On the Vulnerability of Iris-Based Systems to a Software Attack Based on a Genetic Algorithm

Marta Gomez-Barrero, Javier Galbally, Pedro Tome, and Julian Fierrez

Biometric Recognition Group-ATVS, EPS, Universidad Autonoma de Madrid
C/ Francisco Tomas y Valiente 11, 28049 Madrid, Spain

Abstract. The vulnerabilities of a standard iris verification system to a novel indirect attack based on a binary genetic algorithm are studied. The experiments are carried out on the iris subcorpus of the publicly available BioSecure DB. The attack has shown a remarkable performance, thus proving the lack of robustness of the tested system to this type of threat. Furthermore, the consistency of the bits of the iris code is analysed, and a second working scenario discarding the fragile bits is then tested as a possible countermeasure against the proposed attack.

Keywords: Security, vulnerabilities, iris recognition, genetic algorithm, countermeasures.

1 Introduction

Due to their advantages over traditional security approaches, biometric security systems are nowadays being introduced into many applications where a correct identity assessment is a crucial issue, such as access control or sensitive data protection [1]. These systems perform automatic recognition of individuals based on anatomical (e.g., fingerprint, face, iris, etc.) or behavioural characteristics (e.g., signature, gait, keystroke dynamics). Among these traits, the iris has been traditionally regarded as one of the most reliable and accurate [1].

However, biometric systems are vulnerable to external attacks, that can be divided into two different groups, namely: *i) direct attacks*, carried out against the sensor using synthetic traits [2]; and *ii) indirect attacks*, carried out against one of the inner modules of the system [3], and thus requiring some knowledge about the inner working of the system. Several works have already studied the robustness of iris-based biometric systems against direct attacks, including attackers wearing contact lenses with artificial textures printed onto them [4] and fake iris images [5].

In the present paper, a novel indirect attack based on a genetic algorithm is presented. The point of attack are binary templates, as depicted in Fig. 1 (top), where a general hill-climbing attack is shown. Although other hill-climbing attacks have been proposed [6,3,7], none of them work on binary templates, but on real-valued feature vectors or directly on the sample images.

Although in commercial systems the number of consecutive unsuccessful access attempts is usually restricted, this countermeasure has been circumvented in different occasions or may even be used to compromise the system by performing an *account*

L. Alvarez et al. (Eds.): CIARP 2012, LNCS 7441, pp. 114–121, 2012.

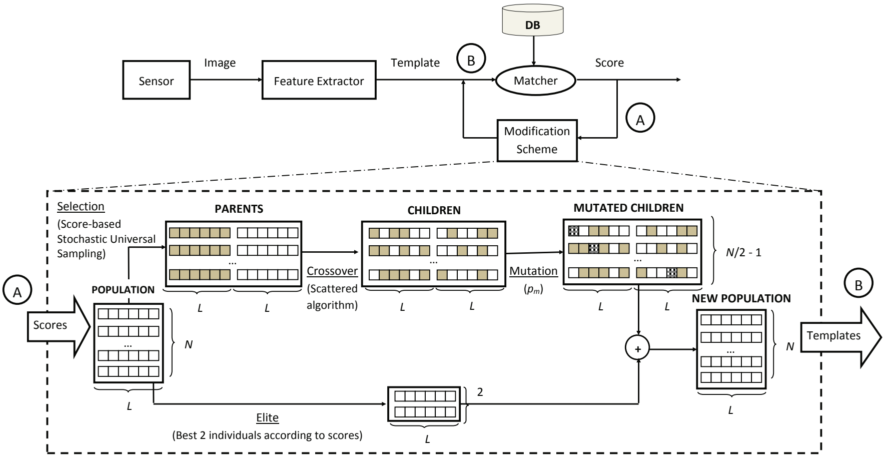

Fig. 1. Diagram of the general structure of a hill-climbing attack (top), with the specific modification scheme here implemented based on a genetic algorithm (bottom)

lockout attack (i.e., the intruder tries to access multiple accounts blocking all of them and collapsing the system). In the present work the consistency of the bits of the iris code is studied, and the use of the most consistent bits is analysed as a possible pure biometric countermeasure against the proposed attack.

The performance of the attack is evaluated on an iris recognition system adapted from the one developed by L. Masek [8] using the iris subcorpus of the BioSecure multimodal database [9]. The results show that most client accounts can be broken at the different operating points tested, even at a very high security one, requiring a similar number of matchings.

The paper is structured as follows: the attacking algorithm is introduced in Sect. 2. The system attacked is presented in Sect. 3, while the experimental protocol followed and the performance evaluation of the system are described in Sect. 4. The results obtained are shown in Sect. 5. Finally conclusions are drawn in Sect. 6.

2 Indirect Attack Based on a Genetic Algorithm

Most iris recognition systems use binary templates [10,11]. Therefore, given the good performance of genetic algorithms in binary optimization problems, they may be a very powerful tool to attack iris-based systems.

In the proposed attack to iris-based systems, the objective is to find an individual x (binary template), which is similar enough to the client being attacked, $\mathcal{C}$, according to a fitness function, $\mathcal{J}$, in this case being the similarity score (s) given by the matcher: $s = \mathcal{J}(\mathcal{C}, x)$

For this purpose, a genetic algorithm is used to optimize the similarity score (i.e., fitness function) starting from a randomly generated population, comprising a fixed number (N) of binary individuals (i.e., iris templates) of length L (in our particular case L will be the length of the iris code). As can be seen in Fig. 1 (bottom), four types of rules are used at each iteration to create the next optimized generation of individuals (templates) from the current population (being the input to the genetic algorithm the scores of the current population, and the output, the new templates):

- **Elite:** the two individuals (templates) with the maximum values for the fitness function (similarity scores) are kept for the next generation.
- **Selection rules:** certain individuals, the *parents*, are chosen by stochastic universal sampling. This way, the individuals with the highest fitness values (i.e., similarity scores) are more likely to be chosen as parents for the next generation: one subject can be selected 0 or many times.
- **Crossover rules:** parents are combined to form $N-2$ *children* following a scattered crossover method, where a random binary vector is created and the genes are selected from the first parent where the bit is a 1, and from the second when it is a 0 (vice versa for the second child).
- **Mutation rules:** random changes are applied to the new children with a mutation probability p_m.

The genetic algorithm is used to iteratively produce new generations following the rules given above. Each of the generations will contain individuals more similar each time to the objective ($\mathcal{C}$) until one of them produces a score higher than the verification threshold (i.e., the account is broken) or until one of the other stopping criteria is fulfilled: the maximum number of generations allowed is reached or the fitness values vary less than a certain pre-established amount.

It should be noticed that the present work is not focused on the study of genetic algorithms, but on the evaluation of the vulnerabilities of iris recognition systems to attacks based on these optimization tools. Therefore, a thorough analysis of the different specific GA parameters fall out of the scope of the work. For a more detailed description of different architectures for genetic algorithms the reader is referred to [12,13].

3 Iris Verification System Attacked

In our experiments, we have used the iris recognition system developed by L. Masek[1] [8], which is widely used to obtain base results in many iris related publications. Although the performance of this system has been largely improved by different systems over the last years [14], the experimental results shown in Sect. 5 are still fully meaningful since, as will be explained in Sect. 4, they have been obtained at operating points corresponding to False Acceptance Rates (FAR) typical of any better recognition system (i.e., with lower False Rejection Rates, FRR, for those same FARs).

The system comprises four different steps, namely: $i)$ segmentation, where the iris and pupil boundaries are modelled as two circles and detected using a circular Hough transform, as in [5]; $ii)$ normalization, which maps the segmented iris region into a 2D array using a technique based on Daugman's rubber sheet model; $iii)$ feature encoding,

[1] The source can be freely downloaded from
`www.csse.uwa.edu.au/pk/studentprojects/libor/sourcecode.html`

		BioSecure DS2 DB (210 Users)	
Session	Sample	170 Users	40 Users
1	1	Training	Test (Impostors)
	2		
2	1		
	2	Test (Clients)	

Fig. 2. Diagram showing the partitioning of the BioSecure DS2 DB according to the performance evaluation protocol defined in the present work

where the normalized iris pattern is convolved with 1D Log-Gabor wavelets, in order to produce a binary template of $20 \times 480 = 9,600$ bits and a corresponding noise mask that represents the eyelids areas; $iv)$ matching, where the inverse of the Hamming distance, $1/HD$, is used for matching (a higher score implies a higher degree of similarity). This Hamming distance is modified so that it incorporates the noise mask, using only the significant bits:

$$HD = \frac{\sum_{j=1}^{L} X_j(XOR)Y_j(AND)\bar{X}n_j(AND)\bar{Y}n_j}{L - \sum_{k=1}^{L} Xn_k(OR)Yn_k}$$

where X_j and Y_j are the two bitwise templates to compare, Xn_j and Yn_j are the corresponding noise masks for X_j and Y_j, and L is the number of bits comprised by each template. $\bar{X}n_j$ denotes the logical not operation applied to Xn_j.

4 Experimental Protocol

The experiments are carried out on the iris subcorpus included in the DS2 of the BioSecure multimodal database [9]. BioSecure DB, publicly available through the BioSecure Foundation [2], was acquired thanks to the joint effort of 11 European institutions and has become one of the standard benchmarks for biometric performance and security evaluation.

The database comprises three datasets captured under different acquisition scenarios. The Desktop Dataset, DS2, comprises voice, fingerprints, face, iris, signature and hand of 210 users, captured in two time-spaced acquisition sessions. The iris subset used in this work includes four grey-scale images (two per session) per eye, all captured with the Iris Access EOU3000 sensor from LG.

The performance of the evaluated system is computed using the experimental protocol shown in Fig. 2. The database is divided into: *i)* a training set comprising the first three samples of 170 clients (enrolment templates); and *ii)* an evaluation set formed by the fourth image of the previous 170 clients (used to compute the genuine scores) and all the 4 images of the remaining 40 users (used to calculate the impostor scores).

The final score given by the system is the average of the scores obtained after matching the input binary vector to the three templates (i.e., iris codes) of the attacked client model $\mathcal{C}$. For the experiments, we consider the left and right eyes of one person as different clients, thus having twice as many clients (340) and impostors (80). The system

[2] `http://biosecure.it-sudparis.eu/AB`

has an Equal Error Rate (EER) of 3.82%. The vulnerability of the system to the attack is evaluated at three operating points corresponding to: FAR = 0.1%, FAR = 0.05%, and FAR = 0.01%, which, according to [15], correspond to a low, medium and high security application. For completeness, the system is tested at a very high security operating point corresponding to FAR $\ll$ 0.01%.

4.1 Experimental Protocol for the Attacks

In order to generate the user accounts to be attacked with the genetic algorithm, we use the train set defined in the performance evaluation protocol (Fig. 2). The performance of the attack will be evaluated in terms of: $i)$ Success Rate (SR) or expected probability of breaking a given account, indicating how dangerous the attack is (the higher the SR, the bigger the threat); Efficiency (Eff) or inverse of the average number of matchings needed to break an account, thus giving an estimation of how easy it is for the attack to break into the system in terms of speed (the higher the Eff, the faster the attack). The SR is computed as the ratio between the number of broken accounts (A_B) and the total number of accounts attacked $(A_T = 170)$: SR $= A_B/A_T$, and the Eff is defined as Eff $= 1/\left(\sum_{i=1}^{A_B} n_i/A_B\right)$, where n_i is the number of matchings computed to bypass each of the broken accounts.

5 Results

The experiments have two different goals, namely: *i)* study the vulnerability of an automatic iris recognition system to the proposed attack, and *ii)* find the most consistent bits in the iris code and analyse whether the use of those bits increases the robustness of the system to the attack.

5.1 Attack Performance

The performance of the attack is measured at four different operating points, namely: *i)* FAR = 0.10%, *ii)* FAR = 0.05%, *iii)* FAR = 0.01%, and *iv)* FAR $\ll$ 0.01%. As can be observed in Table 1, the attacking algorithm proposed in this work successfully breaks most of the attacked accounts: around 80% SR on average, and as many as 50% of broken accounts for an unrealistically high operating point (FA $\ll$ 0.01%). It is also worth noting the fact that the efficiency barely depends on the operating point attacked: the number of comparisons needed increases only about 25% between the operating points FAR = 0.1% and FAR = 0.01% (while a brute force attack using randomly chosen real irises to access the system would need about ten times as many matchings, $\simeq$ 1/FAR).

5.2 Analysis of the Most Consistent Bits in the Iris Code

The results achieved by the hill-climbing attack based on a genetic algorithm against the iris recognition system considered in the experiments have shown its high vulnerability

Table 1. Eff and SR of the attack at the operating points tested

FAR	SR	Eff ($\times10^{-4}$)
0.10%	91.18%	1.400
0.05%	80.89%	1.255
0.01%	62.36%	1.102
$\ll$0.01%	52.06%	1.051

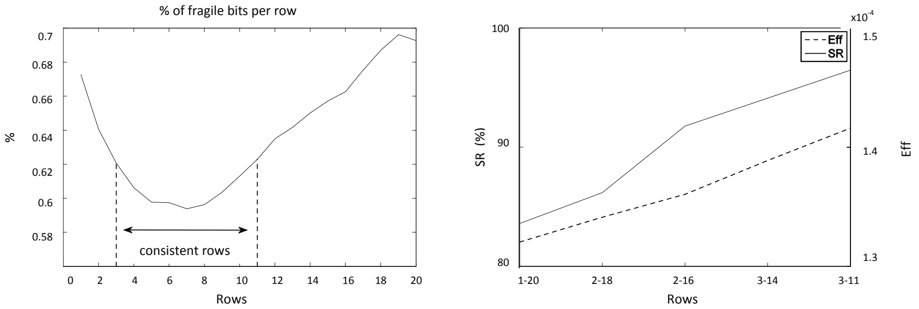

Fig. 3. Percentage of fragile bits (the ones flipping at least once across the images of an iris) in a row (left), and SR and Eff of the attack varying the number of rows used by the matcher (right)

against this type of attacking approach and the need to incorporate some attack protection method that increases its robustness against this threat. In this section we analyse the performance of using only the most consistent bits of the iris code for verification.

According to the analysis made by Hollingsworth *et al.* in [16], there are some bits more fragile than others in an iris code, that is, bits that flip between 0 and 1 in different images of the same iris with a high probability. Here we consider that a bit is consistent, (i.e., not fragile), when it does not flip in any of the four images available for each user. In order to determine the most consistent rows of bits in the iris code, we follow the method described in [16]: we compute the frequency (that must lie between 0% and 50%) that each unmasked bit flips, and take the average frequency across all bits in a row for each subject. All the codes of each user are previously aligned, keeping the rotation that gives the minimum Hamming distance to the first code of that user. In Fig. 3 (left), where the mean percentage of bits considered fragile in each row across all users is depicted, we can observe that rows 3 to 11 are the more consistent ones, having the lowest percentages of fragile bits.

Based on these results, we run some experiments testing the impact of reducing the number of rows of the iris codes: from using all rows (1 - 20) to only the best ones (3 - 11). The results, all obtained at an operating point of FAR = 0.05%, can be observed in Fig. 3 (right). The main reason for the increase in the performance of the attack (both in terms of SR and Eff) is that, by decreasing the number of rows, the number of bits drops drastically while the number of individuals in the population remains the same, thus increasing the diversity of the population and thereby enabling the genetic algorithm to find a maximum faster. Therefore, we may conclude that using only the most consistent bits in the iris code does not improve the robustness of the system against the proposed attacking algorithm.

6 Conclusions

In the present work, a novel indirect attack based on a genetic algorithm has been presented and used to evaluate a standard iris verification system to this type of threat. As many as 90% of the accounts are successfully broken in a similar number of generations for all the operating points considered, proving the vulnerabilities of such systems to this new attacking scheme.

The consistency of the bits of the iris code is then analysed as a possible countermeasure against the proposed attack, and a new scenario discarding the most fragile bits is considered. However, the algorithm reaches higher SRs needing even less comparisons.

Different analysis concerning the optimization of the specific genetic algorithm parameters may be considered in future works. However, these or other improvements fall outside the scope of this study, whose main objective is not to design a perfect method to break the security of biometric systems, but to encourage developers of algorithms and systems to seriously take into account this kind of attack and to implement specific protections and countermeasures.

The main objective of the work, is not to perform a thorough analysis of the different specific GA parameters, but to demonstrate the feasibility of such attacks and to encourage developers to take this security flaw seriously into account.

Acknowledgements. This work has been partially supported by projects Contexts (S2009/TIC-1485) from CAM, Bio-Challenge (TEC2009-11186) from Spanish MICINN, TABULA RASA (FP7-ICT-257289) and BEAT (FP7-SEC-284989) from EU, and *Cátedra UAM-Telefónica*.

References

1. Jain, A.K., Ross, A., Pankanti, S.: Biometrics: a tool for information security. IEEE TIFS 1(2), 125–143 (2006)
2. Matsumoto, T.: Gummy finger and paper iris: an update. In: Proc. WISR, pp. 187–192 (2004)
3. Martinez-Diaz, M., Fierrez, J., et al.: An evaluation of indirect attacks and countermeasures in fingerprint verification systems. Pattern Recognition Letters 32, 1643–1651 (2011)
4. Wei, Z., Qiu, X., et al.: Counterfeit iris detection based on texture analysis. In: Proc. ICPR, pp. 1–4 (2008)
5. Ruiz-Albacete, V., Tome-Gonzalez, P., Alonso-Fernandez, F., Galbally, J., Fierrez, J., Ortega-Garcia, J.: Direct Attacks Using Fake Images in Iris Verification. In: Schouten, B., Juul, N.C., Drygajlo, A., Tistarelli, M. (eds.) BIOID 2008. LNCS, vol. 5372, pp. 181–190. Springer, Heidelberg (2008)
6. Soutar, C., Gilroy, R., Stoianov, A.: Biometric system performance and security. In: Proc. IEEE AIAT (1999)
7. Galbally, J., McCool, C., Fierrez, J., Marcel, S.: On the vulnerability of face verification systems to hill-climbing attacks. Pattern Recognition 43, 1027–1038 (2010)
8. Masek, L., Kovesi, P.: Matlab source code for a biometric identification system based on iris patterns. Master's thesis, School of Computer Science and Software Engineering, University of Western Australia (2003)
9. Ortega-Garcia, J., Fierrez, J., others: The multi-scenario multi-environment BioSecure multimodal database (BMDB). IEEE TPAMI 32, 1097–1111 (2010)

10. Daugman, J.: How iris recognition works. IEEE TCSVT 14(1), 21–30 (2004)
11. Daugman, J.: 4. In: Iris Recognition, pp. 71–90. Springer (2008)
12. Goldberg, D.E.: Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley Longman Publishing Co., Inc. (1989)
13. Goldberg, D.: The design of innovation: lessons from and for competent genetic algorithms. Kluwer Academic Publishers (2002)
14. Grother, P., Tabassi, E., Quinn, G.W., Salamon, W.: Irex i: Performance of iris recognition algorithms on standard images. Technical report, National Institute of Standards and Technology (2009)
15. ANSI: Ansi.x9.84 ANSI X9.84-2001, Biometric Information Management and Security
16. Hollingsworth, K.P., Bowyer, K.W., Flynn, P.J.: The best bits in an iris code. IEEE TPAMI 31(6), 964–973 (2009)