Maximum Key Size and Classification Performance of Fuzzy Commitment for Gaussian Modeled Biometric Sources

Emile J. C. Kelkboom, Jeroen Breebaart, *Member, IEEE*, Ileana Buhan, and Raymond N. J. Veldhuis

***Abstract*—Template protection techniques are used within biometric systems in order to protect the stored biometric template against privacy and security threats. A great portion of template protection techniques are based on extracting a key from, or binding a key to the binary vector derived from the biometric sample. The size of the key plays an important role, as the achieved privacy and security mainly depend on the entropy of the key. In the literature, it can be observed that there is a large variation on the reported key lengths at similar classification performance of the same template protection system, even when based on the same biometric modality and database. In this work, we determine the analytical relationship between the classification performance of the fuzzy commitment scheme and the theoretical maximum key size given as input a Gaussian biometric source. We show the effect of the system parameters such as the biometric source capacity, the number of feature components, the number of enrolment and verification samples, and the target performance on the maximum key size. Furthermore, we provide an analysis of the effect of feature interdependencies on the estimated maximum key size and classification performance. Both the theoretical analysis, as well as an experimental evaluation using the MCYT fingerprint database showed that feature interdependencies have a large impact on performance and key size estimates. This property can explain the large deviation in reported key sizes in literature.**

***Index Terms*—Analytical models, biometrics, template protection.**

I. Introduction

In recent years, interest in biometric systems has significantly increased. Examples include 1) the planned introduction of the United Kingdom National Identity Card based on biometrics required by the Identity Cards Act 2006 [1] and 2) the recommendation by the International Civil Aviation Organization (ICAO) [2] to adopt the ePassport that also includes biometric data.

A biometric system used for authentication primarily consists of an enrolment and verification phase. In the enrolment phase, a biometric sample is captured and a reference template is created and stored. In the verification phase, a new biometric sample is captured and compared to the stored reference template. The subject is considered as being genuine if the new biometric sample is sufficiently similar to the stored reference template. A biometric system requires the storage of a reference template of the biometric data. Hence, the widespread use of biometrics introduces new security and privacy risks such as 1) *identity fraud* where an adversary steals the stored reference template and impersonates the genuine subject of the system by some spoofing mechanism, 2) *limited-renewability* implying the limited capability to renew a compromised reference template due to the limited number of biometric instances (for example we only have ten fingers, two irises or retinas, and a single face), 3) *cross-matching* linking reference templates of the same subject across databases of different applications, and 4) *(sensitive) personal or medical information leakage, implying* that biometric data may reveal the gender, ethnicity, or the presence of certain diseases.

The field of template protection is focused on mitigating these privacy risks by developing template protection techniques that provide 1) *irreversibility* implying that it is impossible or at least very difficult to retrieve the original biometric sample from the reference template, 2) *renewability* or the ability to renew the reference template when necessary, and 3) *unlinkability* which prevents cross-matching.

A. Overview of the Template Protection Field

As described in Jain *et al.* (2008) [3], the template protection techniques proposed in the literature can be divided into two categories, namely 1) *feature transformations* and 2) *biometric cryptosystems*.

The most common technique based on feature transformations is known as *cancelable biometrics* [4], [5]. With cancelable biometrics, the reference template is generated by applying a noninvertible transformation on the enrolment sample. Due to the noninvertible property of the transformation, it is impossible to obtain the original biometric sample from the reference template. In the verification phase, the same noninvertible transformation is applied on the verification sample, and the matching is thus performed on the transformed version of both the enrolment and verification sample.

Biometric cryptosystem techniques can be subdivided into 1) *key binding* and 2) *key generation* methods. In the enrolment phase, the key binding techniques combine the key with a biometric sample into auxiliary data such that the same key

Manuscript received September 29, 2010; revised December 23, 2011; accepted January 04, 2012. Date of publication April 03, 2012; date of current version July 09, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Arun Ross.

E. J. C. Kelkboom is with Philips Research, Eindhoven 5656AE, The Netherlands (e-mail: Emile.Kelkboom@philips.com).

J. Breebaart is with Civolution, Eindhoven 5656AE, The Netherlands (e-mail: Jeroen.Breebaart@civolution.com).

I. Buhan is with Riscure, Delft 2628XJ, The Netherlands (e-mail: Ileana.Buhan@gmail.com).

R. N. J. Veldhuis is with the University of Twente, Fac. EEMCS, Enschede 7500AE, The Netherlands (e-mail: R.N.J.Veldhuis@utwente.nl).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2012.2191961

can be successfully released in the verification phase by using a new biometric sample and the stored auxiliary data. Examples of the key binding techniques are the *Fuzzy Commitment Scheme* (FCS) [6], the *Helper Data System* (HDS) [7], and the *Fuzzy Vault* [8]. Most key binding schemes first extract a binary vector from the biometric sample before the binding process. Key generation techniques extract a robust key from the biometric sample in the enrolment phase, with auxiliary data if necessary. In the verification phase, the same key has to be extracted using a new biometric sample and, when available, the auxiliary data. *fuzzy extractors* are the most common key generation techniques, which can be created using *secure sketches* [9].

In line with the standardization activities in ISO [10], the hash of the key is referred to as the *pseudonymous identifier* (PI) and the protected reference template is the collection of the auxiliary data (AD) and (PI).

B. Privacy, Security, and Convenience

We previously mentioned the security risks of identity fraud and limited-renewability and the privacy risks of cross-matching and leaking (sensitive) medical information. We also mentioned that most key binding schemes first extract a binary vector from the biometric sample before the binding process. Hence, retrieving this binary vector from the stored protected template may facilitate a possible replay attack (makes identity fraud possible) or a cross-matching attack and, therefore, allows for a security or privacy breach. Besides the cross-matching privacy breach, the binary vector could also reveal sensitive or medical information of the subject.

It is known from the key binding technique that given the protected template, an adversary could retrieve the binary vector extracted from the biometric sample by guessing the key and inverting the key binding process. Therefore, the achieved privacy and security protection depends on the entropy of the key, i.e., the difficulty of guessing it. Considering the key to consist of independent and uniform bits, its entropy is then determined by its size. Having a key of k_c bits on average will take 2^{k_c-1} guesses in order to obtain the correct one, hence adding a single bit to the key doubles the adversary's effort.

On the other hand, the classification performance of the template protection system also determines the effort of inverting the key-binding process. In the remainder of this work, we refer to the classification performance of the template protection system as the *system performance*. The system performance can be expressed by the *false match rate* (FMR) and the *false nonmatch rate* (FNMR). Given an enrolment sample, the FMR is the probability of incorrectly classifying a verification sample from a different subject as similar and genuine, hence leading to a false match. Thus, the FMR also indicates the likelihood of finding a random verification sample, e.g., from existing databases, that will lead to a match and, therefore, a security breach, which is also known as the FMR attack. The work of Korte and Plaga (2007) [11] and Buhan *et al.* [12] describe a relationship between the FMR and the key size, namely $k_c \leq -\log_2(FMR)$. The FNMR, on the other hand, is the probability of incorrectly classifying a verification sample from the same subject as different, thus leading to a false nonmatch. We consider the FNMR as part of the convenience factor of the biometric system, because it determines the probability that subjects have to repeat the verification process which is considered as an unpleasant experience. It is also known that increasing the FNMR usually results in a decrease of the FMR, and consequently an increase in the key size. In other words, the security and convenience of a biometric key binding system are often subject to a trade-off. Another factor that influences the security and performance of a biometric system is the number of biometric samples that is used for enrolment and verification. Acquiring multiple biometric samples will improve the system performance as shown in Kittler *et al.* (1997) [13], Faltemier *et al.* (2008) [14], and Kelkboom *et al.* [15], and therefore, also the key size. However, increasing the number of biometric samples increases the acquisition time which could be experienced as inconvenient by the subject, and is another parameter to influence the convenience-security trade-off.

C. Reported Performances With Corresponding Key Size

In the literature, there is a significant variability in the reported key sizes when compared to the key size deducted from the FMR. Table I shows an overview of the reported system performance and key size for different template protection techniques, databases, and feature extraction methods. It is difficult to find a relationship between the system performance and the key size. For example, consider cases 4 and 7 that use the same template protection technique and modality, and a similar database. While having similar reported performance, the key size in case 7 is almost three times larger than in case 4. Likewise, when comparing cases 1c and 6a with a similar template protection technique, modality, database, and performance, the reported key size in case 6a is almost double of the one of case 1c. As a final example, the separate cases 5 and 6 show that using exactly the same template protection technique on the same modality but different databases may lead to a different performance at an equal key size as in case 5 or different key sizes at similar performance as in case 6.

Comparing the performance and the key size of template protection schemes based on different error-correcting code (ECC) implementations, databases, biometric modalities, or feature extraction algorithms is not straightforward. Different ECC implementations may lead to different error-correcting capabilities and, therefore, a possible difference between the system performance and the key size. Different databases, biometric modalities, or feature extraction algorithms influence the quality of the extracted features and, therefore, the system performance. In the comparisons made above, we tried to minimize these differences. From the significant differences observed between the reported system performance (especially the FMR) and the key size, we may conclude that there seems to be no clear relationship between the system performance and the key size.

D. Related Work and Contributions

We are interested in determining the relationship between the maximum key size and the system performance. Furthermore, we also investigate the influence of the system parameters on the key size and the system performance. The system parameters are the discriminating power of the biometric Gaussian source,

TABLE I
OVERVIEW OF THE KEY SIZE AND THE CLASSIFICATION PERFORMANCE OF DIFFERENT BIOMETRIC CRYPTOSYSTEMS TECHNIQUES, MODALITIES AND DATABASES FOUND IN THE LITERATURE. THE BIOMETRIC CRYPTOSYSTEMS UNDER CONSIDERATION ARE THE FUZZY EXTRACTORS, THE FUZZY COMMITMENT SCHEMES (FCS), THE HELPER DATA SYSTEMS (HDS), THE FUZZY VAULT, AND THE CODE-OFFSET CONSTRUCTION

Work	Case	Method	Modality	Database (# samples / instance)	FMR	FNMR	Key size [bits]
Bringer et al. [16]	1a	FCS	iris	ICE 2005 (2953/244)	$< 10^{-5}$	0.0562	42
	1b	FCS	iris	CASIA (756/108)	≈ 0	0.0665	42
	1c	FCS	fingerprint	FVC2000 DB2a (800/100)	0.0553	0.0273	42
Change et al. [17]	2	FCS	fingerprint	NIST 4 (4000/2000)	≈ 0.001	≈ 0.10	10
Sutcu et al. [18]	3	FCS	fingerprint	Mitsubishi (1035/69)	$1.19 \cdot 10^{-4}$	0.11	30
Kelkboom et al. [19]	4	HDS	3D face	FRGC v2 subset (2347/145)	0.0019	0.16	35
Kevenaar et al. [20]	5a	HDS	2D face	FERET ($> 948/237$)	≈ 0	0.35	58
	5b	HDS	2D face	Caltech ($> 209/19$)	≈ 0	0.035	58
Tuyls et al. [21]	6a	HDS	fingerprint	FVC2000 DB2a&b (880/110)	0.052	0.054	76
	6b	HDS	fingerprint	Univ. Twente (2500/500)	0.035	0.054	40
Zhou et al. [22]	7	HDS	3D face	FRGC v1 subset ($> 396/99$)	0.004	0.12	107
Hao et al. [23]	8	code-offset	iris	private (700/70)	≈ 0	0.0047	140
					0.02%	0.15%	112
Clancy et al. [24]	9	fuzzy vault	fingerprint	-	-	0.20-0.30	69
Nandakumar et al. [25]	10a	fuzzy vault	fingerprint	FVC2000 DB2a (800/100)	$\approx 10^{-4}$	0.09	≈ 40
	10b	fuzzy vault	fingerprint	MSU-DBI (640/160)	$\approx 2 \cdot 10^{-4}$	0.175	≈ 40
Arakala et al. [26]	11	fuzzy extractors	fingerprint	FVC2000 DB1a (800/100)	0.15	0.15	34

the number of feature components extracted from the biometric sample, and the number of enrolment and verification samples.

An analysis about the maximum key size given a discrete biometric source is done in Ignatenko and Willems (2009) [27] (which is an extended version of Ignatenko and Willems (2008) [28]) and a similar work of Lai *et al.* (2008) [29], where they estimated the secret-key rate. The work of Willems and Ignatenko (2009) [30] analyzed the secret-key rate for a Gaussian distributed continuous biometric source. The framework of these works assumes that if the number of feature components goes to infinity, the discriminating power of each component remains constant. Assuming independent feature components, this would imply that the biometric source has an infinite discriminating power. This would not hold for a biometric system, where the discriminating power of a biometric trait is limited due to its practical nature, namely measurement noise or biometric variability.

In our work, we use a Gaussian model for a continuous biometric source with a limited discriminating power (or input capacity) that can be distributed over a limited number of feature components. We present five contributions. First, we analytically determine the classification performance of the FCS where the input is a Gaussian modeled biometric source. We also include the number of enrolment and verification samples. Second, from the estimated performance we analytically determine the theoretical maximum key size at the operating point determined by the target FNMR, assuming an ECC with decoding capabilities at Shannon's bound. We also verify the known relationship between the maximum key size and the FMR. Third, we investigate by means of numerical analysis the effect of the parameters such as the capacity of the Gaussian biometric source, the number of enrolment and verification samples, and the target FNMR on the maximum key size. Fourth, we provide an analysis of the effect of feature interdependencies and differences in their quality. Finally, we analyze these findings on the MCYT fingerprint database using two feature extraction algorithms.

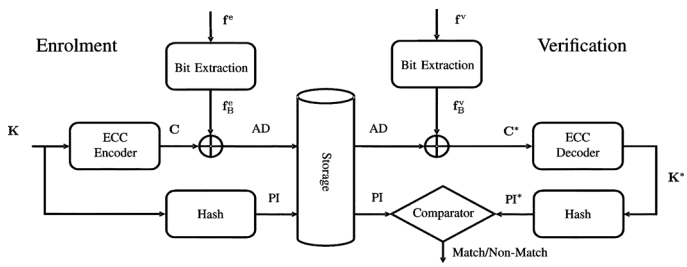

Fig. 1. FCS construction combined with a Bit Extraction module.

E. Outline

The outline of this paper is as follows. We briefly describe the FCS construction in Section II. In Section III, we present the analytical framework that models the biometric source as parallel Gaussian channels. Furthermore, we derive the analytical system performance and the theoretical maximum key size at the target FNMR. Section IV illustrates by means of numerical analysis the effect of the system parameters and feature interdependencies on the maximum key size. The experimental setup using the MCYT database and the obtained results are discussed in Section V. Our final remarks and conclusions are given in Section VI.

II. FUZZY COMMITMENT SCHEME

The FCS construction combined with a *Bit Extraction* module is depicted in Fig. 1.

In the enrolment phase or the key-binding process, the real-valued column *feature vector* $\mathbf{f}^e \in \mathbb{R}^{N_F}$ is extracted from each of the N_e biometric enrolment samples by the feature extraction algorithm. A single binary column vector $\mathbf{f}_B^e \in \{0,1\}^{N_F}$ is created from the mean of the N_e feature vectors within the Bit Extraction module, which we will discuss in Section III. Furthermore, a random key $\mathbf{K} \in \{0,1\}^{k_c}$ is created and encoded by the *ECC Encoder* module into a codeword $\mathbf{C} \in \mathcal{C}$ of size $\{0,1\}^{n_c}$, where $\mathcal{C}$ is the ECC codebook (the set of codewords).

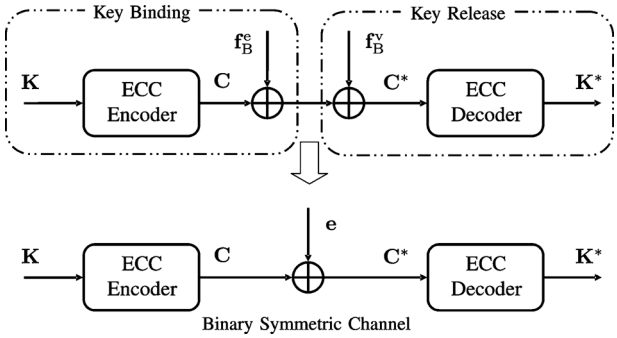

Fig. 2. Modeling the key binding and release process by a binary symmetric channel (BSC).

The codeword is XOR'd with the binary vector $\mathbf{f}_B^e$, creating the auxiliary data (AD). AD is stored as part of the protected template together with the hash of $\mathbf{K}$. Because of the XOR operation and the fact that a single bit is extracted from each feature component, it implies that the size of the extracted real-valued and binary vector are equal to the codeword size, namely $n_c = N_F$, and in the remainder of this work we will only use n_c.

In the verification phase or the key-release process, the binary vector $\mathbf{f}_B^v$ is created by quantizing the mean of the N_v verification feature vectors $\mathbf{f}^v$. Hereafter, the auxiliary data AD is XOR'd with $\mathbf{f}_B^v$ resulting in the possibly corrupted codeword $\mathbf{C}^*$. Decoding $\mathbf{C}^*$ by the *ECC Decoder* module leads to the candidate secret $\mathbf{K}^*$. The candidate pseudonymous identifier PI^* is obtained by hashing $\mathbf{K}^*$. A match is returned by the *Comparator* module if PI and PI^* are equal, which occurs only when $\mathbf{K}$ and $\mathbf{K}^*$ are equal, i.e., the key-release process was successful.

Under the assumption that the bit errors are mutually independent, the channel between the encoder and decoder of the key-binding and key-release process can be modeled by a binary symmetric channel (BSC) as portrayed in Fig. 2, with an error pattern $\mathbf{e} = \mathbf{f}_B^e \oplus \mathbf{f}_B^v$ of weight $\epsilon = \|\mathbf{e}\| = d_H(\mathbf{f}_B^e, \mathbf{f}_B^v)$, where d_H is the Hamming distance, corrupts the original codeword used in the key-binding process. The bit-error probability P_e, which is the probability that a bit of $\mathbf{e}$ is "1," determines the number of bit-errors that have to be corrected by the ECC decoder in order to return a match and therefore also the system performance. The bit-error probability depends on the quantization method being used, the quality of the features, and the number of samples (see Section III-B) and is different for imposter and genuine comparisons.

III. Analytical Framework

In this section, we present the analytical framework for modeling the biometric source, the quantization method, the system performance, and the maximum key size that can be extracted. An overview of this framework is depicted in Fig. 3. The *Source Modeling* module models the biometric source from which the enrolment and verification feature vectors $\mathbf{f}$ are derived. Given the input capacity C_{in} and the number of feature components n_c as it parameters the Source Modeling module outputs the quality of feature component j defined by the within-class and between-class standard deviation ratio $\sigma_b[j]/\sigma_w[j]$, referred to as the feature quality. With the quantization method under consideration,

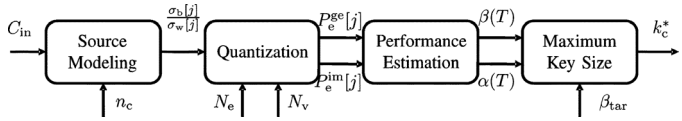

Fig. 3. Overview of the framework used to model the biometric source defined by the feature quality $\sigma_b[j]/\sigma_w[j]$ of the jth component, the resulting bit-error probabilities $P_e^{\mathrm{ge}}[j]$ and $P_e^{\mathrm{im}}[j]$, the corresponding performance defined by the FMR $\alpha(T)$ and the FNMR $\beta(T)$ at the operating point T, and the maximum key size k_c^* that can be extracted.

the number of enrolment N_e and verification N_v samples, and the feature quality $\sigma_b[j]/\sigma_w[j]$, the *Quantization* module estimates the bit-error probability of the extracted bit from feature component j at genuine $P_e^{\mathrm{ge}}[j]$ and imposter $P_e^{\mathrm{im}}[j]$ comparisons. Knowing the bit-error probabilities, the *Performance Estimation* module estimates the analytical system performance defined by the false match rate (FMR) $\alpha(T)$ and the false non-match rate (FNMR) $\beta(T)$ at all possible operating points T. Given the system performance and the target FNMR β_{tar}, the maximum extracted key size k_c^* is determined in the *Maximum Key Size* module. In the remainder of this section, we discuss each module in more detail.

A. Biometric Source Modeling With Parallel Gaussian Channels

The input of the FCS template protection system is a real-valued column feature vector $\mathbf{f} = [f[1], f[2], \ldots, f[n_c]]'$ of dimension n_c, where "$'$" is the transpose operator. The feature vector $\mathbf{f}$ is extracted from a biometric sample by the feature extractor and is likely to be different between two measurements, even if they are acquired immediately after each other. Causes for this difference include sensor noise, environmental conditions, and biometric variabilities. To model these variabilities, we use the parallel Gaussian channels (PGCs) as portrayed in Fig. 4(a). This approach has been successfully used in estimating the performance of two biometric databases in Kelkboom *et al.* (2010) [31] in which the validity of the PGC approach is shown. We assume an ideal *Acquisition and Feature-Extraction* module which always produces the same feature vector $\boldsymbol{\mu}_i$ for subject i. Such an ideal module is thus robust against all aforementioned variabilities. However, the variability of component j is modeled as an additive zero-mean Gaussian noise $w[j]$ with its pdf $p_{w[j],i} \sim \mathcal{N}\left(0, \sigma_{w,i}^2[j]\right)$. Adding the noise $w[j]$ with the mean $\mu_i[j]$ results in the noisy feature component $f[j]$, in vector notation $\mathbf{f} = \boldsymbol{\mu}_i + \mathbf{w}$. The observed variability within one subject is characterized by the variance of the *within-class* pdf and is referred to as within-class variability. We assume that each subject has the same within-class variance, i.e., homogeneous within-class variance $\sigma_{w,i}^2[j] = \sigma_w^2[j], \forall\, i$. We also assume the noise to be independent across components j, subjects i, and across measurements. Hence, the feature vector extracted from each biometric sample is equivalent to retransmitting $\boldsymbol{\mu}_i$ over the same PGC channels.

Each subject should have a unique set of means in order to be distinguishable. Across the population, we assume $\mu_i[j]$ to be another Gaussian random variable with density $p_{b[j]} \sim \mathcal{N}\left(\mu_b[j], \sigma_b^2[j]\right)$. The variability of $\mu_i[j]$ across the population

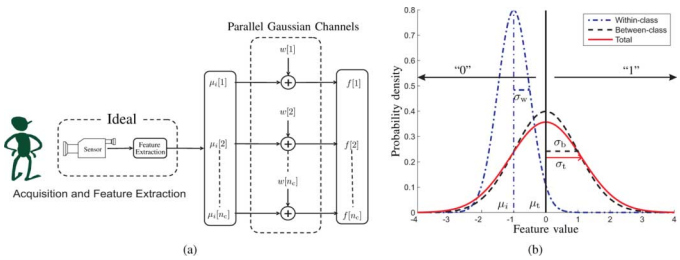

Fig. 4. (a) Parallel Gaussian channels modeling the real-valued features and (b) the within-class, between-class and the total density and the quantization method based on thresholding.

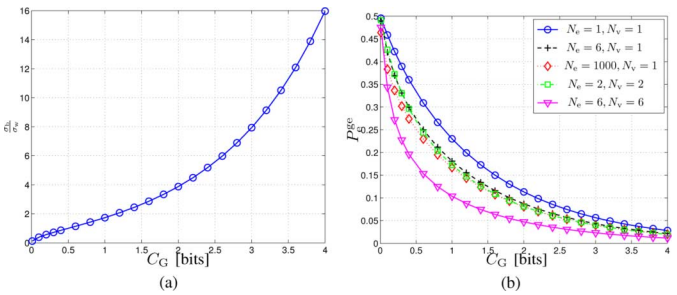

Fig. 5. (a) Feature quality σ_b/σ_w as a function of the Gaussian channel capacity C_G and (b) the genuine bit-error probability P_e^{ge} as a function of C_G for different values of the number of enrolment N_e and verification N_v samples.

is referred to as the *between-class* variability. Fig. 4(b) shows an example of the within-class and between-class pdfs for a specific component and a given subject. The *total* pdf describes the observed real-valued feature value $f[j]$ across the population and is also Gaussian with $p_{t[j]} \sim \mathcal{N}\left(\mu_t[j], \sigma_t^2[j]\right)$, where $\mu_t[j] = \mu_b[j]$ and $\sigma_t^2[j] = \sigma_w^2[j] + \sigma_b^2[j]$. For simplicity but without loss of generality we consider $\mu_t[j] = \mu_b[j] = 0$.

The capacity of each channel is given by the Gaussian channel capacity $C_G[j]$ as defined in Cover and Thomas (1991) [32]

$$C_G[j] = \frac{1}{2}\log_2\left(1 + \left(\frac{\sigma_b[j]}{\sigma_w[j]}\right)^2\right) \tag{1}$$

which in fact states that a maximum of $C_G[j]$ bits could be sent per transmission. Note that the Gaussian channel capacity only depends on the ratio $\sigma_b[j]/\sigma_w[j]$ and in Section III-B we will also show that the bit-error probability P_e depends on this ratio. Therefore, we can define the ratio $\sigma_b[j]/\sigma_w[j]$ as the feature quality of component j and taking its inverse of (1) we obtain

$$\frac{\sigma_b[j]}{\sigma_w[j]} = \sqrt{2^{2C_G[j]} - 1} \tag{2}$$

where the relationship is graphically represented in Fig. 5(a).

With the capacity of feature component j equal to the Gaussian channel capacity $C_G[j]$, we can define the total capacity of the input biometric source C_{in} as the following sum:

$$C_{\text{in}} = \sum_{j=1}^{n_c} C_G[j]. \tag{3}$$

The input capacity C_{in} thus represents the amount of discriminating information in a biometric sample across the population and is distributed among the n_c components. Is this work, we consider the input capacity C_{in} to be uniformly distributed among the n_c components. Hence, the Gaussian capacity of each component $C_G[j]$ is equal to C_{in}/n_c. By substituting $C_G[j] = C_{\text{in}}/n_c$ in (2), the feature quality parameter σ_b/σ_w related to the total capacity C_{in} as

$$\frac{\sigma_b}{\sigma_w} = \sqrt{2^{2C_{\text{in}}/n_c} - 1} \tag{4}$$

and is thus equal for each component.

B. Quantization Module Based on Thresholding

Fig. 4(b) depicts the quantization method under consideration, which is a binarization method based on thresholding, where the mean of the total density μ_t is taken as the threshold [19]–[21]. If the real-valued feature is larger than the threshold, then a bit of value "1" is allocated; otherwise "0." To estimate the analytical system performance we need to estimate the bit-error probability $P_e[j]$ for each component j at imposter and genuine comparisons. In this section, we analytically estimate $P_e[j]$ given the quantization scheme, the feature quality $\sigma_b[j]/\sigma_w[j]$, and the number of enrolment N_e and verification N_v samples.

1) Imposter Bit-Error Probability $P_e^{\text{im}}[j]$: At imposter comparisons, each bit is compared with the bit extracted from a randomly selected feature value from the total density. Because μ_t is the binarization threshold, there is a probability of 1/2 that a randomly selected bit from the population will be equal, hence $P_e^{\text{im}}[j] = 1/2$. Note that both the number of enrolment and verification samples do not have an influence on $P_e^{\text{im}}[j]$, and $P_e^{\text{im}}[j]$ is equal for each component.

2) Genuine Bit-Error Probability $P_e^{\text{ge}}[j]$: At genuine comparisons, the analytical bit-error probability $P_e^{\text{ge}}[j]$ has been derived in Kelkboom *et al.* (2008) [15], namely

$$P_e^{\text{ge}}[j] = \frac{1}{2} - \frac{1}{\pi}\arctan\left(\frac{\sigma_b[j]}{\sigma_w[j]}\frac{\sqrt{N_e N_v}}{\sqrt{N_e + N_v + \left(\frac{\sigma_b[j]}{\sigma_w[j]}\right)^{-2}}}\right) \tag{5}$$

which shows that the standard deviation ratio $\sigma_b[j]/\sigma_w[j]$ (the feature quality) and the number of enrolment N_e and verification N_v samples determine $P_e^{\text{ge}}[j]$. Note that $P_e^{\text{ge}}[j]$ is the average bit-error probability across the population. Some subjects have a larger bit-error probability because their mean $\mu_i[j]$ is closer to the quantization threshold $\mu_t[j]$, while others have a smaller bit-error probability because their mean is further away. However, for estimating the analytical system performance across an infinite number of subjects, it is only necessary to compute the average bit-error probability as shown in Kelkboom *et al.* (2010) [31]. With the assumption that the feature quality is equal for each component, substituting (4) into (5) we obtain

$$P_e^{\text{ge}} = \frac{1}{2} - \frac{1}{\pi}\arctan\left(\frac{\sqrt{\left(2^{2C_{\text{in}}/n_c} - 1\right) N_e N_v}}{\sqrt{N_e + N_v + \left(2^{2C_{\text{in}}/n_c} - 1\right)^{-1}}}\right). \tag{6}$$

With (5) or (6) it is easy to show that P_e^{ge} for the $N_e = N_v = 2X$ case converges to the $\{N_e = \infty, N_v = X\}$ case when the feature quality increases. For example, the argument of the $\arctan$ function in (5) for N_e approaching infinity becomes

$$\lim_{N_e \to +\infty} \frac{\sigma_b[j]}{\sigma_w[j]} \frac{\sqrt{N_e N_v}}{\sqrt{N_e + N_v + \left(\frac{\sigma_b[j]}{\sigma_w[j]}\right)^{-2}}} = \frac{\sigma_b[j]}{\sigma_w[j]} \sqrt{N_v}. \quad (7)$$

Furthermore, under the assumption that $\sigma_b[j]/\sigma_w[j] \gg (N_e + N_v)^{-2}$, we can approximate the argument of the $\arctan$ function as

$$\frac{\sigma_b[j]}{\sigma_w[j]} \frac{\sqrt{N_e N_v}}{\sqrt{N_e + N_v + \left(\frac{\sigma_b[j]}{\sigma_w[j]}\right)^{-2}}} \approx \frac{\sigma_b[j]}{\sigma_w[j]} \sqrt{\frac{N_e N_v}{N_e + N_v}}. \quad (8)$$

For the first case we consider the number of enrolment and verification samples to be equal, namely $N_{e,1} = N_{v,1}$, while for the second case we consider $\{N_{e,2} = \infty, N_{v,2}\}$. For these two cases, the error probability is equal if the argument of the $\arctan$ function is equal. This results in

$$\begin{aligned} \frac{\sigma_b[j]}{\sigma_w[j]} \sqrt{\frac{N_{e,1} N_{v,1}}{N_{e,1} + N_{v,1}}} &= \frac{\sigma_b[j]}{\sigma_w[j]} \sqrt{N_{v,2}} \\ \frac{N_{e,1} N_{v,1}}{N_{e,1} + N_{v,1}} &= N_{v,2} \\ N_{v,1} &= 2N_{v,2}, \text{ with } N_{e,1} = N_{v,1}. \end{aligned} \quad (9)$$

Hence, we have shown that P_e^{ge} converges for the cases $N_e = N_v = 2X$ and $\{N_e = \infty, N_v = X\}$ when the feature quality increases. Note, that the convergence also holds for the $\{N_e = X, N_v = \infty\}$ case.

Fig. 5(b) depicts the bit-error probability P_e^{ge} as a function of C_G for different settings of N_e and N_v as defined by (6). By increasing N_e, P_e^{ge} decreases because the bits extracted in the enrolment phase are more stable, i.e., a smaller within-class variance. However, when increasing N_e further to infinity, P_e^{ge} stays close to the $N_e = N_v = 2$ case and converges when C_G increases. To further decrease P_e^{ge}, it is thus necessary to also increase N_v.

These findings can help the designer of the biometric system when determining the number of enrolment and verification samples. These findings show that the reduction of the bit-error probability (and thus an improvement of the system performance) is limited when increasing only the number of enrolment or verification samples. Above a certain number of enrolment (verification) samples, the improvement of the system performance is minimal and it would be more advantageous to increase the number of verification (enrolment) samples.

C. System Performance

In Section II, we have modeled the channel between the encoder and decoder of the FCS template protection system as a binary symmetric channel with bit-error probability $P_e[j]$. The bit-error probability determines the probability mass function (pmf) of the number of bit errors or Hamming distance $\epsilon = d_H(\mathbf{f}_B^e, \mathbf{f}_B^v)$. As presented in Kelkboom *et al.* (2010) [31], the pmf is defined by the convolution

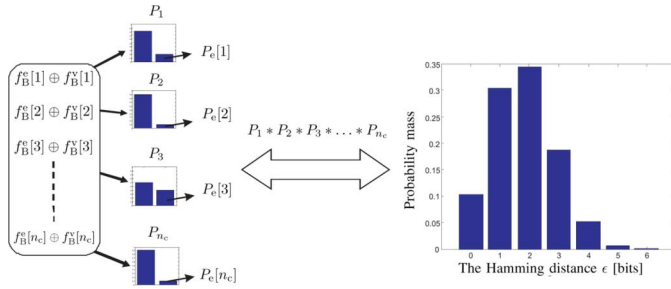

Fig. 6. Toy example of the convolution method given by (7). (From Kelkboom (2010) [31].)

$$\begin{aligned} \phi(\epsilon) &\stackrel{\text{def}}{=} \mathcal{P}\{d_H(\mathbf{f}_B^e, \mathbf{f}_B^v) = \epsilon\} \\ &= (P_1 * P_2 * \cdots * P_{n_c})(\epsilon) \end{aligned} \quad (10)$$

where $P_j = [1 - P_e[j], P_e[j]]$ is the marginal pmf of the single bit extracted from component j. A toy example is depicted in Fig. 6. The toy example shows the marginal pmf at comparisons between the enrolment and verification bits $f_B^e[1]$ and $f_B^v[1]$, respectively. Taking the convolution of all marginal pmf leads to the pmf of the Hamming distance ϵ.

Because we consider the input capacity to be uniformly distributed across the n_c components, $P_e[j]$ is equal for each component, namely P_e. Hence, the convolution in (10) becomes a binomial pmf $P_b(\epsilon; N, p)$ as discussed in Daugman (2003) [33]

$$P_b(\epsilon; N, p) = \binom{N}{\epsilon} p^{\epsilon} (1-p)^{(N-\epsilon)} \quad (11)$$

with dimension $N = n_c$ and probability $p = P_e$.

1) False Match Rate: The false match rate (FMR) depends on the pmf of the Hamming distance ϵ at imposter comparisons, where we have the bit-error probability P_e^{im} that is equal for each extracted bit. Therefore, the pmf of the Hamming distance ϵ is the binomial pmf with p equal to P_e^{im}. Hence, the FMR at the operating point T, $\alpha(T)$, is the probability that ϵ is smaller or equal to T (see Fig. 7), namely

$$\begin{aligned} \alpha(T) &\stackrel{\text{def}}{=} \mathcal{P}\{\epsilon \leq T \mid \text{imposter comparisons}\} \\ &= \sum_{i=0}^{T} P_b(i; n_c, P_e^{\text{im}}) \\ &= 2^{-n_c} \sum_{i=0}^{T} \binom{n_c}{i}. \end{aligned} \quad (12)$$

2) False Nonmatch Rate: In general, P_e^{ge} is not equal for each bit and therefore the pmf of the Hamming distance ϵ at genuine comparisons is defined by the convolution of (10) with marginal pmf's $P_j^{\text{ge}} = [1 - P_e^{\text{ge}}[j], P_e^{\text{ge}}[j]]$. Hence, the false

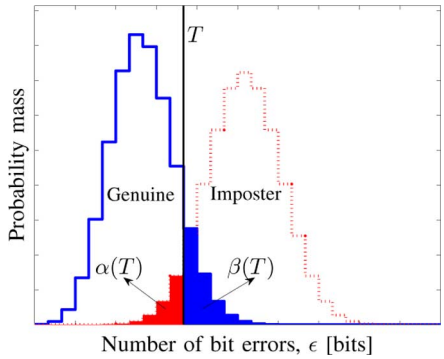

Fig. 7. False match rate (FMR) and the false nonmatch rate (FNMR) given the probability mass function of the number of errors ϵ at imposter and genuine comparisons.

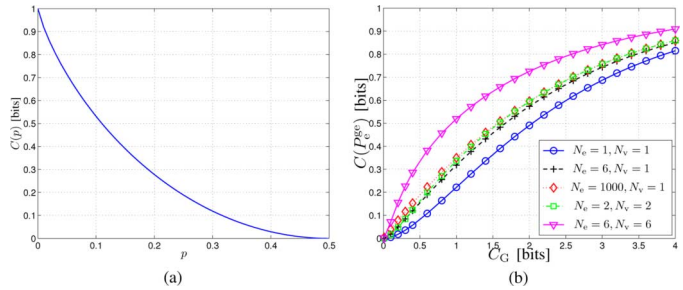

Fig. 8. (a) Binary symmetric channel (BSC) capacity as a function of the bit-error probability p, and (b) the BSC capacity $C(P_e^{\mathrm{ge}})$ as a function of the uniformly distributed input capacity C_{in}/n_c at different values of the number of enrolment N_e and verification N_v samples.

nonmatch rate at the operating point T, $\beta(T)$, is the probability that ϵ is larger than T (see Fig. 7), namely

$$\begin{aligned}\beta(T) &\stackrel{\mathrm{def}}{=} \mathcal{P}\{\epsilon > T \mid \text{genuine comparisons}\} \\ &= \sum_{i=T+1}^{n_c} \left(P_1^{\mathrm{ge}} * P_2^{\mathrm{ge}} * \cdots * P_{n_c}^{\mathrm{ge}}\right)(i). \end{aligned} \tag{13}$$

With the input capacity uniformly distributed among the n_c components, the pmf of ϵ is given by the binomial pmf with probability $p = P_e^{\mathrm{ge}}$, namely

$$\begin{aligned}\beta(T) &= \sum_{i=T+1}^{n_c} P_b\left(i; n_c, P_e^{\mathrm{ge}}\right) \\ &= \sum_{i=T+1}^{n_c} \binom{N}{i} \left(P_e^{\mathrm{ge}}\right)^i \left(1 - P_e^{\mathrm{ge}}\right)^{(N-i)}. \end{aligned} \tag{14}$$

D. Maximum Key Size

As discussed in Section II, the ECC has to decode the corrupted codeword in order to retrieve the encoded key from the enrolment phase. A decoding error occurs when the number of corrupted bits is larger than the error-correcting capability of the ECC. Hence, the decoding error probability determines the FNMR and FMR of the biometric system. Furthermore, the size of the encoded key depends on the number of bits the ECC has to correct, referred to as the operating point, and the codeword size. We assume an ideal binary ECC that corrects up to t_c random bit errors of equal bit-error probability and the ECC operates at the theoretical maximum, e.g., Shannon's bound.

In this section, we investigate the relationship between the bit-error probabilities corrupting the codeword, the maximum key size that can be encoded in the enrolment phase, and the performance of the biometric system given by the FMR and FNMR given the ideal ECC we defined above.

First we discuss Shannon's theorem on which the decoding properties of our ideal ECC is based. We will show that for a biometrics system with a limited codeword size n_c, the FNMR at the operating point stipulated by Shannon's theorem will be close to 50%. Such an FNMR is unacceptable for a biometric system. Hence, we analyze the key size achieved at other operating points such as the equal-error rate (EER), where the FMR is equal to the FNMR, and the operating point determined by the target FNMR, β_{tar}. We define the maximum key size as the key size obtained at the operating point β_{tar}. We conclude with the comparison between the maximum key size at a given operating point and the upper bound given by the corresponding FMR as published in Korte and Plaga (2007) [11] and Buhan *et al.* [12].

1) Shannon's Theorem: With the code rate R equal to the ratio of the key size and the codeword size k_c/n_c, Shannon's noisy channel decoding theorem [34] shows that there exists a decoding technique that can decode the corrupted codeword with a bit-error rate p with an arbitrary small probability of a decoding error when

$$R < C(p) \tag{15}$$

for a sufficiently large value of n_c, where $C(p)$ is the channel capacity defined as

$$C(p) = 1 - h(p) \tag{16}$$

with $h(p)$ being the binary entropy function

$$h(p) = -p\log_2 p - (1-p)\log_2(1-p). \tag{17}$$

Hence, the key size k_c has an upper limit given by Shannon's bound with $p = P_e^{\mathrm{ge}}$ as

$$k_c = n_c R < n_c C\left(P_e^{\mathrm{ge}}\right). \tag{18}$$

With use of (6), we have the relationship between the uniformly distributed input capacity C_{in}/n_c and the BSC channel capacity $C(P_e^{\mathrm{ge}})$ as illustrated in Fig. 8(b) for different numbers of enrolment N_e and verification N_v samples settings. Increasing the number of samples decreases of the genuine bit-error probability P_e^{ge} and, therefore, increases the BSC channel capacity $C(P_e^{\mathrm{ge}})$.

With a code rate close to the bound given by (18), the decoding error is negligible only when n_c is large enough. In a biometric system, however, n_c is not very large. As described in Daugman (2003) [33], the intrinsic degrees of freedom of the binary iris code is 249, which has been derived by fitting the imposter Hamming distance pmf with a binomial pmf with probability $p = 0.5$ and dimension $N = 249$. The impact of this small dimension on the FNMR is depicted by the toy example in Fig. 9. The figure illustrates the achieved FNMR when choosing the operating point $T/n_c = 0.2$ close to $P_e^{\mathrm{ge}} = 0.19$

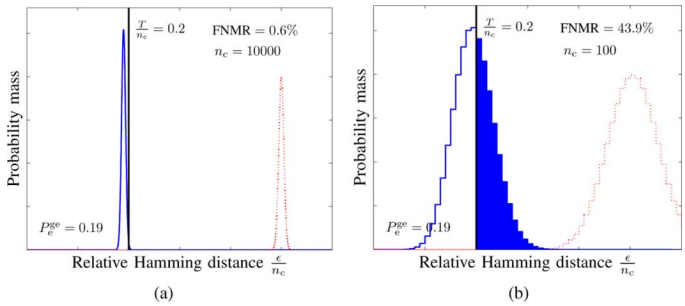

Fig. 9. Toy example of the achieved FNMR when choosing the operating, $T/n_c = 0.2$ close to $P_e^{ge} = 0.19$ as stipulated by Shannon's theorem for different values of n_c. The solid (blue) curve portrays the pmf of the Hamming distance ϵ at genuine comparisons, while the dotted (red) curve depicts the pmf at imposter comparisons.

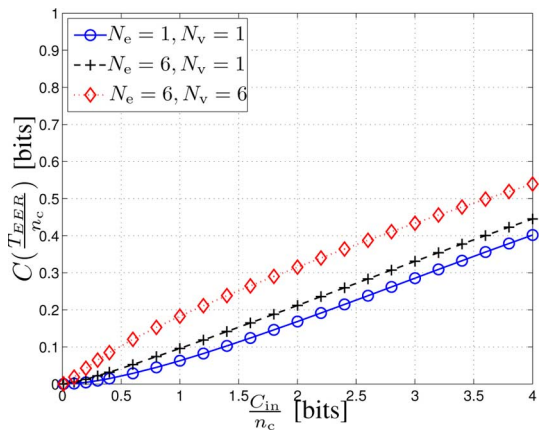

Fig. 10. BSC channel capacity at the EER operating point $C(T_{\mathrm{EER}}/n_c)$ as a function of the uniformly distributed input capacity C_{in}/n_c at different values of N_e and N_v.

as stipulated by Shannon's theorem for different values of n_c. At a large codeword size of $n_c = 10\,000$ bits, the achieved FNMR is 0.6%, which is acceptable. Note, however, that the FNMR significantly increases once n_c decreases, namely 43.9% at $n_c = 100$ bits, respectively. Hence, when the iris has 249 independent bits and is known as one of the best biometrics modality, we can conclude that the codeword size is expected to be too small to achieve an acceptably small FNMR. To lower the FNMR we have to correct more bits. In Section IV, we describe two alternative operating points, namely at the EER operating point or at the target FNMR β_{tar}.

2) The EER Operating Point With Gaussian Approximation: In order to find an analytical expression of the EER operating point, T_{EER}, we approximate the binomial density used for modeling the pmf of the Hamming distance ϵ by a Gaussian density. The EER operating point in terms of P_e^{ge} becomes

$$\frac{T_{\mathrm{EER}}}{n_c} = \frac{\sqrt{P_e^{\mathrm{ge}}\left(1-P_e^{\mathrm{ge}}\right)} + P_e^{\mathrm{ge}}}{2\sqrt{P_e^{\mathrm{ge}}\left(1-P_e^{\mathrm{ge}}\right)} + 1}. \tag{19}$$

where the complete derivation is presented in Section A. Note that the relative operating point T_{EER}/n_c is fully determined by P_e^{ge} and therefore also the uniformly distributed input capacity C_{in}/n_c. The relationship between the BSC channel capacity at the EER operating point $C(T_{\mathrm{EER}}/n_c)$ and C_{in}/n_c is depicted in Fig. 10.

3) Operating Point at the Target FNMR β_{tar}: We have shown that the operating point stipulated by Shannon's theory leads to an optimistic upper bound with a high FNMR, while the EER operating point may not be the ideal operating point of a biometric system in terms of FMR, which consequently leads to a smaller maximum key size. In this section, we present a different operating point determined by the target performance, namely the target FNMR, β_{tar}. Hence, instead of correcting $t_c = n_c P_e^{\mathrm{ge}}$ or T_{EER} bits, we will correct $t_c = T_{\mathrm{tar}}$ bits, where T_{tar} is the operating point in order to reach β_{tar}, namely

$$T_{\mathrm{tar}} = \arg\min_T(|\beta(T) - \beta_{\mathrm{tar}}|). \tag{20}$$

Hence, the theoretical maximum key size assuming an ECC at Shannon's bound with $p = T_{\mathrm{tar}}/n_c$ is then equal to

$$k_c^* \stackrel{\mathrm{def}}{=} n_c C\left(\frac{T_{\mathrm{tar}}}{n_c}\right) = n_c\left(1 - h\left(\frac{T_{\mathrm{tar}}}{n_c}\right)\right). \tag{21}$$

Because T_{tar}/n_c is larger than P_e^{ge} and will not exceed $1/2$, we know that k_c^* will be smaller than the upper bound $n_c C(P_e^{\mathrm{ge}})$ from (18). However, if β_{tar} is larger than the EER, then k_c^* will be larger than $C(T_{\mathrm{EER}}/n_c)$.

We have defined the maximum key size k_c^*, which we will use in the remainder of this work. In Section IV, we study the effect of the system parameters of the framework shown in Fig. 3 on k_c^*.

4) Relationship Between the Maximum Key Size k_c^* and the Target FMR α_{tar}: The work of Korte and Plaga (2007) [11] showed the relationship between the key size k_c and the FMR to be $k_c \leq -\log_2(\alpha(T))$ by using the Hamming bound theorem. Namely, [35, Th. 6, p. 19] [MacWilliams and Sloane (1977)] (the sphere packing or Hamming bound) states: *A t_c-error binary code of length n_c containing M codewords must satisfy*

$$M\left(1 + \binom{n_c}{1} + \binom{n_c}{2} + \cdots + \binom{n_c}{t}\right) \leq 2^{n_c}. \tag{22}$$

With the FMR defined in (12) as $\alpha(T) = 2^{-n_c}\sum_{i=0}^{T}\binom{n_c}{T}$ with $t_c = T$ and $M = 2^{k_c}$, we obtain

$$\begin{aligned} k_c &\leq -\log_2(\alpha(T)) \\ &\leq -\log_2(\alpha_{\mathrm{tar}}), \text{ with } T = T_{\mathrm{tar}} \end{aligned} \tag{23}$$

where we define the FMR at the target operating point T_{tar} as α_{tar}. Thus, we have two upper bounds for the key size at a given operating point, namely $\log_2(\alpha_{\mathrm{tar}})$ from the Hamming bound theorem from (23) and k_c^* from Shannon's theorem from (21). We compare the difference between the two bounds $(-\log_2(\alpha_{\mathrm{tar}}) - k_c^*)$ as a function of the relative operating point T/n_c at a fixed number of components n_c, as illustrated in Fig. 11 for different n_c settings. We observe that if no errors have to be corrected, $T = 0$, then there is no difference because $(-\log_2(\alpha_{\mathrm{tar}}) - k_c^*) = 0$. However, if errors have to be corrected, we observe a difference, where its maximum is around $T_{\mathrm{tar}}/n_c = 0.2$. A larger maximum is observed for larger n_c values.

Hence, $-\log_2(\alpha_{\mathrm{tar}})$ is an upper bound of the key size k_c at the target operating point. However, given the example of Fig. 11, $-\log_2(\alpha_{\mathrm{tar}})$ is 2–4 bits larger than the maximum key

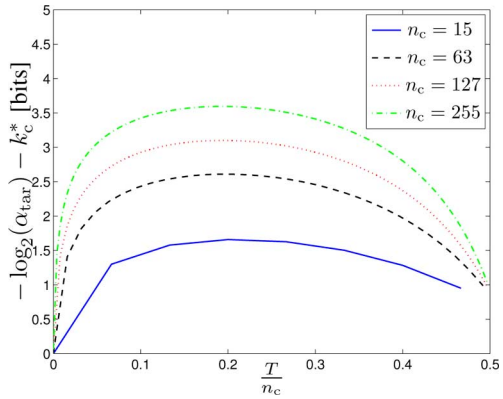

Fig. 11. Difference $(-\log_2(\alpha_{\rm tar}) - k_c^*)$ as a function of relative operating point $T_{\rm tar}/n_c$ with n_c fixed at different n_c settings.

size k_c^* defined by (21). Furthermore, the difference between the two bounds increases when there are more components. For example, the difference can be around 3 bits when the codeword is 127 bits long.

IV. Numerical Analysis of the System Performance and the Maximum Key Size

By means of a numerical analysis we illustrate the effect of the system parameters on both the system performance and the theoretical maximum key size k_c^*. As the system parameters, we have the input capacity $C_{\rm in}$, the number of enrolment N_e and verification N_v samples, and the target FNMR $\beta_{\rm tar}$. In Section IV-A, we analyze the case where the feature components are independent, while in Section IV-B, some feature components are dependent. An extended version of the numerical analysis can be found in [36].

A. Biometric Source With Independent Feature Components

First, we discuss the effect of the parameters $\{C_{\rm in}, \beta_{\rm tar}\}$ on the maximum key size at the target FNMR. Note that we compute the optimal number of components n_c^* for the given input capacity $C_{\rm in}$. The optimal number of components is defined as the number of components, across which $C_{\rm in}$ is uniformly distributed, that leads to the best system performance in terms of the FMR and the FNMR. Fig. 12(a) and (b) portrays the effect of the target FNMR $\beta_{\rm tar}$ and the input capacity $C_{\rm in}$ on the maximum key size k_c^* with a single enrolment and verification sample $N_e = N_v = 1$, where Fig. 12(a) depicts k_c^* as a function of $C_{\rm in}$ with different $\beta_{\rm tar}$ settings and Fig. 12(b) shows k_c^* as a function of $\beta_{\rm tar}$ with different $C_{\rm in}$ settings. Similarly, the effect of $\beta_{\rm tar}$ and $C_{\rm in}$ on the relative operating point $T_{\rm tar}/n_c^*$ and the optimal number of components n_c^* are illustrated in Fig. 12(c)–(f), respectively. The results show that increasing either the input capacity $C_{\rm in}$ or the target FNMR $\beta_{\rm tar}$ increases the maximum key size k_c^* and the optimal number of components n_c^*, but decreases the relative operating point $T_{\rm tar}/n_c^*$. Both the increase of n_c^* and the decrease of $T_{\rm tar}/n_c^*$ have a positive effect on the maximum key size k_c^*. Doubling $\beta_{\rm tar}$ from 10% to 20% on average adds around 2 bits to k_c^*, but from 2.5% to 5% on average adds 1 bit. Furthermore, doubling $C_{\rm in}$ roughly doubles k_c^* for the case when $\beta_{\rm tar} = 20\%$ and almost triples for

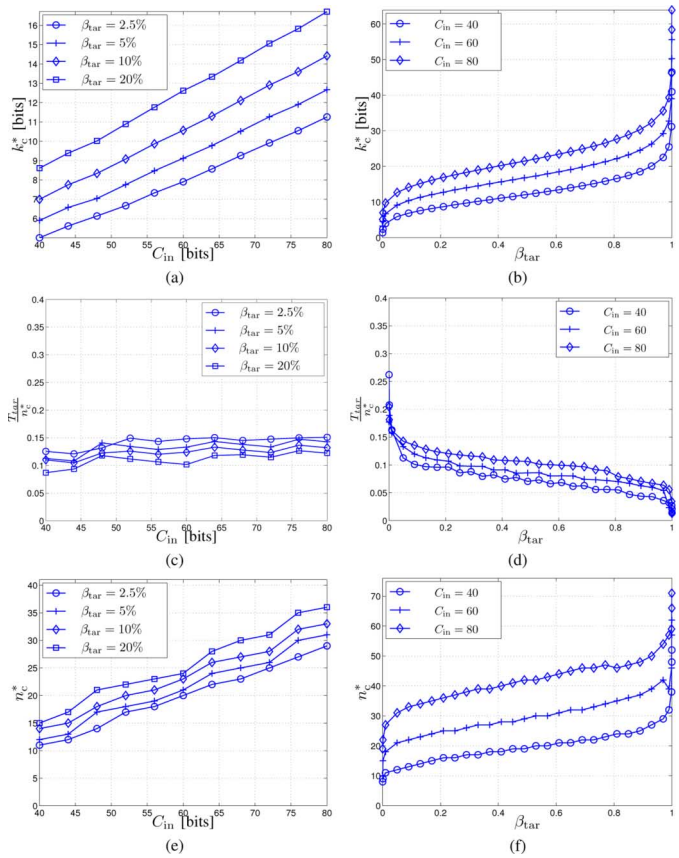

Fig. 12. Subfigures (a), (c), and (e) depict the maximum key size k_c^*, the relative targeted operating point $T_{\rm tar}/n_c$, and the optimal number of components n_c^* as a function of the input capacity $C_{\rm in}$ at different target FNMR $\beta_{\rm tar}$ settings, respectively. Similarly, (b), (d), and (f) depict them as a function of $\beta_{\rm tar}$ with different $C_{\rm in}$ settings.

the case when $\beta_{\rm tar} = 2.5\%$. Also, Fig. 12(b) shows that if $\beta_{\rm tar}$ is small, namely $\leq 5\%$, there is a significant drop of k_c^* when $\beta_{\rm tar}$ decreases further. At smaller $\beta_{\rm tar}$, it is required to correct more bits (as shown in Fig. 12(c) by the increase in $T_{\rm tar}/n_c$), hence it is important to extract bits with smaller bit-error probabilities $P_e^{\rm ge}[j]$. Therefore, at a fixed $C_{\rm in}$, there have to be less components in order for each component to have a better feature quality σ_b/σ_w or Gaussian channel capacity $C_G[j]$ leading to a smaller $P_e^{\rm ge}[j]$. On the contrary, when $\beta_{\rm tar}$ is close to 1, there is a significant increase in k_c^*. If $\beta_{\rm tar}$ converges to 1, k_c^* goes to infinity. In this case, because of the large target FNMR it is not necessary to correct many bits with its extreme case where no bits at all have to be corrected. Hence, many components [see Fig. 12(f)] can be extracted with a worse feature quality or a smaller $C_G[j]$.

Second, we show the effect of the parameters $\{C_{\rm in}, N_e, N_v\}$ on the system performance and the maximum key size. Fig. 13 depicts the effect of the $\{N_e, N_v, C_{\rm in}\}$ parameters on the maximum key size k_c^*, the relative operating point $T_{\rm tar}/n_c$, and the optimal number of components n_c^*. The effect of the input capacity $C_{\rm in}$ is similar as illustrated in Fig. 12(a). Furthermore, increasing either the number of enrolment N_e or verification N_v samples leads to an increase of k_c^*. However, keeping either N_e or N_v fixed while increasing the other shows that k_c^* increases asymptotically and is limited [see Fig. 13(b)]. Changing both

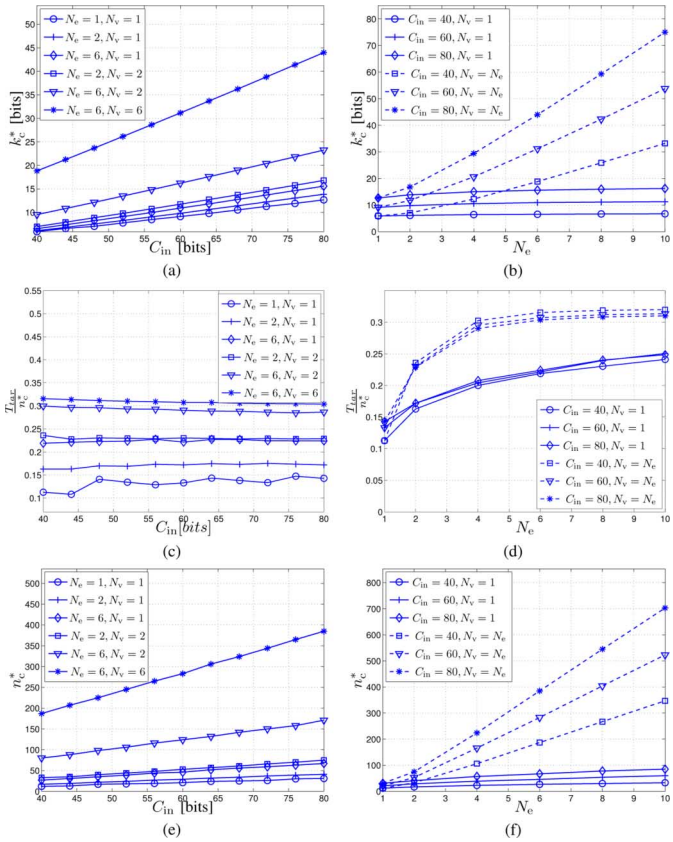

Fig. 13. Subfigures (a), (c), and (e) depict the maximum key size k_c^*, the relative targeted operating point T_{tar}/n_c^*, and the number of components n_c^* as a function of input capacity C_{in} at different $\{N_e, N_v\}$ settings, respectively. Similarly, (b), (d), and (f) depict them as a function of $\{N_e, N_v\}$ with different C_{in} settings. In all cases we have $\beta_{\text{tar}} = 5\%$.

N_e and N_v significantly increase k_c^*. In general, increasing the number of samples enables the use of components with a worse feature quality, hence increasing the optimal number of components n_c^* when the input capacity C_{in} is fixed. Consequently, the relative operating point T_{tar}/n_c^* increases because of the lower quality leading to a larger bit-error probability. A larger T_{tar}/n_c^* leads to a smaller channel capacity and, therefore, a smaller possible key size. However, the optimal number of components increases stronger leading to a net increase of the maximum key size k_c^*.

Some examples of the maximum key size increase are as follows. Within the specific range of target FNMR $2.5\% \leq \beta_{\text{tar}} \leq 20\%$ and the input capacity $40 \leq C_{\text{in}} \leq 80$, doubling the target FMR adds 1 to 2 bits to the maximum keys size k_c^*. Doubling the input capacity C_{in} doubles the maximum key size k_c^* when $\beta_{\text{tar}} = 20\%$ and almost triples when $\beta_{\text{tar}} = 2.5\%$. Furthermore, for the case where the target FNMR is at $\beta_{\text{tar}} = 5\%$, increasing the number of enrolment samples N_e from one to six samples increases the maximum key size k_c^* with 0.6 bits (from 5.9 to 6.5) at $C_{\text{in}} = 40$ bits and 2.9 bits (from 12.7 to 15.6) bits at $C_{\text{in}} = 80$ bits. Keeping $N_e = 6$ and increasing the number of verification samples N_v from one to two samples increases k_c^* with 3.0 bits at $C_{\text{in}} = 40$ and 7.6 bits at $C_{\text{in}} = 80$ bits. A further increase of N_v from two to six samples increases k_c^* with 9.3 bits at $C_{\text{in}} = 40$ and 20.8 bits at $C_{\text{in}} = 80$ bits.

B. Biometric Source With Dependent Feature Components

Until now we have assumed the extracted feature vector components and the channel noise to be independent across components and measurements. However, in practice the components may be dependent. In this section, we will show that the defined maximum key size is an overestimation when components are dependent. Differences in key size estimates due to dependent feature components may have caused the large deviations between the reported key size and FMR as outlined in Table I.

In the following analysis, only a limited number of feature components is assumed to be fully dependent, while the remainder of the feature set is assumed to be independent, because a detailed analysis of the dependencies is beyond the scope of this work. Consider a feature vector with N_F components. We assume that the first n_ρ components have in addition κ_ρ components that are fully dependent (duplicate or identical components), while the remaining $n_{\overline{\rho}}$ components have no duplicates. Hence, it holds that $N_F = n_\rho + n_{\overline{\rho}}$ and the total number of components n_c is equal to $n_c = n_\rho(\kappa_\rho + 1) + n_{\overline{\rho}}$. Furthermore, we define the array with n zeros as $\mathbb{O}_n = [0_1, 0_2, \ldots, 0_n]$. With the assumed dependency model, the pmf of the number of bit errors ϵ as defined by (10) becomes

$$\begin{aligned}\phi(\epsilon) &\stackrel{\text{def}}{=} \mathcal{P}\{d_H(\mathbf{f}_B^e, \mathbf{f}_B^v) = \epsilon\} \\ &= (P_{\rho,1} * P_{\rho,2} * \cdots * P_{\rho,n_\rho} * P_{\overline{\rho},n_\rho+1} * \cdots * P_{\overline{\rho},n_c})(\epsilon)\end{aligned} \tag{24}$$

where $P_{\rho,j} = [1 - P_e[j], \mathbb{O}_{\kappa_\rho}, P_e[j]]$ is the marginal pmf of the Hamming distance from the extracted bits from the set of $\kappa_\rho + 1$ identical components for the first n_ρ components and $P_{\overline{\rho},j} = [1 - P_e[j], P_e[j]]$ is the pmf for the extracted bit from the last $n_{\overline{\rho}}$ components without duplicates. For the set of $\kappa_\rho + 1$ identical bits it is only possible to have zero or $\kappa_\rho + 1$ bit errors with probability $1 - P_e$ and P_e, respectively. As in the previous sections, we can use the same equations for estimating the performance and the maximum key size at the target FNMR.

The results for the case where $N_F = 50$ with input capacity $C_{\text{in}} = 80$ bits and target FNMR at $\beta_{\text{tar}} = 5\%$ is portrayed in Fig. 14, where the first n_ρ components have a single duplicate $\kappa_\rho = 1$. The ROC performance curve deteriorates once duplicate components are added as shown in Fig. 14(a). In other words, the FMR α_{tar} at the target FNMR β_{tar} increases, as illustrated by the decrease of $-\log_2(\alpha_{\text{tar}})$ in Fig. 14(b). Furthermore, the relative operating point T_{tar}/n_c also increases. Although the increase of T_{tar}/n_c reduces the capacity $C(T_{\text{tar}}/n_c)$, we observe that the maximum key size k_c^* increases due to the increase of n_c. However, further increasing n_ρ until each component has κ_ρ duplicates ($n_\rho = N_F$) leads to the same α_{tar} and T_{tar}/n_c as for the case where no components have a duplicate ($n_\rho = 0$). Although the performance is similar, the maximum key size k_c^* has doubled.

The effects of changing κ_ρ are shown in Fig. 15. When all feature components have a duplicate, $n_\rho = N_F$, we can see from Fig. 15(a) that the maximum key size k_c^* increases by $(\kappa_\rho + 1)$ when compared to the case where no feature components have a duplicate $n_{\overline{\rho}} = 0$. Furthermore, Fig. 15(b) shows that the FMR deviation increases when increasing the number of duplicates κ_ρ. Note that the largest FMR, hence the

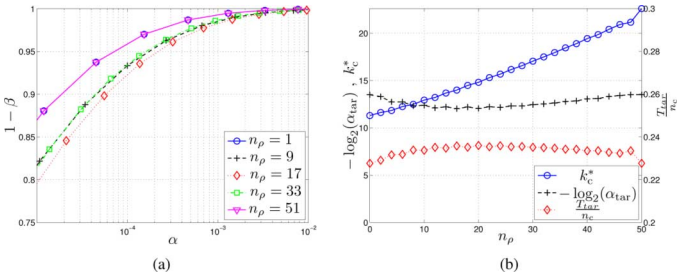

Fig. 14. (a) Performance ROC curve for different n_ρ settings and (b) the maximum key size k_c^*, the log of the FMR at the operating point $-\log_2(\alpha_{\rm tar})$, and the relative operating point $T_{\rm tar}/n_c$ as a function of the number of dependent components n_ρ. For both cases, the input capacity is $C_{\rm in} = 80$ bits with the target FNMR at $\beta_{\rm tar} = 5\%$.

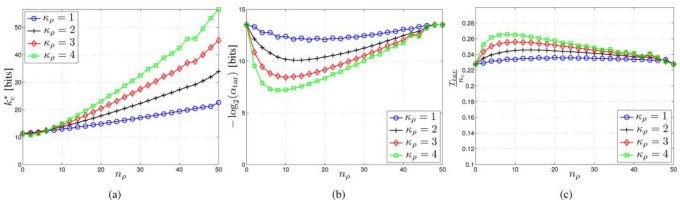

Fig. 15. (a) Maximum key size k_c^*, (b) the log of the FMR at the operating point $-\log_2(\alpha_{\rm tar})$, and the relative operating point $T_{\rm tar}/n_c$ as a function of the number of duplicates κ_ρ.

smallest $-\log_2(\alpha_{\rm tar})$, is achieved at the point where the average Hamming distance from the dependent and independent bits are equal, namely $(\kappa_\rho + 1)n_\rho = n_{\overline{\rho}}$. With $n_{\overline{\rho}} = N_F - n_\rho$, we obtain the point $n_\rho = N_F/\kappa_\rho + 2$. Not only does κ_ρ influence the FMR at the target FNMR and, therefore, also the maximum key size k_c^*, it also influences the relative operating point $T_{\rm tar}/n_c$, which increases with κ_ρ.

Hence, it seems that the maximum key size k_c^* could be increased by adding identical components. However, we argue that the protection actually does not increase because the FMR $\alpha_{\rm tar}$ at the target FNMR $\beta_{\rm tar}$ is either kept unchanged or even decreases. We also observed in Section III-D4 that another upper bound for the key size is $-\log_2(\alpha_{\rm tar})$, which is smaller than the maximum key size when identical bits are added by either increasing n_ρ or κ_ρ. This discrepancy between the FMR bound $-\log_2(\alpha_{\rm tar})$ and the maximum key size is caused by the fact that the ECC is modeled as a Hamming distance classifier that considers each bit to be independent. Hence, the space $\{0,1\}^{n_c}$ is assumed to be fully used and only under this assumption the maximum key size could be achieved. By adding identical components the space $\{0,1\}^{n_c}$ is not fully used, but is reduced to $\{0,1\}^{N_F}$.

We can conclude that by adding multiple κ_ρ identical components to the feature vector the maximum key size can be increased artificially; however, the actual protection indicated by the FMR will at most stay equal. We conjecture that this effect may have caused the large deviations between the reported key size and FMR as outlined in Table I.

V. Experiments

By means of numerical analysis, previous sections illustrated the effects of the system parameters such as the number of enrolment N_e and verification N_v samples on the performance and the maximum key size k_c^*. In this section, we will analyze these findings using an actual biometric database and two feature extraction algorithms.

A. Biometric Modality and Database

The database we use is the Ministerio de Ciencia y Tecnología (MCYT) containing fingerprint images from a capacitive and optical sensor as described in Ortega-Garcia *et al.* (2003) [37]. It contains 12 images of all 10 fingers from 330 subjects for each sensor. However, we limit our dataset to only the images of the right-index finger from the optical sensor, hence there are in total 3960 fingerprint images.

B. 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 prealignment step is performed during the feature extraction process. Such prealignment requires extraction of the core point which is performed according to the algorithm described in Ignatenko *et al.* (2002) [38]. 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. Our feature extraction algorithm failed to extract a feature vector from one subject, due to the failure of finding a core point, so we excluded it from the dataset. Therefore, there are effectively $N_s = 329$ subjects with a total of 3948 fingerprint images.

1) Directional Field Feature: 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 perpendicularly positioned to the gradient vector. In order to extract directional field features from a fingerprint, the algorithm described in Gerez and Bazen (2002) [39] is applied on each grid point. The directional field features have a dimension of $N_F = 578$ and are referred to as the DF features.

2) Gabor Filters Feature: The second type of extracted features are the Gabor filters (GF) features, described in Bazen and Veldhuis (2004) [40], where each grid point is filtered using a set of four 2-D Gabor filters at angles of $\{0, \pi/4, \pi/2, 3\pi/4\}$, respectively. The feature vector is the concatenation of the modulus of the four complex responses at each grid point, resulting in a feature vector dimension of $N_F = 1156$.

3) Dimension Reduction: To decorrelate and reduce the number of 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. The PCA and LDA transformation matrices are computed using the training set. $N_{\rm PCA}$ is the reduced dimension after applying the PCA transformation and $N_{\rm LDA}$ is the reduced dimension after applying the LDA transformation. We limit $N_{\rm LDA}$ to the number of subjects within the training set from which the transformation matrices are determined.

C. Testing Protocol

The performance testing protocol consists of randomly selecting 219 out of $N_s = 329$ subjects as the training set and the remaining 110 subjects as the evaluation set, which is referred to as the training-evaluation-set split. The template protection system parameters such as the quantization thresholds used within the Quantization module of Fig. 1 and the PCA and LDA transformation matrices are estimated using the training set.

From the evaluation set, N_e samples of each subject are randomly selected as the enrolment samples while the remaining samples are considered as the verification samples. This split is referred to as the enrolment-verification split. The protected template is generated using all the enrolment samples and compared with the average of N_v verification samples. 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. Note that the number of genuine and imposter comparisons depends on the number of enrolment and verification samples. For the genuine case, we have 30 250 comparisons for the $N_e = N_v = 1$ case, 16 500 for the case of $N_e = 6$, and 2750 comparisons for $N_e = N_v = 6$ case. For the imposter case, we have 3 297 250, 1 798 500, and 299 750 comparisons, respectively.

The training-evaluation-set split is performed five times, while for each of these splits the enrolment-verification split is also performed five times. From each enrolment-verification split, we estimate the operating point T_{tar} at the target FNMR β_{tar} and the corresponding FMR α_{tar}. 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.

D. Results

First we determine the Gaussian channel capacity $C_G[j]$ of component j of the feature vector obtained after applying the PCA/LDA transformation with use of (1) and estimating the feature quality $\sigma_b[j]/\sigma_w[j]$. We consider both on the training set and the evaluation set. The capacities for the 218 components are illustrated in Fig. 16 for both the directional field DF and the Gabor filters GF features indicating that the capacity is not equal for each component. Note that the capacity is greater for the transformed training set than the transformed evaluation set, because the PCA/LDA transformation matrix is determined on the same set and can thus be perfectly trained and the training and evaluation sets are disjunct. This perfect training is also confirmed by the fact that the last components of the training set have a capacity $C_G[j]$ close or equal to zero, while they are larger than zero for the evaluation set. By assuming all components to be independent, we observe that the DF feature has an input capacity $C_{\text{in}} = 162$ bits on the training set and $C_{\text{in}} = 186$ bits on the evaluation set, while $C_{\text{in}} = 193$ and $C_{\text{in}} = 207$ bits for the GF features. Because the capacities are not equally divided, we already know that the achieved performance and the maximum key size will be suboptimal.

With the known capacity of each component, we can thus compare the maximum key size k_c^* and the log of the FMR at the target FNMR $-\log_2(\alpha_{\text{tar}})$ from the theoretical performance and the experimental performance. The theoretical performance is obtained using the analytical framework. These results are shown in Fig. 17 for different numbers of enrolment N_e and verification N_v samples for both the DF and GF features. Note that due to the limited number of imposter comparisons, it is not possible to obtain a α_{tar} smaller than $1/299\,750 = 3.3 \times 10^{-6}$ except zero for the experimental case with $N_e = N_v = 6$. From the results we observe four effects. First of all, both the experimental and theoretical results confirm the finding in Section IV that the components with a smaller capacity have a greater improvement when more samples are used. For the single enrolment and verification sample case, the experimental results even show that the last components with a much smaller capacity deteriorates the performance and, therefore, also the maximum key size. However, an improvement is observed when we increase the number of enrolment samples to $N_e = 6$, and a greater improvement is observed for when we also increase the number of verification samples to $N_v = 6$. Second, the results also indicate that the estimated k_c^* and $-\log_2(\alpha_{\text{tar}})$ are much greater for the theoretical case than for the experimental one. The results in Fig. 17(e) and (f) portray the significant difference between the obtained relative operating point T_{tar}/n_c between the theoretical and experimental cases. This clearly indicates that the FNMR curve is not correctly estimated, because the target FNMR for the experimental case is at a larger relative operating point than for the theoretical case. As discussed in Kelkboom *et al.* (2010) [31], estimation errors are introduced by deviations from the underlying assumptions such as the Gaussian distribution, an equal and independent within-class for each subject, and independent feature components. They proposed a modified analytical framework for relaxing these assumption; however, this approach is out of the scope of this

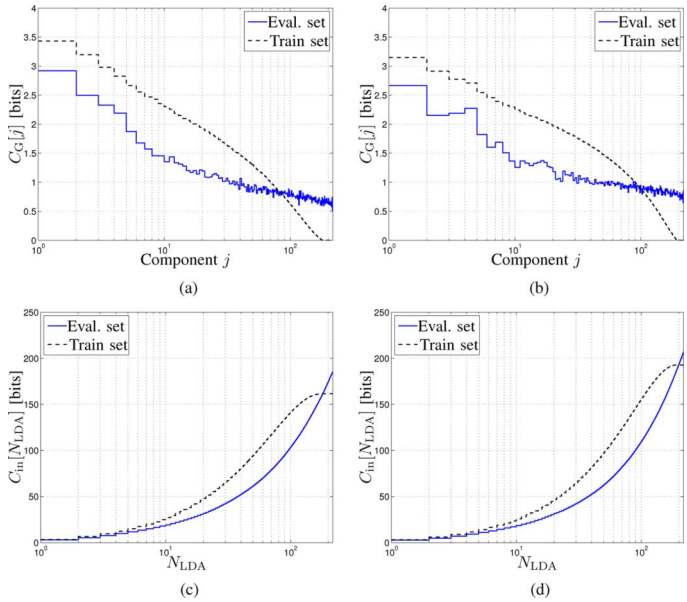

Fig. 16. For both the GF and DF features, (a) and (b) illustrate the Gaussian channel capacity $C_G[j]$ of each component from the training set and evaluation set, and (c) and (d) the input capacity C_{in} taken as the cumulative sum of $C_G[j]$ of all N_{LDA} components, namely $C_{\text{in}} = \sum_{m=1}^{N_{\text{LDA}}} C_G[m]$. (a) DF: Gaussian capacity. (b) GF: Gaussian capacity. (c) DF: Input capacity, (d) GF: Input capacity.

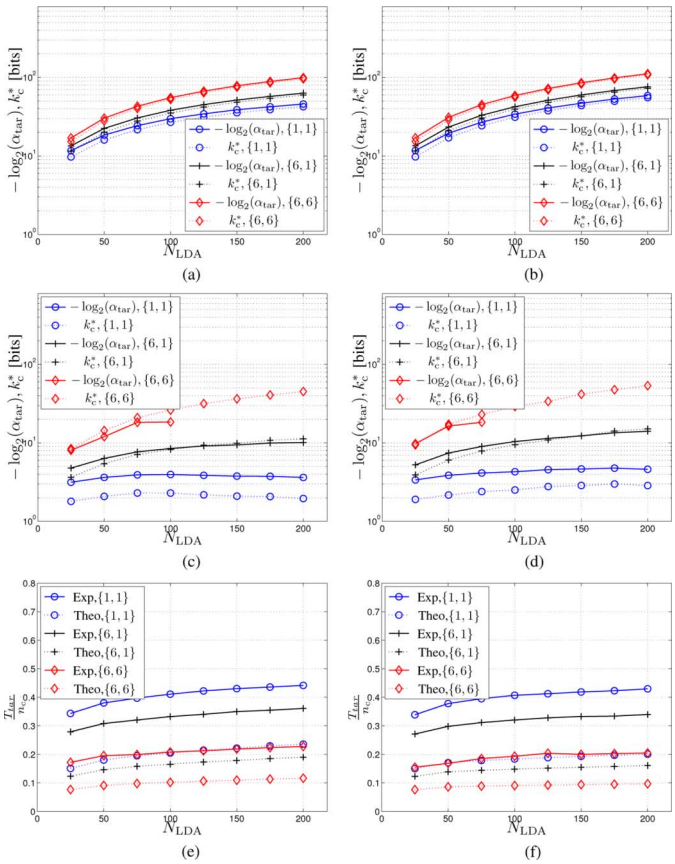

Fig. 17. Maximum key size k_c^*, the log of the FMR at the target FNMR $-\log_2(\alpha_{\mathrm{tar}})$, and the relative operating point T_{tar}/n_c as a function of the LDA dimension N_{LDA} at different N_e and N_v settings indicated as $\{N_e, N_v\}$ in the legend. Subfigures (a) and (b) are for the theoretical case for the DF and GF features, respectively; similarly subfigures (c) and (d) are for the experimental case, and (e) and (f) are the theoretical and experimental case combined. (a) DF: Theoretical. (b) GF: Theoretical. (c) DF: Experimental. (d) GF: Experimental. (e) DF: T_{tar}/n_c. (f) GF: T_{tar}/n_c.

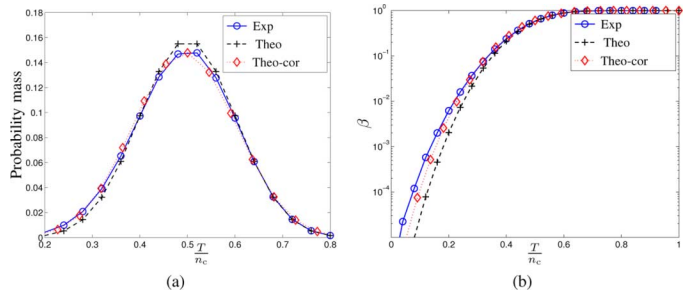

Fig. 18. (a) Hamming distance pmf at imposter comparisons from the experimental case ("Exp"), from the theoretical case ("Theo"), and the corrected theoretical case ("Theo-cor") where the experimental data is fitted with a binomial distribution with dimension $\hat{n}_c$ and bit-error probability $\hat{P}_e^{\mathrm{im}}$. Furthermore, (b) shows the corresponding FMR β curve for the three cases in (a).

work. Third, we observe that the relative difference between the theoretical and experimental results is greater for the $N_e = N_v = 1$ case and decreases when increasing N_e and N_v. It has also been shown in Kelkboom *et al.* (2010) [31] that an increase in the number of samples results in a better Gaussian approximation of the feature distributions. Hence, a better Gaussian approximation due to the increase of the number of samples may be the cause behind the improvement of the estimation error. The forth and last difference we observed between the theoretical and experimental results in Fig. 17(a)–(d) is the relationship between $-\log_2(\alpha_{\mathrm{tar}})$ and the maximum key size k_c^*. We have shown in Section III-D4 that they are related to each other, namely $k_c^* < -\log_2(\alpha_{\mathrm{tar}})$, and this relationship is confirmed by the theoretical case in Fig. 17(a) and (b). However, the results in Fig. 17(c) and (d) show that for the experimental cases $-\log_2(\alpha_{\mathrm{tar}})$ are not always larger than k_c^*. These deviations are caused by the estimation errors of the FMR curve, leading to an optimistically smaller FMR and thus a larger $-\log_2(\alpha_{\mathrm{tar}})$ at the same operating point.

As discussed in Kelkboom *et al.* (2010) [31], having dependent feature components has a great influence on the FMR curve estimation. Due to the dependencies, the variance of the relative Hamming distance (the Hamming distance relative to n_c) distribution at imposter comparisons is larger than the expected variance of the binomial distribution. Because the variance of the relative Hamming distance is inverse proportional to the dimension, namely $\sigma^2 = p(1-p)/N$, the intrinsic dimension decreases when there is a stronger dependency. Similar to the work of Daugman (2003) [33], we will estimate the intrinsic dimension by fitting the imposter Hamming distance distribution with a binomial distribution with a dimension smaller than n_c and a bit-error probability smaller than $1/2$. Given the relative Hamming distances at each comparison, we estimate its variance $\hat{\sigma}^2_{\mathrm{im}}$ and mean $\hat{\mu}_{\mathrm{im}}$, from which we can estimate the new binomial dimensions $\hat{n}_c$ with bit-error probability $\hat{P}_e^{\mathrm{im}}$ as

$$\begin{aligned} \hat{P}_e^{\mathrm{im}} &= \hat{\mu}_{\mathrm{im}} \\ \hat{n}_c &= \left\lfloor \frac{\hat{P}_e^{\mathrm{im}}\left(1-\hat{P}_e^{\mathrm{im}}\right)}{\hat{\sigma}^2_{\mathrm{im}}} \right\rfloor . \end{aligned} \tag{25}$$

An example of this approximation is shown in Fig. 18(a) for the pmf of the relative Hamming distances and in Fig. 18(b) for the FNMR curve. The experimentally obtained curves are indicated with "Exp," while the original theoretical model curve is indicated with "Theo," and its corrected version for the intrinsic dimension by "Theo-cor." Note that we multiplied the pmf for the "Theo-cor" case with $\hat{n}_c/n_c$ in order for its area under the curve to be as large for the other two cases for a fair comparison. From these results, we observe that the corrected pmf "Theo-cor" approximates the experimentally obtained results much better. However, the estimation errors are now mainly at the tails of the pmf and thus at the smallest values of the FNMR.

The estimated bit-error probability $\hat{P}_e^{\mathrm{im}}$ and the intrinsic dimension $\hat{n}_c$ at imposter comparisons for different LDA dimensions N_{LDA} and number of enrolment N_e or verification N_v samples are depicted in Fig. 19 for both the DF and GF features. Instead of the actual estimated intrinsic dimension $\hat{n}_c$, we show the ratio $\hat{n}_c/n_c$. The results from Fig. 19(a) and (b) indicate that when adding more components by increasing N_{LDA}, the relative intrinsic dimension decreases while the bit-error probability converges towards $1/2$. Note that the relative intrinsic dimension also decreases when more samples are used, hence taking the average of N_e or N_v samples increases the dependencies between the bit errors at imposter comparisons.

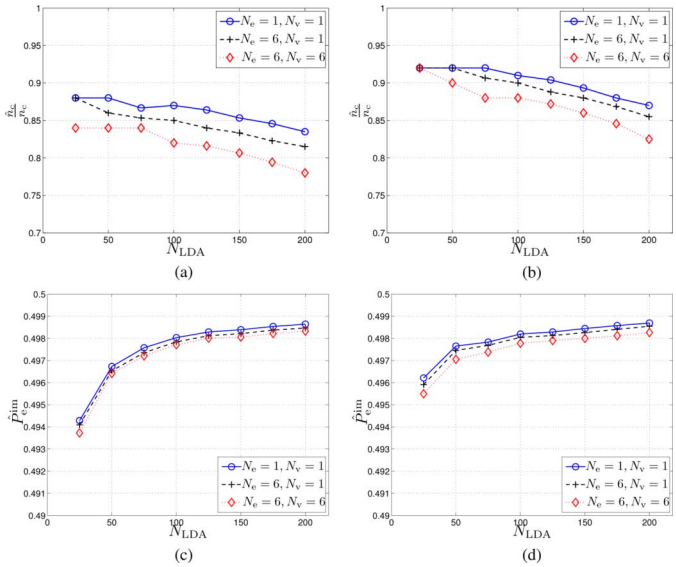

Fig. 19. (a) and (b) Estimated relative intrinsic degrees of freedom or dimension $\hat{n}_c/n_c$ of the Hamming distance pmf at imposter comparisons for different LDA settings N_{LDA} and number of enrolment N_e or verification N_v samples, and (c) and (d) the corresponding estimate bit-error probability $\hat{P}_e^{\text{im}}$ for both the DF and GF features. (a) DF; (b) GF; (c) DF; (d) GF.

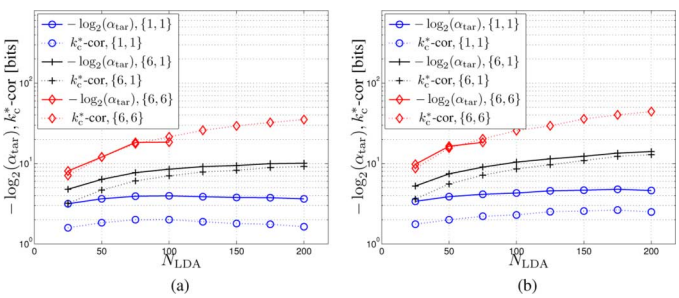

Fig. 20. Corrected maximum key size k_c^*-cor, the log of the FMR at the target FNMR $-\log_2(\alpha_{\text{tar}})$ as a function of the LDA dimension N_{LDA} at different N_e and N_v settings for the DF and GF features. The number of samples is indicated in the legend with $\{N_e, N_v\}$. (a) DF; (b) GF.

The maximum key size estimation can be improved by incorporating the intrinsic dimension as

$$k_c^*\text{-cor} \stackrel{\text{def}}{=} \hat{n}_c C\left(\frac{T_{\text{tar}}}{n_c}\right) = \frac{\hat{n}_c}{n_c} k_c^* \tag{26}$$

where the corrected maximum key size k_c^*-cor is the relative intrinsic dimension $\hat{n}_c/n_c$ times the original maximum key size k_c^*. The improved results are illustrated in Fig. 20. Now also for the $N_e = 6$, $N_v = 1$ case, the corrected maximum key size is always smaller than $-\log_2(\alpha_{\text{tar}})$. The estimation has also improved for the $N_e = 6$, $N_v = 6$ case; however, there are still some deviations, which may be caused by the limited database.

VI. Discussion and Conclusion

The FCS is a well-known template protection scheme in the literature and is based on a key-binding and key-release mechanism, where the entropy of the key is indicative for the amount of privacy and security. Considering the key to consist out of independent and uniform bits, its entropy is then mainly determined by its size. We have analytically determined the classification performance and the maximum key size of the FCS given a Gaussian modeled biometric source, a single bit extraction quantization scheme, the number of enrolment and verification samples, an ECC with a decoding capability at Shannon's bound, and the target FNMR. Furthermore, we modeled the FCS as a binary symmetric channel with its corresponding bit-error probability.

We have analytically derived the bit-error probability as a function of the feature quality denoted by the ratio of the between-class and within-class variance, and the number of enrolment and verification samples. We have shown that having infinite enrolment samples with X verification samples approximates the performance when both are equal to $2X$, if the feature quality is large enough.

We estimated the maximum key size at the target FNMR assuming an ideal binary ECC that corrects up to t_c random bit errors of equal bit-error probability and its decoding capability at Shannon's bound. First, we showed that the FNMR is close to 50% when the operating point of the ECC is set at the point stipulated by Shannon's theory. The high FNMR is caused by the fact that the size of the codeword in the biometric system is not large enough as required by Shannon's theorem. We proposed two other operating points, namely the analytical operating point at the EER and the operating point given the target FNMR. The key size at the EER is always smaller than at the operating point from Shannon's theory. At the EER point, more bits have to be corrected due the smaller FNMR requirement; consequently, the operating point is larger leading to a smaller key size. The operating point at the target FNMR is a compromise between the two aforementioned cases, and leads to the maximum key size at the desired FNMR. We also discussed the relationship between the maximum key size and the target FMR at the target FNMR. We showed that the upperbound from literature $-\log_2(\text{FMR})$ is larger than the maximum key size when errors have to be corrected. The difference increases when using larger codewords, and could be around 3 bits when the codeword is 127 bits long.

We studied the effect of the capacity of the Gaussian biometric source, the number of biometric samples, and the target FNMR on the FMR and maximum key size. There are two main scenarios that we investigated, namely the scenarios where the components are 1) independent or 2) dependent.

For the first scenario, we found the following results for the cases where the input capacity is 40 bits and 80 bits. Doubling the input capacity roughly tripled the key size at a target FNMR of 2.5%, while doubling the target FNMR from 2.5% to 5% on average added around 1 bit. Increasing the number of enrolment samples from one to six added 2.9 bits. With six enrolment samples and increasing the number of verification samples from one to two added 7.6 bits, while increasing from two to six samples added 20.8 bits. Thus, if the subjects of the biometric system have no issue with a less convenient system where the target FNMR has increased or more biometric samples have to be acquired, we could create a protected template that is more

difficult to break by an adversary. Doubling the target FNMR also doubles the search space of the key. Moreover, an increase from 1 to 6 enrolment and verification samples increased the key size by almost 32 bits. Supplying six samples during enrolment seems acceptable, because it only needs to be done once. Although capturing six samples during verification may be considered inconvenient, it still gives a good insight into what can be achieved by such a system. In both the first and second case, we observed that the maximum key size significantly reduces if the target FNMR is smaller than 5%.

In the second scenario, we showed that adding fully dependent bits does not improve the system performance, but artificially increases the maximum key size. The discrepancy between the FMR and the maximum key size increases when more components are dependent.

We presented experimental results on the MCYT fingerprint database using two feature extraction algorithms, namely one based on directional field and one on Gabor filters. For both algorithms, we observed that the difference between the FMR and the maximum key size changed when increasing the number of components. The difference can be made more constant when the dependency between feature components is taken into account.

In the Introduction, Table I presents the reported key size and the system performance from similar template protection schemes from the literature. The table shows the differences between the reported FMR and key size. From the results presented in this work, we conjecture that these discrepancies may be primarily caused by the dependencies between feature components. Hence, both the reported key size and FMR have to be taken into account when analyzing the actual privacy protection and security of a template protection system.

The main contribution of this paper is the analytical relationship between the system performance and the maximum key size given the system parameters. With the analytical framework and experimental results we showed that dependencies between feature components lead to a difference between the reported FMR and key size. Furthermore, we revealed a trade-off between the convenience of the biometric system, determined by the target FNMR and the number of samples to be acquired, and the maximum key size. Essentially, if desired, a larger key size can be achieved by sacrificing some convenience.

APPENDIX

In order to find an analytical expression of the EER operating point T_{EER}, we approximate the binomial density used for modeling the pmf of the Hamming distance ϵ by a Gaussian density as proposed by the Moivre–Laplace theorem [41]. Hence, instead of (11) we use

$$\begin{aligned} P_G(\epsilon; N, p) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\epsilon-\mu/\sigma\sqrt{2}\right)^2} \\ &= \frac{1}{\sqrt{n_c(1-p)(p)}\sqrt{2\pi}} \\ &\quad \times e^{-\left(\epsilon-n_c p/\sqrt{2n_c(1-p)p}\right)^2} \end{aligned} \tag{27}$$

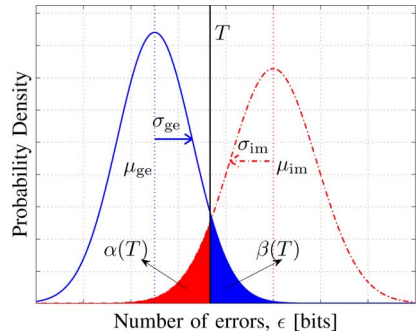

Fig. 21. Gaussian approximation of the pmf of the number of errors ϵ at genuine (the solid blue curve) and imposter (the dashed–dotted red curve) comparisons from Fig. 7.

where we use the mean and the variance of the binomial density, namely the mean $\mu = n_c p$ and standard deviation $\sigma = \sqrt{n_c(1-p)p}$. The resulting approximated probability density as a function of the Hamming distance ϵ is shown in Fig. 21.

Thus given the operating point T, the FNMR from (14) can be rewritten as

$$\begin{aligned} \beta(T) &= \int_{i=T}^{\infty} P_G\left(i; n_c, P_e^{\mathrm{ge}}\right) di \\ &= \int_{i=T}^{\infty} \frac{1}{\sigma_{\mathrm{ge}}\sqrt{2\pi}} e^{-\left(i-\mu_{\mathrm{ge}}/\sigma_{\mathrm{ge}}\sqrt{2}\right)^2} di \end{aligned} \tag{28}$$

with $\mu_{\mathrm{ge}} = n_c P_e^{\mathrm{ge}}$ and $\sigma_{\mathrm{ge}} = \sqrt{n_c(1-P_e^{\mathrm{ge}})P_e^{\mathrm{ge}}}$. By applying the following change of variable $\tau = i - \mu_{\mathrm{ge}}/\sigma_{\mathrm{ge}}$ with $di = \sigma_{\mathrm{ge}} d\tau$, we obtain

$$\beta(T) = \int_{\tau=z_{\mathrm{ge}}(T)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-1/2\tau^2} d\tau \tag{29}$$

where we have the genuine z-score $z_{\mathrm{ge}}(T) = T - \mu_{\mathrm{ge}}/\sigma_{\mathrm{ge}}$ that fully determines the FNMR. Similarly, for the FMR, we have

$$\begin{aligned} \alpha(T) &= \int_{i=-\infty}^{T} P_G\left(i; n_c, P_e^{\mathrm{im}}\right) di \\ &= \int_{\tau=-\infty}^{z_{\mathrm{im}}} \frac{1}{\sqrt{2\pi}} e^{-1/2\tau^2} d\tau \\ &= \int_{\tau=-z_{\mathrm{im}}(T)}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-1/2\tau^2} d\tau \end{aligned} \tag{30}$$

where we applied the same variable change, defined the imposter z-score $z_{\mathrm{im}}(T) = T - \mu_{\mathrm{im}}/\sigma_{\mathrm{im}}$, and used the property that the integral is symmetric. Because $P_e^{\mathrm{im}} = 1/2$, we have $\mu_{\mathrm{im}} = n_c/2$ and $\sigma_{\mathrm{im}} = \sqrt{n_c}/2$ Being at the EER operating point T_{EER} implies that $\alpha(T_{\mathrm{EER}}) = \beta(T_{\mathrm{EER}})$. Hence,

(29) and (30) have to be equal. Both equations are equal when $z_{\rm ge}(T_{\rm EER}) = -z_{\rm im}(T_{\rm EER})$, thus $T_{\rm EER}$ becomes

$$\begin{aligned} z_{\rm ge}(T_{\rm EER}) &= -\, z_{\rm im}(T_{\rm EER}) \\ \frac{T_{\rm EER}-\mu_{\rm ge}}{\sigma_{\rm ge}} &= -\,\frac{T_{\rm EER}-\mu_{\rm im}}{\sigma_{\rm im}} \\ T_{\rm EER} &= \frac{\mu_{\rm im}\sigma_{\rm ge}+\mu_{\rm ge}\sigma_{\rm im}}{\sigma_{\rm im}+\sigma_{\rm ge}}. \end{aligned} \tag{31}$$

Substituting the genuine parameters $\mu_{\rm ge} = n_c P_e^{\rm ge}$ and $\sigma_{\rm ge} = \sqrt{n_c(1-P_e^{\rm ge})P_e^{\rm ge}}$, and the imposter parameters $\mu_{\rm im} = n_c/2$ and $\sigma_{\rm im} = \sqrt{n_c}/2$, we obtain

$$T_{\rm EER} = \frac{n_c\left(\sqrt{P_e^{\rm ge}(1-P_e^{\rm ge})}+P_e^{\rm ge}\right)}{2\sqrt{P_e^{\rm ge}(1-P_e^{\rm ge})}+1}$$

or

$$\frac{T_{\rm EER}}{n_c} = \frac{\sqrt{P_e^{\rm ge}(1-P_e^{\rm ge})}+P_e^{\rm ge}}{2\sqrt{P_e^{\rm ge}(1-P_e^{\rm ge})}+1}. \tag{32}$$

Note that the relative operating point $T_{\rm EER}/n_c$ and thus the BSC channel capacity at the EER operating point $C(T_{\rm EER}/n_c)$ is fully determined by $P_e^{\rm ge}$

REFERENCES

[1] Identity Cards Act 2006 [Online]. Available: http://www.opsi.gov.uk/acts/acts2006/ukpga_20060015_en_1

[2] International Civil Aviation Organization ICAO [Online]. Available: http://www.icao.int/cgi/goto_m_atb.pl?icao/en/atb/fal/mrtd/MRTD_Rpt_V1N1_2006.pdf

[3] A. K. Jain, K. Nandakumar, and A. Nagar, "Biometric template security," *EURASIP J. Adv. Signal Process.*, vol. 2008, pp. 1–17, 2008.

[4] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM Syst. J.*, vol. 40, no. 3, pp. 614–634, 2001.

[5] B. Yang, C. Busch, P. Bours, and D. Gafurov, "Robust minutiae hash for fingerprint template protection," *Proc. SPIE*, vol. 7541, no. 75410R, 2010.

[6] A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in *Proc. 6th ACM Conf. Computer and Communications Security*, Nov. 1999, pp. 28–36.

[7] J.-P. Linnartz and P. Tuyls, "New shielding functions to enhance privacy and prevent misuse of biometric templates," in *Proc. 4th Int. Conf. AVBPA*, 2003, pp. 393–402.

[8] A. Juels and M. Sudan, "A fuzzy vault scheme," *Designs, Codes, Cryptography*, vol. 38, no. 2, pp. 237–257, Feb. 2006.

[9] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," *SIAM J. Comput.*, vol. 38, no. 1, pp. 97–139, 2008.

[10] FCD 24745—Information Technology—Security Techniques—Biometric Template Protection ISO/IEC JTC1 SC27.

[11] U. Korte and R. Plaga, "Cryptographic protection of biometric templates: Chance, challenges and applications," *BIOSIG 2007: Biometrics and Electronic Signatures*, pp. 33–45, 2007.

[12] I. R. Buhan, J. M. Doumen, P. H. Hartel, and R. N. J. Veldhuis, "Fuzzy extractors for continuous distributions," in *Proc. 2nd ACM Symp. Information, Computer and Communications Security (ASIACCS)*, Singapore, Mar. 2007, pp. 353–355.

[13] J. Kittler, J. Matas, K. Jonsson, and M. U. R. Sánchez, "Combining evidence in personal identity verification systems," *Pattern Recognit. Lett.*, vol. 18, pp. 845–852, 1997.

[14] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn, "Using multi-instance enrollment to improve performance of 3D face recognition," *Comput. Vis. Image Understand.*, vol. 112, no. 2, pp. 114–125, Nov. 2008.

[15] E. J. C. Kelkboom, G. G. Molina, T. A. M. Kevenaar, R. N. J. Veldhuis, and W. Jonker, "Binary biometrics: An analytic framework to estimate the bit error probability under gaussian assumption," in *Proc. 2nd IEEE Int. Conf. Biometrics: Theory, Applications and Systems (BTAS '08)*, Sep. 2008, pp. 1–6.

[16] J. Bringer, H. Chabanne, G. Cohen, B. Kindarji, and G. Zemor, "Theoretical and practical boundaries of binary secure sketches," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 4, pp. 673–683, Dec. 2008.

[17] E.-C. Chang and S. Roy, "Robust extraction of secret bits from minutiae," in *Proc. Int. Conf. Biometrics*, Seoul, South Korea, Aug. 2007, pp. 750–759.

[18] Y. Sutcu, S. Rane, J. S. Yedidia, S. C. Draper, and A. Vetro, "Feature extraction for a slepian-wolf biometric system using LDPC codes," in *Proc. IEEE Int. Symp. Information Theory, 2008 (ISIT 2008)*, 2008, pp. 2297–2301.

[19] E. J. C. Kelkboom, B. Gökberk, T. A. M. Kevenaar, A. H. M. Akkermans, and M. van de Veen, "3D face: Biometric template protection for 3D face recognition," in *Proc. Int. Conf. Biometrics*, Seoul, South Korea, Aug. 2007, pp. 566–573.

[20] T. A. M. Kevenaar, G.-J. Schrijen, A. H. M. Akkermans, M. van de Veen, and F. Zuo, "Face recognition with renewable and privacy preserving binary templates," in *Proc. 4th IEEE Workshop on AutoID*, Buffalo, NY, Oct. 2005, pp. 21–26.

[21] P. Tuyls, A. H. M. Akkermans, T. A. M. Kevenaar, G.-J. Schrijnen, A. M. Bazen, and R. N. J. Veldhuis, "Practical biometric authentication with template protection," in *Proc. 5th Int. Conf., AVBPA*, Rye Brook, NY, Jul. 2005.

[22] X. Zhou, "Template protection and its implementation in 3D face recognition systems," in *Proc. SPIE 07, Biometric Technology for Human Identification IV*, 2007, vol. 6539, no. 65390L.

[23] F. Hao, R. Anderson, and J. Daugman, "Combining crypto with biometrics effectively," *IEEE Trans. Comput.*, vol. 55, no. 9, pp. 1081–1088, Sep. 2006.

[24] T. C. Clancy, N. Kiyavash, and D. J. Lin, "Secure smartcard-based fingerprint authentication," in *Proc. 2003 ACM SIGMM Workshop Biometrics Methods and Application (WBMA)*, 2003, pp. 45–52.

[25] K. Nandakumar, A. Nagar, and A. K. Jain, "Hardening fingerprint fuzzy vault using password," in *Proc. 2nd Int. Conf. Biometrics*, Seoul, South Korea, Aug. 2007, pp. 927–937.

[26] A. Arakala, J. Jeffers, and K. J. Horadam, "Fuzzy extractors for minutiae-based fingerprint authentication," in *Proc. Int. Conf. Biometrics*, Seoul, South Korea, 2007, pp. 760–769.

[27] T. Ignatenko and F. M. J. Willems, "Biometric systems: Privacy and secrecy aspects," *IEEE Trans. Inf. Forensics Security*, vol. 4, no. 4, pp. 956–973, Dec. 2009.

[28] T. Ignatenko and F. M. J. Willems, "Privacy leakage in biometric secrecy systems," in *Proc. 46th Ann. Allerton Conf. Communication, Control, and Computing*, 2008, pp. 850–857.

[29] L. Lai, S.-W. Ho, and H. V. Poor, "Privacy-security tradeoffs in biometric security systems," in *Proc. 46th Ann. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Sep. 2008.

[30] F. M. J. Willems and T. Ignatenko, "Quantization effects in biometric systems," in *Proc. Information Theory and Applications Workshop*, San Diego, CA, Feb. 2009, pp. 372–379.

[31] E. J. C. Kelkboom, G. G. Molina, J. Breebaart, R. N. J. Veldhuis, T. A. M. Kevenaar, and W. Jonker, "Binary biometrics: An analytic framework to estimate the performance curves under gaussian assumption," *Special Issue on Advances in Biometrics: Theory, Applications and Systems, IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 40, no. 3, pp. 555–571, May 2010.

[32] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ: Wiley, 1991.

[33] J. Daugman, "The importance of being random: Statistical principles of iris recognition," *Pattern Recognit.*, vol. 36, no. 2, pp. 279–291, 2003.

[34] C. Shannon, "The zero error capacity of a noisy channel," *Trans. Inf. Theory*, vol. 2, no. 3, pp. 8–19, Sep. 1956.

[35] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North-Holland, 1977.

[36] E. Kelkboom, "On the Performance of Helper Data Template Protection Schemes," Ph.D. thesis, University of Twente, Enschede, The Netherlands, 2010 [Online]. Available: http://eprints.eemcs.utwente.nl/18568/

[37] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, 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*, Dec. 2003, pp. 395–401.

[38] M. van der Veen, A. Bazen, T. Ignatenko, and T. Kalker, "Reference point detection for improved fingerprint matching," in *Proc. SPIE*, 2006, pp. 60720G.1–60720G.9.

[39] S. H. Gerez and A. M. Bazen, "Systematic methods for the computation of the directional fields and singular points of fingerprints," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 905–919, Jul. 2002.

[40] A. M. Bazen and R. N. J. Veldhuis, "Likelihood-ratio-based biometric verification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 86–94, Jan. 2004.

[41] A. de Moivre, *The Doctrine of Chances, or, a Method of Calculating the Probabilities of Events in Play*, 3rd ed ed. New York: Chelsea, 2000, reprint of 1756 3rd ed. Original ed. published 1716.

Emile J. C. Kelkboom was born in Oranjestad, Aruba, in 1980. He received the M.Sc. degree in electrical engineering from Delft University of Technology, The Netherlands, in June 2004. Since August 2006, he has been working toward the Ph.D. degree at Philips Research and the Department of Electrical Engineering Mathematics and Computer Science, University of Twente, The Netherlands. His focus is on safeguarding the privacy of the biometric information of subjects within biometric systems, namely the field of template protection.

From October 2004 to July 2006, he worked as an Application Engineer on CD, DVD, and Blu-ray drives within the Storage Engines Department of Philips Semiconductors. His research interests include biometrics, pattern recognition, signal processing, and security.

Mr. Kelkboom won the European Biometrics Forum (EBF) Research Award among Ph.D. students in Europe in 2009.

Jeroen Breebaart (M'07) received the M.Sc. degree in biomedical engineering from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 1997, and the Ph.D. degree in auditory psychophysics from the same university in 2001.

From 2001 to 2007, he was with the Digital Signal Processing Group at Philips Research, conducting research in the areas of spatial hearing, parametric audio coding, automatic audio content analysis, and audio effects processing. Since 2007, he has been the leader of the biometrics cluster of the Information and System Security Group at Philips Research, expanding his research scope toward secure and convenient identification.

Dr. Breebaart is a member of the AES. He contributed to the development of audio coding algorithms as recently standardized in MPEG and 3GPP such as HEAAC, MPEG Surround, and the upcoming standard on spatial audio object coding. He also actively participates in the ISO/IEC IT security techniques standardization committee and is significantly involved in several EU-funded projects. He published more than 50 papers at international conferences and journals.

Ileana Buhan received the Ph.D. degree from the University of Twente in 2008.

She has conducted research into security applications involving noisy data and secure spontaneous interaction. While at Philips Research in the Netherlands, she worked on developing techniques for the protection of biometric data. She is now Security Evaluation Manager at Riscure in the Netherlands.

In 2008, Dr. Buhan received the EBF European Biometric Research Industry Award for her work on combining secure spontaneous interaction with biometrics.

Raymond N. J. Veldhuis received the engineer degree in electrical engineering from the University of Twente, Enschede, The Netherlands, in 1981, and the Ph.D. degree from Nijmegen University, Nijmegen, The Netherlands, in 1988. His dissertation was titled "Adaptive restoration of lost samples in discrete-time signals and digital images."

From 1982 until 1992, he worked as a Researcher at Philips Research Laboratories, Eindhoven, The Netherlands, in various areas of digital signal processing, such as audio and video signal restoration and audio source coding. From 1992 until 2001, he worked at the Institute of Perception Research (IPO), Eindhoven, in speech signal processing and speech synthesis. From 1998 until 2001, he was program manager of the Spoken Language Interfaces research program. He is now an Associate Professor at the University of Twente, working in the fields of biometrics and signal processing. His expertise involves digital signal processing for audio, images and speech, statistical pattern recognition, and biometrics. He has been active in the development of MPEG standards for audio source coding.