How to Evaluate Transformation Based Cancelable Biometric Systems?

Rima Belguechi, Estelle Cherrier, Christophe Rosenberger

How to Evaluate Transformation Based Cancelable Biometric Systems?

(Submitted to IBPC 2012)

Rima Belguechi, Estelle Cherrier and Christophe Rosenberger
Université de Caen Basse-Normandie, UMR 6072 GREYC, F-14032 Caen, France
ENSICAEN, UMR 6072 GREYC, F-14050 Caen, France
CNRS, UMR 6072 GREYC, F-14032 Caen, France
Email: christophe.rosenberger@ensicaen.fr

I. Introduction

The concept of *cancelable biometrics* has been defined for the first time in the pioneering article [17]. It is aimed at enhancing privacy protection and template security, as detailed in the recent reference [8]. Two main approaches can be distinguished dealing with cancelable biometrics. On the one hand, biometric cryptosystems or secure sketches, such as those presented in [9], [7], [24], [6], [5], [2], resort to cryptography. On the other hand, we find feature transformations approaches. 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 [23], [3], [18]. In order to validate their proposition, authors generally provide some experimental results based on performance evaluation (EER value, ROC curves, *etc.*) and sometimes through a security analysis by considering different scenarios [19]. None standard methodology has been defined in order to qualify these privacy by design biometric systems [21] even if some previous research works have been proposed recently [15].

We intend in this paper to clearly define the properties that are requested for the evaluation of cancelable biometric systems, and we propose different attacks that can be simulated to assess how the targeted system fulfills these properties. The plan of the paper is the following. Section 2 first gives an overview of definitions. We then list security and privacy properties in the state of the art for the evaluation of cancelable biometric systems. We present different attacks from the impostor point of view in order to assess the previous properties. Some measures are also given to complete this security and privacy analysis of a transformation based cancelable biometric system. We conclude and give some perspectives in section 4.

II. Definitions and properties

In the sequel, we focus on BioHashing since some weaknesses have been reported in the former approach in [10], [12], [20]. We suppose having a biometric modality where the template is represented by a vector of real values (it can be generalized to any representation like a map of interested points).

We use the following notations like in the paper [15]. Let b_z and $b_z^{'}$ represent the template and query biometric features of user z, respectively. Let f be the feature transformation function. We denote n the dimension of the $f(b_z)$ biocode for user z. Let K_z be a set of transformation parameters corresponding to user z. Let D_O denotes a distance function between the biometric features in the untransformed (original) domain and D_T be 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 threshold denoted as ϵ:

$$R_z = 1_{\{D_T(f(b_z,K_z),f(b_z^{'},K_z))\leq\epsilon\}} \quad (1)$$

Very few works have been dedicated to the evaluation of such biometric systems in the literature [1], [25], [15]. The ISO/IEC 24745 "Information technology Security techniques Biometric information protection" defines the security properties of a biometric system, we use the same terms. Cancelable systems must fulfill several properties as also mentioned in [14]:

- *Revocability/Renewability:*
  It should be possible to revoke a biometric template and to generate a new one from the original data. Given the biometric template of user z, through a transformation based cancelable biometric system, it is possible to compute one biocode $f(b_z, {K_z}^1)$ given parameters ${K_z}^1$ and to revoke it by computing $f(b_z, {K_z}^2)$ with other parameters ${K_z}^2$. As only the reference biocode is stored, the revocability can be achieved easily.

- *Performance:*
  The template protection shall not deteriorate the performance of the original biometric system. As the performance is related to the security of the authentication process (*e.g.*, minimizing the number of false acceptance), a cancelable biometric system must be as efficient as possible. For transformation based cancelable biometric

systems, the seed (contained in K_z for user z) can be seen as an *a priori* information (or a secret key). For this reason, a gain of performance is expected. To assess the efficiency of a biometric system (without any transformation), we generally consider two error metrics:

$$FRR_O(\epsilon) = P(D_O(b_z, b_z^{'}) \leq \epsilon) \quad (2)$$

$$FAR_O(\epsilon) = P(D_O(b_z, b_z^{'}) > \epsilon) \quad (3)$$

Where FRR_O is the false reject rate and FAR_O is the false accept rate of the original biometric system (without any template protection). For a transformation based cancelable biometric system, we consider the two following metrics:

$$FRR_T(\epsilon) = P(D_T(f(b_z, K_z), f(b_z^{'}, K_z)) \leq \epsilon) \quad (4)$$

$$FAR_T(\epsilon) = P(D_T(f(b_z, K_z), f(b_z^{'}, K_z)) > \epsilon) \quad (5)$$

Where FRR_T is the false reject rate and FAR_T is the false accept rate of the cancelable biometric system (with template protection).

- *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 [16]). 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:

  $$FAR_A(\epsilon) = P(D_T(f(b_z, K_z), A_z) \leq \epsilon) \quad (6)$$

  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. This is an important property for privacy issues as it avoids the possibility to trace an individual based on the authentication information. Let be $B_z = \{f(b_z, {K_z}^1), .., f(b_z, {K_z}^Q)\}$ a set of Q generated biocodes for user z and ${K_z}^i$ the set of parameters for user z for the ith revocation, it shall constitute a random sub-sampling of $\{0,1\}^Q$. This property prevents also the linkage attack consisting in using different biocodes of an user to predict an admissible one. This is related to an attack consisting in for an impostor to listen different realizations of biocodes for the same user.

These properties are well known and often cited in papers from the state of the art. We go further in this paper: given a cancelable biometric system, how can we verify if these properties are fulfilled ? Is it possible to quantify the risk associated to the feasibility of an attack limiting one of these properties ? We propose in the next section some measures and attacks to answer these two questions.

III. Security and privacy analysis

Based on some of the early works [17], [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 [22], [8], [15], [18]). 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...).

Note that the following study requires that the decision threshold ϵ to be fixed. In this paper, we set the decision threshold ϵ_{EER_T} to the EER value of the cancelable biometric system (without any template protection). Even if this functioning point of the biometric system has no operational meaning, it is often used and can always be estimated. Different other values can be used for ϵ depending on the security requirements of the application. In order to quantify the robustness of the studied cancelable biometric system, we suppose having a biometric database with multiple biometric samples for each user. Some samples permit to generate the biometric template of each user while the others are used for the tests. We detail below how we quantify if the properties previously mentioned are fulfilled: eight criteria are described below. For each criterion, a value $A_i \in [0,1]$, $i = 1, \ldots, 8$ is computed on the BioHashing scheme (there is at least one criterion for one required property). These risk values will be gathered in an eight-dimensional vector to characterize the evaluation of the studied cancelable biometric scheme.

The security and privacy analysis can be done for the two classical steps in biometrics. We first focus on an *authentication* problem (one against one matching): we develop different attack scenarios that an intruder would manage to impersonate a particular genuine user. In a second step, the *identification* (one against many matchings) problem is considered. In this case, the impostor tries to impersonate one of the individuals in the database.

A. Authentication

1) Performance (A_1):
   To verify if the efficiency is not decreased by using the template protection scheme, we propose to compute the following measure:

   $$A_1 = 1 - \frac{\mathrm{AUC}(\mathrm{FAR_T}, \mathrm{FRR_T})}{\mathrm{AUC}(\mathrm{FAR_O}, \mathrm{FRR_O})} \quad (7)$$

   where AUC denotes the area under the ROC curve for both systems. Many cases are interesting to consider.

First, it may happen that $A_1 = 1$ meaning that the cancelable biometric system provides a perfect performance (without any error or $EER = 0\%$). Second, if the value A_1 is negative, it means that the efficiency of the biometric system is deteriorated by the template protection scheme. Otherwise, the scheme improves performance as expected.

2) Non-invertibility or Irreversibility (A_2 to A_5):
This essential property can be evaluated through different attacks. For all these attacks, we use one or multiple biometric samples to generate an admissible query $\acute{b_z}$ of the user z. Based on the principle of each attack, we generate many fake attempts A_z of the genuine user (as described in equation 6):

- *Zero effort attack* (A_2):
an impostor user x provides its biometric feature $\acute{b_x}$ and parameter K_x to be authenticated as the user z : $A_z = f(\acute{b_x}, K_x)$

- *Brute force attack* (A_3):
An impostor tries to be authenticated by trying different random values of A: $A_z = A$

- *Stolen token attack* (A_4):
An impostor has obtained the token K_z of the genuine user z and tries different random values b to generate: $A_z = f(b, K_z)$

- *Stolen biometric data attack* (A_5):
An impostor knows $\acute{b_z}$ (directly or after computation of the feature on a biometric raw data) and tries different random numbers K to generate: $A_z = f(\acute{b_z}, K)$

To evaluate the efficiency of these four attacks, we propose to compute for each of them, the following criteria:

$$A_i = \mathrm{FAR_A}(\epsilon_{\mathrm{EER_T}}),\ i = 2, \ldots, 5 \tag{8}$$

Indeed, from the impostor point of view, the FAR is the relevant value: the intruder has to generate $f(\acute{b_z}, K_z)$ using different available data (K_z, $\acute{b_z}\ldots$). Recall that the threshold has been fixed to the value ϵ_{EER_T} (obtained by computation of the EER of the cancelable biometric system).

From the impostor's point of view, the values A_i, $i = 2, \ldots, 5$ must be as high as possible.The obtained value for each attack A_i, $i = 2, \ldots, 5$ allows us a ranking of the different attacks and directly gives the risk for the system that an impostor can be authenticated as a genuine user.

3) Diversity or Unlinkability (A_6 to A_8) :
A prominent feature of a cancelable biometric system is its ability to produce different biocodes for the same individual and for different applications. The first criterion we want to assess concerns the unlinkability property for privacy issues.

- *Mutual information of biocodes:*
In order to measure the diversity property in this case, we propose to compute the mutual information provided by several biocodes issued from same biometric data as defined in (9):

$$I(X,Y) = \sum_x \sum_y P(x,y) \log(\frac{P(x,y)}{P(x)P(y)}) \tag{9}$$

where X and Y are two random variables and P the estimation of the probability. In order to measure the diversity property, we quantify the highest value of the mutual information among different biocodes for each individual. The value A_6 is then computed using the average value for all users of the highest value of mutual information, according to equation 10:

$$A_6 = \frac{1}{N} \sum_z \sum_{j=1}^{M} \max(I(f(b_z, K_z), f(b_z^j, K_z))) \tag{10}$$

Where:
- b_z: denotes the biometric template of the individual z in the database,
- b_z^j: denotes the j^{th} biometric query of the individual z in the database,
- N: the number of individuals in the database,
- M: the number of generated biocodes for each individual,
- P: the estimation of the probability.

The second criterion we want to assess concerns the unlinkability property for security issues.

- *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 biocode by predicting an admissible value (as for example by setting each bit to the most probable value). These attacks are tested given by the following process:
  - Generation of Q biocodes for user z: $B_z = \{f(b_z, {K_z}^1), .., f(b_z, {K_z}^Q)\}$
  - Prediction of a possible biocode value by setting the most probable value of each bit given B_z,
  - Computation of equation (8).
  $\Rightarrow A_7$ value for $Q = 3$ and A_8 for $Q = 11$

Of course, more complex prediction methods of the biocode given B_z could be proposed. This is one perspective of this work. An evolution of the efficiency of this attack (depending on the evolution of Q) may be used to predict how many interceptions are necessary for the intruder to achieve an authentication.

These criteria allow us to quantify the robustness of cancelable biometric verification systems.

B. Identification

For identification, the properties are the same but the criteria computation are slightly different. First, the values $A_i, i = 1, 6$ are the same for the authentication and identification contexts. Concerning the non-invertibility and irreversibility properties, equation (6) must be modified to:

$$FAR_A(\epsilon) = \max_z P(D_T(f(b_z, K_z), A_z) \leq \epsilon) \qquad (11)$$

Indeed, in the identification case, the impostor tries to impersonate all individuals in the database. Attacks are similar but quantified in a different way. The computation of $A_i,\ i = 2 : 5, 7 : 8$ is realized considering equation 11.

As a conclusion of the proposed methodology, the security and robustness of a cancelable biometric system are characterized by an eight-dimensional vector $(A_i, i = 1, \ldots, 8)$. The key benefit of this quantitative presentation is to allow easily the comparison of cancelable biometric systems.

IV. Illustrations

We illustrate this methodology on two biometric modalities for an authentication application: Finger knucle print and fingerprint. We use generic features based on Gabor filtering on the raw images. We obtain 128 parameters (16 scales, 8 orientations) for each image. We use a single enrollment process for both biometric systems. Without any template protection, we compute a simple distance (measured with the Minkowski's one) between the reference template b_z and $b_z^{'}$ for each user z. Of course, these biometric systems are far to be the best in the state of the art, they can be seen as generic systems. In this study, we are interested in quantifying the robustness of these systems with template protection using the classical BioHashing process (size of the biocode: 128 bits).

In this paper, we use two biometric benchmarks:

- PolyU FKP Database [11]: 4 fingers of 165 volunteers, each individual has provided 12 images acquired during 2 sessions,
- FVC2002 benchmark [13] dB3: composed of 8 fingerprints (resolution 355 x 390 pixels) for 100 individuals.

Table I presents the values of the different criteria $A_i,\ i = 1..8$ describing the robustness of the studied cancelable biometric systems. These results correspond to the average value of each criterion for 10 tries. It is not possible to compare results from these two biometric systems as the used benchmarks are different. But, it is possible to quantify which attack is the most efficient.

First, we remark that the protection scheme as expected increases the performance of the two biometric systems ($A_1 > 0$). For the fingerprint modality and for this benchmark, we obtain an EER value equals to 0%. Using fingerprints, none attack is efficient. If the threshold is set to a higher value, all attacks would be efficient (as it can be seen in Figure 1). It is interesting to notice that the less efficient attack is the one using the less a priori information (brute force). The most efficient attacks (represented by a curve the most on the left) are the stolen token and listening (N=11) ones. As the performance of the cancelable biometric system using finger knuckle print is lower, the threshold value to obtain the EER is higher. In this case, all attacks are possible with the highest value of the risk equals to 59%. That shows the importance of the setting of the threshold value for efficiency but also for security and privacy reasons.

V. Conclusion and perspectives

We propose in this paper a new methodology to quantify how security and privacy properties of a cancelable biometric system are verified. Eight criteria are proposed to quantitatively measure the robustness properties detailed in this paper. They allow a rigorous comparison of cancelable biometric systems. The key benefit of the retained quantitative-based approach is to easily allow the comparison of new cancelable biometric systems. The perspectives of this work are to compare the obtained results with the assessment of expert in biometrics for the privacy analysis of different cancelable biometric systems. We think also on new attacks and on more sophisticated approaches to generate a fake biocode given all the known information.

References

[1] A. Adler, *Biometric system security*, ser. Handbook of biometrics. Springer ed., 2007.

[2] M. Barni, T. Bianchi, D. Catalano, M. D. Raimondo, R. D. Labati, P. Failla, D. Fiore, R. Lazzeretti, V. Piuri, A. Piva, and F. Scotti, "A privacy-compliant fingerprint recognition system based on homomorphic encryption and fingercode templates," in *BTAS 2010*, 2010.

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

[4] R. Bolle, J. Connell, and N. Ratha, "Biometric perils and patches," *Pattern Recognition*, vol. 35, no. 12, pp. 2727–2738, 2002.

[5] J. Bringer, H. Chabanne, M. Izabachène, D. Pointcheval, Q. Tang, and S. Zimmer, "An application of the Goldwasser-Micali cryptosystem to biometric authentication," in *ACISP'07*, ser. Lecture Notes in Computer Science, vol. 4586. Springer, 2007, pp. 96–100.

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

[7] J. Daugman, "Iris recognition and anti-spoofing countermeasures," in *7-th International Biometrics conference*, 2004.

[8] A. Jain, K. Nandakumar, and A. Nagar, "Biometric template security," in *EURASIP Journal on Advances in Signal Processing*, 2008.

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

Biometric systems	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
Fingerprint	1.0	0	0	0	0	0.4384	0	0
Finger Knuckle print	0.1035	0.2473	0.4003	0.5368	0.3730	0.5834	0.5145	0.5908

TABLE I
EVALUATION RESULTS OF THE CANCELABLE BIOMETRIC SYSTEMS.

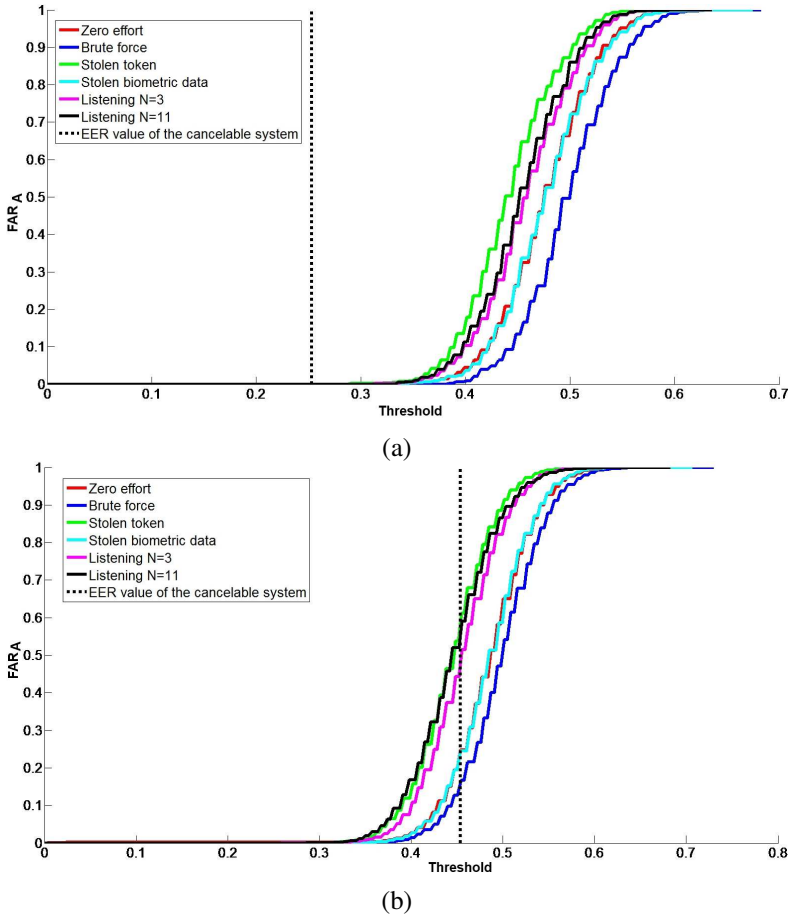

Fig. 1. Evolution of the efficiency of attacks for both biometric systems: (a) Fingerprint, (b) Finger knuckle print

[10] A. Kong, K. Cheung, D. Zhang, M. Kamel, and J. You, "An analysis of biohashing and its variants," *Pattern Recognition*, vol. 39, 2005.

[11] D. Z. Lin Zhang, Lei Zhang, "Finger-knuckle-print verification based on band-limited phase-only correlation," *Proceedings of the International Conference on Computer Analysis of Images and Patterns*, pp. 141–148, 2009.

[12] A. Lumini and L. Nanni, "Empirical tests on biohashing," *NeuroComputing*, vol. 69, pp. 2390–2395, 2006.

[13] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, "Fvc2002: Second fingerprint verification competition," in *International Conference on Pattern Recognition (ICPR'02)*, vol. 3, 2002, pp. 811 – 814.

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

[15] A. Nagar, K. Nandakumar, and A. K. Jain, "Biometric template transformation: A security analysis," *Proceedings of SPIE, Electronic Imaging, Media Forensics and Security XII*, 2010.

[16] D. M. R. Cappelli, A. Lumini and D. Maltoni, "Fingerprint image reconstruction from standard templates," *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 29, pp. 1489–1503, 2007.

[17] N. Ratha, J. Connelle, and R. Bolle, "Enhancing security and privacy in biometrics-based authentication system," *IBM Systems J.*, vol. 37, no. 11, pp. 2245–2255, 2001.

[18] N. Saini and A. Sinha, "Soft biometrics in conjunction with optics based biohashing," *Optics Communications*, vol. 284, no. 3, pp. 756 – 763, 2011.

[19] W. Scheirer and T. Boult, "Cracking fuzzy vaults and biometric encryption," in *Proc. of Biometrics Symposium*, 2007.

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

[21] D. Solove, *Understanding privacy*. Harvard university press, 2009.

[22] A. Teoh, Y. Kuanb, and S. Leea, "Cancellable biometrics and annotations on biohash," *Pattern recognition*, vol. 41, pp. 2034–2044, 2008.

[23] A. Teoh, D. Ngo, and A. Goh, "Biohashing: two factor authentication featuring fingerprint data and tokenised random number," *Pattern recognition*, vol. 40, 2004.

[24] U. Uludag and A. Jain, "Securing fingerprint template: fuzzy vault with helper data," in *Computer Vision and Pattern Recognition Workshop*, 2006.

[25] X. Zhou, S. Wolthusen, C. Busch, and 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.