Jean-Paul M.G. Linnartz, Boris Skoric and Pim Tuyls, "A Communication-Theoretical View on Fuzzy Extractors, pp. 57-78, Chapter 4 in "Security with Noisy Data: Private Biometrics, Secure Key Storage and Anti-Counterfeiting (Hardcover)", Edited by Pim Tuyls, Boris Skoric, Tom Kevenaar, Springer, 2008, ISBN 978-1-84628-983-5

1

A Communication Theorical View on Biometrics

Jean-Paul M.G. Linnartz[1][2], Boris Škorić[1] and Pim Tuyls[1][3]

[1] Philips Research j.p.linnartz@philips.com, boris.skoric@philips.com, pim.tuyls@philips.com
[2] Eindhoven University of Technology j.p.linnartz@tue.nl
[3] Katholieke Universiteit Leuven

[4]

1.1 Introduction

The recent achievements in enhanced throughput, efficiency and reliability of wireless communication systems can largely be contributed to the availability of a versatile mathematical framework for the behavior and performance of digital transmission schemes. The key foundation was Shannon's 1948 paper [16] which introduced the notion of capacity. The term capacity is defined as the maximum achievable rate of information exchange, where the maximization is conducted over all possible choices of transmission and detection techniques. The existence of a fundamental limit has acted as an irresistable target for ambitious engineers. However, it was only until the 1990s that the signal processing capabilities allowed a true exploitation of these insights and the throughput of practical systems closely reached the capacity limits. Another important condition was met earlier: the availability of sufficiently realistic statistical models for signals, the noise and the channel.

The research area of biometrics is presumably less mature in this respect, but strong progress is being made into the statistical modelling of biometric measurements and of sensor imperfections. Most importantly, the notion of a distance between two biometric measurements appears to exist, where a larger distance indicates a lesser likelihood of a statistical deviation. A further refinement is that errors in measurements can be modelled with well behaved joint probability functions.

Anticipating on the further sophistication and verification of such models, this chapter proposes a framework that models the capacity and performance

[4] Published as Chapter 4 in: Tuyls, Skoric and Kevenaar, Security with Noisy Data, 2007 Springer

of such systems, initially assuming generic probabilistic models for biometric sources and sensor aberrations.

We argue that the maximum information rate of the biometric measurement channel can be related directly to the *identification capacity* of the biometric system. The statistical behaviour of the biometrics, as well as the variability between different (imperfect) measurements are assumed to be known in the form of a statistical model. This reveals a commonality between commucation systems and biometric systems that can covered by a variation on Shannon's theory.

In communications, reliable transfer of data can be separated from security functions such as encryption for confidentiality. The usual approach is to start with source coding to compress the input data into an efficient representation, followed by encryption. Thirdly, redundancy is added in the form of channel coding to allow for error correction. In this regard, the biometric system is different. Biometric signals can not be compressed, encrypted, protected by error correction or in any other form be pre-conditioned before being offered to a sensor. Nonetheless, one needs to involve non-linear operations during the detection, for instance to prevent impersonation and leakage of confidential personal information. The second part of this chapter shows that security operations can be performed to shield important information from untrusted parties without significantly affecting the user capacity or security of the system.

1.2 Preliminaries

We distinguish between identification and verification. *Identification* estimates which person is present by searching for a match in a data base of reference data for many persons. The decoder a priori does not know whether he sees "Peggy" or "Petra". The outcome of the identification is either the (most likely) identity of the person present, or an erasue, i.e., the decoder can not establish who the person is with reasonable accuracy.

When assessing the user capacity of an identification system, we are interested in knowing how many individuals can be identified reliably by a biometrical identification system, in particular how this is a function of the amount of observed data and the quality of the observations.

On the other hand, *verification* attempts to establish whether the prover, i.e., the person who is undergoing the verification test, truly is Peggy, which she claims to be. The prover provides not only biometric data but also a message in which she claims to be Peggy. In security language, the decoder is called the verifier and is named Victor. He is assumed to have some a priori knowledge about Peggy, for instance in the form of certified reference data, but at the start of the protocol he is not yet sure whether Peggy is present or another person performing an impersonation attack. The outcome of the verification is binary: either 'the prover is Peggy' or 'the prover is not Peggy.'

In identification, the verifier must have access to a data base of reference data from all individuals. In verification, this is not necessary, and typically the reference data of only Peggy suffices. An identification algorithm can, at least in theory, be modified into a verification algorithm by grouping the set of all outputs except 'the person is Peggy' into the single outcome 'the person is not Peggy'. However, in general the optimization and decision regions are usually chosen differently for identification and verification. Identification systems make the most likely choice, while verification systems are designed around false positive and false negative probabilities.

Private verification, which we will address from Section 1.8 onwards, is a special form of verification in which certain security requirements are also met. In particular, the outcome of private verification can be that the person not only shows biometrics that fit with Peggy, but also that she knows a secret that only Peggy is supposed to know. After the private verification session, Victor preferably doesn't know what the secret is, although Victor can be convinced that the prover knows Peggy's secret.

1.3 Model: Biometrics as Random Codewords

Biometrical systems in general involve two phases. In an *enrolment phase* all individuals are observed and for each individual $p \in \{1, \cdots, M\}$ a record $\mathbf{Y}(p)$ is added to a database. This record is called 'reference data', 'enrolment data' or 'template' and contains L symbols from the alphabet $\mathcal{Y}$. The enrolment data is a noisy version of the biometrical data $\mathbf{X}(p) \in \mathcal{X}^L$ corresponding to the individual p. The set of enrolment data for all users is denoted as the entire M by L matrix $\mathsf{Y} = (\mathbf{Y}(1), \mathbf{Y}(2), \cdots, \mathbf{Y}(M))$.

In the *operational phase*, an unknown individual is observed again. The resulting identification-data $\mathbf{Z}$, another noisy version of the biometrical data $\mathbf{X}$ of the unknown individual, is compared to (all or a subset of) the enrolment data in the database and the system has to come up with an estimate of the individual. An essential fact in this procedure is that both in the enrolment phase and in the operational phase noisy versions of the biometrical data are obtained. The precise biometrical data $\mathbf{X}(p)$ remain unknown.

We use capital notation $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ for random variables and bold face to denote vectors, in this section of dimension L. Moreover, $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$, denote realizations or the random variables. Each individual has a biometrical data sequence $\mathbf{x} = (x_1, x_2, \cdots, x_L)$ with components $x_i \in \mathcal{X}$ for $i = 1, \cdots, L$. The sequence $\mathbf{x}(p)$ is the sequence for person p.

For the development of the theory, preferably we assume that the components of each sequence are independent and identically distributed (IID), and that the biometric source can generate arbitrarily long sequences ($L \to \infty$). In practice, most physical or biological parameters show correlation, but often a set of biometric measurements can be transformed into a sequence of IID random variables. Communication engineers usually consider data to arrive

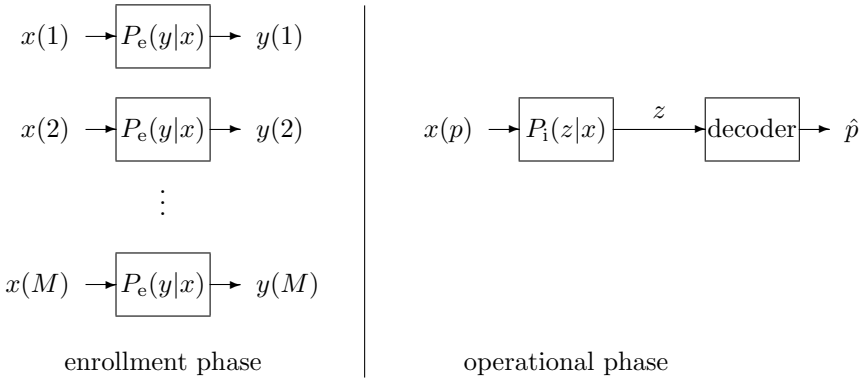

Fig. 1.1. Model of a biometrical identification system. Each biometric measurement is represented as a noisy channel.

sequentially, and usually speak of a 'memoryless source' if $\mathbf{X}$ is IID and of a 'memoryless time-invariant channel' if for a memoryless source $\mathbf{Y}$ (resp. $\mathbf{Z}$) is also IID.

The measured output sequence $\mathbf{Z}$ is used by a *decoder*. In identification, the decoder has access to all M enrolment sequences stored in the database Y. The decoder produces an estimate $\hat{p}$ of the index of the unknown individual, $\hat{p} = \mathsf{Dec}(\mathbf{z}, \mathbf{y})$. An erasure $\perp$ is also a valid decoder output. Hence $\hat{p} \in \{\perp, 1, 2, \cdots, M\}$. Two relevant system parameters are the *maximal error probability* $P_{\max}$ and the *rate* R

$$P_{\max} \stackrel{\Delta}{=} \max_{1 \le p \le M} \mathbb{P}\Big[\hat{P} \neq p | P = p\Big] \quad \text{and} \quad R \stackrel{\Delta}{=} \frac{1}{L} \log_2 M. \tag{1.1}$$

For an ideal binary ($\mathcal{X} = \{0, 1\}$) biometric system, we have $M = 2^L$, so $R = 1$. The *Identification Capacity*, to be defined later, describes asymptotic system properties that theoretically apply only for the case of infinitely long sequences $L \to \infty$. In fact, we will argue that there exists a rate C_{id} such that $P_{\max}$ is arbitrarily small for rates below capacity ($R < C_{\mathrm{id}}$) and that $P_{\max}$ necessarily tends to unity for rates above C_{id}. This C_{id} will be called the identification capacity.

1.3.1 Source and Channel Model

We will denote the probabilities on $\mathbf{X}$ as the source model. We define the 'enrolment channel' as the probability of $\mathbf{Y}(p)$ conditioned on the biometric $\mathbf{X}(p)$. Similarly, we define the operational (or identification) channel as the probability of $\mathbf{Z}(p)$ conditioned on $\mathbf{X}(p)$.

In communication systems, the channel is mostly modelled as a statistical operation that is independent of the source. In our generic biometric framework that is not necessarily the case, although our examples assume this.

IID Source

Each biometrical data sequence is assumed to be generated by an independent identically distributed source according to the symbol distribution $Q(\kappa) = \mathbb{P}[x_l(p) = \kappa]$. The probability distribution of the full sequence is

$$P_{\mathbf{X}(p)}(\mathbf{x}) = \mathbb{P}[\mathbf{X}(p) = \mathbf{x}] = \prod_{l=1}^{L} Q(x_l). \tag{1.2}$$

Note that the the distribution Q does not depend on p.

IID Independent Memoryless Channel

In the enrolment phase all biometrical data sequences are observed via an enrolment channel $\{\mathcal{Y}, P_{\mathrm{e}}(y|x), \mathcal{X}\}$. If we can write

$$\mathbb{P}[\mathbf{Y}(p) = \mathbf{y}|\mathbf{X}(p) = \mathbf{x}] = \prod_{l=1}^{L} P_{\mathrm{e}}(y_l|x_l) \tag{1.3}$$

for any fixed $p \in \{1, \cdots, M\}$, then the channel is memoryless, so we have the same channel $\{\mathcal{Y}, P_{\mathrm{e}}(y_l|x_l), \mathcal{X}\}$ for each symbol number l. Moreover, for this channel, P_{e} does not depend on p. In the identification phase the biometrical data sequence $\mathbf{z}(p)$ of an unknown individual p is observed via a memoryless identification channel $\{\mathcal{Z}, P_{\mathrm{i}}(z_l|x_l), \mathcal{X}\}$. Here $\mathcal{Z}$ is the operational output alphabet. Now

$$\mathbb{P}[\mathbf{Z}(p) = \mathbf{z}|\mathbf{X}(p) = \mathbf{x}] = \prod_{l=1}^{L} P_{\mathrm{i}}(z_l|x_l). \tag{1.4}$$

1.4 Identification Capacity

Definition 1.1. *The identification capacity of a biometrical parameter is the largest value of* C_{id} *such that for any (arbitrarily small)* $\delta > 0$ *and sufficiently large* L *there exist decoders that achieve a rate* R_{id} *of*

$$R_{\mathrm{id}} \geq C_{\mathrm{id}} - \delta \tag{1.5}$$

at vanishingly small error rate $P_{\max} \leq \delta$.

Theorem 1 *The* identification capacity *of a biometrical system with IID source and independent IID channel is given by the mutual information*

$$C_{\mathrm{id}} = \mathbf{I}(\mathbf{Y}; \mathbf{Z}) \tag{1.6}$$

for arbitrary fixed p.

We use the identity $\mathbf{I}(A;B) = \mathsf{H}(A) + \mathsf{H}(B) - \mathsf{H}(A,B)$. The entropies $\mathsf{H}(\mathbf{Y})$, $\mathsf{H}(\mathbf{Z})$ and $\mathsf{H}(\mathbf{Y},\mathbf{Z})$ are computed from the probability of the 'true' biometric $\mathbf{X}$ and the transition properties over the enrolment and identification channel, where

$$\mathbb{P}[\mathbf{Y}(p) = \mathbf{y}] = \prod_{j=1}^{L} \sum_{x_j \in \mathcal{X}} Q(x_j) P_{\mathrm{e}}(y_j|x_j)$$

$$\mathbb{P}[\mathbf{Z}(p) = \mathbf{z}] = \prod_{j=1}^{L} \sum_{x_j \in \mathcal{X}} Q(x_j) P_{\mathrm{i}}(z_j|x_j)$$

$$\mathbb{P}[\mathbf{Y}(p) = \mathbf{y}, \mathbf{Z}(p) = \mathbf{z}] = \prod_{j=1}^{L} \sum_{x_j \in \mathcal{X}} Q(x_j) P_{\mathrm{e}}(y_j|x_j) P_{\mathrm{i}}(z_j|x_j).$$

Note that none of these probabilities depend on p. For an IID (memoryless) source, $\mathsf{H}(\mathbf{X}) = \sum_{l=1}^{L} \mathsf{H}(x_l) = L\mathsf{H}(x_l)$. Similarly, the entropies of $\mathbf{Y}$ and $\mathbf{Z}$ can be calculated component-wise.

1.5 Proof outline for Theorem 1

As in communication theory, the proof of the capacity theorem consists of a part that there exist rates that achieve capacity and a part that proves the non-existence of rates above C_{id} with a low average error rate.

Another important step is the random coding argument [5]. This is the observation that we can prove that the average over all randomly chosen sequences achieves an error rate that vanishes. In the development of a theoretical framework for communication systems, this was an innovative step that laid the foundation for several proofs. Here Shannon's theory apparently fits biometrics naturally, while for communication systems it was a creative, initially believed to be somewhat artificial assumption to support the derivations of bounds. As first exploited in [20], random coding is a very natural and appropriate model for biometric systems, where the source model is one of randomly generated sequences[5]. In communication systems, an engineer can choose an optimum code sequence for every potential message to be transmitted. Shannon postulated that if the engineer just picks random sequences to represent messages on average he achieves capacity, so there must be codes that do at least as good. Note that the generation of the IID biometrical data yields the randomness that makes it all work. We get the random code $\{\mathbf{y}(1), \mathbf{y}(2), \cdots, \mathbf{y}(M)\}$ by the very nature of biometrics.

[5] Yet, the distribution and statistical dependence of biometrics can be questioned.

1.5.1 Achievability

In this part we prove that for all $\delta \geq 0$ the rate $R_{\mathrm{id}} \geq \mathbf{I}(\mathbf{X}, \mathbf{Y}) - \delta$ can be achieved.

To prove that there are rates that achieve low error rates we postulate a decoder that is based on typical sequences for $\mathbf{Y}$, $\mathbf{Z}$ and $\mathbf{Y}, \mathbf{Z}$ jointly. Typical sets are explained in more detail in textbooks such as [5]. The core idea is that a random sequence is with probability 1 a typical sequence. This implies that any typical sequence has a probability that is close to $2^{-\mathsf{H}(\mathbf{Y})}$, $2^{-\mathsf{H}(\mathbf{Z})}$ or $2^{-\mathsf{H}(\mathbf{Y},\mathbf{Z})}$, respectively. More precisely, the jointly typical set $\mathcal{A}_\varepsilon$ is defined as the collection of sequences $(\mathbf{Y}, \mathbf{Z})$ such that $P_{YZ}(\mathbf{y}, \mathbf{z})$ satisfies

$$2^{-L(\mathsf{H}(y_l,z_l)+\varepsilon)} \leq P_{YZ}(\mathbf{y}, \mathbf{z}) \leq 2^{-L(\mathsf{H}(y_l,z_l)-\varepsilon)} \tag{1.7}$$

and similarly for sequences of $\mathbf{Y}$ and $\mathbf{Z}$ separately. An important property is that a sequence $\mathbf{Y}, \mathbf{Z}$ chosen randomly according to the underlying biometric statistical model is a typical sequence with probability higher than $1 - \varepsilon$.

For our proof, we postulate a decoder that generates as its output the unique index $\hat{p}$ satisfying

$$(\mathbf{y}(\hat{p}), \mathbf{z}) \in \mathcal{A}_\varepsilon. \tag{1.8}$$

If no unique $\hat{p}$ exists the decoder outputs an erasure.

Two kinds of error can occur. An error of the first kind (c.f. a false rejection if we had addressed a verification system) occurs when the enrolment sequence of the tested individual p is not jointly typical with his identification sequence resulting from the test. We define the event that the enrolment $\mathbf{Y}(p)$ and an observed identification $\mathbf{Z}$ are jointly typical as

$$E_p = \{(\mathbf{Y}(p), \mathbf{Z}) \in \mathcal{A}_\varepsilon\}$$

Without loss of generality we denote the test sequence as $p = 1$. Thus a false rejection corresponds to $\neg E_1$. An error of the second kind, c.f. a false acceptance, occurs if the enrolment sequence of some other individual $p' \neq p$ is typical with p's identification sequence. This corresponds to $E_2, E_3, ...E_M$ For errors of the first kind, the probability $\mathbb{P}[(\mathbf{Y}(p), \mathbf{Z}(p)) \notin \mathcal{A}_\varepsilon] \leq \varepsilon$ for all large enough L. For errors of the second kind, we calculate the probability that two randomly chosen sequences $\mathbf{Y}$ and $\mathbf{Z}$ match, where the sequences are not necessarily taken from a specific realization of a population but produced by a statistical process that ramdomly generates sequences over $\mathcal{X}^L$. Let $\mathbf{z}$ be the output of the identification channel that is caused by $\mathbf{X}(p)$. For all $\mathbf{y} \in \mathcal{Y}^L$ and $\mathbf{z} \in \mathcal{Z}^L$ and $p' \neq p$ we have

$$\mathbb{P}[\mathbf{Y}(p') = \mathbf{y}, \mathbf{Z}(p) = \mathbf{z}] = \prod_{l=1}^{L} \sum_{\kappa \in \mathcal{X}} Q(\kappa) P_{\mathrm{e}}(y_l|\kappa) \sum_{\lambda \in \mathcal{X}} Q(\lambda) P_{\mathrm{i}}(z_l|\lambda). \tag{1.9}$$

Using the proporties of typical sequences (e.g. Theorem 8.6.1. in [5]), the false acceptance probability of two randomly chosen sequences satisfies

$$\mathbb{P}[(\mathbf{Y}(p'), \mathbf{Z}(p)) \in \mathcal{A}_\varepsilon] \leq 2^{-\mathbf{I}(\mathbf{Y}(p);\mathbf{Z}(p))+3L\varepsilon}.$$

An error of any kind occurs with probability $\mathbb{P}[\mathrm{e}] = \mathbb{P}[\mathrm{e}|p=1] = \mathbb{P}[\neg E_1 \cup E_2 \cup E_3 \cup ... \cup E_M]$. By applying the union bound one obtains

$$\mathbb{P}[\mathrm{e}|p] = \mathbb{P}\left[\neg E_1 \cup \bigcup_{p=2}^{M} E_p\right] \leq \mathbb{P}[\neg E_1] + \sum_{p=2}^{M} \mathbb{P}[E_p]\,.$$

Hence for sufficiently large L we have

$$\mathbb{P}[\mathrm{e}] \leq \varepsilon + \sum_{i=2}^{2^{LR}} 2^{-\mathbf{I}(\mathbf{Y}(p);\mathbf{Z}(p))+3L\varepsilon} \leq \varepsilon + 2^{3L\varepsilon} 2^{-[\mathbf{I}(\mathbf{Y}(p);\mathbf{Z}(p))-LR]} \leq 2\varepsilon.$$

Thus $\mathbb{P}\Big[\hat{P} \neq p | P = p\Big]$ can be made smaller than 2ε by increasing L.

1.5.2 Converse

In this part we show that there exists no identification scheme that can identify more than $2^{\mathbf{I}(\mathbf{Y};\mathbf{Z})}$ persons with negligible error probability.

$$\mathbb{P}\Big[\hat{P} \neq P\Big] \leq \max_{p=1,\cdots,M} \mathbb{P}\Big[\hat{P} \neq p | P = p\Big]\,, \tag{1.10}$$

which we require to remain arbitrarily small. Applying Fano's inequality we get for the entropy in P, knowing the data base Y and observation $\mathbf{Z}$

$$\mathsf{H}(P|\mathsf{Y}, \mathbf{Z}) \leq 1 + \mathbb{P}\Big[\hat{P} \neq P\Big] \log_2 M. \tag{1.11}$$

Note that we did not assume any a priori distribution over the individuals that are to be identified. Let us see what happens if we assume that P is uniformly distributed over $\{1, 2, \cdots, M\}$. Using inequality (1.11) we obtain

$$\begin{aligned} \log_2 M &= \mathsf{H}(P) = \mathsf{H}(P|\mathsf{Y}) = \mathsf{H}(P|\mathsf{Y}) - \mathsf{H}(P|\mathbf{Z}, \mathsf{Y}) + \mathsf{H}(P|\mathbf{Z}, \mathsf{Y}) \\ &\leq \mathbf{I}(P; \mathbf{Z}|\mathsf{Y}) + 1 + \mathbb{P}\Big[\hat{P} \neq P\Big] \log_2 M. \end{aligned} \tag{1.12}$$

Another useful inequality is obtained as follows,

$$\begin{aligned} \mathbf{I}(P; \mathbf{Z}|\mathsf{Y}) &= \mathsf{H}(\mathbf{Z}|\mathsf{Y}) - \mathsf{H}(\mathbf{Z}|P, \mathsf{Y}) \leq \mathsf{H}(\mathbf{Z}) - \mathsf{H}(\mathbf{Z}|P, \mathsf{Y}) \\ &= \mathsf{H}(\mathbf{Z}) - \mathsf{H}(\mathbf{Z}(P)|\mathbf{Y}(P)) = \mathbf{I}(\mathbf{Z}(P); \mathbf{Y}(P)) = L\mathbf{I}(Y_l; Z_l) \end{aligned} \tag{1.13}$$

Combining (1.12) and (1.13) we get

$$\log_2 M \leq L\mathbf{I}(Y_l; Z_l) + 1 + \mathbb{P}\Big[\hat{P} \neq P\Big] \log_2 M$$

or

$$\frac{\log_2 M}{L} \leq \frac{\mathbf{I}(Y_l; Z_l) + 1/L}{1 - \mathbb{P}\Big[\hat{P} \neq P\Big]}.$$

When we take the limit $L \to \infty$, we obtain $R_{\text{id}} \leq \mathbf{I}(Y_l; Z_l)(1+\delta)$. Now as we let $\delta \to 0$ (where δ is defined in Def. 1.1), which implies $\mathbb{P}\Big[\hat{P} \neq P\Big] \to 0$, we have that $R_{\text{id}} \leq \mathbf{I}(Y; Z)$. By combination with the first part of the proof, it follows that the capacity $C_{\text{id}} = \mathbf{I}(Y; Z)$.

1.5.3 Example: Bernoulli variables

Let's consider a hypothetical biometric that gives balanced IID binary values. Let the biometric X form a Bernoulli random variable ($\mathcal{X} = \{0, 1\}$) with parameter $p = \mathbb{P}[X = 1] = 0.5$. Moreover let

$$Y = X \oplus N_{\text{e}}, \qquad Z = X \oplus N_{\text{i}}, \tag{1.14}$$

with Bernoulli noise variables N_{e} and N_{i} having parameters d_{e} and d_{i}, respectively. The addition $\oplus$ is modulo 2. Then

$$\mathbb{P}[Y_l \neq Z_l] = d = d_{\text{e}}(1 - d_{\text{i}}) + (1 - d_{\text{e}})d_{\text{i}}. \tag{1.15}$$

The mutual information per symbol is given by

$$\mathbf{I}(Y_l; Z_l) = \mathsf{H}(Z_l) - \mathsf{H}(Z_l|Y_l) \tag{1.16}$$

which yields $\mathbf{I}(\mathbf{Y}, \mathbf{Z}) = L[1 - \mathsf{h}(d)]$, with $\mathsf{h}(d)$ the binary entropy function, defined as $\mathsf{h}(d) = -d \log_2 d - (1-d)\log_2(1-d)$. Note that in this example we can conceptually think of the enrolment process as error free, and the identification process as distorted by the concatenation of the original channels $\mathbf{X} \to \mathbf{Y}$ and $\mathbf{X} \to \mathbf{Z}$, yielding a binary symmetric channel with probability of error d.

1.5.4 Example: IID Gaussian variables

As a second example, we consider the case that $\mathbf{X}$ is IID Gaussian, zero-mean, with variance σ_0^2. Moreover, for every dimension l, let

$$Y_l = X_l + N_{\text{e}}, \qquad Z_l = X_l + N_{\text{i}}, \tag{1.17}$$

with zero-mean Gaussian noise variables N_{e} and N_{i} having variances σ_{e}^2 and σ_{i}^2 respectively. The covariance matrix Σ_{YZ} is given by

$$\Sigma_{YZ} = \begin{pmatrix} \mathbb{E}[Y_l^2] - \mathbb{E}[Y_l]^2 & \mathbb{E}[Y_l Z_l] - \mathbb{E}[Y_l]\mathbb{E}[Z_l] \\ \mathbb{E}[Z_l Y_l] - \mathbb{E}[Z_l]\mathbb{E}[Y_l] & \mathbb{E}[Z_l^2] - \mathbb{E}[Z_l]^2 \end{pmatrix} = \begin{pmatrix} \sigma_0^2 + \sigma_{\text{e}}^2 & \sigma_0^2 \\ \sigma_0^2 & \sigma_0^2 + \sigma_{\text{i}}^2 \end{pmatrix}. \tag{1.18}$$

Hence, using $\mathsf{H}(Y_l, Z_l) = \frac{1}{2}\log|\det \Sigma_{YZ}|$, it follows that

$$\mathbf{I}(Y_l; Z_l) = \frac{1}{2}\log_2\left(1 + \frac{\sigma_0^2}{\sigma_\mathrm{e}^2 + \sigma_\mathrm{i}^2 + \sigma_\mathrm{e}^2\sigma_\mathrm{i}^2/\sigma_0^2}\right). \tag{1.19}$$

Note that in this example, in contrast to Section 1.5.3, the combined channel $Y_l \to X_l \to Z_l$ with $\sigma_\mathrm{e}^2 > 0$ cannot be represented as a noiseless enrolment followed by an additive Gaussian channel with some noise power depending *only* on σ_e^2 and σ_i^2. This phenomenon finds its cause in the fact that in general the backward channel of an additive channel is non-additive.

Often, enrolment can be performed under 'ideal' circumstances, or can be repeated several times to reduce the noise. Then the noise-free biometric $\mathbf{Y} = \mathbf{X}$ becomes available. In that case $C_\mathrm{id} = 1 - \mathsf{h}(d_i)$ in the example of Bernoulli variables and $C_\mathrm{id} = \frac{1}{2}\log(1 + \sigma_0^2/\sigma_\mathrm{i}^2)$ for Gaussian variables.

1.6 Hypothesis testing; maximum likelihood

We have seen before that a decoder which is based on typicality achieves capacity. Nevertheless such a decoder may not be optimal in the sense of minimizing the maximum error probability for finite L. In a more general detection theoretical setting, we may see our identification problem as a hypothesis testing procedure, i.e. a procedure that aims at achieving the best trade-off between certain error probabilities. An optimal hypothesis testing procedure is based on the likelihoods of the observed data (enrolment data and identification data) given the individual p. The *maximum likelihood decoder* selects

$$\hat{p} = \arg\max_p \mathbb{P}[\mathbf{Z}(p) = \mathbf{z} | \mathbf{Y}(p) = \mathbf{y}(p)] \tag{1.20}$$

where the observation $\mathbf{z}$ is fixed, i.e. not a function of p. For the decoder, the relevant probability can be written as

$$\mathbb{P}[\mathbf{Z}(p) = \mathbf{z} | \mathbf{Y}(p) = \mathbf{y}(p)] = \prod_{j=1}^{L} \frac{\sum_{x\in\mathcal{X}} Q(x) P_\mathrm{e}(y_j(p)|x) P_\mathrm{i}(z_j|x)}{\sum_{x\in\mathcal{X}} Q(x) P_\mathrm{e}(y_j(p)|x)}. \tag{1.21}$$

Here the sum is over all possible x, which is of modest complexity. Particularly if $\mathbf{X}$ contains IID elements, and if the channel models $P_\mathrm{e}(y_j|x)$ and $P_\mathrm{i}(z_j|x)$ are known, e.g. from Chapter 1.5.3 and 1.5.4, the complexiy of this calculation is small. Yet this decision variable has to be calculated for all p in the data base.

This illustrates how the enrolment output sequences $\mathbf{Y}(1), \mathbf{Y}(2), \cdots, \mathbf{Y}(M)$ act as as codewords. These codewords are observed via a memoryless channel $\{\mathcal{Z}, P(z|y), \mathcal{Y}\}$.

Note that decoding according to our achievability proof involves an exhaustive search procedure. It is not known how an identification scheme can be modified in such a way that the decoding complexity is decreased. However, the helper data proposed in the second half of this chapter has the side effect of accelerating the recognition process.

1.7 Private templates

Identification inherently requires that a verifier searches for matches with the measured $\mathbf{Z}$ in a data base Y that contains data about the entire population. This introduces the security and privacy threat that the verifier who steals biometric templates from some (or even all) persons in the data base can perform impersonation attacks. This threat was recognized by several researchers [3, 11, 18]. When a private verification system is used on a large scale, the reference data base has to be made available to many different verifiers, who, in general, cannot be trusted. Matsumoto et al. [12] showed that information stolen from a data base can be misused to construct artificial biometrics to impersonate people. Creation of artificial biometrics is possible even if only part of the template is available. Hill [7] showed that if only minutiae templates of a fingerprint are available, it is still possible to successfully construct artificial biometrics that pass private verification.

To develop an insight in the security aspects of biometrics, we distinguish between verification and private verification. In a typical verification situation, access to the reference template allows a malicious Victor to artificially construct measurement data that will pass the verification test, even if Peggy has never exposed herself to a biometric measurement after the enrolment.

In private verification, the reference data should not leak relevant information to allow Victor to (effectively) construct valid measurement data. Such protection is common practice for storage of computer passwords. When a computer verifies a password, it does not compare the password $\mathbf{Y}$ typed by the user with a stored reference copy. Instead, the password is processed by a cryptographic one-way function F and the outcome is compared against a locally stored reference string $F(\mathbf{Y})$. So $\mathbf{Y}$ is only temporarily available on the system hardware, and no stored data allows calculation of $\mathbf{Y}$. This prevents attacks from the inside by stealing unencrypted or decryptable secrets.

The main difference between password checking and biometric private verification is that during biometric measurements it is unavoidable that noise or other aberrations occur. Noisy measurement data are quantized into discrete values before these can be processed by any cryptographic function. Due to external noise, the outcome of the quantization may differ from experiment to experiment. In particular if one of Peggy's biometric parameters has a value close to a quantization threshold, minor amounts of noise can change the outcome. Minor changes at the input of a cryptographic function are amplified and the outcome will bear no resemblance to the expected outcome. This property, commonly refered to as 'confusion' and 'diffusion', makes it less trivial to use biometric data as input to a cryptographic function. The notion of *near matches* or *distance* between enrolment and operational measurements vanishes after encryption or any other cryptographically strong operation. Hence, the comparison of measured data with reference data can not be executed in the encrypted domain without prior precautions to contain the effect of noise.

Furthermore, with increasing M and L, the probability that a 'randomly created' natural biometric vector lies near a decision boundary goes to unity for any a priori defined digitization scheme. Error correction coding does not help, because the biometrics are generated randomly thus do not naturally lie centered inside decision regions, as codewords would do.

A common misperception is that encryption of $\mathbf{Y}$, and decryption prior to the verification solves this security threat. This would not prevent a dishonest Victor from stealing the decrypted template $\mathbf{Y}$, because Victor knows the decryption key.

The next section presents an algorithm to resolve these threats. It is based on a 'helper data scheme' that resembles 'fuzzy extractors', covered in Chapter **??**.

Besides private verification, a further application can be the generation of a secret key. We illustrate this by the example of access to a data base of highly confidential encrypted documents to which only a (set of) specific users is allowed access. The retrieval system authenticates humans and retrieves a decryption key from their biometric parameters. This system must be protected against a dishonest software programmer Mallory who has access to the biometric reference data from all users. If Mallory downloads the complete reference data file, all encrypted documents, and possibly reads all the software code of the system, she should not be able to decrypt any document.

Meanwhile, it is important to realize that protection of the reference data stored in a database is not a complete solution to the above-mentioned threats. After having had an opportunity to measure operational biometric data, a dishonest Victor uses these measurement data. This can happen without anyone noticing it: Victor grabs the fingerprint image left behind on a sensor. This corresponds to grabbing all keystrokes including the plain passwords typed by a user. We do not address this last attack in this chapter.

1.7.1 The Helper Data Architecture

We observe that a biometric private verification system does not need to store the original biometric templates. Examples of systems that use other architectures and achieve protection of templates are *private biometrics* [6], *fuzzy commitment* [9], *cancelable biometrics* [15], *fuzzy vault* [8], *quantizing secret extraction* [10] and *secret extraction from significant components* [19]. The systems proposed in [6, 8–10, 19] are all based on architectures that use helper data.

In order to combine private biometric verification with cryptographic techniques, we derive *helper data* during the enrolment phase. The helper data $\mathbf{W}$ guarantees that a unique string $\mathbf{S}$ can be derived from the biometrics of an individual during the private verification as well as during the enrolment phase. The helper data serves two purposes. On the one hand it is used to reduce the effects of noise in the biometric measurements. More precisely, it ensures that with high probability the measured noisy biometric always falls within

the same decision region taken by the detector. As a result, exactly the same string $\mathbf{S}$ is always extracted from the same p. Since the string $\mathbf{S}$ is not affected by noise anymore, it can be used as input to cryptographic primitives without causing avalanches of errors. Thus S can be handled in the same secure manner as computer passwords.

However, since random biometrics are usually not uniformly distributed, the extracted string $\mathbf{S}$ is not guaranteed to be uniform. Therefore, another part of the helper data is used to extract the randomness of the biometric measurements. Usually this is done by letting the helper data being a pointer to a randomly chosen function from a universal set of hash functions (see Chapter **??**). The left-over hash lemma guarantees that the hashed string is indistinguishable from a uniformly random string. The error correction phase is usually called *information reconciliation* and the randomness extraction the *privacy amplification* phase.

Private biometric verification consists of two phases: *enrolment* and *private verification*. During the enrolment phase, Peggy visits a Certification Authority (CA) where her biometrics are measured and reference data (including helper data) are generated. For on-line applications, such protected reference data can be stored in a central data base (possibly even publicly accessible), or these data can be certified with a digital signature of the CA, and given to Peggy. In the latter case, it is Peggy's responsibility to (securely) give this certified reference data to Victor. Thus, the reference data consists of two parts: the cryptographic key value $\mathbf{V} = F(\mathbf{S})$ against which the processed measurement data is compared, and the data $\mathbf{W}$ which assists in achieving reliable detection.

Assuming that these data are available as $\mathbf{V}(\text{Peggy}) = \mathbf{v}; \mathbf{W}(\text{Peggy}) = \mathbf{w}$, Peggy authenticates herself as follows:

- When she claims to be Peggy, she sends her identifier message to Victor.
- Victor retrieves the helper data $\mathbf{w}$ from an on-line trusted data base. Alternatively, in an off-line application Peggy could provide Victor with reference data $(\mathbf{v}, \mathbf{w})$ certified by the CA.
- Peggy allows Victor to take a noisy measurement $\mathbf{z}$ of her biometrics.
- Victor calculates $\mathbf{s}' = G(\mathbf{w}, \mathbf{z})$. Here G is a 'shielding' function, to be discussed later.
- *Optional for key establishment:* Victor can extract further cryptographic keys from $\mathbf{s}'$, for instance to generate an access key.
- Victor calculates the cryptographic hash function $\mathbf{v}' = F(\mathbf{s}')$.
- $\mathbf{v}'$ is compared with the reference data $\mathbf{v}$. If $\mathbf{v}' = \mathbf{v}$, the private verification is successful.

Here, we used lower-case $\mathbf{n}, \mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{v}, \mathbf{w}$ to explicitly denote that the protocol operates on realizations of the random variables $\mathbf{N}$, $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$, $\mathbf{V}$, and $\mathbf{W}$, respectively. The length of these vectors is denoted as $L_N, L_X, L_Y, L_Z, L_V, L_W$. Often, the same type of measurement is done for enrolment and for verification, thus $L_Y = L_Z$ and $\mathcal{Y} = \mathcal{Z}$. Further, $\mathbf{S}$ and $F(\mathbf{S})$ are discrete-valued

(typically binary) vectors of length L_S and L_F, resp. Note that here we make an exact match. Checking for imperfect matches would not make sense because of the cryptographic operation F. Measurement imperfections (noise) are eliminated by the use of $\mathbf{W}$ and the so-called δ-contracting property of the shielding function G.

1.7.2 Definitions

During enrolment, $\mathbf{Y}(\text{Peggy}) = \mathbf{y}$ is measured. Some secret $\mathbf{S}(\text{Peggy}) = \mathbf{s} \in \mathcal{S}^{L_S}$ is determined, and the corresponding $\mathbf{V} = F(\mathbf{s}) \in \mathcal{V}^{L_V}$ is computed. In later sections, we will address whether $\mathbf{s}$ can be chosen arbitrarily ($\mathbf{s} \in \mathcal{S}^{L_s}$) by the CA, or that the enrolment algorithm explicitly outputs one specific value of $\mathbf{s}$ based on $\mathbf{y}$. Also, a value for $\mathbf{W}(\text{Peggy}) = \mathbf{w}$ is calculated such that not only $G(\mathbf{w}, \mathbf{y}) = \mathbf{s}$ but also during private verification $G(\mathbf{w}, \mathbf{z}) = \mathbf{s}$ for $\mathbf{z} \approx \mathbf{x}$, more precisely for distances $d(\mathbf{z}, \mathbf{x}) \leq \delta$. We call a function that supports this property δ-contracting.

Definition 1.2. *Let $G : \mathcal{W}^{L_w} \times \mathcal{Y}^{L_y} \to \mathcal{S}^{L_S}$ be a function and $\delta > 0$ be a nonnegative real number. The function G is called δ-contracting if and only if for all $\mathbf{y} \in \mathcal{Y}^{L_Y}$ there exists (an efficient algorithm to find) at least one vector $\mathbf{w} \in \mathcal{W}^{L_W}$ and one $\mathbf{s} \in \mathcal{S}^{L_S}$ such that $G(\mathbf{w}, \mathbf{y}) = G(\mathbf{w}, \mathbf{z}) = \mathbf{S}$ for all $\mathbf{z} \in \mathcal{Y}^{L_Y}$ such that $d(\mathbf{z}, \mathbf{y}) \leq \delta$.*

We now argue that helper data is an essential attribute to make a secure biometrics system. We show this by contradiction, namely by initially assuming that $\mathbf{W}$ does not depend on p.

Theorem 1.1. *If $G(\mathbf{w}, \mathbf{z}) = f(\mathbf{z})$ for all $\mathbf{w}$, then either the largest contracting range of G is $\delta = 0$ or $G(\mathbf{w}, \mathbf{z})$ is a constant independent of $\mathbf{z}$.*

Proof: Take $\mathbf{W} = \mathbf{w}_0$. Assume G is δ-contracting, with $\delta > 0$. Choose two points $\mathbf{z}_1$ and $\mathbf{z}_2$ such that $G(\mathbf{w}_0, \mathbf{z}_1) = \mathbf{s}_1$ and $G(\mathbf{w}_0, \mathbf{z}_2) = \mathbf{s}_2$. Define a vector $\mathbf{r} = \lambda(\mathbf{z}_2 - \mathbf{z}_1)$ such that $0 < d(\mathbf{0}, \mathbf{r}) < \delta$. Then $\mathbf{s}_1 = G(\mathbf{w}_0, \mathbf{z}_1) = G(\mathbf{w}_0, \mathbf{z}_1 + \mathbf{r}) = G(\mathbf{w}_0, \mathbf{z}_1 + 2\mathbf{r}) = ... = \mathbf{s}_2$. Thus $G(\mathbf{w}_0, \mathbf{z}_1) = G(\mathbf{w}_0, \mathbf{z}_2)$ is constant. □

Corrolary: The desirable property that biometric data can be verified in the encrypted domain (in an information theoretic sense) cannot be achieved unless person-specific data $\mathbf{W}$ is used. Private biometric verification that attempts to process $\mathbf{Z}$ without such helper data is doomed to store decryptable user templates.

Any function is 0-contracting. If the radius δ is properly chosen as a function of the noise power, the δ-contracting property ensures that despite the noise, for a specific Peggy all likely measurements $\mathbf{Z}$ will be mapped to the same value of $\mathbf{S}$. This can particularly be guaranteed if $L \to \infty$. For private verification schemes with large $L_Y = L_Z = L$, $d(\mathbf{Z}, \mathbf{Y}) \to \sigma_n \sqrt{L}$, where σ_n^2 is the noise power. So one needs to ensure that δ is sufficiently larger than $\sigma_n \sqrt{L}$.

Definition 1.3. *Let $G : \mathcal{W}^{L_W} \times \mathcal{Y}^{L_Y} \to \mathcal{S}^{L_S}$ be a δ-contracting function with $\delta > 0$ and let $\epsilon > 0$ be a non-negative real number. The function G is called "ϵ-revealing" if and only if for all $\mathbf{y} \in \mathcal{Y}^{L_Y}$ there exists (an efficient algorithm to find) a vector $\mathbf{w} \in \mathcal{W}^{L_W}$ such that $\mathbf{I}(\mathbf{w}; G(\mathbf{w}, \mathbf{y})) < \epsilon$.*

Hence $\mathbf{W}$ conceals $\mathbf{S}$: it reveals only a well-defined, small amount of information about $\mathbf{S}$. Similarly, we require that $F(\mathbf{S})$ conceals $\mathbf{S}$. However we do not interpret this in the information-theoretic sense but in the complexity-theoretic sense, i.e., the computational effort to obtain a reasonable estimate of $\mathbf{X}$ or $\mathbf{S}$ from $F(\mathbf{S})$ is prohibitive, even though in the information theoretic sense $F(\mathbf{S})$ may (uniquely) define $\mathbf{S}$.

The above definitions address properties of the shielding function G. Efficient enrolment requires an algorithm $\Gamma(\mathbf{Y}) \to (\mathbf{W}, \mathbf{S})$ to generate the helper data and the secret. The procedure Γ is a randomized procedure and only used during enrolment.

1.7.3 Example: Quantization Indexing for IID Gaussian

Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{N}_\mathrm{e}, \mathbf{N}_\mathrm{i}$ be Gaussian variables as defined in Section 1.5.4. Moreover $L_X = L_Y = L_Z = L_W = L_S = L$, and $\mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathcal{W} = \mathbb{R}$. The core idea is that measured data are quantized. The quantization intervals are alternatingly mapped to $s_l = 0$ and $s_l = 1$. The helper data w_l acts as a bias in the biometric value to ensure detection of the correct value of s_l. This example resembles strategies known in the literature on Quantization Index Modulation (QIM) [4], which is a specific form of electronic watermarking, and on writing on dirty paper. QIM was applied to biometrics in [10].

Enrolment in QIM

During enrolment, y_l is measured and the CA generates w_l such that the value of $y_l + w_l$ is pushed to the center of the nearest quantization interval that corresponds to the correct s_l value,

$$w_l = \begin{cases} (2n + \frac{1}{2})q - y_l \text{ if } s_l = 1 \\ (2n - \frac{1}{2})q - y_l \text{ if } s_l = 0 \end{cases} \tag{1.22}$$

where $n \in \mathbb{Z}$ is chosen such that $-q < w_l < q$ and q is an appropriately chosen quantization step size. The value of n is discarded, but the values w_l are released as helper data. Fig. 1.2 illustrates the quantization.

Fig. 1.2. Quantization levels for the shielding function G defined in (1.23). The helper data w pushes y towards the center of a quantization interval (indicated by dots).

Private verification in QIM

For the l-th component of $\mathbf{Z}$, the δ-contracting function is

$$s'_l = G(w_l, z_l) = \begin{cases} 1 \text{ if } 2nq < z_l + w_l \leq (2n+1)q \text{ for any } n \in \mathbb{Z} \\ 0 \text{ if } (2n-1)q < z_l + w_l \leq nq \quad \text{for any } n \in \mathbb{Z} \end{cases} \tag{1.23}$$

The contraction range δ equals $q/2$.

Error Probability in QIM

The probability of a bit error in the component s_l in the case of an honest pair Peggy-Victor is given by

$$P_{\mathrm{e}} = 2\sum_{b=0}^{\infty} \left\{ Q([2b+\frac{3}{2}]\frac{q}{\sigma_{\mathrm{n}}}) - Q([2b+\frac{1}{2}]\frac{q}{\sigma_{\mathrm{n}}}) \right\}, \tag{1.24}$$

where $Q(x)$ is the $\int_0^x$ integral over the Gaussian pdf with unit variance, and $\sigma_{\mathrm{n}} = \sqrt{\sigma_{\mathrm{e}}^2 + \sigma_{\mathrm{i}}^2}$ is the strength of the noise $N_{\mathrm{i}} - N_{\mathrm{e}}$. An error-correcting code can be used to correct the bit errors. The maximum achievable code rate is $1 - h(P_{\mathrm{e}})$. In practice this rate is only approached for $L \to \infty$. Large values of q ensure reliable detection, because P_{e} becomes small. However, we will show now that the information leakage is minimized only if q is small.

Information leakage in QIM

Using Bayes' rule, for given w_l we can express the a posteriori probability of the event $S_l = 1$ as

$$\mathbb{P}[S_l = 1 | W_l = w_l] = \frac{f(w_l|S_l=1)}{f(w_l)} \mathbb{P}[S_l = 1]. \tag{1.25}$$

Here f is the probability density function of W. Information leaks whenever $f(w_l|S_l = 1) \neq f(w_l|S_l = 0)$. Since the pdf of X_l is not flat, some values of w_l are more likely than others even within $-q < w_l < q$. This gives an imbalance in the above a posteriori probability.

We now quantify the information leakage given our assumptions on the statistical behavior of the input signal X_l. The statistics of W_l are determined by those of X_l and S_l. We observe that for $s_l = 1$, $w_l = (2n + 1/2)q - y_l$, so

$$f(w_l|S_l=1) = \begin{cases} 0 & \text{for } |w_l| \geq q \\ \sum_{n=-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{([2n+1/2]q - w_l)^2}{2\sigma_y^2}\right) & |w_l| < q. \end{cases} \tag{1.26}$$

Here we defined $\sigma_y^2 = \sigma_0^2 + \sigma_{\mathrm{e}}^2$. An expression similar to (1.26) is obtained for $f(w_l|S_l = 0)$. We have the symmetry relations $f(w_l|S_l = s) = f(q - w_l|S_l = s)$ and $f(w_l|S_l = 0) = f(-w_l|S = 1)$ [10]. The mutual information follows from

$$\mathbf{I}(W_l; S_l) = \mathsf{H}(S_l) - \int_{-q}^{q} \mathsf{H}(S_l|W_l = w) f(w) \mathrm{d}w. \tag{1.27}$$

Using Bayes' rule, the symmetry properties of f, and the uniformity of $\mathbf{S}$, we obtain

$$\mathbf{I}(W_l; S_l) = \int_{-q}^{q} f(w|S_l = 1) \log_2 f(w|S_l = 1) \mathrm{d}w - \int_{-q}^{q} f(w) \log_2 f(w) \mathrm{d}w. \tag{1.28}$$

Fig. 1.7.3 shows that quantization values as crude as $q/\sigma_y = 1$ are sufficient to ensure small leakage ($< 10^{-4}$). Crude quantization steps are favorable as these allow reliable detection (i.e., a large contracting range).

Fig. 1.3. Mutual information $\mathbf{I}(W_l; S_l)$ as a function of the quantisation step size q/σ_y.

1.8 Secrecy and Identification Capacity

It is natural to ask what the maximum length is of the secret key that can be extract from a biometric measurement. The size of the secrets is expressed as the rate R_s, expressed as the effective key size in bits per biometric symbol (entropy bits / symbol). The maximum achievable rate is defined accordingly by the secrecy capacity C_s.

Definition 1.4 (Secrecy Capacity). *The secrecy capacity C_s is the maximal rate R_s, such that for all $\epsilon > 0$, there exist encoders and decoders that, for sufficiently large L_x, achieve*

$$\mathbb{P}[\mathbf{S}' \neq \mathbf{S}] \leq \epsilon, \tag{1.29}$$

$$\mathbf{I}(\mathbf{W}; \mathbf{S}) \leq \epsilon, \tag{1.30}$$

$$\frac{1}{L}\mathsf{H}(\mathbf{S}) \geq (R_\mathrm{s} - \epsilon). \tag{1.31}$$

Eq. (1.29) ensures correctness of the secret, Eq. (1.30) ensures secrecy with respect to eavesdropping of the communication line and Eq. (1.31) guarantees high entropy in the secret. Eq. (1.31) is a stronger requirement than versatility.

If $\mathbf{I}(\mathbf{W}; \mathbf{S})$ is small and H is large, an impersonation attack based on artificial biometrics that pass an private verification. We remark that in general $\mathbf{I}(\mathbf{V}, \mathbf{W}; \mathbf{X})$ is large in the strict information-theoretic sense. In the computational sense, however, it is infeasible to derive information about $\mathbf{S}$ from

V. Hence from a computational point of view **V** does not reveal information about **X**.

The uncertainty expressed by $\mathsf{H}(\mathbf{S}|\mathbf{W}) = \mathsf{H}(\mathbf{S}) - \mathbf{I}(\mathbf{W};\mathbf{S})$ defines a security parameter κ for impersonation. It gives the number of attempts that have to be performed in order to achieve successful impersonation.

In order to compute the secrecy capacity the following lemma is needed, which we present here for the sake of completeness.

Lemma 1.1. *For continuous random variables X, Y and $\epsilon > 0$, there exists a sequence of discretized random variables X_d, Y_d that converge pointwise to X, Y (when $d \to \infty$) such that for sufficiently large d,*

$$\mathbf{I}(X;Y) \geq \mathbf{I}(X_d;Y_d) \geq \mathbf{I}(X;Y) - \epsilon. \tag{1.32}$$

With some modifications to the results from [2,13], the following theorem can be proven using lemma 1.1.

Theorem 1.2. *The secrecy capacity of a biometric system equals*

$$C_{\mathrm{s}} = \mathbf{I}\left(\mathbf{Y}(p);\mathbf{Z}(p)\right). \tag{1.33}$$

Proof. We start with the achievability argument. The proof that $\mathbf{I}(\mathbf{Y};\mathbf{Z})$ can be achieved if **Y** and **Z** are discrete variables, is analogous to the proof in [2]. In order to prove achievability in the continuous case, we choose $\epsilon \geq 0$, and approximate the random variables **Y**, **Z** by discretized (quantized) versions, $\mathbf{Y}_d, \mathbf{Z}_d$ such that $\mathbf{I}(\mathbf{Y};\mathbf{Z}) - \mathbf{I}(\mathbf{Y}_d;\mathbf{Z}_d) \leq \epsilon$. (The fact that such a quantization exists follows from lemma 1.2). Then, taking the encoder that achieves the capacity for the discrete case $(\mathbf{Y}_d, \mathbf{Z}_d)$ it follows that we can achieve $\mathbf{I}(\mathbf{Y}_d;\mathbf{Z}_d)$. Since this can be done for any $\epsilon \geq 0$ the proof follows.

The fact that $\mathbf{I}(\mathbf{Y};\mathbf{Z})$ is an upper bound for C_{s} for discrete random variables, follows from the Fano inequality and some basic entropy inequalities. For the continuous case this follows again by an approximation argument using Lemma 1.1. □

It was proven in [17] that there exists a biometric private verification algorithm that achieves both the secrecy capacity C_{s} and the identification capacity C_{id} at the same time.

1.8.1 Identification Capacity, Revisited

We have derived the secrecy capacity for secure private verification systems with helper data. Yet, the identification capacity was up to this section only established for systems without helper data. In this section we show that the identification capacity is equal to the channel capacity of the biometric sensor if helper data and shielding functions are applied.

Definition 1.5 (Identification Capacity). *The identification capacity* C_{id} *is the maximal rate* R_{id}*, such that for every* $\epsilon > 0$*, for sufficiently large* L*, there exists an identification strategy that achieves*

$$\text{avg } \mathbb{P}[\hat{P} \neq P] \leq \epsilon, \quad \textit{and } \frac{1}{L} \log M \geq R_{\mathrm{id}} - \epsilon, \tag{1.34}$$

where the average is over all individuals and over all random realizations of all biometrics.

A private verification scheme with helper data can be used for identification: for a biometric measurement $\mathbf{y}$ the verifier performs an exhausive search over the entire population of candidates $p' \in \{1, \ldots, M\}$ by retrieving from a database the values $\mathbf{w}, \mathbf{v}$ for each candidate and checking if $F(G(\mathbf{w}, \mathbf{y})) = \mathbf{v}$. In practice such a system can be computationally more efficient than a straightforward identification scheme mentioned in the first part of this chapter: it does not need to consider *near* matches, but only exact matches. The exact matches are performed in the binary domain and therefore very efficient. The above definition addresses such a system.

For systems without enrolment noise, it can be shown [14, 20] that if the δ-contracting range is chosen such that it matches the sphere that verification noise creates around $\mathbf{X}$, the biometric identification systems, including template protecting systems, satisfy $C_{\mathrm{id}} = \mathbf{I}(\mathbf{X}; \mathbf{Y})$. This result can be interpreted merely as a statement that helper data does not negatively influence the performance.

1.9 Relation with Fuzzy Extractors

In this book several chapters deal with key extraction from noisy data in general and biometrics in particular. A well-known technique that is treated in Chapter **??**, is called *Fuzzy Extractors*. Here we prove that the helper data technique developed in this chapter is equivalent to that of a Fuzzy Extractor. We need some details of the construction of a Fuzzy Extractor. For those details we refer to Chapter **??**. We define

$$\mathsf{Gen}(\mathbf{y}) = \Gamma(\mathbf{y}) \quad \text{and} \quad \mathsf{Rep}(\mathbf{z}, \mathbf{w}) = G(\mathbf{w}, \mathbf{z}).$$

With this definition, the following two theorems have been proven [1].

Theorem 1 *Suppose that there exists a* $(\mathcal{Y}, m, l, \delta, \epsilon \leq 1/4)$ *fuzzy extractor with generation and reproduction procedures* Gen *and* Rep *constructed by using a secure sketch and with* K *uniformly distributed over* $\{0,1\}^l$ *statistically independent from* (X, Y)*. Then, there exists a* δ*-contracting,* η*-revealing function* G*, with counterpart* Γ*, with*

$$\eta = h(2\epsilon) + 2\epsilon(|\mathsf{Gen}(\mathbf{y})| + |K|) + h(\epsilon) + \epsilon|K|,$$

This theorem proves that a Fuzzy Extractor implies a helper data algorithm which is η-revealing.

Furthermore, we have the following converse.

Theorem 2 *Let G be a δ-contracting ϵ-revealing function creating a uniformly random key K on $\{0,1\}^l$ wrt to a probability distribution P_{XY} with $\mathsf{H}_\infty(X) > m$. Then there exists a $(\mathcal{X}, m, l, \delta, \sqrt{\epsilon})$ fuzzy extractor.*

For the proofs of theorems 1 and 2 we refer the reader to [1].

Theorem 2 explains that a helper data algorithm leads to a Fuzzy Extractor whose key is 'only' $\sqrt{\epsilon}$ distinguishable from random if the helper data algorithm was ϵ-revealing.

Theorems 1 and 2 show that Fuzzy Extractors and Helper Data algorithms are equivalent up to parameter values.

1.10 Conclusion

We have developed an information-theoretic framework for biometrics. We described biometric identification in terms of two communication channels both having the same biometric source $\mathbf{X}$, but with the enrolment data $\mathbf{Y}$ and the operational measurement $\mathbf{Z}$ as destinations. We have shown that it is possible to derive bounds on the capacity of biometric identification systems with relatively simple methods. The main result is that capacity can be computed as the mutual information $\mathbf{I}(\mathbf{Y};\mathbf{Z})$ between a input source $\mathbf{Y}$ and an output source $\mathbf{Z}$ that are related by the concatenation of the backward enrolment channel and the forward identification channel. The base-2 logarithm of the number of persons that can be distinguished reliably is expressed as the number of symbols in the observation, multiplied by the rate R. For rates R smaller than $\mathbf{I}(\mathbf{Y};\mathbf{Z})$ this probability can also be made smaller than any $\varepsilon > 0$ by increasing L.

We showed that the secrecy capacity measures the entropy available in a key derived from the person. This result has been connected to a protocol that satisfies privacy and security requirements, in particular the protection of templates to prevent misuse by a dishonest verifier. We have introduced the notion of δ-contracting and ϵ-revealing shielding functions, where the δ-contraction describes the robsutness against noise in the biometric sensor. The ϵ-revelation description the absence of any leakage of information via publicly available templates.

The identification capacity appears to be determined by the 'channel capacity' of the biometric sensor, also for schemes that involve template protection. Similarly the entropy of a secret that can be derived from the biometric measurement depends on the channel capacity of the biometric sensor.

References

1.
2. Rudolf Ahlswede and Imre Csiszár. Common randomness in information theory and cryptography - part i: Secret sharing. *IEEE Transactions on Information Theory*, 39(4):1121–1132, 1993.
3. Rudolf M. Bolle, Jonathan Connell, Sharathchandra Pankanti, Nalini K. Ratha, and Andrew W. Senior. Biometrics 101. Report RC22481, IBM Research, 2002.
4. Brian Chen and Gregory W. Wornell. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Transactions on Information Theory*, 47(4):1423–1443, 2001.
5. Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* Wiley series in telecommunications. Wiley interscience, New York, 1991.
6. George I. Davida, Y. Frankel, and B.J. Matt. On enabling secure applications through off-line biometric identification. In *IEEE 1998 Symposium on Research in Security and Privacy*, pages 148–157, May 1998.
7. Christopher J. Hill. Risk of masquerade arising from the storage of biometrics. thesis, Australian National University, `http://chris.fornax.net/download/thesis/thesis.pdf`, November 2001.
8. Ari Juels and Madhu Sudan. A fuzzy vault scheme. In *IEEE International Symposium on Information Theory*, page 408, Lausanne, Switzerland, 2002.
9. Ari Juels and Martin Wattenberg. A fuzzy commitment scheme. In *ACM Conference on Computer and Communications Security*, pages 28–36, 1999.
10. Jean-Paul M. G. Linnartz and Pim Tuyls. New shielding functions to enhance privacy and prevent misuse of biometric templates. In *4th International Conference on Audio- and Video-Based Biometric Person Authentication*, pages 393–402, 2003.
11. Davide Maltoni, Dario Maio, Anil K. Jain, and Salil Prabhakar. *Handbook of Fingerprint Recognition.* Springer-Verlag, New-York, 2003.
12. Tsutomu Matsumoto, Hiroyuki Matsumoto, Koji Yamada, and Satoshi Hoshino. Impact of artificial "gummy" fingers on fingerprint systems. In R. L. van Renesse, editor, *Proc. SPIE Vol. 4677, Optical Security and Counterfeit Deterrence Techniques IV*, pages 275–289, April 2002.
13. Ueli M. Maurer and Stefan Wolf. Information-theoretic key agreement: From weak to strong secrecy for free. In *EUROCRYPT 2000: International Conference on the Theory and Application of Cryptographic Techniques*, volume 1807 of *Lecture Notes in Computer Science*.
14. Joseph A. O'Sullivan and Natalia A. Schmid. Large deviations performance analysis for biometrics recognition. In *Proc. of the 40th Allerton Conference*, 2002.
15. Nalini K. Ratha, Jonathan H. Connell, and Ruud M. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM Systems Journal*, 40(3):614–634, 2001.
16. Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, July and October.
17. Pim Tuyls and Jasper Goseling. Capacity and examples of template-protecting biometric authentication systems. In *ECCV Workshop BioAW*, pages 158–170, 2004.

18. Ton van der Putte and Jeroen Keuning. Biometrical fingerprint recognition: Don't get your fingers burned. In *IFIP TC8/WG8.8 Fourth Working Conference on Smart Card Research and Advanced Applications*, pages 289–303. Kluwer Academic Publishers, 2000.
19. Evgeny Verbitskiy, Pim Tuyls, Dee Denteneer, and Jean-Paul M. G. Linnartz. Reliable biometric authentication with privacy protection. In *24th Benelux WIC Symposium on Information Theory*, pages 125–131, Veldhoven, the Netherlands, May 2003.
20. Frans M. J. Willems, A. A. C. M. Kalker, Jasper Goseling, and Jean-Paul M. G. Linnartz. On the capacity of a biometrical identification system. In *IEEE International Symposium on Information Theory*, page 82, Yokohama, Japan, June 2003.