Identity Leakage Mitigation on Asymmetric Secure Sketch

Chengfang Fang[1] Qiming Li[2] Ee-Chien Chang[1]
[1] National University of Singapore
{c.fang, changec}@comp.nus.edu.sg

[2] Institute for Infocomm Research, Singapore
Qiming.Li@ieee.org

Abstract

We consider secure sketch construction in an asymmetric setting, that is, multiple samples are acquired during enrollment, but only a single sample is obtained during verification. Known protection methods apply secure sketch constructions on the average of the samples, while publishing the auxiliary information extracted from the set of samples, such as variances or weights of the features, in clear. Since the auxiliary information is revealed, an adversary can potentially use it to determine the relationship among multiple sketches, and gather information on the identity of the sketches. In this paper, we give a formal formulation of secure sketch under the asymmetric setting, and propose two schemes that mix the identity-dependent auxiliary information within the sketch. Our analysis shows that while our schemes maintain similar bounds of information loss compared to schemes that reveal the auxiliary information, they offer better privacy protection by limiting the linkages among sketches.

1. Introduction

Protection techniques for biometric templates, such as fuzzy commitment [6], fuzzy vault [5], and secure sketch [3], publish small pieces of data to aid reconstruction of the biometric secrets under inevitable noises. One important goal of these techniques aims to minimize the information loss of the published data, which are also known as secure sketches. Essentially, during enrollment, after the biometric data X is obtained, its sketch S is constructed, typically based on some error-correcting code. From another X' obtained during verification, if X and X' are sufficiently close, the original X can be reconstructed from X' and S. With the sketch released as public data and available during verification, the exact X can be reconstructed whenever a close enough X' is presented. Thus, X can be used as a consistent secret in cryptographic operations. Since the sketch is to be revealed, it must not leak important information on X and the identity of the enrollee. Known constructions of secure sketches, such as for fingerprints [1, 2] or faces [13], generally handle biometric data that are modeled as a set of feature points (either ordered or unordered) under two types of noises. The first type perturbs each biometric feature point by a small amount; and the second type adds and removes some feature points. The differences under the first type of noise can be measured using Euclidean distance, whereas the second type can be measured using set difference.

The constructions of sketches are typically applied in a symmetric setting, that is, only one sample is acquired during both enrollment and verification. To improve the performance in terms of relative operating characteristic (ROC), many applications [4, 12, 7] adopt an asymmetric setting. During enrollment phase, multiple samples are obtained, whereby an average sample and auxiliary information such as variances or weights of features are derived. During verification, only one sample is acquired. The derived auxiliary information can be helpful in improving ROC. For example, it could indicate that a particular feature point is relatively inconsistent and should not be considered, and thus reducing the false reject rate. Note that the auxiliary information is identity-dependent in the sense that different identity would have different auxiliary information. Li et al. [10] observed that by using the auxiliary information in the asymmetric setting, the "key strength" could be enhanced due to the improvement of ROC, but there could be higher leakage on privacy.

Current known works, for example, the schemes given by Li et al. [10] and by Kelkboom [7], store the auxiliary information in clear. Li et al. [10] employ a scheme that carefully groups the feature points to minimize the differences of variance among the groups. The derived grouping is treated as auxiliary information and is published in clear. The scheme proposed by Kelkboom et al. [7] computes the means and variances of the features from the multiple enrolled face images, and selects the k features with least variances. The selection indices are also published in clear. The revealed auxiliary information could potentially leak important identity information as an adversary could distinguish whether a few sketches are of from the same identity by comparing the auxiliary information. Such leakage is similar to the sketch distinguishability in the typical symmetric setting[11]. Therefore, it is desired to have a sketch construction that can protect the auxiliary information as well.

In this paper, we construct two schemes where the auxiliary information is protected by "mixing" it into the sketch. We extend the notation of entropy loss [3] and give a formulation on information loss for secure sketch under asymmetric setting (Section 2.2). We give two sketch constructions for asymmetric setting under the two types of noise. The first construction handles the Euclidean noise with auxiliary information modeled by level of the noise (Section 3). The second construction handles set difference with a weight vector indicating the consistency of the biometric features (Section 4).

We analyze the proposed schemes under two security notions, namely, the average min-entropy loss of the identity information, and the linkages of sketches. Our analysis shows that, and yet our schemes have similar bound on information loss compared to the straightforward methods, they offer better privacy protection (Section 3.2, Section 4.2).

2. Preliminaries and Formulation

2.1. Symmetric Sketch

Let $\mathcal{M}$ be the set of biometric data with a closeness relation $\mathbf{D}$ defined over $\mathcal{M} \times \mathcal{M}$, two biometric samples X and Y are closed to each other if (X, Y) is in $\mathbf{D}$. The secure sketch for symmetric setting is define as follow:

Secure sketch [3]. A secure sketch scheme is a tuple $(\mathcal{M}, \mathbf{D}, Enc, Dec)$, where $Enc : \mathcal{M} \rightarrow \{0,1\}^*$ is an encoder and $Dec : \mathcal{M} \times \{0,1\}^* \rightarrow \mathcal{M}$ is a decoder such that for all $(X, X') \in \mathbf{D}, Dec(X', Enc(X)) = X$. The output $S = Enc(X)$ of the encoder, is called the sketch of X.

Since the exact X can be reconstructed with a close enough X', X can be used as a consistent secret in known cryptographic techniques, for example, a key in a encryption scheme.

2.2. Asymmetric Sketch

We now extend the definition of secure sketch to the asymmetric setting. Let B and X be the information obtained during registration and verification respectively, and let $\mathcal{V}$ be the set of all B's, and $\mathcal{M}$ be the set of all X's. Note that in the asymmetric setting, $\mathcal{V}$ is not the same as $\mathcal{M}$. Let $\mathbf{C} \subset \mathcal{V} \times \mathcal{M}$ be a relation where, $(B, X) \in \mathbf{C}$ if the biometric data X obtained during verification should be considered to be from the same enrollee who provides B. We define asymmetric secure sketch as follow:

Asymmetric secure sketch. An asymmetric secure sketch scheme is a tuple $(\mathcal{M}, \mathcal{V}, \mathbf{C}, P, Enc, Dec)$, where $Enc : \mathcal{V} \rightarrow \{0,1\}^*$ is an encoder and $Dec : \mathcal{M} \times \{0,1\}^* \rightarrow \mathcal{M}$ is a decoder such that for all $(B, X) \in \mathbf{C}, Dec(X, Enc(B)) = P(B)$, where P is a projection from $\mathcal{V}$ to $\mathcal{M}$.

Similarly, $P(B)$ can be used as a consistent secret as it can be reconstructed exactly from S and X if B and X are provided by the same person.

2.3. Entropy Loss from Sketches

The security of a sketch scheme relies on how much information of the biometric data is leaked from the sketch. We follow Dodis et al.[3] notion which quantifies the lost information based on *average min-entropy*. The min entropy $\mathbf{H}_\infty(A)$ of a variable A is defined as $-\log(max_a(Pr[A = a]))$, and the average min-entropy of A given B is defined as: $\widetilde{\mathbf{H}}_\infty(A|B) = -\log(\mathbb{E}_{b \leftarrow B}[2^{-\mathbf{H}_\infty(A|B=b)}])$. Under this notion, the remaining entropy of X given the sketch $S = Enc(X)$ is expressed by $\widetilde{\mathbf{H}}_\infty(X|S)$, and the entropy loss is measured by $\mathbf{H}_\infty(X) - \widetilde{\mathbf{H}}_\infty(X|S)$.

It can be shown [3] that:

$$\mathbf{H}_\infty(X) - \widetilde{\mathbf{H}}_\infty(X|S) \leq |S| - \mathbf{H}_\infty(R) \qquad (1)$$

where R is the invested randomness in sketch construction that can be recovered from X and the sketch S. This is a useful inequality in bounding the (worst case) entropy loss w.r.t any distribution on X.

In the asymmetric setting, we care about the leakage on identity dependent information, and thus consider B $\mathbf{H}_\infty(B) - \widetilde{\mathbf{H}}_\infty(B|S)$ as the security measurement. By the same argument in (1), we can give a similar bound on the entropy loss in asymmetric setting:

$$\mathbf{H}_\infty(B) - \widetilde{\mathbf{H}}_\infty(B|S) \leq |S| - \mathbf{H}_\infty(R) \qquad (2)$$

2.4. Privacy of Schemes

The quantity of remaining entropy is not the only security concern. Even if the remaining entropy is large, important information on identity might have been leaked by a sketch. There are also other concerns, for example, cross matching [8] and correlation attack [9]. Simoens et al.[11] give a security model on the sketch distinguishability. They examine the probability that an attacker can determine whether two documents were encrypted using the same biometric. We give a similar model in the asymmetric setting: the adversary has two sketches and he wants to determine whether they belong to the same identity. Formally, for a sketch scheme, consider the following game between the challenger $\mathcal{C}$ and an attacker $\mathcal{A}$:

1. $\mathcal{C}$ randomly picks a biometrics B_0 from the space $\mathcal{V}$ and a bit a from $\{0,1\}$.

2. (Same identity) If $a = 0$, $\mathcal{C}$ chooses B_1 which is B_0 perturbed by random noise;
(Different identity) If $a = 1$, $\mathcal{C}$ randomly selects a B_1 from $\mathcal{V}$.

3. $\mathcal{C}$ computes the asymmetric sketches on B_0 and B_1 and sends $Enc(B_0), Enc(B_1)$ to $\mathcal{A}$.

4. $\mathcal{A}$ on receiving $Enc(B_0), Enc(B_1)$, outputs a bit a'.

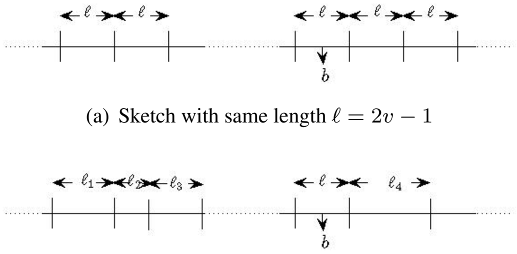
5. $\mathcal{A}$ wins if $a' = a$.

The random noise model in second step of the above game is determined by the enrollment process, which models the noise between two enrollments of the same identity. The value $|Pr[a' = a] - \frac{1}{2}|$ reflects effectiveness of the attacker in distinguishing the identity given two sketches. In Section 3, we will show that compared to the straightforward scheme, our construction is able to hide the auxiliary information while the entropy loss remain unchanged. Note that when $a = 1$, there is a chance that the randomly selected B_1 is close to B_0. Such probability corresponds to the false accept rate and should be small.

3. Asymmetric Sketch under Euclidean Distance

In this section, we give a construction for Euclidean distance. Let us illustrate our idea with a simple case where a biometric sample is represented as an integer in $[0, n)$. During enrollment, multiple samples are acquired and the information B are derived and represented by two integers (b, v): $b \in [0, n)$ is the mean of the samples and $v \in [1, q]$ is a threshold based on the variance of the samples. During the verification, a sample X in $[0, n)$ is acquired. X is consider to be from the same identity who enrolls B if $|X - b| < v$, i.e. the close relation $\mathbf{C} = \{(X, B) \mid |X - b| < v\}$. The choice on the value of v will determine the false reject rate, where a larger v gives lower false reject rate but lowers the key strength.

Let us first describe a straightforward sketch scheme SS_1 as follow: (1) Enc_1 on input $B = (b, v)$ outputs two integers $c = (b \bmod (2v - 1))$ and v; (2) Dec_1 on X outputs the integer in the set $\{a | a \equiv c(mod 2v - 1)\}$ that is closest to X; and (3) the function $P(B)$ projects $B = (b, v)$ to b. It is easy to verify that for all $(B, X) \in \mathbf{C}$, $Dec_1(X, Enc_1(B)) = P(B)$. Essentially, scheme SS_1 divides $[0, n)$ into intervals of length $\ell = 2v - 1$ where b is at the center of one of the intervals as shown in Figure 1(a).



(a) Sketch with same length $\ell = 2v - 1$

(b) Sketch with different lengths

Figure 1. Two sketch schemes over a simple 1D case.

One weakness of scheme SS_1 is that the auxiliary information v is revealed in clear. We give a simple sketch scheme SS_2 which protects the auxiliary information. The main idea of the construction is to partition the domain $[0, n)$ into intervals of different lengths, with the interval $(b - v, b + v)$ among one of them, as shown in Figure 1(b). Given $B = (b, v)$, the encoder Enc_2 constructs the sketch S in the following steps:

1. Let G be a set of two integers $\{b - v, b + v\}$.

2. Randomly generates an integer r from $[1, q]$ and inserts the value $(\min(G) - (2r - 1))$ into G, repeats Step 2 until $\min(G) \leq 0$.

3. Randomly generates an integer r from $[1, q]$ and inserts the value $(\max(G) + (2r - 1))$ into G, repeats Step 3 until $\max(G) \geq n$.

4. Let k be the number of elements in G, sorts G in ascending order and let the sorted list be $\langle g_1, \ldots, g_k \rangle$, note that g_1 is negative and $g_k > n$.

5. Let $\ell_i = g_{i+1} - g_i$ for i in $[1, k)$, returns the sequence $\langle g_1, \ell_1, \ell_2, \ldots, \ell_{k-1} \rangle$ as the sketch S.

Intuitively, the ℓ_i's are the lengths of the intervals as shown in Figure 1(b). The Dec_2 algorithm on X and S, reconstructs the set $G = \langle g_1, g_2, \ldots, g_k \rangle$, and finds the first i such that $g_i > X$ (note that $i > 1$ since $g_1 < 0$), and returns $(g_{i-1} + g_i)/2$. The projection $P(B)$ outputs b as in scheme SS_1.

The correctness (i.e. $Dec(X, Enc(B)) = P(B)$) of the scheme can be easily verified: if X and B are from the same identity, then we have $b - v < X < b + v$, and thus $g_i = b + v$, $g_{i-1} = b - v$ and $(g_{i-1} + g_i)/2 = b$. This leads to $Dec(X, S) = P(B)$ as required.

3.1. Analysis of Entropy Loss

The following analysis gives a bound on the entropy loss (Section 2.3) and gives the comparison on privacy (Section 2.4) of scheme SS_1 and SS_2. Note that such bound holds for any distribution on X.

Lemma 1 *The entropy loss of the sketch produced by SS_2 is at most $1 + 2\log q$.*

Proof Since Step 2 of Enc_2 completes when $\min(G) \leq 0$, g_1 is in $(-2q+1, 0]$ and thus $\mathbf{H}_\infty(g_1) \leq \log 2q$. The $\ell_i = g_{i+1} - g_i$ are odd numbers in $[1, 2q-1]$, thus $\mathbf{H}_\infty(\ell_i) \leq \log q$. When Dec reconstructs G from X and S, the randomness (i.e. the r added in each iteration of Step 2 and 3 in Enc_2) used in generating the $k-2$ intervals can be recovered. By equation (2), the entropy loss is at most $\log 2q + (k-1)\log q - (k-2)\log q = 1 + 2\log q$. □

For the scheme SS_1, since the number of bits required to describe the sketch is $|v| + |c|$, and there is no randomness involved, the entropy loss is bounded by $|v| + |b \bmod 2v|$. Note that v is in range $[1, q]$, $\mathbf{H}_\infty(v) \leq \log q$, and $\mathbf{H}_\infty(c) = \mathbf{H}_\infty(b \bmod 2v - 1) \leq \log 2q$, thus, the entropy loss of scheme SS_1 is bounded by $1 + 2\log q$.

3.2. Analysis of Sketch Distinguishability

While the entropy loss of schemes SS_1 and SS_2 are bounded by a same value, scheme SS_1 reveals the auxiliary information in clear, whereas scheme SS_2 protects the auxiliary information by "mixing" it with the biometric secret and giving different sketches for different enrollments. In this section we will analyze the impact of such difference.

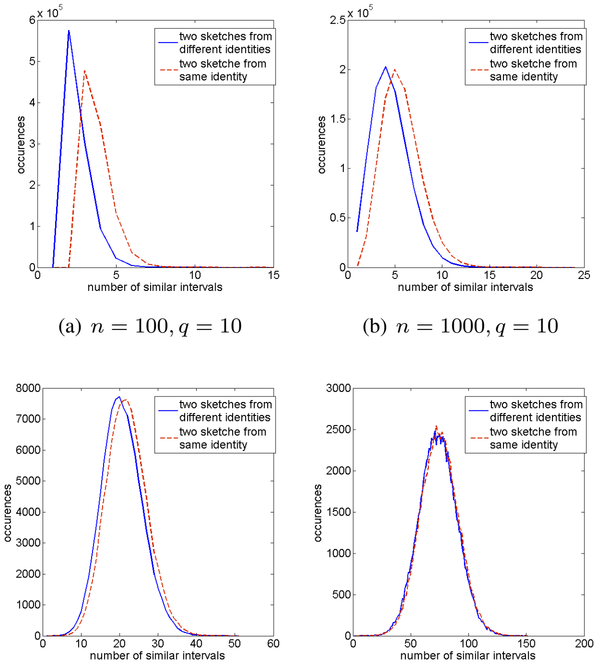
For the discussion to be meaningful, let us assume that the auxiliary information is identity dependent and there are two thresholds t and ϵ such that for two biometric data $B = (b, v)$ and $B' = (b', v')$ obtained from the same identity, we will have $Pr[|v - v'| \geq t] < \epsilon$, and for two biometric data $B = (b, v)$ and $\widetilde{B} = (\widetilde{b}, \widetilde{v})$ obtained from two different identities, $Pr[|v - \widetilde{v}| < t] < \epsilon$.

For scheme SS_1, let $\mathcal{C}_1$ and $\mathcal{A}_1$ be the challenger and attacker described in Section 2.4, let $B_0 = (b_0, v_0)$, $B_1 = (b_1, v_1)$ be the two sketches output by $\mathcal{C}_1$. There is an effective algorithm $\mathcal{A}_1$ in guessing a': it outputs 0 if and only if $|v_0 - v_1| < t$. In this case, the probability $Pr[a' = a \mid a = 0] \geq 1 - \epsilon$ and $Pr[a' = a \mid a = 1] \geq 1 - \frac{2t-1}{q} - \epsilon$.

For scheme SS_2, one strategy of $\mathcal{A}_2$ is to count the number of "similar intervals": two overlapping intervals are similar if the ratio between the length of their intersection and the length of their union is greater than the threshold $\frac{t+q}{q}$. $\mathcal{A}_2$ outputs 0 if the number of "similar intervals" between $Enc(B_0)$ and $Enc(B_1)$ is larger than a threshold it learnt, and output 1 otherwise.

The intuition of the above strategy is that, when $a = 0$, the count is expected to be larger. However, when n is large and q is small, the domain $[0, n)$ is divided into many intervals and this will reduce the effectiveness of the strategy of $\mathcal{A}_2$. Thus, the attack will depends not only on the parameter q, t but also on n.

Figure 2 shows how the parameters will affect the privacy protection. We implement the scheme SS_2 and for different values of n and q with $t = 1$ and $\epsilon = 0.001$, we randomly generated 10^6 biometrics B_0, construct $Enc(B_0)$, $Enc(B_1)$ with different randomness then count the number of similar intervals, where B_1 is a noisy version of B_0. The histogram of the counts is shown by the red dotted line in the figure. We then randomly generated 10^6 pairs of B_0, B_1, construct $Enc(B_0)$ and $Enc(B_1)$ and count the number of similar intervals, where B_0 and B_1 are two different biometric templates. The histogram of the counts is shown by the blue solid line in the figure.



(a) $n = 100, q = 10$

(b) $n = 1000, q = 10$

(c) $n = 1000, q = 5$

(d) $n = 10000, q = 5$

Figure 2. The histogram of number of intervals for different n and q.

Let us consider Figure 2(c) where $n = 1000$ and $q = 5$ as an example. When given two sketches with k "similar intervals", $\mathcal{A}_2$ looks at the probability distribution approximated by Figure 2(c). He checks whether the the red dotted line (which approximates the probability that two sketches are from the same identity) is higher than the blue solid line, and then guess the b' which gives the maximum likelihood. When $n = 1000$ and $q = 5$, $\mathcal{A}_2$ should guess $b' = 0$ when $k > 21$ and the probability of $\mathcal{A}_2$ wins the game is less than 0.55. In contrast, under the same parameters, the adversary $\mathcal{A}_1$ for the straightforward Scheme SS_1 is able to distinguish two sketches with a probability at least $(1 + 1 - \frac{2t-1}{q})/2 = 0.9$. Figure 2 also shows that when n gets larger and q gets smaller, the success probability of $\mathcal{A}_2$ approaches $\frac{1}{2}$.

4. Asymmetric Sketch for Set Difference

In this section, we give an extension of fuzzy vault scheme by Juels and Sudan[5] to handle the set difference, where a biometric sample can be represented as a set of elements in a space $\mathbb{Z}_p$. Under asymmetric setting, multiple sets are enrolled and two sets can be extracted: a set $X = \{x_0, x_1, \ldots, x_{m-1}\}$ where $x_i \in \mathbb{Z}_p$ of the elements appeared, and a set V denoting the importance, derived by the consistency, of each element.

Let us first describe the fuzzy vault scheme[5]:

1. Randomly picks a polynomial F of degree $m - 2t - 1$ in field $\mathbb{Z}_p$.
2. Constructs a set $(1, y_1), (2, y_2), \ldots, (p, y_p)$ in this way: For each $i \in Z_p$, if $i \in X$, then y_i is chosen to be $F(i)$, otherwise, randomly picks an element from $\mathbb{Z}_p - \{F(i)\}$ to be y_i.
3. outputs $S = \{(1, y_1), (2, y_2), \ldots, (p, y_p)\}$.

Given a X', the reconstruction process attempts to find the polynomial F using the points $\{(i, y_i) | i \in X'\}$, and then reconstructs X. When there is at least $m - t$ common points in X and X', the polynomial F can be reconstructed. Let us call this Scheme SS_3.

4.1. The Asymmetric Setting

One possible auxiliary information set difference is the importance and consistancy of the elements in the set. Let us consider the case where during enrollment, $B = (X, V)$ is extracted from the multiple samples, where $X = \{x_0, x_1, \ldots, x_{m-1}\}$ is a vector of m elements with $x_i \in \mathbb{Z}_p$, and $V = \{(x_0, v_0), (x_1, v_1), \ldots, (x_{m-1}, v_{m-1})\}$ is the corresponding weights of the elements, with each $v_i \in \mathbb{Z}_q$. During verification, $X' = \{x'_0, x'_1, \ldots, x'_{k-1}\}$ is obtained from the single sample of the biometrics. X' and B are in the close relation $\mathbf{C}$ if the sum of the weights of the common elements is larger than a threshold t, i.e. $\sum_{v \in W} v > t$ where $W = \{v | \exists x, (x, v) \in V \text{ and } x \in (X \cap X')\}$. Scheme SS_3 can be considered as a special case where $m = k$ and all the v_i's are 1.

The main idea of our construction is to extend the above scheme by associating the more important elements to more points to the polynomial F so that they will contribute more roots in verification. Let $H(x, y) = (x + qy)$ be a function on $\mathbb{Z}_q \times \mathbb{Z}_p \to \mathbb{Z}_{pq}$, and the sketch construction is as follow:

1. Randomly picks a polynomial F of degree $g - 2t' - 1$ in field $\mathbb{Z}_{pq}$, where $g = \sum_{v \in V}(v)$ and $t' = (g - t)$.
2. Starts with a set $Y = X$ and an empty set S.
3. For $i = 0$ to $m - 1$, computes $G_i = \{H(0, x_i), H(1, x_i), \ldots, H(q - 1, x_i)\}$, and randomly picks v_i elements from G_i and get the set G'_i. Inserts $(H(j, x_i), F(H(j, x_i)))$ to S for $H(j, x_i) \in G_i$ and inserts $(H(j, x_i), y_{j,x_i})$ to S for $H(j, x_i) \in (G_i - G'_i)$ where y_{j,x_i} is randomly chosen from $\mathbb{Z}_p - \{F(H(j, x_i))\}$.
4. For $i = m$ to r, randomly picks $x_i \notin Y$, inserts x_i to Y, computes $G_i = \{H(0, x_i), H(1, x_i), \ldots, H(q - 1, x_i)\}$ and inserts (x_i, y_i) to S, $(H(j, x_i), y_{j,x_i})$ to S for $H(j, x_i) \in (G_i)$ where y_{j,x_i} is randomly chosen from $\mathbb{Z}_p - \{F(H(j, x_i))\}$.
5. Output S.

During verification, given a $X' = \{x'_0, x'_1, \ldots, x'_{k-1}\}$, Dec first computes the set S' of $\{H(j, x'_i) | x'_i \in X', j \in [0, q - 1]$, and then finds the polynomial F of degree $g - 2t' - 1$ with points in the set. If such F is found, the original X can be reconstructed. The projection $P(B)$ maps $B = (X, V)$ to X. Let us call this Scheme SS_4.

4.2. Security Analysis

Now let us bound the entropy loss of sketch by Scheme SS_4. The recoverable randomness involved is the coefficients of the polynomial F, as well as the generated $y_{j,i}$. Thus the amount of randomness is $(g - 2t' - 1) \cdot \log p + (qp - g) \cdot \log(p - 1)$. By setting the parameter $r = p$, we can omit the $H(j, x_i)$ and have a compact description of the sketch. Hence, the size of sketch is $pq \cdot \log p$ and the entropy loss can be bounded as follow:

$$
\begin{aligned}
& \mathbf{H}_\infty(B) - \widetilde{\mathbf{H}}_\infty(B|P) \\
= \quad & pq \log p - (g - 2t' - 1) \log p - (qp - g) \log(p - 1) \\
= \quad & pq \log \frac{p}{p-1} + g \log \frac{p}{p-1} + (2t' + 1) \log p \\
\leq \quad & q \log e + g \log \frac{p}{p-1} + (2t' + 1) \log p
\end{aligned}
$$

When q is small, and p is large, the bound is similar to symmetric case. However, when q is large, i.e. when the auxiliary information has high entropy, and the amount of information leak can be high.

In the work by Juels and Sudan [5], the security strength is given by the number of spurious polynomials, i.e. polynomials that have degree $m - 2t - 1$ and m roots in the sketches. For the symmetric scheme described above, with probability $1 - \mu$, there exists at least $\frac{\mu}{3} p^{(m-2t-1)-m} (\frac{r}{m})^m$ spurious polynomials.

Similarly, in the asymmetric scheme, with probability $1 - \mu$, there will be at least $\frac{\mu}{3} p^{(g-2t'-1)-g} (\frac{qr}{g-2t'-1})^{g-2t'-1}$ polynomials with degree $g - 2t' - 1$ and g roots. Let us call these polynomials in asymmetric setting the spurious polynomials. However, the analysis of the spurious polynomials is not sufficient for asymmetric setting as the likelihood of a spurious polynomial to be F depends on the distribution of the roots. Let us call a spurious polynomial a *candidate polynomial* if the number of distinct D_i's that contains the roots of the polynomial is less than a threshold a.

The probability that a random spurious polynomial is a candidate polynomial can be view as a variance of the birthday attack analysis. For example, the probability of the case when $q = 2$ (i.e. the consistent elements are twice important as the inconsistent) is as follow:

$$\frac{1}{\binom{2r}{g}} \sum_{x=g-a}^{g/2} \left(2^{g-x} \cdot \binom{r}{g-x}\binom{g-x}{x}\right).$$

For $r = p = 10^4, t = 2, m = 22$ there is 9.7629×10^{33} spurious polynomials with probability $1 - 1/10^4$ in symmetric setting; and with $g = 35$ and $a = 32$, (i.e. the sum of weights is 35, and polynomials with weight higher than 32 are candidate polynomials). There is in total 2.6996×10^{47} spurious polynomials with probability $1 - 1/10^4$. Note that the reason it has more spurious polynomials than symmetric setting is because each element contributes two (chaff) points. Therefore, approximately 2.4113×10^{-5} of the spurious polynomials are candidate polynomials, which is 6.5095×10^{46}.

5. Conclusion

We pointed out that, sketches that reveal auxiliary information could leak important information leading to sketch distinguishability. To reduce the linkages among sketches, we proposed two schemes. The first scheme handles Euclidean distance and it outputs sketches with intervals of unequaled size. The second scheme handles set-differences and caters the different consistency and importance of the set elements. Our schemes and analysis demonstrate that, by mixing the auxiliary information within the biometric data appropriately, we can reduce the linkage of sketches while acquiring the same bound in overall identity information loss measured by entropy loss.

References

[1] A. Arakala, J. Jeffers, and K. Horadam. Fuzzy extractors for minutiae-based fingerprint authentication. *Advances in Biometrics*, pages 760–769, 2007. 1

[2] T.C. Clancy, N. Kiyavash, and D.J. Lin. Secure smartcard-based fingerprint authentication. *ACM Workshop on Biometric Methods and Applications*, pages 45–52, 2003. 1

[3] Y. Dodis, L. Reyzin, and A. Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *Eurocrypt*, pages 523–540, 2004. 1, 2

[4] A.K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *Circuits and Systems for Video Technology*, pages 4–20, 2004. 1

[5] A. Juels and M. Sudan. A fuzzy vault scheme. *Symposium on Information Theory*, pages 408–421, 2002. 1, 5

[6] A. Juels and M. Wattenberg. A fuzzy commitment scheme. *Computer and Communications Security*, pages 28–36, 1999. 1

[7] E. Kelkboom, B. Gókberk, T. Kevenaar, A. Akkermans, and M. van der Veen. "3d face": Biometric template protection for 3d face recognition. *Advances in Biometrics*, pages 566–573, 2007. 1

[8] E.J.C. Kelkboom, J. Breebaart, T.A.M. Kevenaar, I. Buhan, and R.N.J. Veldhuis. Preventing the decodability attack based cross-matching in a fuzzy commitment scheme. *Information Forensics and Security*, pages 107–121, 2011. 2

[9] A. Kholmatov and B. Yanikoglu. Realization of correlation attack against the fuzzy vault scheme. *Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, 2008. 2

[10] Q. Li, M. Guo, and E.C. Chang. Fuzzy extractors for asymmetric biometric representations. *Computer Vision and Pattern Recognition Workshops*, pages 1–6, 2008. 1

[11] K. Simoens, P. Tuyls, and B. Preneel. Privacy weaknesses in biometric sketches. *Symposium on Security and Privacy*, pages 188–203, 2009. 2

[12] U. Uludag, S. Pankanti, S. Prabhakar, and A.K. Jain. Biometric cryptosystems: Issues and challenges. *Proceedings of the IEEE*, pages 948–960, 2004. 1

[13] X. Zhou. Template protection and its implementation in 3d face recognition systems. *Biometric Technology for Human Identification*, pages 214–225, 2007. 1