ON THE (NON-)REUSABILITY OF FUZZY SKETCHES AND EXTRACTORS AND SECURITY IN THE COMPUTATIONAL SETTING*

Marina Blanton and Mehrdad Aliasgari
Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, U.S.A.

Keywords: Biometrics, Fuzzy sketches and extractors, Reusability, Computational setting.

Abstract: Secure sketches and fuzzy extractors enable the use of biometric data in cryptographic applications by correcting errors in noisy biometric readings and producing cryptographic materials suitable for many applications. Such constructions work by producing a public sketch, which is later used to reproduce the original biometric and all derived information exactly from a noisy biometric reading. It has been previously shown that release of multiple sketches associated with a single biometric presents security problems for certain constructions. Through novel analysis we demonstrate that all other constructions in the literature are also prone to similar problems, which hinders their adoption in practice. To mitigate the problem, we propose for each user to store one short secret string for all possible uses of her biometric, and show that simple constructions in the computational setting have numerous security and usability advantages under standard hardness assumptions. Our constructions are generic in that they can be used with any existing secure sketch as a black box.

1 INTRODUCTION

The motivation for this work comes from practical use of biometric-derived data and its suitability for adoption. Secure sketches and fuzzy extractors (Dodis et al., 2004) were introduced as mechanisms of deriving cryptographic material from noisy biometric data, which can be used for authentication, encryption, and other purposes. Such constructions produce a helper string (secure sketch) – which is viewed as public – from a biometric and later re-produce the cryptographic string from a close noisy biometric reading using the helper string. Only minimal information about the biometric should be leaked due to the release of the helper string.

While this powerful concept enables new applications and can be attractive to users who no longer need to maintain secrets to participate in cryptographic protocols, it has been shown that leakage of information associated with the biometric in such constructions is unavoidable (Smith, 2004; Dodis and Smith, 2005). Furthermore, this concept has been more heavily studied in the context when the construction is applied to a biometric only once. Consecutive publications (Boyen, 2004; Simoens et al., 2009) explored the security guarantees of such schemes in terms of their reusability, when a single biometric or its noisy version is used to produce multiple secure sketches using the same or different algorithms. Information leakage prevents such constructions from meeting standard security requirements sought of them in cryptographic applications such as indistinguishability (inability to link two records to the same biometric) and irreversibility (inability to reverse the construction and directly recover information about the biometric). Some of the more popular constructions have been shown to have serious security weaknesses in presence of even very weak adversaries (Simoens et al., 2009). In this work, we analyze other schemes from the literature and show that they also cannot be safely reused. In particular, our novel analysis shows that the remaining constructions fail to satisfy standard security expectations with respect to reusability and therefore cannot be used in security applications.

In such schemes, information leakage is quantified as the entropy loss associated with the release of the helper string, providing a rough upper bound. For the current error rates and typical sets of parameters in biometric data, the information theoretic analysis provides bounds that result in leakage of most or

*This work was partially supported by grant FA9550-09-1-0223 from the Air Force Office of Scientific Research.

Blanton M. and Aliasgari M..
ON THE (NON-)REUSABILITY OF FUZZY SKETCHES AND EXTRACTORS AND SECURITY IN THE COMPUTATIONAL SETTING.
DOI: 10.5220/0003454900680077
In *Proceedings of the International Conference on Security and Cryptography* (SECRYPT-2011), pages 68-77
ISBN: 978-989-8425-71-3
Copyright © 2011 SCITEPRESS (Science and Technology Publications, Lda.)

even all entropy contained in a biometric (see (Blanton and Hudelson, 2009) for a sample iris code analysis). Because this information leakage is unavoidable, it presents problems even in case of weak adversaries.

To overcome the issues of information leakage and unsafe reuse of biometrics, we propose to use the computational setting, where a user stores a single key and the adversary is computationally bounded. The key is introduced for the purpose of avoiding information leakage and improving security of the schemes and does not change the functionality. We believe that keeping a single short key for all possible uses of biometric-based material in different security applications is a small price to pay for achieving significant security improvements (which otherwise are not possible) and the ability to safely use such constructions in a variety of applications. We show that the use of one key and standard computational assumptions (existence of pseudo-random and hash functions) is sufficient for achieving very attractive properties using simple schemes. Our constructions are generic in that they can use any existing secure sketch scheme as a black box for any type of distance metric).

We would like to note that the use of the secret in our schemes should not be confused with multi-factor authentication or the use of shared secrets, as in our schemes the secret never leaves the user and is not shared and a single secret is sufficient for all possible uses including multiple biometric types, multiple applications, and multiple servers.

The security benefits of our schemes are:

- We achieve provably no information leakage.
- Previously, only certain restricted types of error-correcting codes could be used to ensure security of fuzzy sketches and extractors (Boyen, 2004). Our solution lifts such restrictions and can be used with any type of error-correcting code.
- Prior (Simoens et al., 2009) and our analysis of secure sketch constructions shows that they all fail to achieve standard security requirements for cryptographic applications, while our solution is secure in a much stronger adversarial model.
- Previously, exposure of a biometric-derived key was shown to reveal no information about the biometric for a specific construction in the random oracle model (Boyen, 2004). Our construction, on the other hand, achieves this result in the standard model using any existing secure sketch.

In our analysis of existing constructions, we use a very weak adversary. The security of our own schemes, on the other hand, is shown using a very strong adversary (the strongest in the literature).

To summarize, our contributions are two-fold: (i) new analysis of fuzzy sketch schemes that shows that even a weak adversary has a significant advantage in compromising security of existing constructions, and (ii) simple schemes that use a single secret to achieve strong security under standard assumptions.

2 MODEL AND DEFINITIONS

2.1 Fuzzy Sketches and Extractors

Secure (or *fuzzy*) *sketches*, introduced by (Dodis et al., 2004), correct errors in noisy secrets by releasing a helper string S. Let W denote a random variable and w its value.

Definition 1. *A $(\mathcal{M}, m, m', t)$-secure sketch is a pair of randomized algorithms:*

- SS *is a function that, on input w from metric space $\mathcal{M}$ with distance function* dist*, outputs a sketch S.*
- Rec *is a function that, on input $w' \in \mathcal{M}$ and $S = \mathsf{SS}(w)$, recovers and outputs the original w if $\mathsf{dist}(w, w') \leq t$.*

Secure sketches have been constructed for different metric spaces $\mathcal{M}$, for which $\mathsf{dist}(a, b)$ is defined for all $a, b \in \mathcal{M}$. Security of a secure sketch is evaluated in terms of entropy of W before (m) and after (m') releasing the string S, i.e., the entropy loss $m - m'$ associated with making S public. Precise definitions can be found in (Dodis et al., 2008).

Fuzzy extractors allow one to extract randomness from w (to use it as cryptographic material) and later reproduce it using w' close to the original w.

Definition 2. *A $(\mathcal{M}, m, m', t, \varepsilon)$-fuzzy extractor is a pair of algorithms:*

- Gen *is a function that, on input $w \in \mathcal{M}$, outputs extracted random string R and a helper string P.*
- Rep *is a function that, on input w' and P reproduces and outputs R that was generated using $\mathsf{Gen}(w)$ if $\mathsf{dist}(w, w') \leq t$.*

The security requirement is that, for any W of min-entropy m, the statistical distance between the distribution of R and the uniform distribution of strings of the same length is no greater than ε, even after observing P. A fuzzy extractor can be built from a secure sketch using a generic construction from (Dodis et al., 2004):

$\mathsf{Gen}(w)$:

1. Execute $S \leftarrow \mathsf{SS}(w; r_1)$, where r_1 denotes random coins used by SS (if any).
2. Use a strong extractor Ext to extract a random string R from w, i.e., $R \leftarrow \mathsf{Ext}(w; r_2)$, where r_2 denotes random coins used by Ext.
3. Output public $P = (S, r_2)$ and secret R.

$\mathsf{Rep}(w', P = (S, r_2))$:

1. Execute $w \leftarrow \mathsf{Rec}(w', S)$. If Rec fails (i.e., when $\mathsf{dist}(w, w') > t$ such that $S = \mathsf{SS}(w)$), stop.
2. Extract R from w using r_2 as $R \leftarrow \mathsf{Ext}(w, r_2)$ and output R.

Strong extractors (Nisan and Ta-Shma, 1999) can extract at most $m - 2\log(\frac{1}{\varepsilon}) + O(1)$ nearly random bits (m and ε are as defined above). The entropy loss of $2\log(\frac{1}{\varepsilon}) + O(1)$ is in addition to the loss due to release of sketch S, unless the extractor is modeled as a random oracle.

Many constructions utilize error-correcting codes. A code C is a subset of K elements $\{w_0, \ldots, w_{K-1}\}$ of $\mathcal{M}$. The minimum distance of C is the smallest d such that $\mathsf{dist}(w_i, w_j) \geq d$ for all $i \neq j$, which implies that the code can detect up to $d-1$ errors; and the error-correcting distance is $t = \lfloor (d-1)/2 \rfloor$. A linear error-correcting code C over field $\mathbb{F}_q$ is a k-dimensional linear subspace of the vector space $\mathbb{F}_q^n$ which uses Hamming distance as the metric. For any linear code C, an $(n-k) \times n$ parity-check matrix H projects any vector $v \in \mathbb{F}_q^n$ to the space orthogonal to C. This projection is called the syndrome and denoted by $\mathsf{syn}(v) = Hv$. Then $v \in C$ iff $\mathsf{syn}(v) = 0$. The syndrome contains all information necessary for decoding, i.e., when codeword c is transmitted and noisy $w = c + e$ is received, $\mathsf{syn}(w) = \mathsf{syn}(c) + \mathsf{syn}(e) = 0 + \mathsf{syn}(e)$, where $\mathsf{syn}(e)$ can be used to determine the error pattern e.

Metric-specific secure sketch constructions are known for the Hamming distance (used for iris codes), the set difference (used for fingerprints), and the edit distance (used for DNA comparisons). Also, the permutation-based construction is available for any transitive metric (e.g., Hamming distance and set intersection). Schemes for the Hamming distance have been most heavily analyzed, and some schemes are known to have security problems when reused on related biometrics. In this work we analyze remaining known constructions and show their insecurity.

2.2 Secure Sketch Constructions

(Simoens et al., 2009) show that two popular secure sketch constructions – the code offset construction with a linear error-correcting code (the syndrome construction) and the construction based on permutation groups – do not withstand the requirements of indistinguishability and reversibility, i.e., the adversary can win such experiments with overwhelming probability. The former scheme is for the Hamming distance (and is among the most widely studied schemes) and the latter is for any transitive distance metric. We concentrate on the analysis of other schemes and outline schemes for the set difference and edit distance. In what follows, we use $a \xleftarrow{R} A$ to denote that the value a is chosen uniformly at random from the set A.

Fuzzy Vault. The fuzzy vault scheme (Juels and Sudan, 2002) can be used as a fuzzy sketch for set difference. A biometric is comprised of unordered elements $w = \{w_1, \ldots, w_s\}$ (e.g., minutiae points in fingerprints), which are disguised by adding a large number of *chaff points*. The genuine points carry information that allows w to be reconstructed from noisy w'. Here $t \in [1, s]$ and $r \in [s+1, n]$ are system-wide parameters, where n is the set of all possible points, or the universe. Work is over field $\mathbb{F}_n$, where n is a prime power.

To compute $\mathsf{SS}(w)$:

1. Choose a random polynomial $p(\cdot)$ of degree at most $s - t - 1$ over $\mathbb{F}_n$.
2. For each $w_i \in w$, let $x_i = w_i$ and $y_i = p(x_i)$.
3. Choose $r - s$ distinct points $x_{s+1}, \ldots, x_r$ at random from $\mathbb{F}_n \setminus w$ and set $y_i \xleftarrow{R} \mathbb{F}_n \setminus \{p(x_i)\}$ for $i = s+1, \ldots, r$.
4. Output $\mathsf{SS}(w) = \{(x_1, y_1), \ldots, (x_s, y_s)\}$ sorted by the value of x_i's.

To compute $\mathsf{Rec}(w', S)$:

1. Create the set D of pairs (x_i, y_i) such that $x_i \in w'$.
2. Run Reed-Solomon decoding on D to recover the polynomial $p(\cdot)$.
3. Output s points of the form $(x_i, p(x_i))$ from S.

Privacy of the biometric depends on the number and distribution of points S (i.e., the difficulty of identifying the original points and the number of spurious polynomials created by the chaff points). The entropy loss due to the release of S is upper bounded by $t \log n + \log \binom{n}{r} - \log \binom{n-s}{r-s} + 2$.

Improved Fuzzy Vault. (Dodis et al., 2008) observed that the polynomial in the above construction does not need to be random, which allows for a secure sketch with significantly lower entropy loss, $t \log n$.

To compute $\mathsf{SS}(w)$:

1. Compute unique monic polynomial $p(x) = \prod_{w_i \in w}(x - w_i)$ of degree s.
2. Output the coefficients of $p()$ of degree $s-1$ down to $s-t$, which will form $\mathsf{SS}(w) = (c_{s-1}, \ldots, c_{s-t})$.

To compute $\mathsf{Rec}(w', S = (c_{s-1}, \ldots, c_{s-t}))$:

1. Create a new polynomial p_{high} of degree s that shares the top $t+1$ coefficients with $p()$, i.e., $p_{\text{high}}(x) = x^s + \sum_{i=s-t}^{s-1} c_i x^i$.
2. Evaluate p_{high} on points of w' to obtain pairs $(a_1, b_1), \ldots, (a_s, b_s)$.
3. Use Reed-Solomon decoding to find a polynomial p_{low} of degree $s - t - 1$ such that $p_{\text{low}}(a_i) = b_i$ for at least $s - t/2$ values of a_i's. If none are found, output fail.
4. Output the roots of the polynomial $p_{\text{high}} - p_{\text{low}}$.

Another construction for set difference, **Pinsketch**, is suitable for large universe sizes and variable number of elements in w. It is syndrome-based, and its (in)security is not difficult to reduce to the previously analyzed code-offset scheme. We thus omit its analysis. For the edit distance, the only known way to construct a secure sketch is by embedding it into a transitive metric of larger dimension and applying a secure sketch construction to the target metric. An embedding with attractive properties was developed in (Dodis et al., 2008) using Pinsketch. Once again, the insecurity of the resulting scheme can be shown using prior results and is omitted. This covers all secure sketch schemes.

2.3 Security Notions

The original security definitions of fuzzy sketches and extractors were formulated for a single instance of a fuzzy sketch or extractor in isolation (Dodis et al., 2004). Consecutive literature (Boyen, 2004; Simoens et al., 2009) considered a stronger (and more realistic) adversarial model where such constructions can be invoked multiple times and therefore the security guarantees must hold when the constructions are reused. Furthermore, the power granted to the adversary can greatly differ. In this work we use weak adversaries while analyzing existing construction (to show that they do not provide sufficient security guarantees even in presence of weak adversaries) and strong adversaries when proving security of our proposed solution. In a nutshell, a weak adversary is given two fuzzy sketches and tries to determine whether they were produced using related biometrics or what the biometric was, while a strong adversary can adaptively ask for fuzzy sketches and private keys that fuzzy extractors output.

Let t be the maximum amount of errors that the biometric system can tolerate. We define Δ_t to be the set of all perturbation functions that represent differences in sampling biometric data; we get $\Delta_t = \{\delta : \mathcal{M} \to \mathcal{M}$ such that $\mathsf{dist}(w, \delta(w)) \leq t\}$. We next define a security game for weak adversaries with access to public sketches and then proceed with a security game for strong adversaries. Two security properties for weak adversaries were defined in (Simoens et al., 2009): sketch indistinguishability and irreversibility.

2-Indistinguishability Game. (Simoens et al., 2009):

1. The challenger chooses a random variable $W \in \mathcal{M}$ and samples it to obtain $w_1 \in \mathcal{M}$. The challenger computes $S_1 = \mathsf{SS}(w_1)$ and gives S_1 to $\mathcal{A}$.
2. The challenger chooses $b \xleftarrow{R} \{0,1\}$. If $b = 1$, the challenger chooses $\delta \xleftarrow{R} \Delta_t$ and produces related $w_2 = \delta(w_1)$. Otherwise, the challenger samples W to obtain a different w_2. The challenger computes $S_2 = \mathsf{SS}(w_2)$ and gives S_2 to $\mathcal{A}$.
3. The adversary $\mathcal{A}$ eventually produces a bit b' and wins if $b' = b$.

$\mathcal{A}$'s advantage in this game is defined as $\mathsf{Adv}^{\mathsf{ind}}_{\mathcal{A}} = 2\left|\Pr[b' = b] - \frac{1}{2}\right| = 2\left|\Pr[b' \neq b] - \frac{1}{2}\right|$.

Definition 3. *An $(\mathcal{M}, m, m', t)$-secure fuzzy sketch (SS, Rec) is ε-indistinguishable in Δ_t if for any adversary $\mathcal{A}$ it holds that $\mathsf{Adv}^{\mathsf{ind}}_{\mathcal{A}} \leq \varepsilon$. The fuzzy sketch is reusable when ε is negligible.*

The irreversibility property of a fuzzy sketch scheme means that an adversary who obtains access to multiple sketches generated from the same noisy input using possibly different sketching functions is unable to recover the original input. In the current version of this work we do not treat irreversibility, since a failure to achieve the indistinguishability property alone points out weaknesses of a fuzzy sketch scheme.

We now proceed with defining security games for more powerful adversaries using what we term *weak biometric privacy* and *strong biometric privacy*. In both of them the adversary is allowed to query the scheme a large number of times, but the difference is that in the first the adversary obtains access only to the public information, while in the second it also obtains access to the key output by a fuzzy extractor. Thus, we use the first definition for secure sketches and the second one for fuzzy extractors.

The two security games below are roughly equivalent to outsider and insider chosen perturbation security in (Boyen, 2004), but are stronger than the respective definitions in (Boyen, 2004). In particular, in our definition of weak biometric security we require the adversary to only distinguish between two sketches, while the adversary was required to recover the biometric w in (Boyen, 2004). Furthermore, instead of allowing the adversary to query fuzzy sketches for a particular biometric w and then challenging the adversary by asking it to distinguish between a sketch for w and a sketch for a randomly chosen biometric, we setup two biometrics w_0 and w_1 and allow the adversary to query sketches for both. Then during the challenge, the adversary is asked to determine which biometric was used in producing the challenge sketch. This can give the adversary advantage over the prior formulation, especially in the computational setting where different users will possess different keys.

As our schemes work in the computational setting, we use κ to denote the security parameter. All algorithms are assumed to be polynomial time in κ. Then a function $\varepsilon(\kappa)$ is negligible if for all positive polynomials $p(\cdot)$ and sufficiently large κ $\varepsilon(\kappa) < 1/p(\kappa)$.

Weak Biometric Privacy.

1. (Preparation) $\mathcal{A}$ chooses a random variable $W \in \mathcal{M}$ and sends its specification to the challenger.
2. (Sampling) The challenger randomly samples W to obtain $w_0 \in \mathcal{M}$ and $w_1 \in \mathcal{M}$ and initializes two users $\mathcal{U}_0$ and $\mathcal{U}_1$, resp., using that information.
3. (Queries) $\mathcal{A}$ makes up to q possibly adaptive sketching queries: to form query i, $\mathcal{A}$ chooses $\delta_i \in \Delta_t$ and sends it and a bit b_i to the challenger. The challenger computes $S_i \leftarrow \mathsf{SS}(\delta_i(w_{b_i}); r_i)$ using fresh randomness r_i and returns S_i to $\mathcal{A}$.
4. (Challenge) The challenger chooses a bit $b \xleftarrow{R} \{0,1\}$ and $\delta \xleftarrow{R} \Delta_t$, and produces a biometric $w' = \delta(w_b)$. The challenger then computes $S \leftarrow \mathsf{SS}(w'; r)$ using fresh random r and gives S to $\mathcal{A}$.
5. (More queries) $\mathcal{A}$ can run more queries up to the bound q as specified in step 3.
6. (Response) $\mathcal{A}$ eventually produces a bit b' and wins if $b' = b$.

$\mathcal{A}$'s advantage in this game is defined as $\mathsf{Adv}^{\mathsf{wbp}}_{\mathcal{A}}(\kappa) = 2\left|\Pr[b' = b] - \frac{1}{2}\right| = 2\left|\Pr[b' \neq b] - \frac{1}{2}\right|$.

Definition 4. *An $(\mathcal{M}, m, m', t)$-secure fuzzy sketch (SS, Rec) has weak biometric privacy if for any probabilistic polynomial-time (PPT) adversary $\mathcal{A}$ it holds that $\mathsf{Adv}^{\mathsf{wbp}}_{\mathcal{A}}(\kappa) \leq \varepsilon(\kappa)$ for a negligibly small $\varepsilon(\kappa)$.*

Note that unlike previous definitions, we explicitly specify the security parameter κ and define the adversary's advantage as a function of it.

The next definition corresponds to the strongest version of the insider chosen perturbation security definition in (Boyen, 2004). The adversary can query the challenger to obtain sketches on both related and unrelated biometrics and private key corresponding to unrelated biometrics. This time we ask the adversary to distinguish between the secret key output by a fuzzy extractor on a related biometric and a randomly chosen string. Note that we do not ask the adversary to distinguish between biometric-derived keys of two users because the adversary has the choice of the sketch that it can use in the challenge. This means that the adversary will trivially know for which user the secret key will be produced. We, however, note that in order to distinguish secret keys corresponding to two users, the adversary need to be able to distinguish at least one of them from a random string. Thus, our definition of security will imply the security in the game with two users. Let Δ denote all perturbation functions over space $\mathcal{M}$, i.e., $\Delta = \{\delta : \mathcal{M} \rightarrow \mathcal{M}\}$ where $\mathsf{dist}(w, \delta(w))$ can be greater than t.

Strong Biometric Privacy.

1. (Preparation) $\mathcal{A}$ chooses $W \in \mathcal{M}$ and gives its specification to the challenger.
2. (Sampling) The challenger randomly samples W to obtain $w \in \mathcal{M}$.
3. (Public queries) $\mathcal{A}$ makes up to q possibly adaptive generation queries: to form query i, $\mathcal{A}$ chooses $\delta_i \in \Delta$ and sends it to the challenger. The challenger computes $(P_i, R_i) \leftarrow \mathsf{Gen}(\delta_i(w); r_i)$ using fresh random r_i and returns public P_i to $\mathcal{A}$.
4. (Private queries) $\mathcal{A}$ makes up to q' possibly adaptive reproduction queries that can be interspersed with public queries as follows: to form query i, $\mathcal{A}$ chooses $\delta'_i \in \Delta$ and a public data P'_i and sends them to the challenger. The challenger computes $R'_i \leftarrow \mathsf{Rep}(\delta'_i(w); P'_i)$ and returns R'_i to $\mathcal{A}$.
5. (Challenge) $\mathcal{A}$ chooses string $P^* \in \{P_1, \ldots, P_q\}$ from one of the strings returned by the challenger in a public query such that P^* was produced using a public query δ_i with $\mathsf{dist}(w, \delta_i(w)) \leq t$ and in any private query (δ'_i, P^*) the distance $\mathsf{dist}(w, \delta'_i(w)) > t$. $\mathcal{A}$ sends P^* to the challenger. The challenger chooses a bit $b \xleftarrow{R} \{0,1\}$. If $b = 1$, the challenger computes $R \leftarrow \mathsf{Rep}(w, P^*)$ and gives it to $\mathcal{A}$. Otherwise, if $b = 0$, it chooses a random string of the same length and gives it to $\mathcal{A}$ instead.
6. (More queries) $\mathcal{A}$ can run additional queries as specified in steps 3 and 4 (up to q and q' queries, respectively) with the exception that any query (δ, P^*) such that $\mathsf{dist}(w, \delta(w)) \leq t$ is not allowed.
7. (Response) $\mathcal{A}$ eventually produces a bit b' and wins if $b' = b$.

$\mathcal{A}$'s advantage in this game is defined as $\mathsf{Adv}^{\mathsf{sbp}}_{\mathcal{A}}(\kappa) = 2\left|\Pr[b' = b] - \frac{1}{2}\right| = 2\left|\Pr[b' \neq b] - \frac{1}{2}\right|$.

Definition 5. *We say that an $(\mathcal{M}, m, m', t, \varepsilon)$-secure fuzzy extractor (Gen, Rep) has strong biometric privacy if for any PPT adversary $\mathcal{A}$ it holds that $\mathsf{Adv}^{\mathsf{sbp}}_{\mathcal{A}}(\kappa) \leq \varepsilon(\kappa)$ for a negligibly small $\varepsilon(\kappa)$.*

3 ANALYSIS OF EXISTING SCHEMES

Fuzzy Vault. Before proceeding with the analysis, we note that the basic idea for the strategy in attacking the fuzzy vault scheme when two or more sketches are available – computing the intersection of the points – is straightforward and is not new. This attack appeared in (Scheirer and Boult, 2007; Kholmatov and Yanikoglu, 2008; Poon and Miri, 2009). We still analyze the construction here because all previous publications assume that given sketches are related and proceed with identifying original points. Our work, however, assumes a significantly weaker (and perhaps more realistic) adversary that would like to determine if two given sketches are related or not, which is a

much more difficult task. Therefore, we present a rigorous new analysis that shows weaknesses of the scheme even in the presence of the weakest adversary.

The adversary receives two secure sketches $P_1 = \{(x_1,y_1),\ldots,(x_r,y_r)\}$ and $P_2 = \{(x'_1,y'_1),\ldots,(x'_r,y'_r)\}$, and its goal is to determine the coin flip, i.e., whether the biometrics w_1 and w_2 are related or not. Let P_1^x and P_2^x denote projections of P_1 and P_2, resp., on the x-coordinate, i.e., $P_1^x = \{x_1,\ldots,x_r\}$ and $P_2^x = \{x'_1,\ldots,x'_r\}$. The basic attack idea is to compute the intersection of P_1^x and P_2^x and use its size to make a distinction between related and unrelated biometrics. Related sketches will overlap in at least $s-t$ original biometric points, while unrelated sketches will have fewer original biometric points overlap. In addition, a number of chaff points in P_1^x can collide with chaff points in P_2^x or points in $w_2 \setminus (w_1 \cap w_2)$ (similarly, points from $w_1 \setminus (w_1 \cap w_2)$ can collide with chaff points in P_2^x). Thus, the size of $P_1^x \cap P_2^x$ follows a certain distribution, but the expected overlap size is larger for related sketches. We first analyze the properties of such a distribution.

Let $\alpha = |w_1 \cap w_2|$ denote the number of biometric points in the intersection, i.e., $\alpha \geq s-t$ for related biometric samples and $\alpha \leq s-t-1$ otherwise. Let $a = r-\alpha$ and $b = n-\alpha$, i.e., a is the number of sketch points that do not correspond to the overlapping biometric points and b is the overall space for such points. As customary in the literature, we assume that the biometric points of w are distributed uniformly in the space; the chaff points are also drawn uniformly at random from the remaining space. Then to determine how many points from $P'_1 = P_1^x \setminus (w_1 \cap w_2)$ will collide with points from $P'_2 = P_2^x \setminus (w_1 \cap w_2)$, suppose there are $b = n-\alpha$ bins and points from P'_1 occupy $a = r-\alpha$ of them, i.e., there are a random bins with a ball in them. Then we throw another a balls (points from P'_2) into the bins without replacement and count the number of bins with two balls in them (i.e., if a bin has two balls, it is removed, so that no bin has more than two balls; this is dictated by the requirement that all r points in a sketch are distinct). The above can be modeled as hypergeometric experiment. Let X be a random variable that corresponds to the number of collisions in P_1^x and P_2^x (i.e, its size is $|(P_1^x \cap P_2^x) \setminus (w_1 \cap w_2)|$). We obtain:

$$\Pr[X=k] = \tbinom{a}{k}\tbinom{b-a}{a-k} / \tbinom{b}{a}$$

where X can range between 0 and a. This distribution's mean value is $E[X] = a \cdot (a/b)$.

This analysis leads to the following attack strategy: given sketches P_1 and P_2, $\mathcal{A}$ computes P_1^x, P_2^x, and $c = |P_1^x \cap P_2^x|$. Let β denote the value $(r-s+t)^2/(n-s+t)$ rounded to the nearest integer. If $c \geq (s-t+\beta)$, output 1, otherwise, output 0.

Let α_{auth} (α_{imp}) denote a random variable corresponding to the distribution of $|w_1 \cap w_2|$ when w_1 and w_2 are related or authentic (unrelated or impostor, resp.). Adversary $\mathcal{A}$ has the smallest probability of distinguishing between authentic and impostor sketches when the values of α_{auth} and α_{imp} are the closest, i.e, $\alpha_{\mathsf{auth}} = s-t$ and $\alpha_{\mathsf{imp}} = s-t-1$. According to the indistinguishability definition, we have $\mathsf{Adv}^{\mathsf{ind}}_{\mathcal{A}} = 2\left|\Pr[b'=b] - \frac{1}{2}\right|$. If we let X_1 denote the random variable distributed according to the hypergeometric distribution above with $\alpha_1 = s-t$ and X_2 denote a similar random variable with $\alpha_2 = s-t-1$, we obtain that $\mathcal{A}$ is successful with at least:

$$\begin{aligned}
\Pr[b'=b] &= \Pr[b'=1\,|\,b=1]\Pr[b=1] + \\
&+ \Pr[b'=0\,|\,b=0]\Pr[b=0] \geq \\
&\geq \frac{1}{2}\big(\Pr[X_1 \geq c-\alpha_1] + \Pr[X_2 < c-\alpha_2]\big) \\
&= \frac{1}{2}\big(\Pr[X_1 \geq \beta] + \Pr[X_2 < \beta+1]\big) = \\
&= \frac{1}{2}\Bigg(\sum_{i=\beta}^{r-s+t} \frac{\binom{r-s+t}{i}\binom{n-r}{r-s+t-i}}{\binom{n-s+t}{r-s+t}} + \\
&+ \sum_{i=0}^{\beta} \frac{\binom{r-s+t+1}{i}\binom{n-r}{r-s+t+1-i}}{\binom{n-s+t+1}{r-s+t+1}}\Bigg).
\end{aligned}$$

This probability and $\mathsf{Adv}^{\mathsf{ind}}_{\mathcal{A}}$ can be easily computed for a given set of parameters n, r, s, and t. In reality, each parameter above has limitations placed on it by the behavior of the actual biometric data. In particular, (Clancy et al., 2003) study applicability of the fuzzy vault construction to fingerprint data and determines optimal parameters to use to achieve adequate resistance of the construction against brute force search (when an adversary is given a sketch and tries to determine sensitive information by searching through polynomials). While the fuzzy vault construction was not used exactly as a secure sketch in (Clancy et al., 2003) and was generalized, we nevertheless obtain information about the parameters that would be used for fingerprint data. The field $\mathbb{F}_{p^2}$, for prime p, is used for representing fingerprint features in 2-D and the value of p is set to 251 giving us $n = 251^2 = 63001$ (this value of n also provides many choices for the decoding algorithm). The number of biometric points in a fingerprint was empirically determined on average to be $s = 38$ (it can vary based on the equipment and quality of data, but generally is in a similar range). For this value of s, having 20 points overlap would provide excellent distinguishing capability and low false acceptance rate (Pankanti et al., 2002). Finally, the value of r is constrained in that the complexity of decoding for legitimate users can grow as r increases (this is caused by spurious polynomials introduced by the chaff points). In particular, at

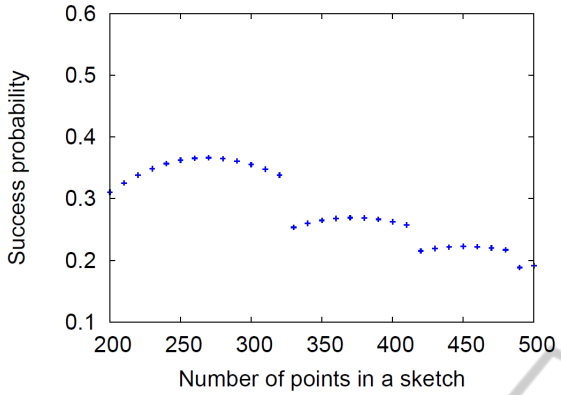

Figure 1: Adversary advantage $\mathsf{Adv}^{\mathsf{ind}}_{\mathcal{A}}$ with parameters $n = 251^2$, $s = 38$, $t = 20$, and varying r.

the decoding time, when a legitimate user computes $w_2 \cap S$, where $S = \mathsf{SS}(w_1)$, the decoding complexity can grow when points from $w_2 \setminus (w_2 \cap w_1)$ coincide with chaff points in S. Since $|w_2 \setminus (w_2 \cap w_1)| \leq t$ for legitimate users, the experiment now consists of throwing t points in $b = n - s + t$ bins, where $a = r - s + t$ bins are already occupied. We want r to be such that the expected (integer-valued) number of collisions $t(a/b)$ is 0.

Figure 1 plots the adversary's advantage $\mathsf{Adv}^{\mathsf{ind}}_{\mathcal{A}}$ for the above parameters as a function of r near the suggested in (Clancy et al., 2003) value of $r \approx 300$. As evident from the figure, the advantage is significant even in the worst (for the adversary) case when only one overlapping point separates authentic data from impostor. The jumps in the plot correspond to the places where the (integer-valued) mean of the distribution, $E[X]$, increases by 1.

Improved Fuzzy Vault. An important observation in designing an attack strategy for this construction is that it is deterministic. This immediately implies that the same biometric will always produce the same secure sketch, giving the adversary the ability to distinguish sketches. Thus, as an important special case we first consider the adversary's ability to win the indistinguishability game when no noise affects multiple sketches of the same w (this arises in several applications, where multiple keys are issued using the same copy of w). Thus, when $\mathcal{A}$ obtains challenge S_2, it outputs 1 if $S_2 = S_1$ and 0 otherwise. This means that when $b = 1$, $\mathcal{A}$ will always guess the bit correctly, but when $b = 0$ it might still sometimes output 1 if the two sketches happened to be the same. The probability of the latter, however, is small and can be bound as follows. Recall that sketch S consists of t coefficients of a polynomial $p(x) = x^s + c_{s-1}x^{s-1} + \ldots + c_1 x + c_0$, where for $w = \{w_1, \ldots, w_s\}$ $c_{s-1} = \sum_i w_i$, $c_{s-2} = \sum_{i \neq j} w_i w_j$, $\ldots$, $c_{s-t} = \sum_{C \subset [1,s], |C| = t} (\prod_{i \in C} w_i)$.

First, for an unrelated random biometric $\hat{w}$, the probability that $\sum_i \hat{w}_i = c_{s-1}$ is $\frac{1}{n}$ (i.e., without any restrictions, there are $\prod_{i=0}^{s-1}(n-i)$ choices for s elements without repetitions from the set of n elements, and when the sum of the elements is fixed (in $\mathbb{F}_n$), the number reduces to $\prod_{i=1}^{s-1}(n-i)$).

Now consider c_{s-2}. We start with a simpler function $x_1 x_2 = b$ in $\mathbb{F}_n$ for a fixed value of b. Recall that $n = p^2$ for a prime p. We enumerate all possible solutions x_1 and x_2 for this function such that $x_1 \neq x_2$ (since all points in a biometric are different). When b is zero, there are $n-1$ unordered pairs (x_1, x_2) with $x_1 \neq x_2$ whose product equals to b (one value is zero and the other can take $n-1$ remaining values). All elements other than zero form a cyclic multiplicative group, and when $b \neq 0$ there are either $\frac{n-1}{2}$ or $\frac{n-1}{2} - 1$ pairs (x_1, x_2) with distinct x_1 and x_2, when b is a quadratic non-residue or quadratic residue, resp.. Therefore, the number of pairs (x_1, x_2) satisfying the congruence for any b is at most $n-1$ from the overall space of $\frac{n(n-1)}{2}$ such pairs, giving us the fraction $(n-1)/\frac{n(n-1)}{2} = \frac{2}{n}$.

Now recall that c_{s-2} is composed of a summation of products $w_i w_j$ for each $i \neq j$. When there is only one product $w_1 w_2$ (i.e., $s = 2$), we obtain that it is equal to 0 more frequently than to other values. When, however, $s > 2$ this is no longer the case. Because all w_i have to be unique and each w_i appears in a number of products $w_i w_j$, the value of the sum tends to be distributed more evenly as s increases. This means that the frequency of the most common value of c_{s-2} approaches $\frac{1}{n}$ when s grows. To illustrate this phenomenon, we plot empirical data for small values of $n = p^2$. In particular, for $s = 2$, 4, and 6 and all possible $w = (w_1, \ldots, w_s) \in \mathbb{F}_n^s$ we find the value of the sum which occurs the highest number of times. Let it be denoted by $count_{\max}$ and the fraction of all biometrics w that results in such value by $f_{\max} = count_{\max}/\binom{n}{s}$. To evaluate how the value of $f_{\max}$ compares to $\frac{1}{n}$, we plot their ratio $f_{\max}/\frac{1}{n}$ in Figure 2. For $s = 2$, $f_{\max} = \frac{2}{n}$ is constant; for $s > 2$ it is clear that $f_{\max}$ rapidly approaches $\frac{1}{n}$ from the above even for very small values of s. This means that $\frac{2}{n}$ is a generous upper bound on the probability that c_{s-2} of a randomly chosen $\hat{w}$ will coincide with a specific value of that coefficient for an unrelated biometric w.

Extending this analysis to $c_{s-3} = \sum w_i w_j w_k$, where i, j, and k are pairwise distinct, we obtain that the most frequently occurring value of c_{s-3} is 0 and when $s = 3$ (i.e., only one product). In that case, the number of possibilities that result in that product is $\frac{(n-1)(n-2)}{2}$ out of $\frac{n(n-1)(n-2)}{2 \cdot 3}$ total choices (and the number of

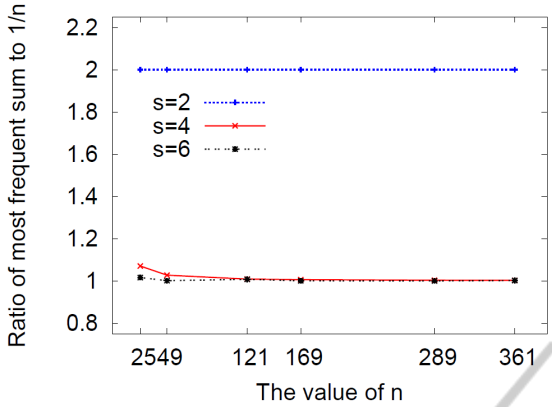

Figure 2: The ratio of the fraction of most frequent value of the sum c_{s-2} to $\frac{1}{n}$ for varying n and s.

possibilities when the product is non-zero is at most $\frac{n-3}{2} \cdot \frac{n-1}{2}$). Thus, the fraction of triples that can result in any given product is $\leq \frac{3}{n}$. For c_{s-4}, the maximum fraction is $\leq \frac{4}{n}$; for c_{s-5}, it is $\leq \frac{5}{n}$, etc. Therefore, the adversarial error is at most $\frac{t!}{n^t}$, and in practice will be close to $\frac{1}{n^t}$ because $s > t$. Both of these quantities are very low even for small values of t (such as 2), and the probability with which the adversary considers two unrelated biometrics to be related is very small. Its advantage in the 2-indistinguishability game is:

$$\begin{aligned}
\mathsf{Adv}^{\mathsf{ind}}_{\mathcal{A}} &= 2\left|\Pr[b'=b]-\frac{1}{2}\right| = \\
&= 2\left|\Pr[b'=1|b=1]\Pr[b=1]+\right. \\
&\left.+\Pr[b'=0|b=0]\Pr[b=0]-\frac{1}{2}\right| = \\
&= \left|2\Pr[b'=1|b=1]\frac{1}{2}+2\Pr[b'=0|b=0]\frac{1}{2}-1\right| \\
&= \left|\Pr[b'=1|b=1]+1-\Pr[b'=1|b=0]-1\right| \\
&= \left|\Pr[b'=1|b=1]-\Pr[b'=1|b=0]\right| > 1-\frac{t!}{n^t}.
\end{aligned}$$

The above analysis addresses an important special case of $w = w'$. We defer analysis of the more general case of related sketches to the full version.

4 OUR CONSTRUCTIONS

In what follows, let $(\mathsf{SS}', \mathsf{Rec}')$ denote any existing fuzzy sketch scheme (for any metric). The key k denotes the long-term user's key of size κ, where κ is the security parameter. This key k is not shared with any parties. We first provide additional definitions.

Definition 6. *Let $F : \{0,1\}^\kappa \times \{0,1\}^{\ell_1(\kappa)} \to \{0,1\}^{\ell_1(\kappa)}$ be a family of functions. For $k \in \{0,1\}^\kappa$, the function $F_k : \{0,1\}^{\ell_1(\kappa)} \to \{0,1\}^{\ell_1(\kappa)}$ is defined as $F_k(x) = F(k,x)$. F is said to be a family of pseudo-random functions (PRF) if for every PPT adversary $\mathcal{A}$ with oracle access to a function F_k and all sufficiently large κ, $|\Pr[\mathcal{A}^{F_k}(1^\kappa) - \Pr[\mathcal{A}^f(1^\kappa)]|$ is negligible in κ, where $k \xleftarrow{R} \{0,1\}^\kappa$ and f is a function chosen at random from all possible functions mapping $\ell_1(\kappa)$-bit inputs to $\ell_1(\kappa)$-bit outputs.*

Definition 7. *A family of functions $h : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^{\ell_2(\kappa)}$ is pairwise independent universal hash function if for all $x, x' \in \{0,1\}^n$, where $x \neq x$, $\Pr[h_y(x) = h_y(x')] = 1/2^{\ell_2(\kappa)}$ for $y \in \{0,1\}^\kappa$.*

In the following secure sketch construction, it is required that $\ell_1(\kappa) \geq |\mathsf{SS}'(w)|$, where $|a|$ is the length of string a. We discuss the choice of parameters later.

To compute $\mathsf{SS}(w,k)$:
1. Choose $r_1 \in \{0,1\}^{\ell_1(\kappa)}$ at random.
2. Output $S = (S_1, S_2) = (r_1, F_k(r_1) \oplus \mathsf{SS}'(w))$.

To compute $\mathsf{Rec}(w', k, S = (S_1, S_2))$:
1. Compute $u \leftarrow F_k(S_1)$.
2. Output what $\mathsf{Rec}'(w', S_2 \oplus u)$ outputs.

Theorem 1. *Assuming that F is a family of PRFs, the above fuzzy sketch scheme achieves weak biometric privacy.*

We omit security proofs due to space constraints.

Note that in our construction deterministic schemes for the underlying SS' are preferred because they produce most concise sketches. So far we assumed that the output length of F, $\ell_1(\kappa)$, is at least as large as the output length of secure sketch $|\mathsf{SS}'(w)|$. While this will hold for many types of biometrics and a reasonable choice of security parameter κ, in some cases the representation of $\mathsf{SS}'(w)$ can be longer. Instead of increasing κ, we suggest modifying the algorithm to use more than one application of F to produce a longer pseudo-random sequence. For instance, if $\ell_1(\kappa) < |\mathsf{SS}'(w)| \leq 2\ell_1(\kappa)$, the sketch can be produced as $(r_1, (F_k(r_1)||F_k((r_1+1) \bmod 2^\kappa)) \oplus \mathsf{SS}'(w))$, where $||$ denotes string concatenation. This increases the number of random values on which F is evaluated and thus the probability of their collision. However, as long as $|\mathsf{SS}'(w)|/\ell_1(\kappa)$ is a constant or polynomial in κ, the security guarantees still hold.

In the fuzzy extractor construction below we split the key k into two keys k_1 and k_2. This is done to simplify the analysis. In practice, the sub-keys k_1 and k_2 can be computed by applying a PRF keyed with k to two different inputs.

To compute $\mathsf{Gen}(w, k_1, k_2)$:
1. Compute $S = \mathsf{SS}(w, k_1)$ using the fuzzy sketch scheme above.

2. Choose $r_2 \xleftarrow{R} \{0,1\}^\kappa$ and compute $s \leftarrow h_{r_2}(w)$.
3. Output $P = (S, r_2)$ and $R \leftarrow F_{k_2}(s)$.

To compute $\mathsf{Rep}(w', k_1, k_2, P = (P_1, P_2))$

1. Run $\mathsf{Rec}(w', k_2, P_1)$ above to recover w. If it fails, output $\perp$.
2. Otherwise, reproduce the key R as $F_{k_2}(s')$, where $s' \leftarrow h_{P_2}(w)$, and output R.

When it is desirable that failures during reconstruction are not reported explicitly, Rep can be modified to output a (wrong) private string, e.g., computed as $R = F_{k_2}(h_{P_2}(w'))$.

We would like to explain the design choices made in our construction. Because a PRF is a powerful primitive, it by itself is sufficient to produce the private string R indistinguishable from random. For example, setting $R \leftarrow F_{k_2}(w||r)$ for random r would satisfy the security game requirements. The reason for including the hash function h in the construction is to compress the biometric w without loosing the amount of its unpredictability. That is, the n-bit representation of biometric is normally substantially longer than the m bits of entropy it contains. For example, for iris the standard values of these parameters are $n = 2048$ and $m = 256$. Because $m \sim \kappa$, we can use a hash function $h : \{0,1\}^\kappa \times \{0,1\}^n \rightarrow \{0,1\}^m$ to reduce the size of w from n to m bits without loosing its entropy. In cases when the value of m exceeds the desired length of the input to a PRF, the hash function output length can be further reduced, i.e., in general $\ell_2(\kappa) \leq m$.

We note that the generic conversion of a secure sketch to a fuzzy extractor (in Section 2.1) uses a strong extractor, which can be built using a universal hash function alone. The use of the hash function in a strong extractor is, however, constrained in that the output length of the extractor must necessarily be smaller than m to be able to meet the requirement of the output being close to the uniform distribution. In particular, at least $2\log(\frac{1}{\varepsilon}) - 2$ bits of entropy are lost, where the parameter ε determines the statistical distance between distribution of the output and the uniform distribution. In our case, no requirements on the uniformity of the output must be met, and therefore no reduction of the output length or entropy loss has to take place.

Theorem 2. *Assuming that F is a family of PRFs and h is a universal hash function, the above fuzzy extractor scheme achieves strong biometric privacy.*

We would like to note that certain constructions of PRFs are known to produce uniformly distributed sequences. For example, (Shparlinski, 2001) shows that PRF in (Naor and Reingold, 1997) has this property for almost all values of parameters. For us this means that the adversary does not obtain advantage in distinguishing pseudo-random strings from random.

We also note that similar results can be achieved by using encryption instead of PRF, and such schemes might be known or used in industry.

5 RELATED WORK

The overall literature on fuzzy sketches and extractors is extensive, and we therefore highlight the most fundamental results and analysis related to this work. (Davida et al., 1998) proposed the first off-line biometric identification scheme, where error-correcting codes were used to reconstruct a biometric from its noisy readings. (Juels and Wattenberg, 1999) developed a fuzzy commitment scheme, which became the basis of the code-offset secure sketch for the Hamming distance. (Juels and Sudan, 2002) proposed a fuzzy vault scheme. (Dodis et al., 2004; Dodis et al., 2008) formalized the notion of secure sketches and fuzzy extractors in their seminal work, which gave a generic conversion from a secure sketch to a fuzzy extractor and developed a number of other schemes.

(Boyen et al., 2005) introduced robust fuzzy extractors secure against active adversaries, where the reconstruction process fails if the sketch has been tampered with. (Dodis et al., 2006) continue that line of research and also study the keyed setting in the bounded storage model. The use of the key in our setting is fundamentally different from that work, where two parties share a long-term secret key and use it to generate a session key for data authentication. Our constructions can potentially be applied to a robust fuzzy extractor to improve reusability properties.

There are also publications that combine fuzzy extractors with passwords to improve their security properties such as (Ballard et al., 2008). This work offers a simpler and more flexible construction.

Security requirements for adequate use of fuzzy sketches and extractors in cryptographic applications have been developing over time. (Boyen, 2004) showed that a number of original constructions cannot be safely applied multiple times to the same biometric. That work developed improved constructions using certain error-correcting codes and permutation groups that satisfy the reusability requirements. Our security definitions for the strong adversary were influenced by that work. Compared to (Boyen, 2004), our solution leaks no information about the biometric data (while leakage is unavoidable in the setting of (Boyen, 2004)) and works for all distance metrics and all secure sketch schemes in the standard model (while Boyen's scheme is limited to special codes and

a particular metric in the random oracle model).

(Scheirer and Boult, 2007) proposed three classes of attacks on secure sketches and fuzzy vault in particular, one of which is equivalent to sketch reusability. It has been empirically evaluated in (Kholmatov and Yanikoglu, 2008) on the fuzzy vault scheme using 200 matching pairs of fuzzy vault sketches. The authors were able to unlock (i.e., reconstruct the polynomial) 118 out of 200 pairs within a short period of time. We note that this evaluation was performed on a specific set of parameters already knowing that two stored sketches are related. Our analysis, on the other hand, is more general and can be applied to a wide variety of parameters. It is also does not assume prior knowledge of related sketches, but rather helps to identify those records. (Poon and Miri, 2009) also describe collusion attacks on the fuzzy vault scheme assuming that the sketches are related. Finally, (Simoens et al., 2009) introduced the notions of indistinguishability and irreversibility for reusable sketches and showed weaknesses of code-offset and permutation groups constructions. We analyze other constructions with respect to the indistinguishability property. (Kelkboom, 2010) also analyzes certain schemes.

6 CONCLUSIONS

This work investigates the reusability properties of secure sketch and fuzzy extractor constructions. Through new analysis we show that, in addition to the schemes that have been previously shown to have security weaknesses, other existing schemes do not meet our security expectations. To mitigate the problem, we propose to use the computational setting. Maintenance of a single key for all uses of such schemes results in solutions with remarkable security and usability improvements which are not possible otherwise. In particular, our general construction works with any existing secure sketch and mitigates information leakage associated with biometrics in the standard model under generic hardness assumptions.

REFERENCES

Ballard, L., Kamara, S., Monrose, F., and Reiter, M. (2008). Towards practical biometric key generation with randomized biometric templates. In *ACM CCS*.

Blanton, M. and Hudelson, W. (2009). Biometric-based non-transferable anonymous credentials. In *ICICS*, pages 165–180.

Boyen, X. (2004). Reusable cryptographic fuzzy extractors. In *ACM CCS*, pages 82–91.

Boyen, X., Dodis, Y., Katz, J., Ostrovsky, R., and Smith, A. (2005). Secure remote authentication using biometric data. In *EUROCRYPT*, pages 147–163.

Clancy, T., Kiyavash, N., and Lin, D. (2003). Secure smartcard-based fingerprint authentication. In *ACM SIGMM Workshop on Biometrics Methods and Applications*, pages 45–52.

Davida, G., Frankel, Y., and Matt, B. (1998). On enabling secure applications through off-line biometric identification. In *IEEE Symposium on Security and Privacy*, pages 148–157.

Dodis, Y., Katz, J., Reyzin, L., and Smith, A. (2006). Robust fuzzy extractors and authenticated key agreement from close secrets. In *CRYPTO*, pages 232–250.

Dodis, Y., Ostrovsky, R., Reyzin, L., and Smith, A. (2008). Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM Journal of Computing*, 38(1):97–139.

Dodis, Y., Reyzin, L., and Smith, A. (2004). Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. In *EUROCRYPT*, pages 523–540.

Dodis, Y. and Smith, A. (2005). Correcting errors without leaking partial information. In *ACM STOC*, pages 654–663.

Juels, A. and Sudan, M. (2002). A fuzzy vault scheme. In *International Symposium on Information Theory*.

Juels, A. and Wattenberg, M. (1999). A fuzzy commitment scheme. In *ACM CCS*, pages 28–36.

Kelkboom, E. (2010). *On the performance of helper data template protection schemes*. PhD thesis, University of Twente.

Kholmatov, A. and Yanikoglu, B. (2008). Realization of correlation attack against the fuzzy vault scheme. In *Proceedings of SPIE*, volume 6819.

Naor, M. and Reingold, O. (1997). Number-theoretic constructions of efficient pseudo-random functions. In *IEEE FOCS*, pages 458–467.

Nisan, N. and Ta-Shma, A. (1999). Extracting randomness: A survey and new constructions. *Journal of Computer and System Sciences*, 58:148–173.

Pankanti, S., Prabhakar, S., and Jain, A. (2002). On the individuality of fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1010–1025.

Poon, H. and Miri, A. (2009). A collusion attack on the fuzzy vault scheme. *ISC International Journal of Information Security*, 1(1):27–34.

Scheirer, W. and Boult, T. (2007). Cracking fuzzy vaults and biometric encryption. In *IEEE Biometrics Symposium*, pages 1–6.

Shparlinski, I. (2001). On the uniformity of distribution of the Naor-Reingold pseudo-random function. *Finite Fields and Their Applications*, 7(2):318–326.

Simoens, K., Tuyls, P., and Preneel, B. (2009). Privacy weaknesses of biometric sketches. In *IEEE Symposium on Security and Privacy*, pages 188–203.

Smith, A. (2004). *Maintaining secrecy when information leakage is unavoidable*. PhD dissertation, MIT.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS