Enhancing the Security of Transformation Based Biometric Template Protection Schemes

Loubna Ghammam, Morgan Barbier, Christophe Rosenberger

To cite this version:

Loubna Ghammam, Morgan Barbier, Christophe Rosenberger. Enhancing the Security of Transformation Based Biometric Template Protection Schemes. CyberWorlds, Oct 2018, Singapour, Singapore. hal-01862157
Enhancing the Security of Transformation Based Biometric Template Protection Schemes

Loubna Ghammam, Morgan Barbier, Christophe Rosenberger
Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France
Loubna.Ghammam@unicaen.fr, {morgan.barbier,christophe.rosenberger}@ensicaen.fr

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 (zero effort 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 two biometric modalities, namely digital fingerprint and finger knuckle print images, show the benefits of the proposed method face to attacks while keeping a good efficiency.

Index Terms—Template protection, biometric authentication, attack, performance evaluation.

I. 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 perceived as personal and sensitive data. In Europe, as 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.) [9].

Three main approaches can be distinguished dealing with template protection in biometrics. First, biometric crypto-systems or secure sketches, such as those presented in [6], [11], [12], resort to cryptography. Second, secure computing methods aim at computing the comparison of two biometric templates by an untrusted party [5], [7]. 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 [2], [25], [28].

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 have been published dealing with the definition of new schemes for the protection of biometric templates (such as those presented in [2], [10], [21], [29]).

Most of such protection schemes lack of robustness considering the stolen token scenario. In this case, an attacker knowing the secret of the transformation protection scheme, has a big advantage to impersonate an user as presented in [3], [16]. This is due to the fact that the used projection matrix is computed only given the secret. The main contributions of the paper are twofolds. First, 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 embed information computed from the biometric data which 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 [24].

The paper is organized as follows. Section II gives the background on template protection schemes based on a transformation. Section III is dedicated to a literature review on template protection schemes based on a transformation. Then, we present the proposed methodology in Section IV. Section V illustrates the benefits of the proposed methodology through experimental results to be compared with those of two cancelable biometric systems. Finally, in section VI, we conclude and we give some perspectives.
II. BACKGROUND

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

We propose to keep the notations of [16]. Let $T_z$ and $\tilde{T}_z$ 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 [16]. 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 $f(T_z, K_z)$ BioCode for the user $z$. Let $D_T$ denotes 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 cancelable biometric system outputs a verification decision denoted $R_z$ if the distance between the reference BioCode and query BioCode is less than a decision threshold denoted as $\epsilon$:

$$R_z = 1_{\{D_T(f(T_z, K_z), f(\tilde{T}_z, K_z)) \leq \epsilon\}}$$  \hspace{1cm} (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.

III. RELATED WORKS

The concept of privacy protection of biometric data has been defined in 2001 in a seminal paper [20]. Since then, many methods have been proposed among random projections approaches [18], BioHashing methods [28], Bloom filters [23]. . . A complete review of cancelable biometric systems can be found in [17]. Very recently, Teoh et al. [10] 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 consequence 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 [24].

In this paper, We detail particularly two popular template protection schemes: BioHashing [28] and BioPhasor [29] and in this context, we detail each algorithm.

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 or equal to the
original length (here, the metric 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 [28], where the fingerprint features are, in a first time, transformed in a real-valued vector of fixed length to generate the biometric template (this step is not useful and not described in this paper). The BioHashing algorithm transforms the biometric template
\[ T = (T_1, \ldots, T_n) \]

in a binary template \[ B = (B_1, \ldots, B_m) \]

with \( m \leq n \), as following:

\section*{Algorithm 1 BioHashing}

1: \textbf{Inputs}
2: \( T = (T_1, \ldots, T_n) \): biometric template,
3: Seed: secret
4: \textbf{Output} \( B = (B_1, \ldots, B_m) \): BioCode
5: Generation with the Seed of \( m \) pseudorandom vectors \( V_1, \ldots, V_m \) of length \( n \),
6: Orthogonalize vectors with the Gram-Schmidt algorithm,
7: for \( i = 1, \ldots, m \) do compute \( h_i = \frac{1}{n} \sum_j \arctan(b_j/V_i) \).
8: Compute BioCode:

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

where \( \tau \) 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 the diversity and revocability properties.

BioPhasor was proposed by Teoh et al. in [29] and it was introduced as a form of cancellable 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 [13], [14], [26]. A main reason is that the projection matrix is only related to the secret. If an impostor obtains the Key (known as stolen token attack), the attack is quite easy especially by combining it with the zero effort one (use of a biometric data belonging to another user). Indeed, with the knowledge of the Key, the impostor will use the same projection basis that 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 aspect.

\section*{IV. PROPOSED METHOD}

We intend in this paper to enhance the security of the BioHashing algorithm by limiting the impact of attacks based on the stolen token. We modified the projection matrix in order to be computed from the biometric template and not only from the secret. 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. In the proposed method, we use:

- A constant number \( C \) that permits us to define a trade-off between performance without any attack and robustness to attacks.
- A polynomial \( P_k(X) = \sum_{i=0}^{3}a_i X^i, k = 1 : m, a_i \in [-C.C, C.C], a_3 \neq 0 \). The choice of a polynomial of degree 3 is well justified. In general, a \( k \)-wise independent function family is presented by a random polynomials of degree \( k - 1 \). In this context, Hoory et al. in [8] have presented an educated conjecture that \( 4 \)-wise independence suffices to achieve pseudorandomness. That is why, in our case, we use a 4-function family, so a polynomial of degree 3.

In the following figure IV, we detail the general principle of the proposition of the BioHashing improvement and the construction of the projection matrix computed from the seed and from biometric data. The proposed algorithm of our method is given in detail 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. On the other hands, we construct the projection matrix \( P \).
Fig. 2. General principle of the proposed method.

**Algorithm 3 Proposal**

1. **Inputs**
   - $T = (T_1, \ldots, T_n)$: biometric template
   - $\mu = E[T]$: average value of biometric templates
   - $\sigma = \sigma[T]$: standard deviation of biometric templates
   - $C$: constant
   - Seed: secret
2. **Output** $B = (B_1, \ldots, B_m)$: BioCode
3. Compute $\tilde{T} = (T - \mu) \ast \sigma$
4. Generation with the Seed of $m$ polynomials $P_k$ with $k = 1 : m$
5. Evaluate $P_{k,i} = \cos(P_k(T_i))$, $\forall i = 1 : n, \forall k = 1 : m$
6. Orthogonalize the matrix $P$ with the Gram-Schmidt algorithm,
7. for $i = 1, \ldots, m$ do compute $x_i = \langle T, P_i \rangle$.
8. Compute BioCode:
   
   $B_i = \begin{cases} 
   0 & \text{if } X_i < \tau \\
   1 & \text{if } X_i \geq \tau,
   \end{cases}$

   where $\tau$ is a given threshold, generally equal to 0.

by the evaluation of these polynomials at each point given by the biometric template; which produces the first method for biometric template protection by a transformation which depends both on a secret and of a biometric template itself.

V. Validation

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

A. Dataset

In order to study the performance and robustness of the proposed transformation, we selected two 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, we obtain in this case 512 real valued features for each fingerprint.

![Fig. 3. Some examples of fingerprint images in the FVC2002DB1 dataset.](image)

The second dataset is the PolyHK which is composed of images of knuckle prints [30] (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, we obtain 256 real valued features for each finger knuckle print.

B. Evaluation

Few works have been dedicated to the evaluation of such biometric systems in the literature [1], [16], [31]. 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 [15], [24]:

- **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.

- **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 [19]). 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:

  \[ \text{FAR}_A(\epsilon) = P(D_T(f(T_z, K_z), A_z) \leq \epsilon) \]  

  Where \( \text{FAR}_A \) is the probability of a successful attack by the impostor for a decision threshold set to \( \epsilon \). 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:**
  It should be infeasible to cross correlate two protected templates.

Based on some of the early works [22], [4] 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 [9], [16], [25], [27]). 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:

- **Zero effort attack:**
  an impostor user \( x \) provides its biometric feature \( T_x \) and parameter \( K_x \) to be authenticated as the user \( z \):
  \[ A_z = f(T_x, K_x) \]

- **Brute force attack:**
  An impostor tries to be authenticated by trying different random values of \( A: A_z = A \)

- **Stolen token attack:**
  An impostor has obtained the token \( K_z \) 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 \( T_z \) (directly or after computation of the feature on a biometric raw data) and tries different random numbers \( K \) to generate: \( A_z = f(T_z, K) \)

- **Worst case attack:**
  An impostor user \( x \) provides its biometric template \( T_x \) and parameter \( K_x \) to be authenticated as the user \( z \) (zero effort attack) and has also obtained the token \( K_z \) of the genuine user \( z \) to generate: \( A_z = f(T_x, K_z) \)

- **Listening attacks:**
  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 \( z \):
    \[ B_z = \{ f(T_z, K_z^{1}), ..., f(T_z, K_z^{Q}) \} \]
  
  - Prediction of a possible BioCode value by setting the most probable value of each bit given \( B_z \), \( \Rightarrow \) computing the \( \text{FAR}_A \) value for \( Q = 3 \) and \( Q = 11 \)

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

### C. Experimental results

In this section, we present the experimental results we obtained with the 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 the fingerprint dataset, we obtained an Equal Error Rate (ERR) of 28.3% and 23.8% for the Finger knuckle print one. We can see clearly that these results are poor and such a biometric system could not been used in real conditions. Transformation protection schemes are known to increase performance if the secret is unknown by impostors.

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 yellow curve presents the evolution of the false rejection rate (FRR) depending on the decision threshold (\( \epsilon \) in equation 2). Other curves correspond to the \( \text{FAR}_A \) value for all considered attacks.
Fig. 5. Security analysis of the BioHashing algorithm on the fingerprint database (a) and the finger knuckle print (b).

For the two datasets and then for the two biometric modalities, the BioHashing and BioPhasor algorithms provide a perfect performance without any attack i.e. when the secret is unknown by the impostor. For the proposed transformation, it is not always the case, in fact, it depends on the value of $C$. When $C$ is set to 10, the EER value equals to 2%. 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. We will see later that we gain of course in term of robustness. We will also show later a study of the impact of the $C$ value on performance and robustness of the proposed transformation scheme.

We then analyze the robustness of the protection schemes related to tested attacks. Of course, the brute force attack (curve in blue) is the less successful attack 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 (curve in green). Only the worst case attack is completely possible ($FAR_A = 100\%$) for these two protection schemes when the $\epsilon$ 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 its computation of the projection matrix. In order to better identify the robustness behavior of these three protection schemes, we propose to plot the FRR value (as a performance indicator) face to the $FAR_A$ one of the worst case (as a robustness indicator in the worst case) in Figure 8. There is a compromise to find between performance and robustness in such transformation schemes. This curve demonstrates that for a given performance (here the FRR value), the robustness evaluated by the $FAR_A$ value is lower for the proposed method.

Finally, we present in Table I the results of the study concerning the impact of the $C$ value in the proposed scheme for the fingerprint dataset. We can see clearly that a higher value of $C$ permits to obtain a better performance but also a lower robustness in the worst case attack. As for example, for a value of $C$ equals to 50, the performance is very good with an EER value of 0.4% and a $FAR_A$ value of 99%. With the BioHashing and BioPhasor algorithms, the $FAR_A$ value was
Perspectives of this work are multiple. First, it concerns the proposal of other combination techniques of the biometric data and the secret (step 10 of the proposed algorithm). Second, we could consider its extension to multi-biometrics where a biometric modality could be used to generate the projection matrix given the secret and the features of the second modality could be projected with it. 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.

Acknowledgements. The authors thank Kevin Atighehchi for comments on this paper.

REFERENCES

Fig. 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.

### TABLE I

<table>
<thead>
<tr>
<th>C value</th>
<th>EER (without attack)</th>
<th>zero effort brute force</th>
<th>stolen token</th>
<th>stolen biometric</th>
<th>worst case</th>
<th>listening N=3</th>
<th>listening N=11</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>7%</td>
<td>6.3%</td>
<td>0%</td>
<td>4.0%</td>
<td>6.7%</td>
<td>94.1%</td>
<td>5.4%</td>
</tr>
<tr>
<td>20</td>
<td>0.6%</td>
<td>0.8%</td>
<td>0%</td>
<td>0%</td>
<td>5.7%</td>
<td>99.3%</td>
<td>0.3%</td>
</tr>
<tr>
<td>50</td>
<td>0.4%</td>
<td>0.3%</td>
<td>0%</td>
<td>0%</td>
<td>0.9%</td>
<td>99%</td>
<td>0.1%</td>
</tr>
</tbody>
</table>


