An Overview on Privacy Preserving Biometrics

Rima Belguechi, Vincent Alimi, Estelle Cherrier, Patrick Lacharme, Christophe Rosenberger

4

An Overview on Privacy Preserving Biometrics

Rima Belguechi, Vincent Alimi, Estelle Cherrier, Patrick Lacharme
and Christophe Rosenberger
Université de Caen Basse-Normandie, UMR 6072 GREYC, F-14032 Caen
ENSICAEN, UMR 6072 GREYC, F-14050 Caen
CNRS, UMR 6072 GREYC, F-14032 Caen
France

1. Introduction

The Internet has consolidated itself as a very powerful platform that has changed the communication and business way. Nowadays, the number of users navigating through Internet is about 1,552 millions according to Internet World Stats. This large audience demands online commerce, e-government, knowledge sharing, social networks, online gaming ...which grew exponentially over the past few years. The security of these transactions is very important considering the number of information that could be intercepted by an attacker. Within this context, authentication is one of the most important challenges in computer security. Indeed, the authentication step is often considered as the weakest link in the security of electronic transactions. In general, the protection of the message content is achieved by using cryptographic protocols that are well known and established. The well-known ID/password is far the most used authentication method, it is widely spread despite its obvious lack of security. This is mainly due to its implementation ease and to its ergonomic feature: the users are used to this system, which enhances its acceptance and deployment. Many more sophisticated solutions exist in the state of the art to secure logical access control (one time passwords tokens, certificates ...) but none of them are used by a large community of users for a lack of simplicity usage (O'Gorman, 2003).
Among the different authentication methods of an individual, biometrics is often presented as a promising solution. Few people know that biometrics has been used for ages for identification or signature purposes. Fingerprints were already used as a signature for commercial exchanges in Babylon (-3000 before JC). Alphonse Bertillon proposed in 1879 to use anthropometric information for police investigation. Nowadays, all police forces in the world use this kind of information to solve crimes. The first prototypes of terminals providing an automatic processing of the voice and digital fingerprints have been defined in the middle of the years 1970. Today, a large number of biometric systems are used for logical and physical access control applications. This technology possesses many favorable properties. First, there is a strong link between the user and its authenticator. As for example, it is not possible to loose its fingerprint as it could be the case for a token. Second, this solution is very usable: indeed, it is very convenient for a user to authenticate himself/herself by putting his/her finger on a sensor or making a capture of the face. Last, biometrics is an interesting candidate to be a unique user's authenticator. A study done by NTA group in 2002 (Monitor, 2002) on 500 users showed that there was approximately 21 passwords per user, 81% of them use

common passwords and 30% of them write their passwords in a file. The uniqueness inherent to any biometric information is a helpful property to solve the aforementioned problems.
Of course, some drawbacks are also inherent to this technology (Bolle et al., 2002). Whereas the uniqueness can be considered as an advantage, it could also allow an attacker to trace operations done by an user through the logging of authentication sessions. Then the biometric verification step ensures with a high probability that the user is the genuine one but there is still some possibilities the user is an attacker. This is a far different approach than checking if a password is correct or not. One of the biggest drawbacks of biometrics is the impossibility to revoke the biometric data of a user if they are compromised (Galbally et al., 2008). This point is related to the users acceptance that need to be sure that their privacy will be respected: how can people be sure that their personal data collected during the enrollment step will not be stolen or diverted and used for other purposes ? This pregnant issue limits the operational use of biometrics for many applications. As for example, in France, it is forbidden to establish a centralized database of biometric data because it is considered too dangerous from a privacy point of view.
The objective of this chapter is to realize an overview on the existing methods to enhance the privacy of biometrics. Section 2 is dedicated to a study of the threats involving privacy in biometric systems, and the ensuing requirements. We present in section 3 some biometrics based secure storage and template protection schemes. Section 4 deals with the remaining challenges in this domain. We conclude this chapter in section 5.

2. Privacy: threats and properties

We present in this section privacy issues concerning authentication schemes and biometric ones.

2.1 Privacy and personal data

The word *privacy* means different things to different people; hence the reference (Solove, 2009) has indicated the complexity of defining privacy. Instead of proposing an overview of different conceptual definitions, we formalize a core definition in which privacy means not only keeping a secret but also covering information and activities involving each person. Referring to some jurisdictions like the European Data Protection Directive (Dir95/46/EU), we give the following definitions:

Definition 1. *Personal data is any information relating to an identified or identifiable natural person (data subject).*

Definition 2. *An identifiable person is someone who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.*

It is clear that biometric systems (detailed in the next section) are designed to identify individuals. So, to examine the implication of privacy using biometric data, it is first necessary to define what is a biometric system and to study to what extent biometric data concerns/threats privacy. Then, we will be able to examine whether this data is personal and to measure the amount of sensitive information it reveals.

2.2 On biometric systems

We begin with a theoretical definition.

Definition 3. *A biometric system can be viewed as a signal detection system with a pattern recognition architecture that senses a raw biometric signal, processes this signal to extract a salient set of features called biometric identifier or template and compares these features against the ones stored in the database (Jain et al., 2006).*

More precisely, all biometric systems involve two steps.

- *Enrollment step*
  Biometric data (fingerprint, face, iris…) are *captured*, *transformed* into a template linked to the individual and *stored* as a reference.
- *Verification step*
  A new template is *issued* from a new capture, and *compared* to the stored reference template.

Given any biometric modality (fingerprint, face, iris…), its representation is not unique. As an illustration, consider fingerprint as a subject of study. Then, there are four different wide-spread fingerprint representations:

- Image-based representation
  Two fingerprints are superimposed and the correlation between corresponding pixels is computed for different alignments.
- Minutiae-based representation
  It is the most popular and widely used technique. A fingerprint appears as a surface alternating parallel ridges and valleys in most regions. Minutiae represent local discontinuities and mark positions where the ridge ends or splits. Minutiae-based matching consists in finding the alignment that results in the maximum number of minutiae pairings.
- Ridge feature-based approach
  Other features of the fingerprint ridge pattern (e.g., local orientation and frequency, ridge shape, texture information) may be extracted more reliably than minutiae in low-quality images.
- Pores-based representation
  With the current development of high resolution sensors, fine fingerprint features such as sweat pores can be considered.

To study privacy issues involved in biometric systems, we explore now how a biometric identifier is personal and sensitive. Two types of errors are present at the verification step:

- *false match*: the verification process outcome is that biometric measurements from two different persons are from the same person
- *false non-match*: the verification process outcome is that two biometric measurements from the same person are from two different persons

These two types of errors are quantified by the *false acceptance rate* and the *false rejection rate*, respectively. Figure 1 presents the error rates of four popular biometric modalities.

2.3 Biometrics and privacy

Biometric data, in its raw or template form (like minutiae template), is in most cases personal data. The reference (Pakanti et al., 2002) estimated a probability that two fingerprints will falsely match as $5.5 * 10^{-59}$. This probability is very low and shows that minutiae information can uniquely identify a person. In practice, as deduced from figure 1, deployment of biometric systems does not imply that the recognition is a fully solved problem. The accuracy changes

Biometric trait	Test	False Rejection Rate	False Acceptance Rate
Fingerprint	FVC 2006	2.2%	2.2%
	FpVTE 2003	0.1%	1%
Face	FRVT 2006	0.8-1.6%	0.1%
Voice	NIST 2004	5-10%	2-5%
Iris	ICE 2006	1.1-1.4%	0.1%

Fig. 1. Illustrations of error rates for different biometric modalities (Teoh et al., 2004b)

depending on different factors (the used modality, the population characteristics, the test conditions and the employed sensor to mention a few) but is never perfect. However, the obtained performances are considered sufficient to conclude that biometric data identifiers can recognize persons. Thus, they are *personal* or *very personal*, in the sense that they consist of information collected from an observation of the individual physical itself.
In return, biometric data are generally considered as sensitive data involving ethical and privacy contests.

2.3.1 Privacy threats in biometric systems

We summarize below potential privacy pitfalls arising when using a biometric identifier (fingerprint modality being again focused on):

1. Biometric information (especially raw images) can expose sensitive information such as information about one's health, racial or ethnic origin and this information can then provide a basis for unjustified discrimination of the individual data subjects (Mordini & Massari, 2008).
2. As revealed in (Schneier, 1999), biometric data are unique identifiers but are not secret: fingerprint is leaved on everything we touch, faces can be easily acquired and voice can be simply recorded. Hence, the potential collection and use of biometric data without knowledge of its owner, without his/her consent or personal control make this information very sensitive.
3. Many proponents of biometric systems claim that it is sufficient to store a compact representation of the biometric (template) rather than the raw data to ensure privacy of individuals. They consider that template is not sensitive information because it does not allow the reconstruction of the initial signal. Recently, several research works showed that this reconstruction is possible. For example, fingerprint can, in fact, be reconstructed from a minutiae template (Cappelli et al., 2007), (Feng & Jain, 2009).
4. The linkage problem which means the possibility to cross matched data across different services or applications by comparing biometric references is another privacy concern. The uniqueness of biometric characteristics allows an intruder to link users between different databases, enabling violations as tracking and profiling individuals.
5. A function creep is another privacy risk. Here, the acquired biometric identifiers are later used for purposes different from the intended ones. For example, an application initially intended to prevent misuse of municipal services may gradually be extended to rights to buy property, to travel, or the right to vote without the consent of individuals.
6. The inherent irrevocability of biometric features in case of data misuse like database compromise or identity theft makes biometrics very sensitive.

With the present risks on privacy violation, carefully handling biometric data becomes more important. Considering the implication of personal sensitive data, the use of biometrics falls within the purview of legislation and laws. In reality, regulations and legislation have codified what Judge Samuel Warren and Louis Brandeis summarized in 1890 as the right of the individual to be alone (Warren & Brandeis, 1890) (this reference is considered as the birthplace of privacy rights), and expanded the notion of data protection beyond the fundamental right to privacy. In the sequel, we are interested in the main attack vectors concerning biometric systems.

2.3.2 Biometric attack vectors

Possible attacks points (or attack vectors) in biometric systems have been discussed from different viewpoints. First, we can mention the scheme of figure 2, provided by the international standard ISO /IEC JTC1 SC37 SD11, which identifies where possible attacks can be conducted.

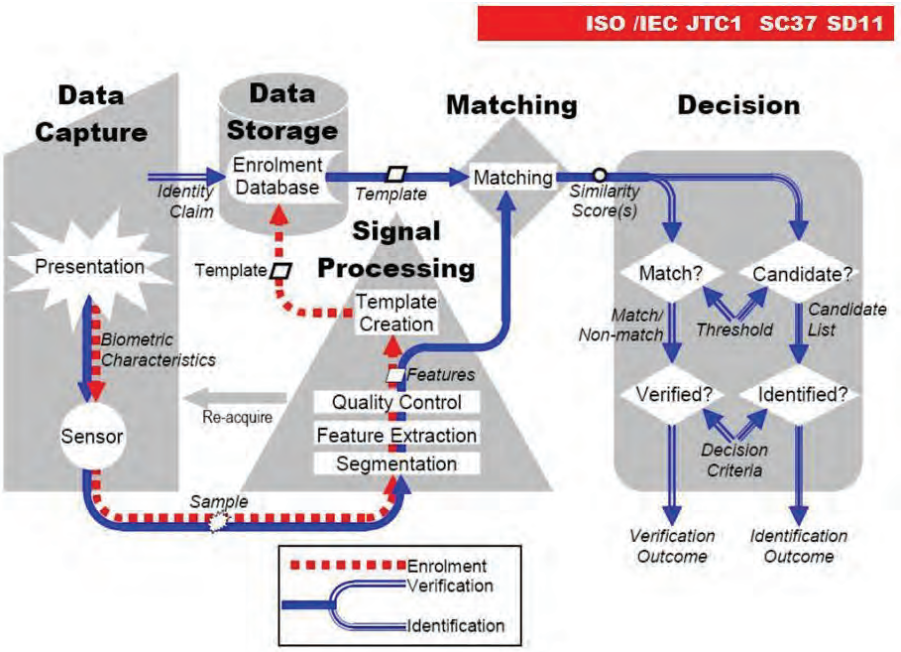

Fig. 2. ISO description of biometric systems

Besides, some of the early works by Ratha, Connell and Bolle (Ratha et al., 2001), (Bolle et al., 2002) identified weak links in each subsystem of a generic authentication system. Eight places where attacks may occur have been identified, as one can see in figure 3.
We do not detail in the present chapter all the types of attacks identified by Ratha. We only focus on attacks concerning privacy. This corresponds to the points 6 and 7 in figure 3. These points are related to attacks violating template protection. Generally, attacks directly threatening biometrics template can be of different types. For instance, an attacker could:

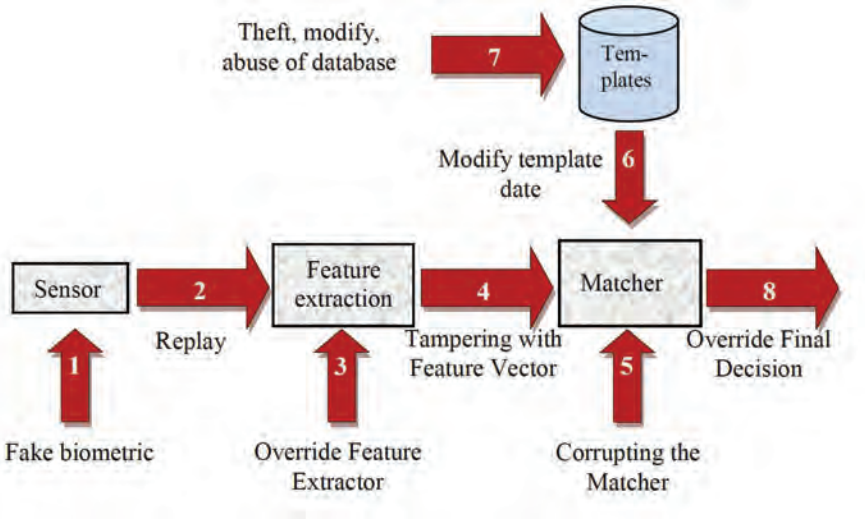

Fig. 3. Ratha's model attack framework

- attempt to capture a reference template
- substitute a template to create a false reference
- tamper the recorded template
- compromise the database by stealing all its records

Such attacks can be very damaging, owing to unavoidable exposure of sensitive personal information and identity theft. Therefore basic requirements that any privacy preserving biometric system should fulfill will be stated in the following.

2.3.3 Requirements of privacy protection

In view of our discussion about biometric systems vulnerabilities and possible threats, a few desirable properties are required, regarding the system safety. A critical issue in the biometric area is the development of a technology to handle both privacy concerns and security goals, see (Jain et al., 2008) for example. We detail now the key considerations for privacy protection.

- All deployments of biometric technology should be implemented with respect to local jurisdictional privacy laws and regulations.
- Today, some legal frameworks introduce the idea of *Privacy by Design*. This new paradigm requires that privacy and data protection should be integrated into the design of information and communication technologies. The application of such principle would emphasize the need to implement Privacy Enhancing Technologies (PET) that we will see after.

As explained in the reference (Adler, 2007), privacy threat is closely related to security weakness. Therefore a particular attention has been paid to privacy enhancing techniques. The aim is to combine privacy and security without any tradeoff between these two basic requirements. Among the techniques related to privacy enhancing, we can mention recent trends:

- Biometric encryption
  Based on cryptographic mechanisms, the ANSI (American National Standards Institute) proposes X9.84 standard as a means to manage biometric information. ANSI X9.84 rules were designed to maintain the integrity and confidentiality of biometric information using encryption algorithms. Even if cryptography has proven its ability to secure data transmission and storage, it becomes inadequate when applied to biometric data. Indeed, owing to the variability over multiple acquisitions of the same biometric trait, one cannot store a biometric template in an encrypted form and then perform matching in the encrypted domain: cryptography is not compatible with intra-user variability. Therefore the comparison is always done in the biometric feature domain which can make it easier for an attacker to obtain the raw biometric data. Since the keys and *a fortiori* the biometric data are controlled by a custodian, most privacy issues related to large databases remain open.
- Template protection schemes
  To solve this problem, recently some algorithms known as template protection schemes have been proposed. These techniques, detailed in section 3.1, are the most promising for template storage protection.
- Anonymous database
  The idea in anonymous data is to verify the membership status of a user without knowing his/her true identity. A key question in anonymous database is the need for secure collaboration between two parties: the biometric server and the user. The techniques presented in sections 3.2 and 3.3 fulfill this requirement.

In this chapter, privacy protection is be considered from two points of view: trusted systems and template protection. Recently, some template protection schemes have been proposed. Ideally, these algorithms aim at providing the following properties as established in the reference (Maltoni et al., 2009), and fulfill the privacy preserving issues 1 to 6 raised at page 4.

- *Non-reversibility*
  It should be computationally infeasible to obtain the unprotected template from the protected one. One of the consequences of this requirement is that the matching needs to be performed in the transformed space of the protected template, which may be very difficult to achieve with high accuracy. This property concerns points 1 to 3.
- *Accuracy*
  Accuracy recognition should be preserved (or degraded smoothly) when protected templates are involved. Indeed, if the accuracy of recognition degrades substantially, it will constitute the weakest link in the security equation. For example, instead of reversing the enrolment template, the hacker may try to cause a false acceptance attack. Thus, it is important that the protection technique does not substantially deteriorate the matching accuracy. This property is a general one.
- *Cancelability and Diversity*
  It should be possible to produce a very large number of protected templates (to be used in different applications) from the same unprotected template. This idea of cancelable biometrics was established for the first time in the pioneering references (Ratha et al., 2001) and (Bolle et al., 2002). To protect privacy, diversity means the impossibility to match protected templates from different applications (this corresponds to the notion of non linkage). Points 4 and 5 are concerned with diversity while point 6 with cancelability.

Compared to (Maltoni et al., 2009), we wish to add an extra property which will be used in the sequel:

- *Randomness*
  The knowledge of multiple revoked templates does not help to predict a following accepted one. This property deals with points 3 and 4.

Since some basic requirements in terms of privacy protection have been stated, biometric techniques fulfilling these requirements are detailed in the next section.

3. Privacy enhancing biometric techniques

In this section, we focus on some recent solutions brought by researchers in biometrics to handle template protection issues. These solutions generally aim at guaranteeing privacy and revocability. First, we detail promising solutions concerned with the storage of biometric templates in secure elements. Then, we emphasize on two approaches dealing with privacy enhancing biometric techniques: the first one is called biometric cryptosystems and the second is known as BioHashing.

3.1 Storage in secure elements

A key question is in relation with the place of storage of data and its security: Is it conserved in a local way (e.g. token)? Or in a central database with different risks of administration, access and misuse of this database? The problem becomes more relevant when dealing with large scale biometric projects such as the biometric passport or the national electronic identity card. Different organisations like the CNIL in France warn against the creation of such databases especially with regard to modality with traces as is the case for fingerprint (it is possible to refer to the central database to find the identity of those who left their traces). The INES debate launched in 2005 is a good illustration of the awareness of such subject (Domnesque, 2004). The use of biometrics may pose significant risks that encourage link ability and tracing of individuals and hence violating the individual liberties. So, the security of the stored biometric data remains challenging and this crucial task is pointed out by experts and legislation authorities. In this section, we study the storage of data in a secure local component: the *Secure Element*.
In (Madlmayr et al., 2007) one can find the following definition:

Definition 4. *The* Secure Element *is a dynamic environment, where applications are downloaded, personalized, managed and removed independently with varying life cycles.*

It is mainly used in smart cards or mobile devices to host sensitive data and applications such as biometrics templates or payment applications. It allows a high level of security and trust (e.g. the required security level for payment applications is set to Common Criteria EAL5+).
The Secure Element can be seen as a set of logical components: a microprocessor, some memory, an operating system and some applications. The secure element operating systems are known as MULTOS, Windows for Smart Cards, ZeitControl, SmartCard .NET and the most widespread: Java Card. Java Card is an adaptation of the well-known Java technology to smart card constraints. Java Card is an open language, which explains its great success. Based on a virtual machine environment, it is very portable (following the famous *Write Once, Run Everywhere*) and allows several applications to be installed and run on the same card.
But some drawbacks are also inherent to this technology: indeed the cohabitation of applications raises some questions. How and when to load the applications? Shall

applications loading be secured ? How to isolate applications from each others ? How long is the life cycle of a single application on the card ? How to determine the privileges of an application ?... Answers to these issues are provided by the GlobalPlatform technology.

3.1.1 GlobalPlatform overview

The GlobalPlatform technology is the fruit of the GlobalPlatform consortium's work. The GlobalPlatform consortium (formerly named Visa Open Platform) is an organization established in 1999 by leading companies from the payment and communications industries, the government sector and the vendor community. The GlobalPlatform specifications cover the entire smart card infrastructure: smart cards, devices and systems. Consistently written, this set of specifications allows developing multi-applications and multi-actors smart cards systems. It specifies the technical models that meet the business models requirements.
The GlobalPlatform card specification (GlobalPlatform, 2006) defines the behavior of a GlobalPlatform Card. As it is shown in figure 4, the GlobalPlatform card architecture comprises security domains and applications. A *Security Domain* acts as the on-card representatives of off-card entities. It allows its owner to control an application in a secure way without sharing any keys nor compromising its security architecture. There are three main types of Security Domain, reflecting the three types of off-card authorities recognized by a card: *Issuer Security Domain*, *Application Provider Security Domain* and *Controlling Authority Security Domain*.

3.1.2 Application to biometric template secure storage

The secure storage of biometric templates on a GlobalPlatform card is realized by an application. This dedicated application is installed, instantiated, selected and pushed a reference biometric template. In the verification case, the minutia are pushed to the application which processes a *match-on-card* verification and returns the result to the outside world.
In order to host this application, an Application Provider Security Domain must previously be created on the card. This security domain is personalized with a set of cryptographic keys and can then provide the application with security services such as cryptographic routines support and secure communications.
The application is involved in both the user enrolment and verification. For those two phases, the application queries the security services of its associated security domain.
During the user enrolment process, a secure communication channel is established between an off-card entity (in the present case a personalization bureau) and the application intended to store the reference biometric template. To this purpose, the security domain handles the handshake between the off-card entity and the application and unwraps the ciphered biometric template prior to forwarding it to the application. Three security levels are available for the secure communication: authentication, integrity and confidentiality.
During the verification process, a secure communication channel is established following the previous scheme. Contrary to the enrolment step, in this phase, the minutiae (and not the reference template) are ciphered and sent to the application. Hence the verification process takes place on card in a secure manner.
We have seen in this section how the Secure Element, a tamper proof component, ensures the secure storage of biometric templates. Indeed, thanks to the GlobalPlatform architecture, the access to the application devoted to the storage of the template and performing the match-on-card verification is secured successively by authentication of the off-card entity, check of data integrity and data ciphering for confidentiality purpose.

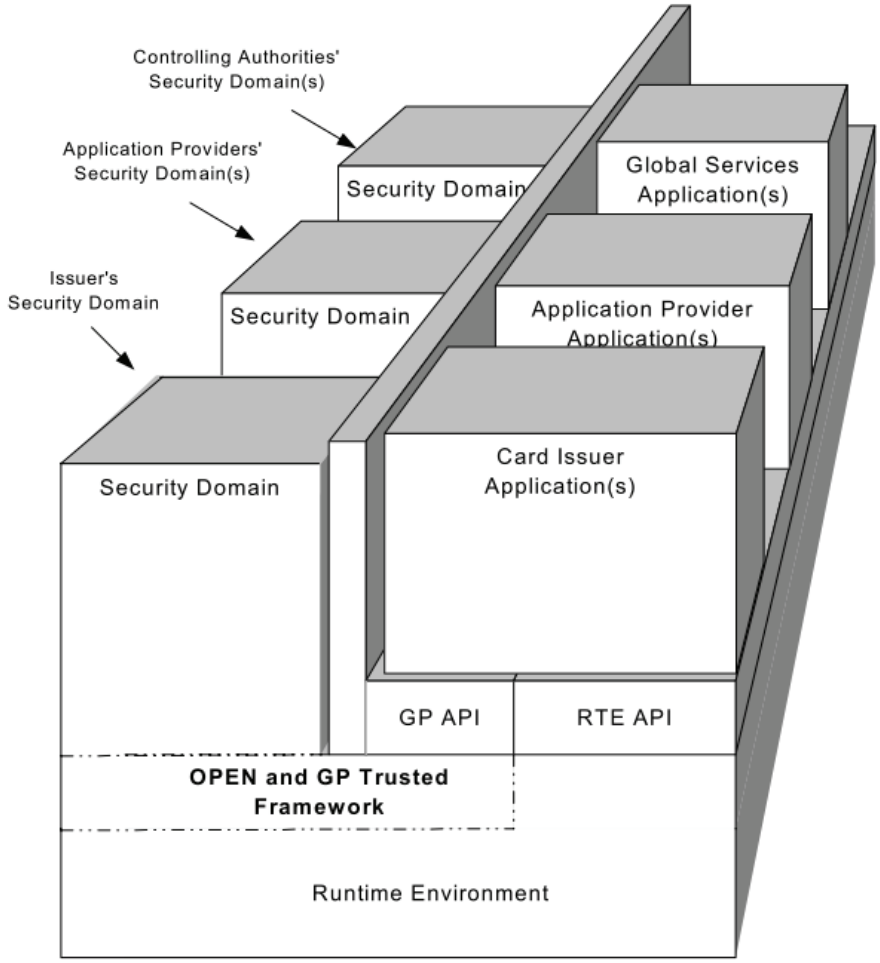

Fig. 4. GlobalPlatform Card Architecture (source: GlobalPlatform)

The next sections are concerned with two algorithmic-based solutions dealing with biometric template protection.

3.2 Cryptographic based solutions

Secure sketches have been introduced by Dodis *et al.* and formalized for a metric space H and the associated distance d, in relation to biometric authentication in (Dodis et al., 2004), (Dodis et al., 2006). A secure sketch considers the problem of error tolerance, existing in biometric authentication context: a template $b \in H$ must be recovered from any sufficiently close template $b' \in H$ and a additional data P. At the same time, the additional data P must not reveal too much information on the original template b. It uses the notion of minimal

entropy $H_\infty(X)$ of a random variable X, defined by the maximal number k such that for all $x \in X$, the probability $P(X = x) \leq 2^{-k}$.

Definition 5. *A (H, m, m', t)-secure sketch is a pair of functions SS and Rec such as:*

- *The randomized function SS takes as input a value $b \in H$ and outputs a public sketch in $\{0,1\}^*$, such that for all random variable B in H, with minimal entropy $H_\infty(B) \geq m$, the conditional minimal entropy $H_\infty(B|SS(B)) \geq m'$.*
- *The deterministic function Rec takes as input a sketch $P = SS(b)$ and a value $b' \in H$ and outputs a value $b'' \in H$ such that $b'' = b$ if the distance $d(b, b') \leq t$.*

The first secure sketch was proposed by Juels et Wattenberg in (Juels & Wattenberg, 1999). This scheme is called *fuzzy commitment* and uses error-correcting codes. The *fuzzy vault* scheme of Juels and Sudan (Juels & Sudan, 2001) is also a secure sketch in an other metric.
A binary linear $[n, k, d]$ code C is a vectorial sub-space of $\{0,1\}^n$ having a dimension k and composed of vectors x having a Hamming weight $w_H(x) \geq d$, where $w_H(x)$ is the number of non-zero coordinates of x. The correction capacity of this code is $t = \lfloor (d-1)/2 \rfloor$. More details on error-correcting codes are given in the book (MacWilliams & Sloane, 1988).
In this construction, the metric space is $\{0,1\}^n$, with the Hamming distance d_H. Let C be a binary linear code with parameters $[n, k, 2t+1]$. Then a $(\{0,1\}^n, m, m-(n-k), t)$-secure sketch is designed as follows:

- The function SS takes as input a value $b \in \{0,1\}^n$ and outputs a sketch $P = c \oplus b$, where $c \in C$ is a randomly chosen codeword.
- The function Rec takes as input a sketch P and a value $b' \in \{0,1\}^n$, decodes $b' \oplus P$ in a codeword c' et returns the value $c' \oplus P$.

The following authentication system is directly related to the previous secure sketch:

Biometric authentication with fuzzy commitment

1. **Enrollment:** the user registers his biometric template b and sends the sketch $P = c \oplus b$, with $H(c)$ to the database, where H is a cryptographic hash function and $c \in C$ is a codeword randomly chosen.
2. **Verification:** the user sends a new biometric template b' to the database which computes $P \oplus b'$. Then the database decodes it in a codeword c' and checks if $H(c) = H(c')$. In cas of equality, the user is authenticated.

According to the minimum distance $2t+1$ of the code, the new biometric template b' is accepted if and only if the Hamming distance $d_H(b, b') \leq t$. The authentication system is based on the following property: if the distance $d_H(b, b') = \epsilon$ is lower than the correction capacity of the code, then it is possible to recover the original codeword c from the word $c \oplus \epsilon$.
Applications of this protocol are proposed in face recognition (Kevenaar et al., 2005) or fingerprints (Tuyls et al., 2005), using BCH codes. This fuzzy commitment scheme is also used for iris recognition, where iris templates are encoded by binary vectors of length 2048, called IrisCodes (Daugman, 2004a;b). For example a combination of Hadamard and Reed-Solomon codes is proposed in (Hao et al., 2006), whereas a Reed-Muller based product code is chosen in (Chabanne et al., 2007).
The previous scheme ensures the protection of the biometric template if the size of the code C is sufficient, whereas the loss of entropy of the template is directly connected to the

redundancy of the code. Security of this system is however limited: biometrics templates are not perfectly random and their entropy is difficult to estimate. Moreover, the protection of the biometric template is related to the knowledge of the random codeword c. This codeword is direclty used by the database during the verification phase.
In order to enhance the security of the previous scheme, Bringer *et al.* have proposed a combination of a secure sketch with a probabilistic encryption and a PIR protocol[1] in (Bringer & Chabanne, 2009; Bringer et al., 2007). The following biometric authentication protocol gives a simplified description of their scheme (without PIR protocols) and illustrates nicely the possibilities proposed by homomorphic encryptions for privacy enhancement in biometric authentication.
The Goldwasser-Micali probabilistic encryption scheme is the first probabilistic encryption scheme proven to be secure under cryptographic assumptions (Goldwasser & Micali, 1982; 1984). The semantic security of this scheme is related to the intractability of the quadratic residuosity problem . The Goldwasser-Micali scheme is defined as follows:

Definition 6. *Let p and q be two large prime numbers, N be the product $p.q$ and x be a non-residue with a Jacobi symbol 1. The public key of the crypto-system is $p_k = (x, N)$ and the private key is $s_k = (p, q)$. Let y be randomly chosen in $\mathbf{Z}_n^*$. A message $m \in \{0, 1\}$ is encrypted in c, where $c = Enc(m) = y^2 x^m \bmod n$. The decryption function Dec takes an encrypted message c and returns m, where $m = 0$ if c is a quatratic residue and 1 otherwise.*

This scheme encrypts a message bit by bit. The encryption of a message of n bits $m = (m_1, \ldots, m_n)$ with the previous scheme is denoted $Enc(m) = (Enc(m_1), \ldots, Enc(m_n))$, where the encryption mechanism is realized with the same key. The Goldwasser-Micali scheme clearly verifies the following property:

$$\mathrm{Dec}(\mathrm{Enc}(m, pk) \times \mathrm{Enc}(m', pk), sk) = m \oplus m'.$$

This homomorphic property is used for the combination of this cryptosystem with the secure sketch construction of Juels and Wattenberg.
The biometric authentication scheme uses the following component: the user U who needs to be authenticated to a service provider SP. The service provider has access to a database where biometrics templates are stored. These templates are encrypted with cryptographic keys, generated and stored by a key manager KM who has no access to the database. For privacy reasons, the service provider SP has never access to the private keys.
For each user U, the key manager KM generates a pair (p_k, s_k) for the Goldwasser-Micali scheme. The public key p_k is published and the private key is stored in a secure way. The biometric authentication system is described as follows:

Biometric authentication with homomorphic encryption

1. **Enrollment:** The user U registers his biometric template b to the service provider. The service provider randomly generates a codeword $c \in C$, computes $H(c)$ where H is a cryptographic hash function and encrypts $Enc(c \oplus b)$ with the Goldwasser-Micali scheme and the public key p_k, and finally stores it in the database.
2. **Verification:** the user U encrypts his biometrics template $Enc(b')$ with p_k and sends it to the service provider. The service provider recovers $Enc(c \oplus b)$ and $H(c)$ from the database, computes and sends the products $Enc(c \oplus b) \times Enc(b')$ to the key manager. The key manager decrypts

[1] Private Information Protocol, see (Chor et al., 1998).

$Dec(Enc(c \oplus b) \times Enc(b')) = c \oplus b \oplus b'$ with its private key s_k and sends the result to the service provider who decodes it in a codeword c'. The service provider finally checks if $H(c) = H(c')$.

Homomorphic property of the Goldwasser-Micali scheme ensures that biometrics templates are never decrypted during the verification phase of the authentication protocol. Moreover, the service provider who has access to the encrypted biometric data does not possess the private key to retrieve the biometric templates and the key manager who generates and stores the private keys has never access to the database.
Other encryption schemes with suitable homomorphic property can be used as the Paillier cryptosystem (Paillier, 1999) or the Damgard-Jurik cryptosystem (Damgard & Jurik, 2001). Homomorphic cryptosystems have been recently used for several constructions of privacy-compliant biometric authentication systems. For example, a face identification system is proposed in (Osadchy et al., 2010), whereas iris and fingerprint identification mechanisms are described in (Barni et al., 2010; Blanton & Gasti, 2010).

3.3 BioHashing

The previous cryptosystems represent promising solutions to enhance the privacy. However, the crucial issues of cancelability and diversity seem to be not well addressed by these techniques (Simoens et al., 2009).
Besides biometric cryptosystems design, transformation based approaches seem more suited to ensure the cancelability and diversity requirements and more generally, fulfill the additional points raised page 7: non-reversibility, accuracy and randomness. The principle of transformation based methods can be explained as follows: instead of directly storing the raw original biometric data, it is stored after transformation relying on a non-invertible function. So, the prominent feature shared by these techniques takes place at the verification stage, which is performed in the transformation field, between the stored template and the newly acquired template. Moreover, these techniques are able to cope with the variability inherent to any biometrics template.
The pioneering work (Ratha et al., 2001) introduces a distortion of the biometric signal by a chosen transformation function. Hence, cancelability is ensured: each time a transformed biometric template is compromised, one has just to change the transformation function to generate a new transformed template. The diversity property is also guaranteed, since different transformation functions can be chosen for different applications.
Among the transformation based approaches, we detail in this chapter the principle of BioHashing. BioHashing is a two factor authentication approach which combines pseudo-random number with biometrics to generate a compact code per person. The first work referencing the BioHashing technique is presented on face modality in (Goh & Ngo, 2003). Then the same technique has been declined to different modalities in the references (Teoh et al., 2004c), (Teoh et al., 2004a), (Connie et al., 2004) and more recently (Belguechi, Rosenberger & Aoudia, 2010), to mention just a few.
Now, we detail the general principle of BioHashing.

3.3.1 BioHashing principle

All BioHashing methods share the common principle of generating a unitary BioCode from two data: the biometric one (for example texture or minutiae for fingerprint modality) and a random number which needs to be stored (for example on a usb key, or more generally on a token), called *tokenized random number*. The same scheme (detailed just below) is applied both:

- at the enrollment stage, where only the BioCode is stored, instead of the raw original biometric data
- at the verification stage, where a new BioCode is generated, from the stored random number

Then the verification relies on the computation of the Hamming distance between the reference BioCode and the newly issued one. This principle allows BioCode cancelability and diversity by using different random numbers for different applications.
More precisely, the BioHashing process is illustrated by the figure 5. One can see that it is a two factor authentication protection scheme, in the sense that the transformation function combines a specific random number whose seed is stored in a token with the biometric feature expressed as a fixed-length vector $F = (f_1, \ldots, f_n)$, $F \in \mathbb{R}^n$.

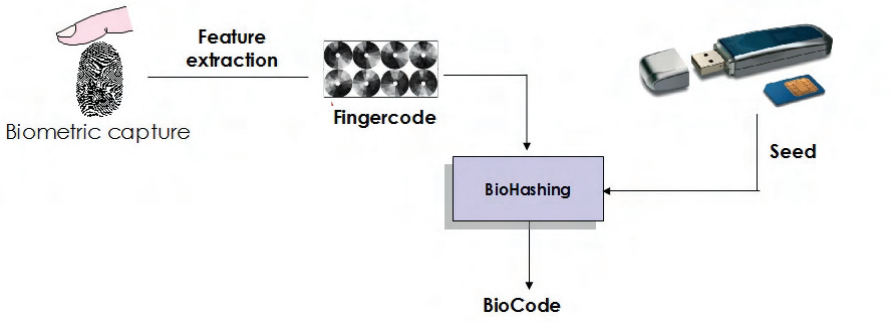

Fig. 5. BioHashing: Ratha's method

Then BioHashing principle, detailed in (Belguechi, Rosenberger & Aoudia, 2010) for example, consists in the projection of the (normalized) biometric data on an orthonormal basis obtained from the random number. This first step somehow amounts to hide the biometric data in some space. The second step relies on a quantization which ensures the non-invertibility of BioHashing: from the final BioCode, it is impossible to obtain the original biometric feature.
Let us give more details on the involved stages: random projection and quantization.

- *Random projection*
  It has been shown in (Kaski, 1998) that random mapping can preserve the distances in the sense that the inner product (which represents a way of measuring the similarity between two vectors from the cosine of the angle between them) between the mapped vectors closely follows the inner product of the initial vectors. One condition is that the involved random matrix R consists of random values and the Euclidean norm of each column is normalized to unity. The reference (Kaski, 1998) proves that the closer to an orthonormal matrix the random matrix R is, the better the statistical characteristics of the feature topology are preserved. As a consequence, in the BioHashing process, the tokenized random number is used as a seed to generate m random vectors r_i, $i = 1, \ldots, m$. After orthonormalization by the Gram-Schmidt method, these vectors are gathered as the column of a matrix $O = \left(O_{ij}\right)_{i,j \in [1,n] \times [1,m]}$.

  The following Johnson-Lindenstrauss lemma (1984), studied in (Dasgupta & Gupta, 1999), (Teoh et al., 2008) is at the core of the BioHashing efficiency:

Lemma 1. *For any* $0 < \varepsilon < 1$ *and any integer* k*, let* m *be a positive integer verifying* $m \geq \frac{4\log(k)}{\varepsilon^2/2-\varepsilon^3/3}$*. Then, for any set* S *containing* k *points in* $\mathbb{R}^n$*, there exists a map* $f : \mathbb{R}^n \to \mathbb{R}^m$ *such that:*

$$\forall x, y \in S,\ (1-\varepsilon)||x-y||^2 \leq ||f(x)-f(y)||^2 \leq (1+\varepsilon)||x-y||^2 \quad (1)$$

In other words, Johnson-Lindenstrauss lemma states that any n point set in Euclidian space can be embedded in suitably high (logarithmic in k, independent of n) dimension without distorting the pairwise distances by more than a factor of $1 \pm \varepsilon$. As a conclusion of this first step, we can say that the pairwise distances are well conserved by random projection under the previous hypotheses on the random matrix. Notice that this distance conservation becomes better when m increases, therefore one can consider $m = n$.

The resulting vector is denoted $W = (W_1, \ldots, W_m)$, with $W = F.O \in \mathbb{R}^m$, see figure 6.

- *Quantization*
  This step is devoted to the transformation in a binary-valued vector of the previous real-valued vector resulting from the projection of the original biometric data on an orthonormalized random matrix. A reinforcement of the non-invertibility (also relying on the random projection process) of the global BioHashing transformation ensues from this quantization. It requires the specification of a threshold τ to compute the final BioCode $B = (B_1, \ldots, B_m)$ following the formula:

$$B_i = \begin{cases} 0 \text{ if } W_i \leq \tau \\ 1 \text{ if } W_i > \tau \end{cases} \quad (2)$$

  In practice, the threshold τ is chosen equal to zero so that half of W_i are larger than the threshold and half smaller. This, in order to maximize the information content of the extracted m bits and to increase the robustness of the resultant template. To this purpose, one may compute the median of the referenced vectors W and use it as a threshold.

These two steps are illustrated by the figure 6.

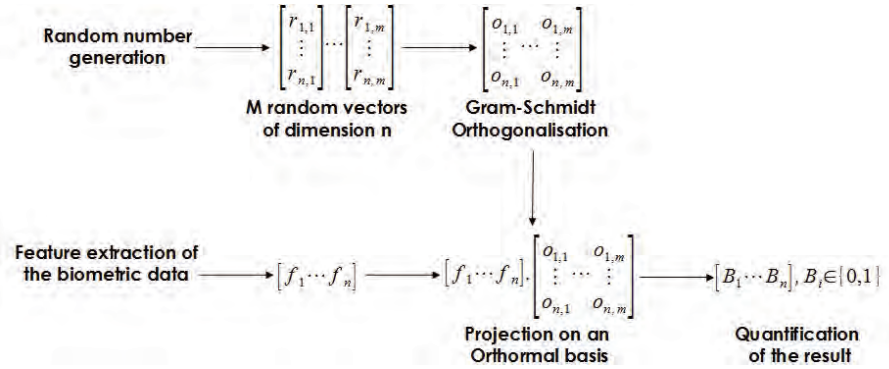

Fig. 6. BioCode generation

In the literature, one can find that the previous general technique has been applied to several biometric modalities to obtain the biometric template F. In (Teoh et al., 2004a), integrated Wavelet Fourier-Mellin transform is applied to fingerprint raw image. This requires the a

priori detection of the core point and produces a translation and rotation-invariant feature. Besides, integrated Wavelet Fourier-Mellin transform has been applied to face raw image in (Teoh & Ngo, 2005), while Fisher Discriminant Analysis (FDA) for face images is developed in (Teoh et al., 2004b), with a slightly different quantification step. Both Principal Component Analysis (PCA) and Fisher Discriminant Analysis are at the core of PalmHashing developed in (Connie et al., 2004) from the ROI of the raw palmprint image. In two recent papers (Belguechi, Rosenberger & Aoudia, 2010), (Belguechi, Hemery & Rosenberger, 2010), we propose to extract biometric templates from minutiae representation, by using a Gabor filterbank, after a detection process of the region of interest.

3.3.2 Discussion

The conclusion shared by the previously mentioned references is that BioHashing has significant advantages over solely biometrics or token usage, such as extremely clear separation of the genuine and the imposter population and zero EER (Equal Error Rate) level. But, among other papers, the reference (Kong et al., 2005) reveals that the outstanding 0% EER achievement of BioHashing implies the unrealistic assumption that the tokenized random number (TRN) would never be lost, stolen, shared or duplicated. In this paper, it is also pointed out that if this assumption held, the TRN alone could serve as a perfect password, making biometrics useless. The presented results show that the true performance of BioHashing is far from perfect and even can be worst than the basic biometric system. The results of some tests on different modalities are given in (Lumini & Nanni, 2006). For fingerprint texture template for example, the authors have demonstrated that the performance of the system in term of EER moves from 7.3% when no hashing is performed to 10.9% when basic BioHashing is operated under the hypothesis that the token is always stolen, while EER is evaluated to 1.5% in case where no TRN is stolen (FVC2002-DB2). These scores are also discussed in our papers (Belguechi, Rosenberger & Aoudia, 2010), (Belguechi, Hemery & Rosenberger, 2010) where different cases are considered, depending on whether the token is stolen or not.

3.4 Summarization

We saw in the previous sections different solutions to protect the biometric information. On the one hand using a Secure Element to store a biometric template is one convenient solution and is already operational. Even if it is technically possible to cancel a biometric template (by updating the content of the SE), this solution does not give any guarantee about the required cancelability properties. The security of a SE is often evaluated by a certification level (as for example EAL4+ for common criteria) giving some elements about the possibility for an hacker to obtain the biometric template.

On the other hand algorithmic solutions propose nice solutions to protect biometric templates privacy. Cryptography based approaches avoid the transmission of biometric templates but does not solve the non revocability problem. BioHashing reveals itself as a promising solution and seems to respect many privacy properties defined previously. This approach needs to be further studied, especially considering attacks.

4. Research challenges

Even if some solutions exist, there are many challenges to deal with in the future.

How can we evaluate cancelable biometric systems ?

Before proposing new privacy protection schemes for biometric templates, it becomes urgent to define objective evaluation methods for these particular systems. Of course, this type of

biometric systems can be evaluated through existing standards in performance evaluation (see (El-Abed et al., 2010) for example) but it is not sufficient. Computing the EER value or the ROC curve does not give any information on how the system protects the privacy of users. Some researchers try to analyze the security and privacy of these systems by taking into account some scenarios. The robustness to an attack is often quantified as for example by the resulting EER or FRR values when the attacker caught some additional information that he/she was not supposed to have. There is a lot of work on this subject.

How to increase the BioCode length?

In order to prevent brute force attack consisting in testing different values of the BioCode, it is necessary to increase the size of the generated BioCode. Many solutions to this problem exist. First, one simple solution is to use different biometric information. One can generate a BioCode for the fingerprint of each hand finger. There is no reason to have a statistical correlation between information provided by the template of each fingerprint. This solution solves the problem of the size and the associated entropy but it is less usable for an user as he/she has to provide as for example the fingerprint of each hand. Second, it is possible to increase the size of the BioCode by using an adapted representation. As for example, computing a BioCode from minutiae (where 30 are detected in average for a fingerprint) provides smaller BioCode compared to a texture representation. So it is necessary to carefully study biometric data representation.

How many times can we cancel a BioCode ?

The objective of a cancelable biometric information is to be able to generate it again in case of known attack. The question is to quantify the possibility to revoke this data a certain amount of times. Suppose an attacker is listening to the authentication session and can have different values of the BioCode, the problem is to know if he/she is able to predict an authorized BioCode. It is necessary to analyze as for example if some bits keep the same value in the BioCode after regeneration.

To what extent is it usable in an operational context ?

There are some (not so much) publications in this domain but very few industrial solutions (except those using a SE). This domain is not enough mature. Using a secure element to store the biometric data and realizing a match on card is well known. It could be interesting to combine a hardware solution using a secure element and an algorithmic solution. We are currently working on this aspect. The next step is also to be able to make a capture on card with an embedded biometric sensor to limit the transmission of the biometric data.

5. Conclusion

Biometrics is a very attractive technology mainly because of the strong relationship between the user and its authenticator. Unfortunately, many problems are also associated with this authentication solution. The main one concerns the impossibility to revoke a biometric data. Besides there is a major concern for ethical and security reasons. We presented in this chapter the main issues in this field and some solutions in the state of the art based on secure storage of the biometric template or using algorithmic solutions. Even if these methods bring some

improvements, many things need to be done in order to have a totally secure solution. We detailed some trends to work on in the near future.

6. References

Adler, A. (2007). *Biometric system security*, Handbook of biometrics, Springer ed.

Barni, M., Bianchi, T., Catalano, D., Raimondo, M. D., Labati, R. D., Failla, P., Fiore, D., Lazzeretti, R., Piuri, V., Piva, A. & Scotti, F. (2010). A privacy-compliant fingerprint recognition system based on homomorphic encryption and fingercode templates, *BTAS 2010*.

Belguechi, R., Hemery, B. & Rosenberger, C. (2010). Authentification révocable pour la vérification basée texture d'empreintes digitales, *Congrès Francophone en Reconnaissance des Formes et Intelligence Artificielle (RFIA)*.

Belguechi, R., Rosenberger, C. & Aoudia, S. (2010). Biohashing for securing minutiae template, *Proceedings of the 20th International Conference on Pattern Recognition*, Washington, DC, USA, pp. 1168–1171.

Blanton, M. & Gasti, P. (2010). Secure and efficient protocols for iris and fingerprint identification, Cryptology ePrint Archive, Report 2010/627. `http://eprint.iacr.org/`.

Bolle, R., Connell, J. & Ratha, N. (2002). Biometric perils and patches, *Pattern Recognition* 35(12): 2727–2738.

Bringer, J. & Chabanne, H. (2009). An authentication protocol with encrypted biometric data, *AfricaCrypt'09*.

Bringer, J., Chabanne, H., Izabachène, M., Pointcheval, D., Tang, Q. & Zimmer, S. (2007). An application of the Goldwasser-Micali cryptosystem to biometric authentication, *ACISP'07*, Vol. 4586 of *Lecture Notes in Computer Science*, Springer, pp. 96–100.

Cappelli, R., Lumini, A., Maio, D. & Maltoni, D. (2007). Fingerprint image reconstruction from standard templates, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(9): 1489–1503.

Chabanne, H., Bringer, J., Cohen, G., Kindarji, B. & Zemor, G. (2007). Optimal iris fuzzy sketches, *IEEE first conference on biometrics BTAS*.

Chor, B., Kushilevitz, E., Goldreich, O. & Sudan, M. (1998). Private information retrieval, *J. ACM* 45(6): 965–981.

Connie, T., Teoh, A., Goh, M. & Ngo, D. (2004). Palmhashing: a novel approach for dualfactor authentication, *Pattern analysis application* 7: 255–268.

Damgard, I. & Jurik, M. (2001). A generalisation, a simplification and some applications of paillier's probabilistic publickey system, *PKC'01*, Vol. 1992 of *Lecture Notes in Computer Science*, Springer, pp. 119–136.

Dasgupta, S. & Gupta, A. (1999). An elementary proof of the Johnson-Lindenstrauss Lemma. UTechnical Report TR-99-006, International Computer Science Institute, Berkeley, CA.

Daugman, J. (2004a). How iris recognition works, *Circuits and Systems for Video Technology, IEEE Transactions on* 14(1): 21–30.

Daugman, J. (2004b). Iris recognition and anti-spoofing countermeasures, *7-th International Biometrics conference*.

Dodis, Y., Katz, J., Reyzin, L. & Smith, A. (2006). Robust fuzzy extractors and authenticated key agreement from close secrets, *CRYPTO'06*, Vol. 4117 of *Lecture Notes in Computer Science*, Springer, pp. 232–250.

Dodis, Y., Reyzin, L. & Smith, A. (2004). How to generate strong keys from biometrics and other noisy data, *EUROCRYPT'04*, Vol. 3027 of *Lecture Notes in Computer Science*, Springer, pp. 523–540.

Domnesque, V. (2004). Carte d'identité électronique et conservation des données biométriques. Master thesis, Lille university.

El-Abed, M., Giot, R., Hemery, B. & Rosenberger, C. (2010). A study of users' acceptance and satisfaction of biometric systems, *IEEE International Carnahan Conference on Security Technology (ICCST'10)*, pp. 170–178.

Feng, J. & Jain, A. (2009). Fm model based fingerprint reconstruction from minutiae template, *International conference on Biometrics (ICB)*.

Galbally, J., Cappelli, R., Lumini, A., Maltoni, D. & Fiérrez-Aguilar, J. (2008). Fake fingertip generation from a minutiae template, *ICPR*, pp. 1–4.

GlobalPlatform (2006). *GlobalPlatform Card Specification Version 2.2*.

Goh, A. & Ngo, C. (2003). *Computation of Cryptographic Keys from Face Biometrics*, Vol. 2828 of *Lecture Notes in Computer Science*, Springer, Berlin.

Goldwasser, S. & Micali, S. (1982). Probabilistic encryption and how to play mental poker keeping secret all partial information, *Proceedings of the fourteenth annual ACM symposium on Theory of computing*, pp. 365–377.

Goldwasser, S. & Micali, S. (1984). Probabilistic encryption, *Journal of Computer and System sciences* 28(2): 270–299.

Hao, F., Anderson, R. & Daugman, J. (2006). Combining crypto with biometrics effectively, *IEEE Transactions on Computers* 55(9): 1081–1088.

Jain, A., Nandakumar, K. & Nagar, A. (2008). Biometric template security, *EURASIP J. Adv. Signal Process 2008*.

Jain, A., Ross, A. & Pankanti, S. (2006). Biometrics: A tool for information security, *IEEE Transactions on Information Forensics and Security* 1(2): 125–143.

Juels, A. & Sudan, M. (2001). A fuzzy vault scheme, *IEEE International Symposium on Information Theory*.

Juels, A. & Wattenberg, M. (1999). A fuzzy commitment scheme, *ACM conference on Computer and communication security*, pp. 28–36.

Kaski, S. (1998). Dimensionality reduction by random mapping: fast similarity computation for clustering, *Proc. of the International Joint Conference on Neural Networks*, Vol. 1, pp. 413–418.

Kevenaar, T., Schrijen, G., van der Veen, M., Akkemans, A. & Zuo, F. (2005). Face recognition with renewable and privacy preserving binary templates, *IEEE workshop on Automatic Identification Advanced Technologies*, pp. 21–26.

Kong, A., Cheung, K., Zhang, D., Kamel, M. & You, J. (2005). An analysis of biohashing and its variants, *Pattern Recognition* 39.

Lumini, A. & Nanni, L. (2006). Empirical tests on biohashing, *NeuroComputing* 69: 2390–2395.

MacWilliams, F. & Sloane, N. (1988). *The Theory of Error-correcting codes*, North-Holland.

Madlmayr, G., Dillinger, O., Langer, J. & Schaffer, C. (2007). The benefit of using sim application toolkit in the context of near field communication applications, *ICMB'07*.

Maltoni, D., Maio, D., Jain, A. & Prabhakar, S. (2009). *Handbook of Fingerprint Recognition*, Springer.

Monitor, N. (2002). 2002 nta monitor password survey.

Mordini, E. & Massari, A. (2008). Body, biometrics and identity, *Bioethics Journal* 22(9): 488–494.

O'Gorman, L. (2003). Comparing passwords, tokens, and biometrics for user authentication, *Proceedings of the IEEE* 91(12): 2021 – 2040.

Osadchy, M., Pinkas, B., Jarrous, A. & Moskovich, B. (2010). Scifi - a system for secure face identification, *IEEE Symposium on Security and Privacy*.

Paillier, P. (1999). Public-key cryptosystems based on composite degree residuosiy classes, *EUROCRYPT'99*, Vol. 1592 of *Lecture Notes in Computer Science*, Springer, pp. 223–238.

Pakanti, S., Prabhakar, S. & Jain, A. K. (2002). On the individuality of fingerprint, *IEEE Trans. Pattern Anal. Machine Intell.* 24(8): 1010–1025.

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

Schneier, B. (1999). Inside risks: the uses and abuses of biometrics, *Commun. ACM* 42: 136.

Simoens, K., Chang, C. & Preneel, B. (2009). Privacy weaknesses in biometric sketches, *30th IEEE Symposium on Security and Privacy*.

Solove, D. (2009). *Understanding privacy*, Harvard university press.

Teoh, A., Kuanb, Y. & Leea, S. (2008). Cancellable biometrics and annotations on biohash, *Pattern recognition* 41: 2034–2044.

Teoh, A. & Ngo, D. (2005). Cancellable biometrics featuring with tokenised random number, *Pattern Recognition Letters* 26: 1454–1460.

Teoh, A., Ngo, D. & Goh, A. (2004a). Biohashing: two factor authentication featuring fingerprint data and tokenised random number, *Pattern recognition* 40.

Teoh, A., Ngo, D. & Goh., A. (2004b). An integrated dual factor authenticator based on the face data and tokenised random number, *1st International conference on biometric authentication (ICBA), Hong Kong*.

Teoh, A., Ngo, D. & Goh, A. (2004c). Personalised cryptographic key generation based on facehashing, *Computers and Security Journal* 23(07): 606–614.

Tuyls, P., Akkemans, A., Kevenaar, T., Schrijen, G., Bazen, A. & Veldhuis, R. (2005). Practical biometric authentication with template protection, *Audio and Video based Personal Authentication*, pp. 436–446.

Warren & Brandeis (1890). The right to privacy. Harvard Law Review (IV).