

Pitfall of the Detection Rate Optimized Bit Allocation within Template Protection and a Remedy

E.J.C. Kelkboom, K.T.J. de Groot, C. Chen, J. Breebaart, R.N.J. Veldhuis

Abstract—One of the requirements of a biometric template protection system is that the protected template ideally should not leak any information about the biometric sample or its derivatives. In the literature, several proposed template protection techniques are based on binary vectors. Hence, they require the extraction of a binary representation from the real-valued biometric sample. In this work we focus on the Detection Rate Optimized Bit Allocation (DROBA) quantization scheme that extracts multiple bits per feature component while maximizing the overall detection rate. The allocation strategy has to be stored as auxiliary data for reuse in the verification phase and is considered as public. This implies that the auxiliary data should not leak any information about the extracted binary representation. Experiments in our work show that the original DROBA algorithm, as known in the literature, creates auxiliary data that leaks a significant amount of information. We show how an adversary is able to exploit this information and significantly increase its success rate on obtaining a false accept. Fortunately, the information leakage can be mitigated by restricting the allocation freedom of the DROBA algorithm. We propose a method based on population statistics and empirically illustrate its effectiveness. All the experiments are based on the MCYT fingerprint database using two different texture based feature extraction algorithms.

I. INTRODUCTION

The widespread use of biometric systems introduces new privacy risks, for example identity theft or cross-matching. These risks can be mitigated by applying template protection techniques. An overview of the privacy risks and template protection techniques are presented in [1]. A subclass of template protection techniques is based on a transformation of a biometric measurement to a binary vector as initial step. Hence, they require the extraction of a binary representation from the real-valued biometric sample. In the literature, numerous quantization schemes have been proposed. They vary from a simple method of extracting a single bit per feature component [2][3] to a more complex, multiple bits per feature component, extraction method [4][5][6][7]. If the quantization scheme is *subject-specific* the information has to be stored as *auxiliary data* for further use in the verification phase.

One of the requirements of a template protection system is that the stored auxiliary data ideally should not leak any information about the binary representation or the biometric sample itself. Hence, the subject-specific quantization

scheme stored as the auxiliary data should not reveal any information that may facilitate an adversary on increasing its success rate guessing the binary representation of the biometric sample in order to obtain a false accept.

The work of [8] showed that the quantization schemes proposed in [9] and [10] do indeed leak information that could be exploited by an adversary. Their attack model is to guess the secret key in an off-line mode by using the auxiliary data and population statistics. They use the guessing distance, consisting of the number of attempts required for a correct guess, as the measure of the degree of difficulty. Their results showed that the guessing distance is much smaller than what is expected based on the claimed security in [9] and [10], respectively. We focus on the Detection Rate Optimized Bit Allocation (DROBA) quantization scheme proposed in [7] that extracts multiple bits per feature component. For each enrolled subject the optimization algorithm allocates the optimal number of bits per component while maximizing the overall detection rate. The bit allocation strategy has to be stored as *auxiliary data* for further use during the verification phase.

Contribution: Our contribution is threefold. Firstly, we show that if the DROBA quantization scheme is not correctly implemented it will leak information about the binary representation of the biometric sample. Secondly, we illustrate an attack method an adversary could use in order to increase its success rate on reproducing a binary representation that leads to a false accept. Instead of using the guessing distance, we use the *false-acceptance* rate (FAR, α) as the degree of difficulty. We consider the template protection technique known as the helper-data system [2][3][11]. However, *any template protection technique incorporating the DROBA quantization scheme is susceptible to this vulnerability*. Thirdly, we outline a solution and propose an implementation guideline as a remedy. The remedy significantly mitigates the information leakage and guarantees a more private template.

The outline of this paper is as follows. In Section II we briefly discuss the considered template protection system with the DROBA quantization scheme. In Section III we describe our experimental setup concerning a fingerprint database, two feature extraction algorithms, and a testing protocol followed by the analysis of the information leakage due to the improper implementation of the DROBA quantization scheme. With use of the information leakage we demonstrate an attack method in Section IV that significantly increases

E.J.C. Kelkboom, K.T.J. de Groot, and J. Breebaart are with Philips Research, The Netherlands {Emile.Kelkboom, Koen.de.Groot, Jeroen.Breebaart}@philips.com

C. Chen and R.N.J. Veldhuis are with the University of Twente, Fac. EEMCS, The Netherlands {C.Chen, R.N.J.Veldhuis}@utwente.nl

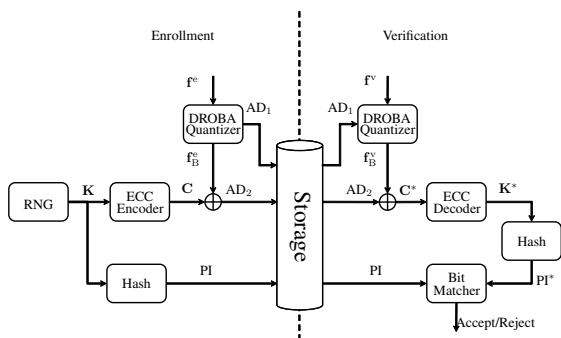


Fig. 1. Template protection scheme with DROBA implementation.

the false accept probability. As a remedy, we propose an implementation guideline in Section V and show that it significantly mitigates the information leakage. We finish with the conclusions in Section VI.

II. TEMPLATE PROTECTION SCHEME WITH DROBA

The template protection technique under consideration is known as the helper-data system [2][3] [11] and is portrayed in Fig. 1. As input we have the real-valued feature vector of dimension N_F , $\mathbf{f} \in \mathbb{R}^{N_F}$, which is extracted from the biometric sample by the feature extraction algorithm. Subsequently, a binary vector $\mathbf{f}_B \in \{0, 1\}^{N_B}$ is extracted by the DROBA quantization module and outputs the first auxiliary data AD_1 containing the allocation strategy. Many template protection schemes are based on the capability of generating a robust binary vector or key out of different biometric measurements of the same subject. However, the binary vector \mathbf{f}_B itself cannot be used as the key because it is most likely not exactly the same in both the enrollment and verification phase ($\mathbf{f}_B^e \neq \mathbf{f}_B^v$), due to measurement noise and biometric variability that lead to *bit errors*. The number of bit errors between two binary vectors is also referred to as the Hamming distance (HD) $d_H(\mathbf{f}_B^e, \mathbf{f}_B^v)$. Therefore, ECCs are used to deal with these bit errors. As shown in Fig. 1, the ECC and hash function are integrated using the well-known Fuzzy Commitment scheme [12]. For the sake of coherence we use the terminology proposed in [13].

Within the fuzzy commitment scheme we use the linear block type ECC “Bose, Ray-Chaudhuri, Hocquenghem” (BCH) that corrects random errors. The codeword \mathbf{C} corresponding to a randomly generated secret \mathbf{K} is XOR-ed with the \mathbf{f}_B^e in order to obtain the auxiliary data AD_2 . Furthermore, the hash of \mathbf{K} is taken in order to obtain the pseudo identity PI . In the verification phase this process is reversed with help of the auxiliary data resulting into a candidate pseudo identity PI^* . Only when $d_H(\mathbf{f}_B^e, \mathbf{f}_B^v) \leq t_c$ then PI and PI^* are equal, thus resulting into an accept. Hence, the Fuzzy Commitment scheme can be considered as a HD-classifier. More details about the template protection system can be found in [2][3].

As mentioned previously, the binary vector \mathbf{f}_B is extracted from the real-valued input vector \mathbf{f} by the DROBA quantization scheme and algorithm proposed in [7]. The DROBA

algorithm has the flexibility to extract multiple bits from a single component. The number of bits extracted from component i is given by b_i . The quantization schemes for the $b_i \in \{1, 2, 3\}$ cases are shown in Fig. 2(a), (b), and (c), respectively. For convenience we refer the $b_i = 1$ case as b_1^* , and b_2^* and b_3^* for the $b_i = 2$ and $b_i = 3$ cases, respectively. The 2^{b_i} quantization intervals are defined as such that the occurrence of each interval is equiprobable with respect to the *total* density, which we assume to be Gaussian distributed $p_t \sim \mathcal{N}(\mu_t, \sigma_t^2)$ with mean μ_t and variance σ_t^2 . The total density defines the observed variability of that component across the whole population. Each quantization interval is assigned a unique b_i bits Gray code [14]. Furthermore, we model the observed biometric variability and measurement errors of the feature vector component of a specific subject with the *within-class* density, which for simplicity is assumed to be another Gaussian density $p_w \sim \mathcal{N}(\mu_w, \sigma_w^2)$. Note that μ_w and σ_w^2 can be different for each component or subject. From [7] the detection rate γ is defined as the probability that the next measurement of the feature component will be in the same quantization interval. For component i the detection rate is computed as

$$\gamma_i(b_i) = \int_{Q_{\mu_w}(b_i)} p_w(v) dv, \quad (1)$$

where $Q_{\mu_w}(b_i)$ is the quantization interval corresponding to μ_w and also depends on the number of bits b_i to be extracted. Thus, the detection rate is the part of the within-class density within the quantization interval corresponding to μ_w , portrayed by the shaded area in Fig. 2. For the case where no bits are extracted ($b_i = 0$) the detection rate is defined as $\gamma_i(0) = 1$. Note that the detection rate decreases when b_i increases. Under the assumption that the N_F feature components are independent, the overall detection rate is defined as

$$\gamma_t = \prod_{i=1}^{N_F} \gamma_i(b_i). \quad (2)$$

The DROBA algorithm has to create a binary vector of length N_B , hence it has to allocate N_B bits across all components. We also refer to N_B as the bit-budget. With use of the multiple (N_e) enrollment samples, the DROBA algorithm analyzes the subject-dependent feature statistics (μ_w and σ_w^2) of each component and allocates the optimal number of bits b_i to component i with the constraints of maximizing the overall detection rate γ_t and allocating the bit-budget $\sum_{i=1}^{N_F} b_i = N_B$. The optimal allocation strategy is stored as auxiliary data $AD_1 = [b_1, b_2, \dots, b_{N_F}]$ for reuse at the verification phase. The optimization is implemented using the dynamic programming approach presented in [7].

III. EXPERIMENTS

If the DROBA implementation is correct, auxiliary data AD_1 should not leak any information about the enrolled binary vector \mathbf{f}_B^e . We will empirically analyze whether there is any information leakage by means of a fingerprint database and two feature extraction algorithms. We first discuss the

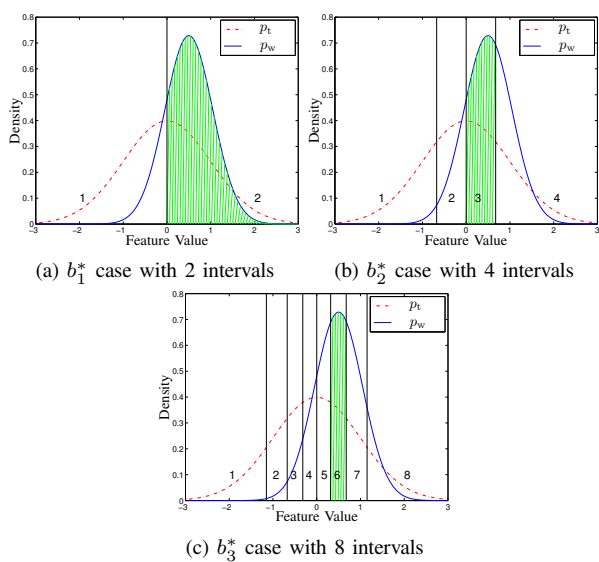


Fig. 2. The total density p_t with an example of a within-class density p_w and the corresponding detection rate γ_i at different quantization scheme where (a) $b_i = 1$ (b_1^*), (b) $b_i = 2$ (b_2^*), (c) $b_i = 3$ (b_3^*) bits are extracted.

experiment setup including the testing protocol followed by the information leakage analysis.

A. Experiment Setup

1) *Biometric Modality and Database*: The database we use is the MCYT (Ministerio de Ciencia y Tecnología) containing fingerprint images [15]. It contains 12 images of all 10 fingers from $N_s = 330$ subjects. However, we limit our dataset to the images of the right-index finger only.

2) *Feature Extraction Algorithms*: Two types of texture based features are extracted from a fingerprint, namely *directional field* and *Gabor* features. In order to compensate for possible translations between enrolled and verification measurements, a translation-only pre-alignment step is performed during the feature extraction process. Such pre-alignment requires extraction of the core point which is performed according to the algorithm described in [16]. Around the core point we define a 17×17 grid with eight pixels between each grid point. The following feature extraction algorithms extract a feature value on each grid point.

The first feature extraction algorithm is based on directional fields. A directional field vector describes the estimated local ridge-valley edge orientation in a fingerprint structure and is based on gradient vectors. The orientation of the ridge-valley edge is orthogonal to the gradient's angle. Therefore a directional field vector that signifies the orientation of the ridge-valley edge is perpendicular positioned to the gradient vector. In order to extract directional field features from a fingerprint the algorithm described in [17] is applied on each grid point. The direction field features have a dimension of $N_F = 578$ and are referred to as the DF features.

The second type of extracted features are the Gabor (GF) features, described in [18], where each grid point is

filtered using a set of four 2D Gabor filters at angles of $\{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$. The feature vector is the concatenation of the modulus of the four complex responses at each grid point, resulting into a feature vector dimension of $N_F = 1156$.

3) *Testing Protocol*: The performance testing protocol consists of randomly selecting 220 out of N_s subjects as the training set and the remaining 110 subjects as the evaluation set, which is referred to as the training-evaluation-set split. To decorrelate the feature components we use the principle component analysis (PCA) and the linear discriminant analysis (LDA) techniques, where the LDA transformation is also used to obtain more discriminating feature components from which we expect to extract more bits from. The PCA and LDA transformation matrices are computed using this training set, where N_{PCA} is the reduced dimension after applying the PCA transformation and N_{LDA} is the reduced dimension after applying the LDA transformation. To avoid singularities we ensure that $N_{LDA} \leq 220$. Furthermore, the template protection system parameters such as the quantization thresholds, used within the *Bit Extraction* module, are also estimated on the training set. From the evaluation set, 6 samples of each subject are randomly selected as the enrollment samples while the remaining samples are considered as the verification samples. This split is referred to as the enrollment-verification split. The protected template is generated using all the enrollment samples and compared with each individual verification sample. When the verification sample is from the same subject as of the protected template, it is referred to as a genuine comparison, otherwise it is an imposter comparison.

The training-evaluation-set split is performed five times, while for each of these splits the enrollment-verification split is performed 3 times. From each enrollment-verification split we estimate the β_{tar} (the *false-rejection rate* (FRR, β) at the targeted FAR of $\alpha_{tar} = 0.1\%$) and the *equal-error rate* (EER) where the FAR is equal to the FRR. Note, that the splits are performed randomly, however the seed at the start of the protocol is always the same, hence all the splits are equal for the performance tests at different settings. Hence, the splitting process does not contribute to any performance differences.

B. Analysis of the Information Leakage

First of all we empirically derive the $\{N_{PCA}, N_{LDA}, N_B\}$ setting leading to the optimal performance in terms of β_{tar} . We evaluate the performance for the settings of $N_{PCA} \in \{50, 100, \dots, 300\}$ and $N_B \in \{50, 100, \dots, \min(N_{PCA} \cdot b_{max}, 300)\}$, while the N_{LDA} parameter is set to $N_{LDA} = \min(N_{PCA}, 220)$ as discussed in Section III-A.3. The achieved β_{tar} performance for the different $\{N_{PCA}, N_{LDA}, N_B\}$ settings are depicted in Fig 3(a) and (b) for the DF and GF features, respectively.

For the DF features the optimal setting is achieved at $\{150, 150, 100\}$, while at $\{200, 200, 100\}$ for the GF features. At the optimal performance settings, the error-rate (α and β) curves with respect to the relative Hamming distance

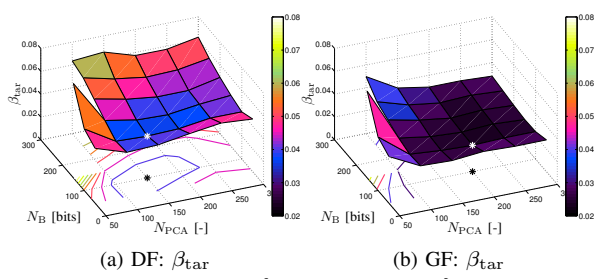
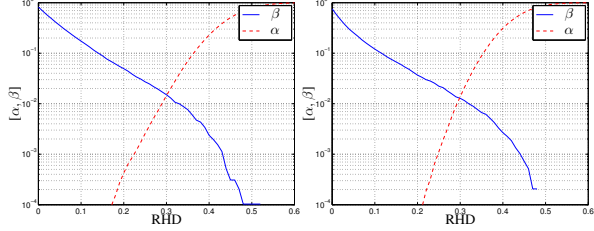
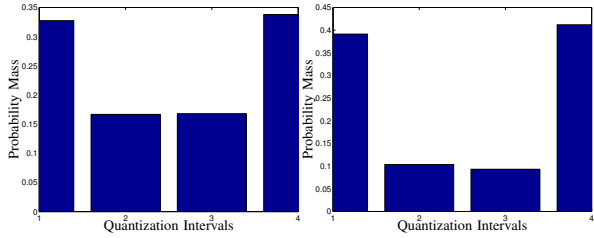


Fig. 3. The β_{tar} for different $\{N_{\text{PCA}}, N_{\text{LDA}}, N_{\text{B}}\}$ settings for the DF and GF features. The optimal performance for each case is indicated by both the black and white star.



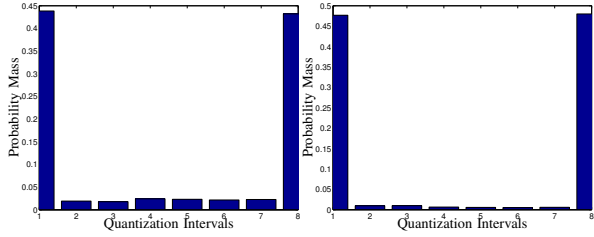
(a) DF: α and β curves

(b) GF: α and β curves



(c) DF: pmf of Q for b_2^*

(d) GF: pmf of Q for b_2^*



(e) DF: pmf of Q for b_3^*

(f) GF: pmf of Q for b_3^*

Fig. 4. The error-rate curves for and the pmf of Q for the b_2^* and b_3^* cases, for the DF and GF features.

(RHD) between $\mathbf{f}_{\text{B}}^{\text{e}}$ and $\mathbf{f}_{\text{B}}^{\text{v}}$ is portrayed in Fig. 4(a) and (b) for the DF and GF features, respectively. The β_{tar} is 3.66% for the DF features and 2.30% for the GF features, while the EER is 1.49% and 1.29%, respectively.

If the DROBA implementation is correct, AD_1 should not leak any information about the enrolled binary vector $\mathbf{f}_{\text{B}}^{\text{e}}$. We know that AD_1 is a concatenation of b_i of each feature component, hence knowing b_i should not leak any information about the actual b_i allocated bits. The allocated bits are equal to the Gray code assigned to the quantization interval in which the sample mean μ_{w} of the subject is measured. This implies that the probability of each quantization interval across the population should be equal irrespective of b_i . Hence, we analyze the probability of each quantization interval, referred to as the probability mass function (pmf) of Q , where we represent the quantization intervals by a discrete random variable Q . For the b_1^* case the pmf is

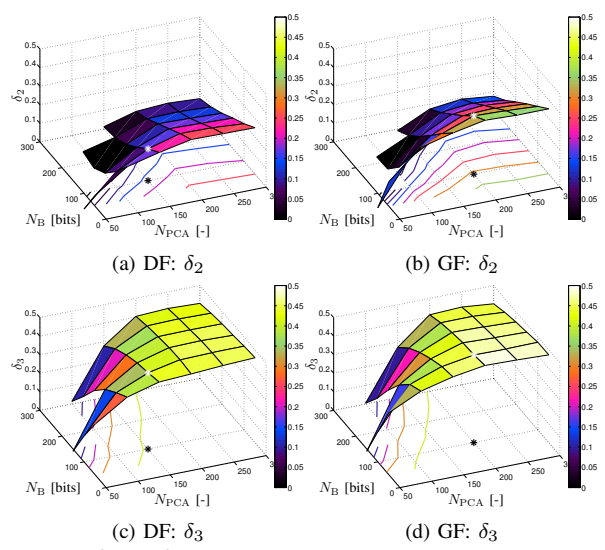


Fig. 5. The δ_2 and δ_3 for different settings of N_{PCA} and N_{B} for the DF and GF features. The optimal performance setting is indicated with both the black and white star.

uniform, however for the b_2^* and b_3^* cases a significantly non-uniform pmf is observed, see Fig. 4(c-f). For the b_2^* case roughly 66% of the cases μ_{w} is found to be in the outer quantization intervals for the DF features, while 80% for the GF features. For the b_3^* case it is around 87% for the DF feature and around 96% for the GF features. Due to the cyclic nature of Gray codes, the binary codes assigned to the outer quantization intervals differ in only a single bit. Hence, if multiple bits are extracted it is an advantage for the adversary to randomly select the binary code corresponding to one of the outer quantization intervals when guessing the binary vector $\mathbf{f}_{\text{B}}^{\text{e}}$.

In order to illustrate at which $\{N_{\text{PCA}}, N_{\text{LDA}}, N_{\text{B}}\}$ settings the most non-uniform pmf of Q is obtained, we define δ as the difference between the average probability of the two outer quantization intervals and the average probability of the remaining inner intervals. Hence, the closer δ is to zero the more the pmf is uniform and its maximum value is $\frac{1}{2}$. Furthermore, δ_2 is defined for the b_2^* case and δ_3 is for the b_3^* case. The δ values for the different settings are depicted in Fig. 5. From the figures we can observe that the non-uniformity is stronger when N_{B} decreases or N_{PCA} increases, which corresponds to the cases where the DROBA algorithm has more freedom to allocate the N_{B} bits. The maximum observed values are $\delta_2 = 0.256$ and $\delta_3 = 0.458$ of the DF features and $\delta_2 = 0.360$ and $\delta_3 = 0.485$ for the GF features. The pmf is close to uniform when $N_{\text{B}} \approx b_{\text{max}} N_{\text{PCA}}$, which is the case where the maximum number of bits is mostly extracted from each component. Note that at the optimal setting (indicated by the black and white star) the non-uniformity is close to its strongest.

Furthermore, we define $p(b_x^*)$ to be the average probability that a bit is derived from a b_x^* case. The $p(b_x^*)$ probabilities are different for each $\{N_{\text{PCA}}, N_{\text{LDA}}, N_{\text{B}}\}$ setting as shown in Fig 6 for the $p(b_2^*)$ and $p(b_3^*)$ cases for the DF and GF features. Because the sum of the probabilities is one, the probability $p(b_1^*)$ can be derived from $p(b_2^*)$ and $p(b_3^*)$. The

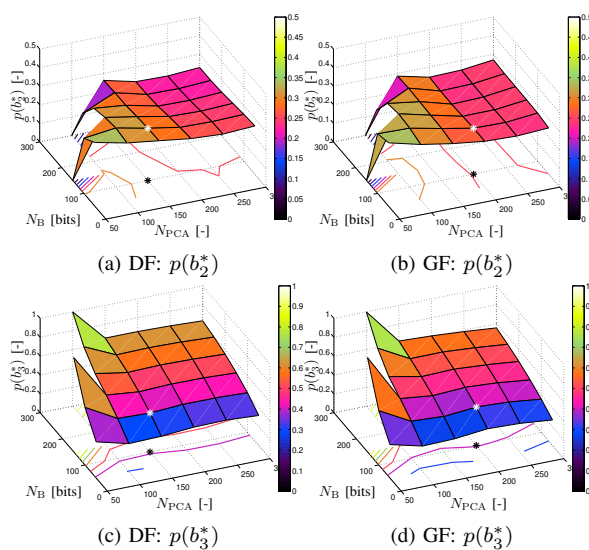


Fig. 6. The $p(b_2^*)$, $p(b_3^*)$ for different settings of N_{PCA} and N_B for the DF and GF features. The optimal performance setting is indicated with both the black and white star.

figures show that if N_B increases, more bits are extracted from the b_3^* case and less from the b_1^* case. The number of bits extracted from the b_2^* case stays relatively stable. For the optimal setting we have the probabilities $p(b_1^*) = 0.345$, $p(b_2^*) = 0.247$, and $p(b_3^*) = 0.408$ for the DF features, and $p(b_1^*) = 0.304$, $p(b_2^*) = 0.282$, and $p(b_3^*) = 0.414$ for the GF features, respectively. Note that the majority of the bits are extracted from a multiple-bits extraction case, from which we know that information is leaked as shown in Fig. 5. More precisely, the largest portion of bits are extracted from the b_3^* case, which leaks the most information.

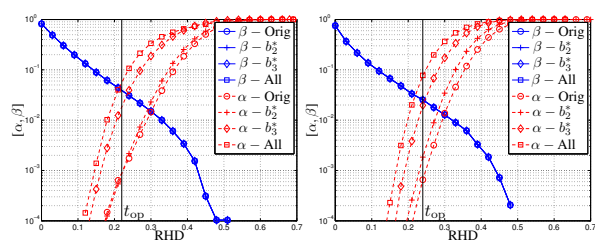
IV. EXPLOITATION OF THE LEAKAGE

In the previous section we have shown that the information leakage from the auxiliary data AD_1 about the enrolled binary vector \mathbf{f}_B^e is significant even at the optimal performance setting. However, it does not show what the actual practical advantage is for the adversary. In this section we propose a simple method the adversary could use in order to take advantage of the leaked information.

We consider the attack scenario where the adversary has the protected template, which is the collection of public auxiliary data AD_1 , AD_2 and PI , of an unknown subject and tries to obtain a false accept by the biometric system. As defined in [19] we focus on the attack level of “*overriding the feature extraction process*”. A possible attack method would be a dictionary attack, where a random image sample from a publicly available fingerprint database is selected, its feature vector \mathbf{f} is extracted and send to the next modules as

TABLE I
THE $p(b_1^*)$, $p(b_2^*)$, $p(b_3^*)$, δ_2 , δ_3 VALUES FOR THE DF AND GF FEATURES.

Features	EER [%]	β_{tar} [%]	$p(b_1^*)$	$p(b_2^*)$	$p(b_3^*)$	δ_2	δ_3
DF	1.49	3.66	0.345	0.247	0.408	0.1706	0.4106
GF	1.29	2.30	0.304	0.282	0.414	0.3136	0.4727



(a) DF: α and β curves (b) GF: α and β curves

Fig. 7. The error-rate curves pmfs for the (a) DF and (b) GF features when using the proposed attack at the imposter comparisons.

if it is authentic. The probability of an accept is equal to the FAR of the template protection system, because the imposter comparisons in fact do represent a dictionary attack. In our work, the targeted FAR is $\alpha_{\text{tar}} = 0.1\%$, thus on average $\frac{1}{\alpha_{\text{tar}}} = 1000$ attempts are expected in order to obtain a successful accept.

In our proposed attack method we also consider the *DROBA Quantizer* module to be compromised. Hence, the binary vector \mathbf{f}_B^e is generated and send to the next module. The leaked information can be exploited in the following way. We change the *DROBA Quantizer* module as such that if multiple bits are extracted (the b_2^* and b_3^* cases indicated by AD_1), we randomly select one of the two outer quantization intervals and return the corresponding Gray code. Hence, if AD_1 indicates that it is a b_2^* case, then either quantization intervals 1 or 4 are selected with 50% probability and when it is a b_3^* case the quantization intervals 1 or 8 are selected at random.

The attack results are given by the error-rate curves in Fig. 7(a) and (b) for the DF and GF features, respectively. Note that the attack is only carried out on the imposter comparisons and hence only the FAR curves are influenced. The original FAR is indicated with the “Orig” suffix, which is previously shown in Fig. 4 and represents the case where the attacker plainly selects a random sample from the database for the verification comparison without using any available knowledge and is the common FAR reported in the literature. For the attacks including the knowledge of the information leakage, we first study the method where only the information leakage from the b_2^* cases are exploited, hereafter we consider the method where only the b_3^* cases are exploited, and as the last method both the b_2^* and b_3^* cases are exploited. These attack methods are indicated with the suffix “ b_2^* ”, “ b_3^* ”, and “All”, respectively.

The operating point of a biometric system is determined using the α -Orig curve. The closest operating point t_{op} where the FAR reaches the targeted $\alpha_{\text{tar}} = 0.1\%$ without exceeding

TABLE II
THE OPERATING POINT t_{op} AT α_{tar} OF THE ORIGINAL CASE AND THE FAR OBTAINED AT THE DIFFERENT ATTACK SCENARIO.

Features	Orig case		FAR at t_{op} at attack scenario		
	t_{op} [RHD]	$\approx \alpha_{\text{tar}}$ [%]	b_2^* [%]	b_3^* [%]	All [%]
DF	0.22	$8.71 \cdot 10^{-2}$	$8.23 \cdot 10^{-2}$	1.89	5.78
GF	0.23	$6.56 \cdot 10^{-2}$	$1.84 \cdot 10^{-1}$	1.97	7.75

it, is portrayed with the solid vertical line. The operating point is at a RHD = 0.22 with $\alpha = 8.71 \cdot 10^{-2}\%$ for the DF features and RHD = 0.23 with $\alpha = 6.56 \cdot 10^{-2}\%$ for the GF features. The FAR obtained at the operating point for the different attack methods are given in Table II. The results show that $\alpha \cdot b_3^*$ is larger than $\alpha \cdot b_2^*$, which confirms the fact that the information leakage of the b_3^* cases is significantly larger than of the b_2^* cases. Furthermore, the advantage of the adversary is further increased by using the information leakage of both cases, because α -All is even larger. Hence, the largest achieved α is 5.78% for the DF features and 7.75% for the GF features. For the DF features the FAR has increased with a gain factor $G_\alpha = 66$, while for the GF features $G_\alpha = 118$. Thus, for both features the adversary gain is around two orders of magnitude. The necessary effort for the adversary to obtain an accept has significantly decreased from on average 1148 attempts to 17 attempts for the DF features and from 1524 to 13 for the GF features. Hence, the gain factor G_α can be seen as the gain of the adversary by exploiting the information leakage.

V. AN IMPLEMENTATION GUIDELINE AS REMEDY

In the previous section we have shown that if no precaution is taken, an adversary with knowledge of the DROBA implementation could significantly increase its false-acceptance rate with two orders of magnitude by exploiting the information leakage embedded in the auxiliary data AD_1 of the protected template. In this section we will address the cause of the information leakage and propose an implementation guideline for mitigating the leakage.

A. The Cause

Recall the fact that the DROBA algorithm is allowed to extract multiple bits from all feature components of \mathbf{f} , irrespective of its discriminating power or quality. Using the Gaussian model for describing the feature distribution of \mathbf{f} (see Section II), we can analyze the detection rate at different subject's mean μ_w for the b_1^* , b_2^* , and b_3^* cases and at different qualities of the feature components. As a measurement of the feature quality we use the Gaussian channel capacity or entropy H_G as defined in [20]

$$H_G = \frac{1}{2} \log_2 \left(1 + \frac{\sigma_b^2}{\sigma_w^2} \right), \quad (3)$$

which only depends on the ratio $\frac{\sigma_b^2}{\sigma_w^2}$ and where σ_b^2 is the variance of the between-class Gaussian density p_b describing the variability of the mean μ_w across the population and σ_w^2 is the variance of the within-class Gaussian density p_w .

Assuming the total density p_t to have a unit variance and using $\sigma_t^2 = \sigma_w^2 + \sigma_b^2$ we can rewrite H_G as

$$\begin{aligned} H_G &= \frac{1}{2} \log_2 \left(1 + \frac{\sigma_t^2 - \sigma_w^2}{\sigma_w^2} \right) \\ &= \frac{1}{2} \log_2 \left(\frac{\sigma_t^2}{\sigma_w^2} \right) \\ &= -\log_2(\sigma_w). \end{aligned} \quad (4)$$

Hence, feature components with $H_G = 1$ have a within-class standard deviation of $\sigma_w = \frac{1}{2^{H_G}} = \frac{1}{2}$, similarly for the cases $H_G = [2, 3, 4]$ we have $\sigma_w = [\frac{1}{4}, \frac{1}{8}, \frac{1}{16}]$, respectively.

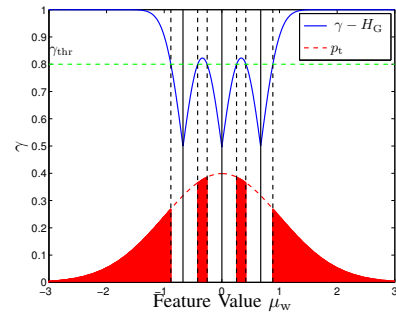


Fig. 9. The probability of selecting each quantization interval leading to a detection rate γ larger than a threshold γ_{thr} .

Using (1) the detection rate γ for different values of μ_w for different b_x^* cases and feature qualities $H_G \in \{1, 2, 3, 4\}$ are shown in Fig. 8. Note that the quantization intervals are fixed because of the unit variance assumption of p_t . The figures show that for the b_2^* and b_3^* cases the maximum detection rate γ for the inner quantization intervals are much lower than for the outer intervals, because the width of the inner quantization intervals are much smaller in order to be equiprobable with respect to the total density. The detection rate difference between the inner and outer quantization bins depend on the feature quality H_G and on the b_x^* case. A larger γ difference is observed for smaller H_G values and when more bits are extracted.

As discussed in Section II, the DROBA algorithm maximizes the overall detection rate γ_t as given by (2). Due to the optimization criteria, the DROBA algorithm tends to allocate multiple bits mostly for the cases where the subject's mean μ_w is in the outer quantization intervals due to the larger γ values. This behavior is stronger for the lower quality feature components because γ is significantly larger for the outer quantization intervals as shown in Fig. 8.

We illustrate the non-uniformity effect introduced by the DROBA algorithm with the following simplified case. Consider the case where there are three feature components of equal quality of $H_G = 2$ from which four bits ($N_B = 4$) have to be extracted and only two bits are allowed to be extracted from each component (b_2^* case). Assume, the first component analyzed has a detection rate of $\gamma_1 = 0.8$. The probability that the next component has a detection rate γ_2 larger than threshold $\gamma_{thr} = \gamma_1$ is portrayed by the shaded area of the p_t density shown in Fig. 9 which is $\Pr(\gamma_2 > \gamma_{thr}) \approx 0.5$. Note that the probability of each quantization interval is not equiprobable. For the outer quantization intervals we obtain $p(q_1) = p(q_4) = 0.38$, while for the inner quantization intervals $p(q_2) = p(q_3) = 0.12$. Hence the difference is $\delta_2 = 0.26$. If it turns out that $\gamma_2 > \gamma_1$, then when analyzing the third component the threshold becomes $\gamma_{thr} = \gamma_2$. Because of the larger γ_{thr} for the third component, the probability of obtaining a higher γ_2 in one of the quantization intervals becomes more uniform and δ_2 is thus larger. Note that this effect is stronger for lower quality feature components with a smaller H_G or when more bits are extracted.

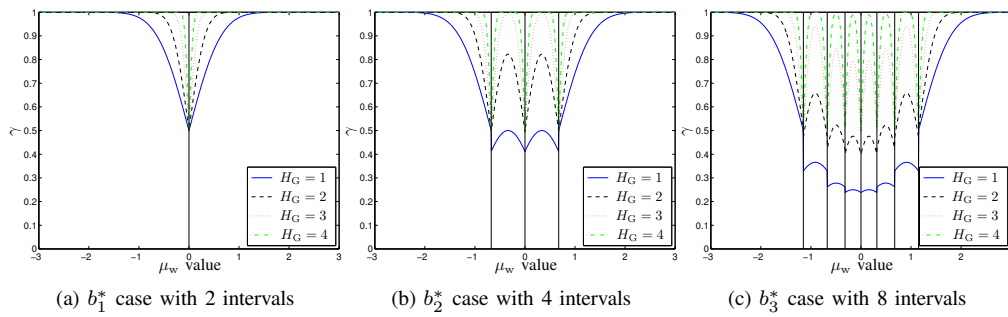


Fig. 8. The detection rate γ for different values of μ_w for the (a) b_1^* , (b) b_2^* , (c) b_3^* case with different feature qualities $H_G \in \{1, 2, 3, 4\}$.

B. The Remedy: Restricting DROBA

As remedy we propose to restrict the DROBA algorithm. The maximum number of bits b_{\max} that the DROBA algorithm is allowed to extract from a component should depend on the overall feature quality of the corresponding component. For each component, we compute the overall feature quality using (3) where we take the average of the subject dependent within-class variance across the population. We introduce the thresholds $\delta_{H_G,2}$ and $\delta_{H_G,3}$, where $\delta_{H_G,2}$ defines the minimum overall feature quality requirement of the component for extracting two bits and similarly $\delta_{H_G,3}$ for the case of extracting three bits. We empirically estimate the optimal threshold settings that minimize the information leakage, i.e. induce δ_2 and δ_3 to be close to zero. The δ_2 and δ_3 values for different $\delta_{H_G,2}$ and $\delta_{H_G,3}$ settings are shown in Fig. 10 for both features. For the δ_2 case we obtain $\delta_2 \approx 0$ by setting $\delta_{H_G,2} = 2.35$ for the DF features and $\delta_{H_G,2} = 2.95$ for the GF features. However, for the δ_3 case it does not reach zero. By increasing $\delta_{H_G,3}$ even further has the consequence that there are only a few b_3^* cases, even less than one case per subject for the GF features as shown by Fig. 10(f). Eventually we select $\delta_{H_G,3}$ with the biggest drop in δ_3 , which is at $\delta_{H_G,3} = 4.05$ for the DF features and $\delta_{H_G,3} = 4.15$ for the GF features.

We implement the proposed remedy to the DROBA algorithm and evaluate the performance and information leakage on the optimal performance setting obtained in Section III-B of $\{150, 150, 100\}$ and $\{200, 200, 100\}$ for the DF and GF features, respectively. The pmf of Q for the b_2^* and b_3^* cases, and the error-rate curves are shown in Fig. 11. The pmf of Q for the b_2^* case for both the DF and GF features are very close to uniform, while for the b_3^* case they tend to become more uniform. Because the threshold $\delta_{H_G,3}$ was limited, otherwise no bits would have been extracted from a b_3^* case, the pmf of Q is not uniform.

Comparing the error-rate curves, we observe that the β -Remedy curve has shifted to the right compared to the original curve, β -Orig. However, the α -Remedy curve has also shifted to the right with the consequent that the EER and β_{tar} values are very similar to the original case, namely 1.76% and 3.87% for the DF features, and 1.27% and 2.17% for the GF features. The FRR curve shift can be caused by the fact that the DROBA algorithm is restricted by the proposed remedy. The allocation strategy may then be sub-

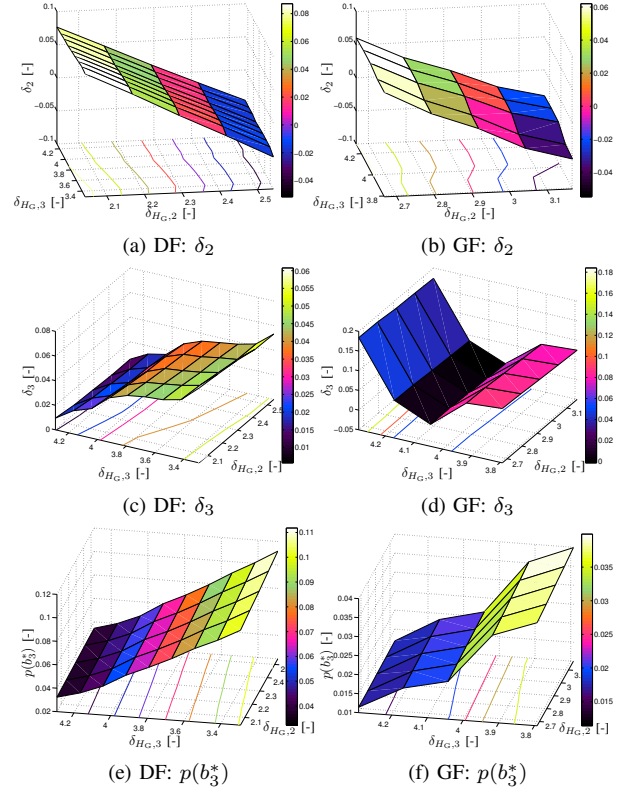


Fig. 10. The δ_2 , δ_3 , and $p(b_3^*)$ for different settings of $\delta_{H_G,2}$ and $\delta_{H_G,3}$ for the DF and GF features.

optimal for the performance. The shift of the FAR curve can be explained in the following way. Note that the variance of p_t is larger during the verification phase, because there are less verification samples than enrollment samples, while the quantization intervals are defined equiprobable on the p_t during the enrollment phase. Hence, when randomly selecting fingerprint images at the verification comparisons the outer quantization intervals are always more probable. When using the original DROBA algorithm, the outer quantization intervals during the enrollment phase are also more probable (the information leakage we have shown). Consequently, there are less bit errors at the imposter comparisons leading to a larger FAR at the same operating point. In other words, it is easier to find a random fingerprint image that leads to an accept. When applying the DROBA remedy, the quantization intervals during the enrollment phase become more equiprobable, consequently eliminating the previously

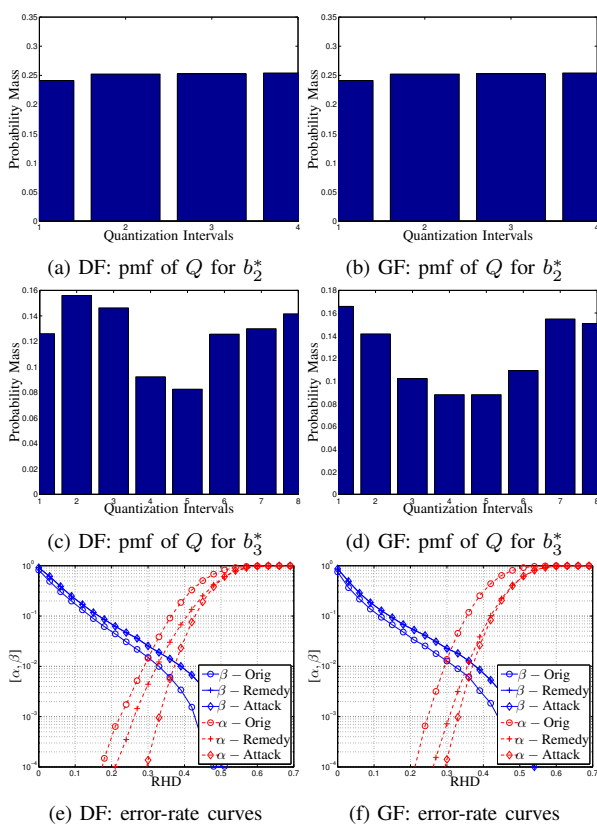


Fig. 11. The pmf of Q for the b_2^* and b_3^* cases, and the error-rate curves for the DF and GF features.

mentioned effect, therefore decreasing the FAR at the same operating point. Furthermore, the α -Attack obtained when using the proposed attack method did not increase with respect to α -Remedy, it has actually decreased. Hence, the adversary does not gain any advantage by using the proposed attack when the DROBA is correctly implemented. The decrease of the α -Attack can be explained by the fact that the attack method does not consider the correlations between the feature components when randomly guessing one of the outer quantization intervals for the b_2^* and b_3^* cases.

VI. CONCLUSION

In this work we have shown that great care has to be taken when designing an DROBA quantization scheme in order to guarantee that its auxiliary data does not leak any information about the binary representation of the biometric sample. If no care is taken, the information leakage can be significant and an adversary is able to exploit this information. We have shown that the adversary is able to increase its success rate of obtaining a false accept by two orders of magnitude.

Fortunately, there is a solution to mitigate the information leakage. We proposed a remedy which in fact is a guideline on how to restrict the allocation freedom of the DROBA algorithm. The maximum allowed bits to be allocated to each component has to depend on the overall feature quality across the population of that component. We empirically estimated the minimum overall feature quality boundaries for allocating two or three bits, respectively. Given the biometric database and the feature extraction algorithms, the proposed

remedy significantly reduced the information leakage without influencing the performance in terms of the EER or the FRR at the targeted FAR of the biometric system.

REFERENCES

- [1] A. K. Jain, K. Nandakumar, and A. Nagar, "Biometric template security," *EURASIP Journal on Advances in Signal Processing*, 2008.
- [2] E. J. C. Kelkboom, B. Gökberk, T. A. M. Kevenaar, A. H. M. Akkermans, and M. van der Veen, "'3D face': Biometric template protection for 3d face recognition," in *Int. Conf. on Biometrics*, Seoul, Korea, August 2007, pp. 566–573.
- [3] T. A. M. Kevenaar, G.-J. Schrijen, A. H. M. Akkermans, M. van der Veen, and F. Zuo, "Face recognition with renewable and privacy preserving binary templates," in *4th IEEE workshop on AutoID*, Buffalo, New York, USA, October 2005, pp. 21–26.
- [4] Y.-J. Chang, W. Zhang, and T. Chen, "Biometrics-based cryptographic key generation," in *IEEE Int. Conf. on Multim. and Expo*, vol. 3, June 2004, pp. 2203 – 2206.
- [5] C. Chen, R. Veldhuis, T. Kevenaar, and A. Akkermans, "Multi-bits biometric string generation based on the likelihood ratio," in *IEEE Conf. on Biometrics: Theory, Applications and Systems*, Washington DC, September 2007.
- [6] W. Zhang, Y.-J. Chang, and T. Chen, "Optimal thresholding for key generation based on biometrics," in *International Conference on Image Processing*, *ICIP '04*, vol. 5, 2004, pp. 3451–3454.
- [7] C. Chen, R. N. J. Veldhuis, T. A. M. Kevenaar, and A. H. M. Akkermans, "Biometric quantization through detection rate optimized bit allocation," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, 2009.
- [8] L. Ballard, S. Kamara, and M. K. Reiter, "The practical subtleties of biometric key generation," in *USENIX Security*, 2008.
- [9] C. Vielhauer, R. Steinmetz, and A. Mayerhofer, "Biometric hash based on statistical features of online signatures," in *Proceedings of the Sixteenth International Conference on Pattern Recognition*, vol. 1, 2002, pp. 123–126.
- [10] F. Hao and C. Chan, "Private key generation from on-line handwritten signatures," in *Information Management & Computer Security*, vol. 10, no. 4, 2002, pp. 159–164.
- [11] P. Tuyls, A. H. M. Akkermans, T. A. M. Kevenaar, G.-J. Schrijen, A. M. Bazen, and R. N. J. Veldhuis, "Practical biometric authentication with template protection," in *5th International Conference, AVBPA*, Rye Brook, New York, July 2005.
- [12] A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in *6th ACM Conference on Computer and Communications Security*, November 1999, pp. 28–36.
- [13] J. Breebaart, C. Busch, J. Grave, and E. Kindt, "A reference architecture for biometric template protection based on pseudo identities," in *BIOSIG*, Darmstadt, Germany, September 2008.
- [14] M. Gardner, *Knotted Doughnuts and Other Mathematical Entertainments*. W.H. Freeman & Company, 1986.
- [15] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, M. F. J. Gonzalez, V. Espinosa, A. Satue, I. Hernaez, J. J. Igarza, C. Vivaracho, D. Escudero, and Q. I. Moro, "MCYT baseline corpus: A bimodal biometric database," in *IEE Proc. Vision, Image and Signal Processing, Special Issue on Biometrics on the Internet*, December 2003, pp. 395–401.
- [16] M. van der Veen, A. Bazen, T. Ignatenko, and T. Kalker, "Reference point detection for improved fingerprint matching," in *Proceedings of SPIE*, 2006, p. 60720G.160720G.9.
- [17] S. Gerez and A. Bazen, "Systematic methods for the computation of the directional fields and singular points of fingerprints," in *IEEE Transactions on pattern analysis and machine intelligence*, July 2002, pp. 905–919.
- [18] A. M. Bazen and R. N. J. Veldhuis, "Likelihood-ratio-based biometric verification," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 14, no. 1, pp. 86–94, 2004.
- [19] N. K. Ratha, J. H. Connell, and R. M. Bolle, "An analysis of minutiae matching strength," in *Proceedings of the 3rd International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA 01)*, Halmstad, Sweden, June 2001, pp. 223–228.
- [20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.