A biometric verification system addressing privacy concerns

Stelvio Cimato, Marco Gamassi, Vincenzo Piuri, Roberto Sassi and Fabio Scotti
Dipartimento di Tecnologie dell'Informazione
Università degli Studi di Milano
Via Bramante 65
26013 Crema, Italy
e-mail: {cimato, gamassi, piuri, fscotti, sassi}@dti.unimi.it

Abstract

Biometric techniques are more and more exploited in order to fasten and make more reliable the identification process. Recently, many proposals have been formulated combining cryptography and biometrics in order to increase the confidence in the system when biometric templates are stored for verification.

In this work we present a biometric authentication technique based on the combination of multiple biometric readings. The authentication control can be performed offline and the stored identifier does not disclose any information on the biometric traits of the identified person, so that even in case of loss or steal of the document, privacy is guaranteed.

Keywords: Biometric identification, Privacy, Secure sketch.

1 Introduction

The problem of devising practical systems for personal identification and verification relying on biometrical data has been faced in many recent works [UPPJ04]. Indeed, biometrics such as fingerprints, voice and face are permanently associated with the user and can therefore obviate the need to carry tokens or remember passwords and keys. On the other side, the strict association between each user and his biometric templates raises concerns on possible uses and abuses of such kind of sensible information.

An interesting approach is to use biometric templates to compute cryptographic keys that can be used in differed kind of applications, such as authentication or data protection. In literature several systems combining cryptography and biometrics have been presented, satisfying the needed requirements to develop practical applications [SK06, NL06]. The main problem is to cope with the higher variability within different readings of biometric data that makes them unsuitable to be directly used for cryptographic applications. In fact, cryptographic keys have zero uncertainty and a single-bit difference (in the key or in the message) spoils the possibility of accessing the encrypted data.

In this work we present a biometric authentication system enabling the creation of an identifier associated with each enrolled person. The system we propose has a number of appealing characteristics, which could encourage its use in a wide range of application for authentication:

- privacy of the biometrics templates is protected, since the templates are subjected to randomization transformation such that the derived published identifier do not suffer from information leakage.
- the system is inherently multi-modal, since its functioning relies on the knowledge of at least two biometrics templates during the enrollment phase.
- the system is modular, since it does not rely on a proprietary algorithm but on the composition of basic modules which can be substituted if the requested specification parameters are satisfied.
- the system is usable, since, as we will see in Section 5, prototype can be conceived by assembling available modules and obtain acceptable overall error rates for different combination of biometric features.

In the next section we discuss some related work. In Section 3 we present our method, describing the basic component of our system. In Section 4 we briefly present some considerations on the security of our approach and in Section 5 we report some experimental results.

2 Related work

In literature, a wide range of techniques have been presented based on the combination of biometrics and cryp-

0-7695-3072-9/07 $25.00 © 2007 IEEE
DOI 10.1109/CIS.2007.117

tography, in order to cope with both problems: variability of biometric templates and protection of personal data. A comprehensive review of these approaches can be found in [UPPJ04]. The process of generating cryptographic keys from biometrics generally relies on an error tolerant binary representation of the biometrics features. In Davida et al. [DFM98, DFMP99], hash functions are used to protect the sensitive user template.

Biohashing was proposed in [JLG04] and relies on a two-factor authenticator based on the combination of pseudo-random numbers and a biometric binarized feature The main disadvantage of BioHashing method is in that poor verification performance is possible when an impostor steals the pseudo-random number used to built the ID of a genuine user and tries to authenticate [NL06]. The usage of a multi-modal biometric authentication system where one or two biometric features have been "biohashed" is shown to reduce the effect of this drawback, but the proposed technique increases the overall error rate.

In [JW99], Juels and Wattenberg proposed the "fuzzy commitment" scheme where a secret message is protected using a biometric template. In this case, an error correcting code is used in order to associate a codeword c with a person and compute an offset ($\delta = c \oplus x$) for the biometric template x. The encrypted message (the *fuzzy commitment*) is then represented by the pair $(\delta, h(c))$, where $h(c)$ is a one way hash function. It is worth to notice that neither the biometric feature, nor the associated codeword are publicly stored. The authentication process is correctly performed if a fresh biometric reading y allows the computation of a binary string $c' = \delta \oplus y$ sufficiently close to c so that the code decodes it to c and the comparison between their hash values succeeds. Moving in the same direction, Hao et al proposed a biometric key generation procedure, which is based on an iris code feature extraction algorithm and on the combined use of Hadamard and Reed-Solomon codes [HAD05]. Juels and Sudan also proposed a "fuzzy vault scheme" in [JS02] relying on the polynomial interpolation technique in order to cope with variability of the biometrics template stored.

Recently, a similar approach has been proposed in [SK06] to achieve a biometric system for offline verification of certified, cryptographically secure documents. The presented technique can produce printable IDs obtained from an extracted and compressed iris feature and an arbitrary text.

3 The proposed method

In the proposed method a number of biometric readings are used during the authentication phase. In the following we detail the functioning of the basic modules: The first one (*enrollment* module) creates the non reversible ID starting from the biometric samples. The second one, the (*verification module*) performs the identification process starting from the novel biometric readings and the information contained into the ID.

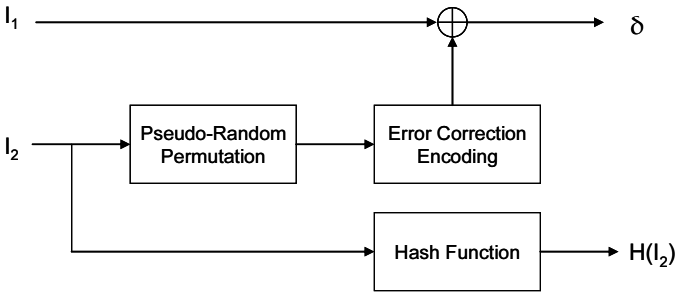

Figure 1. The enroll module.

3.1 Enrollment module

During the *enrollment* phase, a novel identifier is created for each user, by composing the available biometric features. The ID can be then stored and must be provided during the verification phase. The general structure is depicted in Figure 1.

Two biometric readings are separately processed to extract two sets of biometric features as usual. The proposed system relies on at least a couple of feature extraction algorithms $\mathcal{F}_1$ and $\mathcal{F}_2$ which can be selected among the already known feature extraction algorithms in literature. Let us denote with n_i the bit string returned by the feature extraction algorithm and with r_i its error rate, i.e., the rate of bits in the pattern which could be modified without affecting the biometric identification of the subject.

The second biometric feature is given as input to a pseudo random permutation block, which returns a bit string of the same length, having almost uniform distribution. The string is then encoded by using an error correction code with parameters matching the length and the error correction rate previously computed. The resulting codeword is xored with the second biometric feature. To this aim, one or both string could be subjected to a padding in order to derive compatible binary string. The second biometric feature is also given as input to an hash function. The resulting digest, together with the bit string resulting from the xor, and the other additional information needed to invert the transformation are collected and published as the identifier of the enrolled person.

3.2 Verification module

The *verification* phase enables a strong authentication of the subject who has to provide the ID he received during the

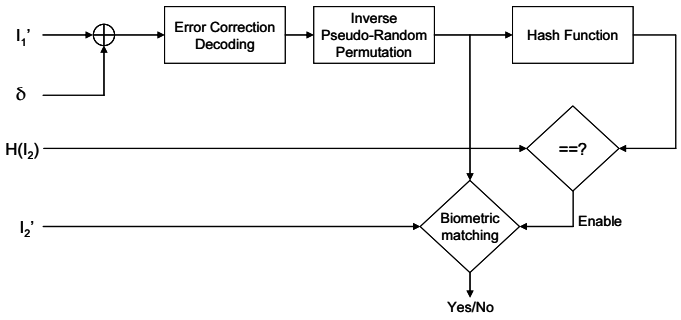

Figure 2. The verification module.

enrollment phase and the biometric traits requested by the procedure. Figure 2 shows the structure of the verification module.

Let us denote with I_1' and I_2' the biometric features freshly collected. The ID provided by the subject is split into δ, the hash $H(I_2)$ (and the additional info x). By xoring the reading I_1' with δ, a bit string is retrieved. Such string should differ from the corresponding string created in the enrollment phase for at most r_1 bits (since the error rate of $\mathcal{F}_1$ is r_1 under the hypothesis that both readings I_1 and I_1' belong to the same subject). The decoding phase of the selected error correction code and the application of the inverse pseudo random permutation, should allow the exact reconstruction of the original reading I_2.

At this point a first check is performed in order to compare the hash of the retrieved value for I_2 with the value $H(I_2)$ stored into the identifier. Only if the check succeeds, a biometric matching is performed using as input the retrieved value I_2 and the fresh biometric reading I_2'. The authentication is successful when the biometring matching is positive.

3.3 Composition of basic modules

The composition of the basic modules enables the creation of authentication application having different levels of security and using a higher number of biometric features. The basic *enrollment* and *verification* modules can be combined in parallel or hierarchically. Figure 3 shows two examples of these basic compositions.

The *parallel composition* offers a simple method to exploit different biometric traits in order to create the ID. This way, the level of multi-modality implemented is higher than in the standard approach since more than two biometric traits are in use. However, the error rate r in the composed input needs particular scrutiny as each biometric method differently contributes to the overall error rate.

Basic modules can be composed also in hierarchical structures. Figure 3 shows an example of a two levels hierarchical composition. Biometric inputs I_1 and I_2 are used

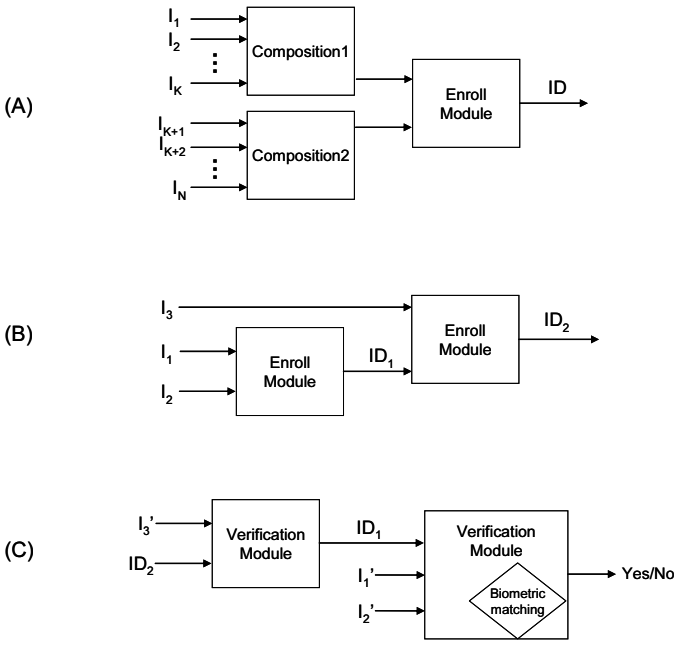

Figure 3. Examples of possible compositions of the proposed basic modules during the enroll and verification phases.

to create $ID1$ by means of a basic enroll module. Then, $ID1$ is used in place of a biometric trait in a second basic enroll module together with a third biometric input I_3 leading to the creation of $ID2$.

The hierarchical composition enables *different levels of security*. In low-security applications only the biometric inputs I_1 and I_2 could be required to verify $ID1$. On the other hand, in high-security applications, a third biometric sample I_3 would be required to verify also $ID2$ (obviously given that the verification of $ID1$ had been successful).

It is worth to notice, that it is possible to build more complex systems by using each method of composition (parallel and hierarchical) recursively or by combining the methods iteratively.

4 Discussion

In the following we informally discuss the security of the proposed method. Our approach builds on the fuzzy commitment scheme presented by Juels and Wattenberg and recasted as secure sketch in [DORS06, Boy04]. Differently from Juel's approach, in our scheme, one biometric reading is xored with a random bit string obtained after a pseudo random permutation from the other biometric reading. Furthermore, in the verification phase, the process is inverted and the second biometric template is reconstructed in order

to be used as preliminary check (by comparing the computed hash with the value stored into the identifier) and as input to a biometric matching. To foolish the authentication system, an adversary has then two ways: recover a biometric template from the public available information (the identifier) associated with the enrolled person; or steal the biometric templates of a genuine user through covert means. In the first case, results on the security of secure sketch constructions still hold and ensure that the adversary cannot take advantage from the knowledge of the identifier. In the second case, the adversary should steal at least two biometric templates to complete the authentication phase. As debated, a higher number of biometrics can be taken into account in the setup of the authentication system, in order to increase the overall security of the application.

The proposed method enables the identification of persons using offline secure documents, in which neither biometrics traits, nor other sensible data are stored. To ensure its validity, the identifier produced during the enrollment phase should be signed using the private key of the issuer using one of the most diffused public key cryptosystem.

5 Experiments and Results

In this section, a practical set up of the method described in the paper will be described. Given the requirements sketched in the previous sections, we selected iris for the first biometric I_1 and fingerprints for the second biometric, I_2. Thus images from the CASIA *iris database* [CAoS03] were coupled to fingerprints extracted from the FVC2000 dataset [MMC$^+$02]. The CASIA (version 1.0) database contains up to seven images of the same eye obtained from 108 subjects. Fingerprint images ('tenprints") were obtained from DB2 (part A and B) of the FVC2000 database.

We adopted a *best of three* approach when selecting the iris template, in order to avoid that segmentation errors might jeopardize the verification stage. Then, the enrolling stage was performed following along the lines of what described in section 3.1. We selected a Reed-Solomon correction code with $n_1 = 9600$ and $r_1 = 0.4$; with this parameters, our scheme allows for up to $k = 1920$ bits for storing the fingerprint minutia. For this reason we first applied a mapping associating each bit in the pseudo-randomized minutia' binary string with a symbol of $m = 14$ bits, having the first $m - 1$ bits randomly selected Then the resulting string was encoded with a $[9600, 1920, 7681]_{2^{14}}$ Reed-Solomon shortened code [Kar02]. A related operation was performed on the iris binary string, which was encoded by prepending $m - 1$ zero bits to each original bit (*zero padding*). Finally, δ was obtained performing a xor between the minutia' encoded message and the coded iris string.

The verification phase was emulated using the remaining eyes and fingerprints pictures. Up to 4 eye's pictures were used to obtain a second iris code with the same numerical code used for enrollment. If one of the iris code was able to unlock the first part of the scheme, also fingerprints were processed. Otherwise, the subject was declared an impostor, and the process stopped. In particular, the iris code underwent a zero padding and was subtracted (xor) to δ. Then it was decoded with a Reed-Solomon decoder. the pseudo random permutation inserted during enrollment removed and the random bits removed to obtain the original set of enrollment minutia. If the SHA-1 hash of the minutia set decoded was identical to what computed at enrollment, this was taken as a positive iris match. Otherwise, first the iris code was shifted in both direction up to 8 bits and eventually the other 3 images were processed. In case of a positive iris match, the ANSI-INCITS fingerprint template decoded was matched with up to 5 fingerprint pictures for each subject. The match was performed using the NIST NBIS-detector `bozorth3`[WGT$^+$07]. In case one of the match resulted positive the remaining comparison were skipped and the verification procedure was judged positive. The overall results we obtained were FRR=1.85% and FAR=0.0087%. The selection of the a good quality iris code and fingerprint template at enrollment proved fundamental in improving the false reject rate of the scheme.

References

[Boy04] X. Boyen. Reusable cryptographic fuzzy extractors. In *11th ACM Conference on Computer and Communication Security (CCS 2004)*, volume 3027. ACM, 2004.

[CAoS03] Chinese Academy of Sciences. Database of 756 greyscale eye images; Version 1.0, 2003.

[DFM98] G. I. Davida, Y. Frankel, and B. J. Matt. On enabling secure applications through off-line biometric. In *Proceedings of the IEEE International Symposium on Security and Privacy, 1998*, pages 148–157. IEEE Press, 1998.

[DFMP99] G. I. Davida, Y. Frankel, B. J. Matt, and R. Peralta. On the relation of error correction and cryptography to an off line biometrics based identification scheme. In *WCC99, Workshop on Coding and Cryptography*, 1999.

[DORS06] Y. Dodis, R. Ostrovsky, L. Reyzin, and A. Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data, 2006.

[HAD05] F. Hao, R. Anderson, and J. Daugman. Combining cryptography with biometrics effectively. Technical Report UCAM-CL-TR-640, University of Cambridge, Computer Laboratory, United Kingdom, July 2005.

[JLG04] A. Teoh Beng Jin, D. Ngo Chek Ling, and A. Goh. Biohashing: two factor authentication featuring fingerprint data and tokenised random number. *Pattern recognition*, 37(11):2245–2255, 2004.

[JS02] A. Juels and M. Sudan. A fuzzy vault scheme. In A. Lapidoth and E. Teletar, editors, *Proceedings of the IEEE International Symposium on Information Theory, 2002*, page 408. IEEE Press, 2002. The full version of the paper is located at http://www.rsasecurity.com/rsalabs/staff/bios/ajuels/publications/fuzzy-vault/fuzzy_vault.pdf.

[JW99] Ari Juels and Martin Wattenberg. A fuzzy commitment scheme. In *CCS '99: Proceedings of the 6th ACM conference on Computer and communications security*, pages 28–36, New York, NY, USA, 1999. ACM Press.

[Kar02] Phil Karn. Reed-solomon encoding and decoding code, 2002.

[MMC+02] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain. FVC2000: Fingerprint verification competition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(3):402–412, 2002.

[NL06] L. Nanni and A. Lumini. Empirical tests on biohashing. *NeuroComputing*, 69(16):2390–2395, October 2006.

[SK06] D. Schonberg and D. Kirovski. Eyecerts. *IEEE Transactions on Information Forensics and Security*, 1:144–153, June 2006.

[UPPJ04] U. Uludag, S. Pankanti, S. Prabhakar, and A. Jain. Biometric cryptosystems: Issues and challenges. In *Proceedings of the IEEE, Special Issue on Enabling Security Technologies for Digital Rights Management*, volume 92, pages 948–960, June 2004.

[WGT+07] Craig I. Watson, Michael D. Garris, Elham Tabassi, Charles L. Wilson, R. Michael McCabe, Stanley Janet, and Kenneth Ko. User's Guide to NIST Biometric Image Software (NBIS). (formerly NISTIR 6813), 2007.