Fundamental limits for biometric identification with a database containing protected templates

Citation for published version (APA):

Ignatenko, T., & Willems, F. M. J. (2010). Fundamental limits for biometric identification with a database containing protected templates. In *Proceedings of the 2010 International Symposium on Information Theory and its Applications (ISITA), 12-15 December 2010, Seattle, Washington* (pp. 54-59). Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/ISITA.2010.5649707>

DOI:

[10.1109/ISITA.2010.5649707](https://doi.org/10.1109/ISITA.2010.5649707)

Document status and date:

Published: 01/01/2010

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Fundamental Limits for Biometric Identification with a Database Containing Protected Templates

Tanya Ignatenko

Electrical Engineering Department
Eindhoven University of Technology
Eindhoven, The Netherlands
Email: t.ignatenko@tue.nl

Frans M.J. Willems

Electrical Engineering Department
Eindhoven University of Technology
Eindhoven, The Netherlands
Email: f.m.j.willems@tue.nl

Abstract—In this paper we analyze secret generation in biometric identification systems with protected templates. This problem is closely related to the study of the biometric identification capacity of Willems et al. 2003 and O’Sullivan and Schmid 2002 and the common randomness generation of Ahlswede and Csiszár 1993. In our system two terminals observe biometric enrollment and identification sequences of a number of individuals. It is the goal of these terminals to form a common secret for the sequences that belong to the same individual by interchanging public (helper) messages for all individuals in such a way that the information leakage about the secrets from these helper messages is negligible. It is important to realize that biometric data are unique for individuals and cannot be replaced if compromised. Therefore the helper messages should contain as little as possible information about the biometric data. On the other hand, the second terminal has to establish the identity of the individual who presented his biometric sequence, based on the helper data produced by the first terminal. In this paper we determine the fundamental tradeoff between secret-key rate, identification rate and privacy-leakage rate in biometric identification systems.

I. INTRODUCTION

O’Sullivan and Schmid [5] and Willems et al. [7] considered biometric identification systems and determined the corresponding identification capacity. They assumed storage of biometric enrollment sequences in the clear. Later Tuncel [6] analyzed the tradeoff between the capacity of a biometric identification system and the storage space (compression rate) required for the biometric templates. It should be noted that Tuncel’s method realizes a kind of privacy protection scheme. Ahlswede and Csiszár [1] introduced the concept of secrecy capacity. This notion can be regarded as the amount of common secret information that can be obtained in an authentication procedure. Helper data, or to put it differently, protected biometric templates, are crucial in this setting. Interestingly the secrecy capacity, which is equal to the mutual information between enrollment and authentication biometric sequences in the biometric setting, equals the identification capacity found by Willems et al. [7].

Important parameter of a biometric system is privacy leakage. Privacy leakage is the amount of information that is contained (leaked) about biometric enrollment sequences in the publicly available data. In [3] the fundamental tradeoff between secret-key rate and privacy-leakage rate was studied for a biometric authentication system. In the present paper we

will investigate the tradeoff between the amount of common secret information and privacy leakage that is achieved in an identification procedure with protected biometric templates. Unlike in biometric authentication systems, here we also take into account the identification rate.

In the system that we investigate in the current paper two terminals observe the enrollment and identification biometric sequences of different individuals. The first terminal forms a secret for each enrolled individual and stores corresponding helper data in a public database. These helper data on one hand facilitate reliable reconstruction of the secret and on the other hand allow determination of the individual’s identity for the second terminal, based on the presented biometric identification sequence. All helper data in the database are assumed to be public. Since the biometric secrets produced by the first terminal are used e.g. to encrypt data, the helper data should provide no information on these secrets. On the other hand, since biometric data are unique for individuals and cannot be replaced if compromised, the helper data should also provide as little as possible information about biometric data. In our identification system we only store the helper data as reference data for identification. Therefore these helper data are also called protected templates. In this paper we determine what identification, secret-key and privacy-leakage rates can be realized by such a biometric identification system.

II. DEFINITIONS

A. Biometrics

A biometric identification system, see Fig. 1, is based on a *biometric source* $\{Q_s(x), x \in \mathcal{X}\}$ and a *biometric channel* $\{Q_c(y|x), y \in \mathcal{Y}, x \in \mathcal{X}\}$.

The system is designed to identify one out of M_I individuals. For each individual $i \in \{1, 2, \dots, M_I\}$ in the system, the biometric source produces a biometric *enrollment sequence* $x^N(i) = (x_1(i), x_2(i), \dots, x_N(i))$ with N symbols from the finite alphabet $\mathcal{X}$. The enrollment sequence x^N occurs with probability

$$\Pr\{X^N = x^N\} = \prod_{n=1}^N Q_s(x_n), \quad (1)$$

hence the symbols $\{X_n, n = 1, 2, \dots, N\}$ are independent of each other and identically distributed according to $Q_s(\cdot)$.

Fig. 1. Model for biometric identification.

Note that the biometric sequences are independent of the individual's identity.

During identification a biometric *identification sequence* of an unknown individual $y^N = (y_1, y_2, \dots, y_N)$ with N symbols from the finite alphabet $\mathcal{Y}$ is observed. This sequence is the output of the biometric channel whose input was the enrollment sequence of this individual. If individual i was observed the sequence y^N occurs with probability

$$\Pr\{Y^N = y^N | X^N = x^N(i)\} = \prod_{n=1}^N Q_c(y_n | x_n(i)), \quad (2)$$

hence the biometric channel is memoryless.

We assume here that all individuals are equally likely to be observed for identification, hence

$$\Pr\{I = i\} = 1/M_I, \text{ for all } i \in \{1, 2, \dots, M_I\}. \quad (3)$$

B. Encoding and Decoding

The enrollment and identification biometric sequences $x^N(i)$ and y^N are observed by an *encoder* and *decoder*, respectively, see Fig. 1. During the enrollment procedure the biometric sequence $x^N(i)$ of individual $i \in \{1, 2, \dots, M_I\}$ is encoded into *helper data* $h(i) \in \{1, 2, \dots, M_H\}$ and a *secret* $s(i) \in \{1, 2, \dots, M_S\}$, hence

$$(H(i), S(i)) = e(X^N(i)), \text{ for } i \in \{1, 2, \dots, M_I\}, \quad (4)$$

where $e(\cdot)$ is the encoder mapping. The helper data $h(i)$ are then stored in a (public) *database* at position i . The secret $s(i)$ is handed over to the individual. The individual can use it e.g. as a key for encryption purposes.

The helper data are stored in the database to make reliable identification possible. They should only contain a negligible amount of information about the corresponding secret, and, moreover, contain as little as possible information about the enrollment biometric sequence.

During identification, upon observing the biometric identification sequence y^N , the decoder forms an estimate $\hat{i}$ of the

identity of the observed individual as well as an estimate of his secret $\hat{s}(\hat{i})$, hence

$$(\hat{I}, \hat{S}(\hat{I})) = d(Y^N, H(1), H(2), \dots, H(M_I)), \quad (5)$$

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

Moreover, the decoder's estimate of the secret $\hat{s}(\hat{i})$ assumes values from the same alphabet as the secret chosen during enrollment, i.e. $\hat{s}(\hat{i}) \in \{1, 2, \dots, M_S\}$. The estimate of the individual's identity $\hat{i}$ takes on values from the set of individuals, i.e. $\hat{i} \in \{1, 2, \dots, M_I\}$.

C. Achievability

Now we are interested to find out what identification, secret-key and privacy-leakage rates can be realized by our identification system with protected templates with negligible error probability, such that individuals' secret keys are close to uniform and for each individual the helper data only provide negligible information on his secret. We give the following definition of achievability.

Definition 1 (Achievability) A secret-key rate, identification rate, and privacy-leakage rate triple (R_S, R_I, R_L) with $R_S \geq 0$ and $R_I \geq 0$ is achievable in a biometric identification setting with protected templates if for all $\delta > 0$ for all N large enough there exist encoders and decoders such that¹

$$\begin{aligned} \Pr\{(\hat{I}, \hat{S}(\hat{I})) \neq (I, S(I))\} &\leq \delta, \\ \frac{1}{N} H(S(i)) + \delta &\geq \frac{1}{N} \log M_S \geq R_S - \delta, \\ \frac{1}{N} \log M_I &\geq R_I - \delta, \\ \frac{1}{N} I(S(i); H(i)) &\leq \delta, \\ \frac{1}{N} I(X^N(i); H(i)) &\leq R_L + \delta, \end{aligned}$$

for all $i \in \{1, 2, \dots, M_I\}$. (6)

Moreover, let $\mathcal{R}_{bi}$ be the region of all achievable secret-key, identification and privacy-leakage rate triples for a biometric identification system with protected templates.

Remark: Note that due to the generation (coding) process, for the secrecy and privacy leakage it holds that $I(S(i); H(i)) = I(S(i); H(1), H(2), \dots, H(M_I))$, since only $H(i)$ can possibly be dependent on $S(i)$, and $I(X^N(i); H(i)) = I(X^N(i); H(1), H(2), \dots, H(M_I))$, since $H(j)$ are independent of $X^N(i)$ if $i \neq j$, for all $i, j \in \{1, 2, \dots, M_I\}$.

III. STATEMENT OF RESULTS

In order to state our result we first define the region $\mathcal{R}$ and then we present our main theorem.

$$\begin{aligned} \mathcal{R} \triangleq \{ (R_I, R_S, R_L) : 0 \leq R_I + R_S \leq I(U; Y), \\ R_L \geq I(U; X) - I(U; Y) + R_I, \\ \text{for } P(u, x, y) = Q_s(x) Q_c(y|x) P(u|x) \\ \text{and } |\mathcal{U}| \leq |\mathcal{X}| + 1 \}. \end{aligned} \quad (7)$$

¹Throughout this paper we take two as base of the log.

Theorem 1 (Biometric Identification, Protected Templates)

$$\mathcal{R}_{bi} = \mathcal{R}. \quad (8)$$

As special cases we can derive the following five theorems from the theorem presented above. These theorems represent already established results that we discuss below. Again we define five regions first.

$$\mathcal{R}_1 \triangleq \{R_S : R_S \leq I(X; Y)\}. \quad (9)$$

$$\begin{aligned} \mathcal{R}_2 \triangleq \{(R_S, R_L) : & 0 \leq R_S \leq I(U; Y), \\ & R_L \geq I(U; X) - I(U; Y), \\ & \text{for } P(u, x, y) = Q_s(x)Q_c(y|x)P(u|x) \\ & \text{and } |\mathcal{U}| \leq |\mathcal{X}| + 1\}. \end{aligned} \quad (10)$$

$$\mathcal{R}_3 \triangleq \{R_I : R_I \leq I(X; Y)\}. \quad (11)$$

$$\begin{aligned} \mathcal{R}_4 \triangleq \{(R_I, R_L) : & 0 \leq R_I \leq I(U; Y), \\ & R_L \geq I(U; X), \\ & \text{for } P(u, x, y) = Q_s(x)Q_c(y|x)P(u|x) \\ & \text{and } |\mathcal{U}| \leq |\mathcal{X}| + 1\}. \end{aligned} \quad (12)$$

$$\begin{aligned} \mathcal{R}_5 \triangleq \{(R_I, R_S) : & 0 \leq R_I + R_S \leq I(X; Y), \\ & \text{for } P(x, y) = Q_s(x)Q_c(y|x)\}. \end{aligned} \quad (13)$$

In the following theorems we set $R_L = \infty$ to indicate that we exclude privacy leakage from our considerations.

Theorem 2 *If we restrict ourselves to $R_I = 0$ and $R_L = \infty$ then*

$$\mathcal{R}_{bi|R_I=0, R_L=\infty} = \mathcal{R}_1. \quad (14)$$

This theorem gives us the Ahlswede and Csiszár [1] result for the amount of common secret information that can be generated by two terminals. Note that in the biometric setting the secrecy capacity can be achieved at privacy leakage rate of $H(X|Y)$.

Theorem 3 *If we restrict ourselves to $R_I = 0$ then*

$$\mathcal{R}_{bi|R_I=0} = \mathcal{R}_2. \quad (15)$$

The region in Thm.3 corresponds to the region for biometric authentication based on secret generation as in [3] and [2].

Theorem 4 *If we restrict ourselves to $R_S = 0$ and $R_L = \infty$ then*

$$\mathcal{R}_{bi|R_S=0, R_L=\infty} = \mathcal{R}_3. \quad (16)$$

The special case in the above theorem corresponds to the identification region for a biometric identification system without protected templates, as in Willems et al. [7] and O'Sullivan and Schmid [5]. Indeed to achieve identification capacity we

have to store all biometric information and thus cannot achieve any privacy protecting as $R_L = H(X)$ then.

Theorem 5 *If we restrict ourselves to $R_S = 0$ then*

$$\mathcal{R}_{bi|R_S=0} = \mathcal{R}_4. \quad (17)$$

Also from the above theorem we can see that if we do not require generation of a secret key, then to achieve identification rate $I(U; Y)$ we have to store the template of rate $I(U; X)$ which results into the privacy-leakage rate $I(U; X)$. This is similar to the Tuncel result [6] if we assume that the underlying biometric source sequence corresponds to the enrollment biometric sequence.

Theorem 6 *If we restrict ourselves to $R_L = \infty$ then*

$$\mathcal{R}_{bi|R_L=\infty} = \mathcal{R}_5. \quad (18)$$

Finally, the last theorem corresponds to the identification setting with secret keys studied in [4].

In general from Thm. 1 we see that the larger identification rate we would like to achieve the smaller the secret-keys rates and the larger the privacy-leakage rates we can realize.

IV. EXAMPLE: BINARY SYMMETRIC DOUBLE SOURCE

Consider a binary symmetric double source (BSDS) with crossover probability $0 \leq q \leq 1/2$, hence $Q(x, y) = Q_s(x)Q_c(y|x) = (1 - q)/2$ for $y = x$ and $q/2$ for $y \neq x$. For such a source

$$\begin{aligned} I(U; Y) &= 1 - H(Y|U), \\ I(U; X) - I(U; Y) &= H(Y|U) - H(X|U). \end{aligned} \quad (19)$$

Mrs. Gerber's Lemma [8] tells us that if $H(X|U) = v$ then $H(Y|U) \geq h(q * h^{-1}(v))$, where $h(a) \triangleq -a \log(a) - (1 - a) \log(1 - a)$ is the binary entropy function. If now $0 \leq p \leq 1/2$ is such that $h(p) = v$ then $H(X|U) = h(p)$ and $H(Y|U) \geq h(q * p)$.

We define privacy-leakage vs. secret-key and identification rate function

$$R_{bi}(R_S, R_I) = \min\{R_L : (R_I, R_S, R_L) \in \mathcal{R}_{bi}\}. \quad (20)$$

Note that for binary symmetric (U, X) with crossover probability p the minimum $H(Y|U)$ is achieved, and consequently for identification rates $R_I \geq 0$ we obtain

$$\begin{aligned} R_{bi}(R_S, R_I) &= h(p * q) - h(p) + R_I, \\ &\text{for } p \text{ satisfying } 1 - h(p * q) - R_I = R_S, \\ &\text{and } R_I \leq 1 - h(p * q). \end{aligned} \quad (21)$$

In Fig. 2 we plot the resulting function for $q = 0.1$ and in Fig. 3-5 the corresponding projections to the secret-key rate and privacy-leakage rate, identification rate and secret-key rate, and identification rate and privacy-leakage rate planes, respectively. These figures demonstrate the tradeoff between three rates.

Fig. 2. Rate function for the crossover probability $q = 0.1$.

Fig. 3. Secret-key vs privacy-leakage rate projection, $q = 0.1$.

Fig. 4. Identification vs secret-key rate projection, $q = 0.1$.

Fig. 5. Identification vs privacy-leakage rate projection, $q = 0.1$.

V. PROOF OF THEOREM 1

The proof of this theorem consists of three parts. The first part, i.e. the converse is treated in detail. The second part concerns the achievability of which we only provide an outline. The third part, the bound on cardinality of U , can be proven using the Fenchel-Eggleston strengthening the Caratheodory lemma, see [9].

A. The Converse

We start by considering the joint entropy $H(I, S(I))$ of the individual's identity and his secret. We use that $(\hat{I}, \hat{S}(\hat{I})) = d(Y^N, H(1), H(2), \dots, H(M_I))$ and Fano's inequality $H(I, S(I)|\hat{I}, \hat{S}(\hat{I})) \leq F$, where $F \triangleq 1 + \Pr\{(\hat{I}, \hat{S}(\hat{I})) \neq (I, S(I))\} \log(M_I M_S)$. Then

$$\begin{aligned} H(I, S(I)) &= I(I, S(I); H(1), H(2), \dots, H(M_I), Y^N) \\ &\quad + H(I, S(I)|H(1), H(2), \dots, H(M_I), Y^N, \hat{I}, \hat{S}(\hat{I})) \\ &\leq I(I, S(I); H(1), H(2), \dots, H(M_I), Y^N) \\ &\quad + H(I, S(I)|\hat{I}, \hat{S}) \end{aligned}$$

$$\begin{aligned} &\leq I(I, S(I); H(1), H(2), \dots, H(M_I)) \\ &\quad + I(I, S(I); Y^N|H(1), H(2), \dots, H(M_I)) + F \\ &= I(I; H(1), H(2), \dots, H(M_I)) \\ &\quad + I(S(I); H(1), H(2), \dots, H(M_I)|I) \\ &\quad + I(Y^N; I, S(I)|H(1), H(2), \dots, H(M_I)) + F \\ &\leq I(S(I); H(1), H(2), \dots, H(M_I)|I) \\ &\quad + I(Y^N; I, S(I), H(1), H(2), \dots, H(M_I)) + F \\ &= I(S(I); H(1), H(2), \dots, H(M_I)|I) \\ &\quad + I(Y^N; I, S(I), H(I)) + F \\ &= I(S(I); H(1), H(2), \dots, H(M_I)|I) + F \\ &\quad + \sum_{n=1}^N I(Y_n; I, S(I), H(I), Y^{n-1}) \\ &\leq I(S(I); H(1), H(2), \dots, H(M_I)|I) + F \\ &\quad + \sum_{n=1}^N I(Y_n; I, S(I), H(I), Y^{n-1}, X^{n-1}(I)) \\ &\leq I(S(I); H(1), H(2), \dots, H(M_I)|I) + F \end{aligned}$$

$$\begin{aligned}
& + \sum_{n=1}^N I(Y_n; I, S(I), H(I), X^{n-1}(I)) \\
& \leq \frac{1}{M_I} \sum_{i=1}^{M_I} I(S(i); H(i)) + NI(U; Y) + F. \quad (22)
\end{aligned}$$

The last steps require some attention. The last equality follows from the fact that biometric sequence Y^N is independent of all the helper data other than the helper data corresponding to the actual individual's identity. The last but one inequality holds since $Y^{n-1} - (I, S(I), H(I), X^{n-1}(I)) - Y_n$. To obtain the last inequality, we, first, define $U_n \triangleq (I, S(I), H(I), X^{n-1})$ for $n = 1, 2, \dots, N$. Then if we take a time-sharing variable T uniform over $\{1, 2, \dots, N\}$ and independent of all other variables and set $U \triangleq (U_n, n)$, $X \triangleq X_n$, and $Y \triangleq Y_n$ for $T = n$, we obtain

$$\begin{aligned}
\sum_{n=1}^N I(I, S(I), H(I), X^{n-1}(I); Y_n) &= \sum_{n=1}^N I(U_n; Y_n) \\
&= NI(U_T; Y_T | T) \\
&= NI((U_T, T); Y_T) \\
&= NI(U; Y). \quad (23)
\end{aligned}$$

Finally, note that $U_n - X_n - Y_n$ and, consequently, $U - X - Y$.

Now for achievable triples (R_S, R_I, R_L) we obtain that

$$\begin{aligned}
\log(M_I M_S) &\leq \log M_I + \min_{i=1,2,\dots,M_I} H(S(i)) + N\delta \\
&\leq H(I) + H(S(I)|I) + N\delta \\
&\leq H(I, S(I)) + N\delta \\
&\leq 2N\delta + NI(U; Y) + 1 + \delta \log(M_I M_S), \quad (24)
\end{aligned}$$

and finally that

$$\begin{aligned}
R_I + R_S - 2\delta &\leq \frac{1}{N} \log(M_I M_S) \\
&\leq \frac{1}{1-\delta} (I(U; Y) + 2\delta + \frac{1}{N}), \quad (25)
\end{aligned}$$

for some $P(u, x, y) = Q_s(x)Q_c(y|x)P(u|x)$.

Now we continue with the privacy leakage.

$$\begin{aligned}
& I(X^N(I); H(1), H(2), \dots, H(M_I) | I) \\
&= I(X^N(I), I; H(1), H(2), \dots, H(M_I)) \\
&= H(X^N(I), I, S(I)) \\
&\quad - H(X^N(I), I, S(I) | H(1), H(2), \dots, H(M_I)) \\
&= H(I) + H(X^N(I), S(I) | I) \\
&\quad - H(I, S(I) | H(1), H(2), \dots, H(M_I)) \\
&\quad - H(X^N(I) | I, S(I), H(1), H(2), \dots, H(M_I)) \\
&= H(I) + H(X^N(I)) \\
&\quad - H(I, S(I) | H(1), H(2), \dots, H(M_I), Y^N) \\
&\quad - I(Y^N; I, S(I) | H(1), H(2), \dots, H(M_I)) \\
&\quad - H(X^N(I) | I, S(I), H(1), H(2), \dots, H(M_I)) \\
&\geq H(I) - H(I, S(I) | \hat{I}, \hat{S}(I)) \\
&\quad + I(X^N(I); I, S(I), H(1), H(2), \dots, H(M_I))
\end{aligned}$$

$$\begin{aligned}
& - I(Y^N; I, S(I), H(1), H(2), \dots, H(M_I)) \\
&\geq \log M_I - F + \sum_{n=1}^N I(X_n(I); I, S(I), H(I), X^{n-1}(I)) \\
&\quad - \sum_{n=1}^N I(Y_n; I, S(I), H(I), Y^{n-1}) \\
&\geq \log M_I + NI(U; X) - NI(U; Y) - F, \quad (26)
\end{aligned}$$

for the joint distribution $P(u, x, y) = Q_s(x)Q_c(y|x)P(u|x)$, mentioned before.

For achievable triples (R_S, R_I, R_L) we get

$$\begin{aligned}
& R_L + \delta \\
&\geq \frac{1}{N} \max_{i \in \{1, 2, \dots, M_I\}} I(X^N(i); H(i)) \\
&\geq \frac{1}{N} \frac{1}{M_I} \sum_{i=1}^{M_I} I(X^N(i); H(i)) \\
&\geq \frac{1}{N} I(X^N(I); H(1), H(2), \dots, H(M_I) | I) \\
&\geq \frac{1}{N} (\log M_I + NI(U; X) - NI(U; Y) - F) \\
&\geq R_I - \delta + I(U; X) - I(U; Y) - \frac{1}{N} (\delta \log(M_I M_S) + 1) \\
&\geq I(U; X) - \frac{1}{1-\delta} I(U; Y) + R_I - \frac{1}{(1-\delta)N} - \frac{\delta(1+\delta)}{1-\delta}, \quad (27)
\end{aligned}$$

here we used Fano's inequality and (25).

If we now let $\delta \downarrow 0$ and $N \rightarrow \infty$, then we obtain the converse from both (25) and (27).

B. Outline of the Achievability Proof

We start by fixing a conditional distribution $\{P(u|x), x \in \mathcal{X}, u \in \mathcal{U}\}$ that determines the joint distribution $P(u, x, y) = Q_s(x)Q_c(y|x)P(u|x)$, for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and $u \in \mathcal{U}$. Then we randomly generate roughly $2^{NI(U; X)}$ sequences u^N . Each of those sequences gets a random s -label and a random h -label. These labels are uniformly chosen. The s -label can assume roughly $2^{(NI(U; Y) - R_I)}$ values, the h -label roughly $2^{(NI(U; X) - (I(U; Y) + R_I))}$ values.

During enrollment for each individual with identity $i \in \{1, 2, \dots, NR_I\}$, the encoder, upon observing the source sequence $x^N(i)$, finds $u^N(i)$ such that $u^N(i)$ and $x^N(i)$ are jointly typical. Then it stores the helper-label $h(i)$ corresponding to $u^N(i)$ at the position i in a *public* database. Moreover, the encoder issues the secret-label $s(i)$ corresponding to this $u^N(i)$ to the individual.

During identification the decoder observes the identification sequence y^N and, checking all the records in the database, determines a unique individual with identity label $\hat{i}$ such that the record of the database $\hat{i}$ contains the label $h(\hat{i}) = h(u^N(\hat{i}))$ for which $u^N(\hat{i})$ and y^N are jointly typical. Then the decoder issues the identity estimate $\hat{i}$ and the secret estimate $\hat{s}(\hat{i})$. It can be shown that the decoder can reliably recover $u^N(\hat{i})$ and thus $\hat{i}$ and $\hat{s}(\hat{i})$.

Finally, it is easy to check that the leakage is not larger than $I(U; X) - I(U; Y) + R_I$. Moreover, to prove the facts that secrecy leakage is negligible and that the secret is close to uniform we can use the property of the encoding procedure that $u^N(i)$ can be reliably reconstructed from $s(i)$ and $h(i)$.

VI. CONCLUSIONS

In this paper we have considered biometric identification systems with protected templates. Biometric data used in such identification systems are also utilized in access control and authentication applications. These applications are typically based on biometric secrets that are used for encryption purposes. To create reliable identification systems, helper data of all enrolled individuals have to be accessed by the decoder. These data are assumed to be public. Thus public information of biometric identification system should provide no information about biometric secrets, though facilitate reliable identification. Also because biometric data cannot be replaced if compromised, the helper data should contain as little as possible information about biometrics.

Here we have analyzed what secret-key, identification and privacy-leakage rates can be realized by biometric identification systems with protected templates. It appears that the larger identification rates we would like to achieve, the smaller secret keys we can generate and the more biometric information we have to leak. We also see that our results are strongly connected to the secret sharing concept of Ahlswede and Csiszár [1]; the biometric identification system without protected templates of Willems et al. [7] and O'Sullivan and Schmid [5]; the biometric identification system with restricted storage of Tuncel [6]; the biometric identification with secret keys of [4] and [2]; and the biometric authentication system with privacy protection of [3]. These results can be derived as special cases in the biometric identification systems with protected templates considered here.

REFERENCES

- [1] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography - part I: Secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, pp. 1121–1132, July 1993.
- [2] T. Ignatenko, "Secret-key rates and privacy leakage in biometric systems," Ph.D. dissertation, Eindhoven University of Technology, 2009.
- [3] T. Ignatenko and F. Willems, "Biometric systems: Privacy and secrecy aspects," *IEEE Trans. Inf. Forensics and Security*, vol. 4, no. 4, December 2009.
- [4] —, "Secret-key and identification rates for biometric identification systems with protected templates," in *Proc. of 31st Symp. Inf. Theory in the Benelux, May 11-12, 2010, Rotterdam, The Netherlands*, 2010.
- [5] J. A. O'Sullivan and N. A. Schmid, "Large deviations performance analysis for biometrics recognition," in *Proc. of 40th Annual Allerton Conference on Communication, Control, and Computing, October 2-4, 2002, Allerton House Monticello, IL, USA*, 2002.
- [6] E. Tuncel, "Capacity/storage tradeoff in high-dimensional identification systems," July 2006, pp. 1929–1933.
- [7] F. Willems, T. Kalker, J. Goseling, and J.-P. Linnartz, "On the capacity of a biometrical identification system," in *Proc. of 2003 IEEE Int. Symp. Information Theory*, 2003.
- [8] A. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications-I," *IEEE Trans. Inf. Theory*, vol. 19, no. 6, pp. 769–772, Nov 1973.
- [9] —, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, January 1976.