GREYCHashing: Combining Biometrics and Secret for Enhancing the Security of Protected Templates

Kevin Atighehchi, Loubna Ghammam, Morgan Barbier, Christophe Rosenberger

GREYCHashing: Combining Biometrics and Secret for Enhancing the Security of Protected Templates

Kevin Atighehchi, Loubna Ghammam, Morgan Barbier, Christophe Rosenberger

Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France

Abstract

Template protection is a crucial issue in biometrics. Many algorithms have been proposed in the literature among secure computing approaches, crypto-biometric algorithm and feature transformation schemes. The BioHashing algorithm belongs to this last category and has very interesting properties. Among them, we can cite its genericity since it could be applied on any biometric modality, the possible cancelability of the generated BioCode and its efficiency when the secret is not stolen by an impostor. Its main drawback is its weakness face to a combined attack (false acceptance with the stolen secret scenario). In this paper, we propose a transformation-based biometric template protection scheme as an improvement of the BioHashing algorithm where the projection matrix is generated by combining the secret and the biometric data. Experimental results on three biometric modalities, namely digital fingerprint, finger knuckle print and hands vein images, show the benefits of the proposed method face to attacks while keeping a good efficiency.

Keywords: Template protection, biometric authentication, attack, performance evaluation.

Email addresses: kevin.atighehchi@unicaen.fr (Kevin Atighehchi), loubna.ghammam@unicaen.fr (Loubna Ghammam), morgan.barbier@ensicaen.fr (Morgan Barbier), christophe.rosenberger@ensicaen.fr (Christophe Rosenberger)

Contents

1. Introduction

Biometrics is an emerging technology for authentication applications. Many biometric modalities are well known and used (such as fingerprints), the design of intelligent sensors is advanced (liveness detection) and algorithms provide very good results. Privacy issues concerning this particular personal information still limit its operational use. The General Data Protection Regulation (GDPR) 2016/679 is a regulation in EU law on data protection and privacy for all individuals within the European Union. Biometric data is of course considered as personal and sensitive data. In Europe, for example, the central storage of biometric data is forbidden or limited to a small amount of users. In order to solve this problem, new biometric systems have been proposed in the last decade based on the "privacy by design" paradigm. These biometric template protection schemes have as objective to guarantee the security and privacy of users to face attacks such as identity theft (e-government applications, border control, *etc.*) [1].

Three main approaches can be distinguished when dealing with template protection in biometrics. First, biometric crypto-systems or secure sketches, such as those presented in [2, 3, 4], resort to cryptography. Second, secure computing methods aim at computing the comparison of two biometric templates by an untrusted party [5, 6]. Last, we find feature transformations approaches for template protection. The BioHashing algorithm is one of the most popular technique and is based on biometric data salting. It has been developed for different biometric modalities such as those presented in [7, 8, 9].

These last systems are called cancelable since the result generated from a biometric template, namely BioCode, can be revoked in case of interception or loss. This BioCode cannot be used as a cryptographic key as the generated BioCode is not exactly the same for each biometric capture. These particular biometric systems must of course address classical issues such as a high level of performance (*i.e.*, minimizing the Equal Error Rate (EER) or Area Under the Curve (AUC) value of the system) but also new constraints concerning privacy. In the literature, many papers [10, 11, 8, 12] have been published dealing with the definition of new schemes for the protection of biometric templates.

Most of such protection schemes lack of robustness considering the stolen token scenario. In this case, an attacker knowing the secret of the protection scheme, has a big advantage to impersonate a user [13, 14]. This is due to the fact that the used projection matrix is computed only given the secret.

Our Contributions. A major drawback of keyed projection-based transformations like BioHashing is that the generated BioCode depends on the key/token used, and less on the input features. In this work, we propose a new transformation for the protection of biometric data. Its main benefit is that the

used projection matrix is not only related to a secret but also embeds information computed from the biometric data. The great advantage of this proposed method is that it limits some attacks. Second, we analyze the behavior of this new transformation in comparison with two other methods by using a recent analysis methodology of such schemes [15]. Emphasis will be placed on both security and privacy aspects, *i.e.* by limiting attacks enabling an adversary to get falsely authenticated and by preventing the recovery of a biometric feature vector from its corresponding BioCode.

Organization. The paper is organized as follows. Section 2 gives the background on feature extraction and template protection schemes. Section 3 is dedicated to a literature review on template protection schemes based on a transformation. More specifically, in this section we recall the details of BioHashing and BioPhasor algorithms. Then, a new transformation algorithm is described in Section 4. Section 5 illustrates the benefits of the proposed method through experimental results to be compared with BioHashing and BioPhasor. Finally, in Section 6, we conclude and give some perspectives.

This invited article supports and improves the results of the original paper entitled "Enhancing the Security of Transformation Based Biometric Template Protection Schemes" [16].

2. Background

The general principle of template protection schemes based on a transformation is depicted in Figure 1. These schemes consist in generating a binary vector called BioCode given a biometric template and a secret.

Feature extraction. In this paragraph, we briefly recall the concept of Gabor filter, proposed by Dennis Gabor in [17], and used for texture analysis and extraction. There are many feature extraction methods described in the literature and it is mentionned in [14] and [18] that Gabor filter is considered as one of the best methods. Gabor features have been used in several image analysis applications. For example, texture classification and segmentation [19], image recognition [20, 21], image registration and motion tracking [22].
Recall that the 2-D Gabor filter can be represented as a complex sinusoidal signal modulated by a Gaussian kernel function as given in the following equation [23]:

$$\Psi_{f,\theta} = \exp\left[-\frac{1}{2}\left(\frac{x_\theta^2}{\sigma_x^2} + \frac{y_\theta^2}{\sigma_y^2}\right)\right] \times \exp(2\pi f \theta_n)$$

with,

- $x_\theta = x\sin\theta_n + y\cos\theta_n$ and $y_\theta = -x\sin\theta_n + y\cos\theta_n$.

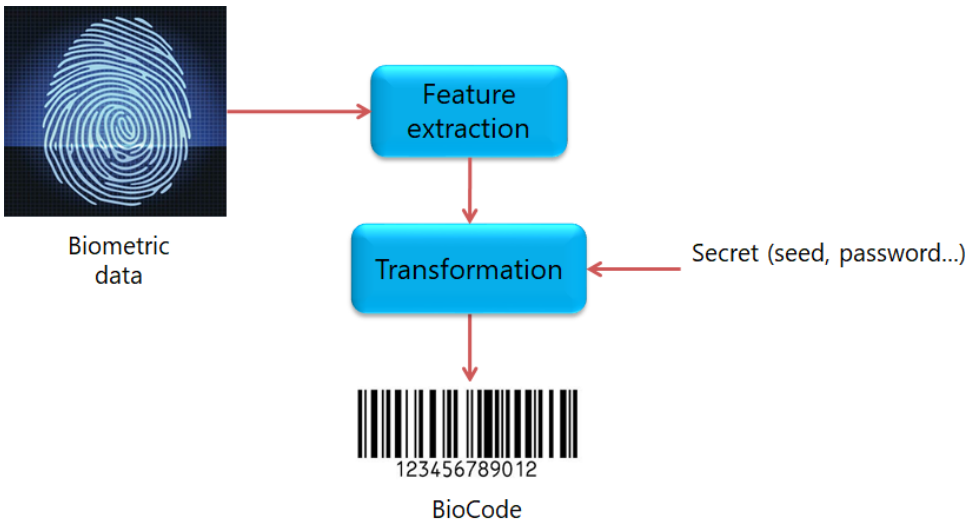


Figure 1: General principle of template protection schemes based on a transformation.

- σ_x and σ_y are the standard derivations of the Gaussian envelope along the x and y dimensions and determine the ratio of the Gaussian window width to wavelength.
- f provides the central frequency of the sinusoidal plane wave at an angle θ_n.
- θ_n is the orientation angle which is $\theta_n = \left(\frac{\pi}{k}\right) \times (n-1)$; $n = 1, 2, \ldots p$. Note that we denote by p the number of orientations.

After doing the Gabor transformations, we obtain a vector of n real numbers, denoted T. For more information, we refer the reader to the example detailed in [24]. A comparative study [14] in the case of fingerprints shows that vectors of $n = 512$ extracted features is nearly optimal. The gap of performance in terms of EER is much smaller when passing from $n = 256$ to $n = 512$, than when passing from $n = 128$ to $n = 256$, suggesting there is no point in increasing further the number of features. This size is then considered in this work. The next step, explained below, is to apply a cancelable transformation on the resulting vectors of Gabor transformations.

Cancelable transformations. We propose to keep the notations of [13]. Let T_z and $\acute{T}_z$ represent respectively the template and query biometric features of user z. Let f be the feature transformation function. Let K_z be a set of transformation parameters corresponding to the user z. A feature transformation is a non-invertible function using a user related key K_z (*i.e* typically a

random seed or a strong password), applied to the used biometric template T_z. The BioCode $f(T_z, K_z)$ is stored in a database or in a personal device. It is generally considered that, given the transformed template $f(T_z, K_z)$ and the key K_z, it is possible to recover the original template T_z (or a close approximation) as presented in [13]. Thus, it is requested to store this key in a second support, even if the reconstruction of the original template strongly depends on the used biometric modality.

Let n denotes the dimension of the BioCode $f(T_z, K_z)$ for the user z. Let D_T denote a distance function between the biometric features in the untransformed (original) domain. However, the comparison is computed in the transformed domain, thus we need D_T a distance function in the transformed one. The verification operation of the cancelable biometric scheme outputs a verification decision $R_z \in \{0, 1\}$ depending whether the distance between the reference BioCode and query BioCode is less than a decision threshold, denoted as ϵ:

$$R_z = 1_{\{D_T(f(T_z,K_z),f(\acute{T_z},K_z))\leq\epsilon\}} \tag{1}$$

The performance of the authentication system is generally estimated with FRR (False Rejection Rate)/FAR (False Acceptance Rate) rates, and the feature transformation should not decline the performance of the system. In fact, this approach tends to improve the performance of the biometric system without any protection even if the key K_z is necessary for the user z to authenticate herself/himself.

3. Related Works

The concept of privacy protection of biometric data has been defined in 2001 in a seminal paper [25]. Since then, many methods have been proposed among random projections approaches [26], BioHashing methods [7], Bloom filters [27], to cite just a few. A complete review of cancelable biometric systems can be found in [28]. Very recently, Teoh *et al.* [12] proposed a new two-factor scheme to protect the biometric template by transformation. Compared with previous works, this method is based on localized random projection and on the rank correlation. Moreover, the obtained results show that this system is strongly resistant against the main attacks. These good results are the consequences of their technical called *Index-Of-Max* which can be viewed as a machine learning on the plain database. For this previous constraint, we do not compare to this method where the BioSystem is tuned for a particular basis. More generally, we can find a security analysis of the biometric system protecting the biometric template based on transformations [15]. In the following paragraphs, we detail two particularly popular template protection schemes, BioHashing [7] and BioPhasor [10].

The BioHashing algorithm is applied on biometric templates that are represented by real-valued vectors of fixed length (so the metric used to evaluate the

similarity between two biometric features is the Euclidean distance). It generates binary templates of length lower than or equal to the original length (here, the metric D_T used to evaluate the similarity between two transformed templates is the Hamming distance). This algorithm has been originally proposed for face and fingerprints by Teoh *et al.* in [7]. The fingerprint images are, in a first extraction step (*e.g.* using the Gabor filter described above), transformed into a real-valued vector of fixed length, a biometric template. Then, the BioHashing algorithm transforms the biometric template $T = (T_1, \dots T_n)$ into a binary template $B = (B_1, \dots B_m)$, with $m \leq n$, as following:

Algorithm 1 BioHashing

1: **Inputs**
2: $T = (T_1, \dots, T_n)$: biometric template,
3: K_z: secret seed
4: **Output** $B = (B_1, \dots, B_m)$: BioCode
5: Generation with the seed K_z of m pseudorandom vectors $V_1, \dots, V_m$ of length n,
6: Orthogonalize vectors with the Gram-Schmidt algorithm,
7: **for** $i = 1, \dots, m$ **do** compute $x_i =< T, V_i >$.
8: Compute BioCode:

$$B_i = \begin{cases} 0 & \text{if} \quad x_i < \tau \\ 1 & \text{if} \quad x_i \geq \tau, \end{cases}$$

where τ is a given threshold, generally equal to 0.

The specificity of the BioHashing algorithm is that it uses a one way function and a random seed of m bits. It is important to note that every enrolled biometric feature uses a different seed in order to create a specific BioCode. The performance of this algorithm is ensured by the scalar products with the orthonormal vectors. The quantization process of the last step ensures the non-invertibility of the data (even if $n = m$, because each coordinate of the input T is a real value, whereas the coordinates of the output B is a single bit). Finally, the random seed guarantees both the diversity and revocability properties.

BioPhasor was proposed by Teoh *et al.* in [10] and was introduced as a form of cancelable biometrics which is based on iterated mixing between the user-specific pseudo-random number and the biometric feature. The BioPhasor algorithm is supposed to be an improvement of the BioHashing one. It is described in Algorithm 2. The step 8 is added in this paper in order to generate a binary output.

An interesting component of these schemes is that no learning phase is required. Nevertheless, some weaknesses have been reported in the former approach in [29, 30, 31]. The main reason is that the projection matrix is only related to the secret key. If an impostor obtains the key (known as stolen token

Algorithm 2 BioPhasor

Inputs
$b = (T_1, \ldots, T_n)$: biometric template
K_z: secret seed
Output $B = (B_1, \ldots, B_m)$: BioCode
Generation with K_z of m pseudorandom vectors $V_1, \ldots, V_m$ of length n,
Orthogonalize vectors with the Gram-Schmidt algorithm,
for $i = 1, \ldots, m$ **do** compute $h_i = 1/n \sum_{j=1}^{n} \arctan({T_j}^2 / V_i^j)$.
Compute BioCode:

$$B_i = \begin{cases} 0 & \text{if} \quad h_i < \tau \\ 1 & \text{if} \quad h_i \geq \tau, \end{cases}$$

where τ is a given threshold, generally equal to 0.

attack), the attack is quite easy especially by combining it with a real biometric data, his own or another user's. Indeed, with the knowledge of the key, the impostor will use the same projection basis as the legitimate user. Thus, it increases the success probability of its attack. In the sequel, we propose a new transformation whose objective is to limit this issue.

4. Proposed Method

We intend in this paper to enhance the security of the BioHashing algorithm by limiting the impact of stolen token based attacks. Geometrically, the classic BioHashing computes a change of basis that is uniquely determined by the seed. From a security standpoint, we can see that the seed has a most important part than the biometric data. Thus, we propose to generate differently the projection matrix: it is derived both from the seed and from the biometric data. So, attackers who steal the seed will have less benefit than with previous methods. This is the point mainly described in [16]. Moreover, we propose a first transformation of the biometric template, which is only a projection onto a hypersphere.

Indeed, we first project the biometric template onto a hypersphere of radius R, a parameter discussed in detail below, which depends only on the maximum of the reference biometric template. This first transformation permits to replace the knowledge of the average and standard deviation of the reference biometric template by only the knowledge of their maximum value. Moreover, this transformation improves the efficiency of the obtained biometric system.

As for BioHashing and BioPhasor methods, we assume that biometric data are given into features forms. Thus, we have the following properties: the features are fixed length, and the biggest versatility part is removed. It is exactly what is assumed for existing method, moreover it is a realistic hypothesis, as for example Gabor filters for fingerprint modality. We suppose in this method

to have some statistics concerning the features (*i.e.* average and standard deviation). This hypothesis is not a high restriction of the operational use of the method and keeps its possible use for any biometric modality. We generate a projection matrix given the seed as secret and the biometric template. For the informative purpose, let us to introduce:

- A hypersphere radius factor R. We take the Euclidean space $\mathbb{R}^n$ with the Euclidean norm given for $x \in \mathbb{R}^n$:

$$||x||_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

  and the hypersphere of $\mathbb{R}^n$. Thus, we consider the hypersphere of radius $R \times max$, where max is the maximum of the biometric template, and the centre is the barycenter of the last one. R is a parameter of the proposed method which modifies the radius of the projection hypersphere. In Section 5.5, we study the impact of R on the proposed system (Figure 10 gives some experiment results for different values).

- A constant number C. This parameter defines the domain interval of the polynomial coefficients. Indeed, the polynomial coefficients are randomly drawn from the interval $[-C.\sigma;\ C.\sigma]$, where σ is the mean of the features. Thus, by the choice of the parameter C, we can define a trade-off between performance without any attack and robustness to attacks. In Figure 11 (Section 5.5), we give some experiment results for different values of C.

- A polynomial P_k such as $P_k(X) = \sum_{i=0}^{3} a_i * X^i, k = 1 : m, a_i \in [-C.\sigma;\ C.\sigma]$, $a_3 \neq 0$. When we manipulate polynomials, a natural question is the choice of their degrees. Since we expect a pseudo-random behavior, we propose to use the conjecture of Hoory *et al.*, which states that the 4-wise independent functions reach pseudo-randomness [32, Part 6]. Thus, we choose polynomials of degree three, which are four-wise independent functions. A three-degree polynomial is completely defined by its four independent monomial coefficients. Moreover, we have analyzed the different results by changing the polynomials degree, and the best behavior is indeed obtained when the degree is set at three.

In the following figure 2, we illustrate the general principle of the proposition of the BioHashing improvement, and in particular, the construction of the projection matrix, computed both from the seed and the biometric data. The transformation operation of our proposal is detailed in Algorithm 3.

The key ingredient of our proposed method is the new construction of the projection matrix P. Firstly, we choose randomly polynomials of degree 3 initialized by the seed. Then, we construct P by evaluating these polynomials at each component of the biometric template (Step 8). This provides a method for biometric template protection, where the transformation depends both on the secret and on the biometric template.

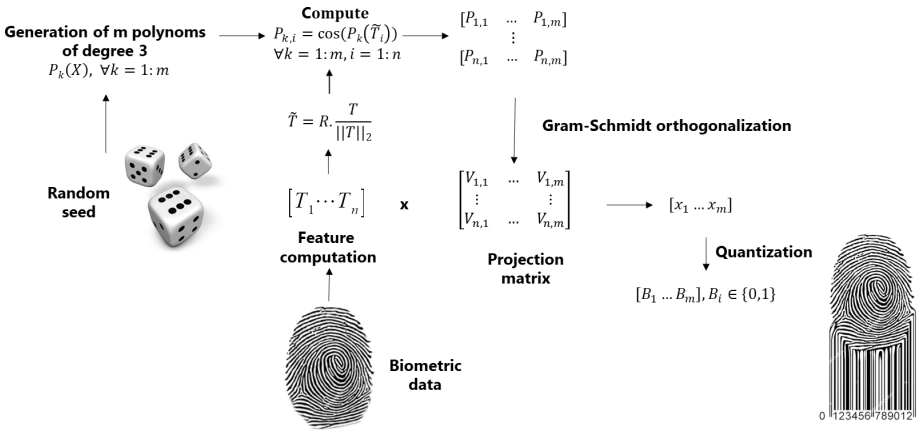


Figure 2: General principle of the proposed method.

5. Performance and Security Analyses

In this section, we present some experimental results demonstrating the benefit of the proposed transformation.

5.1. Dataset

We propose to evaluate our new transformation on databases of different quality. We voluntary give more importance to biometric data. Then, we analyze how much is the impact of database quality on our transformation method. In order to study its performance and robustness, we selected three biometric datasets.

The first one is well known as the FVC2002 DB1 database composed of digital fingerprint images. 100 individuals provided 8 samples of fingerprints. Figure 3 presents some examples of fingerprints. We compute well known Gabor features for each image, and we obtain in this case 512 real-valued features for each fingerprint.

The second dataset is the PolyHK which is composed of images of knuckle prints [33] (see figure 4). The database has been acquired on 4 fingers of 165 volunteers, leading to 660 different classes. Each class contains 12 images acquired during 2 sessions. We compute Gabor features for each image, and we obtain 256 real-valued features for each finger knuckle print.

The third dataset is the VEINEGY database which is composed of images of dorsal hand veins [34]. 100 individuals provided 5 samples of hand veins, for a total of 500 samples. Note that this database is of poor quality with

Algorithm 3 GREYCHashing (transformation)

Inputs
$T = (T_1, \ldots, T_n)$: biometric template
C: constant
K_z: secret seed
Output $B = (B_1, \ldots, B_m)$: BioCode
Compute $\tilde{T} = R \times \frac{T}{||T||_2}$
Generation with the seed K_z of m polynomials P_k with $k = 1 : m$
Evaluate $P_{k,i} = cos(P_k(\tilde{T}_i)), \forall i = 1 : n, \forall k = 1 : m,$
Orthogonalize the matrix P with the Gram-Schmidt algorithm,
for $i = 1, \ldots, m$ **do** compute $x_i =< T, P_i >$.
Compute BioCode:

$$B_i = \begin{cases} 0 & \text{if} \quad X_i < \tau \\ 1 & \text{if} \quad X_i \geq \tau, \end{cases}$$

where τ is a given threshold, generally equal to 0.

Figure 3: Some examples of fingerprint images in the FVC2002DB1 dataset.

small images. We compute Gabor features for each image, and we obtain 512 real-valued features for each of them.

5.2. Security Properties

Few works have been dedicated to the evaluation of such biometric systems in the literature [35, 36, 13]. The ISO/IEC 24745 "Information technology – Security techniques – Biometric information protection" defines the security properties of a biometric system, we use in this paper the same terms. Cancelable systems must fulfill several properties as also mentioned in [37, 15]:

- *Revocability/Renewability:*
  It should be possible to revoke a biometric template and to generate a new one from the original biometric data.

- *Performance:*
  The template protection shall not deteriorate the performance of the original biometric system.

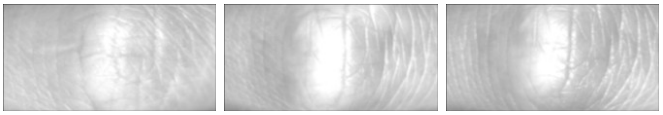


Figure 4: Some examples of finger knuckle print images in the PolyHK dataset.

- *Non-invertibility or Irreversibility:*
  From the transformed data, it should not be possible to obtain enough information on the original biometric data, to prevent any attack consisting in forging a stolen biometric template (as for example, it is possible to generate an eligible fingerprint given minutiae [38]). This property is essential for security purposes. For any attack, an impostor provides an information in order to be authenticated as the legitimate user. The success of the attack is given by:

$$FAR_A(\epsilon) = P(D_T(f(T_z, K_z), A_z) \leq \epsilon) \tag{2}$$

  Where FAR_A is the probability of a successful attack by the impostor for a decision threshold set to ϵ. The A_z BioCode is computed by the impostor by taking into account as much information as possible within different contexts.

- *Diversity or Unlinkability:*
  It should be possible to generate different BioCodes for multiple applications, and no information should be deduced from the comparison or the correlation of different realizations.

- *Indistinguishability:*
  To assure the indistinguishability, it should be infeasible to cross correlate two protected templates.

Based on some of the early works [39], [40] which identified weak links in each subsystem of a generic authentication system, some papers considered the possible attacks in cancelable biometric systems (such as those presented in [41, 1, 13, 9]). We follow the Shannon's maxim ("The enemy knows the system"), we so assume that the impostor has all necessary information on the process used to generate the BioCode (feature generation method, BioCode size...).

Based on the principle of each attack, we generate many fake attempts A_z of the genuine user in an authentication case:

- *False accept attack (zero effort)*:
  For this attack, an impostor user x provides its biometric feature $\acute{T_x}$ and parameter K_x to be authenticated as the legitimate user z:

$$A_z = f(\acute{T_x}, K_x).$$

- *Brute force attack*:
  For the brute force attack, an impostor tries to be authenticated by trying different random values of A until finding:

$$A_z = A.$$

- *Stolen token attack*:
  An impostor has obtained the token K_z (stolen token) of the genuine user z and tries different random values T to generate:

$$A_z = f(T, K_z).$$

- *Stolen biometric data attack*:
  An impostor knows $\acute{T_z}$ (directly or after computation of the feature on a biometric raw data) and tries different random numbers K to generate:

$$A_z = f(\acute{T_z}, K).$$

- *Worst case attack*:
  An impostor user x provides its biometric template $\acute{T_x}$ and parameter K_x to be authenticated as the user z (false accept attack) and has also obtained the token K_z (stolen token) of the genuine user z to generate:

$$A_z = f(T_x, K_z).$$

- *Attacks via Record Multiplicity (ARM):*
  An impostor must not be able to extract any information from different BioCodes issued from the same user. Since BioCodes can be revoked, an impostor can intercept Q of them and issue a new one by predicting an admissible value (as for example by setting each bit to the most probable value). These attacks consist in the following process:

  - Generation of Q BioCodes for user z such that

$$B_z = \{f(T_z, {K_z}^1), .., f(T_z, {K_z}^Q)\}$$

  - Prediction of a possible BioCode value by setting the most probable value of each bit given B_z, $\Rightarrow$ computing the FAR_A value for $Q = 3$ and $Q = 11$.

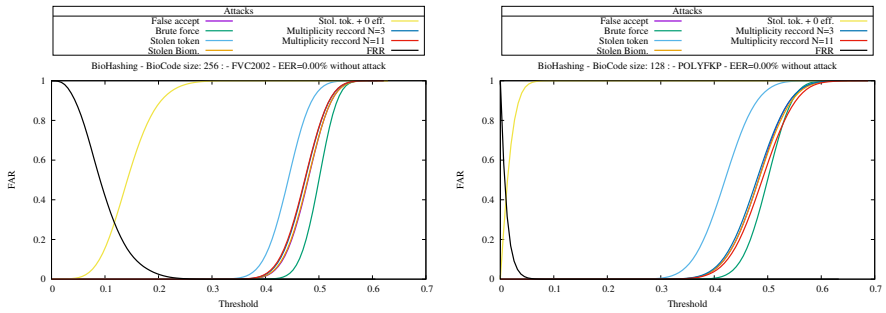
  So, this ARM combining attack consists in computing a new BioCode from a list of ones.

These attacks allow us to quantify the robustness of cancelable biometric verification systems based on feature transformation.

5.3. Performance Evaluation

We now present the experimental results that we obtained when using our proposed transformation face to the BioHashing and BioPhasor protection schemes. Before considering their behaviors, we computed the performance of biometric systems by using Gabor features with the Euclidean distance without any transformation. For different acceptance thresholds, we compute the False Acceptance Rate (FAR), the rate of the illegitimate users who are authenticated; and the False Reject Rate (FRR), the rate of the legitimate users who are rejected. Thus, we obtain the Equal Error Rate (ERR), which is the acceptance threshold when the FAR and FRR are equals. Usually, we set the acceptance threshold to the EER, so that the biometric system optimizes both the FRR and FAR at the same time. Then, smaller is the EER, more accurate is the system.

For the performance without transformation (*i.e.* using unprotected feature vectors), we obtained an EER of 28.3% for the fingerprint dataset and of 25.2% for the finger knuckle print one. We can see clearly that these results are poor and that such a biometric system could not been used in real conditions.



(a) Behavior of the BioHashing system on the fingerprint FVC2002 database

(b) Behavior of the BioHashing system on the knuckle POLYFKP database

Figure 5: Security analysis of the BioHashing algorithm on the fingerprint database (a) and the finger knuckle print database (b).

Protection schemes are known to increase performance if the secret is unknown to impostors. Without any attack, the BioHashing and Biophasor algorithms provide a perfect recognition (*i.e.* ERR=0%). Using the proposed scheme, the performance is a bit lower, with 3.72% for the fingerprint dataset and 14.93% for the finger knuckle print one. We show in the next section that these results are compensated by a better robustness face to attacks.

5.4. Security Evaluation

We analyze the irreversibility, the robustness to attacks and the unlinkability properties of our transformation.

Irreversibility analysis. As stated before, the BioHashing algorithm is prone to feature approximation attacks, a shortcoming mainly due to the linearity of the transformation. An approximation of the feature vector is then sufficient to gain access to the system. In a stolen token scenario, it is possible to reverse partially the transformation using a pseudo-inverse matrix, even if the arrival space of the transformation is of smaller dimension than the feature vector space. The advantages of the proposed scheme are twofold:

- on the one hand, let us summarize the the proposed transformation. We start with evaluation of random polynomials to make a random matrix. Then, we perform a Gram-Schmidt process, which is only a change of basis of the random matrix. Finaly, we perform a vector projection of the biometric vector into this orthonormal basis followed by a quantization of the components to obtain a binary vector. Thus, the proposed transformation can be viewed as functions composition between polynomial evaluations and a vector projection. Since the polynomial evaluation is non-linear and vector projection is linear, the transformation is also non-linear from these inputs: secret and biometric data.

- on the other hand, since the projection matrix is derived from biometrics, finding a pseudo-inverse for a such matrix is difficult, as far as we know.

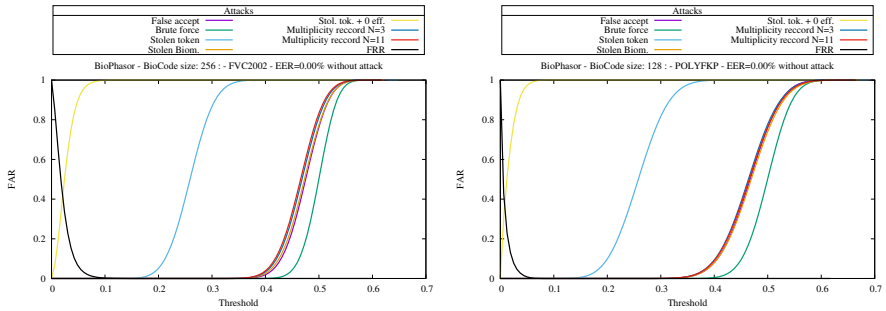
Moreover, we can see our method as an evolution of the BioHashing or BioPhasor one, except that the projection matrix is computed from both of the random secret and the biometric data. This step of the transformation requires the evaluation of three-degree polynomials. As previously seen, this family of polynomials has been chosen for their random behavior. Moreover, an attacker who is trying to guess some information from the BioCode is faced to the polynomial interpolation problem with only the ordinates. Indeed, the traditionnal polynomial interpolation problem is: knowing $\{(x_i, y_i)\}$ for some $i \in \{1, \ldots, n\}$, compute the unique polynomial P of degree strictly lower than n such that

$$\forall i,\ P(x_i) = y_i.$$

In the imposter context, the attacker will only know a transformation of y_i, but not x_i. The computation of the polynomial P is then impossible. The aforementioned arguments prove the one-wayness of the transformation.

Robustness to attacks. Figures 5, 6 and 7 present at the same time the performance of cancelable biometric systems without any attack and the robustness to attacks described in the previous section. The black curve presents the evolution of the false rejection rate (FRR) depending on the decision threshold (ϵ in equation 2). Other curves correspond to the FAR_A value for all considered attacks.

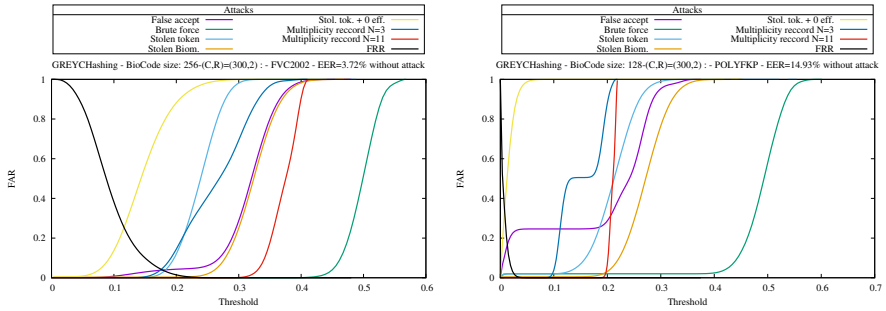
For the biometric modalities of the two first datasets (fingerprints and finger knuckle prints), the BioHashing and BioPhasor algorithms provide a perfect



(a) Behavior of the BioPhasor system on the fingerprint FVC2002 database

(b) Behavior of the BioPhasor system on the knuckle POLYFKP database.

Figure 6: Security analysis of the BioPhasor algorithm on the fingerprint database (a) and the finger knuckle print (b).

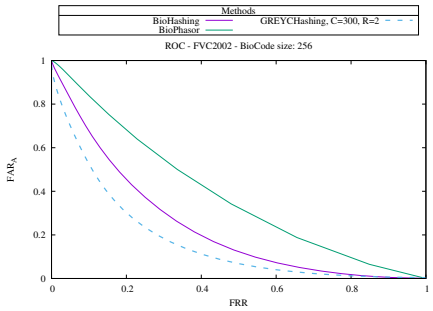


(a) Behavior of the proposed system on the fingerprint FVC2002 database.

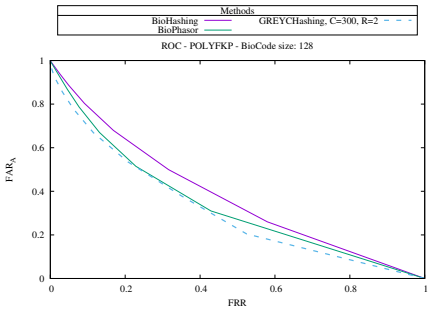
(b) Behavior of the proposed system on the knuckle POLYFKP database.

Figure 7: Security analysis of the proposed algorithm on the fingerprint database (a) and the finger knuckle print (b). Experiments for GREYCHashing have been done with polynomial functions of degree 3, $C = 300$ and $R = 2$.

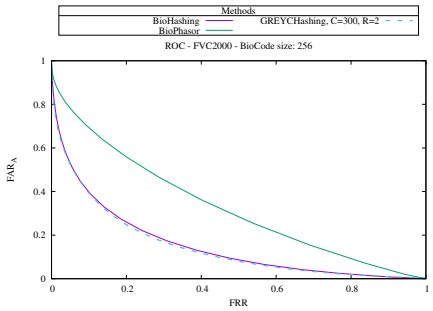
performance without any attack, *i.e.* when the secret is unknown to the impostor. That means that in a non-adversarial environment, these systems work perfectly without error. For the proposed transformation, it is not always the case. It depends on the value of C. When C and R are set to 300 and 2 respectively, the EER value equals to 3.72% for the FVC2002DB1 database, and 14.93% for the POLYFKP database. The main reason of this decrease of performance is due to the fact that the projection matrix is computed by considering biometric data and not only the secret. As the performance when using raw features (*i.e.* without any transformation) is low, it is normal to have such an impact. However, we will see later that we have a significant gain in term of robustness. We will also show a study of the impact of C on the performance and robustness of the proposed transformation scheme.



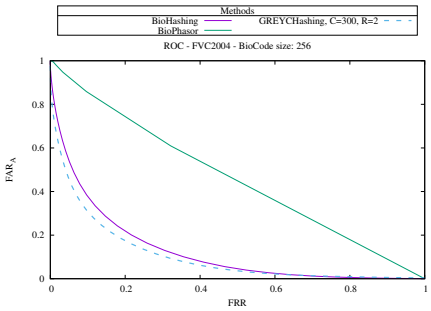
(a) Robustness in function of the performance of studied systems for the FVC2002 fingerprint database.



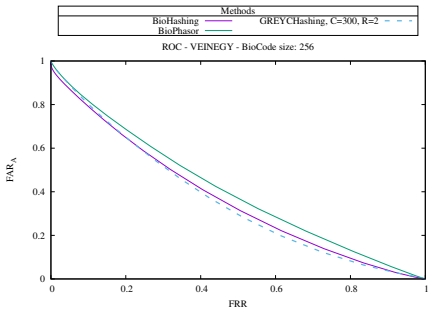
(b) Robustness in function of the performance of studied systems for the POLYFKP knuckle print database.



(c) Robustness in function of the performance of studied systems for the FVC2000 fingerprint database.



(d) Robustness in function of the performance of studied systems for the FVC2004 fingerprint database.



(e) Robustness in function of the performance of studied systems for the VEINEGY vein print database.

Figure 8: Comparison between performance and robustness face to attacks for the three tested template protection schemes. The FRR value is seen as a performance indicator while the FAR_A value gives the probability of successful attack in the worst case.

We then analyze the robustness of the protection schemes for the aforementioned attacks. Of course, the brute force attack (green curve) is the least

effective and all schemes are able to avoid it. The BioHashing and BioPhasor algorithms have a similar behavior for all attacks even if the BioPhasor algorithm is more sensitive to the stolen token attack (light blue curve). Only the worst case attack is completely possible ($\text{FAR}_A = 100\%$) for these two protection schemes when the ϵ decision threshold value is set to have the behavior of the biometric systems at the EER value. The proposed transformation scheme is slightly more robust to attacks (even in the worst case one) thanks to the use of information related to the biometric data in the computation of the projection matrix.

To give a better indicator of comparison of our proposal with BioHashing and BioPhasor, and to identify the robustness behavior of these three protection schemes, we propose in to plot in Figure 8, for several databases, the FAR_A value in function of the FRR value, where A is the worst case attack. Since this attack is the most powerful stolen token attack, it makes sense to give emphasis on it. We can note, for this attack, that the Robustness of BioHashing is equivalent to that of BioPhasor. There is actually a compromise to find between performance and robustness in such transformation schemes. This curve shows clearly that for a given performance (in terms of FRR value), the robustness evaluated by the FAR_A value is lower for the proposed method up to 5%. Our proposal then meets our main objective: providing a better robustness against stolen token attack.

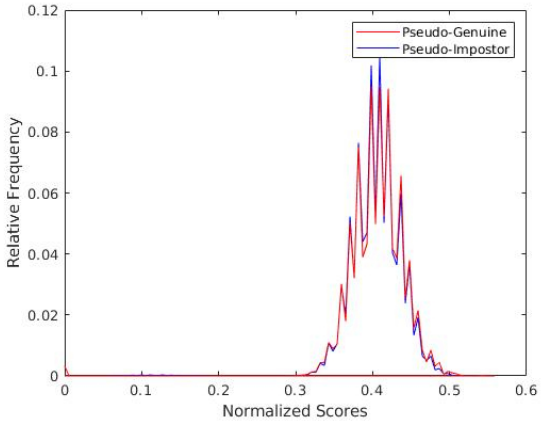


Figure 9: Pseudo-Genuine & Pseudo-Impostor distributions for unlinkability analysis. 256-bit BioCodes were generated using GREYCHashing (with $C = 300$ and $R = 2$) from 512-feature vectors, extracted with Gabor filter on the FVC2002DB1 database. Overlapped distributions indicates indistinction between BioCodes generated from the same user and BioCodes generated from the others.

Unlinkability analysis. We use the same methodology as [12], by comparing the distributions of matching scores, namely the *pseudo-genuine scores* and the

pseudo-impostor scores. The pseudo-impostor scores refer to the matching scores between BioCodes generated from different feature vectors of the same individual (intra-class) but using different seeds. The pseudo-genuine score is computed between BioCodes derived from different (inter-class) feature vectors in stolen token scenario. Figure 9 illustrates the pseudo-impostor and pseudo-genuine distribution plot, where we can see that the pseudo-impostor and pseudo-genuine distributions are sufficiently overlapped, and thus sufficiently indistinguishable.

5.5. Parameters study

Figure 10 shows the performance vs robustness of GREYCHashing for several databases when C is fixed at 300 but R takes value in $\{0.5, 1, 2, 5, 10\}$. The BioCodes are of size 256 for the FVC and VEINEGY databases, and of size 128 for the POLYFKP database. We can see that $R = 2$ performs well for both the fingerprints (whatever is the database) and the finger knuckle prints (POLYFKP), while $R = 0.5$ is better for the VEINEGY database. Figure 11 shows the performance vs robustness of GREYCHashing for the same databases and same BioCode sizes, when R is fixed at 2 but C takes value in $\{10, 50, 100, 200, 250, 275, 300, 325, 350, 500\}$. Except for the VEINEGY database (of lower quality than the others), we can observe that values of C in the range 275-325 are the preferable choices. Setting the C value can adjust the compromise between performance and robustness. We have to recall at this point that the use of raw features lead to an high value of EER, for the three databases. We could expect even better results of the proposed transformation with more efficient features. The pair $(C, R) = (300, 2)$ is not always optimal but is a reasonable tradeoff. As a consequence, this pair has been chosen to conduct our previous experiments.

Table 1 shows the results we obtained in the case of the FVC2002DB1 database for different BioCode sizes (m value), when (C, R) is set to $(300, 2)$. If a too small size of BioCode does not provide perfect results in terms of EER without attack, we observe, however, a good resilience to worst case attacks. This resistance slightly decreases when m increases, but it is still in favour of the proposed method. We recall that in the case of the BioHashing and BioPhasor algorithms, the FAR_A in the *worst case* scenario is 100%. When the size of the BioCode is too similar to the size of the features vector, attacks are more easy to realize (but this is also the case for other transformation approaches).

6. Conclusion and Perspectives

Protecting biometric templates is actually crucial due to new regulation laws on data protection (such as the GDPR in Europe) and the possiblity of using biometric data for authentication in cloud services. We believe that feature-based template protection schemes could have a high impact on security services in the near future. In this context, features, that can be computed without any

Table 1: Study of the impact of the BioCode size (m value) on performance and robustness in the proposed transformation scheme, when $(C, R) = (300, 2)$.

m value	EER (without attack)	false accept	brute force	stolen token
64	28.2%	35.6%	2%	2%
128	30.3%	27.3%	0%	0%
256	3.8%	3.9%	0%	0%
384	2.2%	2.2%	0%	10%
512	4.7%	4.7%	3%	0%

m value	stolen biom.	worst case	ARM comb. N=3	ARM comb. N=11
64	17.6%	57.6%	0%	0%
128	14%	51.8%	0%	0%
256	0.3%	73.7%	3.6%	0%
384	0.2%	86.7%	3.6%	0%
512	0.1%	78.1%	1%	0%

learning with unprotected templates from other users, should be protected by such cancelable schemes. The proposed transformation scheme has the advantage to better combine the secret with the biometric data in order to reduce the impact of the stolen token attack. We obtain with this scheme a configurable transformation by adjusting the compromise between the expected performance assuming an unknown secret and a better robustness otherwise with only a pair of parameters (C and R).

Perspectives of this work are manifold. First, it concerns the proposal of new combination techniques of the biometric data and the secret (the steps 6, 7 and 8 of the proposed algorithm). Second, we could consider its extension to multi-biometrics, where a first biometric modality along with the secret could be used to generate the projection matrix, the latter serving to project the features of the second modality. Last, this scheme can be used in real biometric authentication applications in industry as the computation is very fast while keeping a good privacy protection.

References

[1] A. Jain, K. Nandakumar, A. Nagar, Biometric template security, in: EURASIP Journal on Advances in Signal Processing, 2008.

[2] A. Juels, M. Wattenberg, A fuzzy commitment scheme, in: ACM conference on Computer and communication security, 1999, pp. 28–36.

[3] A. Juels, M. Sudan, A fuzzy vault scheme, Des. Codes Cryptography 38 (2) (2006) 237–257.

[4] H. Chabanne, J. Bringer, G. Cohen, B. Kindarji, G. Zemor, Optimal iris fuzzy sketches, in: IEEE first conference on biometrics BTAS, 2007.

[5] J. Bringer, H. Chabanne, M. Favre, A. Patey, T. Schneider, M. Zohner, Gshade: faster privacy-preserving distance computation and biometric identification, in: Proceedings of the 2nd ACM workshop on Information hiding and multimedia security, ACM, 2014, pp. 187–198.

[6] S. Chatterjee, S. Roy, A. K. Das, S. Chattopadhyay, N. Kumar, A. V. Vasilakos, Secure biometric-based authentication scheme using chebyshev chaotic map for multi-server environment, IEEE Transactions on Dependable and Secure Computing.

[7] A. Teoh, D. Ngo, A. Goh, Biohashing: two factor authentication featuring fingerprint data and tokenised random number, Pattern recognition 40.

[8] R. Belguechi, C. Rosenberger, S. Aoudia, Biohashing for securing minutiae template, in: Proceedings of the 20th International Conference on Pattern Recognition, Washington, DC, USA, 2010, pp. 1168–1171.

[9] N. Saini, A. Sinha, Soft biometrics in conjunction with optics based biohashing, Optics Communications 284 (3) (2011) 756 – 763.

[10] A. Teoh, D. Ngo, Cancellable biometrics realization through biophasoring, in: Proceedings of 9th IEEE International Conference on Control, Automation, Robotics and Vision (ICARCV'06), 2006.

[11] N. Ratha, S. Chikkerur, J. Connell, R. Bolle, Generating cancelable fingerprint templates, IEEE Trans. Pattern Anal. Mach. Intell. 29 (4) (2007) 561–572.

[12] Z. Jin, J. Y. Hwang, Y. L. Lai, S. Kim, A. B. J. Teoh, Ranking-based locality sensitive hashing-enabled cancelable biometrics: Index-of-max hashing, IEEE Transactions on Information Forensics and Security 13 (2) (2018) 393–407.

[13] A. Nagar, K. Nandakumar, A. K. Jain, Biometric template transformation: A security analysis, Proceedings of SPIE, Electronic Imaging, Media Forensics and Security XII.

[14] R. Belguechi, A. Hafiane, E. Cherrier, C. Rosenberger, Comparative study on texture features for fingerprint recognition: application to the biohashing template protection scheme, Journal of Electronic Imaging 25 (1) (2016) 013033–013033.

[15] C. Rosenberger, Evaluation of biometric template protection schemes based on a transformation, in: International Conference on Information Systems Security and Privacy (ICISSP), 2018.

[16] L. Ghammam, M. Barbier, C. Rosenberger, Enhancing the security of transformation based biometric template protection schemes, in: International Conference on CYBERWORLDS (CW), 2018.

[17] D.Gabor, Theory of communications, J. Inst. Elect. Eng 93 (1946) 429–457.

[18] J. G. Daugman, Complete discrete 2-d gabor transforms by neural networks for image analysis and compression, IEEE Trans. Acoustics, Speech, and Signal Processing 36 (7) (1988) 1169–1179.

[19] A. C. Bovik, M. Clark, W. S. Geisler, Multichannel texture analysis using localized spatial filters, IEEE Trans. Pattern Anal. Mach. Intell. 12 (1) (1990) 55–73.

[20] M. Lades, J. C. Vorbrüggen, J. M. Buhmann, J. Lange, C. von der Malsburg, R. P. Würtz, W. Konen, Distortion invariant object recognition in the dynamic link architecture, IEEE Trans. Computers 42 (3) (1993) 300–311.

[21] B. S. Manjunath, R. Chellappa, C. von der Malsburg, A feature based approach to face recognition, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 1992, Proceedings, 15-18 June, 1992, Champaign, Illinois, USA, 1992, pp. 373–378.

[22] B. S. Manjunath, C. Shekhar, R. Chellappa, A new approach to image feature detection with applications, Pattern Recognition 29 (4) (1996) 627–640.

[23] K. M. S. R. T.M. Abhishree, J. Latha, Face recognition using gabor filter based feature extraction with anisotropic diffusion as a pre-processing technique, Procedia Computer Science 45 (2015) 312–321.

[24] A. K. Jain, S. Prabhakar, L. Hong, S. Pankanti, Filterbank-based fingerprint matching, IEEE Trans. Image Processing 9 (5) (2000) 846–859. doi:10.1109/83.841531.
URL https://doi.org/10.1109/83.841531

[25] N. K. Ratha, J. H. Connell, R. M. Bolle, Enhancing security and privacy in biometrics-based authentication systems, IBM systems Journal 40 (3) (2001) 614–634.

[26] J. K. Pillai, V. M. Patel, R. Chellappa, N. K. Ratha, Sectored random projections for cancelable iris biometrics, in: Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on, IEEE, 2010, pp. 1838–1841.

[27] C. Rathgeb, F. Breitinger, C. Busch, H. Baier, On application of bloom filters to iris biometrics, IET Biometrics 3 (4) (2014) 207–218.

[28] V. M. Patel, N. K. Ratha, R. Chellappa, Cancelable biometrics: A review, IEEE Signal Processing Magazine 32 (5) (2015) 54–65.

[29] A. Kong, K. Cheung, D. Zhang, M. Kamel, J. You, An analysis of biohashing and its variants, Pattern Recognition 39.

[30] A. Lumini, L. Nanni, Empirical tests on biohashing, NeuroComputing 69 (2006) 2390–2395.

[31] K. Simoens, C. Chang, B. Preneel, Privacy weaknesses in biometric sketches, in: 30th IEEE Symposium on Security and Privacy, 2009.

[32] S. Hoory, A. Magen, S. Myers, C. Rackoff, Simple permutations mix well, Theor. Comput. Sci. 348 (2-3) (2005) 251–261.

[33] L. Zhang, L. Zhang, D. Zhang, Finger-knuckle-print verification based on band-limited phase-only correlation, in: International Conference on Computer Analysis of Images and Patterns, Springer, 2009, pp. 141–148.

[34] A. Badawi, Hand vein database, At systems and biomedical engineering, Cairo University.

[35] A. Adler, Biometric system security, Handbook of biometrics, Springer ed., 2007.

[36] X. Zhou, S. Wolthusen, C. Busch, A. Kuijper, Feature correlation attack on biometric privacy protection schemes, in: Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2009, pp. 1061–1065.

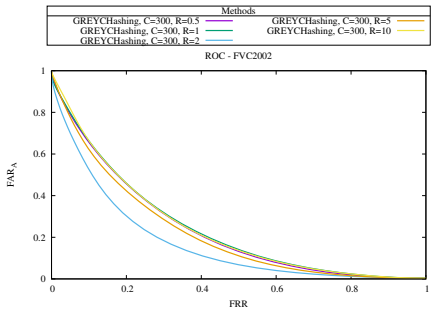
[37] D. Maltoni, D. Maio, A. Jain, S. Prabhakar, Handbook of Fingerprint Recognition, Springer, 2003.

[38] D. M. R. Cappelli, A. Lumini, D. Maltoni, Fingerprint image reconstruction from standard templates, IEEE Transactions on Pattern Analysis Machine Intelligence 29 (2007) 1489–1503.

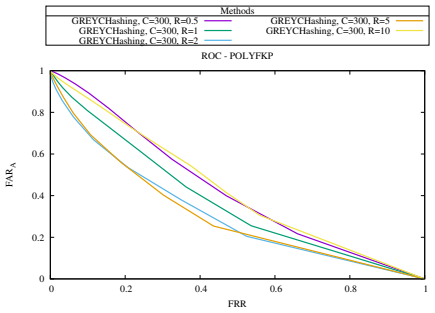
[39] N. Ratha, J. Connelle, R. Bolle, Enhancing security and privacy in biometrics-based authentication system, IBM Systems J. 37 (11) (2001) 2245–2255.

[40] R. Bolle, J. Connell, N. Ratha, Biometric perils and patches, Pattern Recognition 35 (12) (2002) 2727–2738.

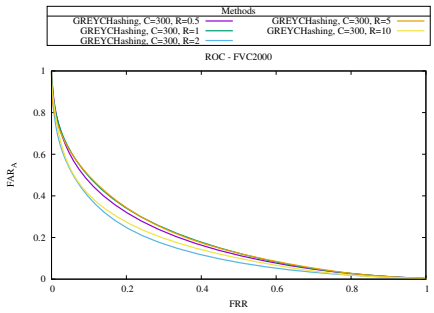
[41] A. Teoh, Y. Kuanb, S. Leea, Cancellable biometrics and annotations on biohash, Pattern recognition 41 (2008) 2034–2044.



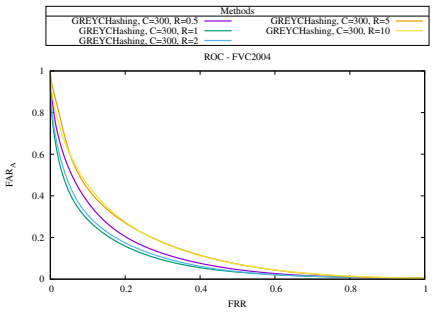
(a) Robustness in function of the performance of studied systems for the FVC2002 fingerprint database.



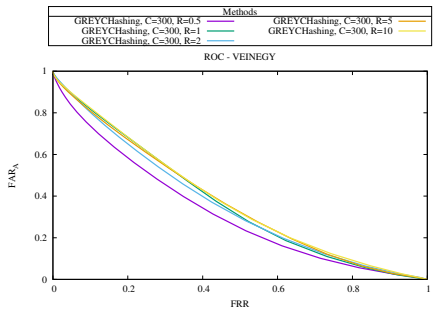
(b) Robustness in function of the performance of studied systems for the POLYFKP knuckle print database.



(c) Robustness in function of the performance of studied systems for the FVC2000 fingerprint database.

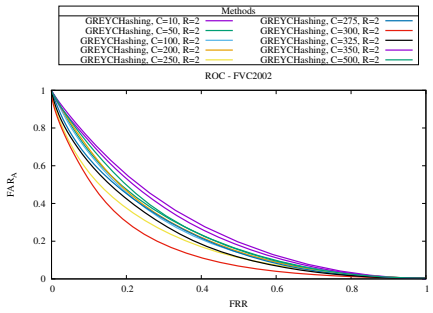


(d) Robustness in function of the performance of studied systems for the FVC2004 fingerprint database.

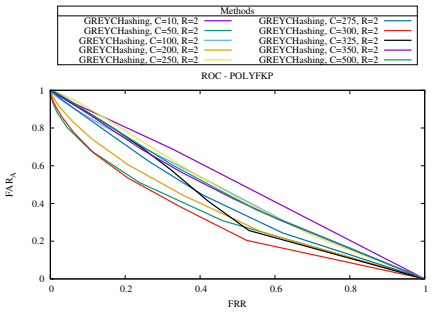


(e) Robustness in function of the performance of studied systems for the VEINEGY vein print database.

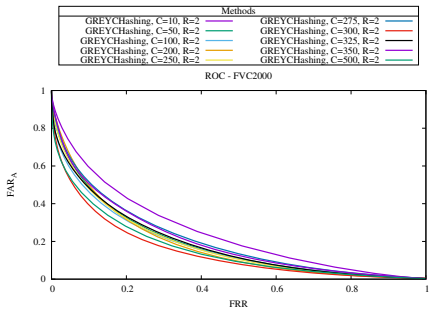
Figure 10: Comparison between performance and robustness face to attacks for the proposed method with different values for the parameter R. The FRR value is seen as a performance indicator while the FAR_A value gives the probability of successful attack in the worst case.



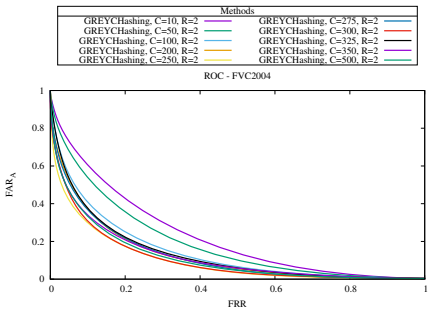
(a) Robustness in function of the performance of studied systems for the FVC2002 fingerprint database.



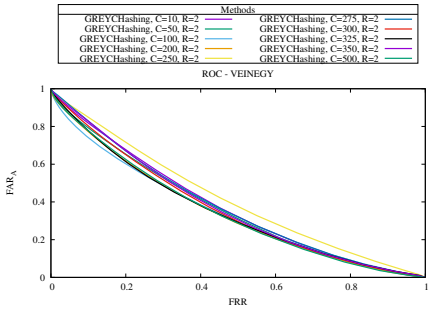
(b) Robustness in function of the performance of studied systems for the POLYFKP knuckle print database.



(c) Robustness in function of the performance of studied systems for the FVC2000 fingerprint database.



(d) Robustness in function of the performance of studied systems for the FVC2004 fingerprint database.



(e) Robustness in function of the performance of studied systems for the VEINEGY vein print database.

Figure 11: Comparison between performance and robustness face to attacks for the proposed method with different values for the parameter C. The FRR value is seen as a performance indicator while the FAR_A value gives the probability of successful attack in the worst case.