Analysis of the Iriscodes Bioencoding Scheme

Patrick Lacharme *patrick.lacharme@ensicaen.fr*
Universite 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

Cancelable biometrics is a technique used to enhance security and user privacy. These schemes are employed to generate multiple revocable data from the original biometric template. In this paper, the security of binary template transformations is evaluated, through a new transformation for iris templates, called bioencoding scheme. This transformation and its security is analyzed, using Boolean functions and non linear Boolean systems. A general discussion on binary template transformations is finally proposed.

Keywords: Cancelable Biometrics, Bioencoding Scheme, Iriscodes

1. INTRODUCTION

Biometric schemes are widely used for identification and authentication. Nevertheless, biometric techniques give rise to many security and privacy concerns, especially because biometric data cannot be revoked. The protection of these sensitive data is a major requirement for the deployment of biometric schemes. Cancelable biometrics is based on a randomized transformation for the generation of a biometric template, from the original biometric data. The additional random number (the *seed*) is used to diversify the template. This seed needs to be carefully stored in the biometric system, and is used during the verification phase. The verification procedure only applies on the transformed template.

Cancelable biometrics is first proposed by Ratha et al. for fingerprint authentication [1]. This concept is later developed on other biometric traits, as iris or face characteristics. This alternative allows the generation of a new biometric template, if the previous template is compromised, or if a new template is required for a new application. Security analysis of these schemes is realized using several properties, as required in [2], [3], [4], including mainly:

- Recognition performance: FAR and FRR of the biometric system do not decrease significantly with the template transformation.
- Non-invertibility: the original template cannot be recovered from a compromised template, even if the random seed is known.
- Unlinkability: the original template cannot be recovered from several compromised templates, even if the corresponding random seeds are known (correlation attack).

For non-invertibility and unlinkability, it should be impossible or computationnaly hard to recover the original template, with the knowledge of the seed. These two properties ensure the user's privacy with the protection of biometric data. Detailed reviews on cancelable biometrics and other biometric cryptosystems are proposed by Jain et al. in [5] and by Rathgeb and Uhl in [6].
A new cancelable biometrics scheme on iriscodes is recently presented by Ouda et al. in [7] [8], called *bioencoding scheme*. The iriscode is partitioned into separate blocks, and each block is treated separately with a pseudorandom sequence. Authors claimed that the performance of their scheme is good, based on experimental results for low block sizes. This scheme ensures diversity and is secure against invertibility. In a second paper, the authors investigated the

unlinkability property of their scheme and found vulnerabilities if several biocodes are compromised, in [9] [10]. Nevertheless, there is no computational evaluation in their analysis.

In this paper, binary template transformations for iriscodes are investigated. The bioencoding scheme and its security are analyzed and criticized. Then, a security proof against correlation attacks is given in relation to non linear Boolean systems. Nevertheless, this security proof is only usable if the block size is high. Thus, the correlation attack is practical for low block sizes. This paper concludes on a discussion of such binary transformations, directly applied on the iriscode.

2. THE BIOENCODING SCHEME

2.1 Iris Cancelable Biometrics

Iris biometrics is known for its very good performance, with low FAR and FRR rates [11] [12]. Iriscode is the most used representation of an iris biometric feature. The iriscode generation is described and improved by Daugman in [13] [14] [15], where a binary vector of 2048 bits is derived from an image of an iris. Hao et al. assume in [16] from 10 to 20 percent of error bits within an iriscode. Several effective constructions of fuzzy commitments schemes on iriscodes are presented with various error correcting codes as the Hadamard and the Reed-Solomon codes in [16], or a Reed-Muller based product code in [17]. Iris cancelable biometrics includes schemes proposed by Chong et al. [18] [19], Zuo et al. [20] or Pillai et al. [21]. More details on iris biometric cryptosystem and cancelable iris biometrics are given in [22]. All these schemes directly work on the iris feature, and not on the binary iriscode, except for the BIN_COMBO and BIN_SALT algorithms of [20]. In these two last schemes, authors propose to use a random secret as a secret permutation or a mask on iriscode. Clearly, diversity and non-invertibility are ensured if the random data is secret. Nevertheless, the iriscode is easily recovered if the secret data is compromised by an attacker in both schemes.

2.2 Description of the Bioencoding Scheme

The bioencoding scheme is a cancelable biometric scheme, with a binary template transformation, applied on the iriscode. In this paper, the iriscode is a n-bits vector. The bioencoding scheme divides the iriscode in n/m blocks of length m and applies a random transformation on each block (more precisely a pseudorandom transformation, defined by 2^m pseudorandom bits, generated from a m-bits random seed). Let S be the pseudorandom sequence of length 2^m which can be made public (or compromised). Let S[i] denotes the i-th bit of the binary sequence S. The transformation uses an address system defined by S and maps independently each of n/m blocks of the iriscode as follows. Each m-bits input block represents a number N between 0 and 2^m-1, and the corresponding output bit is S[N]. Finally, the biocode is composed of the n/m output bits, as illustred in Figure 1.

The bioencoding scheme can be described with random Boolean functions. Let f be a Boolean function with m variables, mapping $\{0,1\}^m$ to {0,1}. A Boolean function is called *balanced* if there are the same number of zeros and ones in its truth table. A Boolean function can also be described with its *algebraic normal form* (ANF), which is just a binary polynomial with m variables.

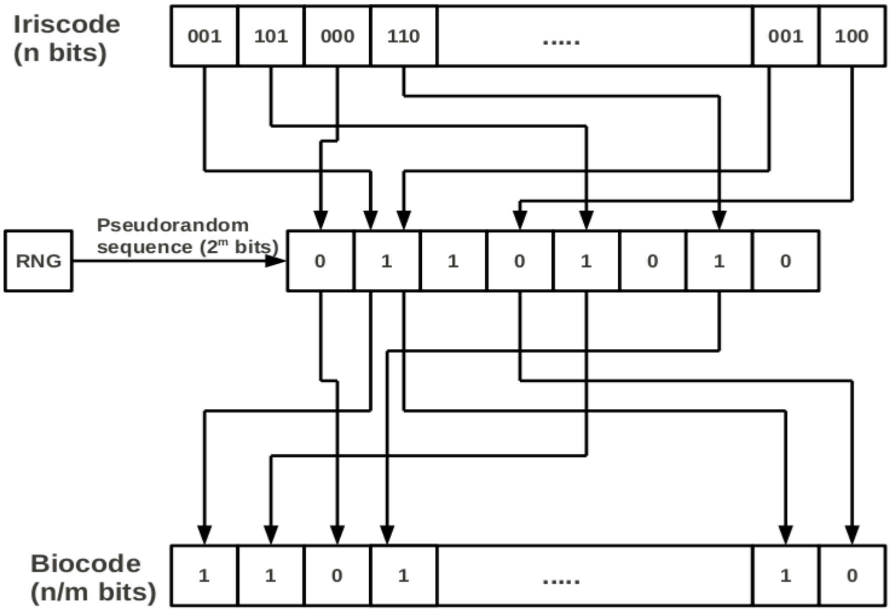

FIGURE 1: The bioencoding scheme.

For example, the balanced Boolean function f with three variables, defined by $f(x_1,x_2,x_3) = x_1+x_2+x_3+x_1.x_2$ mod 2 corresponds to the truth table described in Table 1. The generation of a random Boolean function with m variables requires 2^m random bits for the description of the truth table (or equivalently for coefficients of the ANF polynomial). Consequently, it is not possible to generate a random Boolean function with m variables if the number m is high. More details on Boolean functions can be found in [23].

x_1	x_2	x_3	f(x)
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	1
1	1	1	0

TABLE 1: Truth table of the Boolean function f

The proposed transformation takes m-bits vectors from the original n-bits iriscode and just applies a random Boolean function with m variables, corresponding to the sequence S in order to obtain one output bit. Thus, the function is applied to n/m blocks, and returns a n/m-bits biocode. The address system described by authors is not useful to understand or implement their scheme. For example, the previous Boolean function described in Table 1 corresponds to the pseudorandom sequence described in Figure 1. Thus, the revisited bioencoding scheme is presented below:

- Generate a pseudorandom Boolean function f with m variables from a random seed.
- Divide the n-bits iriscode in n/m blocks of m bits $x^{(0)}, x^{(1)},..., x^{(n/m-1)}$.
- Apply the Boolean function f to each block, such that $b^{(0)} = f(x^{(0)}),..., b^{(n/m-1)} = f(x^{(n/m-1)})$.

- Output the n/m-bits biocode $(b^{(0)},...,b^{(n/m-1)})$.

In the rest of this paper, the bioencoding scheme is described with the Boolean functions terminology. This scheme is only related with two parameters, the size n of the original binary template, divided in block of size m and the (public) Boolean function f, applied to each block.

2.3 Performance of the Bioencoding Scheme

Experimental results on the bioencoding scheme are described in [8] for very small values of m, using the CASIA iris database v_1 and later v_3. The Hamming distance d_{bio} between two biocodes, derived from two iriscodes, is clearly lower than the Hamming distance d_{iris} between the two original iriscodes. However, the length of the biocodes is divided by m. Consequently, the Hamming distance d_{bio} should verify $m.d_{bio} < d_{iris}$, in order to ensure the performance requirement. More precisely, the intra-class variability requires that f(x) + f(x') mod 2 is zero for pairs x, x' with low Hamming distance. For a Boolean function f, the *derivate* of f in a, denoted $D_a(f)$, is defined by the Boolean function $D_a(f)=f(x) + f(x + a \bmod 2) \bmod 2$. Therefore, intra-class variability requires that most of derivates of the Boolean function are zero for elements with low Hamming weight. Unfortunately, there is no general construction for such Boolean functions. Following [16], the percent of error bits in two genuine iriscodes is between 10 and 20. It is not possible to ensure a correct performance requirement, concerning the intra-class variability, with a random balanced Boolean function.

3. NON-INVERTIBILITY AND UNLINKABILITY

3.1 Non Invertibility

Template transformations are designed to produce biometric templates, from which it is computationally hard it is computationally hard to recover the original template, even with the knowledge of the seed [5]. The irreversibility of the bioencoding scheme is based on the compression rate of the Boolean function. If this function is balanced, there are 2^{m-1} inputs for each output bits, providing $2^{(m-1)n/m}$ possible inputs. Consequently, the original iriscode cannot be recovered from one compromised template by an impostor, having the knowledge of the Boolean function.

However, it must be computationally hard to find a biometric template, matching with the given template [24]. This criteria is different to the standard noninvertibility criteria. The construction of another preimage is related to the security of the scheme against spoofing atacks, as in [25] and [2] for the biohashing algorithm. For a given biocode and the knowledge of the Boolean function, it is easy to construct an iriscode which provides the same biocode (directly from the truth table). Consequently, the bioencoding scheme is not protected against spoofing attacks, if the Boolean function is known.

3.2 Unlinkability

The correlation attack proposed by Ouda et al. comes from a basic example with m = 3, where three compromised biocodes are sufficient to recover the original iriscode. This attack is described here with the Boolean functions terminology, using three Boolean functions f_1, f_2, f_3 defined in Table 2.

x_1	x_2	x_3	$f_1(x)$	$f_2(x)$	$f_3(x)$
0	0	0	0	0	1
0	0	1	1	1	0
0	1	0	1	0	0
0	1	1	0	1	1
1	0	0	1	1	1
1	0	1	0	1	0
1	1	0	1	0	1
1	1	1	0	0	0

TABLE 2: Truth table of the Boolean functions f_1, f_2, f_3

Authors consider an example where the first bits of three biocodes, defined by f_1, f_2 and f_3, are 0, 1 and 0. In Table 2, there is only one input x such that $f_1(x) = 0$, $f_2(x) =1$ and $f_3(x) = 0$, then the first block of the orignal iriscode is x = 101. This attack is generalized for all m, by claiming that if m biocodes are compromised, then the original iriscode is recovered. However, the security of a system is generally related to the computational hardness of a given problem, as for NP problems. In this case, there are no description on the method to recover the original iriscode in general case, and the complexity of the attack is not estimated.

The correlation attack is revisited by a Boolean system, in order to evaluate the resistance of the scheme to this attack. Considering that m biocodes $b_1,..., b_m$ are compromised, with m Boolean functions $f_1,..., f_m$. Then, the correlation attack requires the resolution of n/m Boolean systems where the unknown is one of m-bits block of the original template. For example, let $x=(x_1,...,x_m)$ be the first block of the original iriscode and $b_j^{(0)}$ denotes the first bit of the j-th biocode. The following Boolean system must be solved:

$f_1(x_{1,}... ,x_m)=b_1^{(0)}$

...

$f_m(x_1,... ,x_m)=b_m^{(0)}$

A similar Boolean system exists for each bits of biocodes. If the Boolean functions $f_{1,}..., f_m$ are linear and linearly independent, then this system is easily invertible. But the probability that m random Boolean functions are linear is very low. Otherwise, if the Boolean functions are non linear, the resolution of this system is known as a NP problem, providing a security proof on the hardness to realize a correlation attack on this scheme. Nevertheless, if we want to use this NP problem, the number m can not be too small, implying a very high performance degradation.

3.3 Additional Modifications and Discussion

Additional modifications in [7] include a preliminary operation on the iriscode, involving a second random number, before the bioencoding transformation. The first proposition performs the bit-wise XOR between this random number and the iriscode. Nevertheless, the security of this scheme is only related to the secret of the random number. The second proposition uses a secret permutation on the n bits of the iriscode. This proposition is more interesting because the random permutation has not to be secret to ensure the security of biocodes. In this case, the correlation attack is determined by Boolean systems, where the n original bits are possibly involved. Consequently, this attack becomes computationally unfeasible, considering the size of n, even if all permutations are known. Moreover, the random permutation ensures the biocode diversity. Thus, the Boolean function has not to be random and can be determined to optimize the intra-class variability. It is a strong improvment compared to the original scheme, where the Boolean function was generated at random in [8].

Unfortunately, the non-invertibility property is not verified. For a given biocode it is easy to reconstruct a new iriscode which is tranformed to the same biocode with the knowledge of the Boolean function and the permutation. A protection against spoofing attacks requires that the Boolean function (or the permutation) is not compromised. A similar vulnerability is realized by Nagar et al. for fingerprint and face biometrics in [2]. It is the reason why the non-invertibility property ensures that the construction of another preimage should be hard, as suggested in [24]. Consequently, tokenless template transformations should be very carefully designed, especially in the iriscode context, and the protection of the random token in an additional secure element is recommended for many applications.

4. CONCLUSION AND PERSPECTIVES

Binary template transformations are investigated in this paper, through the bioencoding scheme on iriscodes. This scheme is revisited with random Boolean functions and the performance of the transformation is analyzed. Thus, the bioencoding system appears to be a simple application of a random Boolean function on the original iriscode, realized block by block. The bioencoding scheme cannot ensure functional performance requirements for general block sizes. The protection of the original template is related to a Boolean system, possibly enhanced with a random permutation. However, this scheme is invertible because a lot of preimages can be reconstructed from a biocode.

The perspective of this work would be to provide a robust binary template transformation, ensuring a good intra-class variability and a strong preimage resistance in a tokenless environment. Another alternative for iris cancelable biometrics is to transform directly the iris feature, without iriscode transformation.

5. REFERENCES

[1] N. Ratha, J. Connell and R. M. Bolle. “Enhancing Security and Privacy in Biometrics biased Authentication Systems”, IBM Systems, vol 40, N 3, pp. 614-634, 2001.

[2] A. Nagar, K. Nandakumar and A. K. Jain. “Biometric template transformation: A security analysis”, SPIE, Electronic Imaging, Media Forensics and Security XII, 2010.

[3] D. Maio, D. Maltoni, A. Jain and S. Prabhakar. “ Handbook of fingerprint recognition”, Springer, 2009.

[4] R. Belguechi, E. Cherrier and C. Rosenberger. “How to Evaluate Transformation Based Cancelable Biometric Systems?”, IBPC 2012.

[5] A. K. Jain, K. Nandakumar and A. Nagar. “Biometric Template Security”, EURASIP J. Advances in Signal Processing, vol 8, N 2, pp. 1-17, 2008.

[6] C. Rathgeb and A. Uhl. “A Survey on Biometric Cryptosystems and Cancelable Biometrics”, EURASIP J. on Information Security, vol 3, 2011.

[7] O. Ouda, N. Tsumura and T. Nakaguchi. “Tokenless cancelable biometrics scheme for protecting iris codes”, ICPR'10, pp. 882-885, 2010.

[8] O. Ouda, N. Tsumura and T. Nakaguchi. “BioEncoding: A reliable tokenless cancelable biometrics scheme for protecting iriscodes”, IEICE Transaction on Information and Systems, vol E93-D, N 7, pp. 1878-1888, 2010.

[9] O. Ouda, N. Tsumura and T. Nakaguchi. “Securing BioEncoded iriscodes against Correlation Attacks”, IEEE Int. Conference on Communications, ICC'11, pp. 1-5, 2011.

[10] O. Ouda, N. Tsumura and T. Nakaguchi. “On the Security of BioEncoding Based Cancelable Biometrics”, IEICE Trans. on Information and Systems, vol E94-D, N 9, pp. 1768-1778, 2011.

[11] J. Daugman. “Probing the uniqueness and randomness of iris codes: results from 200 billion iris pair comparisons, IEEE, vol 94, N 11, pp. 1927-1935, 2006.

[12] K. W. Bowyer, K. Hollingsworth and P. J. Flynn. “Image understanding for iris biometrics: A survey”, Computer Vision and Image Understanding, vol 110, N 2, pp. 281-307, 2007.

[13] J. Daugman. “High confidence visual recognition of persons by a test of statistical independence”, IEEE Transactions on PAMI, vol 15, N 11, pp. 1148-1161, 1993.

[14] J. Daugman. “The importance of being random: Statistical principles of iris recognition”, Pattern Recognition, vol 36, N 2, pp. 279-291, 2003.

[15] J. Daugman, ”How iris recognition works”, IEEE Transactions on Circiuts and Systems for Video Technology, vol 14, N 1, pp. 21-30, 2004.

[16] F. Hao, R. Anderson and J. Daugman. “Combining crypto with biometrics effectively”, IEEE Transactions on Computers, vol 55, N 9, pp. 1081-1088, 2006.

[17] J. Bringer, H. Chabanne, G. Cohen, B. Kindarji and G. Zemor. “Optimal iris fuzzy sketches”, 1st IEEE Int. Conference on Biometrics: Theory, Applications and Systems, pp. 1-6, 2007.

[18] S. C. Chong, A. B. J. Teoh and D. C. L. Ngo. “High security iris verification system based on random secret integration”, Computer Vision and Image Understanding, vol 102, N 2, pp. 169-177, 2006.

[19] S. C. Chong, A. B. J. Teoh and D. C. L. Ngo. “Iris authentication using privatized advanced correlation filter”, International Conference in Biometrics (ICB), pp. 382-386, 2006.

[20] J. Zuo, N. K. Ratha and J. H. Connell. “Cancelable iris biometric”, Conference on Pattern Recognition (ICPR), pp. 1-4, 2008.

[21] J. K. Pillai, V. M. Patel, R. Chellappa and N. K. Ratha. “Secure and Robust Iris Recognition using Random Projections and Sparse Representations”, IEEE Transactions on pattern analysis and machine inteligence, vol 33, N 9, 2011.

[22] C. Rathgeb and A. Uhl. ” The State-of-the-Art in Iris Biometric Cryptosystems”, InTech, pp. 179-202, 2011.

[23] C. Carlet. “Boolean functions for cryptography and error correcting codes”, Cambridge Univ. Press, pp. 257-397, 2010.

[24] A. Nagar, K. Nandakumar and A. K. Jain. “MultiBiometric Cryptosystems Based on Feature-Level Fusion”, IEEE Transactions on Information Forensics and Security, vol 7, N 1, 2012.

[25] L. Nanni and A. Lumini. “Local binary patterns for a hybrid fingerprint matcher”, Pattern Recognition, vol 41, N 11, pp.3461-3466, 2008.