On the Capacity of a Biometrical Identification System

Frans Willems, Ton Kalker, Stan Baggen and Jean-Paul Linnartz

Philips Research Laboratories, Eindhoven

Abstract

We investigate the fundamental properties of a biometrical identification system. More specially we focus on finding the capacity of such a system, i.e. a measure for the number of individuals that can be reliably identified. We show that this capacity can be computed using standard information theoretic concepts.

1 Introduction

Recent years have shown an increasing awareness of the importance of security in our highly technical society. This interest covers a wide range of topics, from measures for airport security to the protection of digital multimedia content on pre-pressed disks.

The revival of biometrical identification as a relevant research topic fits in this general trend. The objective of a biometrical identification system is to identify individuals on the basis of physical (passive or active) features. One of the oldest and probably best known of such features is the *human fingerprint*. One can safely say that for a long time fingerprint-based identification and biometrical identification have been seen as one and the same thing. The last decade other human features have become practical, and there is now an active research community on iris-based recognition, face recognition, voice recognition and others. A good overview of the general biometrical identification systems, their pros and their cons, can be found in [4].

Very recently the use of biometrical identification methods has been extended to include physical device identification. For example, in the context of the SEARCH project at MIT, Physical Unknown Functions (PUFS) are studied for this purpose [3].

Biometrical identification in general involves two phases. In an enrollment phase all individuals are observed and for each individual a record is added to a database. This record contains enrollment-data, i.e. a noisy version of the biometrical data corresponding to the individual. In the identification phase an unknown individual is observed again. The resulting identification-data, another noisy version of the biometrical data of the unknown individual, is compared to (all) the enrollment-data in the database and the system has to come up with an estimate of the individual. Essential in this procedure is that both in the enrollment-phase and in the identification-phase noisy versions of the biometrical data are obtained. The actual biometrical data of each individual remain unknown.

We are interested in knowing how many individuals can reliably be identified by a biometrical identification system as a function of the amount of observed data and the quality of the observations. We first describe a model for the enrollment and identification procedure. Within this model our question has an answer in terms of an information-theoretical quantity.

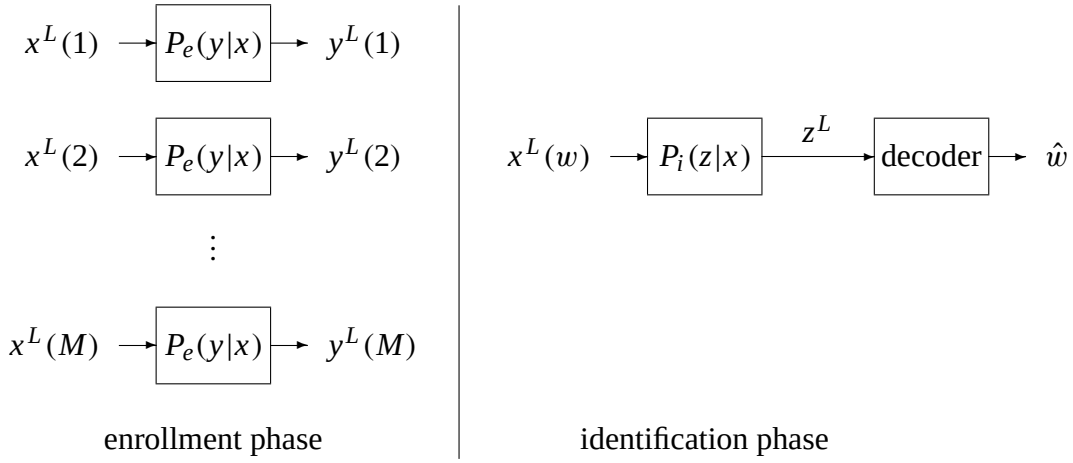

Figure 1: Model of a biometrical identification system.

2 Model description

We assume that there are M individuals. Each individual has an index $w \in \{1, 2, \cdots, M\}$. To each individual there corresponds a biometrical data sequence $x^L = (x_1, x_2, \cdots, x_L)$ with components (symbols) x_l for $l = 1, L$ that assume a value from an alphabet $\mathcal{X}$. The sequence $x^L(w)$ is the sequence for individual w for $w \in \{1, 2, \cdots, M\}$. All these sequences are supposed to be *generated at random*. The corresponding probability distribution is

$$\Pr\{X^L(w) = x^L\} = \prod_{l=1,L} Q(x_l), \text{ for all } x^L \in \mathcal{X}^L, \tag{1}$$

for all individuals $w \in \{1, 2, \cdots, M\}$. Hence each biometrical data sequence is supposed to be generated by an independent identically distributed (i.i.d.) source according to symbol distribution $\{Q(x) : x \in \mathcal{X}\}$.

In the enrollment phase all biometrical data sequences are observed via a memoryless enrollment channel $\{\mathcal{Y}, P_e(y|x), \mathcal{X}\}$. Here $\mathcal{Y}$ is the enrollment output-alphabet. Now

$$\Pr\{Y^L(w) = y^L | X^L(w) = x^L(w)\} = \prod_{l=1,L} P_e(y_l|x_l(w)) \text{ for all } y^L = (y_1, y_2, \cdots, y_L) \in \mathcal{Y}^L, \tag{2}$$

for individuals $w \in \{1, 2, \cdots, M\}$ with biometrical data sequences $x^L(w) = (x_1(w), x_2(w), \cdots, x_L(w))$. The resulting enrollment output sequences $y^L(w)$ for $w \in \{1, 2, \cdots, M\}$ are all stored in a database. Hence it is possible to access $y^L(w)$ from the database for all these w.

In the identification phase the biometrical data sequence $x^L(w)$ of an unknown individual $w \in \{1, 2, \cdots, M\}$ is observed via a memoryless identification channel $\{\mathcal{Z}, P_i(z|x), \mathcal{X}\}$. Here $\mathcal{Z}$ is the identification output alphabet. Now

$$\Pr\{Z^L = z^L | X^L(w) = x^L(w)\} = \prod_{l=1,L} P_i(z_l|x_l(w)) \text{ for all } z^L = (z_1, z_2, \cdots, z_L) \in \mathcal{Z}^L. \tag{3}$$

The resulting identification output sequence z^L is used by a decoder that also has access to all enrollment output sequences $y^L(1), y^L(2), \cdots, y^L(M)$ which were stored in the database. This decoder produces an estimate of the index of the unknown individual, i.e.

$$\hat{w} = d\left(z^L, y^L(1), y^L(2), \cdots, y^L(M)\right). \tag{4}$$

We assume that the estimate $\hat{w} \in \{\epsilon, 1, 2, \cdots, M\}$, thus an erasure ϵ is also a valid decoder output. The two relevant system parameters are the *maximal error probability*[1]

$$P_\epsilon^{\max} \stackrel{\Delta}{=} \max_{w=1,M} \Pr\{\hat{W} \neq w | W = w\} \tag{5}$$

and the *rate*

$$R \stackrel{\Delta}{=} \frac{1}{L} \log_2 M. \tag{6}$$

3 Statement of result

We say that the capacity of a biometrical system is C if for any $\delta > 0$ there exist, for all large enough L, decoders that achieve

$$\begin{aligned} \frac{1}{L} \log_2 M &\geq C - \delta, \\ P_\epsilon^{\max} &\leq \delta. \end{aligned} \tag{7}$$

Theorem 1 *The capacity of a biometrical identification system is given by*

$$C = I(Y; Z), \tag{8}$$

where $P(y, z) = \sum_{x \in \mathcal{X}} Q(x) P_e(y|x) P_i(z|x)$ *for all* $y \in \mathcal{Y}$ *and* $z \in \mathcal{Z}$.

4 Outline of proof

4.1 Achievability

The decoder is based on typicality, see Cover and Thomas [1]. More precisely, the output of the decoder is the unique $\hat{w}$ satisfying

$$(y^L(\hat{w}), z^L) \in \mathcal{A}_\varepsilon^L(Y, Z). \tag{9}$$

If no unique $\hat{w}$ exists the decoder outputs an erasure. The typical set $\mathcal{A}_\varepsilon^L(Y, Z)$ is based on distribution $P(y, z) = \sum_{x \in \mathcal{X}} Q(x) P_e(y|x) P_i(z|x)$ for $y \in \mathcal{Y}$, $z \in \mathcal{Z}$. Parameter $\varepsilon > 0$.

Two kinds of error can occur. An error of the first kind occurs when the enrollment sequence of the tested individual is not typical with the identification sequence resulting from the test. An error of the second kind occurs if the enrollment sequence of some other individual is typical with this identification sequence.

In the outline of the proof below we address errors of the first kind in (A), errors of the second kind in (B) and (C). Part (D) connects our result to the random coding argument.

(A) Note that now for all $y^L \in \mathcal{Y}^L$ and $z^L \in \mathcal{Z}^L$

$$\Pr\{Y^L(w) = y^L, Z^L = z^L | W = w\} = \prod_{l=1,L} \sum_{x \in \mathcal{X}} Q(x) P_e(y_l|x) P_i(z_l|x). \tag{10}$$

Here z^L is the output of the identification channel that is caused by $x^L(w)$. This implies that the probability $\Pr\{(Y^L(w), Z^L) \notin \mathcal{A}_\varepsilon^L(Y, Z)\} \leq \varepsilon$ for all large enough L.

[1] The stochastic processes that play a role here are the generation of the biometrical data sequences and the transmission of these sequences over the enrollment and identification channel.

(B) Again let z^L be the output of the identification channel that is caused by $x^L(w)$. For all $y^L \in \mathcal{Y}^L$ and $z^L \in \mathcal{Z}^L$ and $w' \neq w$

$$\Pr\{Y^L(w') = y^L, Z^L = z^L | W = w\} = \prod_{l=1,L} \sum_{x' \in \mathcal{X}} Q(x') P_e(y_l|x') \sum_{x \in \mathcal{X}} Q(x) P_i(z_l|x). \tag{11}$$

Therefore $\Pr\{(Y^L(w'), Z^L) \in \mathcal{A}_\varepsilon^L(Y, Z)\} \leq 2^{-L[I(Y;Z)-3\varepsilon]}$.

(C) The union bound now yields that for $M = 2^{L[I(Y;Z)-4\epsilon]}$ the error probabilities $\Pr\{\hat{W} \neq w | W = w\}$ can be made smaller than 2ε by increasing L.

(D) Note that the generation of the biometrical data yields the randomness that makes it all work. We get the random code $\{y^L(1), y^L(2), \cdots, y^L(M)\}$ for free!

4.2 Converse

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 W is uniformly distributed over $\{1, 2, \cdots, M\}$. Then

$$\Pr\{\hat{W} \neq W\} \leq \max_{w=1,M} \Pr\{\hat{W} \neq w | W = w\}, \tag{12}$$

and now we can apply Fano's inequality

$$H(W|Y^L(1), Y^L(2), \cdots, Y^L(M), Z^L) \leq 1 + \Pr\{\hat{W} \neq W\} \log_2 M. \tag{13}$$

Now

$$\begin{aligned} \log_2 M &= H(W) \\ &= H(W|Y^L(1), Y^L(2), \cdots, Y^L(M)) \\ &\leq I(W; Z^L|Y^L(1), Y^L(2), \cdots, Y^L(M)) + 1 + \Pr\{\hat{W} \neq W\} \log_2 M. \end{aligned} \tag{14}$$

Another step

$$\begin{aligned} I(W; Z^L|Y^L(1), Y^L(2), \cdots, Y^L(M)) &\leq H(Z^L) - H(Z^L|Y^L(W), W) \\ &= L H(Z) - L H(Z|Y) = L I(Y; Z). \end{aligned} \tag{15}$$

These are the main steps in the converse.

5 Likelihoods, hypothesis testing

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 that it minimizes the maximum error probability. In a more general setting we may see our problem as an 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 (enrollment data and identification data) given the individual $w \in \{1, 2, \cdots, M\}$. For individual w this likelihood can be expressed as

$$\begin{aligned} P(y^L(1), y^L(2), \cdots, y^L(w), \cdots, y^L(M), z^L|w) &= \prod_{w'=1,M} P(y^L(w') \frac{P(y^L(w), z^L|w)}{P(y^L(w)|w)} \\ &= C \cdot \frac{P(y^L(w), z^L|w)}{P(y^L(w)|w)}. \end{aligned} \tag{16}$$

The relevant factor (the factor that depends on w) can now be written as

$$\Pr\{Z^L = z^L | Y^L(w) = y^L, W = w\} = \prod_{l=1,L} \frac{\sum_{x\in\mathcal{X}} Q(x) P_e(y_l|x) P_i(z_l|x)}{\sum_{x\in\mathcal{X}} Q(x) P_e(y_l|x)}. \tag{17}$$

This confirms the fact that the enrollment output sequences $y^L(1), y^L(2), \cdots, y^L(M)\}$ can be viewed as codewords. These codewords are observed via a memoryless channel $\{\mathcal{Z}, P(z|y), \mathcal{Y}\}$.

6 Examples

As a first simple example assume that X is Bernoulli with parameter $p = \Pr\{X = 1\} = 0.5$. Moreover let

$$\begin{aligned} Y &= X + N_e, \\ Z &= X + N_i, \end{aligned} \tag{18}$$

with Bernoulli noise variables N_e and N_i with parameters d_e and d_i, respectively. Addition is modulo-2. Then

$$I(Y; Z) = 1 - H(d), \tag{19}$$

with $d = d_e(1 - d_i) + (1 - d_e)d_i$. Therefore the "channel" between Y and Z is a binary symmetric channel with transition probability d and uniform input on Y. Note that in this example we can conceptually think of the enrollment process as error free, and the identification process as distorted by the concatenation of the original channels $X \to Y$ and $X \to Z$, yielding a binary symmetric channel with probability of error d.

As a second more interesting example consider the case that X is Gaussian, zero-mean, with variance P. Moreover let

$$\begin{aligned} Y &= X + N_e, \\ Z &= Z + N_i, \end{aligned} \tag{20}$$

with zero-mean Gaussian noise variables N_e and N_i having variances σ_e^2 and σ_i^2 respectively. Then

$$I(Y; Z) = \frac{1}{2}\log_2 \frac{(P + \sigma_e^2)(P + \sigma_i^2)}{(P + \sigma_e^2)(P + \sigma_i^2) - P^2} = \frac{1}{2}\log_2 \left(1 + \frac{P}{\sigma_e^2 + \sigma_i^2 + \sigma_e^2\sigma_e^2/P}\right). \tag{21}$$

Note that in this example, in contrast to the first one, we cannot model the enrollment process as error free, and the identification process as an additive Gaussian channel with some noise variance σ^2 depending *only* on σ_e^2 and σ_i^2. One easily sees that this phenomenon find its cause in the fact that in general the backward channel of an additive channel is non-additive.

7 Conclusion and remarks

We have shown that it is possible to derive bounds on the capacity of biometric identification systems with relatively simple methods. The main result of this paper is that capacity can be computed as the mutual information between an input source and an output source that are related by the concatenation of the backward enrollment channel and the forward identification channel.

We have not considered the probability that an individual, that did not undergo the enrollment procedure, is identified as one of the individuals that did enroll properly. For rates R smaller than $I(Y;Z)$ this probability can also be made smaller than any $\varepsilon > 0$ by increasing L.

Note that decoding according to our achievability proof involves an exhaustive search procedure. It is not known how the system should be modified in such a way that the decoding complexity is decreased. This paper did not construct such 'real' biometric codes and this remains a topic of future research.

The problem that we have investigated here suggests that we should increase the block-length L to achieve capacity. However in practise we are more interested in achieving a small error probability for a given number of individuals than to achieve capacity. Still the capacity that we have determined here is a fundamental limit that tells us what we can expect from a certain system.

References

[1] T.M. Cover and J.A. Thomas, *Elements of Information Theory*. Wiley, New York, 1991.

[2] R.G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, 1968.

[3] B. Gassend, D. Clake, M. van Dijk and S. Devadas, *Controlled Physical Unknown Functions: Applications to Secure Smartcards and Certified Execution*, Technical Report MIT-LCS-TR-845 (http://www.lcs.mit.edu/publications), MIT, June, 2002.

[4] S. Pankanti, R. M. Bolle and A. Jain, *Biometrics-The Future of Identification* , IEEE Computer, Volume 33, No. 2, pp. 46 – 49, February, 2002.