Vulnerability Assessment of Fingerprint Matching Based on Time Analysis

Javier Galbally, Sara Carballo, Julian Fierrez, and Javier Ortega-Garcia

Biometric Recognition Group–ATVS, EPS, Universidad Autonoma de Madrid,
C/ Francisco Tomas y Valiente 11, 28049 Madrid, Spain
{javier.galbally,sara.carballo,julian.fierrez,javier.ortega}@uam.es

Abstract. A time analysis of a reference minutiae-based fingerprint matching system is presented. We study the relation between the score generated by the system (NFIS2 from NIST) and the time required to produce the matching score. Experimental results are carried out on a subcorpus of the MCYT database and show a clear correlation between both matching variables (time and score). Thus, a new threat against biometric systems is arisen as attacks based on the matching score could be largely simplified if the time information is used instead.

1 Introduction

In the last recent years important research efforts have been conducted to study the vulnerabilities of biometric systems to direct attacks to the sensor (carried out using synthetic biometric traits such as gummy fingers) [1], and indirect attacks (carried out against some of the inner modules of the system) [2,3]. This research efforts have led to an enhancement of the security level offered by biometric systems through the proposal of new countermeasures to the attacks analyzed. Furthermore, the interest for the analysis of security vulnerabilities has surpassed the scientific field and different standardization initiatives at international level have emerged in order to deal with the problem of security evaluation in biometric systems, such as the Common Criteria through different Supporting Documents [4], or the Biometric Evaluation Methodology [5].

Within the vulnerabilities that have been studied, special attention has been paid to the hill-climbing attacks [6,7,8]. These attacking algorithms produce a number of synthetic templates which are iteratively modified according to the score they produce: if the score increases the changes are kept and otherwise the modifications are discarded. This way the score raises until the acceptance threshold is reached and the system is broken.

Although hill-climbing attacks have proven their efficiency against biometric systems, they still present the strong restriction of needing the score produced by the matcher to be able to break the system. Even if the attacker is able to access the similarity measure (which is not always the case), the attack can still be countermeasured by quantizing the score so that the hill-climbing algorithm does not get the necessary feedback to iteratively increase the similarity measure until the threshold is reached.

J. Fierrez et al. (Eds.): BioID_MultiComm2009, LNCS 5707, pp. 285–292, 2009.

A bigger threat to biometric systems would arise if they could be attacked using some easily measurable information such as the matching time, or the power consumed by the system in the matching process. This type of information, which has already been used to successfully attack cryptographic security systems [9,10], is always accessible to a possible attacker and difficult to be manipulated or distorted by the system designer (in opposition to the similarity score used in traditional hill-climbing algorithms).

In the present work we carry out a time analysis of a reference fingerprint-based recognition system (NFIS2 from NIST [11]), in order to determine if there exists a relation between the similarity score produced by the matching module, and the time required to generate that score. Such a study will permit to determine the feasibility of developing attacks against this type of systems based on the time information, and the necessity or not of taking into account this threat not only when designing biometric systems, but also when evaluating their level of security.

The rest of the paper is structured as follows. In Sect. 2 the system analyzed in the experiments is described. The database used in the experiments and the performance evaluation of the studied system are presented in Sect. 3. Experimental results of the temporal analysis are given in Sect. 4, and conclusions are finally drawn in Sect. 5.

2 Reference System Analyzed (NFIS2)

The system analyzed in the experiments is the minutiae-based NIST Fingerprint Image Software 2 (NFIS2) [11]. This publicly available software is used in many works as a reference system with which to compare new fingerprint verification solutions. It is a PC-based fingerprint processing and recognition system formed of independent software modules. The feature extractor generates a text file containing the location (x and y coordinates) and orientation (angle with respect to the positive x axis) of each minutia from the fingerprint. The matcher uses this file to generate the score. The matching algorithm is rotation and translation invariant since it computes only relative distances and orientations between groups of minutiae.

3 Database and Performance Evaluation

The temporal analysis has been carried out using a subcorpus of the MCYT database [12]. The subcorpus comprises 10 impressions of the right and left index fingers of 75 users ($75 \times 2 \times 10 = 1,500$ images), captured electronically with the optical sensor UareU from Digital Persona (500 dpi, 256×400 images). Six of the samples of each finger were acquired with a high control level (small rotation or displacement of the finger core from the center of the sensor was permitted), another two with a medium control level, and the remaining two with low control level (see Fig. 1 for examples of fingerprint images).

Fig. 1. Examples of typical fingerprint images that can be found in MCYT acquired with a low (left), medium (center), and high degree of control (right)

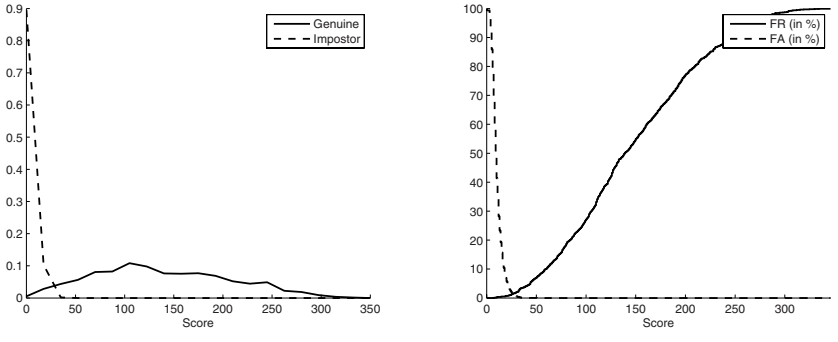

Fig. 2. Score distributions (left), and FA and FR curves (right), for the MoC system

In order to estimate the verification performance of the system a set of genuine and impostor scores were computed. We used one of the low control samples as a template and the other 9 samples from the same finger as probes to test genuine matches, leading to $150 \times 9 = 1,350$ genuine user scores. Impostor scores are obtained comparing each template to one sample from the remaining fingers of the database, thus we have $150 \times 149 = 22,350$ impostor scores. In Fig. 2 we show the two score distributions (left), and the FA and FR curves of the evaluated system (right). The EER of the system is 1.47%.

4 Time Analysis

The objective of the experimental study is to determine if there exists a correlation between the score given by the analyzed system and the matching time required to produce that score. In order to reach this goal two experiments have

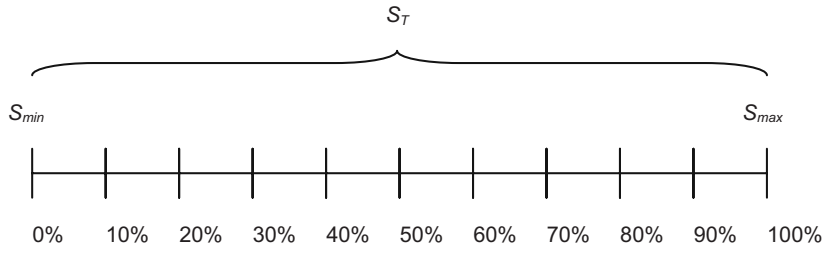

Fig. 3. Division of the total range of scores S_T used in experiment 1

been carried out, the first one to find out if there is a correspondence between score and time, and the second one to identify the behavior of the matching time when the score increases or decreases.

4.1 Experiment 1: Relation between Score and Time

As explained in Sect. 3, with the used database a set of genuine and impostor scores were generated, comprising 1,350 and 22,350 similarity measures respectively. Each of these scores has a time value associated corresponding to the matching time used by the system to produce that score. Thus, there are a total 23,700 scores (irrespective of genuine or impostor) and the same amount of time values.

In this experiment, the whole range of scores (S_T) is divided into ten equally spaced bands (corresponding to different percentages of the total range) as shown in Fig. 3. The distributions of the time values associated to the scores corresponding to those bands are plotted in Fig. 4, from the most distant regions (those time distributions corresponding to the first and last 10% of S_T), to the closest bands (time distributions corresponding to the scores in bands [40-50]% and [50-60%]). This way we will be able to determine if more distant score distributions correspond to more separated time distributions.

No distinction is made between impostor and genuine scores as we want to find the relation between a certain score s and its associated time value t_s, regardless of whether that score has been produced by a genuine or impostor fingerprint.

In Fig. 4 we can see that the degree of overlap of the time distributions through plots *(a)* to *(e)* increases. That is, the time distributions corresponding to the scores that belong to more distant bands are more separated than those corresponding to scores of closer bands. Furthermore, in all the plots, the distribution corresponding to the higher band (drawn with a dashed line) has a higher mean value than that corresponding to the lower band of scores (drawn with a plain line). These observations suggest that bigger scores produce higher time values.

4.2 Experiment 2: Relation between Score and Time Variations

This second experiment has a twofold objective, first to confirm the conclusions extracted from the previous experiment (i.e., that higher time values correspond to higher scores), and second to determine if there exists a correlation between

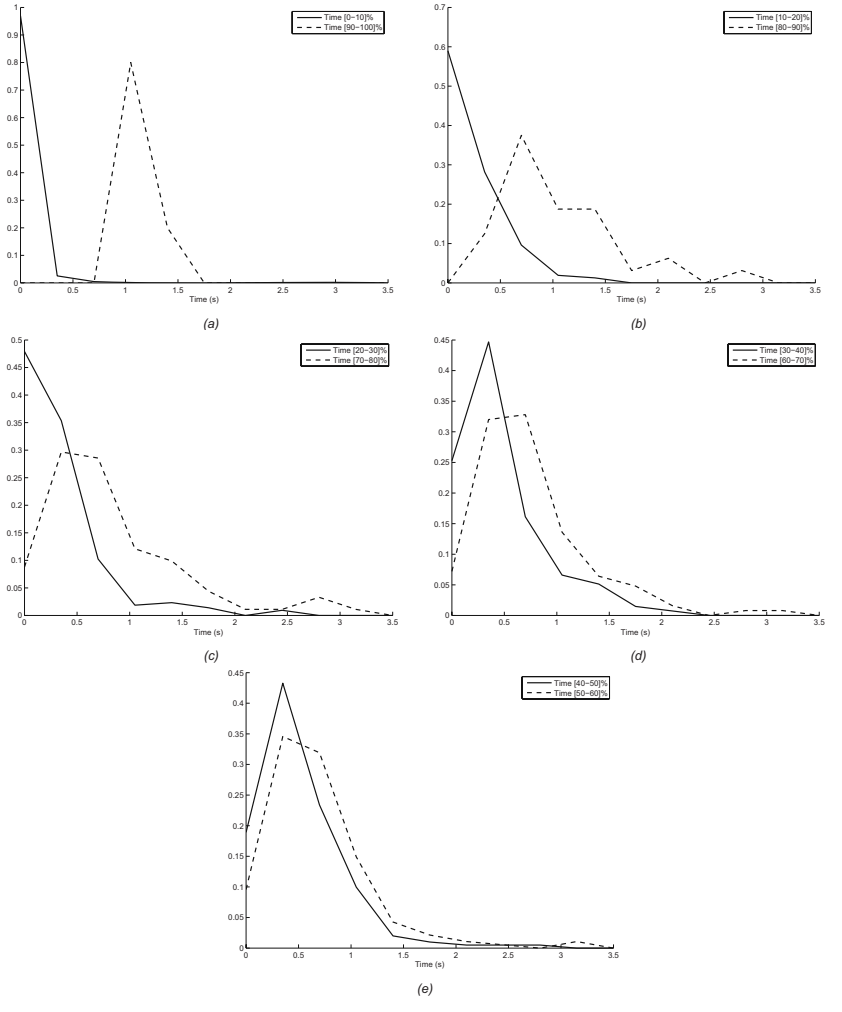

Fig. 4. Distributions of the time values associated with the scores corresponding to: *(a)* bands [0-10]% and [90-100]%, *(b)* bands [10-20]% and [80-90]%, *(c)* bands [20-30]% and [70-80]%, *(d)* bands [30-40]% and [60-70]%, and *(e)* bands [40-50]% and [50-60]%

a variation in the score and a variation in the time needed to produce that score. With this purpose, 50 templates corresponding to 50 different users of the database were slowly degraded following an iterative process. At each iteration of the algorithm one of these two random modifications A) perturbing a minutia or, B) substituting a minutia, are applied to the fingerprint, and the modified template is matched to the original one at the same time that the time needed to produce the score is measured.

The two modifications A and B were taken from [2]. The fact that only perturbing or substituting a minutia are permitted (and not other modifications such as adding or deleting a minutia), assures that the number of minutiae in

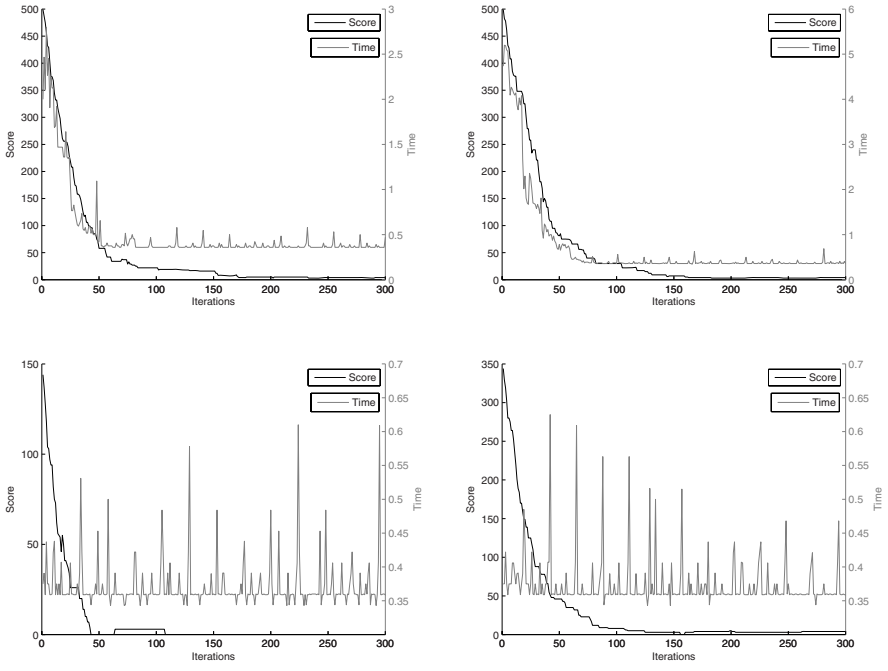

Fig. 5. Evolution of the score (black) and the time (grey) for four of the fingerprints used in experiment 2

the template remains constant through the iterations and that the changes in the matching time are not due to variations in the number of singular points.

This way, in the first iteration of the process the fingerprint is matched against itself reaching the highest possible score for that template. Through the remaining iterations, until reaching the maximum permitted (300), the score slowly decreases as a result of the changes introduced in the template. At the end of the experiment we have 50 sets of 300 matching scores (corresponding to the evolution of the score for each of the 50 fingerprints), and their associated 50 sets of 300 time values. In Fig. 5 we show the evolution of the score (black) and the time (grey) for four of the cases studied. As the score and time ranges are different, the score scale is depicted on the left of each plot (also in black), and the time scale on the right (in grey).

In the top row of Fig. 5 we have depicted two cases in which a clear relation between the score and the time can be seen: the higher the score, the higher the time needed by the system to generate it. However, in the two examples of the bottom row that relation cannot be observed, and score and time seem to be totally uncorrelated. In order to study the behavior of the system from a statistical point of view, and not for each particular case, the mean of the 50 score sets (with 300 matching scores each) and of the corresponding 50 time sets were computed. Both means are shown in Fig. 6 in an analogue way to the one used in the plots of Fig. 5.

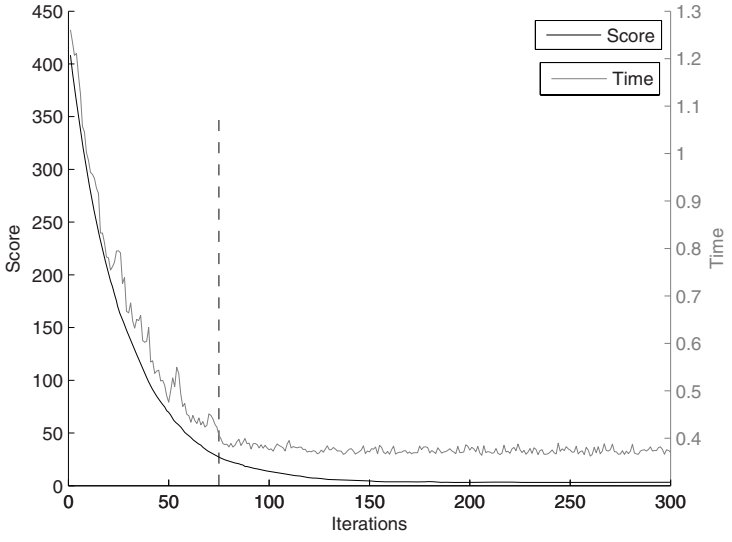

Fig. 6. Mean of the score (black) and the time (grey) evolution for all the 50 fingerprints considered in experiment 2

In the score and time evolution depicted in Fig. 6 we can distinguish two zones (separated with a vertical dashed line) where the system presents different behaviors. On the left of the vertical dashed line we can see that there is a clear correlation between score and time variations: an increase in the score causes a raise in the time (on average). However, on the right of the separating line (once the score has fallen below 30 approximately), time and score seem to be uncorrelated, and while the score keeps decreasing the time fluctuates around a constant value.

5 Conclusions

A time analysis of a reference fingerprint based verification system has been made (NFIS2 from NIST). Experiments were carried out on a subcorpus of the publicly available MCYT database and show that there exists a clear correlation between the score given by the system and the time needed to produce that score. These findings reveal a new type of vulnerability of biometric systems as the matching time (easy to measure) could be used to simplify attacks initially thought to exploit the matching scores (not always easy to access).

The present work might be of special interest not only for designers (in order to include the necessary countermeasures), but also for evaluators of security systems (in order to take into account this new vulnerability), and for those institutions developing evaluation standards such as the Common Criteria [4], or the associated Biometric Evaluation Methodology [5].

Acknowledgements

J. G. is supported by a FPU Fellowship from Spanish MEC and J. F. is supported by a Marie Curie Fellowship from the European Commission. This work was supported by Spanish MEC under project TEC2006-13141-C03-03.

References

1. Galbally, J., Fierrez, J., et al.: On the vulnerability of fingerprint verification systems to fake fingerprint attacks. In: Proc. IEEE of International Carnahan Conference on Security Technology (ICCST), pp. 130–136 (2006)
2. Uludag, U., Jain, A.K.: Attacks on biometric systems: a case study in fingerprints. In: Proc. SPIE-IE, vol. 5306, pp. 622–633 (2004)
3. Hill, C.J.: Risk of masquerade arising from the storage of Biometrics, B.S. Thesis. Autralian National University (2001)
4. CC: Common Criteria for Information Technology Security Evaluation. v3.1 (2006)
5. BEM: Biometric Evaluation Methodology. v1.0 (2002)
6. Adler, A.: Sample images can be independently restored from face recognition templates. In: Proc. Canadian Conference Electrical and Computing Engineering (CCECE), vol. 2, pp. 1163–1166 (2003)
7. Martinez-Diaz, M., Fierrez, J., et al.: Hill-climbing and brute force attacks on biometric systems: a case study in match-on-card fingerprint verification. In: Proc. IEEE of International Carnahan Conference on Security Technology (ICCST), pp. 151–159 (2006)
8. Galbally, J., Fierrez, J., Ortega-Garcia, J.: Bayesian hill-climbing attack and its application to signature verification. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 386–395. Springer, Heidelberg (2007)
9. Kocher, P.C.: Timing attacks on implementations of diffie-hellman, RSA, DSS, and other systems. In: Koblitz, N. (ed.) CRYPTO 1996. LNCS, vol. 1109, pp. 104–113. Springer, Heidelberg (1996)
10. Kocher, P.C., Jaffe, J., Jun, B.: Differential power analysis. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, p. 388. Springer, Heidelberg (1999)
11. Watson, G.I., Garris, M.D., et al.: User's guide to NIST Fingerprint Image Software 2 (NFIS2). National Institute of Standards and Technology (2004)
12. Ortega-Garcia, J., Fierrez-Aguilar, J., et al.: MCYT baseline corpus: a bimodal biometric database. IEE Proc. Vision, Image and Signal Processing 150, 391–401 (2003)