Information-Theoretic Approach to Privacy Protection of Biometric Templates

Jasper Goseling[1] and Pim Tuyls
Philips Research, Prof. Holstlaan 4, 5656 AA Einhoven, The Netherlands.
j.goseling@ieee.org, pim.tuyls@philips.com

I. Introduction

The increasing demand for more reliable and convenient security systems generates a renewed interest in human identification based on biometric identifiers. We identify three security and privacy threats that arise from the use of biometrics: i) Impersonation by means of artificial biometrics based on stolen templates. ii) Compromised biometrics, are compromised forever. iii) Biometric data contain sensitive information about an individual.

In this paper, we present a general algorithm that guarantees privacy protection of biometric templates. Some special examples of this algorithm have been studied in [1, 2, 3] and a more general approach is given in [5, 6]. These systems are all based on the use of a one-way transform and *helper data*, which is used to achieve noise robustness at the input of the one-way transform. We analyze the principle behind all these systems, identify fundamental performance bounds and propose an algorithm that achieves these bounds.

II. Model and Definitions

The biometric templates obtained during enrollment are described by sequences of n i.i.d random variables belonging to some alphabet $\mathcal{X}$ which can be either continuous or discrete. A hashed form of the template is stored together with helper data in a database that is vulnerable to attacks from the outside as well as from the inside (malicious verifier). During the authentication phase an attacker is able to present artificial biometrics at the sensor. Authentication measurements are modelled as observations through a noisy memoryless channel and its results are described by a random vector $Y^n \in \mathcal{Y}^n$.

Definition 1 (Secret Extraction Code (SEC)) *Let $n, \epsilon > 0$. An $(n, |\mathcal{S}|, \epsilon)$ Secret Extraction Code $\mathcal{C}$, defined on $\mathcal{X}^n \times \mathcal{Y}^n$, is an ordered set of pairs of encoding and decoding regions*

$$\mathcal{C} = \left\{ (\mathcal{E}_i, \mathcal{D}_i) \,\middle|\, i = 1, 2, \ldots, |\mathcal{S}| \right\}, \tag{1}$$

where $\mathcal{E}_i \subseteq \mathcal{X}^n$ and $\mathcal{D}_i \subseteq \mathcal{Y}^n$, such that $\mathcal{E}_i \cap \mathcal{E}_j = \emptyset$, $\mathcal{D}_i \cap \mathcal{D}_j = \emptyset$, $\bigcup_i \mathcal{D}_i = \mathcal{Y}^n$, for $i, j = 1, 2, \ldots, |\mathcal{S}|$, $i \neq j$ and $P_{Y^n|X^n}(\mathcal{D}_i|x_i^n) \geq 1 - \epsilon$, for all $x_i^n \in \mathcal{E}_i$ and $i = 1, 2, \ldots, |\mathcal{S}|$.

III. Secure Biometric Authentication Algorithm

We present the Secure Biometric Authentication (SBA) algorithm that generalizes the examples given in [1, 2, 3].

Let $\mathcal{C}$ be a finite collection of **SEC**s on $\mathcal{X}^n \times \mathcal{Y}^n$. The collection of **SEC**s is made available in both the enrollment and the authentication phase. Furthermore, for $x^n \in \mathcal{X}^n$ we define $\Phi_{x^n} \subseteq \mathcal{C}$ as follows. A **SEC** $C = \{(\mathcal{E}_i, \mathcal{D}_i)\}_{i=1}^{|\mathcal{S}|} \in \Phi_{x^n}$ iff $x^n \in \bigcup_i \mathcal{E}_i$. The collection $\mathcal{C}$ is used as follows.

[1]Also with University of Twente, the Netherlands.

Enrollment The following steps are performed for all users $m = 1, 2 \ldots, M$:

i) The biometrics x_m^n of user m are measured.

ii) Choose a **SEC** C at random in Φ_{x^n}. Define the helper data w_m as the index of this **SEC** C. If $\Phi_{x^n} = \emptyset$, a **SEC** is selected at random from $\mathcal{C}$.

iii) Given a **SEC** $C = \{(\mathcal{E}_i, \mathcal{D}_i)\}_{i=1}^{|\mathcal{S}|}$, the secret s_m is defined as follows, $s_m = i$ if $x^n \in \mathcal{E}_i$. For $\Phi_{x^n} = \emptyset$, s_m is chosen at random.

iv) A one-way function F is applied to s_m. The data $F(s_m)$ and w_m are stored in the database.

Authentication Given an identity claim m and a measurement y^n of the biometrics of a user, the following steps are taken:

i) The database information $F(s_m)$ and the helper data w_m are retrieved.

ii) The **SEC** $C(w_m)$ is used to derive a secret v_m as follows $v_m = i$ if $y^n \in \mathcal{D}_i$.

If $F(v_m) = F(s)$ the user is positively authenticated.

In order to prevent impersonation we require that W gives no information on S. The secrecy capacity C_s is defined as the maximal rate $R_s = \frac{1}{n} \log |\mathcal{S}|$ such that for all $\epsilon > 0$ and sufficiently large n there exists algorithms that achieve i) $\Pr\{V \neq S\} \leq \epsilon$, ii) $\mathbf{I}(W; S) < \epsilon$ and iii) $\frac{1}{n}\mathsf{H}(S) \geq R_s - \epsilon$.

Theorem 1 *The secrecy capacity of a biometric authentication system is given by, $\mathsf{C}_s = \mathbf{I}(X; Y)$.*

Let $R_{\mathrm{id}} = \frac{1}{n} \log M$ and define C_{id} the identification capacity according to [4].

Theorem 2 *The SBA-algorithm meets the necessary security requirements i), ii) and iii) given above and achieves $R_s = \mathsf{C}_s$ and $R_{\mathrm{id}} = \mathsf{C}_{\mathrm{id}} = \mathbf{I}(X; Y)$.*

A more in-depth treatment of this work is given in [6].

References

[1] J.P. Linnartz and P. Tuyls, "New Shielding Functions to Enhance Privacy and Prevent Misuse of Biometric Templates," AVBPA Guildford 2003, *LNCS*2688.

[2] E. Verbitskiy, P. Tuyls, J.P. Linnartz and D. Denteneer, "Reliable Biometric Authentication with Privacy Protection", *P*roc. of the 24th Symposium on Information Theory in the Benelux, Veldhoven, The Netherlands (2003), 125–132.

[3] A. Juels and M. Wattenberg, "A Fuzzy Commitment Scheme", *S*ixth ACM Conf. on Comp. and Comm. Sec., (1999), 28–36.

[4] F. Willems, T. Kalker, J. Goseling and J.-P. Linnartz, "On the Capacity of a Biometrical Identification System", *Proc. 2003 IEEE Int. Symp. Inform. Theory*, Japan, 2003.

[5] Y. Dodis, L. Reyzin and A. Smith, "Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data ", accepted at Eurocrypt 2004.

[6] P. Tuyls and J. Goseling, "Capacity and Examples of Template-Protecting Biometric Authentication Systems", accepted at *BIOAW 2004*, Prague, Czech Republic, 2004.