Secret rate - Privacy leakage in biometric systems

Secret Rate - Privacy Leakage in Biometric Systems

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

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

***Abstract*—Ahlswede and Csiszár [1993] introduced the concept of secret sharing. In their source model two terminals observe two correlated sequences. It is the objective of the terminals to form a common secret by interchanging a public message (helper data) in such a way that the secrecy leakage is negligible. In a biometric setting, where the sequences correspond to the enrollment and authentication data, respectively, it is crucial that the public message leaks as little information as possible about the biometric data, since compromised biometric data cannot be replaced. We investigated the fundamental trade-offs for four biometric settings. The first one is the standard (Ahlswede-Csiszár) secret generation setting, for which we determined the secret-key vs. privacy-leakage rate region. Here leakage corresponds to the mutual information between helper data and biometric enrollment sequence. In the second setting the secret is not generated by the terminals but independently chosen, and transmitted using a public message. Again we determined the region of achievable rate-leakage pairs. In setting three and four we consider zero-leakage, i.e. the public message contains only a negligible amount of information about the secret and about the biometric enrollment sequence. To achieve this a private key is needed, which can be observed only by the terminals. We considered again both secret generation and secret transmission and determined for both cases the region of achievable secret-key vs. private-key rate pairs.**

I. Introduction

First, Maurer [10] and slightly later Ahlswede and Csiszár [1] introduced the concept of secret sharing. In their source model two terminals observe two correlated sequences $\underline{X}$ and $\underline{Y}$. It is the objective of both terminals to form a common secret S by interchanging a public message H (helper data) that should only contain a negligible amount of information about the secret. Ahlswede and Csiszár showed that the maximum secret-key rate that can be achieved in this way is equal to the mutual information between the correlated source outputs $I(X;Y)$. Their achievability proofs can be expressed in terms of Slepian-Wolf techniques presented by Cover [3], in which binning of typical sequences plays an important role, see e.g. Ye and Narayan [15] and [7]. The concept of secret sharing is closely related to the generation of common randomness. When two terminals try to generate common randomness the issue of secrecy of the helper data is dropped. Common randomness capacity was first studied in a systematic way by Ahlswede and Csiszár [2]. Later helper terminals were included by Csiszár and Narayan in their investigations in [4].

In a biometric setting, where the X-sequence corresponds to the enrollment and the Y-sequence to the authentication biometric data, it is crucial that the public message H leaks as little information as possible about the biometric data, since compromised biometric data cannot be replaced. Smith [12] has investigated this privacy leakage and came to the conclusion that it cannot be avoided. In our work we determine the trade-off between secret-key rate and privacy leakage for the i.i.d. case.

We also consider secret transmission. We study a model in which a uniformly chosen secret key is transmitted by the first terminal via a public message to the second terminal. The terminals observe two correlated biometric sequences, and the public helper data should be uninformative about the secret and as uninformative as possible about the biometric data. Again we determine the rate-leakage balance for this setting. Recently, Prabhakaran and Ramchandran [11] and Gündüz et al. [5] studied source coding problems where also the issue of (biometric) leakage is addressed. In their work it is not the intention of the users to produce a secret but to communicate a (biometric) source sequence in a secure way from the first terminal to the second terminal.

Next we study a zero-leakage secret generation system. In this system an additional random key is made available only to the two terminals. We now focus on helper data that contain only a negligible amount of information about the secret and biometric sequence. Also for this case we determine the trade-off between private-key rate and the resulting secret-key rate.

Moreover, we address zero-leakage secret transmission. Here again both terminals have access to a private key, but now it is their intention to transmit an independently chosen uniform secret form the first terminal to the second by means of public helper data, that are practically not leaking. The trade-off for this setting is presented here, i.e. we show how the secret-key rate depends on the private-key rate.

In this paper we concentrate on the privacy leakage defined as mutual information between the helper data and the enrollment biometric sequence. However, a stronger definition of the leakage is possible when the leakage corresponds to the conditional version of this mutual information, given the secret. Here we provide the results for all four cases for unconditional privacy leakage, but only prove the results of last two settings, i.e. zero-leakage. The results and proofs for the settings in the conditional case can be found in [6] and [8]. The models with unconditional privacy leakage were studied in [8] and, for the first two settings, also in Lai at al. [9].

978-1-4244-4313-0/09/$25.00 ©2009 IEEE

II. Four Cases, Definitions

A. Basic Definitions

A biometric system is based on a *biometric source* $\{Q(x,y), x \in \mathcal{X}, y \in \mathcal{Y}\}$ that produces an X-sequence $\underline{x} = (x_1, x_2, \ldots, x_N)$ with N symbols from the finite alphabet $\mathcal{X}$ and a Y-sequence $\underline{y} = (y_1, y_2, \ldots, y_N)$ with N symbols from finite alphabet $\mathcal{Y}$. The sequence pair $(\underline{x}, \underline{y})$ occurs with probability

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

hence the source pairs $\{(X_n, Y_n), n = 1, \ldots, N\}$ are independent of each other and identically distributed according to $Q(\cdot,\cdot)$. The biometric source sequences $\underline{x}$ and $\underline{y}$ are in general not independent of each other.

The sequences $\underline{x}$ and $\underline{y}$ are observed by an encoder and decoder, respectively. One of the outputs that the encoder produces is an index $h \in \{1, 2, \ldots, M_H\}$, which is referred to as helper data. The helper data are made public and are used by the decoder.

We can subdivide systems into those in which both terminals are supposed to *generate* a secret, and those in which a uniformly chosen secret is *transmitted* from the encoder to the decoder. The generated or transmitted secret s assumes values in $\{1, 2, \ldots, M_S\}$. The decoder's estimate $\widehat{s}$ of the secret s also assumes values from $\{1, 2, \ldots, M_S\}$. In transmission systems the secret s is a uniformly distributed index, hence

$$\Pr\{S = s\} = 1/M_S \text{ for all } s \in \{1, 2, \ldots, M_S\}. \tag{2}$$

Moreover, we can subdivide systems into systems in which the helper data are allowed to leak some information about the biometric sequence $\underline{X}$, and systems in which this leakage should be negligible. In the so-called *zero-leakage* systems both terminals have access to a private random key p. This key is uniformly distributed, hence

$$\Pr\{P = p\} = 1/M_P \text{ for all } p \in \{1, 2, \ldots, M_P\}. \tag{3}$$

In the next subsections the four resulting combinations (1) secret generation, (2) secret transmission, (3) zero-leakage secret generation, and (4) zero-leakage secret transmission, will be proposed in detail.

There are two types of privacy leakage, (a) unconditional leakage and (b) conditional leakage (not treated here). Unconditional leakage corresponds to bounding $I(\underline{X}; H)$, whereas conditional leakage corresponds to bounding $I(\underline{X}; H|S)$.

B. Secret Generation

In a biometric secret generation system, see Fig. 1, the encoder observes the enrollment biometric source sequence $\underline{X}$ and produces a secret S and helper data H, hence $(S, H) = e(\underline{X})$, where $e(\cdot)$ is the encoder mapping. The public helper data H are sent to the decoder which also observes the authentication biometric source sequence $\underline{Y}$. This decoder now forms an estimate $\widehat{S}$ of the secret that was produced by the encoder, hence $\widehat{S} = d(\underline{Y}, H)$, where $d(\cdot,\cdot)$ is the decoder mapping.

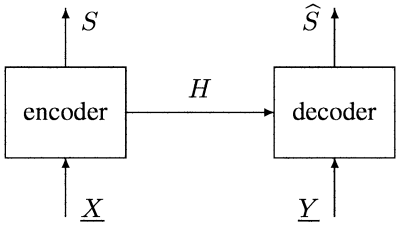

Fig. 1. Model for biometric secret generation.

For biometric secret generation we now give the definition of achievability corresponding to the unconditional privacy leakage.

Definition 1 (Uncond.) *In a biometric secret generation system, a rate-leakage pair (R, L) with $R \geq 0$ is achievable in the unconditional case if for all $\delta > 0$ and for all N large enough there exist encoders and decoders such that*

$$\begin{aligned} & & \Pr\{\widehat{S} \neq S\} &\leq \delta, \\ H(S) + N\delta &\geq \log(M_S) &\geq N(R - \delta), \\ & & I(S; H) &\leq N\delta, \\ & & I(\underline{X}; H) &\leq N(L + \delta). \end{aligned} \tag{4}$$

Moreover, $\mathcal{R}_{sg}^{u}$ is the region of all achievable rate-leakage pairs for a secret generation system in the unconditional case. The corresponding rate-leakage function $R_{sg}^{u}(L)$ is defined as

$$R_{sg}^{u}(L) \triangleq \max\{R : (R, L) \in \mathcal{R}_{sg}^{u}\}. \tag{5}$$

C. Secret Transmission

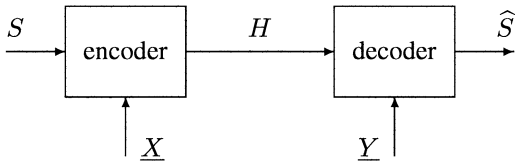

Fig. 2. Model for biometric secret transmission.

In a biometric secret transmission system, see Fig. 2, a secret S that is to be transmitted from an encoder to a decoder is uniformly distributed, see (2). The encoder observes the enrollment biometric source sequence $\underline{X}$ and the secret S and produces helper data H, hence $H = e(S, \underline{X})$, where $e(\cdot,\cdot)$ is the encoder mapping. The public helper data H are sent to the decoder that also observes the authentication biometric source sequence $\underline{Y}$. This decoder forms an estimate $\widehat{S}$ of the secret that was transmitted by the encoder, hence $\widehat{S} = d(H, \underline{Y})$, and $d(\cdot,\cdot)$ is the decoder mapping.

Definition 2 (Uncond.) *In a biometric secret transmission system, a rate-leakage pair (R, L) with $R \geq 0$ is achievable in the unconditional case if for all $\delta > 0$ and for all N large enough there exist encoders and decoders such that*

$$\begin{aligned} \Pr\{\widehat{S} \neq S\} &\leq \delta, \\ \log(M_S) &\geq N(R - \delta), \\ I(S; H) &\leq N\delta, \\ I(\underline{X}; H) &\leq N(L + \delta). \end{aligned} \tag{6}$$

Moreover, $\mathcal{R}_{st}^{u}$ is the region of all achievable rate-leakage pairs for a secret transmission system in the unconditional case. The corresponding rate-leakage function $R_{st}^{u}(L)$ is defined as

$$R_{st}^{u}(L) \stackrel{\Delta}{=} \max\{R : (R, L) \in \mathcal{R}_{st}^{u}\}. \tag{7}$$

D. Zero-Leakage Secret Generation

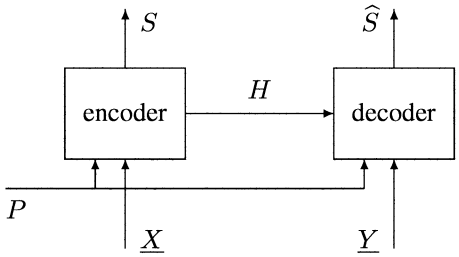

Fig. 3. Model for zero-leakage biometric secret generation.

In a zero-leakage biometric secret generation system, see Fig. 3, a private random key P that is available to both an encoder and decoder, is uniformly distributed. The encoder observes the enrollment biometric source sequence $\underline{X}$ and the private key P and produces a secret S and helper data H, hence $(S, H) = e(\underline{X}, P)$, where $e(\cdot, \cdot)$ is the encoder mapping. The helper data H are sent to the decoder that also observes the authentication biometric source sequence $\underline{Y}$ and that has access to the private key P. This decoder now forms an estimate $\widehat{S}$ of the secret that was produced by the encoder, hence $\widehat{S} = d(H, \underline{Y}, P)$, where $d(\cdot, \cdot, \cdot)$ is the decoder mapping.

Definition 3 (Uncond.) *In a zero-leakage biometric secret generation system, a secret-key vs. private-key rate pair (R, K) with $R \geq 0$ is achievable in the unconditional case if for all $\delta > 0$ and for all N large enough there exist encoders and decoders such that*

$$\begin{array}{rrrcl} & & \Pr\{\widehat{S} \neq S\} & \leq & \delta, \\ H(S) + N\delta & \geq & \log(M_S) & \geq & N(R - \delta), \\ & & \log(M_P) & \leq & N(K + \delta), \\ & & I(S; H) & \leq & N\delta, \\ & & I(\underline{X}; H) & \leq & N\delta. \end{array} \tag{8}$$

Moreover, $\mathcal{R}_{zsg}^{u}$ is the region of all achievable secret-key vs. private-key rate pairs (R, K) for a zero-leakage secret generation system in the unconditional case. The corresponding secret-key vs. private-key rate function $R_{zsg}^{u}(K)$ is defined as

$$R_{zsg}^{u}(K) \stackrel{\Delta}{=} \max\{R : (R, K) \in \mathcal{R}_{zsg}^{u}\}. \tag{9}$$

E. Zero-Leakage Secret Transmission

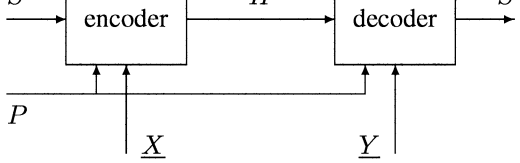

Fig. 4. Model for zero-leakage biometric secret transmission.

In a zero-leakage biometric secret transmission system, see Fig. 4, a private random key P that is available to both an encoder and decoder is uniformly distributed, see (3). Moreover, a secret S that is to be transmitted from the encoder to the decoder is also uniformly distributed, see (2).

The encoder observes the enrollment biometric source sequence $\underline{X}$, the private key P, and the secret S and forms helper data $H = e(S, \underline{X}, P)$, where $e(\cdot, \cdot, \cdot)$ is the encoder mapping. The helper data H are sent to the decoder that also observes the authentication biometric source sequence $\underline{Y}$ and that has access to the private key P. This decoder now forms an estimate $\widehat{S}$ of the secret that was transmitted by the encoder, hence $\widehat{S} = d(H, \underline{Y}, P)$, where $d(\cdot, \cdot, \cdot)$ is the decoder mapping.

Definition 4 (Uncond.) *In a zero-leakage biometric secret transmission system, a secret-key vs. private-key rate pair (R, K) with $R \geq 0$ is achievable in the unconditional case if for all $\delta > 0$ and for all N large enough there exist encoders and decoders such that*

$$\begin{array}{rcl} \Pr\{\widehat{S} \neq S\} & \leq & \delta, \\ \log(M_S) & \geq & N(R - \delta), \\ \log(M_P) & \leq & N(K + \delta), \\ I(S; H) & \leq & N\delta, \\ I(\underline{X}; H) & \leq & N\delta. \end{array} \tag{10}$$

Moreover, $\mathcal{R}_{zst}^{u}$ is the region of all achievable secret-key vs. private-key rate pairs (R, K) for a zero-leakage secret transmission system in the unconditional case. The corresponding secret-key vs. private-key rate function $R_{zst}^{u}(K)$ is defined as

$$R_{zst}^{u}(K) \stackrel{\Delta}{=} \max\{R : (R, K) \in \mathcal{R}_{zst}^{u}\}. \tag{11}$$

III. Statement of Results

In order to state our results we first define the regions $\mathcal{R}_1$ and $\mathcal{R}_2$. Then we will present four theorems.

$$\begin{array}{rcll} \mathcal{R}_1 \stackrel{\Delta}{=} \{(R, L) & : & 0 \leq R \leq I(U; Y), & \\ & & L \geq I(U; X) - I(U; Y), & \\ & & \text{for } P(u, x, y) = Q(x, y)P(u|x) & \\ & & \text{with } |\mathcal{U}| \leq |\mathcal{X}| + 1\}, & (12) \\ \mathcal{R}_2 \stackrel{\Delta}{=} \{(R, K) & : & 0 \leq R \leq I(U; Y) + K, & \\ & & K \geq I(U; X) - I(U; Y), & \\ & & \text{for } P(u, x, y) = Q(x, y)P(u|x) & \\ & & \text{with } |\mathcal{U}| \leq |\mathcal{X}| + 1\}. & (13) \end{array}$$

Theorem 1 (Secret Generation, Uncond.)

$$\mathcal{R}_{sg}^{u} = \mathcal{R}_1. \tag{14}$$

Theorem 2 (Secret Transmission, Uncond.)

$$\mathcal{R}_{st}^{u} = \mathcal{R}_1. \tag{15}$$

Theorem 3 (Zero-Leakage Secret Generation, Uncond.)

$$\mathcal{R}_{zsg}^{u} = \mathcal{R}_2. \tag{16}$$

Theorem 4 (Zero-Leakage Secret Transmission, Uncond.)

$$\mathcal{R}_{zst}^{u} = \mathcal{R}_2. \tag{17}$$

IV. Example: Binary Symmetric Double Source

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

$$\begin{aligned} I(U;Y) &= 1 - H(Y|U), \\ I(U;X) - I(U;Y) &= H(Y|U) - H(X|U). \end{aligned} \quad (18)$$

Mrs. Gerber's Lemma [13] tells us that if $H(X|U) = v$, then $H(Y|U) \geq h(q * h^{-1}(v))$, where $a * b = a(1-b) + b(1-a)$ and $h(a) \stackrel{\Delta}{=} -a\log(a) - (1-a)\log(1-a)$ 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 $H(Y|U) \geq h(q * p)$. For binary symmetric (U, X) with crossover probability p the minimum $H(Y|U)$ is achieved. Hence for private-key rates K we get

$$\begin{aligned} R^{\mathsf{u}}_{\mathsf{zsg}}(K) = R^{\mathsf{u}}_{\mathsf{zst}}(K) &= 1 - h(p), \\ &\text{for } p \text{ such that } h(q*p) - h(p) = K. \end{aligned} \quad (19)$$

For $q = 0.03$, 0.1, and 0.3 we have plotted the resulting secret-key vs. private-key rate functions in Fig. 5. From this figure we can observe that the private-key rate K is never larger than the secret-key rate R, and we can speak of key boosting.

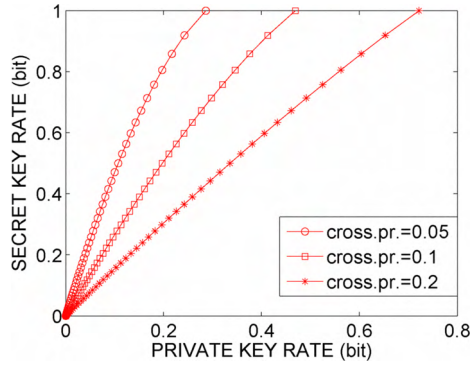

Fig. 5. Secret-key vs. private-key rate functions $R^{\mathsf{u}}_{\mathsf{zsg}}(\cdot)$, $R^{\mathsf{u}}_{\mathsf{zst}}(\cdot)$ for three values of q.

V. Proof of Thm. 3

The proof of this theorem consists of three parts. The first part, the converse, will be treated in detail. The achievability in the second part will only be outlined. The third part, the bound on cardinality of U, can be proven using the Fenchel-Eggleston strengthening the Caratheodory lemma, see [14].

A. Converse

First we consider the entropy of the secret. We use that $\widehat{S} = d(H, \underline{Y}, P)$ and Fano's inequality $H(S|\widehat{S}) \leq F$, where $F \stackrel{\Delta}{=} 1 + \Pr\{\widehat{S} \neq S\}\log(M_S) \leq 1 + \Pr\{\widehat{S} \neq S\}(N\log|\mathcal{X}| + \log(M_P))$. For achievable pairs (R, K) we have that

$$\begin{aligned} H(S) &= I(S; H, Y^N, P) + H(S|H, Y^N, P, \widehat{S}) \\ &\leq I(S; H, Y^N, P) + H(S|\widehat{S}) \\ &\leq I(S;H) + I(S;P|H) + \sum_{n=1}^{N} I(S; Y_n | H, Y^{n-1}, P) + F \\ &\leq I(S;H) + \log(M_P) + \sum_{n=1}^{N} I(S, H, X^{n-1}, P; Y_n) + F \\ &\leq N\delta + N(K+\delta) + NI(U;Y) + 1 + \delta\log(M_S). \end{aligned} \quad (20)$$

We used that $I(S, H, Y^{n-1}, P; Y_n) \leq I(S, H, X^{n-1}, Y^{n-1}, P; Y_n) = I(S, H, X^{n-1}, P; Y_n)$, since $Y^{n-1} \to (S, H, X^{n-1}, P) \to Y_n$. Moreover, we defined $U_n \stackrel{\Delta}{=} (S, H, X^{n-1}, P)$ and T to be time-sharing variable uniformly distributed over $\{1, 2, \ldots, N\}$ and independent of all other variables, and set $U \stackrel{\Delta}{=} (U_n, n)$, $X \stackrel{\Delta}{=} X_n$, and $Y \stackrel{\Delta}{=} Y_n$ for $T = n$. Then $U_n \to X_n \to Y_n$ and consequently $U \to X \to Y$ hold.

Now for achievable pairs (R, K) we have that

$$\begin{aligned} R - \delta &\leq \log(M_S)/N \leq H(S)/N + \delta \\ &\leq I(U;Y) + (1+\delta)K + 3\delta + \delta\log|\mathcal{X}| + \delta^2 + 1/N. \end{aligned} \quad (21)$$

In a similar manner we find for the leakage

$$\begin{aligned} 2N\delta &\geq I(X^N; H) + I(S; H) \geq I(X^N; H) \\ &= I(X^N, S, P; H) - I(P, S; H|X^N) \\ &= H(H) - H(P|X^N) - H(S|P, X^N) + H(P, S|X^N, H) \\ &\geq H(H) - H(P) \geq H(H, \widehat{S}|Y^N, P) - \log(M_P) \\ &= H(S, H, \widehat{S}|Y^N, P) - H(S|Y^N, P, \widehat{S}, H) - \log(M_P) \\ &\geq H(S, H|Y^N, P) - F - H(S, H|X^N, P) - \log(M_P) \\ &= I(S, H; X^N|P) - I(S, H; Y^N|P) - \log(M_P) - F \\ &= \sum_{n=1}^{N} I(S, H; X_n|P, X^{n-1}) - \sum_{n=1}^{N} I(S, H; Y_n|P, Y^{n-1}) \\ &\quad - \log(M_P) - F \\ &\geq \sum_{n=1}^{N} I(S, H, X^{n-1}, P; X_n) - \sum_{n=1}^{N} I(S, H, X^{n-1}, P; Y_n) \\ &\quad - \log(M_P) - F = NI(U;X) - NI(U;Y) - \\ &\quad N(1+\delta)K - N\delta - N\delta\log|\mathcal{X}| - N\delta^2 - 1. \end{aligned} \quad (22)$$

Letting $\delta \downarrow 0$ and $N \to \infty$, we may conclude from (21) that $R \leq I(U;Y) + K$. Also (22) after rearranging yields $K \geq I(U;X) - I(U;Y)$ and hence the converse.

B. Outline of the Achievability Proof

We start by fixing a conditional distribution $\{P(u|x), x \in \mathcal{X}, u \in \mathcal{U}\}$. This determines the joint distribution $P(u, x, y) = Q(x, y)P(u|x)$, 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 $\underline{u}$ with labels s. Each of those sequences also gets a random h-label. The h-label can assume roughly $2^{N(I(U;X)-I(U;Y))}$ values. Moreover, there is also a random uniformly generated private key p that assumes at least $2^{N(I(U;X)-I(U;Y))}$ values.

The encoder, upon observing the source sequence $\underline{x}$, outputs the s-label corresponding to the index of this sequence as a secret, and h-label, corresponding to $\underline{x}$, as helper data. The helper data are made uninformative in a one-time-pad way, using the private key p, resulting in helper data $h \oplus p$, where $\oplus$ denotes addition modulo M_P. The helper data are sent to the decoder. The decoder observes the helper data $h \oplus p$ and, using the private key p, recovers the helper label as $h \oplus p \ominus p$, where $\ominus$ is substraction modulo M_P. It also observes the source

sequence $\underline{y}$ and determines the source sequence $\widehat{\underline{u}}$ with an h-label matching the helper data, such that $(\widehat{\underline{u}}, \underline{y}) \in \mathcal{A}_\epsilon^{(N)}(UY)$. It can be shown that the decoder can reliably recover s now. Using the property of the proof that the index of $\underline{u}$ uniquely defines it, we can show that $\underline{u}$ is uniform, and hence also S is.

Now it is easy to check that the secrecy and privacy leakages are negligible, since $I(S; H \oplus P) \leq \log(M_H) - H(H \oplus P|S) \leq \log(M_H) - H(P|H, S) = 0$ and $I(X^N; H \oplus P) \leq \log(M_H) - H(H \oplus P|X^N) = \log(M_H) - H(P|X^N) = 0$.

Finally, note that if (R, K) is achievable, also $(R+\alpha, K+\alpha)$ for $\alpha > 0$ is. Just use the extra private-key rate α as extra secret symbols. Then with $\alpha = K - I(U;X) + I(U;Y)$ we obtain the achievability.

VI. Proof of Thm. 4

A. Converse

As in the converse for secret generation we obtain for achievable (R, K) that

$$H(S) \leq N(I(U;Y) + K + 2\delta) + 1 + \delta \log(M_S). \quad (23)$$

We used that $I(S, H, P, Y^{n-1}; Y_n) \leq I(S, H, P, X^{n-1}; Y_n)$, since also here $Y^{n-1} \to (S, H, P, X^{n-1}) \to Y_n$. As before we defined $U_n \triangleq (S, H, P, X^{n-1})$ and took a time-sharing variable T uniform over $\{1, 2, \ldots, N\}$ and independent of all other variables and set $U \triangleq (U_n, n)$, $X \triangleq X_n$, and $Y \triangleq Y_n$ for $T = n$. Now again $U_n \to X_n \to Y_n$ and consequently $U \to X \to Y$ hold. For achievable (R, L) we obtain

$$\begin{aligned} R - \delta &\leq \log(M_S)/N = H(S)/N \\ &\leq 1/(1-\delta)\,(I(U;Y) + K + 2\delta + 1/N)\,. \end{aligned} \quad (24)$$

Similarly, we can write for the privacy leakage

$$\begin{aligned} 2N\delta &\geq I(X^N;H) + I(S;H) \geq I(S,H,P;X^N) - I(S,P;X^N|H) \\ &\geq \sum_{n=1}^{N} I(S, H, P, X^{n-1}; X_n) - H(P|H) - H(S|P, H) \\ &\geq NI(X;U) - H(P) - H(S, Y^N|P, H) + H(Y^N|P, H, S) \\ &\geq NI(X;U) - H(P) - I(Y^N; P, H, S) - H(S|P, H, Y^N) \\ &\geq NI(X;U) - NK - N\delta - NI(Y;U) - H(S|\widehat{S}) \\ &\geq N(I(X;U) - I(Y;U) - K - \delta - N\delta \log(M_S) - N), \end{aligned} \quad (25)$$

where U is defined as before.

If we now let $\delta \downarrow 0$ and $N \to \infty$, then we find that $R \leq I(U;Y) + K$ from (24) and that $K \geq I(U;X) - I(U;Y)$ from (25) after rearranging. The converse is now complete.

B. Outline of the Achievability Proof

The achievability proof is based on the achievability proof of Thm. 3. The difference is that we use an additional masking layer that uses the generated secret S_g in a one-time pad system to hide the transmitted secret S_t, such that $H_t = S_t \oplus S_g$ is an additional helper data, where operations $\oplus$ and $\ominus$ are modulo M_S. Such a masking layer was also used by Ahlswede and Csiszár [1].

Now keeping in mind that S_t is uniform on $\{1, 2, \ldots, M_S\}$ and independent of X^N, the generated secret S_g, and the corresponding helper data H_g, and that S_g is an achievable secret-key rate satisfying (8), we obtain $I(S_t; H_g \oplus P, H_t) = I(S_t; H_g \oplus P) + I(S_t; H_t|H_g \oplus P) = H(S_t \oplus S_g|H_g \oplus P) - H(S_t \oplus S_g|H_g \oplus P, S_t) \leq \log(M_S) - H(S_g|H_g \oplus P, S_t) \leq I(S_g; H_g \oplus P) + N\delta \leq 2N\delta$. Moreover, we get $I(X^N; H_g \oplus P, H_t) = I(X^N; H_g \oplus P) + I(X^N; H_t|H_g \oplus P) \leq N\delta + H(S_t \oplus S_g) - H(S_t \oplus S_g|H_g \oplus P, X^N) \leq N\delta + \log(M_S) - H(S_t|P, X^N) = N\delta$. Note that $S_t = \widehat{S}_t$ only if $S_g = \widehat{S_g}$, and thus $Pr\{S_t \neq \widehat{S}_t\} \leq \delta$ for achievable S_g. This finalizes the achievability proof.

VII. Conclusions

In this paper we have considered privacy leakage in biometric systems. We have investigated systems without an extra private key, and determined how the generated secret-key rate relates to the privacy leakage. We have also considered a version of this setup in which the secret-key is chosen uniformly and transmitted.

For the setting in which an extra private key is used by both terminals, we have focussed on the private-key rate needed to guarantee negligible privacy leakage for a certain secret-key rate. We considered both cases where the key is generated and where the key is arbitrarily chosen and then transmitted. For all cases we could determine the fundamental limits.

Detailed proofs can be found in [8].

References

[1] R. Ahlswede and I. Csiszár, "Common randomness in inf. theory and cryptography - part I: Secret sharing," *IEEE Trans. on Inf. Theory*, vol. 39, pp. 1121–1132, July 1993.

[2] ——, "Common randomness in inf. theory and cryptography - part II: CR capacity," *IEEE Trans. on Inf. Theory*, vol. 44, pp. 225–240, 1998.

[3] T. Cover, "A proof of the data compression theorem of Slepain and Wolf for ergodic sources," *IEEE Trans. on Inf. Th.*, vol. 22, pp. 226–228, 1975.

[4] I. Csiszár and P. Narayan, "CR and secret key generation with a helper," *IEEE Trans. on Inf. Th.*, vol. 46, pp. 344–366, 2000.

[5] D. Gündüz, E. Erkip, and H. V. Poor, "Secure lossless compression with side inf." in *In Proc. of the IEEE ITW, Porto, Portugal*, 2008.

[6] T. Ignatenko and F. Willems, "Privacy leakage in biometric secrecy systems," in *Proc. of 46th Ann. Allerton Conf. on Comm., Cont., and Comp., Sept. 23-26 2008, Monticello, IL.*

[7] ——, "On the security of the xor-method in biometric authentication systems," in *Proc. of 27th Symp. on Inf. Theory in the Benelux, Noordwijk, The Netherlands*, 2006, pp. 197–204.

[8] ——, "Biometric authentication systems: Privacy and security aspects," *submitted to IEEE Trans. on Inf. Forensics and Security*, Sept. 19, 2008.

[9] L. Lai, S.-W. Ho, and H. V. Poor, "Privacy-security tradeoffs in biometric security systems," in *Proc. of 46th Ann. Allerton Conf. on Comm., Cont., and Comp., Sept. 23-26 2008, Monticello, IL.*

[10] U. Maurer, "Secret key agreement by public discussion from common inf." *IEEE Trans. on Inf. Theory*, vol. 39, pp. 733–742, May 1993.

[11] V. Prabhakaran and K. Ramchandran, "On secure distributed source coding," *In Proc. of the IEEE ITW*, pp. 442–447, Sept. 2007.

[12] A. Smith, "Maintaining secrecy when inf. leakage is unavoidable," Ph.D. dissertation, MIT, 2004.

[13] A. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications–I," *IEEE Trans. on Inf. Th.*, vol. 19, pp. 769–772, 1973.

[14] ——, "The rate-distortion function for source coding with side inf. at the decoder," *IEEE Trans. on Inf. Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[15] C. Ye and P. Narayan, "Secret and private key constructions for simple multiterminal source models," in *In Proc. of the IEEE ISIT, Adelaide, Australia*, September 4 - 9 2005, pp. 2138 – 2141.