
Biometric Security from an Information-Theoretical Perspective

Biometric Security from an Information-Theoretical Perspective

Tanya Ignatenko

*Eindhoven University of Technology
Eindhoven
The Netherlands
t.ignatenko@ieee.org*

Frans M.J. Willems

*Eindhoven University of Technology
Eindhoven
The Netherlands
f.m.j.willems@tue.nl*

now
the essence of knowledge
Boston – Delft

Foundations and Trends[®] in Communications and Information Theory

Published, sold and distributed by:

now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:

now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is T. Ignatenko and F. M. J. Willems, Biometric Security from an Information-Theoretical Perspective, Foundations and Trends[®] in Communications and Information Theory, vol 7, nos 2–3, pp 135–316, 2010

ISBN: 978-1-60198-522-4

© 2012 T. Ignatenko and F. M. J. Willems

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1-781-871-0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

**Foundations and Trends[®] in
Communications and Information Theory**
Volume 7 Issues 2–3, 2010
Editorial Board

Editor-in-Chief:

Sergio Verdú

*Depart of Electrical Engineering
Princeton University
Princeton, New Jersey 08544*

Editors

Venkat Anantharam (UC. Berkeley)	Amos Lapidoth (ETH Zurich)
Ezio Biglieri (U. Torino)	Bob McEliece (Caltech)
Giuseppe Caire (U. Southern California)	Neri Merhav (Technion)
Roger Cheng (U. Hong Kong)	David Neuhoff (U. Michigan)
K.C. Chen (Taipei)	Alon Orlitsky (UC. San Diego)
Daniel Costello (U. Notre Dame)	Vincent Poor (Princeton)
Thomas Cover (Stanford)	Kannan Ramchandran (UC. Berkeley)
Anthony Ephremides (U. Maryland)	Bixio Rimoldi (EPFL)
Andrea Goldsmith (Stanford)	Shlomo Shamai (Technion)
Dave Forney (MIT)	Amin Shokrollahi (EPFL)
Georgios Giannakis (U. Minnesota)	Gadiel Seroussi (MSRI)
Joachim Hagenauer (TU Munich)	Wojciech Szpankowski (Purdue)
Te Sun Han (Tokyo)	Vahid Tarokh (Harvard)
Babak Hassibi (Caltech)	David Tse (UC. Berkeley)
Michael Honig (Northwestern)	Ruediger Urbanke (EPFL)
Johannes Huber (Erlangen)	Steve Wicker (Cornell)
Hideki Imai (Tokyo)	Raymond Yeung (Hong Kong)
Rodney Kennedy (Canberra)	Bin Yu (UC. Berkeley)
Sanjeev Kulkarni (Princeton)	

Editorial Scope

Foundations and Trends[®] in Communications and Information Theory will publish survey and tutorial articles in the following topics:

- Coded modulation
- Coding theory and practice
- Communication complexity
- Communication system design
- Cryptology and data security
- Data compression
- Data networks
- Demodulation and Equalization
- Denoising
- Detection and estimation
- Information theory and statistics
- Information theory and computer science
- Joint source/channel coding
- Modulation and signal design
- Multiuser detection
- Multiuser information theory
- Optical communication channels
- Pattern recognition and learning
- Quantization
- Quantum information processing
- Rate-distortion theory
- Shannon theory
- Signal processing for communications
- Source coding
- Storage and recording codes
- Speech and Image Compression
- Wireless Communications

Information for Librarians

Foundations and Trends[®] in Communications and Information Theory, 2010, Volume 7, 6 issues. ISSN paper version 1567-2190. ISSN online version 1567-2328. Also available as a combined paper and online subscription.

Biometric Security from an Information-Theoretical Perspective

Tanya Ignatenko¹ and Frans M.J. Willems²

¹ *Eindhoven University of Technology, Den Dolech 2, Eindhoven, The Netherlands, t.ignatenko@ieee.org*

² *Eindhoven University of Technology, Den Dolech 2, Eindhoven, The Netherlands, f.m.j.willems@tue.nl*

Abstract

In this review, biometric systems are studied from an information theoretical point of view. In the first part biometric authentication systems are studied. The objective of these systems is, observing correlated enrollment and authentication biometric sequences, to generate or convey as large as possible secret keys by interchanging a public message, while minimizing privacy leakage. Here privacy leakage is defined as the amount of information that this public message contains about the biometric enrollment sequence. In this setting also the secrecy leakage, that is, the amount of information the public message leaks about the secret key, should be negligible. Next identification biometric systems are investigated. These systems should be able to identify as many individuals as possible while being able to assign as large as possible secret keys to each individual and again minimize the privacy leakage. To realize these systems public reference data are stored in the database.

Leakage is defined with respect to these reference data. For all these biometric systems fundamental limits are determined in the current work. Finally, a popular practical construction for biometric systems, fuzzy commitment, is analyzed with respect to both its theoretical performance and performance related to the code choice.

Contents

1	Introduction	1
1.1	Access Control Systems and Biometrics	1
1.2	From Traditional Biometric Systems to Biometric Secrecy Systems	3
1.3	Modeling Biometric Data	12
1.4	Related Work	12
1.5	Organization of This Review	15
2	Secret Sharing and Biometric Systems	17
2.1	Introduction	17
2.2	Model for Biometric Generated-Secret Systems	18
2.3	Statement of Result	20
2.4	Proof of Theorem 2.1	20
2.5	Stationary Ergodic Case	24
2.6	From Secret-Key Generation to Biometric Authentication	26
2.7	Privacy Leakage for Codes Achieving the Maximum Secret-Key Rate	30
2.8	Chosen Secrets, Masking Layer	31
2.9	Conclusions	33
3	Biometric Authentication. Discrete Biometric Sources	35
3.1	Introduction	35

3.2	Discrete Biometric Source and Setting for Biometric Models	39
3.3	Two Basic Biometric Systems: Generated-Secret and Chosen-Secret Systems	41
3.4	Generated-Secret and Chosen-Secret Systems, and Conditional Privacy Leakage	48
3.5	Systems with Zero-Leakage	53
3.6	Achievability of Special Points	60
3.7	Conclusions and Remarks	61
4	Biometric Authentication. Gaussian Biometric Sources	65
4.1	Gaussian Biometric Source	65
4.2	Results for the Gaussian Case	68
4.3	Properties of the Achievable Region	68
4.4	Conclusions	70
5	Biometric Identification	73
5.1	Introduction	73
5.2	Biometric Source and Biometric Channel	75
5.3	Identification with Secret Generation	76
5.4	Identification with Secret-Binding	78
5.5	Statement of Results	79
5.6	Connection to Other Results	82
5.7	Conclusions	85
6	Practical Constructions. Fuzzy Commitment and Its Properties	87
6.1	Introduction	87
6.2	Fuzzy Commitment	88
6.3	Achievable Region for Fuzzy Commitment	91
6.4	The Totally-Symmetric Memoryless Case	93
6.5	The Input-Symmetric Memoryless Case	96
6.6	The Memoryless Case	99

6.7	The Stationary Ergodic Case	101
6.8	Tighter Bounds with Systematic Parity-Check Codes	103
6.9	Conclusions	105
7	From Gaussian to Binary: Quantization Effects. Case Study for Fuzzy Commitment	107
7.1	Introduction	107
7.2	Binary Symmetric Biometric Systems	108
7.3	Binary Quantization	110
7.4	Coding for Fuzzy Commitment	112
7.5	Conclusions	116
8	Conclusions and Future Directions	117
8.1	Conclusions	117
8.2	Future Directions	119
	Glossary	121
A	Some Results of Information Theory	123
A.1	Typical Sets	123
A.2	Modified Typical Sets	124
A.3	Mrs. Gerber's Lemma	126
B	Proof of the Results from Section 3	127
B.1	Proof of Theorem 3.1	127
B.2	Proof of Theorem 3.2	135
B.3	Proof of Theorem 3.3	138
B.4	Proof of Theorem 3.4	139
B.5	Proof of Theorem 3.5	140
B.6	Proof of Theorem 3.6	142
B.7	Proof of Theorem 3.7	143
B.8	Proof of Theorem 3.8	145
B.9	Bound on the Cardinality of U	146

C Proof of the Results from Section 4	147
C.1 Proof of Theorem 4.1	147
C.2 Proof of Theorem 4.2	151
D Proof of the Results from Section 5	153
D.1 Proof of Theorem 5.1	153
D.2 Proof of Theorem 5.2	162
E Proof of the Results from Section 6	165
E.1 Proof of Theorem 6.1	165
E.2 Proof of Theorem 6.5	168
E.3 Proof of Theorem 6.7	170
E.4 Proof of Lemma 6.9	173
E.5 Proof of Theorem 6.10	174
Acknowledgments	177
References	179

1

Introduction

1.1 Access Control Systems and Biometrics

Nowadays people live in the era of large-scale computer networks connecting huge numbers of electronic devices. These devices execute applications that use networks for exchanging information. Sometimes the information that is transmitted within these networks and stored by the devices is sensitive to misuse. Moreover, the networks and devices cannot always be trusted. This can lead to intrusions into the privacy of users by, for example, hackers, commercial parties, or even by governmental institutions. Also illegal copying of copyrighted content, illegal use of e-payment systems, and identity theft can be foreseen. In order to prevent all such malicious actions the security of networks and devices should be adequate.

Traditional systems for access control, which are based on the possession of secret knowledge (passwords, secret keys, etc.) or on a physical token (ID card, smart-card, etc.), have the drawback that they cannot guarantee that it is the legitimate user who, for example, enters

2 Introduction

a password or presents a smart-card. Moreover, passwords can often be guessed, since people tend to use passwords which are easy to remember. Physical tokens in their turn can be lost, stolen, or copied.

Biometric systems offer a solution to most of the problems mentioned above. They could be either substituted for traditional systems or used to reinforce them. Biometric systems are based on physical or behavioral characteristics of human beings, like faces, fingerprints, voice, irises, gait, see Jain et al. [39]. The results of the measurement of these characteristics are called biometric data. Biometric data have the advantage that potentially they are unique identifiers of human beings, as was argued by Clarke [13]. They provide therefore a closer bond with the identity of their owner than a password or a token does. Moreover, biometric data cannot be stolen or lost. They potentially contain a large amount of information and therefore are hard to guess. All this makes biometrics a good candidate for substitution of traditional passwords and secret keys. A drawback of using biometrics is that the outcome of their measurements is, in general, noisy due to intrinsic variability, varying measurement conditions, or due to the use of different hardware. However, advanced signal-processing and error-correcting techniques can be applied to guarantee reliable overall behavior.

The attractive property of uniqueness, that holds for biometrics, also results in its major weakness. Unlike passwords and secret keys, biometric information, if compromised once, cannot be canceled and easily replaced by other biometric information, since people only have limited resources of biometric data. Theft of biometric data results in a partially stolen identity, and this is, in principle, irreversible. Therefore requirements for biometric systems should include *secure storage* and *secure communication* of biometric data in the applications where they are used.

Although biometric data may provide solutions to the problems discussed above, there are situations when they cannot be used. There is, for example, a small percentage of people whose fingerprints cannot be used due to intrinsic bad quality, see Dorizzi [25]. Also DNA recognition fails for identical twins. In such situations combination with standard cryptographic tools is needed to provide additional security.

1.2 From Traditional Biometric Systems to Biometric Secrecy Systems

1.2.1 Traditional Biometric Systems

The terms “Biometrics” and “Biometry” have been used since the first part of the 20th century to refer to the field of development of statistical and mathematical methods applicable to data analysis problems in biological sciences [8]. Relatively recently the term “Biometrics” has also been used to refer to the field of technology devoted to automatic identification of individuals using biological traits, such as those based on retinal or iris scans, fingerprints, faces, signatures, etc. Such biological traits are unique for individuals as noted in Jain et al. [39].

Traditionally, biometric recognition was used in forensic applications and performed by human experts. However, recent advantages in automated recognition resulted in the spreading of biometric applications, now ranging from border control at airports to access control in Walt Disney amusement parks (see Wayman et al. [83]).

A typical biometric system is essentially a pattern-recognition system, which performs one or more identity checks based on specific physiological or behavioral characteristics possessed by individuals. There are two different ways to resolve an individual’s identity, that is, authentication and identification. Authentication (Am I who I claim I to be?) involves confirming or denying the individual’s claimed identity. In identification, one has to establish the individual’s identity (Who am I?). Each of these approaches has its own characteristics and could probably be solved best by biometric systems.

All biometric technology systems have certain aspects in common. All are dependent upon an accurate reference or enrollment data. If a biometric system is to identify or to authenticate an individual, it first must have these reference data positively linked to the subject. Modern biometric identification systems, based on digital technologies, analyze personal physical attributes at the time of enrollment and distill from them a series of numbers. Once this reference sample or template is in the system, future attempts to identify an individual rest on comparing “live” data to the reference data.

4 Introduction

A perfect system would always recognize an individual, and always reject an impostor. However, biometric data are gathered from individuals under environmental conditions that cannot always be controlled, over equipment that may slowly be wearing out, and using technologies and methods that vary in their level of precision. Consequently, an ideal behavior of biometric systems cannot be realized in practice. Traditionally, the probability that an authorized individual is rejected by a biometric system is called False Rejection Rate (FRR), and the probability that an unauthorized individual is accepted by a biometric system is called False Acceptance Rate (FAR). There are also other performance measures that characterize biometric systems. For a complete overview and similar issues see Jain et al. [39], Maltoni et al. [50], or Wayman et al. [83].

Although biometric technologies have their advantages when they are applied in access control systems, privacy aspects of biometric data should not be ignored. Identification and authentication require storage of biometric reference data in some way. However, people feel uncomfortable with supplying their biometric information to a huge number of seemingly secure databases for various reasons, since

- practice shows that one cannot fully trust implementations of secure algorithms by third parties. Even governmental organizations that are typically trusted by the majority of the population cannot always guarantee that important sensitive data are securely stored;
- databases might be attacked from inside, which allows an owner of a database to abuse biometric information, for example, by selling it to third parties;
- people have limited resources of biometric data, that can be conveniently used for access control. Therefore an “identity theft” of biometric information has much more serious implications than a “simple” theft of a credit card. In the latter case, one can simply block and replace this credit card, while biometric information cannot be easily revoked and replaced by other biometric information.

It is often argued that privacy need not be a real issue in biometric systems, since biometric data are not secret and can easily be captured (faces, irises) or left in public (fingerprints), see Schneier [64]. However, this information, unlike the reference data, is typically of low quality and therefore cannot be easily used for impersonation. Even if it was of good quality, which might be the case with faces, connecting it to the corresponding database is not always an easy task.

Another important point is, that obtaining biometric data of a specific person as well as any other secret information belonging to him, is always possible when sufficient effort is exerted. In contrast, compromising a database, requires a comparable effort, but then provides immediate access to the biometric data of large number of individuals. Therefore, it makes sense to concentrate on protecting the database. It would be ideal if, in case the database becomes public, the biometric reference data could not be recovered.

1.2.2 Types of Security

To assess cryptographic protocols, two notions of security are commonly used, that is, information-theoretical security and computational security.

Computationally secure protocols rely on such an assumption as hardness of mathematical problems, for example, factoring and taking discrete logarithms, and assume that an adversary has bounded computing power. However, hardness of a problem is sometimes difficult to prove, and in practice certain problems are “assumed” to be hard.

Protocols whose security does not rely on computational assumptions, that is, they are secure even when the adversary has unbounded computing power, are called unconditionally or information-theoretically secure. Information-theoretically secure protocols are more desirable, but not always achievable. Therefore, in practice, cryptographers mostly use computational security.

In this review, we will treat security from an information-theoretical point of view. The key mathematical concept on which information theory is built and which is also relevant for considering information-theoretical security, is entropy. The notion of entropy comes from

Shannon [67]. Entropy is a measure of the information contained in a random variable. Although there are a number of alternative entropy concepts, for example, Rényi and min-entropy (Rényi entropy of order 2) [62], and smooth Rényi entropy [61], we will only use the classical (Shannon) notion of entropy here. Another Shannon-type concept is that of mutual information. Mutual information measures by how much the entropy of the first random variable decreases if access to the second random variable is obtained, and this notion can be defined in terms of entropies. For the exact definitions, properties and their proofs of entropy and mutual information we refer to Shannon [67] or, for example, Cover and Thomas [15].

An interesting special case of information-theoretical security is perfect security. This concept was introduced by Shannon [68]. He defined a secrecy system to be perfect if the mutual information between plaintext M and ciphertext C satisfies

$$I(M; C) = 0, \quad (1.1)$$

i.e., if a ciphertext C , which is a function of a plaintext M and a secret key S , provides no information about the plaintext M , in other words, if C and M are statistically independent. Shannon proved that perfect secrecy can only be achieved when the key-entropy and plaintext-entropy satisfy

$$H(S) \geq H(M). \quad (1.2)$$

An example of a perfectly secure system is the one-time pad system, also referred to as the Vernam cipher [80]. In one-time pad, a binary plaintext is concealed by adding modulo-2 (XOR-ing) a random binary secret key.

In practice it is quite possible and common for a secrecy system to leak some information. Although such a system is not perfectly secure, it can be information-theoretically secure up to a certain level.

1.2.3 Biometric Secrecy Systems with Helper Data

A perfect system for a secure biometric access control has to satisfy three requirements. Biometric data have to be private, namely, the reference information stored in a database should not reveal the actual

biometric data. Reference data that are communicated from a database to a point where access can be granted have to be resilient to eavesdropping. Reference data stored in a database have to be resilient to guessing, that is, to brute-force attacks.

Consider a biometric authentication system. A simple naive approach to satisfy both the first and the second requirements would be to use the biometric data as a password in a UNIX-password authentication scheme. In such a scheme, a user possesses a password x that gives access to his account. There is a trusted server that stores some information $y = f(x)$ about the password. The user gains access to the account only if he enters the password x' , such that $f(x') = y$. The scheme has the requirement that nobody can figure out the password x from y in any way other than by guessing. To fulfill this requirement, a UNIX-password scheme relies on one-way functions. A one-way function $f(\cdot)$ is a function that is easy to compute but “hard to invert,” where “hard to invert” refers to the property that no probabilistic polynomial-time algorithm can compute a pre-image of $f(x)$ with a better than negligible probability when x is chosen at random.

Thus, if we would use the UNIX-password authentication scheme and apply a one-way function to the biometric data, the storage of biometric data in the clear would be circumvented. However, there are a number of problems that would arise if we use biometric data in the UNIX scheme. First, the security properties that are guaranteed by one-way functions rely on the assumption that x is truly uniform, while we know that biometric data are far from uniform, although they do contain randomness of course. Moreover, one-way functions, as all cryptographic primitives, require their entries to be exactly reproducible for positive authentication,¹ while biometric data measurements are almost never identical. Therefore additional processing (e.g., error-correction and compression) is needed to realize a biometric UNIX-like authentication scheme that can tolerate a reasonable amount of errors in biometric measurements and results in uniform entries to the one-way function. One way of operating would be to use a collection of error-correcting

¹ Positive authentication can also be a result of an entry that produces a collision. However, here we do not consider collisions, since this is a problem associated with the design of one-way functions and therefore beyond the scope of this review.

codes such that for each observed biometric enrollment template there is a code that contains this template as a codeword. The index to this code is then stored in the database as helper data. Upon observing the individual for a second time, the helper can then be used to retrieve the enrollment template from the authentication template. The error-correcting code should be strong enough to correct the errors between the enrollment and authentication templates. From this, we may conclude that error-correcting techniques and helper data can be applied to combat errors. Subsequently compression methods can be used to achieve almost uniform entries.

Now that we have argued that helper data could be used to create a reliable system, the question arises what requirements ideal helper data should satisfy. Since helper data need to be stored (and communicated) for authentication, it would be advantageous if they could be made publicly available without compromising or leaking any information about the data that are used to get access to the system. We say that secrecy leakage from the helper data has to be negligible. Note that these data could be obtained using a one-way function as in the UNIX-scheme, but better procedures may exist as well. On the other hand, the helper data should leak as little information as possible about the observed biometric enrollment template. This would reduce privacy-related problems. Note that it might be impossible to make this leakage negligible, since helper data should contain some information about the biometric data in order to set up a reliable system. It will become clear later in this review that a notion of secret-key sharing originated from Information Theory, see Ahlswede and Csiszár [2], will be essential in designing and analyzing biometric systems in which public helper data are used. For these secret-key sharing systems, the problem of maximizing the size of the extracted secrets (the data needed to get access) was solved. This provides the solution for our third requirement, resilience to guessing.

In what we have discussed up to now, we have always assumed that keys were obtained as a result of a one-way operation on a password or on a biometric template. A biometric system would however be more flexible if we could choose the keys ourselves. We will show that the helper-data construction will make this possible. Therefore in this

review, we will distinguish between generated-key systems and chosen-key systems. Sometimes their performance will not differ that much, but in other situations the differences can be dramatic.

1.2.4 Protocols

In this subsection, we discuss examples of protocols for biometric authentication. Then we present two generic biometric models that constitute a core of biometric authentication systems. We will focus on these models and their modifications in our investigations of authentication systems. Moreover, we also sketch a protocol for biometric identification.

1.2.4.1 Protocols for Biometric Authentication

Protocol A

One of the typical protocols for secure authentication, also shown in Figure 1.1, reads as follows.

During enrollment, the biometric data of a subject are captured and analyzed, and the template X^N is extracted. A secret S is chosen or generated from these data. Then the template X^N is linked to the

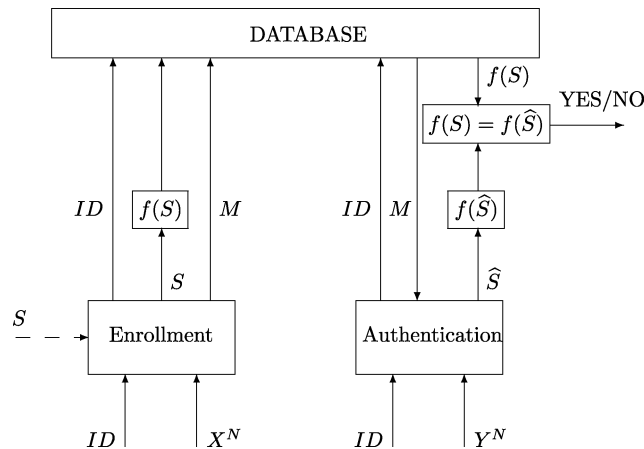


Fig. 1.1 Secure authentication. The dotted arrow indicates the possibility that the secret key is chosen, not generated from X^N .

key S via a helper message M . The key is encrypted using a one-way function and stored in a database as $f(S)$, together with an ID of the subject and the helper message M .

During authentication, the subject claims his ID. His biometric data are captured and preprocessed again, resulting in the template Y^N . The key $\hat{S}$ is estimated based on Y^N and the helper message M that corresponds to the claimed ID. This estimated key is encrypted and then matched against the encrypted key $f(S)$ corresponding to the claimed ID. Only if $f(\hat{S})$ is the same as $f(S)$ the subject is positively authenticated.

Protocol B

A variation of the first protocol is a protocol for biometric authentication with distributed storage. The first step of Protocol B is similar to the enrollment procedure in Protocol A. Here, however, the secret key is not stored anymore in the database, but on a smartcard. The smartcard is given to the individual.

During authentication the subject provides a measurement of his biometrics. This measurement is preprocessed, resulting in a noisy template Y^N . The template and the helper message M are now used to derive a key $\hat{S}$. The key $\hat{S}$ is then compared to the secret key S on the smartcard.

1.2.4.2 Two Generic Settings

From the discussions above we may conclude that in order to design a good biometric secrecy system, we can focus on a number of generic structures, that is, models that constitute the core of any biometric secrecy system. These generic, secret-key sharing models can be subdivided into a class of models with generated keys, see Figure 1.2(a), and a class of models with chosen keys, see Figure 1.2(b). This subdivision also appears in the overview paper of Jain et al. [40]. In both models S is a randomly generated/chosen secret key, X^N and Y^N are biometric enrollment and authentication sequences having length N , M is a helper message, and $\hat{S}$ is an estimated secret key. The channel between an encoder and decoder is assumed to be public. We only

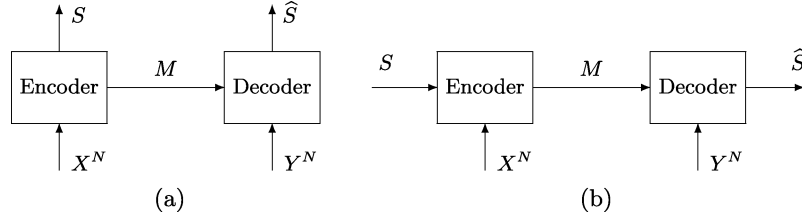


Fig. 1.2 Generic settings, generated and chosen keys.

assume that passive attacks are possible, namely, an attacker can see all public information but cannot change it. The information leakage is characterized in terms of mutual information, and the size of the secret keys in terms of entropy. The generic models must satisfy the following requirements

$$\begin{aligned}
 \Pr(S \neq \hat{S}) &\approx 0 && \text{(reliability),} \\
 \frac{1}{N} H(S) &\approx \frac{1}{N} \log_2 |\mathcal{S}| && \text{(secret uniformity),} \\
 \frac{1}{N} H(S) &\text{ is as large as possible} && \text{(secret-key rate),} \\
 \frac{1}{N} I(S; M) &\approx 0 && \text{(secrecy leakage),} \\
 \frac{1}{N} I(X^N; M) &\text{ is as small as possible} && \text{(privacy leakage).} \quad (1.3)
 \end{aligned}$$

1.2.4.3 Protocol for Biometric Identification

A typical protocol for identification systems, see Figure 1.3 consists of the following steps.

During enrollment, the biometric data of $|\mathcal{V}|$ subjects are captured and analyzed, resulting into biometric templates $X^N(v), v = 1, 2, \dots, |\mathcal{V}|$. Based on these templates, secure templates or helper data $M(1), M(2), \dots, M(|\mathcal{V}|)$ are extracted and then stored in the database together with the subjects IDs.

During identification, a subject presents his biometric data, which are captured and preprocessed again, resulting in the template Y^N . The identity label $\hat{v}$ of the individual is estimated based on Y^N and all helper data $M(1), M(2), \dots, M(|\mathcal{V}|)$ from the database. The result of this step is ID-label $\hat{v}$ of the subject or “no found” error.

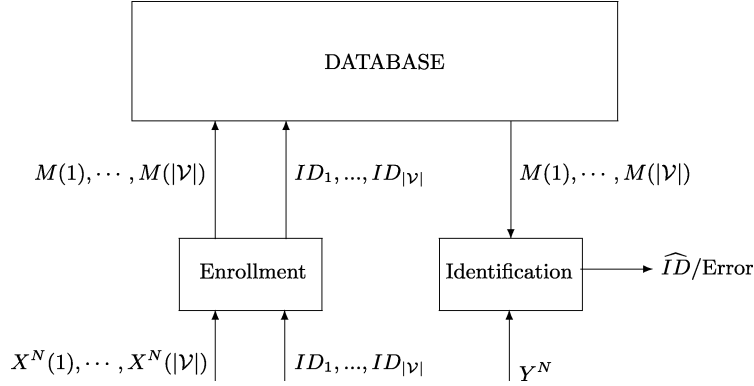


Fig. 1.3 Secure identification.

1.3 Modeling Biometric Data

Throughout this review, we will consider two types of biometric data. First we assume that our biometric sequences (feature vectors) are discrete, independent and identically distributed (i.i.d.). Fingerprints and irises can be modeled as such biometric sources. A discrete representation of other biometric modalities can be obtained using quantization. Then we also consider systems based on Gaussian i.i.d. biometric sequences. We may assume Gaussian biometric data distribution, since it is well-known that most transmission channels can be modeled as additive white Gaussian noise channels. The independence of biometric features is not unreasonable to assume, since Principal Components Analysis (PCA), Linear Discriminant Analysis (LDA) and other transforms, which are applied to biometric measurements during feature extraction (see Wayman et al. [83]) result in more or less independent features. In general, different components of biometric sequences may have different ranges of correlation. However, for reasons of simplicity we will only discuss i.i.d. biometrics here. Finally, we assume that biometric sequences are aligned.

1.4 Related Work

Privacy concerns related to the use of biometric data in various secrecy systems were raised more than a two decade ago. Schneier [64] pointed

out that biometric data are not standard secret keys and cannot be easily canceled. Ratha et al. [60] investigated the vulnerability points of biometric secrecy systems. In Prabhakar et al. [56] security and privacy concerns were raised. Linnartz and Tuyls [48] looked at the problem of achieving biometric systems with no secrecy leakage. Finally, at the DSP forum [81] secrecy- and privacy-protecting technologies were discussed. The extent to which secrecy and privacy problems were investigated in literature also received attention there.

Considerable interest in the topic of biometric secrecy systems resulted in the proposal of various techniques. Recent developments in the area of biometric secrecy systems led to methods grouped around two classes: cancelable biometrics and “fuzzy encryption.” Detailed summaries of these two approaches can be found in Uludag et al. [77] and in Jain et al. [40]. These works concentrate on biometric authentication systems.

It is the objective of cancelable biometrics, introduced by Ratha et al. [59, 60], Ang et al. [3], and Maiorana et al. [49], to avoid storage of reference biometric data in the clear in biometric authentication systems. These methods are based on non-invertible transformations that preserve the statistical properties of biometric data and rely on the assumption that it is hard to exactly reconstruct biometric data from the transformed data and applied transformation. However, hardness of a problem is difficult to prove, and, in practice, the properties of these schemes are assessed using brute-force attacks. Moreover, visual inspection shows that transformed data, for example, the distorted faces in Ratha et al. [59], still contain a lot of biometric information. Therefore, in this review we concentrate on the second class of systems.

The “fuzzy encryption” approach focuses on generation and binding of secret keys from/to biometric data. These secret keys are used to regulate access to, for example, sensitive data, services, and environments in key-based cryptographic applications and, in particular, in biometric authentication systems (all referred to as biometric secrecy systems). In biometric secrecy systems a secret key is generated/chosen during an enrollment procedure in which biometric data are observed for the first time. This key is to be reconstructed after these biometric data are observed again during an attempt to obtain access (authentication).

Since biometric measurements are typically noisy, reliable biometric secrecy systems also extract so-called helper data from the biometric observation at the time of enrollment. These helper data facilitate reliable reconstruction of the secret key in the authentication process. The helper data are assumed to be public, and therefore they should not contain information on the secret key. We say that the secrecy leakage should be negligible. Important parameters of a biometric secrecy system include the size of the secret key and the information that the helper data contain (leak) on the biometric observation. This latter parameter is called privacy leakage. Ideally the privacy leakage should be small, to avoid the biometric data of an individual to become compromised. Moreover, the secret-key length (also characterized by the secret-key rate) should be large to minimize the probability that the secret key is guessed and unauthorized access is granted.

Implementations of “fuzzy encryption” include methods based on various forms of Shamir’s secret sharing [66]. These methods are used to harden passwords with biometric data (Monrose et al. [52, 53]). Methods based on error-correcting codes, that bind uniformly distributed secret keys to biometric data and that tolerate (biometric) errors in these secret keys, were formally defined by Juels and Wattenberg [43]. Less formal approaches can be found in Davida et al. [20, 21]. Error-correction based methods were extended to the set difference metric developed by Juels and Sudan [42]. Some other approaches focus on continuous biometric data and provide solutions which are based on quantization of biometric data as in Linnartz and Tuyls [48], Denteneer et al. [22] (with emphasis on reliable components), Teoh et al. [73], and Buhan et al. [10].

Finally, a formal approach for designing secure biometric systems for three metric distances (Hamming, edit and set), called fuzzy extractors, was introduced in Dodis et al. [24] and further elaborated in [23]. Dodis et al. [23, 24] were the first ones who addressed the problem of code construction for biometric secret-key generation in a systematic information-theoretical way. Although their works provide results on the maximum secret-key rates in biometric secrecy systems, they also give the corresponding results for the maximum privacy leakage. In biometric setting, however, the goal is to minimize the privacy leakage.

The need for quantifying the exact information leakage on biometric data was also stated as an open question in Sutcu et al. [70].

Another branch of work on “fuzzy encryption” focuses on combination of biometric and cryptographic keys. Methods in this direction include attempts to harden the fuzzy vault scheme of Juels and Sudan [42] with passwords by Nandakumar et al. [54] and dithering techniques that were proposed by Buhan et al. [9].

Recently, Prabhakaran and Ramchandran [57], and Gündüz et al. [30] studied source-coding problems where the issue of (biometric) leakage was addressed. In their work, though, it is not the intention of the users to produce a secret but to communicate a (biometric) source sequence in a secure way from the first to the second terminal.

From information-theoretical point of view biometric identification systems were studied by O’Sullivan and Schmid [55] and Willems et al. [87]. They derived the corresponding identification capacity, that is, the maximum number of individuals that a systems can reliably identify. They assumed, however, storage of the biometric enrollment sequences in the clear. Later Tuncel [74] analyzed the trade-off between the capacity of a biometric identification system and the space required to store the biometric templates. Tuncel’s method realizes a kind of template protection.

1.5 Organization of This Review

In the current review, we study a number of problems related to the design of biometric secrecy systems for both authentication and identification.

First, in Section 2 we review the problem of secret sharing in order to set theoretical grounds for our investigation of secret-key rates and privacy leakage in biometric secrecy systems. In this section, we revisit the classical Ahlswede and Csiszár [1] and Maurer [51] problem of generating a secret from two dependent sequences but in the biometric setting. In this section, the biometric source is assumed to be discrete memoryless, however stationary ergodic biometric sources are also discussed there. Moreover, we investigated the question of which FRR and FAR can be achieved with biometric secret generation systems.

Next in Section 3 we continue to study secret-key rates and privacy leakage. We concentrate on biometric authentication systems. In this section we study a more general situation. One of the challenges in designing biometric secrecy systems is to minimize the privacy leakage for a given secret-key rate. Therefore, in Section 3, we focus on finding the fundamental trade-off between secret-key rates and privacy leakage for a number of biometric models. In this section we assume that our biometric source is discrete memoryless. In Section 4 we extend the results found in Section 3 to the Gaussian biometric sources.

The following section, Section 5, is devoted to biometric identification systems with protected templates. Since biometric data are typically used for both identification and authentication purposes, we determine there the trade-off between identification, secret-key and privacy-leakage rates. In this section we again assume our biometric source and channel to be discrete memoryless.

Next we turn to practical constructions. In Section 6 a popular realization of binary biometric authentication systems with chosen secret keys, called fuzzy commitment [43], is analyzed. There we present theoretical performance analysis of the scheme. The following section investigates how binary quantization of biometric sequences influences the performance of biometric secrecy systems with respect to secret-key rates and privacy leakage. Also there we study the effect of the code choice and binary quantization in fuzzy commitment.

Section 8 concludes this review and present some discussions on the future directions in the area of biometric secrecy systems.

2

Secret Sharing and Biometric Systems

In this section we revisit the classical Ahlswede and Csiszár [1] source-type problem for generating a common secret key out of two dependent sequences. Moreover, we discuss here the relation between secret-key generation and biometric authentication systems.

2.1 Introduction

The problem of generating secret keys from biometric data is closely related to the concept of secret sharing, which was introduced by Maurer [51] and (slightly later) by Ahlswede and Csiszár [1]. In the source model of Ahlswede and Csiszár [1] two terminals observe two correlated sequences X^N and Y^N and aim at producing an as large as possible common secret S by interchanging a public message M . This message, to which we refer as helper data, should only provide a negligible amount of information on the secret. It was shown that, in the i.i.d case, the maximum secret-key rate in this model is equal to the mutual information $I(X;Y)$ between the observed sequences. The secret sharing concept is also closely related to the concept of common randomness generation that was studied by Ahlswede and Csiszár [2] and later extended with helper terminals by Csiszár and Narayan [17].

In common randomness schemes the requirement that the helper data should provide only a negligible amount of information on the generated randomness is dropped.

In a biometric authentication system where two terminals need to generate a common secret key S from a biometric enrollment sequence X^N and a biometric authentication sequence Y^N , the helper-message M should provide no information on the generated secret, but it is also crucial that it provides as little information as possible on the biometric data. We call this information the privacy leakage. The interest in privacy leakage is motivated by the fact that stolen biometric data result in a stolen identity. Moreover, compromised biometrics cannot be canceled and easily substituted as people have only limited resources of biometric data. Therefore the use of biometrics raises privacy concerns, as noted by Schneier [64], Ratha et al. [60], Prabhakar et al. [56], Linnartz and Tuyls [48], in DSP forum [81], etc.

This section is devoted to theoretical grounds for our further investigation of secret-key rates and privacy leakage in biometric secrecy systems. In this section we, first, revisit the classical Ahlswede and Csiszár [1] problem of generating a secret from two dependent sequences. We focus on the biometric setting. We provide here the results on the largest achievable secret-key rate and the corresponding privacy leakage for the case when the biometric data are discrete, i.i.d. Our achievability proofs are expressed in terms of the Slepian–Wolf techniques that were presented by Cover [14], in which binning of typical sequences plays an important role. We extend these results to the stationary ergodic case. In the last part of the section we also discuss how typical biometric performance measures, that is, the FRR and the FAR, relate to the results obtained for the biometric generated-secret model. The relation between reliability criteria and secrecy in information-theoretic security problems was also studied in [32]. Finally, we demonstrate how a generated-secret system can be transformed into a chosen secret-key system.

2.2 Model for Biometric Generated-Secret Systems

Consider a biometric generated-secret system, see Figure 2.1, which is based on a biometric source with distribution $\{Q(x, y), x \in \mathcal{X},$

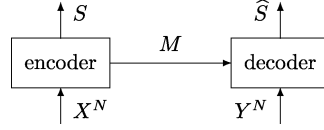


Fig. 2.1 Biometric generated-secret system model.

$y \in \mathcal{Y}\}$. This source produces a first sequence $x^N = (x_1, x_2, \dots, x_N)$ of N symbols from the finite alphabet $\mathcal{X}$ and a second sequence $y^N = (y_1, y_2, \dots, y_N)$ of N symbols from the finite alphabet $\mathcal{Y}$. We refer to the first sequence as enrollment sequence, the second sequence we call the authentication sequence. The sequence pair (x^N, y^N) occurs with probability

$$\Pr\{X^N = x^N, Y^N = y^N\} = \prod_{n=1}^N Q(x_n, y_n), \quad (2.1)$$

in other words, the sequence of pairs (X_n, Y_n) , $n = 1, 2, \dots, N$ is i.i.d according to $Q(x, y)$. The biometric sequences X^N and Y^N are in general not independent of each other. In some sense the second sequence Y^N is a noisy version of the first sequence X^N .

Next consider an encoder that observes the enrollment sequence X^N . From this sequence the encoder generates a secret $S \in \{1, 2, \dots, |\mathcal{S}|\}$ and a public helper-message $M \in \{1, 2, \dots, |\mathcal{M}|\}$. The helper-message is sent to a decoder. Hence

$$(S, M) = e(X^N), \quad (2.2)$$

where $e(\cdot)$ is the deterministic encoder mapping.

The decoder, on its turn, observes the authentication sequence Y^N and produces an estimate $\hat{S}$ of the secret S using the received helper-message M and the observed authentication sequence Y^N , hence

$$\hat{S} = d(Y^N, M), \quad (2.3)$$

where $d(\cdot, \cdot)$ is the deterministic decoder mapping.

It is the goal of the encoder and decoder to produce a secret key as large as possible, in such a way that the probability that the estimated secret $\hat{S}$ is not equal to the generated secret S is close to zero. Moreover, the secret key should be nearly uniform in the entropy sense

and the secrecy-leakage rate, characterizing information that the helper message reveals about the generated secret, should be negligible. The following definition reflects these requirements.

Definition 2.1. A secret-key rate R , for $R \geq 0$, is called achievable if for all $\delta > 0$ and all N large enough, there exist encoders and decoders such that¹

$$\begin{aligned} \Pr\{\hat{S} \neq S\} &\leq \delta, \\ \frac{1}{N}H(S) + \delta &\geq \frac{1}{N}\log|\mathcal{S}| \geq R - \delta, \\ \frac{1}{N}I(S;M) &\leq \delta. \end{aligned} \tag{2.4}$$

2.3 Statement of Result

The main result of this section is stated in the next theorem. It is actually a result found by Ahlswede and Csiszár [1], when only one-way communication is allowed.

Theorem 2.1. For a biometric generated-secret model the largest achievable secret-key rate R is equal to $I(X;Y)$.

Although this theorem was proven by Ahlswede and Csiszár [1], for the sake of completeness, we give a proof of this result in the next section. Unlike the original proof in [1], our proof is based on weak typicality, see Appendix A.1 for definitions and main properties. This proof will be the core part of the achievability proofs presented in the later sections of this work where we deal with more general biometric settings.

2.4 Proof of Theorem 2.1

2.4.1 Achievability Part of Theorem 2.1

Fix an $\varepsilon > 0$ and let $\mathcal{A}_\varepsilon^{(N)}(X)$ and $\mathcal{A}_\varepsilon^{(N)}(XY)$ be the sets of ε -typical and jointly ε -typical N -sequences based on the joint distribution

¹Throughout this review log denotes the base 2 logarithm.

$\{Q(x, y), x \in \mathcal{X}, y \in \mathcal{Y}\}$ of the XY -source. We prove the achievability using a random labeling argument.

Random Code Construction, Encoding and Decoding

Random labeling: The coding strategy is as follows. Let us define a random partition of space $\mathcal{X}^N$, viz. the space of typical X^N -sequences, into $|\mathcal{M}|$ bins. An encoder independently assigns to each sequence x^N a helper-label (index of the bin) $m \in \{1, 2, \dots, |\mathcal{M}|\}$ with probability

$$\Pr\{M(x^N) = m\} = 1/|\mathcal{M}|. \quad (2.5)$$

Furthermore, we define a second random partition over $\mathcal{X}^N$ with $|\mathcal{S}|$ bins, and the encoder also assigns to each sequence x^N a randomness-label (bin-index of this second partition) $s \in \{1, 2, \dots, |\mathcal{S}|\}$ with probability

$$\Pr\{S(x^N) = s\} = 1/|\mathcal{S}|. \quad (2.6)$$

Encoding: The encoder observes the sequence x^N and determines the secret and helper labels s and m . If x^N is not a typical sequence, an error is declared. Moreover, the encoder checks if there exists another X -sequence with the same label pair, if so also an error is declared. The encoder sends the helper label m to the decoder.

Decoding: The decoder, after having observed y^N , looks for a unique sequence x^N with label m such that $(x^N, y^N) \in \mathcal{A}_\epsilon^{(N)}(XY)$. If such a sequence is found, the decoder emits its randomness-label $\hat{s} = S(x^N)$, otherwise an error is declared.

Events, Error Probability

Events: Let X^N and Y^N be the observed biometric source sequences, S and M the actual labels, and $S(x^N)$ and $M(x^N)$ the random labels assigned to x^N , then define the events:

$$\begin{aligned} A &\triangleq \{X^N \in \mathcal{A}_\epsilon^{(N)}(X)\}, \\ B(x^N) &\triangleq \{M(x^N) = M \cap S(x^N) = S\}, \\ C &\triangleq \{(X^N, Y^N) \in \mathcal{A}_\epsilon^{(N)}(XY)\}, \\ D(x^N) &\triangleq \{M(x^N) = M\}. \end{aligned}$$

Error probability: The first problem now is to determine the error probability averaged over the random labeling. We can write for the resulting error probability averaged over the ensemble of random labelings

$$\begin{aligned}
\overline{P_E} &\leq \Pr \left\{ A^c \cup \left(\bigcup_{x^N \neq X^N : x^N \in \mathcal{A}_\epsilon^{(N)}(X)} B(x^N) \right) \cup C^c \right. \\
&\quad \left. \cup \left(\bigcup_{x^N \neq X^N : (x^N, Y^N) \in \mathcal{A}_\epsilon^{(N)}(XY)} D(x^N) \right) \right\} \\
&\leq \Pr\{A^c\} + \sum_{x^N \neq X^N : x^N \in \mathcal{A}_\epsilon^{(N)}(X)} \Pr\{B(x^N)\} + \Pr\{C^c\} \\
&\quad + \sum_{x^N \neq X^N : (x^N, Y^N) \in \mathcal{A}_\epsilon^{(N)}(XY)} \Pr\{D(x^N)\} \\
&\leq \varepsilon + |\{x^N : x^N \in \mathcal{A}_\epsilon^{(N)}(X)\}| \cdot \frac{1}{|\mathcal{M}|} \cdot \frac{1}{|\mathcal{S}|} \\
&\quad + \varepsilon + |\{x^N : (x^N, Y^N) \in \mathcal{A}_\epsilon^{(N)}(XY)\}| \cdot \frac{1}{|\mathcal{M}|} \\
&\leq 2\varepsilon + 2^{N(H(X)+\varepsilon)} \cdot \frac{1}{|\mathcal{M}|} \cdot \frac{1}{|\mathcal{S}|} + 2^{N(H(X|Y)+2\varepsilon)} \cdot \frac{1}{|\mathcal{M}|} \\
&\leq 4\varepsilon, \tag{2.7}
\end{aligned}$$

for N large enough if we take $\frac{1}{N} \log |\mathcal{M}| = H(X|Y) + 3\varepsilon$ and $\frac{1}{N} \log |\mathcal{S}| = I(X;Y) - \varepsilon$. In our derivation we used the union bound, random labeling properties, and (A.3), (A.4), and (A.5).

Wrap up

Secret-key rate and error probability: Since $\overline{P_E} \leq 4\varepsilon$ for N large enough, there exists for all N large enough, at least one pair of labelings $M(\cdot)$ and $S(\cdot)$, such that for the error probability for these labelings it holds $P_E \leq 4\varepsilon$. Now we focus on such a pair of labelings for the rest of the proof. Note that

$$\Pr\{\widehat{S} \neq S\} \leq 4\varepsilon, \tag{2.8}$$

$$H(M) \leq \log |\mathcal{M}| = N(H(X|Y) + 3\varepsilon), \tag{2.9}$$

$$H(S) \leq \log |\mathcal{S}| = N(I(X;Y) - \varepsilon). \tag{2.10}$$

Uniformity: Since the encoder requires the secret and helper labels to be unique for x^N , we can assume that it can reconstruct the sequence based on the label pair, if no error was declared. Now let $\widehat{X^N}$ be the encoder's estimate of X^N based on S and M , then we find that

$$\begin{aligned}
 H(X^N) &= H(X^N, S, M) \\
 &= H(S) + H(M|S) + H(X^N|S, M) \\
 &\leq H(S) + H(M) + H(X^N|S, M, \widehat{X^N}) \\
 &\leq H(S) + H(M) + NP_E \log |\mathcal{X}| + 1,
 \end{aligned} \tag{2.11}$$

where the last step follows from Fano's inequality, see, for example, Cover and Thomas [15], p. 206.

Hence the entropy of the secret is lower bounded by

$$\begin{aligned}
 H(S) &\geq H(X^N) - H(M) - NP_E \log |\mathcal{X}| - 1 \\
 &\stackrel{(a)}{\geq} NH(X) - N(H(X|Y) + 3\varepsilon) - NP_E \log |\mathcal{X}| - 1 \\
 &\geq NI(X; Y) - 3N\varepsilon - 4N\varepsilon \log |\mathcal{X}| - 1 \\
 &\stackrel{(b)}{=} \log |S| - 2N\varepsilon - 4N\varepsilon \log |\mathcal{X}| - 1,
 \end{aligned} \tag{2.12}$$

where (a) follows from the fact that X^N is an i.i.d sequence and from (2.9), and (b) from (2.10).

Secrecy: Next we study the secrecy

$$\begin{aligned}
 I(S; M) &= H(S) + H(M) - H(S, M) \\
 &= H(S) + H(M) - H(S, M, X^N) + H(X^N|S, M) \\
 &\stackrel{(a)}{=} H(S) + H(M) - H(X^N) + H(X^N|S, M, \widehat{X^N}) \\
 &\stackrel{(b)}{\leq} H(S) + H(M) - NH(X) + NP_E \log |\mathcal{X}| + 1 \\
 &\stackrel{(c)}{\leq} 2N\varepsilon + 4N\varepsilon \log |\mathcal{X}| + 1,
 \end{aligned} \tag{2.13}$$

where step (a) holds, since S and M are functions of X^N , and $\widehat{X^N}$ is a function of S and M , (b) follows from the fact that X^N is an i.i.d sequence and from Fano's inequality, and (c) follows from (2.9) and (2.10).

Finally, letting $\varepsilon \downarrow 0$ and $N \rightarrow \infty$, we obtain the achievability from (2.8), (2.10), (2.12), and (2.13).

2.4.2 Converse Part of Theorem 2.1

Assume that the rate R is achievable. Now we consider the entropy of the secret

$$\begin{aligned}
 H(S) &= I(S; Y^N, M) + H(S|Y^N, M) \\
 &\stackrel{(a)}{=} I(S; M) + I(S; Y^N|M) + H(S|Y^N, M, \hat{S}) \\
 &\leq I(S; M) + H(Y^N) - H(Y^N|M, S, X^N) + H(S|\hat{S}) \\
 &\stackrel{(b)}{\leq} I(S; M) + H(Y^N) - H(Y^N|X^N) + \Pr\{\hat{S} \neq S\} \log |\mathcal{S}| + 1 \\
 &\stackrel{(c)}{\leq} I(S; M) + I(X^N; Y^N) + N \Pr\{\hat{S} \neq S\} \log |\mathcal{X}| + 1 \\
 &\stackrel{(d)}{\leq} N\delta + NI(X; Y) + N\delta \log |\mathcal{X}| + 1, \tag{2.14}
 \end{aligned}$$

where step (a) holds, since $\hat{S}$ is a function of M and Y^N , (b) follows from the fact that S and M are functions of X^N and from Fano's inequality, (c) holds, since the encoder is deterministic and therefore $|S| \leq |\mathcal{X}|^N$ (possibly) after renumbering, and step (d) follows from the fact that (X^N, Y^N) is a sequence of i.i.d pairs and the fact that for achievable rates R we have that $I(S; M) \leq N\delta$ and $\Pr\{\hat{S} \neq S\} \leq \delta$.

Then for achievable rates R we obtain

$$R - 2\delta \leq \frac{H(S)}{N} \leq I(X; Y) + \delta + \delta \log |\mathcal{X}| + \frac{1}{N}. \tag{2.15}$$

Letting $\delta \downarrow 0$ and $N \rightarrow \infty$, we obtain the converse part to the theorem.

2.5 Stationary Ergodic Case

In the previous sections we considered the case when the biometric data source is i.i.d. Now we argue that the results, similar to those derived for the i.i.d case, hold for biometric sequences that are generated by jointly stationary ergodic sources. In Csiszár and Narayan [18] it was

also argued that the i.i.d results on strong secrecy capacity for multiple terminals can be extended to the stationary ergodic case.

Let (X^N, Y^N) be a jointly stationary ergodic sequence pair, see, for example, Cover and Thomas [15], p. 644 for exact definitions, and define²

$$\begin{aligned} H_\infty(X) &\triangleq \lim_{N \rightarrow \infty} \frac{1}{N} H(X_1, X_2, \dots, X_N), \\ H_\infty(Y) &\triangleq \lim_{N \rightarrow \infty} \frac{1}{N} H(Y_1, Y_2, \dots, Y_N), \\ H_\infty(X, Y) &\triangleq \lim_{N \rightarrow \infty} \frac{1}{N} H(X_1, Y_1, X_2, Y_2, \dots, X_N, Y_N), \\ H_\infty(X|Y) &\triangleq H_\infty(X, Y) - H_\infty(Y), \\ I_\infty(X; Y) &\triangleq H_\infty(X) + H_\infty(Y) - H_\infty(X, Y). \end{aligned} \quad (2.16)$$

Then the following theorem holds.

Theorem 2.2. For a jointly stationary ergodic sequence pair (X^N, Y^N) the result of Theorem 2.1 holds if we replace $I(X; Y)$ with $I_\infty(X; Y)$.

2.5.1 Proof of Theorem 2.2

The proof of Theorem 2.2 is similar to the proofs of Theorem 2.1. The difference is that now the definitions of typical sets are as in Cover [14].

Moreover, in both the achievability and converse proofs, the steps involving equalities $H(X^N) = NH(X)$ and $I(X^N; Y^N) = NI(X; Y)$ have to be adjusted. Using the definitions of the stationary ergodic sequences and noting that the achievability proof and the converse hold for N large enough, we substitute $NH_\infty(X)$ for $H(X^N)$ and $NI_\infty(X; Y)$ for $I(X^N; Y^N)$ in (2.12), (2.13), and (2.14).

² Note that $H_\infty(\cdot)$ is also used to denote min-entropy, here, however, it denotes the entropy rate.

2.6 From Secret-Key Generation to Biometric Authentication

2.6.1 Authentication Scenario

So far we have discussed secret-key sharing only. Now we will demonstrate how a secret-key sharing system can be incorporated into a biometric authentication system. The idea is to embed the generated-secret system into a biometric authentication system as shown in Figure 2.2.

Consider the following scenario. First individuals are enrolled in a biometric system. By this we mean that for each such individual an observed biometric sequence X^N , that is, the enrollment sequence that the individual presents, is processed by an encoder. The encoder derives a secret-key S and a helper-message M from this X^N . The helper-message is stored in a public database, the secret-key is prepared to be made available to an authenticator at some later time.³

When an authorized individual arrives for authentication, the system has to check if the individual is the one who he claims to be. To do it, individual's biometric data are observed and the resulting sequence Y^N is handed over to a decoder that retrieves from the database the helper message corresponding to the claimed ID. Based on Y^N the decoder forms an estimate of the secret key $\hat{S}$ and this secret key is compared to the secret key S corresponding to the claimed ID. If the secret keys match the authenticator assumes that the person is who

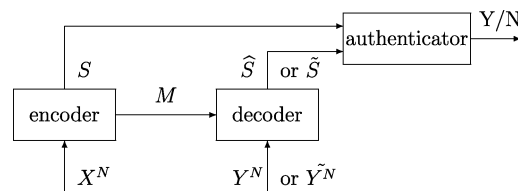


Fig. 2.2 A biometric authentication system based on secret generation.

³ A typical protocol of conveying the secret-key S to the authenticator is to apply a one-way function $f(\cdot)$ to it and store the result $f(S)$ in the public database, together with the helper message M . Thus, when the authenticator need to compare two secrets, it compares values of the corresponding one-way functions. If the secret keys match, the values of their one-way functions also match. We will not further elaborate on this issue here however.

he claims to be. If no match occurs the individual is rejected by the system.

Now consider an imposter who claims to be a certain individual. The imposter can observe the helper-message M corresponding to the individual of interest, and based on this helper-message form a biometric sequence $\tilde{Y}^N$ that he presents for observation to the decoder. The decoder then forms a secret key $\tilde{S}$, making use of the helper data M . The authenticator finally checks whether this $\tilde{S}$ equals the secret-key S corresponding to the claimed ID, and decides based on the outcome of this check whether or not the imposter is accepted by the system.

The FRR and the FAR are typical performance measures for biometric authentication systems. Recall that in standard biometric authentication systems the FRR is defined as the probability that an authorized individual cannot get access to a system, while the FAR is defined as the probability that an unauthorized individual (an imposter) is granted access, hence

$$\begin{aligned}\text{FRR} &= \Pr\{\hat{S} \neq S\}, \\ \text{FAR} &= \Pr\{\tilde{S} = S\}.\end{aligned}\tag{2.17}$$

The result that we prove in what follows can be stated now.

Theorem 2.3. For a biometric authentication model based on secret-generation there exist encoders and decoders that achieve $\text{FRR} \leq \delta$, and $\text{FAR} \leq 2^{-N(I(X;Y)-\delta)}$ for all imposter strategies, for all $\delta > 0$ and all N large enough.

This result immediately links the maximum secret-key rate $I(X;Y)$ to the achievable false-accept exponent $\frac{1}{N} \log \frac{1}{\text{FAR}}$. This relation explains why, throughout this review, we are interested in secret-key rates as large as possible.

2.6.2 Achievability Proof for Theorem 2.3

We start our achievability proof by fixing a $\delta > 0$ and N large enough. Now consider an encoder and decoder that satisfy the conditions of achievability given in Definition 2.1. The achievability proof of

Theorem 2.1, given in Section 2.4.1, guarantees the existence of such an encoder and decoder. We first discuss the FRR of this encoder and decoder. Later we modify the encoder and consider the FAR for an arbitrary imposter, resulting from this modification.

2.6.2.1 FRR

For an authorized individual, our encoder and decoder guarantee that

$$\text{FRR} = \Pr\{S \neq \hat{S}\} \leq \delta, \quad (2.18)$$

for all N large enough.

2.6.2.2 FAR

Next we discuss the situation in which an imposter chooses a sequence $\tilde{y}^N$ to get authenticated. Note that the imposter has access to the helper message M corresponding to the identity of interest, which is stored in the public database. The imposter will try to choose $\tilde{y}^N$, knowing the helper message M , in such a way that the decoder's output is equal to the secret key S corresponding to the claimed ID.

We fix a helper-message index m . Note that the decoder will never decode a sequence x^N which is not typical, that is, $x^N \notin \mathcal{A}_\epsilon^{(N)}(X)$. Therefore an imposter need only care about typical sequences x^N having helper-message index m .

To make it hard for the imposter to get a large FAR, we modify the secret-key labels of all typical sequences x^N with helper-message label m . More specially, we distribute these sequences x^N over all s -labels, in such a way that the number of x^N sequences having a certain s -label differs by at most one.⁴

The number of typical x^N sequences with helper-message index m is upper bounded by $\Pr\{M = m\}/2^{-N(H(X)+\epsilon)}$. Distributing these sequences over all s -labels leads to at most $\lceil \Pr\{M = m\}/(2^{-N(H(X)+\epsilon)} \cdot |S|) \rceil$ typical sequences having a certain secret-key label.

⁴At this point we do not require anymore the uniformity of the secret-key and secrecy leakage $\frac{1}{N}I(S;M)$ to be negligible and just use Slepian–Wolf coding. We only focus on the FAR of an authentication system with helper data and secret keys. Note that the FRR realized by the encoder and decoder does not change if we only modify the s -labels of the typical x^N sequences.

The probability that an imposter, when m occurs, chooses the correct secret, no matter what choice he makes, is therefore upper-bounded by

$$\left\lceil \frac{\Pr\{M = m\}}{2^{-N(H(X)+\varepsilon)} \cdot |\mathcal{S}|} \right\rceil \cdot 2^{-N(H(X)-\varepsilon)}. \quad (2.19)$$

We obtain an upper bound for the FAR, by carrying out the summation over all m . This results in

$$\begin{aligned} \text{FAR} &\leq \sum_{m=1}^{|\mathcal{M}|} \left\lceil \frac{\Pr\{M = m\}}{2^{-N(H(X)+\varepsilon)} \cdot |\mathcal{S}|} \right\rceil \cdot 2^{-N(H(X)-\varepsilon)} \\ &\leq \sum_{m=1}^{|\mathcal{M}|} \left(\frac{\Pr\{M = m\}}{2^{-N(H(X)+\varepsilon)} \cdot |\mathcal{S}|} + 1 \right) \cdot 2^{-N(H(X)-\varepsilon)} \\ &= \sum_{m=1}^{|\mathcal{M}|} \frac{\Pr\{M = m\}}{2^{-N(H(X)+\varepsilon)} \cdot |\mathcal{S}|} \cdot 2^{-N(H(X)-\varepsilon)} + \sum_{m=1}^{|\mathcal{M}|} 2^{-N(H(X)-\varepsilon)} \\ &= 2^{-N(H(X)-3\varepsilon)} + 2^{-N(I(X;Y)-4\varepsilon)} \\ &\leq 2^{-N(I(X;Y)-5\varepsilon)}, \end{aligned} \quad (2.20)$$

for large enough N , for all imposters. Note that we have used Slepian–Wolf coding here. Letting $\varepsilon \downarrow 0$ and $N \rightarrow \infty$, results in the theorem.

2.6.3 An Optimal Imposter Strategy

From the achievability proof of Theorem 2.1 for generated-secret systems, it follows that the estimate of the generated secret will be equal to the generated key only if a biometric authentication sequence is jointly typical with the biometric enrollment sequence. Therefore an imposter can try to generate at random a biometric sequence $\tilde{Y}^N$, that is, jointly typical with the enrolled sequence X^N , according to distribution $\{Q(x, y), x \in \mathcal{X}, y \in \mathcal{Y}\}$. If the sequence $\tilde{y}^N$ is generated i.i.d according to $Q(\tilde{y}) = \sum_x Q(x, \tilde{y})$, it follows from the properties of jointly typical sequences, see (A.7), that this imposter achieves for N large enough

$$\text{FAR} = \Pr\{(X^N, \tilde{Y}^N) \in \mathcal{A}_\epsilon^{(N)}(XY)\} \geq (1 - \varepsilon)2^{-N(I(X;Y)+3\varepsilon)}, \quad (2.21)$$

and from (A.6) that

$$\text{FAR} = \Pr\{(X^N, \tilde{Y}^N) \in \mathcal{A}_\epsilon^{(N)}(XY)\} \leq 2^{-N(I(X;Y)-3\epsilon)}. \quad (2.22)$$

On one hand, it is remarkable that the imposter need not make use of the helper-message in this way. However, on the other hand, we have designed the encoder and decoder for the generated-secret system, such that $\frac{1}{N}I(S;M)$ is negligible.

It is also interesting to see that these bounds on the FAR relates to the identification capacity in biometric systems without secrecy constraints found in Willems et al. [87] and Schmid and O’Sullivan [63]. Achievability for identification is however more related to property (A.6), which corresponds to the probability that some other enrolled biometric sequence is jointly typical with the observed version of the enrolled sequence of the observed individual. This again is strongly related to random coding argument showing that channel capacity is achievable, as for example, in Theorem 7.7.1. in Cover and Thomas [15].

2.7 Privacy Leakage for Codes Achieving the Maximum Secret-Key Rate

In biometric secrecy systems we are also interested in the amount of information that the helper data leak about the biometric data. This leakage is called privacy leakage and $I(M;X^N)/N$ is defined to be the privacy leakage per biometric source symbol. Now we concentrate on the codes that we have determined in the achievability proof of Theorem 2.1. The following proposition gives us the privacy leakage per source symbol corresponding to the maximum achievable secret-key rate in the biometric generated-secret model when such codes are used. A similar result for error-correcting codes was also demonstrated in Dodis et al. [24] and Smith [69] and for binary error-correcting codes in Tuyls and Goseling [76]. The result obtained in the proposition below is somewhat more general.

Proposition 2.4. In a biometric generated-secret model, for the codes that demonstrate the achievability of secret-key rate $I(X;Y)$, the privacy leakage per source symbol is equal to $H(X|Y)$.

Remark: Note that this result also holds for the stationary ergodic case, if we replace $I(X;Y)$ and $H(X|Y)$ with $I_\infty(X;Y)$ and $H_\infty(X|Y)$, respectively.

The above proposition gives us the privacy leakage per source symbol if we apply the coding scheme outlined in the achievability proof. However, it may be possible to achieve a privacy leakage that is smaller if an alternative coding scheme is applied. Note that in the biometric setting, we are interested in a system where the helper data leak as little information as possible about the biometric data. Therefore now the question arises what the minimum achievable privacy leakage in a biometric secrecy system is, for a fixed secret-key rate. The answer to this question, though, we postpone to the next section.

2.7.1 Proof of Proposition 2.4

Note that, since M is a function of X^N , we have that $I(X^N;M) = H(M)$. Hence the privacy leakage is equal to the entropy of the helper data. From the achievability proof corresponding to Theorem 2.1, we can write for the entropy of the helper data, using (2.11), (2.10) and the fact that X^N is an i.i.d sequence, that

$$\begin{aligned} I(X^N;M) = H(M) &\geq H(X^N) - H(S) - NP_E \log |\mathcal{X}| - 1 \\ &\geq NH(X|Y) + N\varepsilon - 4N\varepsilon \log |\mathcal{X}| - 1. \end{aligned} \quad (2.23)$$

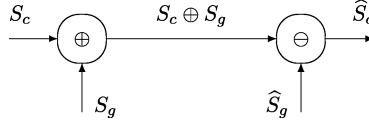
On the other hand, for the helper data (2.9) holds, and thus

$$I(X^N;M) = H(M) \leq N(H(X|Y) + 3\varepsilon). \quad (2.24)$$

Then, dividing both sides of the above expressions by N and letting $\varepsilon \downarrow 0$ and $N \rightarrow \infty$, we finalize the proof.

2.8 Chosen Secrets, Masking Layer

In the previous sections we described a procedure for secure key-generation from biometric data. Some application, however, require that a secret key can be chosen and then bound to a biometric enrollment sequence. This secret key should be reconstructed based on an authentication sequence, making use of helper data stored in a public

Fig. 2.3 Masking layer with additional helper data $S_c \oplus S_g$.

database. Again we require the mutual information between the helper data and the chosen secret key (per source symbol) to be negligible.

To accomplish this a masking layer as in [1], see Figure 2.3, which is based on the one-time pad principle, is introduced. This masking layer is added on top of the biometric generated-secret system discussed in this section. In the masking layer, a uniform secret key S_c is bound to a biometric sequence X^N . The encoder transmits additional helper-data, which is the chosen secret key S_c added modulo $|\mathcal{S}|$ to the generated secret key S_g , to the decoder. The decoder subtracts, modulo $|\mathcal{S}|$, the estimated secret key $\hat{S}_g$ based on Y^N from the additional helper data and uses the resulting secret key $\hat{S}_c$ for authentication purposes.

Theorem 2.5. Using a masking-layer procedure, based on a uniform secret key S_c , the chosen secret-key rate $I(X;Y)$ is achievable and this rate is the maximum achievable rate. Thus, for all $\delta > 0$ and N large enough, there exist encoders and decoders with

$$\begin{aligned} \Pr\{\hat{S}_c \neq S_c\} &\leq \delta, \\ \frac{1}{N}H(S_c) &\geq I(X;Y) - \delta, \\ \frac{1}{N}I(S_c; M, S_c \oplus S_g) &\leq \delta. \end{aligned} \tag{2.25}$$

Note that this theorem can also be generalized to the stationary ergodic case.

2.8.1 Proof of Theorem 2.5

The proof relies on adding a masking layer to the generated-secret scheme described in the achievability proof of Theorem 2.1. Note first that $\Pr\{\hat{S}_c \neq S_c\} = \Pr\{\hat{S}_g \neq S_g\} \leq \delta$.

Take a chosen key S_c to be uniform from $\{1, 2, \dots, |\mathcal{S}|\}$. Then from (2.10), the entropy of the chosen secret key is $H(S_c) = \log |\mathcal{S}| = N(I(X; Y) - \varepsilon)$. Note that $I(S_c; M, S_c \oplus S_g) = H(M, S_c \oplus S_g) - H(M, S_c \oplus S_g | S_c)$. Therefore we first consider

$$\begin{aligned} H(M, S_c \oplus S_g) &= H(M) + H(S_c \oplus S_g | M) \\ &\leq H(M) + H(S_c) \\ &= H(M) + \log |\mathcal{S}|. \end{aligned} \quad (2.26)$$

For the second term we find

$$\begin{aligned} H(M, S_c \oplus S_g | S_c) &= H(M | S_c) + H(S_c \oplus S_g | M, S_c) \\ &= H(M) + H(S_g | M, S_c) \\ &= H(M) + H(S_g | M) \\ &= H(M) + H(S_g) - I(S_g; M). \end{aligned} \quad (2.27)$$

Now we combine the two terms and obtain, using (2.12) and (2.13), that

$$\begin{aligned} I(S_c; M, S_c \oplus S_g) &\leq \log |\mathcal{S}| - H(S_g) + I(S_g; M) \\ &\leq \log |\mathcal{S}| - (\log |\mathcal{S}| - 2N\varepsilon - 4N\varepsilon \log |\mathcal{X}| - 1) \\ &\quad + 2N\varepsilon + 4N\varepsilon \log |\mathcal{X}| + 1 \\ &= N(4\varepsilon + 8\varepsilon \log_2 |\mathcal{X}|) + 2. \end{aligned} \quad (2.28)$$

Letting $\varepsilon \downarrow 0$ and $N \rightarrow \infty$ finalizes the proof.

2.9 Conclusions

In this section we have considered the classical Ahlswede–Csiszár secret generation model in a biometric setting. The maximum secret-key rate achievable for this model is equal to $I(X; Y)$. Although this result was proved using strong typicality before by Ahlswede and Csiszár [1], for the sake of completeness we have provided a proof here too. Our proof, which is based on weak typicality, will be the core part of the achievability proofs given in the later sections of this review where we deal with more general biometric settings.

We have also demonstrated that the privacy leakage that corresponds to the maximum secret-key rate in our biometric secret generation system is roughly equal to $H(X|Y)$. A similar result but for error-correcting codes was also obtained in Dodis et al. [24], Smith [69], and Tuyls and Goseling [76]. The question whether it is possible to achieve smaller privacy leakage will be addressed in the next section.

Next we have investigated the question of which FRR and FAR can be achieved with biometric secret generation systems. We have shown that the FRR can be made arbitrarily small, while at the same time the FAR can be guaranteed to be not larger than roughly $2^{-NI(X;Y)}$.

The biometric generated-secret model was studied for discrete i.i.d biometric sources. We could also extend the i.i.d results to stationary ergodic sources. In addition to generated-secret systems we also studied the relation between chosen-secret systems and generated-secret systems.

3

Biometric Authentication. Discrete Biometric Sources

In this section, we investigate biometric models for generating and choosing secret keys, where the goal is to minimize privacy leakage. These models constitute the core of biometric authentication systems. We start with discrete biometric data case and study fundamental trade-off between secret-key rate and privacy-leakage rate in biometric secrecy systems. We also study systems with no privacy leakage. In this type of systems private key is required and for them secret-key rate versus private-key rate trade-off is investigated.

3.1 Introduction

In the previous section, we have started the analysis of biometric systems with generated secrets. We have considered the setup where the system was optimized with respect to achieving the maximum secret-key rate, and for this case we analyzed the corresponding privacy leakage. In the current section we consider a more general situation. As noted in the introduction to the previous section, the use of biometrics brings about privacy concerns, and therefore in a biometric setting the goal is to minimize the privacy leakage for a given secret-key rate.

Thus we are interested in finding the trade-off between the secret-key rate and the privacy leakage.

In the overview paper of Jain et al. [40] it was noted that biometric authentication secrecy systems can be subdivided into those that generate secret keys from biometric data and those that bind secret keys to biometric data. Therefore in this section, we consider two types of biometric secrecy systems, that is, systems with generated secret keys and systems with chosen (bound) secret keys.

3.1.1 Eight Models

In this section we consider four biometric settings. The first one is the standard Ahlswede–Csiszár secret-generation setting. There two terminals observe two correlated biometric sequences. It is their objective to form a common secret by interchanging a public message. This message should contain only a negligible amount of information about the secret, but, in addition, we require here that it should leak as little information as possible about the biometric data. For this first case the fundamental trade-off between the secret-key rate and the privacy-leakage rate will be determined. It should be noted that this result is in some ways similar to and a special case of the SK (secret-key) part of Theorem 2.4 in Csiszár and Narayan [17].

The second setting that we consider is a biometric model with chosen keys, where the secret key is not generated by the terminals but independently chosen at the encoder side and conveyed to the decoder. This model corresponds to key-binding described in the overview paper of Jain et al. [40]. For the chosen-key setting we will also determine the fundamental secret-key rate versus privacy-leakage balance.

The other two biometric settings that we analyze correspond to biometric secrecy systems with zero privacy leakage. Solely, biometrics may not always satisfy the security and privacy requirements of certain systems. In this case the performance of biometric systems can be enhanced using standard cryptographic keys. Although this reduces user convenience, since for example, extra cryptographic keys need to be stored on external media or memorized, such systems may offer a higher level of secrecy and privacy. Practical methods in this direction include

attempts to harden the fuzzy vault scheme of Juels and Sudan [42] with passwords by Nandakumar et al. [54] and dithering techniques that were proposed by Buhan et al. [9]. In our model we assume that only the two terminals have access to an extra independent private key, which is observed together with the correlated biometric sequences. The private key is used to achieve a negligible amount of privacy leakage (zero leakage). We investigate both the generated-secret model with zero-leakage and the model with chosen keys and zero-leakage. For both models we will determine the trade-off between the private-key rate and the resulting secret-key rate.

For the four settings outlined above, the fundamental balance will be determined for both unconditional and conditional privacy leakage. This results in eight biometric models. Unconditional leakage corresponds to the unconditional mutual information between the helper data and the biometric enrollment sequence, while conditional leakage relates to this mutual information conditioned on the secret. The latter notion of privacy leakage is somewhat stricter than the unconditional one, since it leads to characterization of the amount of information that the helper data leak about the pair — the secret key and the biometric data.

We assume that biometric sequences (feature vectors) are discrete, independent and i.i.d. Fingerprints and irises are typical examples of such biometric sources. A discrete representation of other biometric modalities can be obtained using quantization. The independence of biometric features is not unreasonable to assume, since PCA, LDA and other transforms, which are applied to biometric measurements during feature extraction (see Wayman et al. [83]) result in more or less independent features. In general, different components of biometric sequences may have different ranges of correlation. However for reasons of simplicity we will only discuss identically distributed biometrics here.

3.1.2 A Motivating Example

Before we turn to a more formal part of this section, we first discuss an example. Consider an i.i.d biometric binary symmetric double source

$\{Q(x, y), x \in \{0, 1\}, y \in \{0, 1\}\}$ with crossover probability $0 \leq q \leq 1/2$ such that $Q(x, y) = (1 - q)/2$, for $y = x$ and $Q(x, y) = q/2$, for $y \neq x$. In this example we use $q = 0.1$. In the classical Ahlswede–Csiszár [1] secret-generation setting the maximum secret-key rate for this biometric source is $R = I(X; Y) = 1 - h(q)$, where $h(\cdot)$ is the binary entropy function expressed in bits. The corresponding privacy-leakage rate in this case is $H(X|Y) = h(q)$. Then the ratio between secret-key rate and privacy-leakage rate is equal to $(1 - h(q))/h(q) = 1.1322$.

Now suppose that we want to reduce the privacy-leakage rate to a fraction of α of its original size. We could apply a trivial method in which we only use a fraction α of the biometric symbols, but then the secret-key rate is also reduced to a fraction of α of its original size, and there is no effect on the key-leakage ratio. A question now arises of whether it is possible to achieve a larger key-leakage ratio at reduced privacy leakage.

We will demonstrate next that we can achieve this goal using the binary Golay code as a vector quantizer. This code consists of 4096 codewords of length 23 and has minimum Hamming distance 3. It is also perfect, that is, all 4096 sets of sequences having a distance of at most 3 from a codeword are disjoint and their union is the set of all binary sequences of length 23. A decoding sphere of this code contains exactly 2048 sequences and within a decoding sphere there are 254 sequences that are different from the codeword at a fixed position. This perfect code is now used as a vector quantizer for $\{0, 1\}^{23}$, hence each binary biometric enrollment sequence x^{23} is mapped onto the closest codeword u^{23} in the Golay code. Now we consider the derived biometric source whose enrollment output is the quantized sequence U^{23} of X^{23} and whose authentication output is the sequence Y^{23} .

Again we are interested in the key-leakage ratio $I(U^{23}; Y^{23})/H(U^{23}|Y^{23})$ for which we can now write

$$\begin{aligned} \frac{I(U^{23}; Y^{23})}{H(U^{23}|Y^{23})} &= \frac{H(Y^{23}) - H(Y^{23}|U^{23})}{H(U^{23}) + H(Y^{23}|U^{23}) - H(Y^{23})} \\ &= \frac{23 - H(Y^{23}|U^{23})}{H(Y^{23}|U^{23}) - 11}. \end{aligned} \quad (3.1)$$