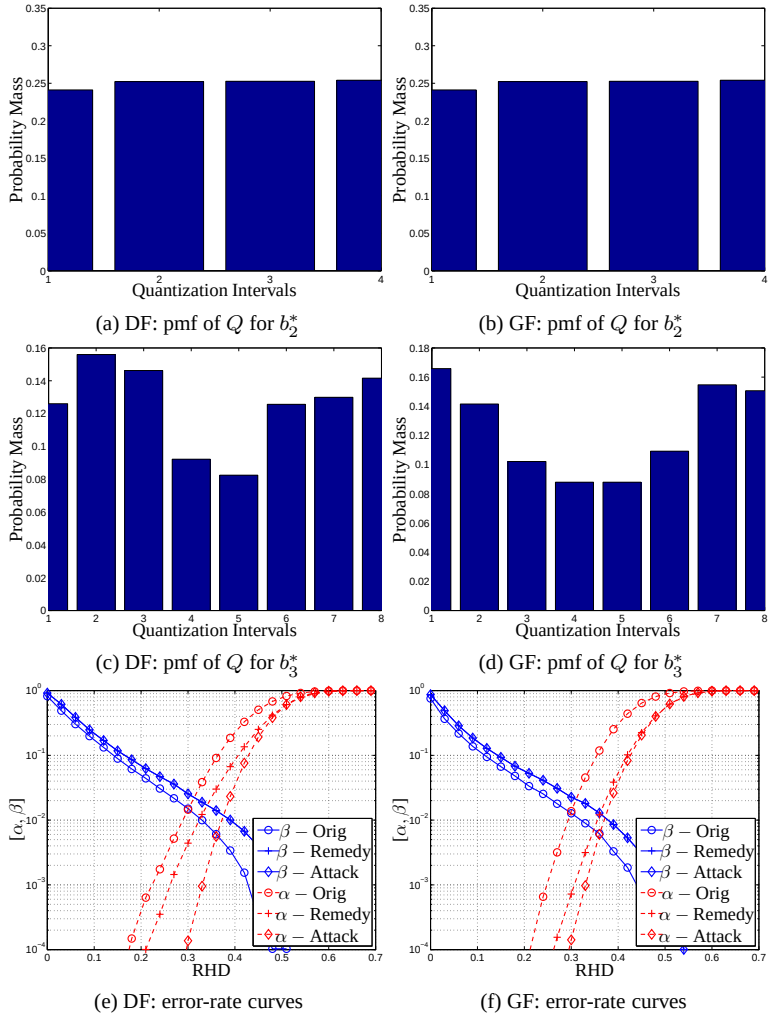
(a) DF: δ_2 (b) GF: δ_2

(c) DF: δ_3 (d) GF: δ_3

(e) DF: $p(b_3^*)$ (f) GF: $p(b_3^*)$

Figure 6.10: The δ_2, δ_3, and $p(b_3^*)$ for different settings of $\delta_{H_G,2}$ and $\delta_{H_G,3}$ for the DF and GF features.

shifted to the right with the consequent that the EER and β_{tar} values are very similar to the original case, namely 1.76% and 3.87% for the DF features, and 1.27% and 2.17% for the GF features. The FRR curve shift can be caused by the fact that the DROBA algorithm is restricted by the proposed remedy. The allocation strategy may then be sub-optimal for the performance. The shift of the FAR curve can be explained in the following way. Note that the variance of p_{t} is larger during the verification phase, because there are less verification samples than enrollment samples, while the quantization intervals are defined equiprobable on the p_{t} during the enrollment phase. Hence, when randomly selecting fingerprint images at the verification comparisons the outer quantization intervals are al-



(a) DF: pmf of Q for b_2^*

(b) GF: pmf of Q for b_2^*

(c) DF: pmf of Q for b_3^*

(d) GF: pmf of Q for b_3^*

(e) DF: error-rate curves

(f) GF: error-rate curves

Figure 6.11: The pmf of Q for the b_2^* and b_3^* cases, and the error-rate curves for the DF and GF features.

ways more probable. When using the original DROBA algorithm, the outer quantization intervals during the enrollment phase are also more probable (the information leakage we have shown). Consequently, there are less bit errors at the imposter comparisons leading to a larger FAR at the same operating point. In other words, it is easier to find a random fingerprint image that leads to an accept. When applying the DROBA remedy, the quantization intervals during the enrollment phase become more equiprobable, consequently eliminating the previously mentioned effect, therefore decreasing the FAR at the same operating point. Furthermore, the α-Attack obtained when using the proposed attack

method did not increase with respect to α-Remedy, it has actually decreased. Hence, the adversary does not gain any advantage by using the proposed attack when the DROBA is correctly implemented. The decrease of the α-Attack can be explained by the fact that the attack method does not consider the correlations between the feature components when randomly guessing one of the outer quantization intervals for the b_2^* and b_3^* cases.

6.2.7 Conclusions

In this work we have shown that great care has to be taken when designing an DROBA quantization scheme in order to guarantee that its auxiliary data does not leak any information about the binary representation of the biometric sample. If no care is taken, the information leakage can be significant and an adversary is able to exploit this information. We have shown that the adversary is able to increase its success rate of obtaining an false accept by two orders of magnitude.

Fortunately, there is a solution to mitigate the information leakage. We proposed a remedy which in fact is a guideline on how to restrict the allocation freedom of the DROBA algorithm. The maximum allowed bits to be allocated to each component has to depend on the overall feature quality across the population of that component. We empirically estimated the minimum overall feature quality boundaries for allocating two or three bits, respectively. Given the biometric database and the feature extraction algorithms, the proposed remedy significantly reduced the information leakage without influencing the performance in terms of the EER or the FRR at the targeted FAR of the biometric system.

6.3 Analysis of the System and Cross-Matching Performance of Bit Extraction Schemes with Template Protection

6.3.1 Abstract

The field of template protection focusses on safeguarding the privacy and security of the stored reference template within a biometric system. Many of the template protection schemes require the extraction of a binary representation from the biometric sample. Because the extracted features from biometric samples are in many cases real-valued, it is necessary to design a bit extraction process for the conversion to a binary vector. Different types of bit extraction schemes have been proposed in the literature, including schemes that use subject-specific statistics to determine the bit allocation strategy in order to extract a more robust binary vector. If the bit extraction scheme is subject-specific, auxiliary data has to be stored as part of the protected template. In this work we (i) demonstrate that the use of subject-specific bit extraction schemes can improve the system performance, (ii) show that the auxiliary data that has to be stored can be used for cross-matching subjects between different applications and compare the system and cross-matching performance for different bit extraction schemes, (iii) show that reconstructing the bit allocation strategy from the verification samples significantly deteriorates the system performance, and (iv) investigate whether the system performance can be improved by fusion of the system and the cross-matching performance.

6.3.2 Introduction

The plethora of passwords that we have to remember for our work or daily life is often leading to frustrations. According to the studies [5] [6], roughly 20% of the participants have to remember 15 or more passwords for their job, while roughly 35% to 57% have between six and 15 passwords to remember. Overall, 82% of the participants are frustrated with managing their passwords. On top of the passwords used at work, there are also many passwords necessary for private use, for example for social networking or commercial websites. It would be much more convenient to replace the use of passwords with biometrics as proposed in [139]. In contrast to passwords it is not possible to "forget" your biometric data.

However, the widespread use of biometrics increases the security or privacy risks such as (i) *identity fraud*, (ii) *limited-renewability*, (iii) *cross-matching*, and (iv) *(sensitive) medical information leakage*. In case of identity fraud an adversary impersonates the genuine subject by some spoofing mechanism. Limited-renewability implies the limited capability to renew a compromised reference template due to the small number of biometric instances, for example we only have ten fingers, two irises or retinas, and a single face. Cross-matching refers to the ability of linking reference templates of the same subject across databases of different applications. It is known that biometric data may reveal the gender, ethnicity, or the presence of certain diseases [20–22].

To mitigate these risks, numerous template protection methods such as the *Fuzzy Com-*

mitment Scheme (FCS) [36], *Helper Data System* (HDS) [33–35, 43, 48], *Fuzzy Extractors* [64, 65], *Fuzzy Vault* [80, 84] and *Cancelable Biometrics* [59] have been proposed in the literature.

Several of the proposed template protection schemes depend on the extraction of a binary vector from the biometric sample. Figure 6.12 depicts a generic template protection scheme that contains a bit extraction scheme. In the enrolment phase the feature extraction algorithm extracts a real-valued *feature vector* $\mathbf{f}^{\mathrm{e}} \in \mathbb{R}^{N_{\mathrm{F}}}$ of N_{F} components from each of the N_{e} enrolment biometric samples. From the N_{e} feature vectors, a single *binary vector* $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \in \{0,1\}^{N_{\mathrm{B}}}$ is created within the *Bit Extraction Generator* module. Note that the size of the binary vector N_{B} can differ from the size of the real-valued feature vector N_{F}. The bit extraction scheme could be subject-specific and therefore has to store some auxiliary data AD_1 as part of the protected template for use in the verification phase. We refer to AD_1 as the *bit extraction auxiliary data*. The final step in the enrolment phase is the protection of the binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ by the *Bit Protection Generator* module whose output is the protected version of the binary vector, namely $[\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}]$.

In the verification phase, the feature extraction algorithm extracts $\mathbf{f}^{\mathrm{v}}$ from each of the N_{v} verification samples. The *Bit Extraction Reproduce* module derives the binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$ with help of the stored auxiliary data AD_1 from the enrolment phase. The binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$ together with the protected version of the enrolment binary vector $[\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}]$ are used within the *Bit Protection Comparator* module in order to derive the decision of either a match or a non-match. Because of this generic approach, all template protection schemes mentioned earlier can be considered as long as they have a binary vector as input. The protected template thus entails the pair AD_1 and $[\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}]$. We assume that the classification performance of the template protection scheme, referred to as the *system performance*, to be equal to the classification performance of the binary vectors. For example, this assumption holds for the fuzzy commitment scheme if the Hamming distance between $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ and $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$ is smaller than the error-correcting capability of the t_{c}-error error-correcting code (ECC).

In the literature, numerous bit extraction schemes have been proposed in order to extract more robust bits. The proposed bit extraction schemes vary from methods that extract a single bit per feature component [33–35, 43] to more complex, multiple-bits extraction methods [42, 44]. The bit extraction scheme from [33–35], also known as *reliable component selection (RCS)*, selects and publishes the most reliable components for each subject in order to extract bits that are more robust. The *quantization index modulation* (QIM) bit extraction scheme from [43] shifts the binarization intervals according to the subjects mean and publishes the required offset. The multi-bits extraction scheme from [42] determines and publishes the number of bits to be extracted based on the detection rate optimized bit allocation (DROBA) algorithm using quantization intervals that are subject-independent, while the scheme from [44] adapts the quantization intervals according to the statistics of the subject similarly to the work of [45].

Due to the subject-specific characteristic of the bit allocation strategy, auxiliary data AD_1 has to be stored as part of the protected template for use in the verification phase. Consequently, there is a risk that the auxiliary data may be used for cross-matching as shown in [43].

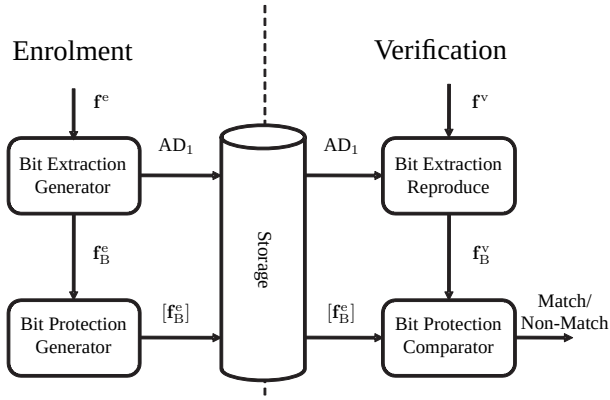


Figure 6.12: Template protection scheme including a *Bit Extraction* module.

Contributions: In this work we have four contributions. Firstly, we demonstrate that the use of subject-specific information can improve the system performance. Secondly, we determine the cross-matching performance of the bit extraction auxiliary data and illustrate the difference between the system and cross-matching performance with respect to the number of enrolment and verification samples. Thirdly, we show that reconstructing the bit allocation strategy from the verification samples, in order to prevent cross-matching, significantly deteriorates the system performance. Fourthly, we investigate if the system performance can be improved by fusion of the system and the cross-matching performance. We focus mainly on the RCS and DROBA bit extraction schemes because they can be considered as a family, namely as auxiliary data they store only the number of bits that have to be extracted. Furthermore, it was shown in [41] that the auxiliary data of bit extraction scheme of [44] leaks a significant amount of information about the extracted binary vector $\mathbf{f}_B^e$. With this known vulnerability, the adversary could easily increase its success rate of impersonation. A similar vulnerability was also discovered for the DROBA scheme in [133], however they also proposed a remedy which we incorporate.

The outline of this paper is as follows. In Section 6.3.3 with discuss the different bit extraction schemes that we analyze. The schemes vary in the amount of subject-specific information that they used. In Section 6.3.4 we discuss the different dissimilarity scores that could be used for cross-matching based on the bit extraction auxiliary data AD_1. We present the experimental results in Section 6.3.5 for both the system and cross-matching performance. Section 6.3.6 investigates the possibility of reconstructing AD_1 using the biometric samples in the verification phase. Furthermore, Section 6.3.7 determines whether the system and cross-matching scores can be fused in order to improved the classification performance. We conclude with the discussion and conclusions in Section 6.3.8.

6.3.3 Bit Extraction Schemes

We consider three types of bit extraction schemes, namely (i) the *simple binarization scheme* (SimpBin), (ii) the *reliable component scheme* (RCS) [33] [34], and the *detection rate optimized bit allocation* (DROBA) scheme [42]. Each bit extraction scheme is described in more details.

Simple Binarization

In case of the simple binarization scheme, referred to as SimpBin, a single bit is extracted from each component of the real-valued feature vector $\mathbf{f} \in \mathbb{R}^{N_\mathrm{F}}$ implying that $N_\mathrm{B} = N_\mathrm{F}$. An illustration of the binarization scheme is depicted in Figure 6.13(a). We model the observed biometric variability and measurement errors of the feature value of a specific subject with the *within-class* density, which we assume to be a Gaussian density $p_\mathrm{w} \sim \mathcal{N}(\mu_\mathrm{w}, \sigma_\mathrm{w}^2)$ with mean μ_w and variance σ_w^2. Note that the mean and variance can be different for each component or subject. We use a single bit binarization scheme based on thresholding, where the quantization threshold δ is equal to the mean of the *adjusted-total* density, which we assume to be Gaussian distributed $p_\mathrm{t}^* \sim N(\mu_\mathrm{t}, \sigma_\mathrm{t}^{2^*})$ with mean μ_t and variance $\sigma_\mathrm{t}^{2^*}$ and adjusted for the averaging of the N_e enrollment samples. The adjusted-total density defines the observed variability of the real-valued feature, averaged from N_e enrolment samples, across the whole population. The variance of the adjusted-total density is $\sigma_\mathrm{t}^{2^*} = \sigma_\mathrm{b}^2 + \frac{\sigma_\mathrm{w}^2}{N_\mathrm{e}}$, where the Gaussian between-class density with variance σ_b^2 models the variability of the mean μ_w across the whole population.

In order for the enrolment bits to be *uniform*, i.e. the probability of a bit value of '0' is equal to '1', the threshold is set equal to the mean of adjusted-total density. Consequently, the threshold creates two binarization intervals labeled '1' and '2', respectively.

Reliable Component Selection

The *reliable component selection* (RCS) scheme selects the N_B most reliable components based on the detection rate of each extracted bit. Similar to the simple binarization scheme, a single bit is extracted from each component of the real-valued representation of the biometric sample $\mathbf{f} \in \mathbb{R}^{N_\mathrm{F}}$, therefore $N_\mathrm{F} > N_\mathrm{B}$. The most reliable components are the ones having a larger detection rate. From [42] the detection rate γ is defined as the probability that the next measurement of the feature component will result in the binarization interval Q_{μ_w} corresponding to the mean μ_w of the within-class density p_w of the subject as portrayed in Figure 6.13(a), more formally

$$\gamma = \int_{Q_{\mu_\mathrm{w}}} p_\mathrm{w}(v)\mathrm{d}v. \tag{6.5}$$

Hence, the detection rate γ is the part of the within-class density within the selected quantization interval in the enrolment phase as portrayed by the shaded area in Figure 6.13(a). For the single bit extraction case, an equivalent method of determining the most reliable component is by means of the z-score. For each component the z-score is estimated as the ratio between the distance of the estimated mean with respect to the binarization threshold

and an estimated standard deviation, namely $z = \frac{|\hat{\mu}-\delta|}{\hat{\sigma}_{\mathrm{w}}}$ [33]. If the standard deviation is estimated from the N_{e} enrolment samples and is therefore subject-specific we refer to the bit extraction scheme as RCS-E. If the standard deviation is estimated from the training set, where we assume each subject to have the same within-class variance, we refer to the bit extraction scheme as RCS-T. The N_{B} components with the largest z-score are selected and stored as auxiliary data $\mathrm{AD}_1 = [b_1, b_2, \ldots, b_{N_{\mathrm{F}}}]$ with $b_i \in \{0, 1\}$. If component j is selected, its corresponding index b_j is one, otherwise zero. Therefore it holds that $\sum_{j=1}^{N_{\mathrm{F}}} b_j = N_{\mathrm{B}}$.

Detection Rate Optimized Bit Allocation

The detection rate optimized bit allocation (DROBA) bit extraction scheme as proposed in [42] has the flexibility to extract multiple bits from a single component. The number of bits extracted from component j is indicated by b_j. The quantization intervals for the $b_j \in \{1, 2, 3\}$ cases are shown in Figure 6.13(a), (b), and (c), respectively. For convenience we refer the $b_j = 1$ case as b_1^*, and b_2^* and b_3^* for the $b_j = 2$ and $b_j = 3$ cases, respectively. The 2^{b_j} quantization intervals are defined as such that the occurrence of each interval is equiprobable with respect to the *adjusted-total* density p_{t}^*. By using the adjusted-total density we can guarantee that the bits extracted in the enrolment phase are uniform. Each quantization interval is assigned a unique b_j bits Gray code [137]. Assuming the within-class density to be Gaussian, $p_{\mathrm{w}} \sim N(\mu_{\mathrm{w}}, \sigma_{\mathrm{w}}^2)$, and the quantization interval corresponding to μ_{w} is selected in the enrolment phase, the detection rate γ_j for component j computed as

$$\gamma_j(b_j) = \int_{Q_{\mu_{\mathrm{w}}}(b_j)} p_{\mathrm{w}}(v)\mathrm{d}v, \tag{6.6}$$

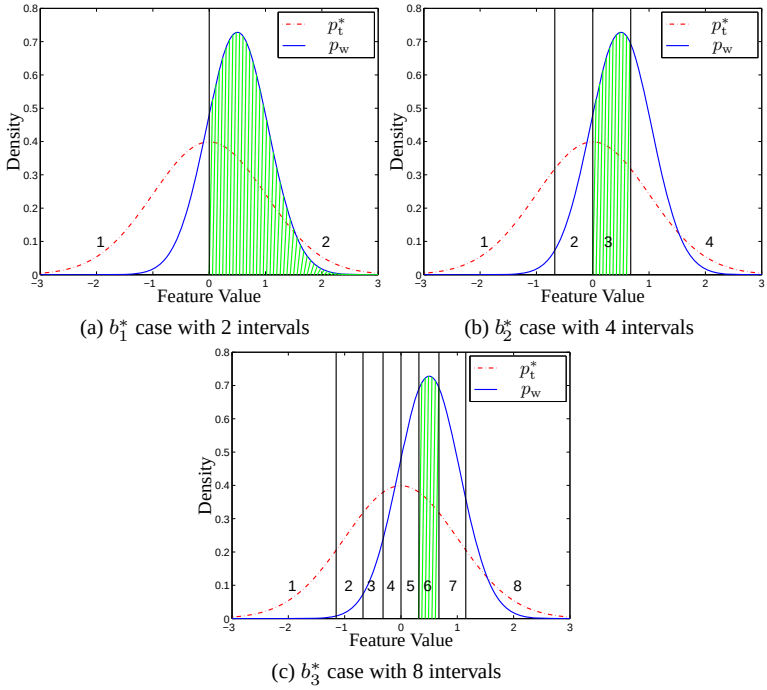
where $Q_{\mu_{\mathrm{w}}}(b_j)$ is the quantization interval corresponding to μ_{w} and for the DROBA scheme also depends on the number of bits b_j to be extracted. For the case where no bits are extracted ($b_j = 0$) the detection rate is defined as $\gamma_j(0) = 1$. Note that the detection rate decreases when b_j increases.

Under the assumption that the N_{F} feature components are independent, the overall detection rate is defined as

$$\gamma_t = \prod_{j=1}^{N_{\mathrm{F}}} \gamma_j(b_j). \tag{6.7}$$

The DROBA algorithm has to create a binary vector of length N_{B}, hence it has to allocate N_{B} bits across all components. We also refer to N_{B} as the bit-budget. With use of the multiple (N_{e}) enrollment samples, the DROBA algorithm analyzes the subject-dependent feature statistics (μ_{w} and σ_{w}^2) of each component and allocates the optimal number of bits b_j to component j with the constrains of maximizing the overall detection rate γ_t from (6.6) and allocating the bit-budget $\sum_{j=1}^{N_{\mathrm{F}}} b_j = N_{\mathrm{B}}$. The optimal allocation strategy is stored as auxiliary data $\mathrm{AD}_1 = [b_1, b_2, \ldots, b_{N_{\mathrm{F}}}]$ for use at the verification phase. The optimization is implemented using the dynamic programming approach presented in [42].

In practice, the mean μ_{w} and variance σ_{w}^2 have to be estimated for each subject and component. If the variance is estimated from the N_{e} enrolment samples we refer to the



(a) b_1^* case with 2 intervals

(b) b_2^* case with 4 intervals

(c) b_3^* case with 8 intervals

Figure 6.13: The adjusted-total density p_t^* with an example of a within-class density p_w where (a) $b_j = 1$ (b_1^*), (b) $b_j = 2$ (b_2^*), (c) $b_j = 3$ (b_3^*) bits are extracted. The corresponding detection rate γ_j is portrayed by the shaded area.

bit extraction scheme as DROBA-E. If the variance is estimated from the training set, where we assume each subject to have the same within-class variance, we refer to the bit extraction scheme as DROBA-T.

The work in [133] shows that if the original DROBA algorithm as presented in [42] is used, its allocating strategy stored in the auxiliary data AD_1 may leak information about the extracted binary string $\mathbf{f}_B^e$. If multiple bits are extracted, it is most likely that the mean of the subject lies in the outer quantization intervals. Because, neighboring Gray codes differ in only one bit, approximately $b_j - 1$ bits are revealed. As shown in [133] this information leakage could be simply exploited by an adversary to increase its rate of success to impersonate another person by two orders of magnitude. In this work we use the updated DROBA allocation algorithm to prevent this information leakage as presented in [133]. The prevention is based on limiting the DROBA allocation algorithm from extracting more bits if its overall feature quality defined by the Gaussian capacity does not exceed a given threshold. It is allowed to extract two bits when the feature quality exceeds the threshold $\delta_{H_G,2}$ and three bits when larger than $\delta_{H_G,3}$. We use the same thresholds that were empirically estimated in [133].

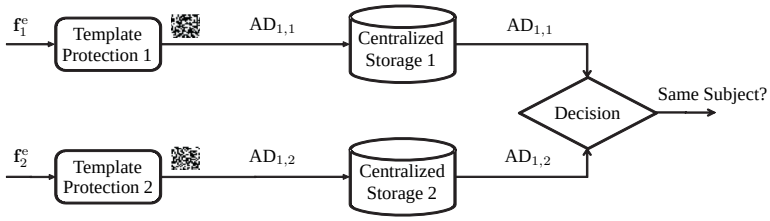


Figure 6.14: The cross-matching attack scenario between two application databases that are accessible by the adversary. For our work we only assume that bit extraction auxiliary data AD_1 is stored in the database accessible by the adversary, while the protected binary vector $[\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}]$ is assumed to be securely stored and not accessible.

6.3.4 Cross-matching Performance

In this section we discuss the cross-matching possibilities concerning the bit extraction auxiliary data AD_1 from the presented bit extraction schemes in Section 6.3.3. Our cross-matching scenario is depicted in Figure 6.14. We consider to have two applications using identical template protection schemes with their own centralized storage, where the auxiliary data for the first and second application is referred to as $\mathrm{AD}_{1,1}$ and $\mathrm{AD}_{1,2}$, respectively. We consider only the auxiliary data AD_1 to be accessible by the adversary from which the adversary tries to find subjects that are enrolled in both applications. Between each database, the auxiliary data $\mathrm{AD}_{1,1}$ and $\mathrm{AD}_{1,2}$ are compared by a *cross-matching classifier* in the *Decision* module. The cross-matching classifier computes a cross-matching dissimilarity score s_{CM} on which to base its decision on whether the two protected templates belong to the same subject (genuine) or not (imposter). The comparison between the protected templates of the same subject is referred to as a genuine comparison and between different subjects as an imposter comparison.

Cross-matching Dissimilarity Scores

The cross-matching classifier used in the *Decision* module of Figure 6.14 extracts a score from both auxiliary data $\mathrm{AD}_{1,1}$ and $\mathrm{AD}_{1.2}$ in order to base its decision on. In this work we investigate the effectiveness of three different dissimilarity scores.

The most obvious score is to check whether the number of bits extracted from each component ($\mathrm{AD}_1[j]$) is equal in both applications. The number of equal occurrences would be a similarity score and in order to obtain a dissimilarity score $s_{\mathrm{CM}}^{a}(\mathrm{AD}_{1,1}, \mathrm{AD}_{1,2})$ we take the difference with its maximum equal to N_{F}, namely

$$s_{\mathrm{CM}}^{a}(\mathrm{AD}_{1,1}, \mathrm{AD}_{1,2}) = N_{\mathrm{F}} - \sum_{j=1}^{N_{\mathrm{F}}} \mathbb{1}_{\{0\}}(\mathrm{AD}_{1,1}[j] - \mathrm{AD}_{1,2}[j]), \tag{6.8}$$

where $\mathbb{1}_{\{A\}}(x)$ is the indicator function that returns a one when $x \in A$, otherwise zero. Note that (6.8) can be used for both the RCS and DROBA bit extraction schemes. How-

ever, because the DROBA bit extraction scheme extracts more than a single bit and therefore $\mathrm{AD}_1[j]$ could be larger than one, we will investigate the effectiveness of two alternative cross-matching scores.

The first alternative score checks whether bits, irrespective the exact number, are extracted from the same component in both applications and count the number of occurrences. This would be a similarity score and in order to make it a dissimilarity score we count the number of occurrences where only one of the applications extracts bits, namely

$$s^b_{\mathrm{CM}}(\mathrm{AD}_{1,1}, \mathrm{AD}_{1,2}) = \sum_{j=1}^{N_\mathrm{F}} \left| \mathbb{1}_{\{0\}}(\mathrm{AD}_{1,1}[j]) - \mathbb{1}_{\{0\}}(\mathrm{AD}_{1,2}[j]) \right|. \tag{6.9}$$

The second alternative score looks at the difference between the number of bits that are extracted from the same components. We use the p-norm defined as

$$s^c_{\mathrm{CM}}(\mathrm{AD}_{1,1}, \mathrm{AD}_{1,2}, p) = \left(\sum_{j=1}^{N_\mathrm{F}} \left| \mathrm{AD}_{1,1}[j] - \mathrm{AD}_{1,2}[j] \right|^p \right)^{\frac{1}{p}}, \tag{6.10}$$

where the norms of interest are the Manhattan norm with $p = 1$ and the Euclidean norm with $p = 2$.

6.3.5 Experiments

In this section we will analyze the performance of the different bit extraction schemes using a biometric database with a feature extraction algorithm. Furthermore, we also determine and analyze the cause of the cross-matching performance of the auxiliary data of each bit extraction scheme.

Biometric Modality and Database

The database we use is the MCYT (Ministerio de Ciencia y Tecnología) containing fingerprint images from a capacitive and optical sensor as described in [128]. It contains 12 images of all 10 fingers from 330 subjects for each sensor. However, we limit our dataset to only the images of the right-index finger from the optical sensor.

Feature Extraction Algorithms

In order to compensate for possible translations between enrolled and verification measurements, a translation-only pre-alignment step is performed during the feature extraction process. Such pre-alignment requires extraction of the core point which is performed according to the algorithm described in [129]. Around the core point we define a 17×17 grid with eight pixels between each grid point. The following feature extraction algorithms extract a feature value on each grid point. Our feature extraction algorithm failed to extract a feature vector from one subject, so we excluded it from the dataset, hence there are effectively only $N_\mathrm{s} = 329$ subjects.

Gabor Filter Response We use the Gabor filter response features extraction algorithm, described in [107], where each grid point is filtered using a set of four 2D Gabor filters at angles of $\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$, respectively. The feature vector is the concatenation of the modulus of the four complex responses at each grid point, resulting into a feature vector dimension of $N_{\mathrm{F}} = 1156$.

Dimension Reduction To decorrelate and reduce the number of feature components we use the principle component analysis (PCA) and the linear discriminant analysis (LDA) techniques, where the LDA transformation is also used to obtain more discriminating feature components. The PCA and LDA transformation matrices are computed using the training set. N_{PCA} is the reduced dimension after applying the PCA transformation and N_{LDA} is the reduced dimension after applying the LDA transformation. We limit N_{LDA} to the number of subjects within the training set from which the transformation matrices are determined. Note that the final real-valued feature vectors are of length $N_{\mathrm{F}} = N_{\mathrm{LDA}}$.

Testing Protocol

The performance testing protocol consists of randomly selecting 219 out of $N_{\mathrm{s}} = 329$ subjects as the *training set* and the remaining 110 subjects as the *evaluation set*, which is referred to as the *training-evaluation-set split*. The bit extraction parameters such as the quantization thresholds, and the PCA and LDA transformation matrices are estimated using the training set.

From the evaluation set we evaluate both the system and cross-matching classification performance.

- For the *system performance* evaluation, N_{e} samples of each subject are randomly selected as the enrolment samples while the remaining samples are considered as the verification samples. The protected template is generated using all the N_{e} enrolment samples and compared with disjoint groups of N_{v} verification samples where the mean of the feature vectors is taken prior to the bit extraction.

- For the *cross-matching performance* evaluation, we randomly select N_{e} samples for the enrolment for the first application and another random N_{e} samples for the second application as such that we have distinct samples for each application. For each application we create the bit extraction auxiliary data AD_1 and compare all AD_1 using the cross-matching classifier. Because of the limitation of 12 samples of each subject we are are limited to $N_{\mathrm{e}} \leq 6$ for the cross-matching performance.

This split of creating the enrolment and verification set or the enrolment set for application one and two is referred to as the enrolment-verification splits. If the verification sample is from the same subject as of the protected template, it is referred to as a genuine comparison, otherwise it is an imposter comparison.

Both the training-evaluation-set and the enrolment-verification split are performed five times. For each spilt we estimate the false match rate (FMR), the false non-match rate (FNMR) and the equal-error rate (EER) where the FMR and the FNMR are equal. Note that the splits are performed randomly, however the seed at the start of the protocol

is always the same, hence all the splits are equal for the performance tests at different settings. Therefore, the splitting process does not contribute to any performance differences.

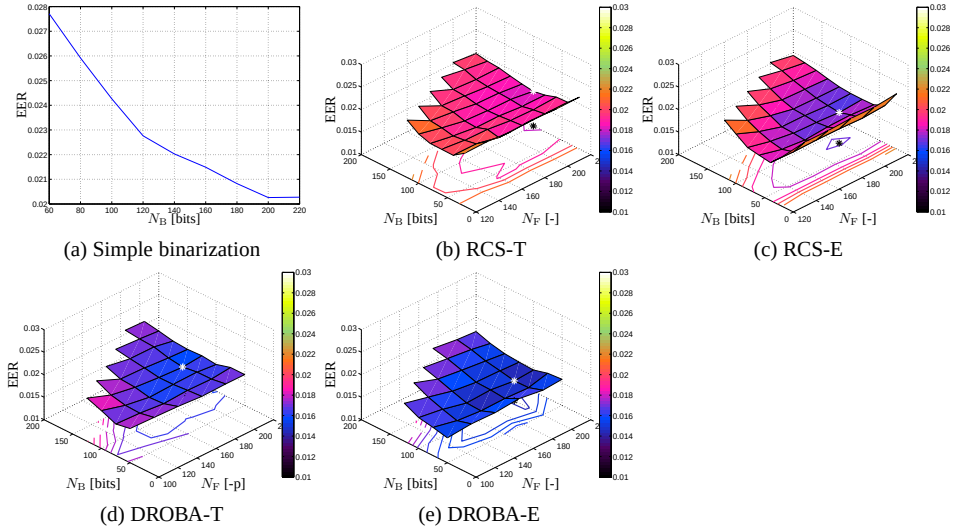
Bit Extraction Schemes Performance

Figure 6.15 depicts the performance in terms of the equal error rate (EER) at different $N_{\mathrm{F}} = N_{\mathrm{LDA}}$ and N_{B} settings with size enrolment and a single verification samples $\{N_{\mathrm{e}} = 6, N_{\mathrm{e}} = 1\}$ for the (a) the simple binarization scheme where it also holds that $N_{\mathrm{B}} = N_{\mathrm{F}}$, (b) the RCS scheme (RCS-T) where the average within-class variance across the population is estimated from the training set, (c) the RCS scheme (RCS-E) where the subject-specific within-class variance is estimated from the enrolment set, and similarly for the (d)(e) DROBA schemes DROBA-E and DROBA-T, respectively. From these results we observe that the optimal setting for the simple binarization scheme is at $N_{\mathrm{B}} = N_{\mathrm{F}} = 200$, while for both RCS and DROBA schemes their optimal setting is around $N_{\mathrm{F}} = 200,\ N_{\mathrm{B}} = 100$, see Table 6.3. The optimal setting for the RCS-E, RCS-T and DROBA-T schemes are slightly different, namely $\{N_{\mathrm{F}} = 220,\ N_{\mathrm{B}} = 80\}$, $\{N_{\mathrm{F}} = 220,\ N_{\mathrm{B}} = 100\}$, and $\{N_{\mathrm{F}} = 200,\ N_{\mathrm{B}} = 125\}$, respectively. Because of the flatness of the EER curves we consider in the remainder of this work the *same setting* of $N_{\mathrm{F}} = 200$ and $N_{\mathrm{B}} = 100$ for both the RCS and DROBA schemes.

With the optimal $\{N_{\mathrm{F}}, N_{\mathrm{B}}\}$ setting determined for $\{N_{\mathrm{e}} = 6, N_{\mathrm{v}} = 1\}$, the influence of the number of enrolment samples N_{e} on the EER for each bit extraction scheme is illustrated in Figure 6.16. Note that for the $N_{\mathrm{e}} = 1$ case it is not possible to estimate the within-class variance of the enrolment samples, hence there is no EER measurement Figure 6.16 for the RCS-E and DROBA-E schemes. Increasing the number of enrolment samples decreases the EER for each bit extraction scheme. Because of averaging the N_{e} real-valued feature vectors prior to the bit extraction process, having more samples further decreases the within-class variance. A smaller within-class variances leads to a binary vector with smaller bit-error probabilities, hence obtaining a better performance. The most significant decrease of the EER is obtained when changing N_{e} from one to two. Further increasing N_{e} still decreases EER and improves the performance, however its impact is of a lesser extent and the EER curve becomes more flat. The work in [33] showed a similar influence of N_{e} on the EER for a biometric database of 3D faces. Additionally, the results in Figure 6.16 also show that when $N_{\mathrm{e}} > 3$ the RCS-E and DROBA-E cases have a smaller EER than their counterpart RCS-T and DROBA-T, respectively. Hence,

Table 6.3: Optimal settings of $N_{\mathrm{F}} = N_{\mathrm{LDA}}$ and N_{B} and the corresponding EER found for the different bit extraction schemes with $N_{\mathrm{e}} = 6$ and $N_{\mathrm{v}} = 1$.

Bit Extraction Scheme	$N_{\mathrm{F}} = N_{\mathrm{LDA}}$	N_{B}	EER [%]
SimpBin	200	200	2.03
RCS-T	220	100	1.84
RCS-E	200	80	1.71
DROBA-T	200	125	1.64
DROBA-E	200	100	1.44



(a) Simple binarization (b) RCS-T (c) RCS-E

(d) DROBA-T (e) DROBA-E

Figure 6.15: The EER at different $N_F = N_{LDA}$ and N_B settings for the (a) simple binarization scheme where $N_B = N_F$, (b)(d) the RCS and DROBA scheme with the average within-class variance across the population estimated from the training set and referred to as RCS-T and DROBA-T, respectively, (c)(e) the RCS and DROBA schemes with the subject-specific within-class variance estimated from the enrolment set referred to as RCS-E and DROBA-E, respectively. Note that we used six enrolment samples and a single verification sample, $\{N_e = 6, N_v = 1\}$.

for $N_e > 3$, the within-class variance estimation $\hat{\sigma}_w^2$ is more accurate when estimating it from the enrolment samples than from the training set.

Cross-Matching Performance

In Section 6.3.4 we discussed three different dissimilarity scores the cross-matching classifier could use to base its decision on. For the bit extraction schemes RCS-T and RCS-E, the three proposed scores are all equal and it is not necessary to compare them. However, for the DROBA-T and DROBA-E bit extraction schemes the cross-matching performance based on the proposed scores for different N_e values are depicted in Figure 6.17(a) and Figure 6.17(b), respectively. For all proposed scores, the cross-matching performance improves when the number of enrolment samples N_e increases. This implies that the bit allocation strategy becomes more stable across applications as more enrolment samples are available. For both DROBA-T and DROBA-E cases, the dissimilarity score s^a_{CM}, which indicates the number of occurrences that the number of extracted bits is not equal, leads to the best cross-matching performance. Furthermore, the cross-matching performance for the DROBA-T case is better than for the DROBA-E case. In contrast to what we have observed in Figure 6.16, having six enrolment samples available is not enough for the cross-

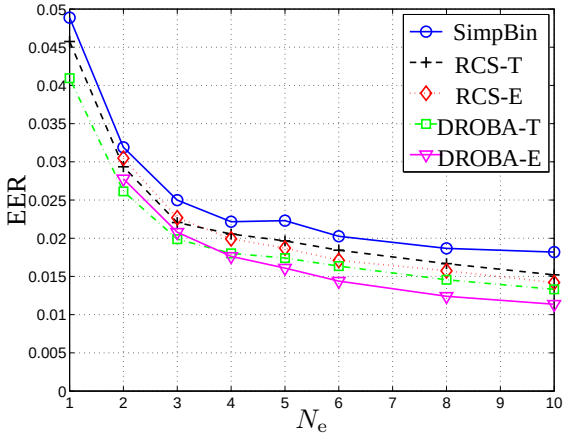


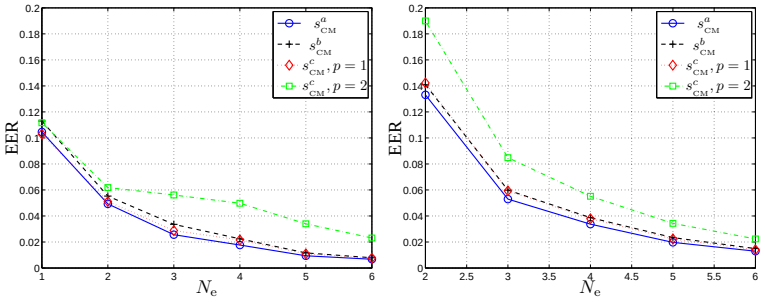
Figure 6.16: System performance expressed by the EER for different number of enrolment samples N_e for the SimpBin, RCS-T, RCS-E, DROBA-T, DROBA-E bit extraction schemes.

matching performance for the DROBA-E case to outperform the DROBA-T case. Hence, the bit allocation strategy at $N_e = 6$ for the DROBA-E case is more subject-specific and accurate enough to improve the system performance with respect to the DROBA-T case, however not robust enough to have a better cross-matching performance.

For the RCS-T, RCS-E, DROBA-T, and DROBA-E bit extraction schemes, Figure 6.18 illustrates the cross-matching performance using the cross-matching score s^{a}_{CM} from (6.8) as function of N_e. From the results we can observe that the cross-matching performance is best for the DROBA-T case followed by the RCS-T case, while both the RCS-E and DROBA-E cases using the variance estimated from the N_e enrolment samples have the worst cross-matching performance. The DROBA-E case still outperforms the RCS-E case. We conjecture that because the DROBA bit extraction scheme in general is more subject-specific than the RCS scheme due to its option to extract multiple bits, therefore its cross-matching performance should be better.

Information Leakage

We have shown in Section 6.3.5 that subject-specific bit extraction auxiliary data AD_1 for both the RCS and DROBA schemes leaks information that can be used for cross-matching. In case of the RCS scheme, the auxiliary data only reveals which components are more reliable and does not reveal anything about the actual extracted bit value. As shown in [133], for the DROBA scheme it does not always hold and depends on the implementation of the DROBA allocation algorithm. Because we use the proposed remedy in [133] of restricting the allocation algorithm, we can show with Figure 6.19 that the information leakage is close to zero because the selected quantization intervals in the enrolment phase are close to uniformly distributed. Only the quantization interval for the



(a) DROBA-T (b) DROBA-E

Figure 6.17: For different number of enrolment samples N_e, the cross-matching performance of the auxiliary data for the (a) DROBA-T and (b) DROBA-E bit extraction schemes using the proposed dissimilarity scores s^a_{CM} of (6.8), s^b_{CM} of (6.9), and s^c_{CM} of (6.10) with $p = 1$ and $p = 2$.

cases where three bits are extracted b^*_3 are not close to uniform, but because only 1% of the extracted bits are derived from a b^*_3 case its impact on the information leakage is minimal. As discussed in [133], increasing the feature quality restriction threshold of extracting three bits $\delta_{H_G,3}$ will make the distribution of the quantization intervals more uniform, however it will further decrease the frequency of the b^*_3 cases. We can also conclude that for this biometric data the maximum number of bits that should be extracted is two.

Thus, the information that is being leaked is not of the enrolled binary vector $\mathbf{f}^e_B$ and therefore it has to leak information of the enrolled real-valued feature vector $\mathbf{f}^e$. For the RCS scheme, the N_B most reliable components are selected based on the z-score $z = \frac{|\hat{\mu}-\delta|}{\hat{\sigma}_w}$, where δ is the binarization threshold, $\hat{\mu}$ is the estimated mean, and $\hat{\sigma}_w$ is the estimated standard deviation. Because the N_B component with the largest z-score are selected, we know that if a component belongs to the N_B most reliable components, its mean will most likely be at least at a certain distance from the binarization threshold δ in order to have a large enough z-score. A visualization of the information leakage is shown in Figure 6.20(a). For a given scenario the shaded area indicates where we know with a high probability that the real-valued feature value of the selected components is most likely distant from the binarization threshold. The distance would depend on the estimated standard deviation and on the smallest z-score selected as a reliable component. At an equal z-score, the distance between the mean and threshold $|\hat{\mu} - \delta|$ is smaller when the standard deviation is smaller and vice versa. If the smallest selected z-score increases, the distance would also increase. The smallest z-score obtained from the N_B selected components would depend on the ratio between the number of selected components and the feature vector length $\frac{N_B}{N_F}$. Decreasing the ratio $\frac{N_B}{N_F}$, i.e. selecting a smaller fraction of the most reliable components, increases the smallest z-score that has been selected. The exact distance would be difficult to estimate due its dependency of many parameters, therefore it is out of the scope of this work. We can only illustrate the type of information leakage but not its quantity. For the components that were not selected we know the op-

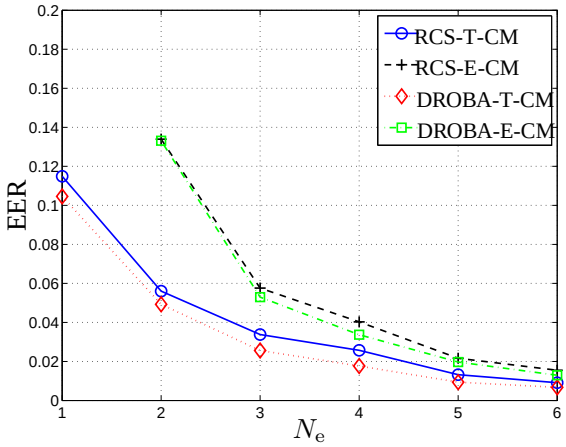


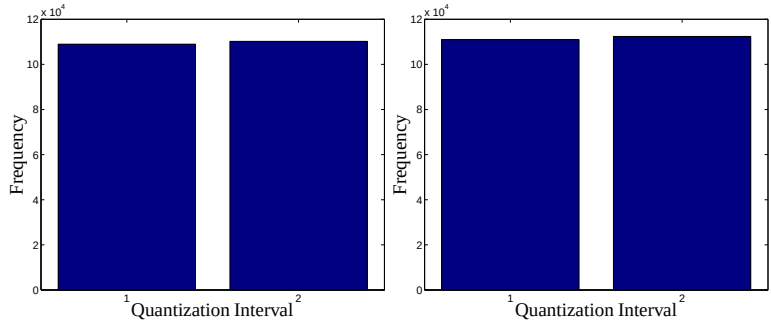
Figure 6.18: The cross-matching performance (EER) as function of the number of enrolment samples N_e for the RCS-T, RCS-E, DROBA-T, and DROBA-E bit extraction schemes using the cross-matching distance score s^a_{CM} from (6.8).

posite information, namely that their real-valued feature value is close to the binarization threshold.

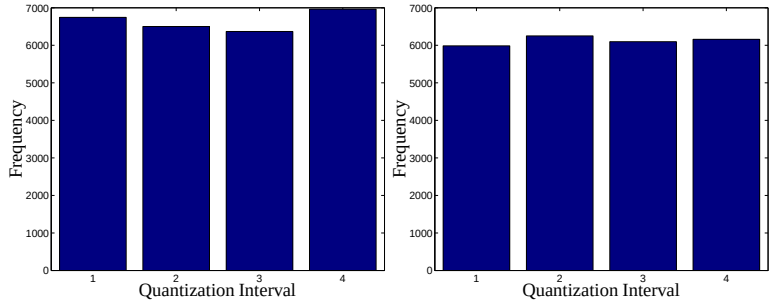
Similarly, this information leakage also holds for the DROBA scheme. Because of the freedom of allocating the number of bits to be extracted, we know that the real-valued feature value of the components are at least at a certain distance from the quantization threshold. When multiple bits are extracted, Figure 6.20(b) and Figure 6.20(c) visualize the information leakage for the cases where two bits b^*_2 and three bits b^*_3 are extracted, respectively. The real-valued feature will be most likely in the shaded areas. It is out of the scope of this work to derive the exact boundaries of the shaded areas.

System vs Cross-Matching Performance

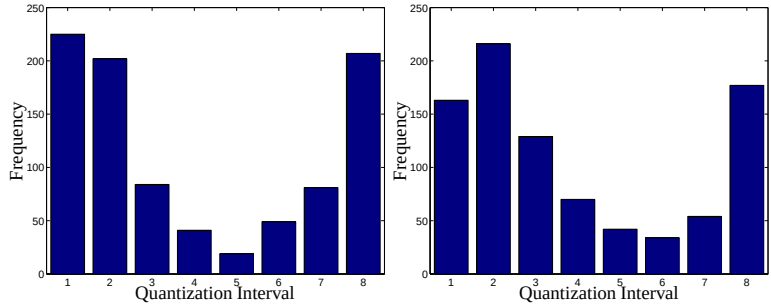
The comparisons between the system and the cross-matching performance (indicated with the suffix 'CM') are shown in Figure 6.21. Figure 6.21(a) and (b) illustrate the achieved EER for different number of enrolment samples N_e for the RCS and DROBA schemes, respectively. For both schemes, the system performance is better than the cross-matching performance at small N_e values, but the difference decreases when N_e increases. If N_e is large enough, the cross-matching performance can even be better than the system performance. For the RCS-T and DROBA-T schemes, where the variance is estimated from the training set, the cross-over point occurs at a smaller N_e value than for the RCS-E and DROBA-E schemes, where the variance is estimated from the enrolment samples. Estimating the variance from the limited enrolment set introduces more uncertainty in the bit allocation strategy and reduces the cross-matching performance. When comparing the ROC curves between the system and cross-matching performance with $N_e = 6$, see Figure 6.21(c) and (d) for the RCS and DROBA schemes respectively, the cross-matching



(a) DROBA-T, b_1^* case (b) DROBA-E, b_1^* case



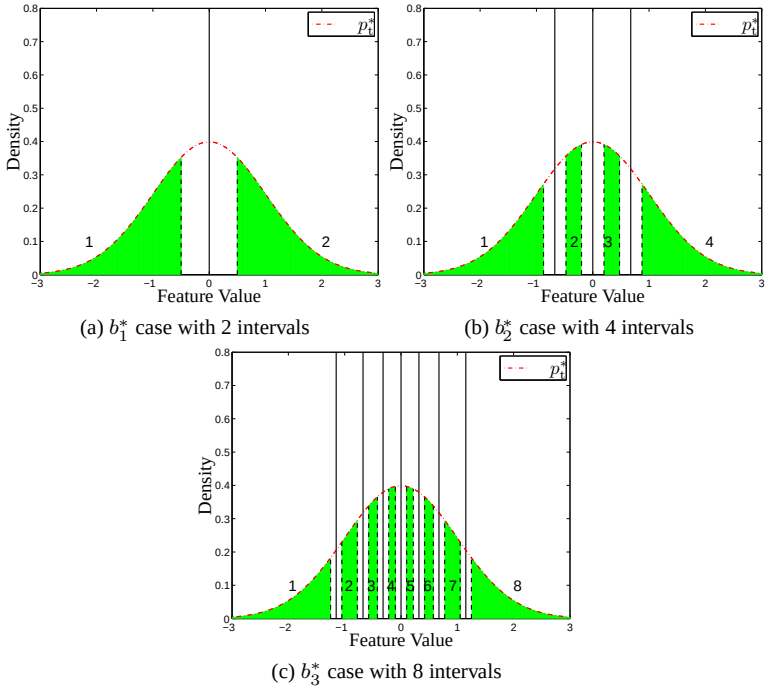
(c) DROBA-T, b_2^* case (d) DROBA-E, b_2^* case



(e) DROBA-T, b_3^* case (f) DROBA-E, b_3^* case

Figure 6.19: The frequency of the selected quantization intervals for the case when 1 bit (b_1^*), two bits (b_2^*) , and three bits (b_3^*) are allocated.

performance is consistently better than the system performance except for the RCS-E scheme for small FMR values. The main reason why the cross-matching performance outperforms the system performance for the case of $N_{\mathrm{e}} = 6$ with $N_{\mathrm{v}} = 1$, is because of the imbalance of the system in terms of the number of enrolment and verification samples. Note that the cross-matching classifier compares the auxiliary data created in the enrolment phase where $N_{\mathrm{e}} = 6$ samples are used in *both* application, while the system performance compares the binary vector from the enrolment phase based on $N_{\mathrm{e}} = 6$ sam-



(a) b_1^* case with 2 intervals

(b) b_2^* case with 4 intervals

(c) b_3^* case with 8 intervals

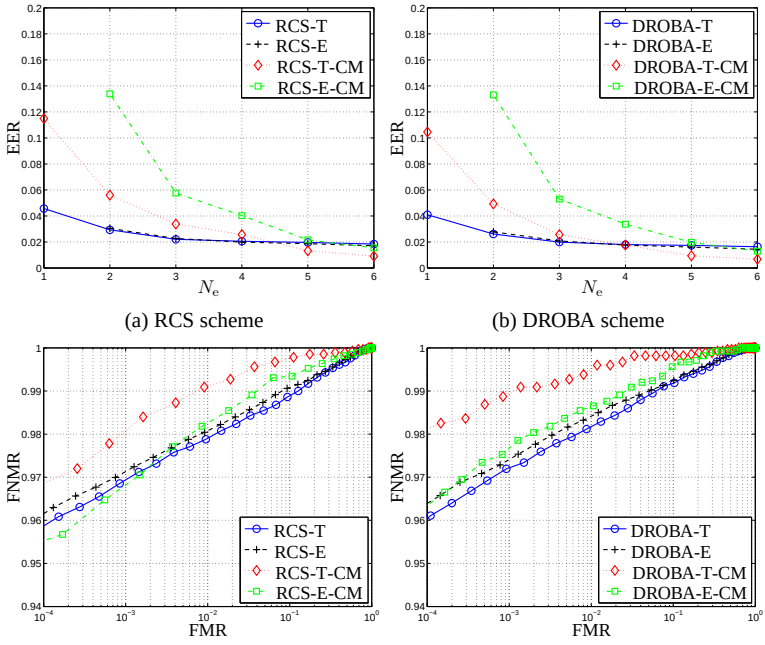
Figure 6.20: The information leakage of the real-valued feature vectors with respect to the adjusted-total density p_{t}^* adjusted for the averaging of N_{e} enrollment samples. For each of the three bit extraction cases, we know with a high probability that the real-valued feature vector will be in the shaded area, and thus not close to the binarization threshold.

ples with the binary vector from the verification phase based on $N_{\mathrm{v}} = 1$ samples. Hence, because of the unbalanced scenario where $N_{\mathrm{e}} > N_{\mathrm{v}}$ the cross-matching performance can be better than the system performance.

6.3.6 Reconstruction of AD_1 in the Verification Phase

In the previous section we have experimentally shown that the cross-matching performance of the auxiliary data from the bit extraction scheme can be in the same order of magnitude of the system performance or even better. In this section, we consider not to store the subject-specific bit allocation strategy AD_1 from the enrolment phase, but instead we investigate whether the same allocation strategy can be reconstructed in the verification phase in order to resolve the issue of cross-matching based on AD_1.

Consider the scheme depicted in Figure 6.22. The auxiliary data AD_1 from the *Bit Extraction Generator* module in the enrolment phase is discarded and thus not stored as part of the protected template. In the verification phase however, the auxiliary data AD_1^* is reconstructed and used to extract the binary vector in the verification phase $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$. The



(a) RCS scheme (b) DROBA scheme

(c) RCS scheme (d) DROBA scheme

Figure 6.21: The comparison between the system performance and the cross-matching performance of the RCS and DROBA schemes where cross-matching performance is indicated by the suffix 'CM'. Sub-figures (a) and (b) depicts the EER as function of the number of enrolment samples N_e with a single verification sample $N_v = 1$ for the RCS and DROBA schemes, respectively. Sub-figures (c) and (d) illustrates the corresponding ROC curves for the case where $N_e = 6$ and $N_v = 1$.

number of bit errors between the binary vectors from the enrolment $\mathbf{f}_B^e$ and verification $\mathbf{f}_B^v$ will now also depend on the accuracy of the reconstructed auxiliary data AD_1^*.

The system performance expressed by the EER as function of the number of enrolment and verification samples $N_e = N_v$ of the proposed bit extraction scheme from Figure 6.22 is illustrated in Figure 6.23 for both the RCS and DROBA schemes and denoted with the suffix 'Rec'. As reference we also include the performance of the original bit extraction scheme from Figure 6.12. From these results we can conclude that the reconstructed auxiliary data in the verification phase is not sufficiently similar to the one from the enrolment phase, because its performance is two orders of magnitude worse when compared with the original scheme. Note that due to the significant performance improvement for the $N_v = N_e > 4$ cases the dataset has to be considered to be too small to accurately measure the EER.

To have an impression on the accuracy of the reconstructed auxiliary data AD_1^*, we show in Figure 6.24 the average ratio between the score s_{CM}^a from (6.8) and N_B across the

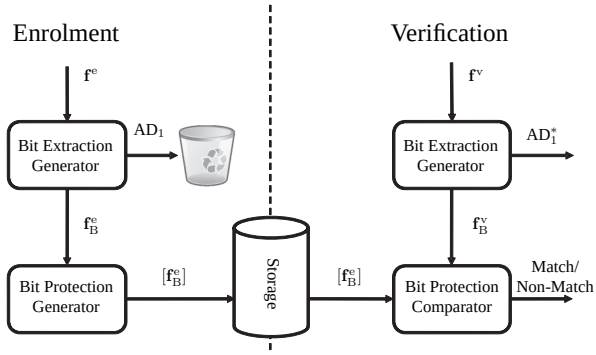


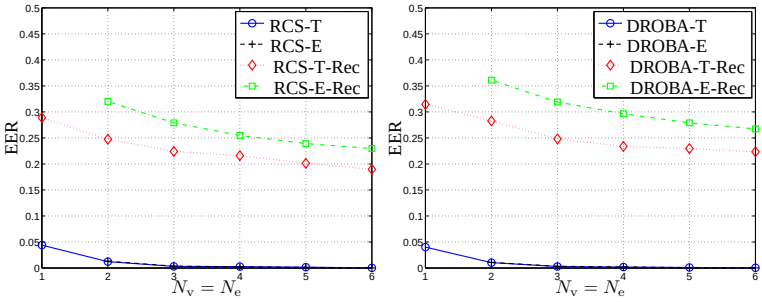
Figure 6.22: Adjusted bit extraction scheme where the auxiliary data AD_1 in the enrolment phase is destroyed and not stored as the protected template. In the verification phase, the auxiliary data is reconstructed as AD_1^* using the verification samples instead.

evaluation set population as function of the number of enrolment and verification samples with $N_{\mathrm{e}} = N_{\mathrm{v}}$. The ratio $\frac{s^a_{\mathrm{CM}}}{N_{\mathrm{B}}}$ indicates the fraction of the bit allocation strategy that is different between enrolment $\mathrm{AD}_{1,1}$ and verification $\mathrm{AD}_{1,1}^*$. Note that for these cases at least 16% of the bit allocation strategy is different. This implies that on average half of those differences will lead to a bit error between the binary vector in the enrolment $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ and verification $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$ phase, assuming that the bits are uniformly distributed across the population. At smaller numbers of enrolment and verification samples, the difference ratio increases towards 30%-40% thus leading to at least 15%-20% relative bit errors between $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ and $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$. Furthermore, estimating the within-class variance from the enrolment samples or the DROBA scheme in general consistently has a larger difference ratio. The large difference ratio clearly explains the significant performance difference between the original system and the proposed system reconstructing the auxiliary data.

Hence, we can conclude that using this method of reconstructing the bit allocation strategy in the verification phase cannot be used to prevent cross-matching, because its system performance is unacceptable. Unless a better reconstruction method is developed, it is *necessary* to store the auxiliary data from the enrolment phase for use in the verification phase in order to maintain a good system performance. Consequently, the auxiliary data could be used for cross-matching when stored in clear (unencrypted).

6.3.7 Increasing the Difference between Cross-matching and System Performance

In the previous sections we have compared the system and the cross-matching performance for different bit extraction schemes. If the number of enrolment samples is large enough, i.e. $N_{\mathrm{e}} > 4$ for these experiments, the cross-matching performance is better than the system performance. Furthermore, we have shown that in order to harvest the system performance gain from the RCS and DROBA schemes, it is necessary to store the



(a) RCS scheme (b) DROBA scheme

Figure 6.23: The performance in terms of the EER of the bit extraction scheme proposed in Figure 6.22 denoted with the suffix 'Rec' and the original bit extraction scheme from Figure 6.12. Note that due to the significant performance improvement of the original scheme for the $N_v = N_e > 4$ cases the dataset may tend to be too small to accurately measure the EER.

auxiliary data for later use in the verification phase. Hence, there is the trade-off between improving the system performance and having a cross-matching performance. In this section we investigate several methods to increase the difference between the system and cross-matching performance. First, we investigate the scenario of having a reversed unbalanced system where the number of verification samples N_v is larger than the number of enrolment samples N_e and a balanced system where $N_e = N_v$. Furthermore, we investigate the possibility of fusion the system and cross-matching performance.

Reversed Unbalanced and Balanced System

As discussed in Section 6.3.5, the main reason why the cross-matching performance can outperform the system performance because of the unbalanced scenario where $N_e > N_v$. By reversing the scenario such that $N_v > N_e$, referred to as the reversed unbalanced system, the system performance is always better than the cross-matching performance as depicted for the $N_e = 1$ with $N_v >= 1$ cases in Figure 6.25(a) and Figure 6.25(b) for the RCS-T and DROBA-T schemes, respectively. Note that because $N_e = 1$ we cannot investigate the RCS-E or DROBA-E schemes. Because the bit extraction auxiliary data is derived from a single enrolment sample only, its bit allocation strategy is not as stable across applications as is observed for the $N_e = 6$ case. Consequently the cross-matching performance deteriorates. Note that the system performance mostly remains unaffected when swapping N_e and N_v as can be observed by comparing Figure 6.21(a) and Figure 6.21(b) with Figure 6.25(a) and Figure 6.25(b), respectively.

Furthermore, by considering a balanced system where $N_v = N_e$, the system performance is also better than the cross-matching performance as illustrated by Figure 6.25(c) and Figure 6.25(d) for the RCS and DROBA schemes, respectively. Note that the cross-matching performance is equal to the ones shown in Figure 6.21(a) and Figure 6.21(b),

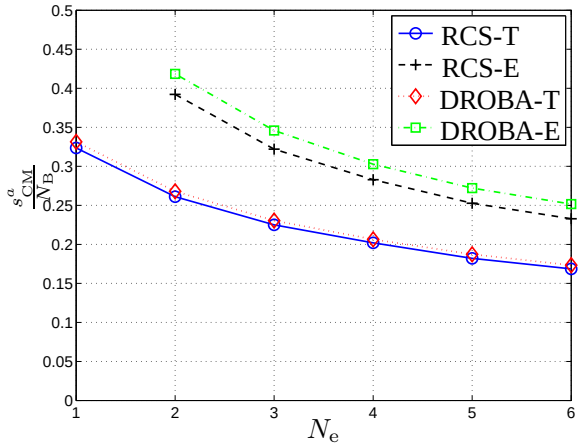


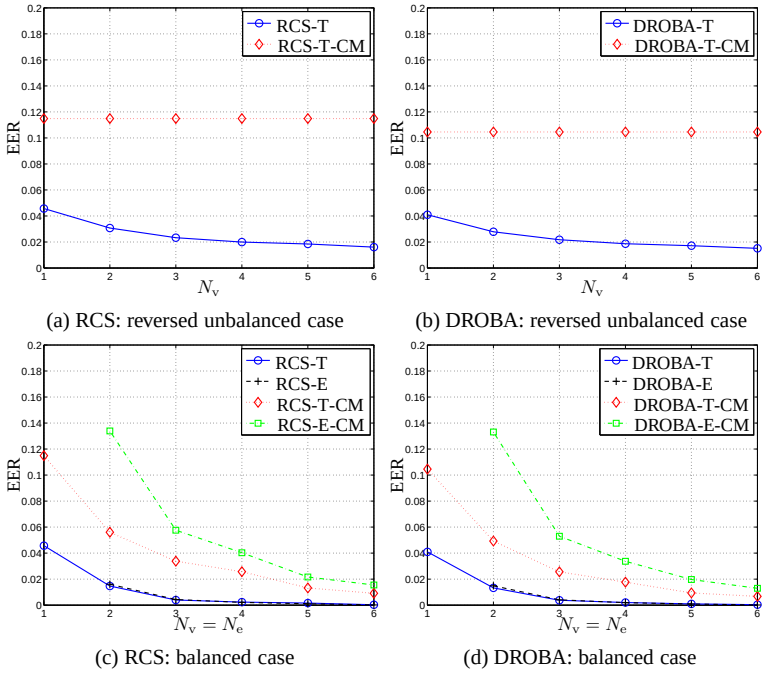
Figure 6.24: The ratio $\frac{s^a_{\mathrm{CM}}}{N_{\mathrm{B}}}$ as function of the number of enrolment samples N_{e} with $N_{\mathrm{v}} = 1$. The ratio $\frac{s^a_{\mathrm{CM}}}{N_{\mathrm{B}}}$ indicates the fraction of the bit allocation strategy that is different between enrolment $\mathrm{AD}_{1,1}$ and verification $\mathrm{AD}^*_{1,1}$.

but the system performance has significantly increased due the increase of the number of verification samples equal to the number of enrolment samples. The system performance increase is so great that the current dataset may tend to be too small for an accurate performance estimation.

Fusion System and Cross-Matching Performance

Another potential method to increase the difference between the system and cross-matching performance is to exploit the cross-matching performance by fusion with the system performance. This method is portrayed in Figure 6.26. In the verification phase, besides extracting the verification binary vector $\mathbf{f}^{\mathrm{v}}_{\mathrm{B}}$ the auxiliary data AD_1 is also reconstructed from the verification samples as AD^*_1. In contrast to Figure 6.12, the *Bit Protection Comparator* module outputs the system dissimilarity score s_{TP} considered to be the Hamming distance between $\mathbf{f}^{\mathrm{e}}_{\mathrm{B}}$ and $\mathbf{f}^{\mathrm{v}}_{\mathrm{B}}$, $d_{\mathrm{H}}(\mathbf{f}^{\mathrm{e}}_{\mathrm{B}}, \mathbf{f}^{\mathrm{v}}_{\mathrm{B}})$. The work of [140] describes the possibility of extracting a score when using the fuzzy commitment scheme as the bit protection scheme, which is generally known to output only a decision. Furthermore, the *Cross-Matching Classifier* module determines the cross-matching dissimilarity score s_{CM} given AD_1 and AD^*_1 with use of (6.8). The *Fusion* module bases its decision on the dissimilarity scores s_{TP} and s_{CM}.

This concept of fusion of the system and cross-matching scores is similar to the one presented in the work [141]. In [141] they proposed to combine the *fragile bit distance* (FBD) with the Hamming distance, both derived from the enrolment and verification iris code. By applying fusion with either the Weighted-Sum-rule or Product-rule, they managed to obtain a significant improvement of the EER from 9.4×10^{-3} to 8.55×10^{-3}. They



(a) RCS: reversed unbalanced case

(b) DROBA: reversed unbalanced case

(c) RCS: balanced case

(d) DROBA: balanced case

Figure 6.25: The system and cross-matching performance expressed by the EER for the (a) RCS and (b) DROBA schemes as function of N_v with $N_e = 1$ referred to as the reversed unbalanced case, and as function of $N_e = N_v$ referred to as the balanced case for the (c) RCS and (d) DROBA schemes.

considered the 25% most inconsistent bits as the fragile bits and the FBD is the fraction of the non-overlapping fragile bits from both the enrolment and verification iris code. Fragile bits within iris codes have been investigated in the works of [142], [143], and [144] in which they all agree that due to the coarse bit extraction schemes some bits are less consistent and can thus be labeled as fragile. They proposed methods to determine the most fragile bits and observed that excluding the fragile bits led to a better performance. These recent developments on iris codes are very similar to the concept of the RCS bit extraction scheme discussed in this work, which has frequently been used in context of template protection in the work of for example [35], [34], and [33].

We use the Weighted-Sum-rule as the score-level fusion method to derive the fused score s_f

$$s_f = \lambda \frac{s_{\text{TP}}}{N_{\text{B}}} + (1-\lambda)\frac{s_{\text{CM}}}{N_{\text{F}}}, \text{ with } \lambda \in [0,1]. \tag{6.11}$$

where we normalized the system score s_{TP} with respect to the length of the binary vector N_{B} and the cross-matching score s_{CM} with respect to the length of the real-valued feature vector N_{F}. The scatter plot of the normalized s_{TP} and s_{CM} scores are depicted in

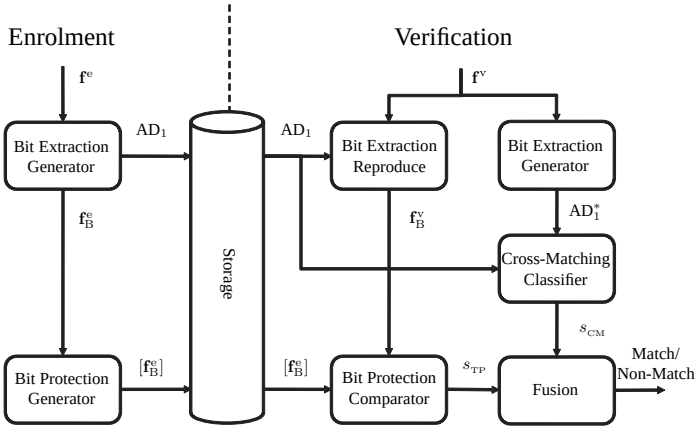


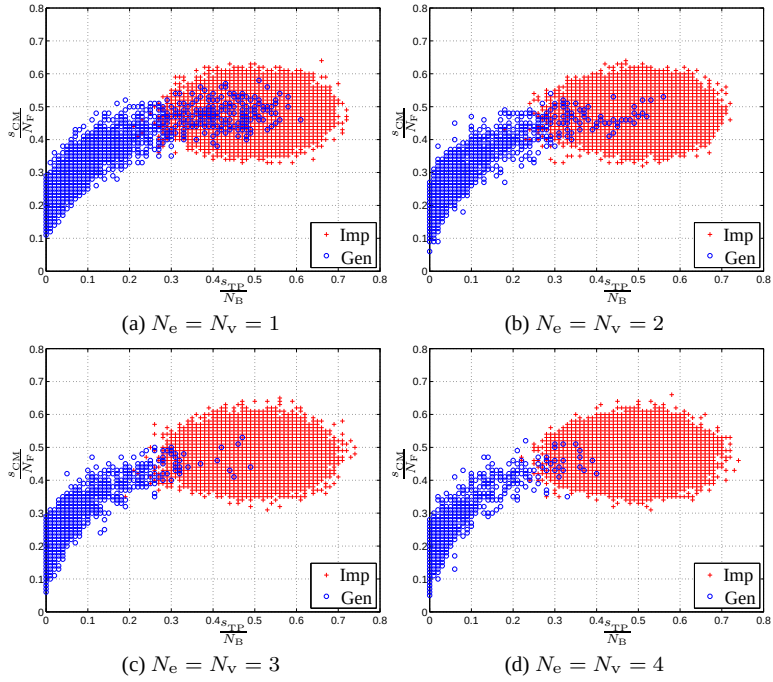
Figure 6.26: Scheme for fusion of the system and cross-matching performance.

Figure 6.27 for the RCS-T scheme for different $N_{\mathrm{e}} = N_{\mathrm{v}}$ settings. Note that we omitted the cases with $N_{\mathrm{e}} = N_{\mathrm{v}} = 5$ and $N_{\mathrm{e}} = N_{\mathrm{v}} = 6$ because as observed in Figure 6.23 the performance improvement is so significant making the evaluation dataset to be too small. The scores for imposter comparisons are indicated by red crosses '+', while we use blue discs 'o' for genuine comparisons. Note that the cluster of imposter scores is stable with respect to the number of enrolment and verification samples, while the genuine scores shift to the lower-left corner when the number of samples is increases. The scatter plots also clearly indicate that the system scores s_{TP} are more discriminating then the cross-matching scores s_{CM}. Figure 6.28 depicts the EER obtained with fusion using the Weighted-Sum-rule from (6.11) as function of the weight λ for both the RCS and DROBA schemes and for different cases of equal number of enrolment and verification samples, i.e. $N_{\mathrm{e}} = N_{\mathrm{v}}$. Note that if $\lambda = 1$ the EER of the system performance is obtained, while the cross-matching performance is obtained when $\lambda = 0$. In all cases there is no significant improvement in performance. When changing λ from one to zero, the EER mainly remains equal or increases.

We can thus conclude that in contrast to the findings in [141], fusion of the system and cross-matching performance using the Weighted-Sum-rule does not lead to a significant performance improvement in our experimental setup.

6.3.8 Discussion and Conclusions

Extracting a binary vector from the biometric sample is an essential element for many template protection schemes. We compared several bit extraction schemes varying in their use of subject-specific information, namely (i) the simple bit extraction scheme SimpBin where no subject-specific information is used, (ii) the reliable component selection (RCS) scheme that uses subject-specific statistics to select the most reliable components, and (iii) the detection rate optimized bit allocation (DROBA) scheme where multiple bits



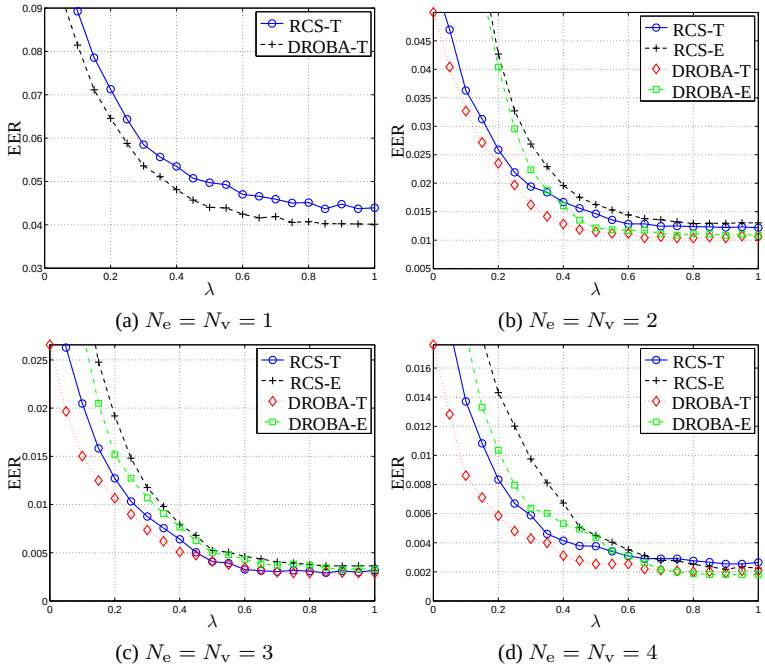
(a) $N_{\mathrm{e}} = N_{\mathrm{v}} = 1$ (b) $N_{\mathrm{e}} = N_{\mathrm{v}} = 2$

(c) $N_{\mathrm{e}} = N_{\mathrm{v}} = 3$ (d) $N_{\mathrm{e}} = N_{\mathrm{v}} = 4$

Figure 6.27: Scatter plot of the normalized system score $\frac{s_{\mathrm{TP}}}{N_{\mathrm{B}}}$ and the cross-matching score $\frac{s_{\mathrm{CM}}}{N_{\mathrm{F}}}$.

can be allocated based on subject-specific statistics. Because of its option to allocate multiple bits, the DROBA scheme can be considered to be more subject-specific than the RCS scheme. The degree of using subject-specific statistics is increased by estimating the variance from the enrolment samples instead of using the average variance from the training set. The auxiliary data containing the bit allocation strategy is stored as part of the protected template. Storing the auxiliary data as part of the protected template can lead to cross-matching between protected templates from the same subject within different applications. We also investigate the cross-matching performance for each bit extraction scheme.

From the experimental results, we have shown that the system performance improves when the bit extraction scheme is more subject-specific, because the DROBA scheme obtained the best system performance. We have also illustrated that the best cross-matching performance is also achieved with the DROBA scheme. Furthermore, we have demonstrated that estimating the variance from the enrolment samples instead from the training set leads to an improvement of the system performance only if the number of enrolment samples is larger than three. In contrast, the cross-matching performance deteriorates when the variance is estimated from the enrolment samples, because the bit allocation strategy is less stable due to spread of the variance estimation. For both the RCS and



(a) $N_e = N_v = 1$ (b) $N_e = N_v = 2$

(c) $N_e = N_v = 3$ (d) $N_e = N_v = 4$

Figure 6.28: EER of the fused scores s_f for the cases with (a) $N_e = N_v = 1$, (b) $N_e = N_v = 2$, (b) $N_e = N_v = 3$, and (b) $N_e = N_v = 4$.

DROBA schemes, increasing the number of enrolment samples greater than three with a single verification sample may cause the cross-matching performance to outperform the system performance. Hence, caution has to be taken when increasing the number of enrolment samples.

We investigated an alternative bit-extraction scheme without any cross-matching, because the bit allocation strategy is not stored but reconstructed from the verification samples. We have shown that its system performance is significantly worse because at least 16% of the bit allocation strategy is different leading to at least 8% of bit errors. These results indicate that the bit allocation strategy has to be stored as the auxiliary data part of the protected template in order to increase the system performance through subject-specific bit allocation.

From the experimental results we have seen that the difference between the system cross-matching performance can be made increased by using (i) a *reversed unbalanced* system where the number of verification samples is larger than the number of enrolment samples or (ii) a *balanced system* where the number of enrolment and verification samples are equal. Furthermore, inspired by the results from the literature, we attempted to increase the difference between the system and cross-matching performance by combining both with score-level fusion using the Weighted-Sum-rule. However, our performance

at fusion was not significantly better than the system performance.

In conclusion, when designing a template protection system with a bit extraction scheme, the benefits of the system performance gain by using a more subject-specific bit extraction scheme has to be weighted against the drawbacks corresponding to the obtained or increased cross-matching performance. If the risks related to cross-matching are too high it may be advised to use a bit extraction method without or with less subject-specific information but with a system performance loss as its consequence. We would advice not to use more than four enrolment samples when there is a single verification sample, because the cross-matching performance may outperform the system performance. Furthermore, the difference between the system and cross-matching performance can be increased by using a reversed unbalanced or balanced system approach. The drawback of the reversed unbalanced or balanced system approach is the time-expensive and inconvenience of acquiring multiple samples in the verification phase.

To further mitigate the privacy risks of cross-matching even when template protection is used, the ISO guidelines [25] recommend (i) the practice of *data separation* where the most privacy sensitive information such as auxiliary data AD_1 is stored on an individual smartcard or token, and (ii) the use of *classical encryption* techniques such as DES and AES to augment the confidentiality of the reference template.

6.4 Chapter Conclusions

Firstly, in Section 6.2, we have shown that great care has to be taken when designing the DROBA bit extraction scheme in order to guarantee that its auxiliary data does not leak information about the binary representation of the biometric sample. When not properly designed, the information leakage can be significant and an adversary is able to exploit this information and increase its success rate of impersonation by two orders of magnitude. As a solution to reduce this information leakage, we propose a redesign of the DROBA algorithm. Specifically, we propose a remedy which in fact is a guideline on how to restrict the allocation freedom of the DROBA algorithm. Experimental results showed that the proposed remedy significantly reduces the information leakage without influencing the system performance.

Secondly, in Section 6.3, we investigated the relationship between the improvement of the HDS classification performance by using bit extraction schemes with more subject-specific information and the corresponding cross-matching performance based on the bit extraction auxiliary data AD_1. If more subject-specific information is used within the bit extraction scheme the HDS performance improves, however the cross-matching performance also improves. Furthermore, we showed that the cross-matching performance can be better than the HDS performance in case of an unbalanced system, which has more enrolment samples than verification samples. On the other hand, we have also shown that the HDS performance saturates with increasing number of enrolment samples if the number of verification samples is fixed. Therefore, we would advice not to use more than four enrolment samples when there is a single verification sample, because the cross-matching performance could become better than the HDS performance. The cross-matching performance can be degraded with respect to the HDS performance by using a balanced system with equal number of verification as enrolment samples, or a reversed unbalanced system where there are more verification than enrolment samples.

In general, due to the information leakage we have identified, it is advisable to protect the auxiliary data AD_1 part by data separation principles (stored on a token) or by using encryption techniques.

Chapter 7

Multi-Sample and Multi-Algorithm Fusion

7.1 Chapter Introduction

In this chapter the fourth and last research question will be addressed, namely

> **Given the HDS template protection scheme: How can one realize fusion with protected templates and to what extend can it improve the classification performance?**

Not being able to apply fusion at score-level with the HDS system has been frequently emphasized as its limitation, see Maiorana et al. (2010) [47]. However, we show in this chapter that with some modification of the verification phase of the HDS it is possible to apply fusion at score-level, however, there are some limitations on the match and non-match regions that can be created in the score space. Furthermore, we compare the fusion classification performance at score-level with the one obtained at feature-level and decision-level fusion. We applied this comparison in context of multi-sample and multi-algorithm fusion, which are published in Kelkboom et al (2009) [125] [1] and Kelkboom et al (2009) [140][2], respectively. Despite the aforementioned limitations of fusion at score-level, fusion at score-level outperforms fusion at feature- and decision-level in case of multi-algorithm fusion, while no significant performance differences were found on the three fusion levels in case of multi-sample fusion.

[1] E. J. C. Kelkboom, X. Zhou, J. Breebaart, R. N. J. Veldhuis, and C. Busch, "Multi-sample fusion with template protection," in Proc. of BIOSIG 2009: Biometrics and Electronic Signatures, Darmstadt, Germany, 2009, pp. 55 - 67.

[2] E. J. C. Kelkboom, X. Zhou, J. Breebaart, R. N. J. Veldhuis, and C. Busch, "Multi-algorithm fusion with template protection," in IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems (BTAS 09), Washington DC, U.S.A., September 2009, pp. 1-8.

7.2 Multi-Sample Fusion with Template Protection

7.2.1 Abstract

The widespread use of biometrics and its increased popularity introduces privacy risks. In order to mitigate these risks, solutions such as the helper-data system, fuzzy vault, fuzzy extractors, and cancelable biometrics were introduced, also known as the field of template protection. Besides these developments, fusion of multiple sources of biometric information have shown to improve the verification performance of the biometric system. Our work consists of analyzing multi-sample fusion in the context of the template protection framework using the helper-data system. We verify the results using the FRGC v2 database and two feature extraction algorithms.

7.2.2 Introduction

More applications are using biometrics ranging from simple home or business applications with a small and limited group of enrolled people (for example access control to buildings or rooms) to large-scale systems used by an entire nation or even the entire world (for example identity cards with biometrics or the electronic passport ePassport). Unfortunately, its widespread use increases the related privacy risks such as identity fraud or activity monitoring by cross-matching between biometric databases of different applications. However, the field of template protection provides the technology that enables the mitigation of these privacy risks by transforming the biometric template with a one-way operation in order to guarantee the irreversibility property and by randomizing the biometric template that guarantees that multiple protected templates from the same biometric sample cannot be linked to each other. In the literature, different types of technologies have been presented, for example the *Helper-Data Systems* (HDS) [33–35], *Fuzzy Vaults* [76, 84], *Fuzzy Extractors* [64, 65], and *Cancelable Biometrics* [59].

Besides the template protection developments, fusion of multiple sources of biometric information has shown to improve the verification performance of the biometric system. As described in [18], the basic principle of fusion is the reconciliation of evidence presented by multiple sources of biometric information in order to enhance the classification performance. Furthermore, different sources of biometric information can be extracted from the same biometric modality by: (i) capturing a sample of multiple instances (left and right index fingerprint or iris) with the same sensor, (ii) using different types of sensors to acquire a different biometric sample from the same instance, (iii) capturing several samples using the same sensor and instance, and (iv) extracting dissimilar feature representations of the same biometric sample using different algorithms. These cases are referred to as the multi-instance, multi-sensor, multi-sample, and multi-algorithm systems, respectively. Furthermore, the fifth type is the multi-modal system, which is the fusion of sources of biometric information from multiple modalities, for example fingerprint, face, iris, voice, palm or retina. To complete the summary from [18], the sixth type is referred to as the hybrid system, which consists of a combination of the aforementioned fusion types. The most common implementations of multi-biometric systems address fusion at the feature-level, score-level or decision-level.

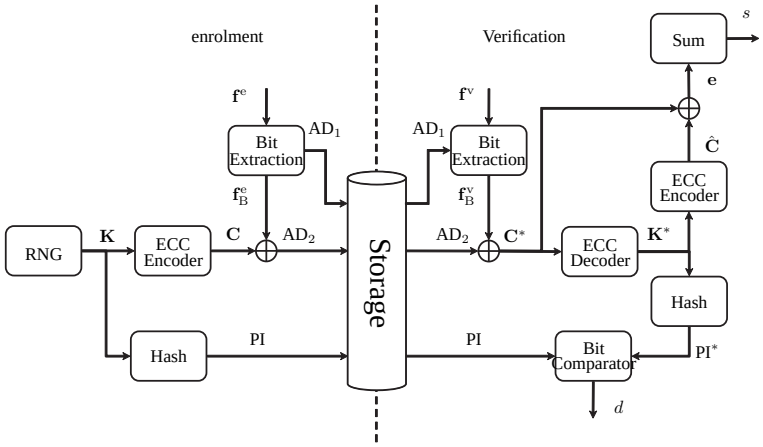


Figure 7.1: The HDS template protection scheme.

In the work of [46], the Fuzzy Vault template protection system is used for applying multi-sample, multi-instance, and multi-modal fusion. In case of multi-sample fusion, they create a single mosaiced template from multiple fingerprint impressions from which they construct the vault. For multi-instance fusion they take the union of the minutiae sets of the left and right index fingers for constructing the vault. For multi-modal fusion, a fingerprint and an iris sample are combined by concatenating the unordered minutiae set with the transformed iriscode extracted from the fingerprint and iris samples, respectively. The vault is constructed using the concatenated unordered set. The verification performance has improved for all three cases as well as the claimed security.

Furthermore, the works of [33, 34, 48] based on the HDS template protection system inherently apply multi-sample fusion at feature-level by averaging the multiple enrolment samples. However, no arguments are provided for applying feature-level fusion instead of either score-level or decision-level.

Our work also consists of applying multi-sample fusion using the HDS, but we analyze the performance improvements of fusion at feature-, score-, and decision-level fusion. We use 3D face range images of the FRGC v2 dataset [99] and verify the performance improvement on two recognition algorithms.

The outline of this paper is as follows. In Section 7.2.3 we briefly discuss the HDS system, while in Section 7.2.4 we discuss the application of multi-sample fusion at feature-, score-, and decision-level using the HDS system together with the experimental setup and results. We finish with the conclusions in Section 7.2.5.

7.2.3 Template Protection Scheme

In the literature, many presented template protection schemes are based on the capability of generating a robust binary vector or key from biometric measurements of the same subject. This also holds for the HDS system we consider and is depicted in Figure 7.1.

For the sake of coherence we use the terminology *auxiliary data* (AD) and *pseudonymous identifier* (PI) proposed in [102], which is in line with standardization activities in ISO [25]. Within the *Bit Extraction* module, a binary vector $\mathbf{f}_{\mathrm{B}} \in \{0,1\}^{N_{\mathrm{B}}}$ is extracted from the real-valued representation of the biometric sample, $\mathbf{f} \in \mathbb{R}^{N_{\mathrm{F}}}$. We use a single bit quantization scheme based on thresholding and the *reliable component selection* (RCS) algorithm. We select the N_{B} most reliable components based on the estimated z-score of each component. With use of the multiple (N_{e}) enrolment samples, the z-score is estimated as the ratio between the distance of the estimated mean with respect to the quantization threshold and the estimated standard deviation, see [33] for a more detailed description of the z-score estimation and the quantization scheme. The index information of the selected reliable components is stored as auxiliary data AD_1.

The binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ could be used as a key for any encryption purposes, however it is not considered as being practical because of the high probability that it is not exactly the same in both the enrolment and verification phase ($\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \neq \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$), due to measurement noise and biometric variability that lead to *bit errors*. The number of bit errors is also referred to as the Hamming distance $d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}})$. To deal with the bit errors, we use error-correcting codes (ECC). The combination of the ECC with a cryptographic hash function forms the scheme also known as the Fuzzy Commitment scheme [36]. In the enrolment phase, a binary secret or message vector $\mathbf{K} \in \{0,1\}^{k_{\mathrm{c}}}$ is randomly generated by the *Random-Number-Generator* (*RNG*) module. A codeword $\mathbf{C}$ of an error-correcting code is obtained by encoding $\mathbf{K}$ in the *ECC-Encoder* module. As the ECC we use the linear block type code "Bose, Ray-Chaudhuri, Hocquenghem" (BCH) [145], which is specified by the codeword length (n_{c}), secret length (k_{c}), and the corresponding number of bits that can be corrected (t_{c}), in short $[n_{\mathrm{c}}, k_{\mathrm{c}}, t_{\mathrm{c}}]$. Some practical BCH settings are provided in Table 7.1, where the bit error rate (BER) is the ratio $t_{\mathrm{c}}/n_{\mathrm{c}}$. The codeword is XOR-ed with $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ in order to obtain auxiliary data AD_2. Hence, $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ should have the same dimension as $\mathbf{C}$, implying $N_{\mathrm{B}} = n_{\mathrm{c}}$. Furthermore, the hash of $\mathbf{K}$ is taken in order to obtain the pseudonymous identifier PI. Under the assumption that the bits of $\mathbf{f}_{\mathrm{B}}$ are independent, from [146] we can use the secret size k_{c} as a measurement of the difficulty of guessing the enrollment binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ from the protected template $\{\mathrm{AD}_1, \mathrm{AD}_2, \mathrm{PI}\}$, hence safeguarding the privacy. The larger the secret size the more difficult it is to either guess $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ or $\mathbf{K}$ from PI.

Table 7.1: Some examples of the BCH code given by the codeword (n_{c}) and secret (k_{c}) length, the corresponding number of correctable bits (t_{c}), and the bit error rate (BER) $t_{\mathrm{c}}/n_{\mathrm{c}}$.

n_{c}	k_{c}	t_{c}	BER $= t_{\mathrm{c}}/n_{\mathrm{c}}$
127	8	31	24.4%
	15	27	21.3%
255	9	63	24.7%
	21	55	21.6%
511	10	127	24.9%
	31	109	21.3%

In the verification phase, a new biometric sample is taken and transformed into its binary representation within the *Bit Extraction* module with help of auxiliary data AD_1. The new word $\mathbf{C}^*$ is computed by XOR-ing $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$ with AD_2, and for a genuine case it is expected that $\mathbf{C}^*$ is close to $\mathbf{C}$. The candidate secret $\mathbf{K}^*$ is obtained by decoding $\mathbf{C}^*$ in the *ECC-Decoder* module. Subsequently, the candidate pseudo identity PI^* is computed by hashing $\mathbf{K}^*$. The decision in the *Bit-Comparator* module is based on whether PI and PI^* are bitwise identical.

The *Bit-Comparator* module outputs a match as its decision d only if PI and PI^* are identical, which occurs when the number of bit errors between the binary vectors $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ and $\mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$ is smaller or equal to the error-correcting capability t_{c} of the ECC. Thus, there is a match when the Hamming distance is smaller than t_{c}, $d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}) = ||\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \oplus \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}||_1 \leq t_{\mathrm{c}}$. Therefore, the fuzzy commitment scheme can be considered as a Hamming distance classifier with threshold t_{c}. Note, that the maximum number of bits that the BCH can correct t_{c}^* is close to 25% of the codeword length. In the remainder of the text, we indicate this limitation as the *ECC-limitation*.

As a distance score s we use the number of bits that had to be corrected by the ECC decoder. The candidate secret $\mathbf{K}^*$ is encoded to its corresponding codeword $\hat{\mathbf{C}}$ and is XOR-ed with $\mathbf{C}^*$ in order to obtain the error pattern $\mathbf{e}$. The error pattern is equal to the bit differences between the enrolment and verification binary feature vectors ($\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \oplus \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$) as follows

$$\begin{aligned} \mathbf{e} &= \hat{\mathbf{C}} \oplus \mathbf{C}^* \\ &= \hat{\mathbf{C}} \oplus (\mathbf{f}_{\mathrm{B}}^{\mathrm{v}} \oplus \mathrm{AD}_2) \\ &= \hat{\mathbf{C}} \oplus (\mathbf{f}_{\mathrm{B}}^{\mathrm{v}} \oplus (\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \oplus \mathbf{C})) \\ &= (\hat{\mathbf{C}} \oplus \mathbf{C}) \oplus (\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \oplus \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}) \\ &= (\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \oplus \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}) \text{ if } \hat{\mathbf{C}} = \mathbf{C}, \end{aligned} \tag{7.1}$$

where $\hat{\mathbf{C}}$ is equal to $\mathbf{C}$ when there is a match, i.e. $\mathbf{K}$ and $\mathbf{K}^*$ are equal. The distance score s is thus the sum of the error pattern, hence equal to $d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}})$ and only a valid score when there is a match, i.e. $d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}) \leq t_{\mathrm{c}}$. If the score is not valid we only know that $d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}) > t_{\mathrm{c}}$.

7.2.4 Experiments

In this section we present the methods for multi-sample fusion at feature-, score-, and decision-level and empirically validate the best performance achieved at each level by means of a biometric database and two feature extraction algorithms.

Experiment Setup

Biometric Databases All the results in this work are obtained using the FRGC v2 dataset [99] containing a total of 4007 3D shape samples from 465 subjects.

However, one of the two 3D shape recognizers we used could not successfully extract a feature vector out of each sample, hence reducing the dataset to 3507 samples from 454 subjects. As the template protection algorithm works best at multiple enrolment samples, the subset of subjects with at least 6 (5 as enrolment samples with at least one for the

verification) samples or more is created. This resulted into a subset of 261 subjects with in total 2970 samples.

Feature Extraction Algorithms The first algorithm is the shape-based 3D face recognizer from [106] and is referred to as Algo1. It has two main steps: 1) the alignment of faces, and 2) the extraction of surface features from 3D facial data. In the alignment step, each face is registered to a generic face model (GFM) and the central facial region is cropped. The GFM is computed by averaging correctly aligned images from a training set. After the alignment step, we can assume that all faces are transformed in such a way that they best fit the GFM, and have the same position in the common coordinate system.

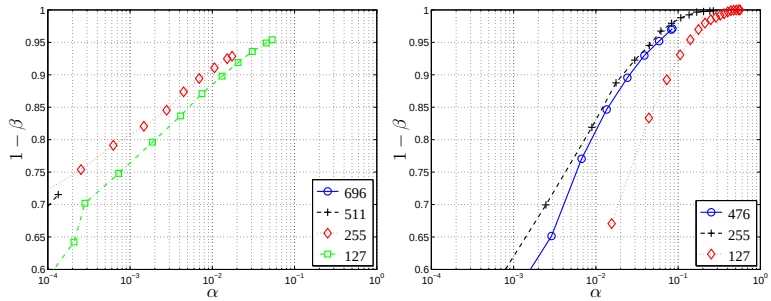
After alignment, the facial surface is divided into 174 local regions. For each region, the maximum and minimum principal curvature direction are computed. Each of the two directions is presented by the azimuthal and the polar angle in the spherical coordinate system. Combining all the regions leads to a feature vector dimension $N_{\mathrm{F}} = 174 \times 2 \times 2 = 696$.

The second algorithm, Algo2, is a histogram-based feature extraction method. After the pre-registration of the face data, a frontal view of the face model is obtained, where the tip of the nose is at the origin in the Cartesian coordinate system. The distribution of depth values of the normalized face model describes the characteristics of an individual facial surface. In order to obtain more detailed information about the local geometry, the 3D model is divided into several sub areas which are orthogonal to the symmetry plane of the face. The features are extracted from the depth value distribution in each sub-area. The feature vector dimension is $N_{\mathrm{F}} = 476$. A full description of this algorithm is provided in [147].

For both feature extraction algorithms, the raw feature vectors they produce are used as input of the template protection system as described in Section 7.2.3. Hence, no further signal processing is performed.

Testing Protocols The performance testing protocol consists of randomly selecting 50% (130) subjects as the training set and the other subjects as the test set, this is referred to as the training-test-set split. The template protection system parameters such as the quantization thresholds, used within the *Bit Extraction* module, are estimated on this training set. Hereafter, the test set is randomly split into an equally sized fusion-training and evaluation set containing around 65 subjects each. All the training needed for fusion is thus performed on the fusion-training set and the reported performance is obtained from the evaluation set. From the evaluation set, 5 samples of each subject are randomly selected as the enrolment samples while the remaining samples are considered as the verification samples. This split is referred to as the enrolment-verification split. The protected template is generated using all the enrolment samples and compared with each verification sample.

The training-test-set split is performed five times, while for each split the enrolment-verification split is performed five times. From each enrollment-verification split we measure the β_{tar} (the false non-match rate (FNMR, β) at the targeted false match rate (FMR, α) of $\alpha_{\mathrm{tar}} = 0.25\%$) and the equal-error rate (EER), which is the error rate achieved at the



(a) Algo1 (b) Algo2

Figure 7.2: ROC curves at feature-level fusion for different n_{c} settings for the (a) Algo1 and (b) Algo2 algorithm.

operating point where both FNMR and FMR are equal. With use of the 25 measurements we estimate the 95% confidence interval (ci) defined as $ci = 1.96\sigma_{EER}/\sqrt{(25)}$ for the EER case while using $\sigma_{\beta_{\mathrm{tar}}}$ for the β_{tar} case, respectively. Note, that the splits are performed randomly, however the seed at the start of the protocol is always the same, hence all the splits are equal for the performance tests at feature-, score-, and decision-level fusion. Hence, the splitting process does not contribute to any performance differences.

Experiment Results

Feature-level Fusion Similar to the works [33, 34, 48], we average the $N_{\mathrm{e}} = 5$ enrolment samples before entering the template protection scheme. By averaging the samples the measurement noise and the biometric variability are suppressed. Hence there will be less bit-errors and the binary representation will be more robust.

The achieved performances for different n_{c} settings are portrayed by the ROC curves in Figure 7.2(a) and (b) for algorithms Algo1 and Algo2, respectively. Furthermore, the EER and β_{tar} details are given in Table 7.2. The table provides the ci for both EER and β_{tar} and their operating point provided as the relative Hamming distance (RHD). The right column of the table provides the effective secret size $|\mathbf{K}_f|$ of the template protection system at the specific fusion level. Because a single protected template is created at feature-level fusion, $|\mathbf{K}_f|$ is equal to k_{c} of the ECC. On the other hand, k_{c} is determined by the t_{c} setting that leads to a α close to the target α_{tar}, but smaller. This is exactly the ECC setting with a BER just larger than the operating point in RHD corresponding to β_{tar}. Entries in the table indicated with quotes cannot be reached in practice because of the ECC-limitation, however we are able to estimate them because of the Hamming distance classifier assumption as discussed in Section 7.2.3. Entries with "x" can neither be reached nor estimated.

Note that the ROC curves are limited because of the ECC-limitation. In order to reach larger α and smaller β values it is required to tolerate and thus correct more bit errors. However, the error correcting capability of an ECC is limited. From the results

we can conclude that both algorithms perform optimally at a codeword size of $n_c = 255$. These settings are used in the score- and decision-level fusion analysis. Compared to the Algo2 algorithm, Algo1 has a better performance but a smaller secret size (see Table 7.2, right column).

Score-Level Fusion A general implementation of the template protection system at score- or decision-level fusion is depicted in Figure 7.3. A protected template is created for each of the N_e enrolment samples. Note that the RCS quantization scheme as discussed in Section 7.2.3 uses multiple enrolment samples in order to estimate the necessary statistics, hence we use all the N_e enrolment samples to determine the N_B most reliable components and is used as such in each N_e template protection systems portrayed in Figure 7.3. Within the *Score- or Decision-level Fusion* module the scores $\{s_1, s_2, \ldots, s_{N_e}\}$ are combined into a single fused score s_f from which the decision d_f is taken based on a score threshold. Note that a score is valid only when there is a match from the corresponding template protection system and occurs when $s_i \leq t_c$. Therefore we set the error-correcting capacity t_c to its maximum (t_c^*) in order to obtain a valid score for the largest range possible. Consequently, the secret size used for each of the N_e protected templates is equal to nine bits and does not depend on the score threshold. Hence, at score-level fusion the score threshold determines the operating point of the ROC curve and not the ECC setting. Combination methods such as the minimum (MIN), the maximum (MAX), and the mean (MEAN) of the scores are used in order to obtain s_f. For the MEAN method we take the mean of the valid scores only, while the MIN and MAX methods are based on all the scores. We take the maximum based on all the scores because if there is a single invalid score it should lead to a non-match. Furthermore, for each method, if all the scores are not valid it will automatically lead to a non-match.

The ROC curves at the optimal setting of $n_c = 255$ are depicted in Figure 7.4 with the details in Table 7.3. As a comparison, we included the ROC curve obtained at feature-level fusion indicated as "FTR". Because it suffices to guess a single $\mathbf{f}_B^e$ from one of the N_e

Table 7.2: The EER and β_{tar}, and their ci and operating point for the individual algorithms Algo1 and Algo2 at different settings of n_c. The last column is the effective secret size $|\mathbf{K}_f|$ which is equal to the secret size k_c of the ECC at the operating point t_c for achieving α_{tar}.

n_c	*EER* [%]	*RHD* [%]	β_{tar} [%]	*RHD* [%]	$\lvert\mathbf{K}_f\rvert$ [bits]
Algo1					
696	"3.76 ± 0.25"	"38.8"	"16.13 ± 1.93"	"33.62"	x
511	"3.69 ± 0.30"	"35.2"	"15.19 ± 1.79"	"28.77"	x
255	"4.02 ± 0.41"	"27.5"	15.84 ± 2.10	19.61	21
127	4.88 ± 0.47	23.6	18.95 ± 2.01	14.96	29
Algo2					
476	5.44 ± 0.35	22.1	37.69 ± 3.14	11.76	x
255	5.06 ± 0.30	10.2	30.25 ± 2.88	1.96	215
127	8.92 ± 0.33	3.9	89.57 ± 1.20	0.00	120

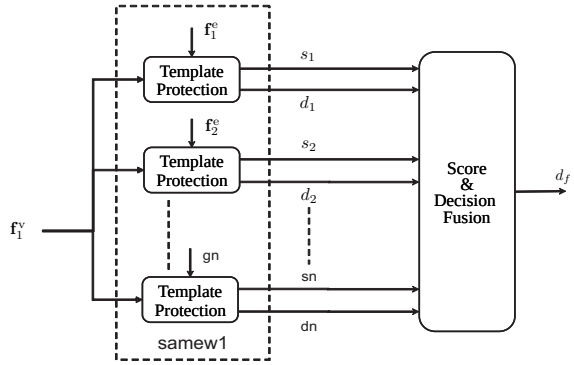
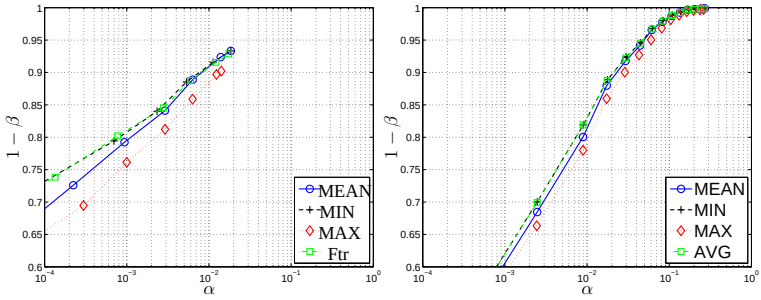


Figure 7.3: The general implementation of multi-sample fusion at score- or decision-level.



(a) Algo1 (b) Algo2

Figure 7.4: ROC curves at score-level fusion compared to the feature-level (FTR) curves for the (a) Algo1 and (b) Algo2 algorithm.

protected templates to breach your privacy, the effective secret size $|\mathbf{K}_f|$ of the template protection system at score-level fusion for each method is also nine bits. Consequently we have omitted them from the table. The results indicate that taking the MIN method leads to the best performance, however the difference is not significant when considering the ci. Furthermore, the MIN method ROC curve is very close to the ROC from feature-level fusion (FTR). Note that for the Algo1 algorithm it is not possible to estimate the EER for all the methods, because the EER is at an operating point greater than t_c^*, hence there are no valid scores.

We also observed that the ROC curves, especially for Algo2, are very similar. At further analysis we discovered that the ROC curves converge to a single one when decreasing n_c. This can be explained as follows. When selecting the most reliable components many enrolment samples from the same subject have an identical binary representation $\mathbf{f}_B$. For example, for the $n_c = 255$ case 75% of the enrolled subjects have no differences between the binary representation $\mathbf{f}_B$ of its N_e enrolled samples for the Algo1 algorithm and 92%

for the Algo2 algorithm, respectively. For the $n_{\mathrm{c}} = 127$ case, the likelihood increases to 99% and 100%, respectively.

Multi-Sample Fusion at Decision Level Similar to the score-level fusion case a protected template is created for each N_{e} samples and compared with the single verification sample. However, the *Score- or Decision-level Fusion* module combines the decision $\{d_1, d_2, \ldots, d_{N_{\mathrm{e}}}\}$ into a single fused decision d_f. Methods such as the OR-rule, AND-rule, and majority voting (MV) are used in order to obtain d_f. For the AND-rule method, all the decisions have to be a match in order for the final one to be a match too, while for the OR-rule case only a single match leads to a final match. For the MV method more than half of the decisions should be a match in order to have a final match.

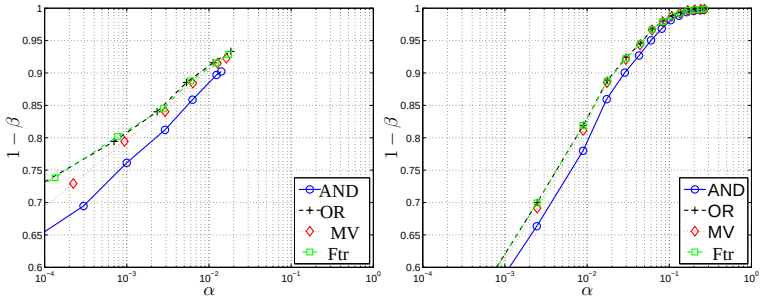
Again, it suffices to break a single protected template for the adversary to know $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$, hence the effective secret size $|\mathbf{K}_f|$ is equal to the secret k_{c} corresponding to the ECC setting.

The experimental results are portrayed in Figure 7.5 with the performance details in Table 7.4. As a comparison, we included the ROC curve obtained at feature-level fusion indicated as "FTR". From these results we can conclude that the OR-rule fusion method consistently leads to a better performance, followed by the MV method, and the worst performance is with the AND-rue method. However, the difference is not significant. Compared to feature-level fusion results, the OR-rule methods leads to a similar ROC curve. The ROC curves, especially for the Algo2 algorithm, are very similar due to the same reason as as discussed in the previous section where it was noticed that the reliable binary representation $\mathbf{f}_{\mathrm{B}}$ is very similar for every N_{e} samples.

Summary and Discussions We have compared performances of multi-sample fusion at feature-, score-, and decision-level. At the optimal setting of $n_{\mathrm{c}} = 255$ we do not observe a significant performance differences between either feature-, score-, and decision-level fusion method. The effective secret size $|\mathbf{K}_f|$ is the same at feature- and decision-level fusion, and at its smallest at score-level fusion. Taking into account that at score and decision level fusion a protected template has to be made and stored for each N_{e} enrolment

Table 7.3: The EER and β_{tar}, and their ci and operating point for the score-level fusion experiments with $n_{\mathrm{c}} = 255$.

Method	*EER* [%]	*RHD* [%]	β_{tar} [%]	*RHD* [%]
Algo1, $n_{\mathrm{c}} = 255$				
MEAN	x	x	16.45 ± 2.08	20.00
MIN	x	x	15.74 ± 2.09	19.61
MAX	x	x	19.48 ± 2.08	20.39
Algo2, $n_{\mathrm{c}} = 255$				
MEAN	4.96 ± 0.28	10.6	31.46 ± 3.23	2.35
MIN	4.87 ± 0.30	10.2	29.90 ± 3.29	2.35
MAX	5.49 ± 0.29	11.4	33.49 ± 3.08	2.35



(a) Algo1 (b) Algo2

Figure 7.5: ROC curves at decision-level fusion compared to the feature-level (FTR) curves for the (a) Algo1 and (b) Algo2 algorithm.

sample but only a single one at feature level, we can conclude that the best multi-sample fusion method is at feature level. For security and privacy reasons it is also not desired to store multiple protected templates, which could facilitate the attacker with hacking the protected template and either obtain the secret or the biometric data itself. Furthermore, a single protected template has a smaller storage capacity requirement.

When carefully analyzing the score- and decision-level fusion results, we can also conclude that the MIN-score and OR-decision methods have precisely the same performance, similarly for the MAX-score and AND-decision methods. The explanation for the MAX-score and AND-decision case is that if the maximum score is a match it would imply that all the other $N_{\mathrm{e}} - 1$ scores are also a match, which is also the requirement for the AND-decision fusion method. The MIN-score and OR-decision performance similarity can be explained by the fact that both methods require at least a single individual comparison to be a match in order for the final decision to be a match.

Table 7.4: The EER and β_{tar}, and their ci and operating point, and the effective secret size $|\mathbf{K}_f|$ for the decision-level fusion experiments with $n_{\mathrm{c}} = 255$.

Method	*EER* [%]	*RHD* [%]	β_{tar} [%]	*RHD* [%]	$\lvert\mathbf{K}_f\rvert$ [bits]
Algo1, $n_{\mathrm{c}} = 255$					
AND	"4.76 ± 0.40"	"29.0"	19.48 ± 2.08	20.39	21
OR	"3.95 ± 0.39"	"27.1"	15.74 ± 2.09	19.61	21
MV	"4.11 ± 0.44"	"27.8"	16.62 ± 2.05	20.00	21
Algo2, $n_{\mathrm{c}} = 255$					
AND	5.49 ± 0.29	11.4	33.49 ± 3.08	2.35	207
OR	4.87 ± 0.30	10.2	29.90 ± 3.29	2.35	207
MV	4.89 ± 0.28	10.2	30.78 ± 3.27	2.35	207

7.2.5 Conclusions

With this work we have shown that it is possible to apply multi-sample fusion with the HDS system at feature-, score-, and decision-level. Because the HDS system inherently has only a decision as the output, we adapted the system accordingly in order to have a score as output for the score-level fusion. As a distance score we took the number of bits the ECC had to correct. Furthermore, applying fusion with template protection at feature- or decision-level is straightforward and conventional. However, fusion at score-level is different due to the use of an ECC, which has a limited error-correcting capability. Consequently, for each template protection system there is only a valid score when there is a match.

Given the biometric database and feature extraction algorithms, our experimental results showed that at the optimal setting of $n_{\mathrm{c}} = 255$ there are no significant differences between the best performance (ROC curves) obtained at feature-, score-, and decision-level. Because at feature-level fusion only a single protected template is created, which is better in terms of privacy and security protection and storage, we can conclude that the optimal multi-sample fusion is at feature-level.

Acknowledgment

The authors would like to acknowledge the support of the partners within the 3DFACE project, a European Integrated Project funded under the European Commission IST FP6 program.

7.3 Multi-Algorithm Fusion with Template Protection

7.3.1 Abstract

The popularity of biometrics and its widespread use introduces privacy risks. To mitigate these risks, solutions such as the helper-data system, fuzzy vault, fuzzy extractors, and cancelable biometrics were introduced, also known as the field of template protection. In parallel to these developments, fusion of multiple sources of biometric information have shown to improve the verification performance of the biometric system. In this work we analyze fusion of the protected template from two 3D recognition algorithms (multi-algorithm fusion) at feature-, score-, and decision-level. We show that fusion can be applied at the known fusion-levels with the template protection technique known as the Helper-Data System. We also illustrate the required changes of the Helper-Data System and its corresponding limitations. Furthermore, our experimental results, based on 3D face range images of the FRGC v2 dataset, show that indeed fusion improves the verification performance.

7.3.2 Introduction

There is a growing popularity of using biometrics in applications ranging from simple home or business applications with a small and limited group of enrolled people (for example access control to buildings or rooms) to large-scale systems used by an entire nation or even the entire world (for example identity cards with biometrics or the electronic passport ePassport). However, its widespread use increases the privacy risks such as identity fraud or activity monitoring by cross-matching between biometric databases of different applications. The field of template protection provides the technology that mitigates these privacy risks by transforming the biometric template with a one-way function in order to guarantee the irreversibility property and by randomizing the biometric template in order to guarantee that multiple protected templates from the same biometric sample cannot be linked with each other. In the literature, multiple solutions have been presented to solve these problems. Some examples are the *Fuzzy Commitment Scheme* [36], *Helper-Data Systems* (HDS) [33–35], *Fuzzy Vaults* [76, 84], *Fuzzy Extractors* [64, 65], and *Cancelable Biometrics* [58].

In parallel to these developments, fusion of multiple sources of biometric information has shown to improve the recognition performance of the biometric system. As stated in [18], the basic principle of fusion is the reconciliation of evidence presented by multiple sources of biometric information in order to enhance the classification performance. As described in [18], multiple sources of biometric information can be extracted from the same biometric modality by (see Figure 7.6 for the case of fingerprints): (i) capturing a sample of multiple instances (left and right index fingerprint or iris) with the same sensor, (ii) using different sensors to acquire a different type of biometric samples from the same instance, (iii) capturing multiple samples using the same sensor and instance, and (iv) extracting multiple feature representations of the same biometric sample using different algorithms. These cases are referred to as the multi-instance, multi-sensor, multi-sample, and multi-algorithm systems, respectively. Further more, the fifth type is the multi-modal

system, which is the fusion of sources of biometric information from multiple modalities, for example fingerprint, face, iris, voice, palm or retina. To complete the summary from [18], the sixth type is referred to as the hybrid system, which consists of a combination of the aforementioned fusion types. Each multi-biometric fusion type can be implemented at feature-level, score-level, or decision-level of the biometric system.

In [46], multi-sample, multi-instance, and multi-modal fusion has been applied using the Fuzzy Vault as the template protection system. For multi-sample fusion a single mosaiced template is obtained from multiple fingerprint impressions from which the vault is constructed. For multi-instance fusion the union of the minutiae sets of the left and right index fingers is used to construct the vault. For multi-modal fusion, a fingerprint and an iris sample are combined by concatenating the unordered minutiae set with the transformed iriscode extracted from the fingerprint and iris samples, respectively. The concatenated unordered set is used to construct the vault. The recognition performance improved for all three cases as well as the claimed security.

Our Contribution: Our work consists of applying multi-algorithm fusion with the Helper-Data System. We show that fusion can be applied at feature-, score-, and decision-level and illustrate the required changes of the Helper-Data System and its corresponding limitations. We experimentally determine the performance of different fusion methods at each level. The experiments are based on 3D face range images of the FRGC v2 dataset [99], where we use two recognition algorithms from different vendors.

The outline of this paper is as follows. In Section 7.3.3 we briefly discuss the HDS system, while in Section 7.3.4 we discuss the application of multi-algorithm fusion at feature-, score-, and decision-level using the HDS system. The experimental setup and results are provided in Section 7.3.5. We finish with the conclusions in Section 7.3.6.

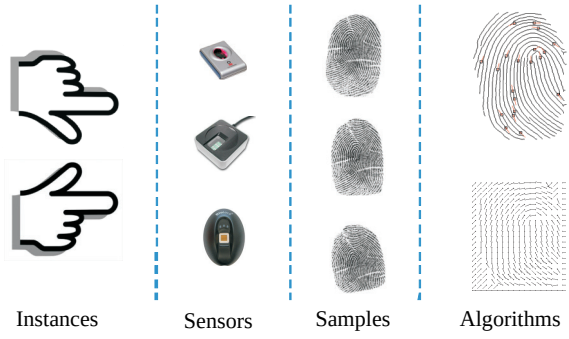


Figure 7.6: Multiple sources of biometric information using fingerprints as the single modality.

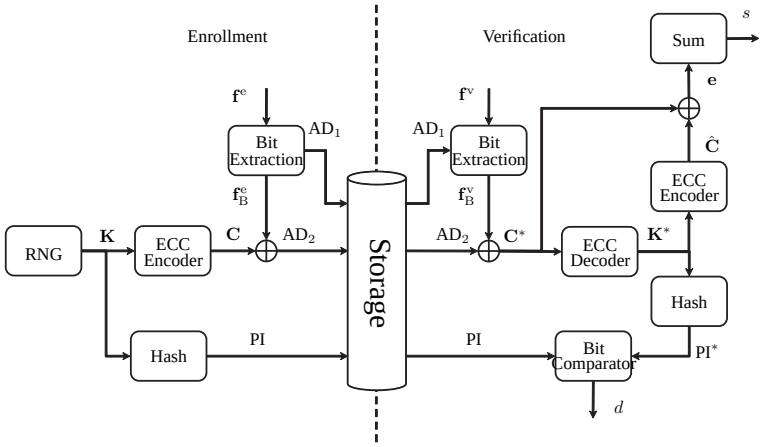


Figure 7.7: The HDS template protection scheme.

7.3.3 Template Protection Scheme

Many template protection schemes presented in the literature are based on the capability of generating a robust binary vector or key from biometric measurements of the same subject. The HDS system we consider is depicted in Figure 7.7. For the sake of coherence we use the terminology *auxiliary data* (AD) and *pseudonymous identifier* (PI) proposed in [102], which is in line with standardization activities in ISO. From the real-valued representation of the biometric sample, $\mathbf{f} \in \mathbb{R}^{N_{\mathrm{F}}}$, a binary vector $\mathbf{f}_{\mathrm{B}} \in \{0,1\}^{N_{\mathrm{B}}}$ is extracted within the *Bit Extraction* module. We use a single bit quantization scheme based on thresholding and the *reliable component selection* (RCS) algorithm. The N_{B} most reliable components are selected based on the estimated z-score for each component. With use of the multiple enrollment samples, the z-score is estimated as the ratio between the distance of the estimated mean with respect to the quantization threshold and the estimated standard deviation, see [33] for a more detailed description of the z-score estimation and the quantization scheme. The auxiliary data AD_1 contains the index information of the selected reliable components.

The binary vector $\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}$ could be used as a key for any encryption purposes, however it is not considered as being practical because of the high probability that it is not exactly the same in both the enrollment and verification phase ($\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \neq \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}$), due to measurement noise and biometric variability that lead to *bit errors*. The number of bit errors is also referred to as the Hamming distance $d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}})$. Therefore, error-correcting codes (ECC) are used to deal with the bit errors. Combining the ECC with a cryptographic hash function forms the scheme also known as the Fuzzy Commitment scheme [36]. In the enrollment phase, a binary secret or message vector $\mathbf{K}$ is randomly generated by the *Random-Number-Generator* (*RNG*) module. A codeword $\mathbf{C}$ of an error-correcting code is obtained by encoding $\mathbf{K}$ in the *ECC-Encoder* module. As the ECC we use the linear block type code "Bose, Ray-Chaudhuri, Hocquenghem" (BCH) [145], which is specified

by the codeword length (n_c), message length (k_c), and the corresponding number of bits that can be corrected (t_c), in short $[n_c, k_c, t_c]$. Some practical BCH settings are provided in Table 7.5, where the bit error rate (BER) is the ratio t_c/n_c. The codeword is XOR-ed with $\mathbf{f}_B^e$ in order to obtain the auxiliary data AD_2. Hence, $\mathbf{f}_B^e$ should have the same dimension as $\mathbf{C}$ implying $N_B = n_c$. Furthermore, the hash of $\mathbf{K}$ is taken in order to obtain the pseudonymous identifier PI. The larger the secret size the more difficult it is to guess $\mathbf{K}$ from PI.

In the verification phase, a new biometric sample is taken and transformed into its binary representation within the *Bit Extraction* module with help of auxiliary data AD_1. The new word $\mathbf{C}^*$ is computed by XOR-ing $\mathbf{f}_B^v$ with AD_2. The candidate secret $\mathbf{K}^*$ is obtained by decoding $\mathbf{C}^*$ in the *ECC-Decoder* module. Subsequently, the candidate pseudonymous identifier PI^* is computed by hashing $\mathbf{K}^*$. The decision in the *Comparator* module is based on whether PI and PI^* are bitwise identical.

The *Comparator* module yields identical PI and PI^* when the number of bit errors between the binary vectors $\mathbf{f}_B^e$ and $\mathbf{f}_B^v$ is smaller or equal to the error-correcting capability t_c of the ECC. Thus, there is an accept when the Hamming distance is smaller than t_c, $d_H(\mathbf{f}_B^e, \mathbf{f}_B^v) = ||\mathbf{f}_B^e \oplus \mathbf{f}_B^v||_1 \leq t_c$. Therefore, the fuzzy commitment scheme can be considered as a Hamming distance classifier with threshold t_c. Note, that the maximum number of bits that the BCH can correct t_c^* is close to 25% of the codeword length. In the remainder of the text, we indicate this limitation as the *ECC-limitation*.

As a distance score s we use the number of bits that had to be corrected by the ECC decoder. The candidate secret $\mathbf{K}^*$ is encoded to its corresponding codeword $\hat{\mathbf{C}}$ and is XOR-ed with $\mathbf{C}^*$ in order to obtain the error pattern $\mathbf{e}$. The error pattern is equal to the bit differences between the enrollment and verification binary feature vectors ($\mathbf{f}_B^e \oplus \mathbf{f}_B^v$) as

Table 7.5: Some examples of the BCH code given by the codeword (n_c and message (k_c) length, the corresponding number of correctable bits (t_c), and the bit error rate (BER) t_c/n_c.

n_c	k_c	t_c	BER $= t_c/n_c$
127	8	31	24.4%
	15	27	21.3%
255	9	63	24.7%
	21	55	21.6%
511	10	127	24.9%
	31	109	21.3%
1023	11	255	24.9%
	46	219	21.4%

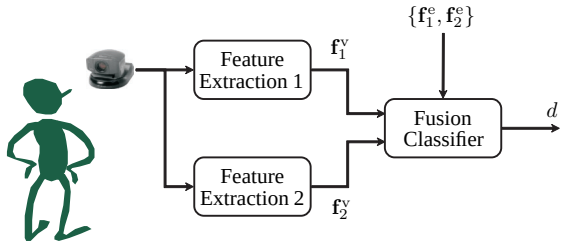


Figure 7.8: A toy-example of a multi-algorithm fusion system.

follows

$$\begin{aligned}\mathbf{e} &= \hat{\mathbf{C}} \oplus \mathbf{C}^* \\ &= \hat{\mathbf{C}} \oplus (\mathbf{f}_{\mathrm{B}}^{\mathrm{v}} \oplus \mathrm{AD}_2) \\ &= \hat{\mathbf{C}} \oplus (\mathbf{f}_{\mathrm{B}}^{\mathrm{v}} \oplus (\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \oplus \mathbf{C})) \\ &= (\hat{\mathbf{C}} \oplus \mathbf{C}) \oplus (\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \oplus \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}) \\ &= (\mathbf{f}_{\mathrm{B}}^{\mathrm{e}} \oplus \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}) \text{ if } \hat{\mathbf{C}} = \mathbf{C},\end{aligned} \tag{7.2}$$

where $\hat{\mathbf{C}}$ is equal to $\mathbf{C}$ when there is an accept, i.e. $\mathbf{K}$ and $\mathbf{K}^*$ are equal. The distance score s is thus the sum of the error pattern, hence equal to $d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}})$ and only a valid score when there is an accept, i.e. $d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}) \leq t_{\mathrm{c}}$. If the score is not valid we only know that $d_{\mathrm{H}}(\mathbf{f}_{\mathrm{B}}^{\mathrm{e}}, \mathbf{f}_{\mathrm{B}}^{\mathrm{v}}) > t_{\mathrm{c}}$.

7.3.4 Applying Template Protection at Different Fusion Levels

In this work we are interested in the multi-algorithm fusion system as depicted in Figure 7.8, where a 3D image is taken of the face of the subject from which the feature vectors $\mathbf{f}_1^{\mathrm{v}}$ and $\mathbf{f}_2^{\mathrm{v}}$ are extracted using two different feature extraction algorithms. These features are compared with their enrolled version $\{\mathbf{f}_1^{\mathrm{e}}, \mathbf{f}_2^{\mathrm{e}}\}$ within the *Fusion Classifier* module and a decision d is made whether to accept or reject the identity claim of the subject.

The comparison within the *Fusion Classifier* module can occur at different levels, namely at feature-, score-, or decision-level. In the following sections we discuss the implementation of the template protection system at the different fusion levels.

Feature-Level Fusion

Applying the template protection scheme at feature-level fusion is straightforward, the two feature vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ are concatenated before entering the template protection scheme, thus $\mathbf{f} = [\mathbf{f}_1; \mathbf{f}_2]$. The fused feature vectors have more components and most likely more components that have discriminating and robust properties. Hence, it is expected that more robust and discriminating bits can be extracted, which allows the use of larger binary vectors $\mathbf{f}_{\mathrm{B}}$ and thus larger codewords. It is known from the BCH code that larger codewords are more efficient, they have a larger secret at the same bit error rate (BER), see Table 7.5.

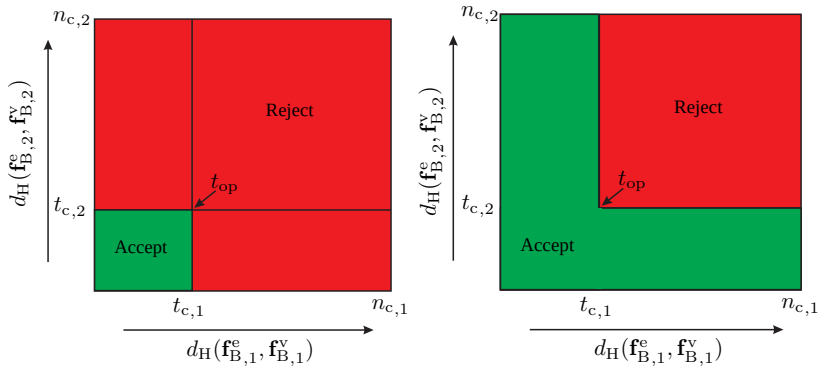


Figure 7.9: Decision boundaries for the (a) AND and (b) OR decision fusion rule. The operating point t_{op} is at the intersection of the decision boundary given by $t_{\mathrm{c}1}$ and $t_{\mathrm{c}2}$.

Decision-Level Fusion

At decision-level fusion there is a template protection system for each source of biometric information with an individual decision for each system. The two decisions can be fused into a single decision d_f using a AND-rule or OR-rule. For the AND-rule, there is a final accept if and only if both template protection systems lead to an accept, thus $d_{\mathrm{H}}(\mathbf{f}^{\mathrm{e}}_{\mathrm{B},1}, \mathbf{f}^{\mathrm{v}}_{\mathrm{B},1}) \leq t_{\mathrm{c},1}$ and $d_{\mathrm{H}}(\mathbf{f}^{\mathrm{e}}_{\mathrm{B},2}, \mathbf{f}^{\mathrm{v}}_{\mathrm{B},2}) \leq t_{\mathrm{c},2}$. The acceptance region is the intersection defined by the individual decision boundaries crossing the operating point $t_{\mathrm{op}} = \{t_{op,1},\ t_{op,2}\} = \{t_{\mathrm{c},1}, t_{\mathrm{c},2}\}$ as shown in Figure 7.9(a). For the OR-rule, there is a final accept if at least a single template protection system gives an accept. Hence, the acceptance region is the union of both as portrayed in Figure 7.9(b).

Under the assumption that the binary vectors $\mathbf{f}_{\mathrm{B}}$ are randomly distributed in $\{0,1\}^{N_{\mathrm{B}}}$, it follows from the results in [146] that the maximum amount of privacy information that the HDS system can preserve is equal to the secret size $|\mathbf{K}| = k_c$ from the ECC. The average number of attempts necessary for the adversary to randomly guess the secret $\mathbf{K}$ from its hashed version PI is equal to $\frac{1}{2}2^{k_{\mathrm{c}}}$. For the first source the secret size is $|\mathbf{K}_1| = k_{\mathrm{c},1}$ and $|\mathbf{K}_2| = k_{\mathrm{c},2}$ for the second source. For the OR-rule fusion, only one of the hash values has to be guessed correctly for a successful attack, hence the effective secret size in the fused setup is equal to the smallest secret size $|\mathbf{K}_f| = \min(k_{\mathrm{c},1}, k_{\mathrm{c},2})$. In case of the AND-rule fusion, both hash values have to be guessed correctly independently, thus the effective secret size is $|\mathbf{K}_f| = \log_2(2^{k_{\mathrm{c},1}} + 2^{k_{\mathrm{c},2}}) \leq \max(k_{\mathrm{c},1}, k_{\mathrm{c},2}) + 1$, where the equality holds only when $k_{\mathrm{c},1} = k_{\mathrm{c},2}$. This can be improved by combining or concatenating both secrets prior to hashing. In that case, the effective secret size is $|\mathbf{K}_f| = |\mathbf{K}_1| + |\mathbf{K}_2| = k_{\mathrm{c},1} + k_{\mathrm{c},2}$.

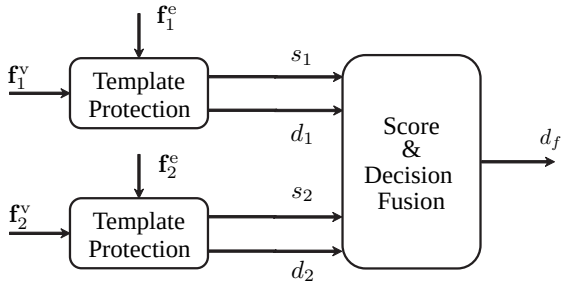
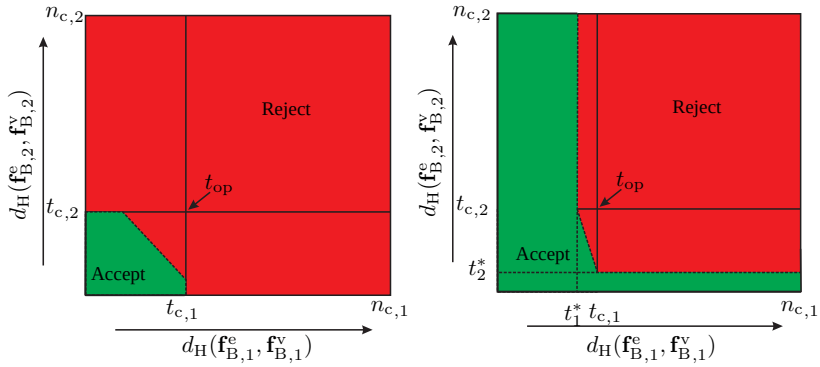


Figure 7.10: Score fusion with template protection.



(a) AND-configuration (b) OR-configuration

Figure 7.11: Examples of the decision boundaries for the score-level fusion case with (a) the AND- and (b) OR-configuration.

Score-Level Fusion

A general implementation of the template protection system at score-level fusion is depicted in Figure 7.10. Each source of biometric information has a separate template protection system with a decision and score value as output. Note that we are using the number of corrected bits within the ECC as the distance score that is valid only when there is an accept, see Section 7.3.3. Both scores (s_1 and s_2) and decisions (d_1 and d_2) are combined in the Score & Decision Fusion module into a single decision d_f. With the available scores, more flexible decision boundaries can be defined when compared to the AND-rule and OR-rule decision-level fusion cases that were presented in Figure 7.9. Similar to the decision-level fusion case, an AND- or OR-rule can be used based on the decision d_i, which is now extended by incorporating the scores s_i to determine the final decision d_f. Hence, there are two cases we refer to as the AND-configuration and the OR-configuration case.

For the AND-configuration case the initial acceptance region is similar to the AND-

rule case as shown in Figure 7.9(a). However, with use of the scores s_i a more refined decision boundary given by a function $f(s_1, s_2)$ can be defined. We mainly focus on the Sum-rule and Weighted-Sum-rule given as

$$s_f = w_1 s_1 + w_2 s_2, \text{ with } w_1 + w_2 = 1, \tag{7.3}$$

where the Sum-rule is a degenerate case of the Weighted-Sum-rule by using weights equal to $\frac{1}{2}$. If there is an accept for both sources ($d_1 = d_2 = 1 =$ "Accept"), then there is only a final accept ($d_f = 1$) if the scores s_1 and s_2 are in the acceptance region defined by the function $f(s_1, s_2)$, see Figure 7.11(a) for an example of the acceptance region using the Weighted-Sum-rule.

For the OR-configuration case, the same boundaries can be defined as for the AND-configuration case when there is an accept for both sources. However, if there is a single accept it is still possible to give a final accept if the single score s_i is smaller than a stricter threshold t_i^*. We use a stricter threshold because the final decision is now only based on a single source of biometric information. An example of the acceptance region is depicted in Figure 7.11(b). Note that we define the stricter threshold t_1^* (t_2^*) as the intersection of the decision boundary function $f(s_1, s_2)$ with the $t_{c,2}$ ($t_{c,1}$).

7.3.5 Experiments

In the previous section we presented the methods for multi-algorithm fusion at feature-, score-, and decision-level. In this section, we empirically validate the best performance achieved at each level by means of a biometric database and two feature extraction algorithms.

Experiment Setup

Biometric Databases All the results in this work are obtained using the FRGC v2 dataset [99] containing a total of 4007 3D shape samples from 465 subjects.

However, one of the 3D shape recognizer we used could not successfully extract a feature vector out of each sample, hence reducing the dataset to 3507 samples from 454 subjects. As the template protection algorithm works best at multiple enrollment samples, the subset of subjects with at least 6 (5 as enrolment samples with at least one for the verification) samples or more is selected. This resulted into a subset of 261 subjects with in total 2970 samples.

Feature Extraction Algorithms The first algorithm is the shape-based 3D face recognizer from [106] and is referred to as Algo1. It has two main steps: 1) the alignment of faces, and 2) the extraction of surface features from 3D facial data. In the alignment step, each face is registered to a generic face model (GFM) and the central facial region is cropped. The GFM is computed by averaging correctly aligned images from a training set. After the alignment step, we can assume that all faces are transformed in such a way that they best fit the GFM, and have the same position in the common coordinate system.

After alignment, the facial surface is divided into 174 local regions. For each region, the maximum and minimum principal curvature direction are computed. Each of the two directions is presented by the azimuthal and the polar angle in the spherical coordinate system. Combining all the regions leads to a feature vector dimension $N_{\mathrm{F}} = 174 \times 2 \times 2 = 696$.

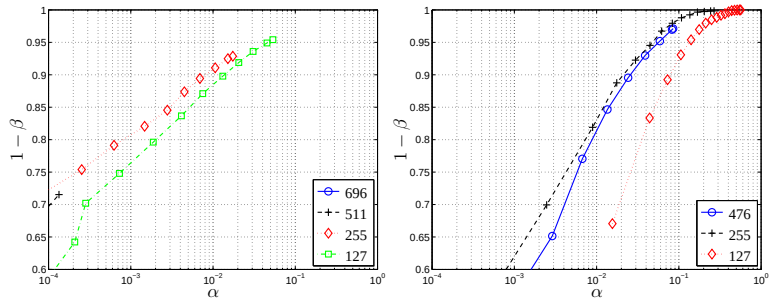
The second algorithm, Algo2, is a histogram-based feature extraction method. After the pre-registration of the face data, a frontal view of the face model is obtained, where the tip of the nose is at the origin in the Cartesian coordinate system. The distribution of depth values of the normalized face model describes the characteristics of an individual facial surface. In order to obtain more detailed information about the local geometry, the 3D model is divided into several sub areas which are orthogonal to the symmetry plane of the face. The features are extracted from the depth value distribution in each sub-area. The feature vector dimension is $N_{\mathrm{F}} = 476$. A full description of this algorithm is provided in [147].

For both feature extraction algorithms, the raw feature vectors they produce are used as input of the template protection system as described in Section 7.3.3. Hence, no signal processing is performed.

Testing Protocols The performance testing protocol consists of randomly selecting 50% (130) subjects as the training set and the other subjects as the test set, this is referred to as the training-test-set split. The template protection system parameters such as the quantization thresholds, used within the *Bit Extraction* module, are estimated on this training set. Hereafter, the test set is randomly split into an equally sized fusion-training and evaluation set containing around 65 subjects each. All the training needed for fusion is thus performed on the fusion-training set and the reported performance is obtained from the evaluation set. From the evaluation set, 5 samples of each subject are randomly selected as the enrollment samples while the remaining samples are considered as the verification samples. This split is referred to as the enrollment-verification split. The protected template is generated using all the enrollment samples and compared with each verification sample.

The training-test-set split is performed five times, while for each split the enrollment-verification split is performed five times. From each enrollment-verification split we measure the β_{tar} (the false rejection rate (FRR, β) at the targeted false acceptance rate (FAR, α) of $\alpha_{\mathrm{tar}} = 0.25\%$) and the equal-error rate (EER), which is the error rate achieved at the operating point where both FRR and FAR are equal. With use of the 25 measurements we estimate the 95% confidence interval (ci) defined as $ci = 1.96\sigma_{EER}/\sqrt{(25)}$ for the EER case while using $\sigma_{\beta_{\mathrm{tar}}}$ for the β_{tar} case, respectively. Note, that the splits are performed randomly, however the seed at the start of the protocol is always the same, hence all the splits are equal for the performance tests at feature-, score-, and decision-level fusion. Hence, the splitting process does not contribute to any performance differences.

Experiment Results



(a) Algo1 (b) Algo2

Figure 7.12: Individual ROC curves for algorithm (a) Algo1 and (b) Algo2 at different settings of n_{c}.

Individual Algorithm Performances Before we start fusing the different biometric sources, we first determine their individual performance as given by the ROC curves in Figure 7.12 for different codeword lengths n_{c} with the EER and β_{tar} details in Table 7.6. The table provides the ci for both EER and β_{tar} and their operating point provided as the relative Hamming distance (RHD). The right column of the table provides the secret size $|\mathbf{K}|$ of the ECC corresponding to the t_{c} setting that leads to closest α but smaller than the target α_{tar}. This is the ECC setting with a BER just larger than the operating point in RHD corresponding to β_{tar}. Entries in the table indicated with quotes cannot be reached in practice because of the ECC-limitation, however we are able to estimate them because of the Hamming distance classifier assumption as discussed in Section 7.3.3. Entries with "x" can neither be reached nor estimated.

Note that we used five enrollment samples ($N_{\mathrm{e}} = 5$) from which the average is taken. Also note that the ROC curves are limited because of the ECC-limitation. In order to

Table 7.6: The EER and β_{tar}, and their ci and operating point for the individual algorithms Algo1 and Algo2 at different settings of n_{c}. The last column is the secret size $|\mathbf{K}|$ of the ECC at the operating point t_{c} for achieving α_{tar}.

n_{c}	*EER* [%]	*RHD* [%]	β_{tar} [%]	*RHD* [%]	$\lvert\mathbf{K}\rvert$ [bits]
Algo1					
696	"3.75 ± 0.21"	"38.8"	"16.02 ± 1.61"	"33.6"	x
511	"3.69 ± 0.26"	"35.0"	"14.91 ± 1.63"	"29.0"	x
255	"3.99 ± 0.35"	"27.5"	15.33 ± 1.84	20.0	21
127	4.84 ± 0.42	23.6	19.18 ± 1.82	15.0	29
Algo2					
476	5.44 ± 0.35	22.1	37.69 ± 3.14	11.8	45
255	5.06 ± 0.30	10.2	30.25 ± 2.88	2.0	215
127	8.92 ± 0.33	3.9	89.57 ± 1.20	0	120

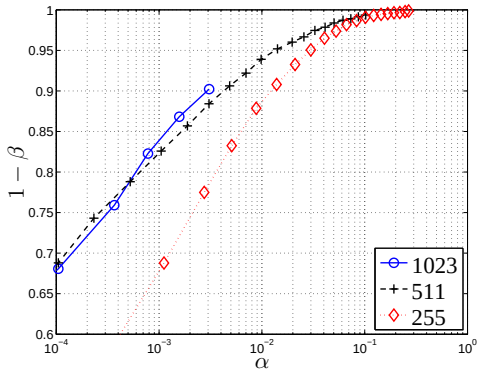


Figure 7.13: ROC curves at feature-level fusion of Algo1 and Algo2 algorithm.

reach larger α and smaller β values it is required to tolerate and thus correct more bit errors. However, the error correcting capability of an ECC is limited. From the results we can conclude that both algorithms perform optimally at a codeword size of $n_{\mathrm{c}} = 255$. These settings are used in the score- and decision-level fusion analysis. Compared to the Algo2 algorithm, Algo1 has a better performance but a smaller secret size (see Table 7.6, right column).

Multi-Algorithm Fusion at Feature-Level At feature-level we concatenate both feature vectors together and consider it as a single feature vector. The new dimension of the feature vector is 1175. Because of the larger dimension it is possible to use larger codeword lengths as in the individual case in Section 7.3.5. The performances at different codeword lengths are shown in Figure 7.13 with the EER and β_{tar} details in Table 7.7. The best performance is achieved by using the largest codeword length of 1023 bits. It is just able to reach the targeted α_{tar} that leads to a $\beta_{\mathrm{tar}} = 11.1\%$.

Multi-Algorithm Fusion at Decision-Level At decision and score-level fusion, the scatter plot of the genuine and imposter scores of both algorithms, as shown in Fig-

Table 7.7: Performance results of multi-algorithm fusion at feature-level.

n_{c}	*EER* [%]	*RHD* [%]	β_{tar} [%]	*RHD* [%]	$\lvert\mathbf{K}\rvert$ [bits]
1023	"2.45 ± 0.24"	"29.6"	11.10 ± 1.70	24.5	11
511	2.89 ± 0.34	18.6	12.88 ± 1.71	11.7	103
255	3.89 ± 0.32	11.8	22.79 ± 2.64	5.1	155

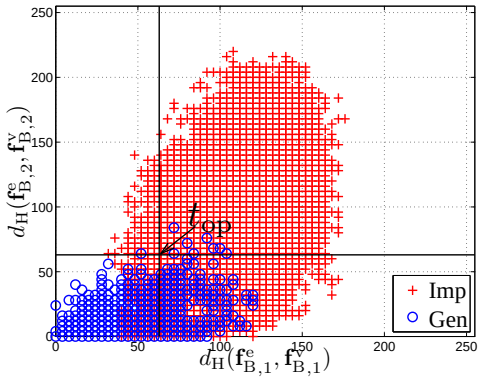
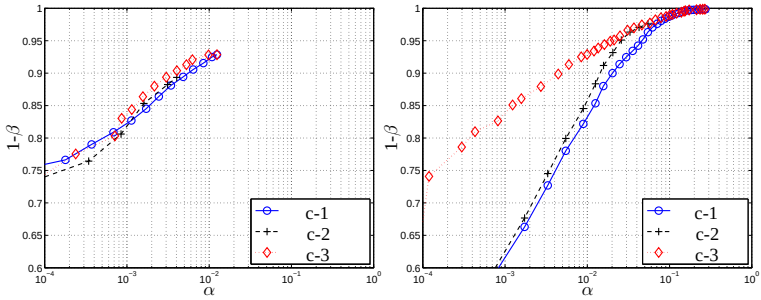


Figure 7.14: Scatter plot of the genuine (Gen) and imposter (Imp) scores of the algorithms Algo1 and Algo2. The operating point t_{op} is at the intersection of the vertical and horizontal decision boundaries of Algo1 and Algo2, respectively.

ure 7.14, may indicate the possible gain when fusing at these levels. The scatter plot also depicts the decision boundary indicated by the operating point t_{op}.

We will investigate both the AND-rule and OR-rule performance at different strategies of moving the operating point $t_{\mathrm{op}} = \{t_{op,1},\ t_{op,2}\}$ on the scatter plot, whose range is $t_{op,1} \in [0,\ t^*_{\mathrm{c},1}]$ and $t_{op,2} \in [0,\ t^*_{\mathrm{c},2}]$ for each axis respectively, with t^*_{c} being the maximum error-correcting capability of the ECC. In the first case (c-1) we consider $t_{op,2} = t_{op,1}$ and vary $t_{op,1}$ from 0 to $t^*_{\mathrm{c},1}$ considering that $t^*_{\mathrm{c},1} = t^*_{\mathrm{c},2}$ because the optimal individual performance is at the same codeword length as observed in Section 7.3.5. In the second case (c-2), the operating point crosses the EER operating point of the individual performances $\{t_{EER,1},\ t_{EER,2}\}$ linearly, hence the operating point is defined as $t_{\mathrm{op}} = \{t_{op,1}, \frac{t_{EER,2}}{t_{EER,1}} t_{op,1}\}$ with $t_{op,1} \in [0,\ \min(t^*_{\mathrm{c},1}, \frac{t_{EER,1}}{t_{EER,2}} t^*_{\mathrm{c},2})]$. In the third and final case (c-3) we use the optimal fusion method from [148], which estimates the performance in terms of α and β at each possible operating point in the scatter plot and takes the operating points on the envelope which leads to the best performance. This optimization process of finding the optimal operating points is in fact a training process and is thus performed on the fusion-training set. The final performance results are obtained by calculating the performance of the test set on the optimal operating points.

The performance results of the three cases are shown in Figure 7.15(a) for the AND-rule and Figure 7.15(b) for the OR-rule respectively with the performance details provided in Table 7.8. Because there are two template protection systems we provide the RHD of the operating point and the secret size for each system. From the results we can conclude that the optimal decision fusion method (c-3) leads to the best performance for both the AND-rule and OR-rule method. The performance differences between the three cases of moving the operating point is very small for the AND-rule method, while significant for the OR-rule method. This difference becomes more evident when analyzing the trajectory of the operating point as depicted in Figure 7.16. The optimal operating points obtained



(a) AND-rule (b) OR-rule

Figure 7.15: Decision fusion results for (a) the AND-rule and (b) the OR-rule for the three cases (c-1, c-2, c-3).

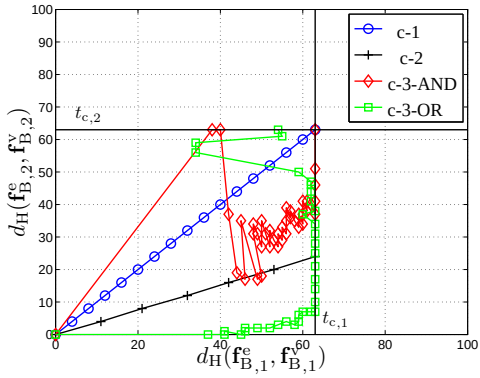


Figure 7.16: The operating points trajectory for the three cases (c-1, c-2, c-3) for the AND-rule and OR-rule decision fusion methods.

Table 7.8: Performance results of multi-algorithm fusion at decision-level. The operating points and secret size are provided for both template protection systems.

n_c	*EER* [%]	*RHD* [%]	β_{tar} [%]	*RHD* [%]	$\lvert\mathbf{K}\rvert$ [bits]
AND-rule					
c-1	x	x	13.45 ± 1.87	$[20.8, 20.8]$	$[21, 21]$
c-2	x	x	12.71 ± 2.59	$[23.9, 9.0]$	$[9, 99]$
c-3	x	x	11.34 ± 2.72	$[22.0, 13.7]$	$[13, 47]$
OR-rule					
c-1	4.78 ± 0.29	$[10.2, 10.2]$	29.83 ± 3.31	$[2.4, 2.4]$	$[207, 207]$
c-2	3.46 ± 0.34	$[21.2, 7.8]$	28.23 ± 3.50	$[6.7, 2.4]$	$[131, 207]$
c-3	3.27 ± 0.38	$[24.7, 5.9]$	12.58 ± 6.27	$[19.2, 0.8]$	$[21, 239]$

by the optimal AND-rule method (c-3-AND) is between the operating points of cases c-1 and c-2. However, for the optimal OR-rule method (c-3-OR) the obtained operating points are significantly different than for case c-1 and c-2. For the first few points the operating points moves to the right, tangent to the x-axis ($t_{op,1}$ increases while $t_{op,2}$ stays relatively constant) and sharply moves up ($t_{op,2}$ increases) once $t_{op,1}$ reaches the limit of $t_{\mathrm{c},1}$. Because the optimal fusion method facilitates more flexibility of the operating points, it significantly improves the performance as is shown in Figure 7.15(b).

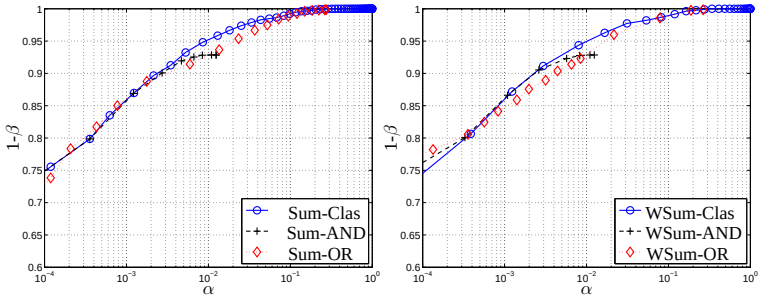
Observe that the OR-rule is able to obtain a greater part of the ROC curve than the AND-rule, as the OR-rule is able to reach the EER operating point while the AND-rule cannot, while both have the same ECC-limitation. The decision boundaries in Figures 7.9, 7.11, and 7.14 clearly show that at the same operating point the OR-rule has a larger Accept area than the AND-rule and can thus achieve a larger α and smaller β.

The effective secret size as discussed in Section 7.3.4 depends on the configuration being used. For the AND-configuration, the total secret size is the sum of the secret size of each template protection system individually. For the OR-configuration case the effective secret size is the minimum of both.

Multi-Algorithm Fusion at Score-Level The scatter plot indicates that using a Sum-rule or Weighted-Sum-rule score fusion method should improve the overall performance with respect to the individual performances. For the Weighted-Sum-rule method given by (7.3), the weighting coefficients are estimated from the disjunct fusion-training set as discussed in Section 7.3.5. The weights are iteratively varied and the values with the best performance in terms of the EER are selected. If the EER cannot be estimated, for example because of the ECC-limitation, we optimize using β_{tar} instead.

The score fusion algorithm can only be applied when the scores of both algorithm are available as portrayed by the accept region in Figure 7.11(a) for the AND-configuration case. The accept region can be extend by using the OR-configuration given in Figure 7.11(b). If only a single score s_1 (s_2) is available a stricter threshold t_1^* (t_2^*) is used. Note that the ECC settings are set to t_{c}^* for both template protection systems in order to have the largest acceptance region where both scores are available, hence fully benefitting from the score-fusion method. Thus, the threshold variable for the ROC curve becomes the weighted sum given by (7.3).

The results for the Sum-rule and Weighted-Sum-rule score fusion methods are depicted in Figure 7.17(a) and (b), respectively. We investigate both the AND- and OR-configuration indicated as Sum-AND and Sum-OR for the Sum-rule and WSum-AND and WSum-OR for the Weighted-Sum-rule. As a comparison, the classical Sum-rule and Weighted-Sum-rule without the ECC limitation are included and referred to as Sum-Clas and WSum-Clas, respectively. The average weights $[\bar{w_1}, \bar{w_2}]$ found during the fusion training are $[0.59, 0.41]$ for the WSum-Clas case, $[0.7, 0.3]$ for the WSum-AND case, and $[0.8, 0.2]$ for the WSum-OR case. More performance details are provide in Table 7.9. Because there are two template protection system we provide the RHD of the operating point and the secret size for each system. In terms of the β_{tar} values, the results indicate that the AND-configuration outperforms the OR-configuration but not the classical results without the ECC-limitations. Within the AND-configuration, the Weighted-Sum-rule has the



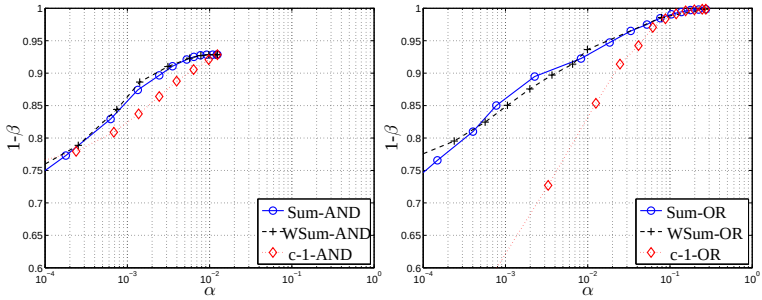
(a) Sum-rule (b) Weighted-Sum-rule

Figure 7.17: ROC curves at score-level fusion using (a) the Sum-rule and (b) the Weighted-Sum-rule. In both cases we compare the classical performance (Clas) where there is no ECC-limitation with the AND- and OR-configuration with ECC-limitation.

best performances, while the Sum-rule has a better performance for the OR-configuration case. Note, that all the measured differences are within the estimated confidence intervals, hence the observed differences cannot be considered as being significant. The results also show that the Sum-AND (WSum-AND) curve follows the Sum-Clas (WSum-Clas) curve at smaller α values, but starts deviating at larger α values. At smaller α values the accept area for the Sum-AND case is not limited by the ECC-limitation and is thus equal to the accept area of the Sum-Clas case. This also holds for the WSum-AND and WSum-Clas scenario only if the weights are equal for both cases. However at larger α values the decision boundary is at a larger Hamming distances with the consequence that the accept area for the WSum-AND and Sum-AND cases are limited by the ECC-limitation as shown in Figure 7.11(a) and approaches the accept area for the AND-rule c-1 decision-level fusion method case as depicted in Figure 7.9(a). Under the same conditions this also holds for the OR-rule cases. The convergence of the score-level fusion ROC curves towards the decision-level curves are portrayed in Figure 7.18.

Table 7.9: Performance results of multi-algorithm fusion at feature-level.

case	*EER* [%]	β_{tar} [%]	$\lvert\mathbf{K}\rvert$ [bits]
Sum			
Clas	2.58 ± 0.30	9.83 ± 1.81	$[9, 9]$
AND	x	10.26 ± 1.80	$[9, 9]$
OR	3.45 ± 0.37	10.38 ± 1.56	$[9, 9]$
WSum			
Clas	2.57 ± 0.32	9.58 ± 1.74	$[9, 9]$
AND	x	9.63 ± 2.20	$[9, 9]$
OR	3.28 ± 0.39	11.68 ± 1.74	$[9, 9]$



(a) AND-rule (b) OR-rule

Figure 7.18: Convergence of the score-level fusion ROC curves (Sum and Wsum) towards the decision-level curves (c-1) for the (a) OR-rule and (b) AND-rule cases.

Table 7.10: Summary of empirical results of multi-algorithm.

Type	*EER* [%]	*RHD* [%]	β_{tar} [%]	*RHD* [%]	k_{c} [bits]
Feature	x	x	11.10 ± 1.70	24.5	11
Score	x	x	9.63 ± 2.20	$[24.7, 24.7]$	$[9, 9]$
Decision	x	x	11.34 ± 2.72	$[22.0, 13.7]$	$[13, 47]$
Algo1	x	x	15.33 ± 1.84	20.0	21
Algo2	x	x	30.25 ± 2.88	2.0	215

Because we fixed the ECC correcting capability at t_{c}^{*} the secret size for each protected template is 9 bits at $n_{\text{c}} = 255$ and the effective secret size is the sum of 18 bits for the AND-configuration when both secrets are concatenated before hashing. For the OR-configuration case the effective secret size is the minimum of both, hence 9 bits.

Summary and Discussions As a summary we compare the performance of the individual algorithms with the best performances obtained at each fusion level, see Figure 7.19 for the ROC curves with the details in Table 7.10. The best performance at feature-level fusion was with a codeword of 1023 bit. At score-level fusion, the best performance is obtained using the Weight-Sum-rule with the AND-configuration, while at decision-level fusion the optimal AND-rule method led to the best performance.

Compared to the individual performances, the performance improvement with fusion in terms of β_{tar} exceeds 6%. The difference can be considered as significant because the combined confidence interval is around 4%. The best performance is obtained at score-level fusion, however the differences with the feature- and decision-level fusion methods are not significant. The effective secret size at score-level fusion is close to the secret size of the best individual algorithm. Hence we can conclude that the performance has been improved while maintaining a similar effective secret size.

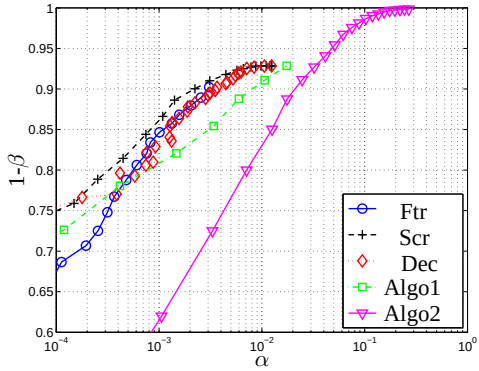


Figure 7.19: Overview of the best ROC curves obtained at feature-, score-, and decision-level fusion, and the individual algorithms Algo1 and Algo2.

7.3.6 Conclusions

We have shown that it is possible to apply fusion with the Helper-Data System at feature-, score-, and decision-level. However, the Helper-Data System inherently has only a decision as the output, hence it had to be adapted in order to have a score as output for the score-level fusion. We took the number of the bits the ECC had to correct as the distance score measurement.

Furthermore, we have also shown that applying fusion with template protection at feature- or decision-level is straightforward and conventional. However, fusion at score-level is different due to the use of an ECC, which has a limited error correcting capability. Consequently, for each template protection system there is only a valid score when there is a match. Hence, this ECC-limitation limits the decision boundaries.

The performance at all fusion levels is significantly better than the performance of the individual biometric sources. The best performance is obtained at score-level fusion, with a β_{tar} improvement of 6% while maintaining a similar secret size.

Acknowledgment

The authors would like to acknowledge the support of the partners within the 3DFACE project, a European Integrated Project funded under the European Commission IST FP6 program.

7.4 Chapter Conclusions

We have shown that it is possible to apply multi-sample and multi-algorithm fusion with the HDS system at feature-, score-, and decision-level. Because the HDS system inherently has only a decision as output of the comparison stage, we adapted the system accordingly in order to have a score as output for fusion at score-level. As the distance score we took the number of bits the ECC had to correct. Furthermore, applying fusion with template protection at feature- or decision-level is straightforward and in line with conventional approaches. However, fusion at score-level is different due to the use of an ECC, which has a limited error-correcting capability. Consequently, for each template protection system there is only a valid score when there is a match. Hence, this ECC-limitation limits the decision boundaries for fusion at score-level, see Figure 7.11.

For multi-sample fusion, no significant classification performance difference has been observed at feature-, score-, and decision-level. Because fusion at feature-level has only a single protected template, which is better in terms privacy and security protection and storage, we can conclude that the optimal multi-sample fusion is at feature-level. This has been the practice of many published paper on of the HDS in which multiple enrolment samples are being used, however a detailed analysis was so far missing.

For multi-algorithm fusion, the classification performance at feature-, score-, and decision-level are better than the performance of the individual biometric sources. Despite the ECC-limitation of fusion at score-level we obtained the best performance, where the absolute difference between the FNMR at the target FMR is 6% while maintaining a similar key size.

Chapter 8

Conclusions, Recommendations, and Future Directions

In this thesis we have analyzed the helper-data (HDS) template protection scheme from different perspectives. From the main research question

What is the performance of the helper-data template protection scheme?

we deduced the four more specific research questions, namely.

Given the helper data template protection scheme:

1. What is the **theoretical classification performance**?
   - i How can we model the classification performance?
   - ii How do the system parameters influence it?
   - iii How does it compare with the classification performance without template protection?
2. What is the **maximum key size** at a given target classification performance and system parameters?
3. How does the **information leakage from the auxiliary data** affect the irreversibility and unlinkability property?
4. How can one realize **fusion with protected templates** and to what extent can it improve the classification performance?

8.1 Answers to the Research Questions

The answers to the four research questions are discussed separately in the following sections.

8.1.1 Theoretical Classification Performance

In Section 3.2, we have shown that it is possible to theoretically determine the classification performance of the HDS based on a single bit extraction scheme employing a single quantization threshold. This was primarily accomplished by deriving a closed-form analytical expression of the average bit-error probability of the bit extracted from a component. The naive model assuming independent feature components with a homogeneous within-class variance has a large deviation, which can be reduced by incorporating the dependent and non-homogeneous feature components. Increasing the system parameters, such as the number of enrolment and verification samples, improve the classification performance by reducing the within-class variance, and also improve the performance estimation due to the central limit theorem.

In Section 3.3 we have shown that the classification performance of the unprotected templates (on continuous level) using the optimal likelihood ratio classifier is better than the performance of the protected templates using the HDS with a single bit extraction scheme based on a single quantization threshold. The results are optimistic, because they are based on the naive Gaussian model of independent components with a homogeneous within-class variance across the population, which seems not to hold in practice as we have shown in Section 3.2.

8.1.2 Maximum Key Size

We determined the maximum key size using the theoretical performance of the Gaussian framework from Section 3.2 and assuming the ECC operating on Shannon's bound. An important finding of this work is the fact that the maximum key size we determined is a couple of bits smaller than the upperbound of the key size for the HDS known in the literature, namely $-\log_2(\alpha_{\mathrm{tar}})$ of (4.21) where α_{tar} is the target FMR. The difference can be a couple of bits and increases with the number of feature components. When the FMR is taken as the target performance, the maximum key size is determined by the upperbound given by $-\log_2(\alpha_{\mathrm{tar}})$. However, when taking the FNMR as the target performance, the maximum key size depends on the target FNMR, the input capacity, the number of feature components, and the number of enrolment and verification samples.

With respect to the number of enrolment and verification samples, we have shown that increasing the number of enrolment samples to infinity leads to a similar performance when doubling both the number of enrolment and verification samples.

Considering the fact that having a larger target FNMR and more enrolment and verification samples do influence the convenience of the biometric system, we have shown a trade-off between the protection capability of the HDS in terms of key size and its convenience.

8.1.3 Information Leakage of the Auxiliary Data

In Section 6.2, we have shown that great care has to be taken when designing the DROBA bit extraction scheme in order to guarantee that AD_1 does not leak information about the binary vector extracted from the biometric sample and therefore affecting the irreversibil-

ity property. When not properly designed, the information leakage can be significant and an adversary is able to exploit this information and increase its success rate of impersonation by two orders of magnitude. As a solution to reduce this information leakage, we proposed and validated a remedy which in fact is a guideline on how to restrict the allocation freedom of the DROBA algorithm.

In Chapter 5 and Section 6.3 we have shown the cross-matching possibility of both AD_1 and AD_2. When having a balanced system, where the number of enrolment and verification samples are equal, the cross-matching performance is worse than the classification performance of the HDS. When there are more enrolment samples the cross-matching performance can become better than the system performance. Therefore, we would advice not to use more than four enrolment samples when there is a single verification sample. On the other hand the cross-matching performance can be made significantly worse with respect to the system performance by only increasing the number of verification samples. The cross-matching possibility due to the decodability attack on AD_2 can be made close to random by introducing an application dependent bit-permutation matrix randomization process (see Figure 8.1). The cross-matching on AD_1 is caused by the use of subject-specific information within the bit extraction scheme, which has to be stored in AD_1. Hence, improving the system performance by using more subject-specific information also improves the cross-matching performance.

In general, due to the information leakage we have identified, it is advisable to protect the bit extraction auxiliary data AD_1 by data separation principles (stored on a token) or by using encryption techniques.

8.1.4 Fusion

In Chapter 7 we have shown that it is possible to apply multi-sample and multi-algorithm fusion with the HDS system at feature-, score-, and decision-level. We adapted the HDS accordingly in order to facilitate fusion at score-level. We took as distance score the number of bits the ECC had to correct. Due to the limitation of the error-correcting capability of the ECC, the decision boundaries for fusion at score-level are restricted, see Figure 7.11. For multi-sample fusion, no significant classification performance difference has been observed at feature-, score-, and decision-level. For multi-algorithm fusion, the classification performances at feature-, score-, and decision-level are better than the performances of the individual biometric sources. Despite the ECC-limitation of fusion at score-level we obtained the best performance, where the absolute difference between the FNMR at the target FMR is 6% while maintaining a similar key size.

8.1.5 The Improved Helper Data System

From the answers to the these research questions we obtained an improved HDS scheme that is portrayed in Figure 8.1. The improvements are twofold, namely (i) from the results of the first part of the third research question (see Chapter 5) we proposed the *Bit Randomizer* module using a bit permutation transformation in order to prevent the decodability attack that leads to cross-matching and (ii) from the results of the fourth research question (see Chapter 7) we introduced the *Score Generation* module in order to generate

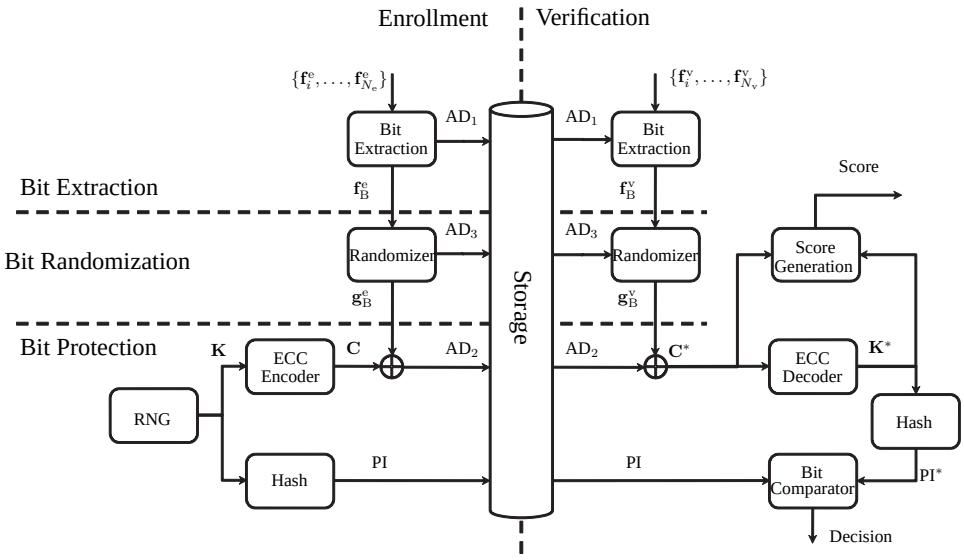


Figure 8.1: The improved helper-data system (HDS) template protection scheme with the required *Bit Randomizer* module using a bit permutation transformation in order to mitigate the cross-matching possibility based on the decodability attack (see Section 5.2) and the *Score Generation* module in order to generate a score that can be used for fusion at score-level (see Chapter 7).

a score that can be used for fusion at score-level.

8.2 Recommendations

Optimal Features An important parameter for protection of the HDS is the extracted key size, which depends on the performance of the underlying biometric recognition system. It is thus of great importance to improve the underlying feature extraction algorithm in order to extract features of better quality, i.e. features with a larger ratio of between-class and within-class variance. Furthermore, the maximum key size can be optimized by adapting the feature extraction algorithm such that the format of the feature vectors is optimized for the template protection scheme, ensuring equal feature quality with the optimal number of feature components.

Multiple Samples By averaging multiple samples the within-class variance will reduce and consequently the classification performance will improve. Furthermore, the within-class distribution becomes more Gaussian and therefore better fits the Gaussian model used for feature selection and quantization parameter. We have shown that increasing the

number of enrolment samples to infinity leads to a similar performance when doubling both the number of enrolment and verification samples. Hence, it is recommended to avoid the case that the number of enrolment samples is significantly larger than the number of verification samples, because the HDS classification performance will not improve significantly, while we have shown that the cross-matching performance can significantly improve and outperform the HDS performance. If there is a single verification sample, it would therefore be advised not to use more than 4 enrolment samples. The cross-matching performance can be degraded with respect to the HDS performance by increasing the number of verification samples. However, capturing more verification samples increases the verification time and may be considered as inconvenient.

Subject-Specific Information The HDS classification performance can be improved by using bit extraction schemes that incorporate more subject-specific information. However, care has to be taken as the subject-specific information has to be stored as part of the protected template and may leak information affecting the irreversibility and unlinkability properties. The use of data separation or encryption is advised in order to mitigate the information leakage.

8.3 Future Directions

The theoretical analysis in Chapter 3 and Chapter 4 are based on a bit extraction scheme that extracts a single bit using a single quantization threshold. It would be of great interest to analytically determine the performance of other bit extraction schemes, for example the reliable component selection (RCS) or the detection rate optimized bit allocation (DROBA) schemes. The complexity for the DROBA scheme will be significantly greater than for the RCS scheme, and it remains questionable whether a closed-form analytical expression, as obtained in Chapter 3, can be found for the bit-error probability. With the theoretical work we could determine the relationship between the HDS classification and cross-matching performance, which has now been experimentally analyzed in Chapter 6.

In contrast to the theoretical work presented in Section 3.3, the classification performance difference between the protected and unprotected templates has to be studied for a practical scenario, where there are dependencies between feature components and the within-class variances are not homogeneous. These parameters would have to be estimated on a training set of a limited size. Due to the limited size, estimation errors will occur and therefore influence the performance difference observed in Section 3.3. We conjecture that the performance difference will decrease due to the estimation errors of the within-class variance and feature component dependencies.

It is known that a Hamming distance classifier works optimally if the bit-error probabilities among the binary vector are equal. In our analysis in Chapter 3 and Chapter 4, the bit-error probabilities are equal only for the imposter comparisons. It would be of interest to analyze the possible classification performance gain by designing a bit extraction scheme that, if possible, enforces the bit-error probabilities at both genuine and imposter comparisons to be equal. A drawback of enforcing equal bit-error probabilities at genuine comparisons is the requirement of storing subject-specific information from the bit

extraction scheme, which we have shown to be vulnerable to cross-matching.

Furthermore, any improvements on the error-correcting capability of the ECC, bringing it closer to Shannon's bound, will directly increase the key size and therefore also the privacy and security protection of the HDS. These improvements have to occur mainly for the case of large bit-error probabilities as the binary vectors extracted from the biometric sample are noisy when compared to bit-error probabilities of modern communication channel.

Bibliography

[1] A. Jain, R. Bolle, and S. Pankanti, Eds., *Biometrics: Personal Identification in Networked Society*. Kluwer Academic Publishers, 1999.

[2] R. M. Bolle, J. H. Connell, and N. K. Ratha, "Biometric perils and patches," *Pattern Recognition*, vol. 35, no. 12, pp. 2727 – 2738, December 2002.

[3] L. O'Gorman, "Comparing passwords, tokens, and biometrics for user authentication," *Proceedings of the IEEE*, vol. 91, no. 12, pp. 2021 – 2040, 2003.

[4] A. K. Jain, A. Ross, and S. Pankanti, "Biometrics: A tool for information security," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 125 – 143, June 2006.

[5] RSASecurity, "Rsa security research shows volume of business passwords overwhelming end users and hindering it security efforts," www.rsa.com/press_release.aspx?id=7296, September 2006.

[6] SearchSecurity, "Survey: Most workers must remember six passwords or more," www.searchsecurity.techtarget.com, May 2003.

[7] K. Haley, "Password survey results," http://www.symantec.com/connect/blogs/password-survey-results.

[8] Identity Cards Act 2006, http://www.opsi.gov.uk/acts/acts2006/ukpga_20060015_en_1.

[9] ICAO, "International Civil Aviation Organization," http://www.icao.int/cgi/goto_m_atb.pl?icao/en/atb/fal/mrtd/MRTD_Rpt_V1N1_2006.pdf.

[10] Privium at Schiphol, http://www.schiphol.com/Travellers/AtSchiphol/PriviumIrisscan/WhyPrivium/FastBorderPassageWithTheIrisScan.htm.

[11] H. Ogata, "Interoperability of biometric transactions in japanese bank systems," in *1st ASEAN-FBI meeting*, Bangkok, Thailand, June 2009.

[12] P. Jones, "Banking on vein at the atm," in *Biometric technology today*, May 2006, pp. 8 – 9.

[13] L. Tan, "Citibank singapore launches biometric payment service," http://www.apacs.org.uk/09_03_19.htm, November 2009.

[14] P. Jones, "Banking on biometrics," in *Biometric technology today*, April 2007, pp. 7 – 8.

[15] R. Clarke, "Human identification in information systems: Management challenges and public policy issues," *Information Technology & People*, vol. 7, no. 4, pp. 6 – 37, 1994.

[16] T. van der Putte and J. Keuning, "Biometrical fingerprint recognition: Don't get your fingers burned," in *Fourth Working Conference on Smart Card Research and Advanced Applications*. Kluwer Academic Publishers, 2000, pp. 289 – 303.

[17] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino, "Impact of artificial "gummy" fingers on fingerprint systems." in *Optical Sec. and Counterfeit Deterrence Techn. IV. Volume 4677 of Proc. of SPIE.*, 2002.

[18] A. A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multibiometrics*, D. D. Zhang and A. K. Jain, Eds. Springer, 2006.

[19] "ISO/IEC JTC1 SC37. TR 24722 - information technology - biometrics - multimodal and other multibiometric fusion," 2009.

[20] H. J. Weinreb, "Fingerprint patterns in alzheimer's disease," *Archives of Neurology*, vol. 42, no. 1, pp. 50–54, January 1985.

[21] S. Rajangam, S. Janakiram, and I. Thomas, "Dermatoglyphics in down's syndrome," *Journal of the Indian Medicine Assocation*, vol. 93, no. 1, pp. 10–13, January 1995.

[22] J. Bolling, "A window to your health," in *Jacksonville Medicine, 51, Special Issue: Retinal deseases*, 2000.

[23] C. Hill, "Risk of masquerade arising from the storage of biometrics," Master's thesis, Australian National University, November 2001.

[24] A. Ross, J. Shah, and A. K. Jain, "From template to image: Reconstructing fingerprints from minutiae points," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 544 – 560, April 2007.

[25] "ISO/IEC JTC1 SC27. FCD 24745 - information technology - security techniques - biometric template protection," 2010.

[26] B. Schneier, "Inside risks: The uses and abuses of biometrics," *Communications of the ACM*, vol. 42, no. 8, p. 136, 1999.

[27] SPEED, "Signal Processing in the EncryptEd Domain," http://www.speedproject.eu/.

[28] B. Schoenmakers and P. Tuyls, "Efficient binary conversion for Paillier encrypted values," in *Advances in Cryptology - EUROCRYPT 2006*, ser. LNCS, vol. 4004. Springer Berlin / Heidelberg, 2006, pp. 522–537.

[29] J. P. Prins, Z. Erkin, and R. L. Lagendijk, "Anonymous fingerprinting with robust qim watermarking techniques," *EURASIP Journal on Information Security*, vol. 2007, p. 13, 2007.

[30] 3DFace, "3DFace EU Project," http://www.3dface.org/home/welcome.

[31] TURBINE, "TrUsted Revocable Biometric IdeNtitiEs EU Project," ttp://www.turbine-project.eu/.

[32] A. Cavoukian and A. Stoianov, "Biometric encryption: A positive-sum technology that achieves strong authentication, security and privacy," White Paper from information and Privacy Commissioner/Ontario, Tech. Rep., March 2007.

[33] 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 *Intl. Conf. on Biometrics*, Seoul, South Korea, August 2007, pp. 566 – 573.

[34] T. A. M. Kevenaar, G.-J. Schrijen, A. H. M. Akkermans, M. van der Veen, and F. Zuo, "Face recognition with renewable and privacy preserving binary templates," in *4th IEEE workshop on AutoID*, Buffalo, New York, USA, October 2005, pp. 21–26.

[35] P. Tuyls, A. H. M. Akkermans, T. A. M. Kevenaar, G.-J. Schrijnen, A. M. Bazen, and R. N. J. Veldhuis, "Practical biometric authentication with template protection," in *5th International Conference, AVBPA*, Rye Brook, New York, July 2005.

[36] A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in *6th ACM Conference on Computer and Communications Security*, November 1999, pp. 28–36.

[37] R. N. J. Veldhuis and A. Bazen, "One-to-template and one-to-one verification in the single and multi-user case," in *26th Symposium on Information Theory in the Benelux*, Brussels, 2005.

[38] F. M. J. Willems and T. Ignatenko, "Quantization effects in biometric systems," in *Proceedings of Information Theory and Applications Workshop*, San Diego, California, February 2009, pp. 372 – 379.

[39] F. Carter and A. Stoianov, "Implications of biometric encryption on wide spread use of biometrics," in *EBF Biometric Encryption Seminar*, Amsterdam, The Netherlands, June 2008. [Online]. Available: http://www.eubiometricsforum.com/pdfs/be/BE-Carter_Stoianov.pdf

[40] K. Simoens, P. Tuyls, and B. Preneel, "Privacy weaknesses in biometric sketches," in *Proceedings of the 2009 IEEE Symposium on Security and Privacy*. IEEE Computer Society, 2009.

[41] L. Ballard, S. Kamara, and M. K. Reiter, "The practical subtleties of biometric key generation," in *in the proceeding of the 17th Annual USENIX Security Symposium*, San Jose, CA, USA, August 2008, pp. 61–74.

[42] C. Chen, R. N. J. Veldhuis, T. A. M. Kevenaar, and A. H. M. Akkermans, "Biometric quantization through detection rate optimized bit allocation," *EURASIP Journal on Advances in Signal Processing*, 2009.

[43] I. R. Buhan, J. Breebaart, J. Guajardo Merchan, K. T. J. de Groot, E. J. C. Kelkboom, and T. Akkermans, "A quantative analysis of indistinguishability for a continuous domain biometric cryptosystem," in *Proc. 4th international workshop on data privacy management (DPM '09)*, ser. Lecture Notes in Computer Science, vol. 5939. St. Malo, France: Springer-Verlag, 2009, pp. 78–92.

[44] Y.-J. Chang, W. Zhang, and T. Chen;, "Biometrics-based cryptographic key generation," in *IEEE Int. Conf. on Multim. and Expo*, vol. 3, June 2004, pp. 2203 – 2206.

[45] C. Vielhauer and R. Steinmetz, "Handwriting: Feature correlation analysis for biometric hashes," *EURASIP Journal on Applied Signal Processing*, vol. 4, pp. 542–558, 2004.

[46] K. Nandakumar and A. K. Jain, "Multibiometric template security using fuzzy vault," in *International Conference on Biometrics: Theory, Applications and Systems*, 2008, pp. 1–6.

[47] E. Maiorana, P. Campisi, J. Fierrez, J. Ortega-Garcia, and A. 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*, pp. 1 – 14, 2010.

[48] J.-P. Linnartz and P. Tuyls, "New shielding functions to enhance privacy and prevent misuse of biometric templates," in *4th Int. Conf. on AVBPA*, 2003, pp. 393 – 402.

[49] A. K. Jain, K. Nandakumar, and A. Nagar, "Biometric template security," *EURASIP Journal on Advances in Signal Processing*, 2008.

[50] E. Barker and A. Roginsky, "DRAFT NIST special publication 800-131, recommendation for the transitioning of cryptographic algorithms and key sizes," http://csrc.nist.gov/publications/drafts/800-131/draft-800-131_transition-paper.pdf, January 2010.

[51] A. B. J. Teoh and D. C. L. Ngo, "Cancellable biometrics featuring with tokenized random number," in *Pattern Recognition Letter*, 2005, pp. 1454–1460.

[52] A. B. J. Teoh, A. Goh, and D. 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*, vol. 28, no. 12, pp. 1892–1901, December 2006.

[53] C. S. Chin, A. T. B. Jin, and D. N. C. Ling, "High security iris verification system based on random secret integration," *Computer Vision and Image Understanding*, vol. 102, no. 2, pp. 169–177, 2006.

[54] T. Connie, A. B. J. Teoh, M. Goh, and D. C. L. Ngo, "Palmhashing: a novel approach for cancelable biometrics," *Information Processing Letters*, vol. 93, no. 1, pp. 1–5, 2005.

[55] T. S. Ong, A. B. J. Teoh, S. E. Khor, and T. Connie, "Reliable template protection technique for biometric authentication," in *IEICE Electronics Express*, vol. 5, no. 8, 2008, pp. 278 – 284.

[56] F. Farooq, R. M. Bolle, J. Tsai-Yang, and N. K. Ratha, "Anonymous and revocable fingerprint recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '07)*, 2007, pp. 1 – 7.

[57] C. Lee and J. Kim, "Cancelable fingerprint templates using minutiae-based bitstrings," *Journal of Network and Computer Applications*, 2010.

[58] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM Systems Journal*, vol. 40, no. 3, pp. 614–634, 2001.

[59] N. K. Ratha, S. Chikkerur, J. H. Connell, and R. M. Bolle, "Generating cancelable fingerprint templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 561 – 572, April 2007.

[60] S. Chikkerur, N. K. Ratha, J. H. Connell, and R. M. Bolle, "Generating registration-free cancelable fingerprint templates," in *IEEE 2nd International Conference on Biometrics: Theory, Applications and Systems (BTAS '08)*, 2008, pp. 1 – 6.

[61] B. Yang, C. Busch, P. Bours, and D. Gafurov, "Robust minutiae hash for fingerprint template protection," in *Proc. SPIE*, vol. 7541, 2010.

[62] J. Bringer, H. Chabanne, and B. Kindarji, "Anonymous identification with cancelable biometrics," in *6th International Symposium on Image and Signal Processing and Analysis (ISPA '09)*, 2009.

[63] J. Zuo, N. K. Ratha, and J. H. Connell, "Cancelable iris biometric," in *19th International Conference on Pattern Recognition (ICPR '08)*, December 2008, pp. 1 – 4.

[64] Y. Dodis, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong secret keys from biometrics and other noisy data," in *Advances in Cryptology - Eurocrypt 2004, LNCS 3027*, 2004, pp. 532 – 540.

[65] E.-C. Chang and S. Roy, "Robust extraction of secret bits from minutiae," in *Int. Conf. on Biometrics*, Seoul, South Korea, August 2007, pp. 750–759.

[66] F. Hao, R. Anderson, and J. Daugman, "Combining crypto with biometrics effectively," *IEEE Transactions on Computers*, vol. 55, no. 9, pp. 1081–1088, September 2006.

[67] J. Bringer, H. Chabanne, G. Cohen, B. Kindarji, and G. Zémor, "Optimal iris fuzzy sketches," in *Biometrics: Theory, Applications, and Systems*, vol. abs/0705.3740, 2007.

[68] L. Zhang, Z. Sun, T. Tan, and S. Hu, "Robust biometric key extraction based on iris cryptosystem," in *Third International Conference on Biometrics, ICB 2009*. Springer-Verlag, 2009, pp. 1060 – 1069.

[69] C. Rathgeb and A. Uhl, "Systematic construction of iris-based fuzzy commitment scheme," in *Third International Conference on Biometrics, ICB 2009*. Springer-Verlag, 2009, pp. 940 – 949.

[70] S. Yang and I. Verbauwhede, "Secure iris verification," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, June 2007, pp. 133–136.

[71] J. Bringer, H. Chabanne, G. Cohen, B. Kindarji, and G. Zemor, "Theoretical and practical boundaries of binary secure sketches," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 4, pp. 673–683, 2008.

[72] M. van der Veen, T. Kevenaar, G.-J. Schrijen, T. H. Akkermans, and F. Zuo, "Face biometrics with renewable templates," in *SPIE Proc on Security, Steganography, and Watermarking of Multimedia Contents*, vol. 6072, San Jose, USA, 2006.

[73] X. Zhou, T. Kevenaar, E. Kelkboom, C. Busch, M. van der Veen, and A. Nouak, "Privacy enhancing technology for a 3D-face recognition system," in *Proc. of BIOSIG 2007: Biometrics and Electronic Signatures*, Darmstadt, Germany, July 2007, pp. 3 – 14.

[74] F. M. Bui, K. Martin, H. Lu, K. N. Plataniotis, and D. Hatzinakos, "Fuzzy key binding strategies based on quantization index modulation (QIM) for biometric encryption (be) applications," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 1, pp. 118 – 132, March 2010.

[75] I. R. Buhan, J. Doumen, P. Hartel, T. Qiang, and R. N. J. Veldhuis, "Embedding renewable cryptographic keys into noisy data," in *10th Int. Conf. on Information and Communications Security (ICICS 2008)*, 2008, pp. 294–310.

[76] A. Juels and M. Sudan, "A fuzzy vault scheme," in *Proc. of the 2002 International Symposium on Information Theory (ISIT 2002)*, Lausanne, 2002, p. 408.

[77] A. Kumar and A. Kumar, "Development of a new cryptographic construct using palmprint based fuzzy vault," *EURASIP Journal on Advances in Signal Processing*, 2009.

[78] Y. J. Lee, K. R. Park, S. J. Lee, K. Bae, and J. Kim, "A new method for generating an invariant iris private key based on the fuzzy vault system," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,*, vol. 38, no. 5, pp. 1302 – 1313, October 2008.

[79] T. C. Clancy, N. Kiyavash, and D. J. Lin, "Secure smartcard-based fingerprint authentication," in *Proc. 2003 ACM SIGMM Workshop Biometrics Methods and Application (WBMA)*, 2003, pp. 45 – 52.

[80] A. Juels and M. Sudan, "A fuzzy vault scheme," *Designs, Codes and Cryptography*, vol. 38, no. 2, pp. 237–257, February 2006.

[81] T. Frassen, Z. Xuebing, and C. Busch, "Fuzzy vault for 3D face recognition systems," in *International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP '08)*, 2008, pp. 1069 –1074.

[82] S. Lee, D. Moon, W. Y. Choi, and Y. Chung;, "Analysis of tradeoffs among verification accuracy, memory consumption, and execution time in the GH-based fuzzy fingerprint vault," in *International Conference on Security Technology (SecTech '08)*, 2008, pp. 75 – 78.

[83] P. Li, X. Yang, K. Cao, P. Shi, and J. Tian, "Security-enhanced fuzzy fingerprint vault based on minutiae's local ridge information," in *Third International Conference on Biometrics, ICB 2009*. Springer-Verlag, 2009, pp. 930 – 939.

[84] K. Nandakumar, A. K. Jain, and S. Pankanti, "Fingerprint-based fuzzy vault: Implementation and performance," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 4, pp. 744 – 757, Decmber 2007.

[85] S. Yang and I. Verbauwhede, "Automatic secure fingerprint verification system based on fuzzy vault scheme," in *Proc. IEEE ICASSP*, vol. 5, Philadelphia, PA, Maart 2005, pp. 609–612.

[86] U. Uludag and A. K. Jain, "Securing fingerprint template: Fuzzy vault with helper data," in *Proc. CVPR Workshop Privacy Research Vision*, New-York, June 2006, pp. 163–171.

[87] G. I. Davida, Y. Frankel, and B. J. Matt, "On enabling secure applications through off-line biometric identification," in *Proc. of the IEEE 1998 Symp. on Security and Privacy*, Oakland, Ca., 1998, pp. 148–157.

[88] A. Nagar, K. Nandakumar, and A. K. Jain, "Securing fingerprint template: Fuzzy vault with minutiae descriptors," in *19th International Conference on Pattern Recognition (ICPR 2008)*, 2008, pp. 1 – 4.

[89] J. Bringer, H. Chabanne, and B. Kindarji, "The best of both worlds: Applying secure sketches to cancelable biometrics," in *Science of Computer Programming*, October 2008.

[90] Q. Li, Y. Sutcu, and N. Memon, "Secure sketch for biometric template," in *Advances in Cryptology - ASIACRYPT*, 2006, pp. 99–113.

[91] Y. Sutcu, Q. Li, and N. Memon, "Protecting biometric templates with sketch: Theory and practice," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 503 – 512, September 2007.

[92] Q. Li, M. Guo, and E.-C. Chang, "Fuzzy extractors for asymmetric biometric representation," in *IEEE Workshop on Biometrics (In association with CVPR)*, 2008, pp. 1 – 6.

[93] A. Arakala, J. Jeffers, and K. J. Horadam, "Fuzzy extractors for minutiae-based fingerprint authentication," in *International Conference on Biometrics*, Seoul, South Korea, 2007, pp. 760–769.

[94] Y. Sutcu, Q. Li, and N. Memon., "Design and analysis of fuzzy extractors for faces," in *Biometric Technology for Human Identification, part of the SPIE International Defense and Security Symposium*, Orlando, USA, 13-17 April 2009.

[95] E. J. C. Kelkboom, G. Garcia Molina, J. Breebaart, R. N. J. Veldhuis, T. A. M. Kevenaar, and W. Jonker, "Binary biometrics: An analytic framework to estimate the performance curves under gaussian assumption," *IEEE Transactions on Systems, Man and Cybernetics Part A, Special Issue on Advances in Biometrics: Theory, Applications and Systems*, vol. 40, no. 3, pp. 555–571, May 2010.

[96] E. J. C. Kelkboom, R. N. J. Veldhuis, and J. Breebaart, "Classification performance comparison of a continuous and binary classifier under gaussian assumption," in *The 31st Symposium on Information Theory in the Benelux*, May 2010, pp. 129 – 136.

[97] E. J. C. Kelkboom, G. Garcia Molina, T. A. M. Kevenaar, R. N. J. Veldhuis, and W. Jonker, "Binary biometrics: An analytic framework to estimate the bit error probability under gaussian assumption," in *2nd IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS '08)*, September 2008, pp. 1–6.

[98] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 22, no. 1, pp. 4–37, January 2000.

[99] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *IEEE CVPR*, vol. 2, June 2005, pp. 454–461.

[100] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, "FVC2000: fingerprint verification competition," *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, vol. 24, no. 3, pp. 402–412, 2002.

[101] C. Chen, R. N. J. Veldhuis, T. A. M. Kevenaar, and A. H. M. Akkermans, "Multi-bits biometric string generation based on the likelihood ratio," in *IEEE Conf. on Biometrics: Theory, Applications and Systems*, Washington DC, September 2007.

[102] J. Breebaart, C. Busch, J. Grave, and E. Kindt, "A reference architecture for biometric template protection based on pseudo identities," in *BIOSIG*, Darmstadt, Germany, September 2008.

[103] N. Delvaux, P. Lindeberg, J. Midgren, J. Breebaart, T. Akkermans, M. van der Veen, R. N. J. Veldhuis, E. Kindt, K. Simoens, C. Busch, P. Bours, D. Gafurov, B. Yang, J. Stern, C. Rust, B. Cucinelli, and D. Skepastianos, "Pseudo idenitities based on fingerprint characteristics," in *IEEE 4th international conference on intelligent information hiding and multimedia signal processing (IIH-MSP)*, 2008.

[104] J. Daugman, "The importance of being random: statistical principles of iris recognition," *Pattern Recognition*, vol. 36, no. 2, pp. 279–291, 2003.

[105] A. P. Prudnikov, Y. A. Brychkov, and O. I. Marichev, *Integrals and Series, Vol. 2: Special Functions*. New York: Gordon and Breach, 1990.

[106] B. Gökberk, M. O. Irfanoglu, and L. Akarun, "3D shape-based face representation and feature extraction for face recognition," *Image and Vision Computing*, vol. 24, no. 8, pp. 857–869, August 2006.

[107] A. M. Bazen and R. N. J. Veldhuis, "Likelihood-ratio-based biometric verification," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 14, no. 1, pp. 86–94, 2004.

[108] C. Chatfield, *Statistics for Technology: A course in Applied Statistics*, third edition ed. Chapman & Hall, 1983.

[109] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, 2nd ed. Addison Wiley, 1994.

[110] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, "A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation," in *Proceedings of ICSLP-98*, 1998.

[111] J. Breebaart, A. H. M. Akkermans, and E. J. C. Kelkboom, "Inter-subject differences in false non-match rates for a fingerprint-based authentication system," *J. Advances in Signal Processing*, January 2009, article ID 896383.

[112] N. Yager and T. Dunstone, "Worms, chameleons, phantoms and doves: New additions to the biometric menagerie," in *AutoID 2007*, 2007.

[113] NRC, "Fingerprints in passports can't be used by the police - yet ," 18 September 2009. [Online]. Available: http://www.nrc.nl/international/Features/article2363938.ece/Fingerprints_in_passports_cant_be_used_by_the_police_-_yet

[114] E. J. C. Kelkboom, J. Breebaart, I. R. Buhan, and R. N. J. Veldhuis, "Analytical template protection performance and maximum key size given a Gaussian modeled biometric source," *Submitted to IEEE Transactions on Information Forensics and Security*, 2010.

[115] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," *SIAM Journal on Computing*, vol. 38, no. 1, pp. 97 – 139, 2008.

[116] U. Korte and R. Plaga, "Cryptographic protection of biometric templates: Chance, challenges and applications," in *BIOSIG 2007: Biometrics and Electronic Signatures*, 2007, pp. 33 – 45.

[117] J. Kittler, J. Matas, K. Jonsson, and M. U. R. Snchez, "Combining evidence in personal identity verification systems," *Pattern Recognition Letters*, vol. 18, pp. 845 – 852, 1997.

[118] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn, "Using multi-instance enrollment to improve performance of 3D face recognition," *Computer Vision and Image Understanding*, vol. 112, no. 2, pp. 114 – 125, November 2008.

[119] K. Nandakumar, A. Nagar, and A. K. Jain., "Hardening fingerprint fuzzy vault using password." in *Proceedings of Second International Conference on Biometrics*, Seoul, South Korea, August 2007, pp. 927 – 937.

[120] Y. Sutcu, S. Rane, J. S. Yedidia, S. C. Draper, and A. Vetro, "Feature extraction for a Slepian-Wolf biometric system using LDPC codes," in *IEEE International Symposium on Information Theory, 2008. ISIT 2008.*, 2008, pp. 2297–2301.

[121] X. Zhou, "Template protection and its implementation in 3D face recognition systems," in *Proceedings of SPIE 07, Biometric Technology for Human Identification IV*, 2007.

[122] T. Ignatenko and F. M. J. Willems, "Biometric systems: Privacy and secrecy aspects," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 4, pp. 956 – 973, December 2009.

[123] ——, "Privacy leakage in biometric secrecy systems," in *46th Annual Allerton Conference on Communication, Control, and Computing*, 2008, pp. 850 – 857.

[124] L. Lai, S.-W. Ho, and H. V. Poor, "Privacy-security tradeoffs in biometric security systems," in *Proceedings of Forty-Sixth Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, September 2008.

[125] E. J. C. Kelkboom, X. Zhou, J. Breebaart, R. N. J. Veldhuis, and C. Busch, "Multi-sample fusion with template protection," in *Proc. of BIOSIG 2008: Biometrics and Electronic Signatures*, Darmstadt, Germany, September 2009, pp. 55 – 67.

[126] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.

[127] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North-Holland, 1977.

[128] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J. J. Igarza, C. Vivaracho, D. Escudero, and Q. I. Moro, "MCYT baseline corpus: A bimodal biometric database," in *IEE Proc. Vision, Image and Signal Processing, Special Issue on Biometrics on the Internet*, December 2003, pp. 395 – 401.

[129] M. van der Veen, A. Bazen, T. Ignatenko, and T. Kalker, "Reference point detection for improved fingerprint matching," in *Proceedings of SPIE*, 2006, p. 60720G.160720G.9.

[130] S. H. Gerez and A. M. Bazen, "Systematic methods for the computation of the directional fields and singular points of fingerprints," in *IEEE Transactions on pattern analysis and machine intellignece*, July 2002, pp. 905 – 919.

[131] E. J. C. Kelkboom, J. Breebaart, T. A. M. Kevenaar, I. R. Buhan, and R. N. J. Veldhuis, "Preventing the decodability attack based cross-matching in a fuzzy commitment scheme," *Submitted to IEEE Transactions on Information Forensics and Security*, 2010.

[132] W. J. Scheirer and T. E. Boult, "Cracking fuzzy vaults and biometric encryption," in *Proceedings of Biometric Symposium*, Baltimore, MD, September 2007.

[133] E. J. C. Kelkboom, K. T. J. de Groot, C. Chen, J. Breebaart, and R. N. J. Veldhuis, "Pitfall of the detection rate optimized bit allocation within template protection and a remedy," in *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, 2009. (BTAS '09)*, 2009, pp. 1 – 8.

[134] E. J. C. Kelkboom, J. Breebaart, and R. N. J. Veldhuis, "Analysis of the system and cross-matching performance of bit extraction schemes with template protection," *Submitted to EURASIP Journal on Advances in Signal Processing*, 2010.

[135] C. Vielhauer, R. Steinmetz, and A. Mayerhofer, "Biometric hash based on statistical features of online signatures," in *Proceedings of the Sixteenth International Conference on Pattern Recognition*, vol. 1, 2002, pp. 123–126.

[136] F. Hao and C. W. Chan, "Private key generation from on-line handwritten signatures," in *Information Management & Computer Security*, vol. 10, no. 4, 2002, pp. 159–164.

[137] M. Gardner, *Knotted Doughnuts and Other Mathematical Entertainments*. W.H. Freeman & Company, 1986.

[138] N. K. Ratha, J. H. Connell, and R. M. Bolle, "An analysis of minutiae matching strength," in *Proceedings of the 3rd International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA 01)*, Halmstad, Sweden, June 2001, pp. 223 – 228.

[139] B. Park, S. Hong, J. Oh, H. Lee, and Y. Won, "One touch login: Replacing multiple passwords with single fingerprint recognition," in *The Sixth IEEE International Conference on Computer and Information Technology (CIT '06)*, 2006.

[140] E. J. C. Kelkboom, X. Zhou, J. Breebaart, R. N. J. Veldhuis, and C. Busch, "Multi-algorithm fusion with template protection," in *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems (BTAS '09)*, Washington DC, U.S.A., September 2009, pp. 1 – 8.

[141] K. P. Hollingsworth, K. W. Bowyer, and P. J. Flynn, "Using fragile bit coincidence to improve iris recognition," in *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems (BTAS '09)*, September 2009, pp. 1 – 6.

[142] ——, "The best bits in an iris code," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 6, pp. 964 – 973, June 2009.

[143] N. Barzegar and M. S. Moin, "A new user dependent iris recognition system based on an area preserving pointwise level set segmentation approach." in *EURASIP Journal on Advances in Signal Processing*, 2009.

[144] G. Dozier, K. Frederiksen, R. Meeks, M. Savvides, K. Bryant, D. Hopes, and T. Munemoto, "Minimizing the number of bits needed for iris recognition via bit inconsistency and grit." in *In Proc. IEEE Workshop on Computational Intelligence in Biometrics: Theory, Algorithms, and Applications*, April 2009.

[145] R. C. Bose and D. K. Ray-Chaudhuri, "On a class of error correcting binary group codes," *Information and Control*, vol. 3, no. 1, pp. 68 – 79, March 1960.

[146] P. Tuyls and J. Goseling, "Capacity and examples of template-protecting biometric authentication systems," in *Biometric Authentication Workshop ECCV*, 2004, pp. 158 – 170.

[147] X. Zhou, H. Seibert, C. Busch, and W. Funk, "A 3D face recognition algorithm using histogram-based features," in *Eurographics 2008 Workshop on 3D Object Retrieval*, Crete, Greece, April 2008, pp. 65–71.

[148] Q. Tao and R. N. J. Veldhuis, "Threshold-optimized decision-level fusion and its application to biometrics," *Pattern Recognition*, vol. 4, no. 5, pp. 823 – 836, May 2009.

Curriculum Vitae

Emile J. C. Kelkboom was born in Oranjestad, Aruba, in 1980. He received his M.Sc. degree in Electrical Engineering from the Delft University of Technology, the Netherlands in June 2004. From October 2004 to July 2006, he worked as an Application Engineer on CD, DVD and Blu-ray drives within the Storage Engines department of Philips Semiconductors. Since August 2006, he is pursuing his Ph.D. Degree at Philips Research and the Department of Electrical Engineering Mathematics and Computer Science of the University of Twente, the Netherlands. His focus is on safeguarding the privacy of the biometric information of subjects within biometric systems, namely the field of template protection. He won the European Biometrics Forum (EBF) Research Award among Ph.D. students in Europe in 2009. His research interests include biometrics, pattern recognition, signal processing, and security.

List of Publications

2010

- E. J. C. Kelkboom, J. Breebaart, I. R. Buhan, and R. N. J. Veldhuis, "Analytical template protection performance and maximum key size given a Gaussian modeled biometric source," *Submitted to IEEE Transactions on Information Forensics and Security*, 2010.

- E. J. C. Kelkboom, J. Breebaart, and R. N. J. Veldhuis, "Analysis of the system and cross-matching performance of bit extraction schemes with template protection," *Submitted to EURASIP Journal on Advances in Signal Processing*, 2010.

- E. J. C. Kelkboom, J. Breebaart, T. A. M. Kevenaar, I. R. Buhan, and R. N. J. Veldhuis, "Preventing the decodability attack based cross-matching in a fuzzy commitment scheme," *Submitted to IEEE Transactions on Information Forensics and Security*, 2010.

- J. Breebaart, I. R. Buhan, K. de Groot, and E. J. C. Kelkboom, "On the integration of biometrics in pin-based payment systems," *Submitted to Electronic Commerce Research and Application*, 2010.
- I. R. Buhan, J. G. Merchan, and E. J. C. Kelkboom, "Efficient strategies for playing the indistinguishability game for fuzzy sketches," in *Accepted at IEEE Workshop on Information Forensics and Security (WIFS 2010)*, 2010.
- I. R. Buhan, E. J. C. Kelkboom, and K. Simoens, "A survey of the security and privacy measures for anonymous biometric authentication systems," in *Accepted at Proceedings of BIOSIG 2010: Biometrics and Electronic Signatures*, 2010.
- E. J. C. Kelkboom, J. Breebaart, I. R. Buhan, and R. N. J. Veldhuis, "Analytical template protection performance and maximum key size given a gaussian modeled biometric source," in *Proc. SPIE, Vol. 7667, Biometric Technology for Human Identification VII*, 2010.
- S. Chindaro, F. Deravi, Z. Zhou, M. W. R. Ng, M. C. Neves, X. Zhou, and E. Kelkboom, "A multibiometric face recognition fusion framework with template protection," in *Proc. SPIE, Vol. 7667, Biometric Technology for Human Identification VII*, 2010.
- E. J. C. Kelkboom, R. N. J. Veldhuis, and J. Breebaart, "Classification performance comparison of a continuous and binary classifier under gaussian assumption," in *The 31st Symposium on Information Theory in the Benelux*, May 2010, pp. 129 - 136.
- E. J. C. Kelkboom, G. Garcia Molina, J. Breebaart, R. N. J. Veldhuis, T. A. M. Kevenaar, and W. Jonker, "Binary biometrics: An analytic framework to estimate the performance curves under gaussian assumption," *IEEE Transactions on Systems, Man and Cybernetics Part A, Special Issue on Advances in Biometrics: Theory, Applications and Systems*, vol. 40, no. 3, pp. 555 - 571, May 2010.

2009

- E. J. C. Kelkboom, X. Zhou, J. Breebaart, R. N. J. Veldhuis, and C. Busch, "Multi-algorithm fusion with template protection," in *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems (BTAS 09)*, Washington DC, U.S.A., September 2009, pp. 1 - 8.
- E. J. C. Kelkboom, K. T. J. de Groot, C. Chen, J. Breebaart, and R. N. J. Veldhuis, "Pitfall of the detection rate optimized bit allocation within template protection and a remedy," in *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, 2009. (BTAS 09)*, 2009, pp. 1 - 8.
- E. J. C. Kelkboom, X. Zhou, J. Breebaart, R. N. J. Veldhuis, and C. Busch, "Multi-sample fusion with template protection," in *Proc. of BIOSIG 2009: Biometrics and Electronic Signatures*, Darmstadt, Germany, 2009, pp. 55 - 67.

- I. R. Buhan, J. Breebaart, J. Guajardo Merchan, K. T. J. de Groot, E. Kelkboom, and T. Akkermans, "A quantative analysis of indistinguishability for a continuous domain biometric cryptosystem," in *Proc. 4th international workshop on data privacy management (DPM 09)*, ser. Lecture Notes in Computer Science, vol. 5939. St. Malo, France: Springer-Verlag, 2009, pp. 78 - 92.

- J. Breebaart, A. H. M. Akkermans, and E. J. C. Kelkboom, "Inter-subject differences in false non-match rates for a fingerprint-based authentication system," *J. Advances in Signal Processing*, January 2009, article ID 896383.

2008

- E. J. C. Kelkboom, G. Garcia Molina, T. A. M. Kevenaar, R. N. J. Veldhuis, and W. Jonker, "Binary biometrics: An analytic framework to estimate the bit error probability under gaussian assumption," in *2nd IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS 08)*, 2008, pp. 1 - 6.

- C. Busch, A. Nouak, X. Zhou, J.-M. Suchier, E. Kelkboom, and T. Kevenaar, "3D face recognition for unattended border control," in *Proceedings Net-ID 2008 Conference*, 2008.

2007

- X. Zhou, T. Kevenaar, E. Kelkboom, C. Busch, M. van der Veen, and A. Nouak, "Privacy enhancing technology for a 3D-face recognition system," in Proceedings of BIOSIG 2007: Biometrics and Electronic Signatures, 2007, pp. 3 - 14.

- 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 Intl. Conf. on Biometrics, Seoul, South Korea, August 2007, pp. 566 - 573.