Issues and Challenges in Storing Biometric Templates Securely

Mots clés

Biometrics,
Security,
Evaluation,
Dependability

Daniel SOCEK[1], Dubravko CULIBRK[1], Vladimir BOZOVIC[2]
CoreTex Systems[1], Florida Atlantic University[2]

1. Introduction

According to a recent study, over half a million identity theft cases were reported yearly just in the United States resulting in annual losses in the billions of dollars [6]. Identify theft is a fastest growing type of fraud in the US [6]. All such criminal endeavors were based on deceiving an underlying authentication system with a stolen identifier. While identity theft may occur due to lack of vigilance on the consumer's part, it can also occur as a result of direct tampering with the authentication system by a criminal.

An authentication system generally consists of two phases: (1) the *enrollment* phase, and (2) the *verification or identification* phase. During the enrollment phase, the user's authentication template is stored in a database within the system. Once enrolled, the user can be verified or identified in the verification or identification phase by presenting his or her authentication template, or identifier. Only if the presented authentication template and the stored one match exactly, or fall within a given similarity bound, the user is successfully authenticated.

Authentication methods based on user's biometric data have several advantages over other authentication methods. The greatest benefits of biometric-based authentication methods are the simplicity of use and a limited risk of losing, stealing, or forging one's biological identifier. However, the main weakness of biometrics-based methods is the inability to renew a stolen biological identifier. This is a particularly significant issue when identity theft is concerned. Moreover, biometric-based authentication with the same biometrics is likely to be used in multiple application systems. For example, a fingerprint-based authentication could be used to gain access to a bank account, but also to gain access to a computer system or a highly secure laboratory. If a biometric template is stolen from one authentication system by an adversary, it can be abused at present or future in multiple other authentication systems the user is or will be enrolled with, that use the same biological identifier. Therefore, unless the system uses highly tamper resistant local storage or a heavily secured and tamper resistant remote storage, storing the user's biometric template in its clear form during the enrollment phase should be avoided.

Note that securing biometric templates is important even if feature vectors contain condensed information about the biometric uniqueness of the user. For instance, in case of fingerprints, the system often stores only the discriminatory set of minutiae points. However, given such points an attacker can construct a fake fingerprint

L'ESSENTIEL

Dans cet article, nous analysons les questions et les défis de sécurité relatifs aux technologies émergentes de stockage de gabarits de signatures biométriques. Plusieurs procédés précédemment proposés sont analysés et leurs points faibles sont mis en évidence dans plusieurs scénarios concrets. Après avoir généralisé les insécurités observées, nous établissons un meilleur modèle de sécurité avec lequel des procédés analogues pourraient être développés. En outre, certaines limitations et défis d'applicabilité sont mis en évidence dans plusieurs procédés. Enfin, nous présentons les bases théoriques pour de nouvelles directions de recherche en conception d'algorithmes pour la sécurisation de gabarits de signatures biométriques, voie qui fournit une meilleure applicabilité que les procédés existants, en se basant sur une mesure plus normale de la similitude dans le domaine de la sécurité.

SYNOPSIS

In this work we discuss security issues and challenges in the emerging technologies for storing biometric templates securely. Several of the previously proposed schemes are analyzed and are shown to have weaknesses in several practical scenarios. We provide a generalization of the observed insecurities and establish a better security model under which similar schemes should be developed. Furthermore, some applicability limitations and challenges are pointed out for a number of schemes. Finally, we give the theoretical grounds for a new research direction in design of algorithms for securing biometric templates, a direction that provides for a better applicability than the existing schemes by relying on a more natural similarity measure in the secure domain.

that has the same discriminatory information. Methods for creating fake fingerprints such as SFINGE by Cappelli, Miao and Maltoni [3] or synthetic generation technique by Araque et al. [1] can be used for exactly that purpose. Uludag et al. [13] described many attacks on fingerprint-based identification systems using a fake fingerprint such as rubber or silicon finger, and similar. Similar arguments are also applicable to the other types of biometrics.

Unfortunately, standard cryptographic primitives are unsuitable for this purpose since the biometric identifiers are not exactly reproducible as a result of imperfections present in both biometric sensors and feature extraction algorithms. For instance, due to their strong avalanche effect, it is not possible to directly use standard cryptographic hash functions to secure biometric templates. In light of this, several schemes for storing biometric templates securely were proposed recently. In this work we analyze security aspects of several such schemes and identify their weaknesses. We also provide a more general security model under which such schemes should be considered. Furthermore, we present several issues in terms of applicability of these schemes, and provide a reserch direction that would potentially be able to overcome some of these limitations.

The rest of this paper is organized as follows. In Section 2 several preliminary concepts from related work are introduced and some new definitions are established. A review of relevant schemes for securing biometric templates is given in Section 3. In Sections 4 and 5 security and applicability issues and challenges regarding these schemes are presented and generalized, respectively. Finally, a promising reserch direction is presented in Section 6, while conclusions are given in Section 7.

2. Preliminaries and New Definitions

In this section we review the concept of secure sketch, fuzzy extractor and a general notion of robust hash function. We also establish a set of requirements that are generally of interest to biometrics-based security applications. Finally we classify biometrics feature vectors into several common types whose properties significantly affect the design of corresponding security schemes.

2.1. Secure Sketches, Fuzzy Extractors and Robust Hashing

In work by Dodis et al. [4] and [5], two types of schemes for securing biometric templates are defined:

- *Secure sketch* - This scheme essentially allows for the precise reconstruction of a noisy input. Given an input x, the scheme produces a public value $f(x)$, called secure sketch, from which no information about x can be deduced (i.e. f is a one-way function). The scheme can recover the original value of x solely from $f(x)$ and y if and only if y is similar to x according to some similarity measure, denoted with $y \sim x$.
- *Fuzzy extractor* - For a given input x this scheme produces a public value $f(x)$ and a secret value k. Function f is a one-way map so that no information about x can be deduced from $f(x)$. The scheme is able to recover k solely from y and $f(x)$ if and only if $y \sim x$. In practice, k is often used as a secret key for further cryptographic processing.

In [4], it was also shown that it is always possible to construct fuzzy extractors from secure sketches. Intuitively this means that secure sketches comply with a stronger condition (or requirement) than fuzzy extractors do. However, in a number of biometrics-based security applications, even fuzzy extractors comply to a stronger requirement than what is actually needed.

Sutcu et al. [12] considered the concept of *secure robust hashing* that is closer to the usual verification or identification requirement. Sutcu et al. defined a *secure robust hash function* as a one-way hash function that for similar inputs gives the same or similar outputs, but for inputs that are not similar gives completely different outputs. In general, one-wayness is a necessary feature for security of the templates, however, hash values need not be strictly smaller than the input.

In order to clarify these concepts and set the stage for later discussion, we define the following three possible biometric security application requirements:

R1. *Ability to recover the original biometric template solely from its public one-way transformed value and a similar template* — This requirement is achieved with a secure sketch scheme by definition. For instance, applications in which it is necessary to analyze the original template at a later time clearly have this requirement.

R2. *Ability to extract the same random secret solely from a template similar to the original one and the public one-way transformed value of the original template* — This requirement is clearly achieved with a fuzzy extractor, and therefore also with a secure sketch since the former can be easily obtained from the later. For example, a secret key could be used to encrypt some additional information in the public domain, and only a person with the similar biometric template could recover this information by extracting that key from its template.

R3. *Ability to measure threshold-based similarity between an original template and a newly presented one solely from a new template and the public one-way transformed value of the original template* — When concerned with pure verification or identification applications, ability to determine whether a new template matches the stored one is a sufficient requirement. In general, a match is declared when two templates are similar, or, in other words, with similarity measure greater than some threshold t (also referred to as the *similarity bound*). Note that the similarity function is not necessarily a metric. We define a threshold-based similarity measuring scheme S to be a scheme that for given one-way transformed value $f(x)$ and a template y determines whether the original template x and y are similar or not:

$$S(h(x),y)=\begin{cases} \text{similar}, & \text{if } s(x,y)>t; \\ \text{not similar}, & \text{if } s(x,y)\le t, \end{cases}$$

where denotes a similarity measure of x and y. Strictly speaking, this kind of scheme is slightly more limited than a scheme that can compute the actual value of $s(x,y)$ from $f(x)$ and y ; however, almost all biometrics security systems are based on a threshold similarity measure approach.

For basic authentication and verification it suffices to have a scheme that satisfies R3 requirement. Robust hash functions by the definition from [12] aim to satisfy this requirement. Also, it is not difficult to see that fuzzy extractors and secure sketches, which aim to satisfy R2 and R1 requirements respectively, are also necessarily threshold-based similarity measuring schemes. Note that it may be of interest to consider schemes that strictly comply to R3 requirement but not to R1 requirement.

2.2. Types of Feature Vectors

The design of a scheme for securing biometric templates is constrained with a type of biometric feature vector that is extracted from the sensory information. Properties of feature vectors representing biometric templates heavily depend on the type of biometric data involved, capability of a sensor, and the corresponding feature extraction algorithm. These properties include the types of errors introduced during data acquisition process, as well as the expected range of values and similarity thresholds.

Let $x=\{x_1,\dots,x_k\}$ and $y=\{y_1,\dots,y_l\}$ denote two biometric feature vectors obtained after the feature extraction phase. The following different types of biometrics feature vectors often appear in practice:

- *Type I* - Constant order and size; i.e., $k = l$ and x_i corresponds to y_i when compared for similarity. For example this type of feature vector is obtained using singular value decomposition (SVD)-based methods as feature extraction which is commonly used for face biometrics. Another common example of this type is IrisCode [7] feature vector template used in many iris recognition systems.
- *Type II* - Variable order and size. For instance, fingerprint and palm-print minutiae-based feature vectors belong to this type.

Schemes for securing biometric templates (features) are in general designed for a particular feature vector type.

3. Overview of Related Schemes

We give an outline of the most prominent schemes regarding security of biometric templates. The general idea is to construct the scheme that provides a certain level of security of stored biometric data and a mechanism for biometric-based authentication. To perform authentication a measure of *similarity* must be established. The most natural idea is to map biometric data into some metric space, which is the case with most of the proposed schemes.

3.1. Fuzzy Commitment Scheme (Juels and Wattenberg)

In [10], Juels and Wattenberg proposed a scheme, called fuzzy commitment, that is applicable to storing biometrics feature vectors of type I securely. This conceptually simple scheme is based on error correcting codes. Let F be a field, and C the set of vectors of some t-error correcting code. Let $x \in F^n$ denote a biometric feature vector. Assuming that all codewords lie in F^n, a codeword c is selected uniformly at random from C and difference $\epsilon = c - x$ is computed. Next, a suitable one-way function h is selected, and the pair $(\epsilon, h(c))$ is published, representing the output of fuzzy commitment scheme.

To reconstruct the original feature vector x, a similar vector y is required, where the measure of similarity is given by a certain metric. If the usual Hamming distance between $c' = \epsilon + v$ and c is less than t, the error correcting capability of the code C, then it is possible to reconstruct c and consequently x. Since the feature vectors are required to be from F^n, the scheme can be applied only to type I feature vectors, where constant size and order is assumed. Fuzzy commitment is a secure sketch scheme and as such can comply to all three aforementioned application requirements R1, R2 and R3. A scheme based on fuzzy vault principle was constructed and successfully applied for securing a particular type of iris templates, called IrisCode, as described in [7].

3.2. Fuzzy Vault Scheme (Juels and Sudan)

Juels and Sudan in [9] proposed a scheme, called *fuzzy vault*, that slightly extends the applicability of a scheme from [10] by allowing for order invariance of feature vector coordinates. This scheme substantially relies on Reed-Solomon error correcting codes, where the codewords are polynomials over a finite field F. Given a feature vector (set) $x \subset \mathsf{F}$ and a secret value k, a polynomial $p \in \mathsf{F}[X]$ is selected so that it encodes k in some way (e.g., has an embedding of k in its coefficients). Then an evaluation of the elements of x against p is computed and, along with these points, a number of random *chaff* points that do not lie on p is added to a public collection R.

To recover k, a set y similar to x must be presented. If $y \sim x$, then y contains many points that lie on p. Using error correction procedure, it is possible to reconstruct p exactly, and thereby p. If is not similar to x, it does not overlap substantially with x and thus it is not possible to reconstruct p using the error correction mechanism of Reed-Solomon code. By observing the public value R, it is infeasible to learn k due to the presence of many chaff points. This is also a secure sketch scheme thus conforming to all application requirements R1, R2 and R3. While fuzzy vault does allow for a variable order, it does require feature vector sizes to be of the fixed length, thus still not fully supporting biometrics feature vectors of type II.

3.3. Mixture of Gaussians Scheme (Sutcu, Sencar and Memon)

The scheme proposed by Sutcu et al. [12] is designed to serve as a robust hash function of securing biometric templates of type I. The user's feature vector x = $[x_1,..., x_n]$ is transformed using a robust hash function based on multiple Gaussians. It is assumed that each coordinate x_i has a δ_i -fuzziness associated with it, i.e., given a *similar* feature vector y = $[y_1,...,y_n]$ it must be that $y_i \in [x_i - \delta_i, x_i + \delta_i]$ for all $i = 1,...,n$. The user randomly assigns the value s_i to each coordinate x_i. Then, given r for each vector coordinate i a Gaussian fitting is performed using the following three points: $P_1 = (x_i - \delta_i, s_i)$, $P_2 = (x_i, s_i + r)$, and $P_3 = (x_i + \delta_i, s_i)$, where r is a randomly selected value between 0 and 1.

After this stage, a number of *chaff* Gaussian functions are generated and combined with the first one. For each coordinate this composite Gaussians, referred to as the *Mixture of Gaussians* (MoG), are stored publicly to serve as the robust one-way hash of the input template. The feature vector assigned values $s_1,...,s_n$ are concatenated and transformed using a standard secure cryptographic hash function. This transformed data is added to the public information.

In the reconstruction phase, a given feature vector is transformed using the stored MoG functions, and the resulting values concatenated and further transformed using a cryptographic hash function. Successful authentication occurs if this result match the concatenated hash value of the assigned values $s_1,...,s_n$ of the original template, which is the case when $y \sim x$. The MoG scheme complies only to R3 requirement.

3.4. PinSketch Scheme (Dodis, Ostrovsky, Reyzin and Smith)

Dodis et al. in **Erreur ! Source du renvoi introuvable**. proposed a scheme that allows for securing biometric feature vectors of type II. This scheme, called *PinSketch*, relies on t-error correcting (BCH) code C. In order to simplify description, let us assume H to be a parity check matrix of the code C over some finite field F. For a given feature vector x which belongs to F^n, the scheme computes output $syn(x) = Hx$, which is referred to as the syndrome of vector x.

In the reconstruction phase, $syn(y)$ is computed for a given vector y. Let $\delta = syn(x) - syn(y)$. It is easy to see that there exists at most one vector y such that $syn(v) = \delta$ and $weight(v) \leq t$. One of the nice features of binary BCH codes is possibility of computing $supp(v)$ given $syn(v)$ and vice versa, where $supp(v)$ represents the listing of positions where v has nonzero coordinate. Computing of $supp(v)$ for a given $syn(v)$ is the key step in the reconstruction phase. If a distance metric $d(x, y) \leq t$ then $supp(v) - x\Delta y$, and in that case the original set could be reconstructed by $x = y\Delta supp(x)$. PinSketch is a secure sketch scheme that supports biometrics feature vectors of type II.

4. Security Analysis

The authors of schemes summarized in Section [9] did not consider the complete security model under which such schemes operate in the real world. One of the most serious attacks that should be considered for biometrics-based authentication systems is the *multiple-use attack*, that is often feasible to launch.

Under the multiple-use attack, the adversary has public information obtained from multiple authentication systems regarding user U. The multiple-use attack is successful if it is possible to compromise the secret information about U (in whole or in part) from analyzing the public information about U from multiple systems. In what follows, we show that the fuzzy vault scheme [9] and the MoG scheme [12] are both weak against multiple-use attack.

4.1. Fuzzy Vault Scheme with Multiple-Use Attack

Suppose the same user is enrolled in k > l authentication systems which are all based on the same kind of biometric (e.g. fingerprint) and which all use the fuzzy vault scheme for securing biometric feature vectors. For simplification, let us assume that the user's biometric feature vector in all systems was $x = \{x_1,..., x_t\}$ since almost the same arguments apply when these vectors are *similar*. Recall that the public information that is stored in system i is a collection $R^{(i)}$ that contains t points $(x_1, p^{(i)}(x_1)),\ldots,(x_t, p^{(i)}(t))$, and $m^{(i)}$ chaff points $(r_1^{(i)}, s_1^{(i)}),\ldots,(r_{m^{(i)}}^{(i)}, s_{m^{(i)}}^{(i)})$. According to the fuzzy vault

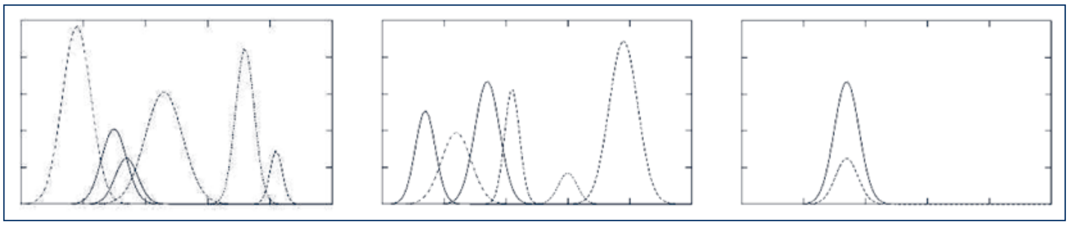


Figure 1. Illustration of the multiple-use attack on the MoG scheme: left figure represents Gaussians of one of the user's feature coordinates in the first system, middle figure shows corresponding Gaussians in the second system, and right figure illustrates removing of chaff by observing the intersection of x-axis values of Gaussian peaks.

specification chaff points are selected uniformly at random from U − *x*, where U de notes the universe of feature vector coordinates. If $R_x^{(i)}$ denotes the restriction of $R^{(i)}$ to the x-axis, then

$$\lim_{k\to\infty}(R_x^{(1)} \cap R_x^{(2)} \cap \ldots \cap R_x^{(k)}) = x$$

unless chaff points always entirely cover the remaining universe or some fixed parts of it. Moreover, if we take a simple case when $r = |R_x^{(i)}| - t = |U|$ for *i* – 1,2, then

$$Prob(R_x^{(1)} \cap R_x^{(2)} = x) = \frac{\binom{|U| - t - r}{r}}{\binom{|U| - t}{r} + 1} \approx 1,$$

where $|U|$ denotes the cardinality of set U. In other words, if the number of randomly selected chaff points is much smaller than the size of the universe U, the intersection of chaff points of the same person taken from two authentication systems will almost certainly be empty.

In [9] it is shown that the number of different polynomials that agree on *t* is small if the size of collection *R* is small. Thus, in order to ensure security from that point of view, the authors recommend taking a large number of chaff points. Yet, the authors do not require to always cover the entire remaining universe U − *x* with chaff. Indeed, this is probably infeasible when dealing with larger universes. However, to avoid the multiple-use attack as described here, the entire remaining universe or fixed part of it must be covered by chaff. That is, $R_x^{(i)} = U'$ for all *i* where U' is a subset of U that provides a large number of polynomials that agree on *t* points and also a computationally infeasible search space.

4.2. MoG Scheme with Multiple-Use Attack

The multiple-use attack on the MoG scheme works in a way analogous to that of an attack on the fuzzy vault scheme. Let us suppose that the user is enrolled in $k > 1$ authentication systems that utilize the *Mixture of Gaussians* scheme to secure the user's biometric feature vectors. Recall that MoG works with feature vectors of type I so that all vectors are of fixed length, say *t*. For each coordinate *j* in *x*, $1 \leq j \leq i$, a set of random Gaussians is generated with only one Gaussian whose peek is attained at x_j. In the peaks of all Gaussians of all *k* MoG's corresponding to x_j, only the peaks corresponding to x_j will significantly overlap. Thus, by looking at the most frequent *x* coordinates of peaks, the attacker obtains with high probability the *j*-th biometric feature vector coordinate. This process is illustrated in figure 1.

In [12] the authors do not specify the number of chaff Gaussians that need to be generated. Due to the multiple-use attack, it must be the case that Gaussians are constructed to cover the entire universe of possible values on the x-axis.

4.3. Generalization and Improved Security Model

It is not difficult to see that the multiple-use attack is in general very effective for schemes that are based on the principle of *chaffing and winnowing* [11]. As long as the security of a scheme relies on a presence of randomly generated chaff that disguises the actual secret data but does not cover the entire universe, the intersection of such information gathered from multiple systems would likely reveal the secret, or at least significantly narrow down the search space. To resist this kind of attack, one possible approach would be to cover the entire universe with chaff points, which is only feasible for relatively small universes.

The security of fuzzy commitment and PinSketch schemes rely on different principles, and multiple-use attacks as described here are not applicable. In [2], Boyen considered the issues of multiple uses of the same fuzzy secret in a general fuzzy extractor scheme. Boyen pointed out that in the security model of fuzzy extractors such issue must be addressed and related security risks accounted for. In that respect he designed an improved security model for practical fuzzy extractors.

5. Applicability Issues

From the mathematical point view, the most suitable method of measuring *similarity* between two sets by their symmetric set difference. However, this quite reasonable mathematical choice is often a limitation for practical use. Let us try to illustrate this problem in the case where it is needed to measure closeness between two sets *A* and *B* that represent biometric (fingerprint) personal data, of not necessarily different persons. Reconstruction of *A*, using *similar* set *B* will be successful if and only if $|A\Delta B|\leq t$, where *t* is a given parameter that controls the closeness between sets. It seems that error correcting codes are a suitable choice for reconstructing *A* from a noisy input *B*. Here, *t* is the error correcting bound of the chosen code.

We argue that the use of error correcting codes and consequently the Hamming distance as a measure of similarity between type II feature vectors is not an adequate choice. For instance, in the PinSketch scheme [4], templates are represented as characteristic vectors with respect to universe U. Therefore, the symmetric difference is simply related to the Hamming distance between characteristic vectors. In a typical application of PinSketch, such as fingerprint identification, the scheme has a substantial applicability issue. The number of minutiae, according to many statistical analyses of fingerprints lies with high probability in the interval between 30 and 100. Thus, choice of the error correcting bound that is used in this scheme seems to be its main shortcoming. Clearly, *t* must be less than 30 and it needs to be set in a way that security is not compromised. Thus, the choice of is primarily related to the size of universe. Let us construct one example.

Suppose $|A|=80$, $|B|=75$, and $|A\cap B|=60$. Since $|A\Delta B|\geq 30$ the scheme will not identify *A* and *B* as biometric templates that belong to the same person. However, in the standard forensic fingerprint, the so-called *twelve points matching rule* indicates that twelve common points confirm that with high probability the two templates belong to the same person.

The authors of fuzzy vault [9] indicate that the scheme is applicable to feature vectors with fixed size and variable ordering which limits the practical use of the scheme to type I vectors. Even if it is possible to extend the fuzzy vault scheme to work with the type II feature vectors, the scheme would face the same aforementioned applicability issues since it is based on error correction approach.

The error correcting code principle can be used to design schemes for type I feature vectors. Fuzzy commitment scheme is an example of a scheme for which no weaknesses in terms of security and applicability are known. The schemes based on the ideas from fuzzy commitment, such as the scheme from [7], are to date secure and applicable in their domains. Further research should be conducted in order to design a secure scheme applicable to type II biometric feature templates, such as fingerprint minutiae, which are the most common templates in the real world.

6. A Promising Research Direction

A close examination of the nature of biometric data of type II (such as a fingerprint templates) reveals certain theoretical properties. Symmetric-key encryption of the entire template requires a secret key which cannot be stored in a non-tamper resistant environment. Standard cryptographic one-way hash functions as well as public-key encryption in the form of a classical identity-based encryption (IBE) cannot be applied due to the fuzzyness present in the acquisition of biometric data. In most biometric applications the size of the universe to which individual elements from the template belong to is relatively small, and thus, it is not possible to secure the template by applying a one-way transformation element-wise; it would be feasible to exhaust the universe and reveal the secret by applying a membership test.

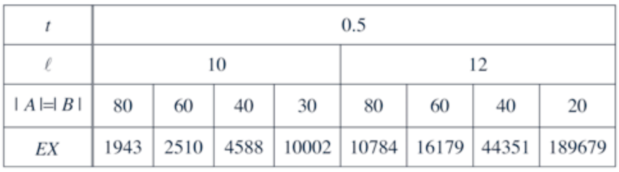
In order to make exhaustive search infeasible, an idea to take a one way transformation on subsets of elements of the template comes naturally. The chosen subset size ℓ should be large enough to provide security, while at the same time providing robustness to the fuzzy input. Clearly, ℓ should be chosen so that $\binom{|\mathsf{U}|}{\ell}$ is over a computational bound, such as, say, 2^{80}. If the size of universe is approximately $|\mathsf{U}|=10{,}000$, then ℓ could be 8, since that gives a key space of $\approx 2^{90}$ considering the exaustive search throughout the set of all ℓ subsets of U. Of course, it is not possible to take ℓ to be the entire size of the stored template because of its fuzziness in the extraction phase. The particularly important case in actual forensic practice is when $\ell = 12$ as we emphasized earlier. We have already argued why the size of intersection between probe and original template is the most appropriate similarity measure for the practical use. Next, a theoretical model from [14] that uses the set intersection as a natural simmilarity measure based on the aforementiond observations is sumarized.

Let *A* be the original and *B* the probe template. Let us adjoin to *A* a secret *s* and suppose that it has been constructed an extractor *R* such that *R*(*C*) = *s*, whenever $C\subset$ A and $|C|=\ell$. We assume that knowing the secret *s* uniquely determine the original template *A*. We store *h*(*s*), where *h* is publicly known collision resistant one-way transformation. Now, we go through all ℓ, subsets of *B* and apply extractor *R* on each. Every time, we get *s'* as a result, we compute *h*(*s'*) and compare it to the stored *h*(*s*). If *h*(s) = *h*(s') then it follows that *s* = *s'* and therefore correspoding subset is in the intersection of *A* and *B*. It could be the case that ℓ, common elements is enough for confirmation that *A* and *B* come from the same person with high probability, like in case, ℓ = 12. In general, *s* uniquely determines *A*, so by reconstructing it, it is possible to obtain $|A\cap B|$ exactly. It should be emphasized that search space is $\binom{|B|}{\ell}$. Therefore, in the case when $|B|$ is large, this search would be inefficient. However, we can always assume a natural request $|A \cap B| \geq \lceil t|A| \rceil$, and $t \geq 0.5$ for matching the probe template A as B. Since the random variable X that describes the probability of getting ℓ subsets from $A \cap B$ has a *negative hypergeometric distribution*, then mathematical expectation is given by:

$$EX = \frac{c+1}{b+1},$$

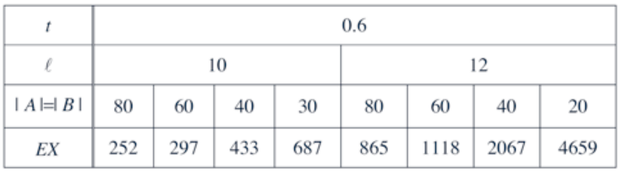
where $b = \binom{\lceil t|A| \rceil}{\ell}$ and $c = \binom{|B|}{\ell}$.

Tables taken from [14] is given to show some of these relationships and demonstrate the feasibility of the proposed research direction.



t	0.5							
ℓ	10				12			
$\lvert A\rvert=\lvert B\rvert$	80	60	40	30	80	60	40	20
EX	1943	2510	4588	10002	10784	16179	44351	189679

Table 1. The expected number of attempts needed to find an ℓ-subset of $|A \cap B|$ for various sizes of A and B when t = 0.5.



t	0.6							
ℓ	10				12			
$\lvert A\rvert=\lvert B\rvert$	80	60	40	30	80	60	40	20
EX	252	297	433	687	865	1118	2067	4659

Table 2. The expected number of attempts needed to find an ℓ-subset of $|A \cap B|$ for various sizes of A and B when t = 0.6.

7. Conclusions

The identity theft problem is significantly exacerbated for the biometric-based authentication systems, since the biometric data and the corresponding feature vectors are non-renewable. Once a biometric data has been compromised the identity of a user is in jeopardy not only in respect to the system from which the data has been compromised, but likely also in respect to all past, present and future systems that rely on the authentication or identification based on the same kind of biometric (e.g. fingerprint). If the system is highly tamper-resistant, the risk of identity theft is considerably reduced. However, highly tamper-resistant devices are expensive to manufacture and algorithmic solutions are needed to facilitate a low-cost production.

A number of algorithms for securing stored biometric templates were proposed recently. In this work, several principal schemes are introduced and analyzed in terms of their security and applicability aspects. Three crucial application requirements are identified, and it is argued that these requirements allow for a clearer applicability categorization of such schemes. The multiple-use attack is described and several prominent algorithms based on the principle of chaffing and winnowing were shown to have serious weaknesses against this type of attack, unless certain strict chaffing conditions are met. The multiple-use attack is often feasible and it should be considered in the general security model for the schemes for securing biometric templates. Applicability issues and challenges are discussed in respect to several schemes, and it is concluded that the schemes based on error correction are inadequate for securing the feature vectors that have variable size and order. Finally, a basis for a promising new research direction based on a natural similarity measure between two type II biometric templates is established.

References

[1] J. ARAQUE, M. BAENA, B. CHALELA, D. NAVARRO, P. VIZCAYA, *"Synthesis of Fingerprint Images"*, In 16th International Conference on Pattern Recognition, volume 2, pp. 422-425, 2002.

[2] X. BOYEN, *"Reusable Cryptographic Fuzzy Extractors"*, In ACM Conference on Computer and Communications Security (CCS 2004), pp. 82-91, New York: ACM Press, 2004.

[3] R. CAPPELLI, D. MAIO, D. MALTONI, "*Synthetic Fingerprint-Database Generation*", In 16th International Conference on Pattern Recognition, volume 3, pp. 744-747, 2002.

[4] Y. DODIS, R. OSTROVSKY, L. REYZIN, A. SMITH, *"Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data"*, April 28 2006, http://www.citebase.org/abstract?id=oai:arXiv.org:cs/0602007.

[5] Y. DODIS, L. REYZIN, A. SMITH, "*Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data*", In EUROCRYPT 2004, Interlaken, Switzerland, pp. 523-540, 2004.

[6] A. J. ELBIRT, *"Who Are You? How to Protect against Identity Theft"*, IEEE Technology and Society Magazine, Summer 2005.

[7] F. HAO, R. ANDERSON, J. DAUGMAN, "*Combining Crypto with Biometrics Effectively*", IEEE Transactions on Computers, 55(9): 1081-1088, 2006.

[8] A. JUELS, M. SUDAN, "*A Fuzzy Vault Scheme,*" In IEEE International Symposium on Information Theory (ISIT 2002), Lausanne, Switzerland, 2002.

[9] A. JUELS, M. SUDAN, *"A Fuzzy Vault Scheme. Designs, Codes and Cryptography"*, 38(2): 237-257, 2006.

[10] A. JUELS, M. WATTENBERG, *"A Fuzzy Commitment Scheme"*, In ACM Conference on Computer and Communications Security, pp. 28-36, 1999.

[11] R. L. RIVEST, "*Chaffing and Winnowing: Confidentiality without Encryption*", April 24 1998, http://theory.lcs.mit.edu/~rivest/chaffing.txt.

[12] Y. SUTCU, H. T. SENCAR, N. MEMON, "*A Secure Biometric Authentication Scheme Based on Robust Hashing,*"

In ACM Multimedia and Security Workshop (MM&Sec'05), August 1-2 2005, New York, pp. 111-116.

[13] U. ULUDAG, S. PANKANTI, S. PRABHAKAR, A. JAIN, *"Biometric Cryptosystems: Issues and Challenges"*, IEEE Special Issue on Enabling Security Technologies for Digital Rights Management, 92(6): 948-960, 2004.

[14] D. SOCEK, D. CULIBRK, V. BOZOVIC, "*Practical Secure Biometrics Using Set Intersection as a Similarity Measure*", In International Conference on Security and Cryptography (SECRYPT 2007), July 28-31 2007, Barcelona, Spain, pp. 25-32.

Les auteurs

Daniel Socek (daniel.socek@coretexsys.com), Ph.D., is a Director of Research at CoreTex Systems, LLC (www.coretexsys.com). He received his Ph.D. in computer science from Florida Atlantic University in 2006. His research include biometrics, cryptography, multimedia security, digital image and video compression, coding and analysis, information aecurity and secure communications, and secure system design.

Dubravko Culibrk (dubravko.culibrk@coretexsys.com), Ph.D., is a Director of Technology Development at CoreTex Systems, LLC (www.coretexsys.com). He received his Ph.D. in computer engineering from Florida Atlantic University in 2006. His fields of specialization and interest are image and video processing, theoretical and applied cryptology, multimedia and multimedia security, biometrics, embedded system design and tools, evolutionary computing and neural networks, data mining, FPGA-based system design.

Vladimir Bozovic (vbozovic@fau.edu), M.Sc., is currently a Ph.D. student in Mathematics in the Department of Mathematical Sciences at Florida Atlantic University (Boca Raton, Florida, USA). His reseach interests include combinatorics, group theory, cryptology and applied mathematics.