ZERO LEAKAGE QUANTIZATION SCHEME FOR BIOMETRIC VERIFICATION

J.A. de Groot and J.-P.M.G. Linnartz

Eindhoven University of Technology
Department of Electrical Engineering
Den Dolech 2, Eindhoven, the Netherlands

ABSTRACT

Biometrics gain increasing interest as a solution for many security issues, but privacy risks exist in case we do not protect the stored templates well. This paper presents a new verification scheme, which protects the secrets of the enrolled users. We will show that zero leakage is achieved if certain criteria are met and we benchmark the performance of this scheme. We quantify performance loss in terms of detection and false acceptance rate and capacity of the biometric channel, which are slightly worse than those of the current leaky methods.

Index Terms— biometrics, verification scheme, leakage, privacy protection

1. INTRODUCTION

Biometrics gain increasing interest as a solution to many security issues. New passports include biometric data of the owner and laptops nowadays almost always include a fingerprint reader for login.

It is likely that biometric templates will be stored in a public database for verification purposes. Identity theft becomes a very real threat if we do not protect the confidentiality of the enrolled data. Applying an encryption-decryption scheme will not work since a dishonest verifier is then able to steal the decrypted secret and use it to his advantage.

A password, which can be considered a secret, can be protected against a dishonest verifier by applying a cryptographic hash function. When the user authenticates himself, the same hash is applied and only the hashed passwords are compared.

Biometric features however are not exactly reproducible. Various solutions have been proposed to handle these deviations by mapping the most likely values for an individual to a single value of the secret. Among others the Fuzzy Commitment Scheme (FCS) [1], Helper Data Scheme (HDS) [2], Fuzzy Extractors [3], Fuzzy Vault [4], Cancellable Biometrics [5] and Likelihood based approaches [6, 7] have been proposed. A common property is that individual prover-dependent 'helper data' is required for mapping [2]. The required helper data might leak information about the secret, which a dishonest verifier could exploit.

Such privacy protection deteriorates the performance in terms of detections [8, 9]. Our experience shows that practical (low-dimensional) biometrics require careful design to avoid undesirable performance losses, although for the limiting case, the secret capacity is not affected by a privacy protection scheme [10].

The second section of this paper introduces a new scheme that is capable of leaking no information about the enrolled secret. In section 3 we study under which conditions zero leakage is achieved. We extend the commonly used model that an attacker only has access to the template database, by assuming that he also has some a priori knowledge about the approximate biometrics statistics of a particular prover. Subsequently we will derive the performance loss of the scheme in terms of detection rate in section 4 and channel capacity in section 5. Finally we will discuss and conclude on our results in sections 6 and 7.

2. A NEW VERIFICATION SCHEME

We consider a verification scheme as introduced in [2], which consists of an enrollment and verification phase. In the Enrollment phase the prover provides his biometric data $\underline{x} = (x_1, \ldots, x_M)$, which the systems stores safely in the hashed form $\{h(\underline{s}), \underline{w}\}$, where $\underline{s}$ is a digitized version of $\underline{x}$. In the verification phase the prover provides his correlated biometric data $\underline{y} = (y_0, \ldots, y_M)$ to prove his identity. The entire scheme, including our modifications, is depicted in Fig. 1.

We have extended the scheme with a pre-distortion function, $u_i = g_i(x_i)$ and $v_i = g_i(y_i)$, before creating or applying the helper data. The arbitrary continuous and bounded probability densities of the biometric features are transformed to uniform distributions $u_i, v_i \sim \mathcal{U}(0,1)$ by the function g_i.

Any r.v. x with a continuous distribution F_{X} can be transformed to a uniform distribution $\mathcal{U}(0,1)$, by distorting it according to $u = F_{\mathrm{X}}(x)$. In fact, since $0 \leq F_{\mathrm{X}}(x) \leq 1$ we know that for $0 \leq u \leq 1$,

$$F_{\mathrm{U}}(u) = \mathrm{P}(U \leq u) = \mathrm{P}(X \leq x) = F_{\mathrm{X}}(x) = u. \qquad (1)$$

So the probability density of u is uniform, namely

$$f_{\mathrm{U}}(u) = 1 \qquad \text{for} \quad 0 \leq u \leq 1, \qquad (2)$$

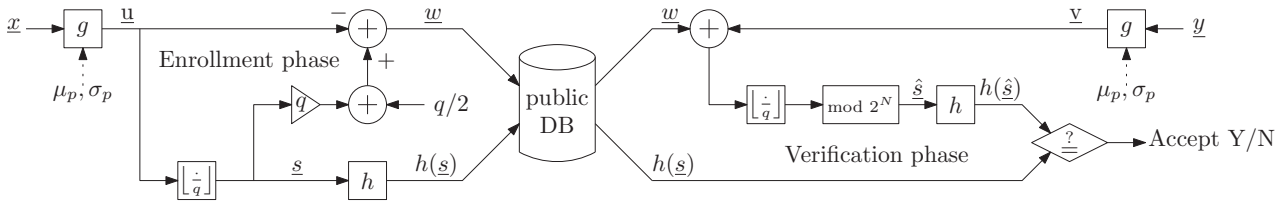

Fig. 1. Zero leakage verification scheme. Extension of [2] with pre-distortion function g.

and zero elsewhere. Therefore we choose $g_i(x) = F_{\mathrm{X_i}}(x)$. This yields zero leakage between helper data w_i and the corresponding enrolled secret s_i, as will be shown in the next section.

On $\underline{u} = (u_1, \ldots, u_M)$ we subsequently define 2^N quantization intervals, each having a width of $q = 2^{-N}$, to obtain a secret $\underline{s} = \lfloor \underline{u}/q \rfloor$.

Helper data $\underline{w} = q\,\underline{s} + q/2 - \underline{u}$ is used to center the pre-distorted enrollment samples $\underline{u}$ on the quantization interval. The helper data might push the verification samples $v_i = g_i(y_i)$ outside the standard interval. Therefore a modulus 2^N operation in applied directly after quantization to obtain the estimated secret $\hat{\underline{s}}$, hence

$$\hat{\underline{s}} = \left\lfloor \frac{g(\underline{y}) - g(\underline{x})}{q} + \frac{1}{2} \right\rfloor + \underline{s} \mod 2^N. \tag{3}$$

3. LEAKAGE ANALYSIS

We assume that the attacker knows the scheme as depicted in Fig. 1, including the applied pre-distortion function g, and that he knows all information in the public database. This knowledge should not leak any information about the provers secret $\underline{s}$.

To prove that the scheme indeed achieves zero leakage we extend the results of [2]. Our helper data is obtained by $w_i = q \cdot \lfloor u_i/q \rfloor + q/2 - u_i$. Since the transformed enrollment sample u_i is distributed uniformly, the helper data will be distributed uniformly as well, moreover it is independent of s,

$$f_{\mathrm{W}}(w|S=n) = f_{\mathrm{W}}(w) = \begin{cases} \frac{1}{q} & \text{for } -\frac{q}{2} < w \leq \frac{q}{2} \\ 0 & \text{otherwise} \end{cases}. \tag{4}$$

The leakage expressed as mutual information between helper data and secret can be calculated as

$$\begin{aligned} I(W;S) &= H(S) - H(S|W) \qquad (5) \\ &= -\sum_{n=0}^{2^N-1} \mathrm{P}(S=n) \log_2 \mathrm{P}(S=n) \\ &\quad + \int_{-q/2}^{q/2} \sum_{n=0}^{2^N-1} f_W(w|S=n)\mathrm{P}(S=n) \\ &\quad \log_2 \frac{f_W(w|S=n)}{f_W(w)} \mathrm{P}(S=n)\, dw. \qquad (6) \end{aligned}$$

Due to the uniform distribution of the helper data (4) the second term simplifies to

$$H(S|W) = \int_{-q/2}^{q/2} \frac{1}{q} \sum_{n=0}^{2^N-1} \mathrm{P}(S=n) \log_2 \mathrm{P}(S=n)\, dw. \tag{7}$$

Working out the integral exactly yields the entropy $H(S)$ of the secret, but with opposite sign in (5), therefore the leakage is zero. Also note that due to the uniform distribution each secret is equally likely, hence

$$P(S=n) = q \qquad \Rightarrow \qquad H(S) = N. \tag{8}$$

The situation changes if an attacker has a better understanding of the biometric distributions than the system designer, who determined the pre-distortion function. To illustrate this case, we will assume the population distribution to be Gaussian $\mathcal{N}(0,1)$. Suppose that the attacker knows that a certain prover belongs to a subgroup with average and variance (μ_g, σ_g^2). The attacker is unaware of the provers individual biometric, but the group statistics are more specific than those of the entire population ($\mu_p = 0, \sigma_p^2 = 1$), hence $\sigma_g^2 < \sigma_p^2$.

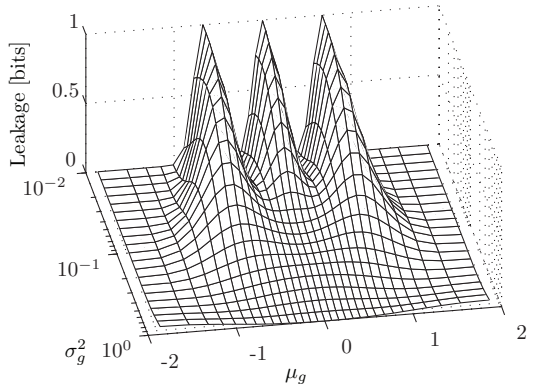

Fig. 2. Leakage for an informed attacker in a mismatched scheme for $N = 2$ bit/dimension.

In this case, the amount of information I_A that an attacker eventually has, consists of a priori knowledge, given by the group distribution, and additional information from the helper data (Fig. 2). Both types of information depend on μ_g and σ_g^2. The a priori knowledge can be calculated as follows:

$$I_A = H(S) - H(S|\mu_g, \sigma_g^2) \tag{9}$$

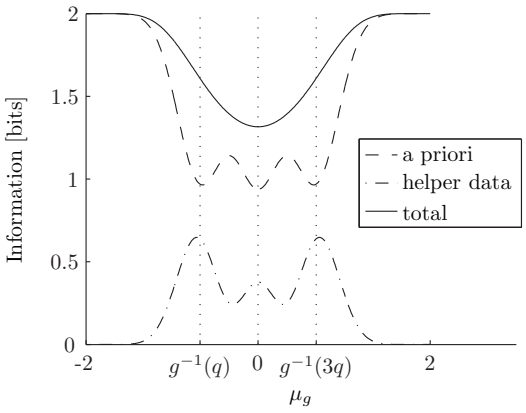

Fig. 3. Information for an informed attacker in a scheme with $N = 2$ and a group variance $\sigma_g^2 = .063$. The quantization interval boundaries are depicted as dotted lines at $g^{-1}(q)$, $g^{-1}(2q) = 0$ and $g^{-1}(3q)$.

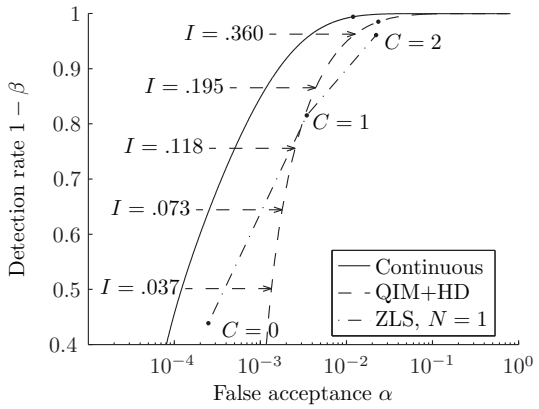

Fig. 4. Receiver operator curves for $M = 10$ and $\rho = .9$.

in which $H(S) = N$, the number of bits, and $H(S|\mu_g, \sigma_g^2)$, the actual entropy of the secret, which can be calculated by the following probability function

$$P(S = n) = F_{\mathrm{X}}\left(\frac{F_{\mathrm{X}}^{-1}(q\,(n+1)) - \mu_g}{\sigma_g}\right) - F_{\mathrm{X}}\left(\frac{F_{\mathrm{X}}^{-1}(q\,n) - \mu_g}{\sigma_g}\right). \quad (10)$$

The leakage can be calculated by using (6) with

$$\mathrm{P}(S = n) = \int_{qn}^{q(n+1)} f_{gp}\left(g^{-1}(\nu)\right) d\nu, \quad (11)$$

$$f_{\mathrm{W}}(w|S = n) = \frac{f_{gp}\left(g^{-1}(q(n + 1/2) - w)\right)}{\mathrm{P}(S = n)}, \quad (12)$$

in which f_{gp} describes the ratio between the actual (group) density and the (population) density assumed in the detector for a Gaussian biometric,

$$f_{gp}(x) = \frac{1}{\sigma_g} \exp\left(\frac{x^2}{2} - \frac{(x - \mu_g)^2}{2\sigma_g^2}\right). \quad (13)$$

Note that (12) is only valid for $n \in \{0, \ldots, 2^N - 1)$ and $-q/2 < w \leq q/2$ and is zero otherwise. By combining (11) and (12) we get $f_{\mathrm{W}}(w)$, which we insert in (6). An attacker who knows, for example, that a provers feature is very close to one of the quantization boundaries, i.e. $\mu \approx g^{-1}(q)$ and $\sigma < 1$, is able to acquire a significant amount of the total information from the helper data. In Fig. 3 this is illustrated for a Gaussian distributed biometric.

4. DETECTION PERFORMANCE

In order to derive the performance in terms of false rejection rate (FRR, β) and false acceptance rate (FAR, α) we derived the probability of accepting the verification for a single dimension first. For ease of analysis we give an example for i.i.d. standard normal features. Furthermore, the biometric features of genuine provers during enrollment and verification are assumed to be correlated by $0 < \rho < 1$, whereas the impostors' features are uncorrelated, i.e. $\rho = 0$.

By integration over the acceptance regions defined by (3), where $\hat{s} = s$, we obtain the acceptance probabilites $p_a(\mathrm{gen})$ and $p_a(\mathrm{imp})$ for a genuine prover and an impostor respectively. For benchmarking we include the performance of a Quantization Index Modulation (QIM) scheme with helper data [2] and a likelihood continuous classifier [8], which can be obtained in a similar way. The other schemes however have the possibility of adjusting some kind of threshold, i.e. the quantization width q or likelihood score s respectively, whereas the zero leakage scheme has none. The q in the zero leakage scheme is determined by the number of bits and cannot be adjusted, since the scheme will lose its zero leakage property in that case. Instead we allow a number of $C < M \cdot N$ errors to occur between enrollment and verification. This can be implemented as a C-error correcting code.

Since the chosen quantization width in the QIM scheme influences the amount of leakage, this amount is also depicted in Fig. 4 for a few operation points. The leakage of the continuous classifier equals 1 bit for every stored bit, since this scheme requires the template to be stored in the clear.

5. CHANNEL CAPACITY

The biometric verification scheme can also be considered as a 2^N wide memoryless channel, since it transfers the bits of the secret from the enrollment to the verification phase. The transition probability p_n, for an order n error, can be derived from (3), yielding for the capacity

$$C_{\mathrm{Z}} = \max_{p(s)} H(S) - H(S|\hat{S}) = N + \sum_{n=0}^{2^N-1} p_n \log p_n. \quad (14)$$

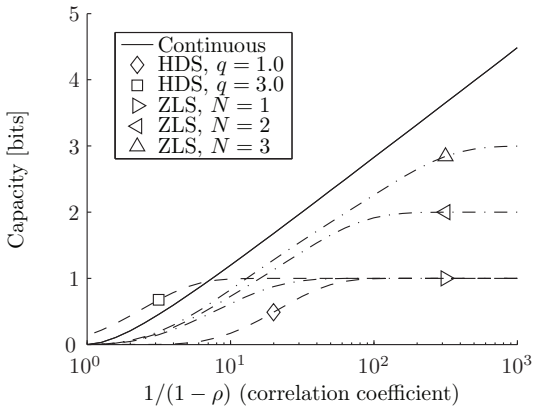

Fig. 5. Channel capacity of the verification schemes.

Again we compare the channel capacity against that of a continuous classifier and a QIM scheme with helper data (HDS) for benchmarking. As can be seen in Fig. 5, the continuous classifier provides an optimal capacity, but does not protect against leakage. The QIM scheme is the most flexible by its quantization width q at the cost of leakage and the zero leakage scheme requires a strongly correlated biometric to achieve capacity close to the number of bits assigned to each dimension.

6. DISCUSSION

We have shown that zero leakage can be obtained by applying a pre-distortion. However, a system designer needs to know the biometrics distribution to obtain the required pre-distortion function, which could be difficult to obtain in a practical situation.

In this paper we have developed a model to compute the leakage for an informed attacker as well as performance loss and illustrated it by assuming Gaussian distributions. However, the method is not limited to Gaussian distributions. In fact, it can be applied to any biometric with a continuous and bounded distribution.

A topic for further study is the relatively high false rejection rate. The quantization width cannot be adjusted to lower this value, since its boundaries are being dictated by the inverse pre-distortion function. One solution, to prevent these errors, is the application of an error correcting code. However, error correcting codes demand a certain distance between them, reducing the total number of available secrets significantly, which, on its turn, will weaken the strength of the entire scheme, despite the absence of leakage.

The scheme provides a means to assign more than just one bit per biometric feature. However, as it turns out, assigning multiple bits is only interesting for biometrics with a strong correlation between enrollment and verification. As can be seen in Fig. 5, the capacity only converges to the number of assigned bits ($N \in \{2, 3\}$) for a correlation coefficient very close to one.

7. CONCLUSION

We have proposed a new biometric verification scheme which achieves zero leakage and therefore protects the identity of the enrolled users essentially better than previously published schemes. We have quantified its performance penalty, in terms of worsening of the false rejection and acceptance ratio, and in terms of capacity of the biometric channel.

Moreover, we have introduced the notion of a pre-informed attacker. For the first time in literature, we have shown that when a priori knowledge about the provers biometrics is available, helper data schemes may leak more than in the case of un-informed attackers.

8. REFERENCES

[1] A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in *CCS '99: Proceedings of the 6th ACM conf on Comp and comm security*, 1999.

[2] J.-P. Linnartz and P. Tuyls, "New shielding functions to enhance privacy and prevent misuse of biometric templates," in *Audio- and Video-Based Biometric Person Authentication*. Springer, 2003.

[3] Y. Dodis, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," in *Lect Notes Comput Sc*. Springer, 2004.

[4] A. Juels and M. Sudan, "A fuzzy vault scheme," *Design Code Cryptogr*, vol. 38, pp. 237–257, 2006.

[5] N.K. Ratha, S. Chikkerur, J.H. Connell, and R.M. Bolle, "Generating cancelable fingerprint templates," *IEEE T Pattern Anal*, vol. 29, pp. 561 –572, 2007.

[6] A.M. Bazen and R.N.J. Veldhuis, "Likelihood-ratio-based biometric verification," *IEEE T Circ Syst Vid*, vol. 14, pp. 86 – 94, 2004.

[7] C. Chen, R.N.J. Veldhuis, T.A.M. Kevenaar, and A.H.M. Akkermans, "Multi-bits biometric string generation based on the likelihood ratio," in *Proc. IEEE Int. Conf. on Biometrics: Theory, Applications, and Systems*, 2007.

[8] E.J.C. Kelkboom, J. Breebaart, and R.N.J. Veldhuis, "Classification performance comparison of a continuous and binary classifier under gaussian assumption," in *Proc of the 31st Symp on Inf Theory in the Benelux*, 2010.

[9] F.M.J. Willems and T. Ignatenko, "Quantization effects in biometric systems," in *Information Theory and Applications Workshop*, 2009.

[10] J.-P.M.G. Linnartz, P. Tuyls, and B. Skoric, *A Communication-Theoretical View on Secret Extraction*, chapter 4, pp. 57–77, Security with Noisy Data. Springer, 2007.