PRIVACY-LEAKAGE CODES FOR BIOMETRIC AUTHENTICATION SYSTEMS

Tanya Ignatenko, Frans M.J. Willems

Electrical Engineering Department
Eindhoven University of Technology
Den Dolech 2, 5612 AZ Eindhoven, The Netherlands

ABSTRACT

In biometric privacy-preserving authentication systems that are based on key-binding, two terminals observe two correlated biometric sequences. The first terminal selects a secret key, which is independent of the biometric data, binds this secret key to the observed biometric sequence and communicates it to the second terminal by sending a public message. This message should only contain a negligible amount of information about the secret key, but also leak as little as possible about the biometric data. Current approaches to realize such biometric systems use fuzzy commitment with codes that, given a secret-key rate, can only achieve the corresponding privacy-leakage rate equal to one minus this secret-key rate. However, the results in Willems and Ignatenko [2009] indicate that lower privacy leakage can be achieved if vector quantization is used at the encoder. In this paper we study the use of convolutional and turbo codes applied in fuzzy commitment and its modifications that realize this.

Index Terms— Biometric authentication, privacy, BCH codes, convolutional codes, turbo codes

1. INTRODUCTION

Privacy problems related to the use of biometric data in various access control systems have been attracting attention of the research community for more than a decade. As pointed out by Schneier [1], an important property of biometric data is that they cannot be easily canceled and substituted with new biometrics, as they are unique for individuals. Therefore secure storage and communication of biometric information in the corresponding access control systems becomes crucial.

Common secret sharing concepts introduced by Maurer [2] and slightly later Ahlswede and Csiszar [3] play important role in biometric systems with template protection, and set the ground for biometric privacy-preserving authentication systems based on key-binding or key transmission. In these systems two terminals observe two correlated biometric sequences. The first terminal chooses a random secret key, which is independent of biometrics, and forms a helper message based on the observed biometric sequence and the chosen secret. This helper message facilitates reliable reconstruction of the selected secret key at the second terminal given the second observation of the biometric sequence. The first terminal here represents enrollment, while the second terminal performs authentication. The helper data and secret key, encrypted using a one-way function, are stored in a biometric database. We assume the biometric database to be public, since one cannot guarantee its robustness to outside or inside attacks. Therefore, to ensure secure system access, the secret-key rates should be as large as possible and, moveover, the helper message has to contain only a negligible amount of information about the secret key. On the other hand, to ensure privacy, the helper data have to contain as little as possible information about the biometrics, i.e. the privacy leakage has to be small. The fundamental trade-offs between secret-key rates and privacy-leakage rates for this type of systems for discrete biometric sources were determined in [4], [5], and for Gaussian biometric sources in [6].

Practical constructions for biometric privacy-preserving authentication systems based on key-binding include fuzzy commitment based schemes, proposed by Juels and Wattenberg [7]. Since fuzzy commitment is designed for binary sequences, enrollment and authentication biometric sequences are binary quantized in such schemes, see e.g. [8]. However, it was demonstrated in [6], that binary quantization results into performance loss. Ye et al. [9] considered the Gaussian case and applied scalar multi-level quantization instead of binary quantization at the encoder side. Moreover, in their schemes they used soft decision during authentication. Therefore the resulting scheme improves upon fuzzy commitment schemes with respect to the secret-key rate.

It should be noted that the above techniques were focussing on secret-key rates only. As a result fuzzy commitment is not optimal with respect to privacy leakage. Although Ye et al. [9] did not concentrate on privacy leakage, their method effects the balance between privacy-leakage and secret-key rate. The scalar quantization that they used is not optimal. It was shown in [6] that to achieve the optimal trade-off vector quantization should be applied.

In this paper we focus on coding schemes for biometric authentication that control privacy leakage. We assume that our biometric data sequences are produced by Gaussian sources, and study the performance of a number of coding

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

techniques that can be used in combination with fuzzy commitment and its modifications. We show how fuzzy commitment can be improved by only using quantization of biometric sequences during enrollment. Moreover, we present a coding scheme with vector quantization that achieves a better secret-key vs. privacy-leakage trade-off than fuzzy commitment.

2. BIOMETRIC AUTHENTICATION BASED ON KEY-BINDING

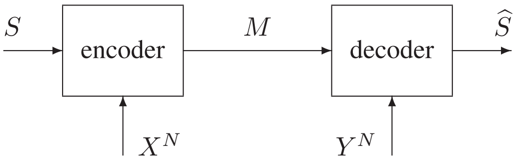

Fig. 1. Model for a biometric system with key-binding.

A Gaussian biometric system is based on a *Gaussian biometric source* $\{G_\rho(x,y), x \in \mathbf{R}, y \in \mathbf{R}\}$ that produces an X-sequence $x^N = (x_1, x_2, \cdots, x_N)$ with N real-valued symbols and a Y-sequence $y^N = (y_1, y_2, \cdots, y_N)$ also having N real-valued components. The density corresponding to sequence pair (X^N, Y^N) is given by

$$p_{X^N,Y^N}(x^N, y^N) = \prod_{n=1}^{N} G_\rho(x_n, y_n), \text{ where} \quad (1)$$

$$G_\rho(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}\right), \quad (2)$$

for $x \in \mathbf{R}, y \in \mathbf{R}$, and correlation coefficient $|\rho| < 1$. Thus, the pairs $\{(X_n, Y_n), n = 1, \cdots, N\}$ are independent of each other and identically distributed (i.i.d.) according to $G_\rho(\cdot,\cdot)$. Note that scaling can always be applied to obtain unit X-variance and unit Y-variance. The signal-to-noise ratio, SNR, for a virtual channel from X to Y relates to the correlation coefficient ρ as

$$\text{SNR} = \rho^2/(1-\rho^2). \quad (3)$$

Consider now a Gaussian biometric system with key-binding, see Fig. 1. In this system a secret key S is chosen uniformly and independently of biometric sequences from alphabet $\{1, 2, \ldots, |\mathcal{S}|\}$, thus

$$\Pr\{S = s\} = 1/|\mathcal{S}|, \text{ for all } s \in \{1, 2, \cdots, |\mathcal{S}|\}. \quad (4)$$

The encoder observes the biometric enrollment source sequence X^N and the secret S and produces helper data M, hence $M = e(S, X^N)$, where $e(\cdot,\cdot)$ is the encoder mapping and the helper data index is from alphabet $\{1, 2, \cdots, |\mathcal{M}|\}$. The helper data are assumed to be public.

The helper data M are sent to the decoder that also observes the biometric authentication sequence Y^N. This decoder forms an estimate $\widehat{S}$ of the chosen secret, hence $\widehat{S} = d(M, Y^N)$, where $d(\cdot,\cdot)$ is the decoder mapping. The decoders estimate of the secret also assumes values from $\{1, 2, \cdots, |\mathcal{S}|\}$.

In this system the helper data are considered to be public. Thus, the goal of this system is to transmit the secret key with negligible error probability and negligible secrecy-leakage rate, while realizing secret-key rates as large as possible and privacy-leakage rates as small as possible. This corresponds to the following definition of achievability.

Definition 1 *In a Gaussian biometric system with key-binding, a secret-key vs. privacy-leakage rate pair (R, L) with $R \geq 0$ is achievable if for all $\delta > 0$ for all N large enough there exist encoders and decoders such that*

$$\begin{aligned} \Pr\{\widehat{S} \neq S\} &\leq \delta, \\ \log_2 |\mathcal{S}|/N &\geq R - \delta, \\ I(S; M)/N &\leq \delta, \\ I(X^N; M)/N &\leq L + \delta. \end{aligned} \quad (5)$$

Moreover, let $\mathcal{R}_\rho$ be the region of all achievable secret-key vs. privacy-leakage rate pairs.

The characterization of $\mathcal{R}_\rho$ is given by the following theorem.

Theorem 1 *[Key-binding based on Gaussian sources, [6]]*

$$\begin{aligned} \mathcal{R}_\rho = \{(R, L) \quad : \quad & 0 \leq R \leq \frac{1}{2}\log_2\left(\frac{1}{\alpha\rho^2 + 1 - \rho^2}\right), \\ & L \geq \frac{1}{2}\log_2\left(\frac{\alpha\rho^2+1-\rho^2}{\alpha}\right), \\ & \textit{for } 0 < \alpha \leq 1\}. \end{aligned} \quad (6)$$

Now if we define the rate-leakage function as follows

$$R_\rho(L) \triangleq \max_{(R,L)\in\mathcal{R}_\rho} R, \quad (7)$$

we can write

$$R_\rho(L) = \frac{1}{2}\log_2\left(1 + \text{SNR}\frac{(2^{2L}-1)}{2^{2L}}\right). \quad (8)$$

From this function we can see that

$$\lim_{L\to\infty} R_\rho(L) = \frac{1}{2}\log_2(1 + \text{SNR}) = I(X; Y). \quad (9)$$

Note that $I(X; Y)$ is the secret-key capacity for our biometric system, and thus this capacity is achievable at infinitely large privacy leakage. Therefore, to obtain a biometric system that has a good control on privacy leakage, one would be interested in operational points (secret-key and privacy-leakage rate pairs) that have large secret-key to privacy-leakage rate ratio for a given SNR.

If we rewrite (8) in the following way

$$\text{SNR}_{\text{min}}(R, L) = (2^{2R} - 1)\frac{2^{2L}}{2^{2L}-1}, \quad (10)$$

we obtain a fundamental limit for biometric authentication. It gives us the minimal SNR required to achieve reliable key reconstruction and thus authentication in a biometric system with given secret-key and privacy-leakage rates.

3. CODING FOR BIOMETRIC AUTHENTICATION WITH KEY-BINDING

Now we turn to the problem of selecting a good code that can realize biometric authentication with privacy protection.

Consider a biometric system whose inputs are biometric data sequences with Gaussian i.i.d. continuous components. Here we analyze as an example a biometric source, which is Gaussian with a target SNR equal to 3. This corresponds to maximum rate 1. This source produces biometric sequences of length $N = 512$. Suppose we need a system that operates at error probability characterized by a word error rate (WER) of roughly 0.01. Note that in this case the WER is equivalent to the false rejection rate (FRR), as WER characterizes the probability of correct reconstruction of the whole secret key.

3.1. Coding with Binary Quantization at Both Sides

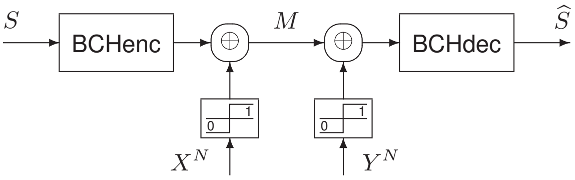

Fig. 2. Fuzzy commitment with BCH codes.

We start with coding for fuzzy commitment, see Fig. 2, introduced by Juels and Wattenberg [7]. Fuzzy commitment is realization of a biometric system with key-binding for binary biometric sources. In fuzzy commitment a chosen secret key S is encoded into a codeword using a certain error-correcting code (ECC) of rate R, and this codeword is then added modulo-2 to the observed biometric enrollment sequence. The result is released as helper data. During authentication the decoder observes a biometric authentication sequence and subtracts it modulo-2 from the helper data. The result is decoded to a closest codeword in the corresponding ECC, which leads to the secret key $\widehat{S}$.

Since fuzzy commitment operates on binary data, we have to apply binary quantization to both the enrollment and authentication sequences. Note that the resulting "channel" crossover probability is $\frac{2}{\pi} \arctan \sqrt{\frac{1-\rho}{1+\rho}}$, see [6] for the details. Now we use the quantized binary sequences in a fuzzy commitment scheme with a BCH code, see e.g. [10], of length $N = 511$ and message length (secret-key length) 31. This code can correct up to $t = 109$ errors. If we look at the biometric system characterization in terms of secret-key rate and privacy leakage, we see that for this code the (code and) secret-key rate is $R = 31/511 = 0.0607$ and the privacy leakage is $L = 480/511 = 0.9393$. For these rates $\text{SNR}_{\text{min}}(0.0607, 0.9393) = -9.2\text{dB}$.

Using computer simulations, we see that this BCH code achieves the target performance of 0.01 at SNR of 4.3dB, see Fig. 3. Thus fuzzy commitment combined with the BCH code is 13.5dB from optimal. The reason for this poor behavior can be explained by two effects. Firstly, binary quantization at two-sides is not optimal and, secondly, a BCH code is not powerful enough here.

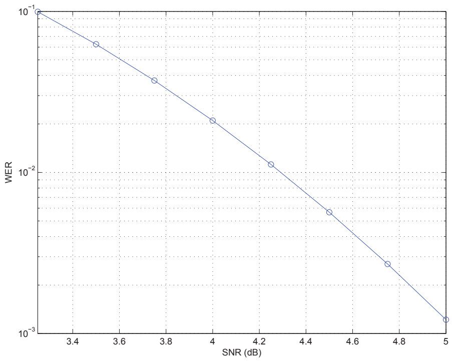

Fig. 3. Performance of $(511, 31, 109)$ BCH code.

3.2. Binary Quantization at the Encoder Only

In order to improve the performance of fuzzy commitment, we modify it by applying quantization only at the encoder side, see Fig. 4.

Then, at first, we use a 16-state non-systematic convolutional code of rate $1/4$ with generator polynomials $25, 27, 33$, and 37 (octal). This code has minimum free distance of 16. For this code we take the trellis length equal to 128, and the codeword length $N = 512$. Decoding is performed using the Viterbi algorithm [11] using soft information based on the sequence Y^N and the helper data M with metrics $m_0 = Q((1-2m)\rho y/\sqrt{1-\rho^2})$ and $m_1 = Q((2m-1)\rho y/\sqrt{1-\rho^2})$, see [9]. For this system, the secret-key rate is $R = 124/512 = 0.2422$, while the privacy leakage is $L = 388/512 = 0.7578$. Thus $\text{SNR}_{\text{min}}(0.2422, 0.7578) = -2.1$ dB.

The performance of this code is shown in Fig. 5. We see that now a WER of 0.01 is achieved at SNR $= 5.3$dB. Therefore the modified scheme with the convolutional code is 7.4dB from optimal and 6dB better than fuzzy commitment with the BCH code. Moreover, observe that now we achieve a secret-key length four times larger than the one that was achieved with the BCH code.

Next, we use a parallel-concatenated turbo code of rate $1/3$. The constituent codes have 8 states, trellis length of 169, and we use a 13×13 block-interleaver. The codeword length

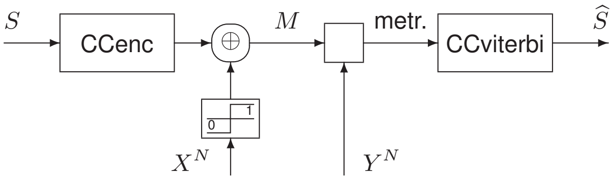

Fig. 4. Quantization at the encoder only, convolutional codes.

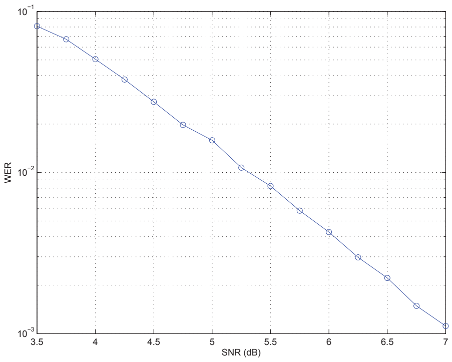

Fig. 5. Performance of rate $1/4$ 16-state convolutional code.

here is $N = 507$, and the secret-key length is 163 bits. Decoding is performed using turbo decoding with soft information based on Y^N and M with the same metrics as before. Now the secret-key rate is $R = 163/507 = 0.3215$ and the privacy-leakage rate is $L = 344/507 = 0.6785$. Thus we have $\text{SNR}_{\text{min}}(0.3215, 0.6785) = -0.4$ dB.

Computer simulations, see Fig. 6, show that now we need $\text{SNR} = 3.9$ dB to achieve a WER of 0.01. Thus with turbo codes we are 4.3dB from the fundamental limit, that is 3.1dB better than the system with convolutional codes. Also the secret-key and privacy-leakage rates are improved again.

3.3. Coding with Vector Quantization at the Encoder

Note that all the methods considered before could only realize secret-key and privacy-leakage pairs for which $R + L = 1$. However, the fundamental regions show that it should be possible to achieve better trade-offs, i.e. with $L < 1 - R$ for a given R. We have to use vector quantization for this.

Instead of scalar binary quantization, we now apply a vector quantizer based on a 4-state convolutional code of rate $2/3$. We use Viterbi decoding to find the corresponding codeword for the enrollment sequence X^N, followed by 13×13 block-interleaver. The result is added modulo-2 to a codeword produced by a 4-state convolutional code of rate $1/2$. This structure allows for serial iteration, assisted with the helper data, see Fig. 7. For this system we have the secret-key rate $R = 167/507 = 0.3294$ and the privacy-leakage rate $L = 173/507 = 0.3412$. This pair corresponds to $\text{SNR}_{\text{min}}(0.3294, 0.3412) = 1.9$dB.

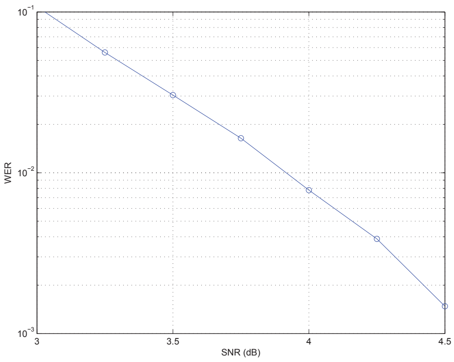

Fig. 6. Performance of turbo code of rate $1/3$.

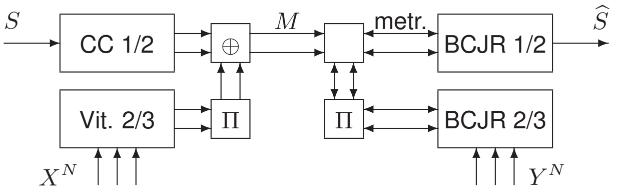

Fig. 7. Vector quantization at the encoder.

Simulations show that $\text{SNR} = 7.1$dB is required to get a WER of 0.01, see Fig. 8. Observe that now we are at 5.2dB from the fundamental limit. Even though we have a slight degradation comparing to the performance of the system with the turbo code and binary quantization at the encoder, here we have a real privacy-leakage controlling system. Observe that the secret-key rate here is roughly the same as in the system with the turbo code, but the privacy leakage is almost twice as low. Thus this approach paves the road to biometric systems with much smaller privacy leakage than in the current fuzzy commitment based systems.

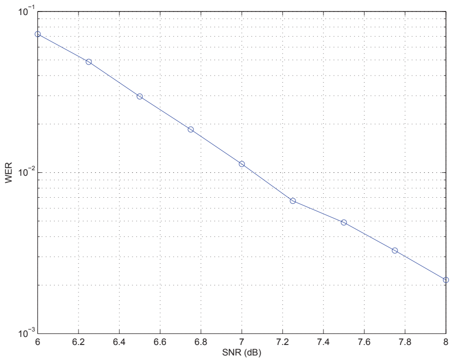

Fig. 8. Performance of the system with vector quantization.

4. CONCLUSIONS

In this paper we considered coding techniques for Gaussian biometric systems with privacy protection. We proposed the concept of minimal SNR for a given secret-key and privacy-leakage rates as a fundamental limit for evaluating the performance of biometric systems. We showed that fuzzy commitment based on a BCH code is 13.5dB from the fundamental limit. Then we modified fuzzy commitment to incorporate soft information at the decoder, and used a convolutional and turbo codes there. Our convolutional code appeared to be 7.4dB from the fundamental limit, while the turbo code only 4.3dB away from it. Next we deployed vector quantization based on Viterbi decoding at the encoder and a convolutional code to create a system that is 5.2dB from the fundamental limit. It is remarkable that this system has privacy-leakage twice as low as the one achieved with the turbo code.

5. REFERENCES

[1] Bruce Schneier, “Inside risks: the uses and abuses of biometrics,” *Communications of the ACM*, vol. 42, no. 8, pp. 136, 1999.

[2] U. Maurer, “Secret key agreement by public discussion from common information,” *IEEE Trans. Inf. Theory*, vol. 39, pp. 733–742, May 1993.

[3] 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.

[4] T. Ignatenko and F. Willems, “Biometric systems: Privacy and secrecy aspects,” *IEEE Trans. Inf. Forensics and Security*, vol. 4, no. 4, December 2009.

[5] Lifeng Lai, Siu-Wai Ho, and H. Vincent Poor, “Privacy-security tradeoffs in biometric security systems,” in *Proc. of 46th Ann. Allerton Conf. on Comm., Control, and Computing, Sept. 23-26 2008, Monticello, IL, USA*, 2008.

[6] F.M.J.Willems and T. Ignatenko, “Quantization effects in biometric systems,” in *Proc. of Workshop ITA (Information Theory and its Applications), 8-13 February 2009, San Diego, CA, USA*, 2009.

[7] A. Juels and M. Wattenberg, “A fuzzy commitment scheme,” in *6th ACM Conf. on Computer and Communications Security*, 1999, pp. 28–36.

[8] P. Tuyls, A.H.M. Akkermans, T.A.M. Kevenaar, G.-J. Schrijen, Bazen A, and R.N.J. Veldhuis, “Practical biometric authentication with template protection,” in *5th Int. Conf. on Audio- and Video-Based Personal Authentication (AVBPA)*, 2005, pp. pp. 436–446.

[9] C. Ye, A. Reznik, and Y. Shah, “Extracting secrecy from gaussian random variables,” in *IEEE Int. Symp. Inf. Theory (ISIT), Seattle, USA*, July 9 - 14 2006, pp. 2593–2597.

[10] William Cary Huffman and Vera Pless, *Fundamentals of Error-Correcting Codes*, Cambridge University Press, New York, NY, USA, 2003.

[11] A. J. Viterbi, “Error bounds for convolutional codes and an asymptotically optimum decoding algorithm,” *IEEE Trans. Inf. Theory*, vol. 13, pp. 260–269, April 1967.