Texture based Fingerprint BioHashing: Attacks and Robustness

Rima Belguechi, Estelle Cherrier, Christophe Rosenberger

Texture based Fingerprint BioHashing: Attacks and Robustness

Rima Belguechi Estelle Cherrier
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

Abstract

BioHashing is a popular biometric template protection scheme defined in the last decade. Most of previous studies on this algorithm focus on the performance optimization or the use of this privacy protection scheme on many types of biometric modalities (face, fingerprint, palmprint...). The objective of this paper is to study the robustness of this algorithm on a texture based representation by testing and simulating operational attacks. We consider in this study some quantitative measures of the robustness of the BioHashing algorithm and we show some new results on its security on fingerprints represented by texture features.

1. Introduction

Literally, biometrics is the science that measures the characteristics of living beings. In recent years, biometrics more specifically refers to the identification or identity verification of individuals based on a morphological analysis (face, fingerprint, iris...) or a behavioral one (e.g. voice, signature dynamics, keystroke dynamics...) [NIS04], [GEAR09]. Many biometric applications have been developed including border control by a biometric electronic passport, physical access control to secure buildings or fingerprint sensors embedded on laptops for logical access control. However, biometrics presents some risks in terms of compliance rights and fundamental freedoms. The fact of capturing and keeping a raw biometric data may be an invasion of privacy. These data are sensitive and are not yet protected by an international standard. Among the considered solutions, it is possible to create anonymous databases and more generally to incorporate the notion of the privacy respect during the design of a biometric system. Another problem concerns an important violation of privacy: the intrinsic non-revocability of the biometric data. Unlike a password or a PIN code, biometric personal characteristics may not be changed in case of theft or forgery. The concept of *cancelable biometrics* has been defined for the first time in the article [RCB01] as the transformation of raw biometric data, using a chosen function such that the transformed data are safe, cancelable and respect the user privacy. The book [MMJP03] gives the essential characteristics of a convenient cancelable biometric system:

- Revocability: one must be able to easily remove the data if compromised,
- Non-invertibility: from the transformed data, it should not be possible to obtain information on the biometric raw data,
- Performance: the fact that biometric data is cancelable shall not deteriorate the performance of the biometric system,
- Diversity: it should be possible to generate different data for multiple applications.

This paper is organized as follows. We present a brief state of the art in section 2 on solutions for the protection of biometric data. Section 3 describes the general principle of BioHashing and its variations given the biometric modality. A comparison of existing methods is realized. Section 4 presents the protocol and the various attacks to analyze the security of this protection scheme. Experimental results are detailed and discussed. We conclude and give some perspectives of this study in section 5.

2. Background

To protect a biometric template, there are several solutions in the literature. When designing a biometric system, the difficulty to meet all conditions outlined above mainly concerns the natural variability of the biometric signal. Therefore, classical cryptography is not really fitted for this problem.

The transformation functions represent an interesting solution to compensate for the variability of the biometric raw data by directly performing the comparison between the capture and the biometric reference in the transformed domain. Based on Ratha's concept of cancelable biometrics, the pioneering article written by Teoh *et al.* [TNG04] presents the general principle of the *BioHashing* algorithm. This method introduces a distortion of the biometric signal using a chosen transformation function to generate a BioCode. Revocability is guaranteed because, when a BioCode is compromised, one simply needs to change the transformation function. Diversity is also ensured by the choice of different functions for each application. However, finding such functions is not easy. Indeed, besides non-invertibility, these functions must exhibit two important properties: intra-class variability robustness (that is to say a robustness face to the variations of the biometric raw data of an individual) and inter-class sensitivity (it should be possible to distinguish two different individuals given their BioCodes).

3. BioHashing description

The general principle of BioHashing is to generate a binary BioCode (used for the enrolment and verification steps) from the representation of the biometric data (such as texture parameters or minutiae for fingerprints) and a random number [TNG04]. This process is used at the enrolment step (where only the generated BioCode is stored) and at the verification one (where the BioCode is recalculated for each verification and the stored random number is required). The verification result is obtained from the computation of a simple Hamming distance between the reference BioCode and the one issued from the new capture. The advantage of this approach lies in the ability to revoke the BioCode (by applying the same process with a different random number). Another interest lies in the possibility to generate different BioCodes to authenticate oneself to different services from the same biometric raw data (fingerprint by example). Figure 1 illustrates the overall process.

In the sequel, we will focus on fingerprint modality. If we detail the Biohashing process, it first consists in projecting the (normalized) raw biometric data (called FingerCode) on an orthogonal basis generated from the random number. The resulting dimension is at most equal to the dimension of the initial representation of the biometric data. This step somehow amounts to hiding the biometric data in a part of the multi-dimensional space. Using an orthogonal basis ensures the conservation of similarity relationships between the BioCodes, as demonstrated by the lemma Johnson-Lindenstrauss (see reference [DG99]). The second step has for objective to quantize this result using a simple

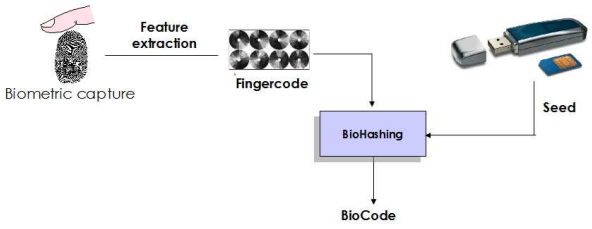


Figure 1. General principle of the BioHashing algorithm

thresholding. This step ensures the non-invertibility of the process (preventing an intruder from finding the raw biometric data from a BioCode) and increases the robustness of the process (by allowing minor differences in the projected vector inherent to the acquisition of the raw biometric data). The general principle is summarized in Figure 2.

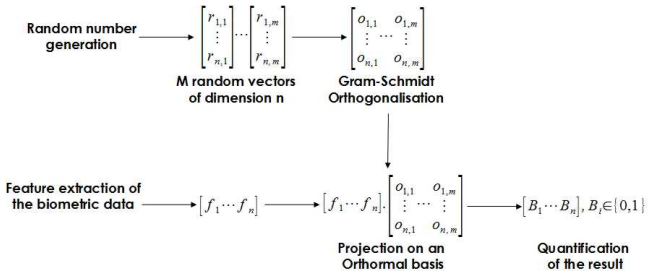


Figure 2. BioCode generation

This two-factor process ensures that it is not possible to retrieve the raw biometric data given the BioCode. To resist the brute-force attack (BioCode prediction of a genuine user), it is necessary to have a representation of the biometric data which provides the largest entropy. Minutiae are commonly used to represent a fingerprint but this representation is too compact in general to be used in this context. Authors in [RB10] proposed to use a representation based on texture features (obtained by Gabor filters) to generate a 384 bit BioCode with the Ratha's method . We use in this paper the same approach that is to say a Gabor filter bank to generate a FingerCode given a fingerprint. This leads to a dimension of representation equal to 128 bits (8 orientations, 16 scales). As detailed previously, the BioHashing method is a generic approach enabling the revocation of raw biometric data. It has been used on multiple biometric modalities (palmprint [CTGN04], face [LN07], fingerprints [NNJ09], finger knuckle prints [BCAR11]...). The issue of protection of biometric data was often discussed in a surprising way considering performance (i.e. error rate minimization and maximization

of the BioCode size). Apart from some works, including [KCZ+06, LCM09, SLMM09, Nag10, KTT10, BCAR11], there are very few studies focusing on the robustness of these algorithms, especially including the definition and test of attacks scenarios. The main contributions of the paper are:

- a rigorous study of the BioHashing scheme using textural features,
- the proposition of a new attack based on multiple listenings,
- the study of the identification case (to estimate if an attacker can be identified as one genuine user).

4. Study of BioHashing robustness

The proposed protocol and the performed attacks are detailed below.

4.1. Protocol

Generally speaking, the performance of biometric systems is equally determined by: the sensor quality, the capture ergonomics and the algorithms performance. In the following, only the third point is dealt with. We suppose the other conditions are fulfilled. When dealing with authentication by password verification, the process is said deterministic in the sense that the output is either positive (if both passwords are found identical), either negative (if both passwords are found different). However, comparing two (transformed or not) biometric data is a statistical process. Indeed, each capture of the same biometric data is different: the verification system has to evaluate a degree of similarity and then, depending on a fixed threshold, has to decide whether the user is authorized or not. Some standard error rates will be used as the FAR (False Acceptance Rate), the FRR (False Rejection Rate) and the EER (Equal Error Rate).

In this paper, we consider the FVC2002 benchmark [FVC] dB3 composed of 8 fingerprints (resolution 355 x 390 pixels) for 100 individuals. The FingerCode of each user is generated following the method presented in section 3, with resort to a Gabor filter bank. Once this is achieved, 8 FingerCodes are available for each person, which means 800 FingerCodes of length 128 to 512 bits. After random projection and quantization, 800 BioCodes are issued. For each person, one BioCode is kept as a reference, the other 7 BioCodes are used to test the different attack scenarios.

4.2. Performance evaluation

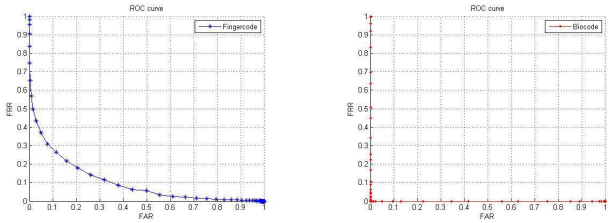
First, the performances of both biometric systems based on the use of the FingerCode and the BioCode respectively are studied through an analysis of their ROC curves.

Test 1 : **FingerCode**

In this case, the biometric system is not cancelable since it uses the raw fingerprint data. The obtained score is measured with the Minkowski distance between the fingerprint kept as reference (*i.e.* the reference FingerCode) and the other fingerprints of the database. The obtained EER is equal to 19%, see figure 3(a). This performance is obviously far from being the best compared to results in the literature. But, this value allows us to set a baseline performance of the system.

Test 2 : **BioCode**

In this case, the performance of a cancelable biometric system is evaluated. Notice that the EER is null as illustrated by figure 3(b), which means that another user in the database will not be able to impersonate another genuine user. This scenario is known as a *zero-effort attack*. The corresponding threshold is equal to 0.32. Compared to the FingerCode performance, this improvement looks very interesting: it is mainly due to the smoothing of the intrinsic intra-class disparity by the random projection (the random value is kept fixed for each user). In other words, that means a reduction of the intra-class variability, leading to better performances. In the sequel, we suppose that the threshold of the BioHashing system remains at this value of 0.32.



(a) FingerCode, EER=19% (b) BioCode, EER=0%

Figure 3. ROC curves

4.3. Robustness evaluation

In this section, we analyze the behavior of the studied system against some attacks.

4.3.1 Description of the different scenarios

Studying BioHashing robustness amounts to evaluate the system performance with respect to:

- unavoidable variations of a biometric data, also called intra-class variations
- the differences between data issued from distinct persons, also called inter-class variations

An ideal cancelable biometric system would exhibit both robustness to intra-class variations and sensibility to inter-class variations. This criterion is very challenging, since these two conditions are quite difficult to conciliate. In this paper, we propose to test some well-known attack scenarios, such as stolen biometric data, stolen-key attack, brute-force attack. But, we also propose to test some new attacks (such as *BioCode correlation attack*) with a view to defining a framework (or a minimum security level) for the further study of future cancelable biometric systems. Let us recall the database configuration: 100 users, 1 reference BioCode for each user (enrolment), 7 other BioCodes for each user (testing). First, known attacks are tested, with an original approach to quantify their operational impact. More precisely, the relevant value we will study is 1-FAR: indeed, given the reference BioCode of a genuine user, the intruder will try to generate an eligible BioCode, using different available data (token, FingerCode...). Recall that the threshold has been fixed to 0.32 to ensure the better performances *i.e.* an EER equal to zero in ideal conditions. Therefore, *the intruder aims at generating a BioCode whose Hamming distance with the reference BioCode is less than* 0.32.

Test 3 : **Brute-force attack**

Now, we suppose that the intruder has no information, more precisely, he/she does not know anything about the FingerCode nor the tokenized random number. The only available information is the length of the binary BioCode. The attack consists in the comparison of randomly issued BioCodes (with a random FingerCode and a random number different each time) to the genuine BioCodes. The figure 4 illustrates the evolution of the rate 1-FAR of the original BioCode and that of the brute-force attack. It can be seen that the brute-force attack 1-FAR is almost always less than 1 for thresholds less than 0.32. There is only a narrow interval in the threshold values for which this value is different from 1: to evaluate the attack performance, one needs to precisely compute the value $1 - FAR$ at the abscissa 0.32, leading to an efficiency of the brute-force attack equal to 0.14%.

Test 4 : **Stolen FingerCode**

In this test, we assume that the intruder has stolen one fingerprint among the 800 in the database. Then, he/she will try to impersonate this authorized user. Then, the intruder is able to generate a FingerCode (we suppose he/she knows how to compute it) and a corresponding BioCode with a random number different from that which has been used to compute the reference BioCode. The figure 4 illustrates the performances of the stolen FingerCode attack. With the same reasoning as in the previous test, the efficiency of this attack is estimated at around 0.14% at the fixed threshold of 0.32. This proves that the knowledge of the FingerCode does not significantly improve the brute-force attack, so the FingerCode is not an interesting information source for the intruder. This is a very good property for privacy.

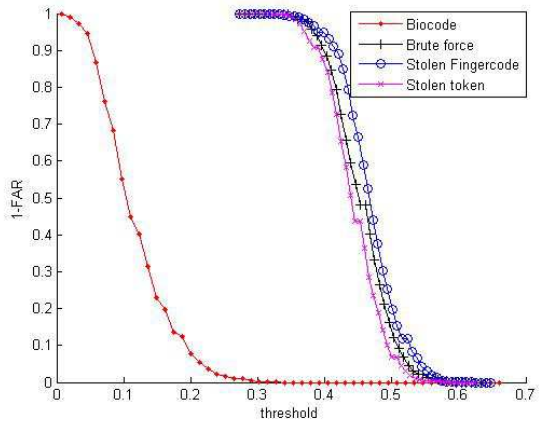


Figure 4. BioHashing robustness analysis against brute-force, stolen FingerCode and stolen token attacks

Test 5 : **Stolen token**

In this test, we suppose that the intruder has stolen the token. In this case, the random number is available, but not the FingerCode. The fake BioCode will be generated from a random FingerCode (in fact from the intruder's fingerprint, considered as random in our numerical simulations). For this test only, the random number is fixed at the same value for all the authorized users of the database. The performances of this attack can be seen at figure 4. Roughly speaking, this attack is again not very effective, since the efficiency rate is estimated at 0.28%. We remark that this rate is twice that of the former attacks. This conclusion underlines the importance of the tokenized random number, and the need for its relevant secure storage.

Contrary to the previous attacks that can be found in the literature (we refer the reader to parts of the works [TKL08] or [Nag10] for example), some new (to the authors knowledge) attack scenarios are detailed now. These attacks ensue from the revocable feature of BioHashing: assume that an intruder has eavesdropped a genuine user so as to collect N different BioCodes $\{B_1, .., B_N\}$ of this user (each BioCode has been generated with a different random number). Then, the intruder generates a new BioCode, from these N interceptions, by statistically setting the bits at the value 0 or 1, depending on the most frequent value among $\{B_1, .., B_N\}$. The following tests aim at analyzing if the information contained in the spurious BioCode is sufficient to be accepted by the system. We chose to study two cases ($N = 3$ and $N = 11$) to analyze if more interceptions (and hence a better statistical analysis) lead to

a better efficiency of this attack. Considering a larger value for N indeed becomes quite unrealistic.

Two readings can be made from these statistical attacks. On the one hand, the value 1-FAR will be examined as in previous tests, to determine if the intruder can impersonate the genuine user corresponding to the intercepted BioCodes. In this case, the attack is called *personal*. On the other hand, the intruder may simply want to impersonate one of the authorized users. In this second case, this attack is called *global*: the ROC curve and the EER must be rather examined.

Test 6 : **Attack by eavesdropping**
In this test, we assume that the intruder has eavesdropped $N = 3$ or $N = 11$ BioCodes of the same user (these BioCodes have been intercepted after revocation). The figure 5 shows that the system is able to resist the personal attack for both values of N. More precisely, the efficiency of the attack does not increase much when N becomes larger: the risk amounts to 0.06% when $N = 3$ and to 0.16% when $N = 11$. The obtained efficiency rates indicate that the personal attack is not the most effective one: the intruder should prefer to steal the token of the user. Even if this is not realistic, more interceptions will lead to a slight improvement of the efficiency of the personal attack.

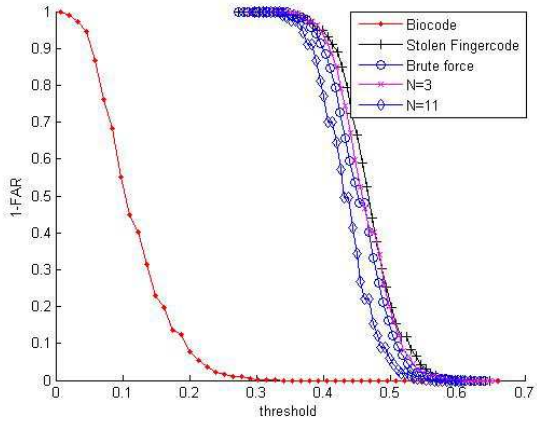


Figure 5. Personal attacks by interception of N BioCodes

Test 7 : **Global attacks**
In this test, assume the intruder only wants to be authorized by the system as one of the 100 genuine users, without a precise identity theft, using N stolen BioCodes. The ROC curves plotted at figure 6 for $N = 3$ and $N = 11$ display an EER equal to 0.5 in both cases. This means that the intruder is likely to be accepted by the system, but without knowing which user he/she impersonates. When the same study of the ROC curves and the EER values is conducted for the other attacks (brute force, stolen token, stolen FingerCode), the same conclusion can be drawn: a global attack is always efficient.

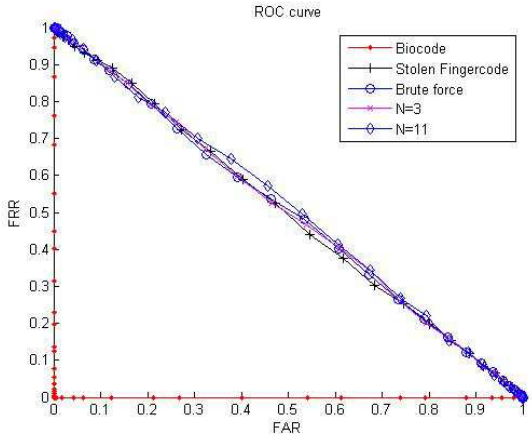


Figure 6. Global attack by interception of N BioCodes

4.4. Synthesis

This section is devoted to a summary of the conducted attacks and a comparison of their efficiency. Considering any personal attack scenario, a very low risk exists, which means a flaw in BioHashing, that will require countermeasures in future works. We compute the same attacks for BioCodes of size 128, 256 and 512 bits. For the three sizes, the FingerCode performance is similar and a perfect performance is achieved using the BioCode in ideal case. In table 2, the obtained efficiency risks of the different attacks are gathered, corresponding to the value of $1 - FAR$ when the threshold is fixed at 0.32 (for a BioCode of size 128 bits). This allows us to classify the scenarios against each other: the most efficient attack corresponds to the *stolen token scenario*, nevertheless the risk is rather weak with 0.28% of success. We tested the same attacks for different sizes of the BioCode (by adapting the number of Gabor filters). For BioCodes higher than 128 bits, none of the attack is operational. Therefore, the size of the BioCode is important for robustness issues.

	128 bits	256 bits	512 bits
FingerCode	19%	18%	17%
BioCode	0%	0%	0%

Table 1. Performance evaluation through the EER value (without attack)

Considering the global attacks, they have been found operational and thus raise privacy issues. Notice that this kind

BioCode size	Brute force	Stolen FingerCode	Stolen token	$N = 3$ interceptions	$N = 11$ interceptions
128 bits	0.14%	0.14%	0.28%	0.06%	0.16%
256 bits	0%	0%	0%	0%	0%
512 bits	0%	0%	0%	0%	0%

Table 2. Attack risk value (1-FAR) for a cancelable biometric system with prescribed threshold ensuring EER $= 0\%$ in ideal case (without attack)

of attack does not allow the intruder to know which user he/she impersonates.

5. Conclusion

Despite active research in recent years in the proposal of protection schemes of biometric data, very few studies have focused on the security and robustness of these protocols. This is however vital in an area such as biometrics which handles highly sensitive data. The main contribution of this paper is to propose an experimental analysis of the robustness of these algorithms for different attack scenarios. We defined a methodology to compare the risk associated to an attack. This work was performed on a popular cancelable algorithm on a significant fingerprint database of the literature. The proposed study has shown a good robustness of the scheme in the context of theft identity. The risk identified for the most severe attack reaches 0.28%. It was also shown that an attacker is able to generate fairly easily a BioCode eligible by the identification system showing a significant weakness of this protection scheme. The prospects of this study are the development of more complex attacks for the identity theft. We intend to analyze the distribution of bits in the BioCode to define heuristic search for the generation of identity theft attacks.

References

[BCAR11] R. Belguechi, E. Cherrier, M. El Abed, and C. Rosenberger. Evaluation of cancelable biometric systems: Application to finger-knuckle-prints. In *IEEE International Conference on Hand-Based Biometrics*, pages 222–227, 2011.

[CTGN04] T. Connie, A. Teoh, M. Goh, and D. Ngo. Palmhashing: a novel approach for dualfactor authentication. *Pattern analysis application*, 7:255–268, 2004.

[DG99] S. Dasgupta and A. Gupta. An elementary proof of the Johnson-Lindenstrauss Lemma, 1999. UTechnical Report TR-99-006, International Computer Science Institute, Berkeley, CA.

[FVC] Fvc2002.

[GEAR09] R. Giot, M. El-Abed, and C. Rosenberger. Keystroke dynamics with low constraints svm based passphrase enrollment. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–6, 2009.

[KCZ+06] A. Kong, K.-H. Cheung, D. Zhang, M. Kamel, and J. You. An analysis of biohashing and its variants. *Pattern Recognition*, 39:1359–1368, 2006.

[KTT10] Y. Kim, A.B.J. Teoh, and K-A. Toh. A performance driven methodology for cancelable face templates generation. *Pattern recognition*, 43:25442559, 2010.

[LCM09] Y. Lee, Y. Chung, and K. Moon. Inverse operation and preimage attack on biohashing. In *IEEE Workshop on Computational Intelligence in Biometrics*, 2009.

[LN07] A. Lumini and L. Nanni. An improved biohashing for human authentication. *Pattern Recognition*, 40:1057–1065, 2007.

[MMJP03] D. Maltoni, D. Maio, A.K. Jain, and S. Prabhakar. *Handbook of Fingerprint Recognition*. Springer, 2003.

[Nag10] K. & Jain A. K. Nagar, A.; Nandakumar. Biometric template transformation: A security analysis. In *Proceedings of SPIE, Electronic Imaging, Media Forensics and Security XII*, 2010.

[NIS04] NIST. Minutiae interoperabilty exchange test 2004, 2004. http://www.nist.gov/itl/iad/ig/minex04.cfm.

[NNJ09] A. Nagar, K. Nandakumar, and A.K. Jain. A hybrid biometric cryptosystem for securing fingerprint minutiae templates. *Pattern Recognition Letters*, 33(8):733–741, 2009.

[RB10] S. Ait Aoudia R. Belguechi, C. Rosenberger. Cancelable authentication based on ngerprints texture. In *ICMOSS Conference proceedings*, 2010.

[RCB01] N.K. Ratha, J.H. Connelle, and R. Bolle. Enhancing security and privacy in biometrics-based authentication system. *IBM Systems J.*, 37(11):2245–2255, 2001.

[SLMM09] S. W. Shin, M.-K. Lee, D. Moon, and K. Moon. Dictionary attack on functional transform-based cancelable fingerprint templates. *ETRI Journal*, 31:628–630, 2009.

[TKL08] A.B.J. Teoh, Y. Kuanb, and S. Leea. Cancellable biometrics and annotations on biohash. *Pattern recognition*, 41:2034–2044, 2008.

[TNG04] A.B.J. Teoh, D. Ngo, and A. Goh. Biohashing: two factor authentication featuring fingerprint data and tokenised random number. *Pattern recognition*, 37:2245–2255, 2004.