An Information-Theoretic Analysis of Revocability and Reusability in Secure Biometrics

Ye Wang
Boston University
Boston, MA.
yw@bu.edu

Shantanu Rane
Mitsubishi Electric Research Laboratories
Cambridge, MA.
rane@merl.com

Stark C. Draper
University of Wisconsin
Madison, WI
sdraper@ece.wisc.edu

Prakash Ishwar
Boston University
Boston, MA.
pi@bu.edu

***Abstract*—Secure biometric systems are designed to allow authentication without requiring a reference biometric sample to be stored in the clear at the access control device. Instead, a template extracted from the reference biometric is stored on the device. An enrolled user can be authenticated by the template combined with a legitimate test biometric. However, an attacker who infiltrates the device only discovers the template, which reveals little or no information about the true biometric. We present a general framework for secure biometric authentication systems, and then provide a comparative information-theoretic analysis of two related realizations: (1) fuzzy commitment, in which authentication is framed as a problem of correcting errors between the reference and test biometrics, and (2) secure sketches, in which authentication is framed as a Slepian-Wolf decoding problem. We derive the false reject rates, false accept rates and successful attack rates for both realizations. We also consider the information leaked about a user's biometric identity when the database of biometric templates is compromised. Finally, we analyze a scenario in which the same biometric has been used to generate templates for several access control devices, some of which have been compromised by an adversary. It is shown that, two-factor versions of fuzzy commitment and secure sketch not only allow revocability, but also provide resistance to attacks in which the adversary compromises several databases at the same time.**

***Index Terms*—Biometrics, Fuzzy Commitment, Slepian-Wolf Coding, Revocability, Reusability**

I. Introduction

Human biometric measurements are attractive tools for verifying a person's identity and for authentication in access control situations. Compared with conventional identifying documents, they are difficult to forge. Compared with passwords traditionally used for access control, they do not have to be remembered. However, biometrics also present some new challenges that are not encountered in traditional methods. A characteristic feature of all human biometric measurements, such as fingerprints, iris scans, face images, ECG measurements, etc., is that each is distinct. Every measurement of a biometric is slightly different from all others.

In authentication or identity verification systems, the issue of noise in biometric measurements is currently solved using pattern recognition. Specifically, a measurement of the biometric is taken at the time of enrollment and stored in a database of enrolled identities or on an access control device. At a later time, the person in question provides a "test" or a "probe" biometric for comparison with the stored enrollment biometric. Using sophisticated methods of pattern recognition, it is possible to determine whether the enrollment and probe biometrics are similar according to a predefined metric. For example, access is granted if the probe fingerprint feature vector is less than a specified threshold in hamming distance from some enrolled fingerprint feature vector on the access control device.

Unfortunately, the above method creates a privacy problem: An adversary who infiltrates the device gains access to the enrollment biometric. This is a serious problem for two reasons. Firstly, it is a security hazard; the attacker can now use the enrollment biometric to gain repeated access to the system and to any other biometric-based systems in which the user has been enrolled. Secondly, it is a privacy hazard; the attacker now has access to the user's identifying information and can therefore impersonate the user illegally. This second problem is made worse by the fact that, since biometrics are inherent properties of a user's body, the user cannot arbitrarily generate a new biometrics when an old enrolled biometric is compromised.

In response to the growing concerns about security hazards and identity theft, new methods of biometric authentication have been proposed, which will be surveyed briefly in the next section, and analyzed in detail in subsequent sections. Based on the concerns outlined above, the goals of these new methods are three-fold. First, the information to be stored at enrollment, often called the biometric template, should provide little or no information about the actual biometric. Second, the stored template should not allow an attacker to gain unauthorized access to the system or to verify the identity as if he is the legitimate user. Third, if the stored template is known to have been compromised, then it should be possible to revoke this template and issue a new template for the user that successfully prevents the adversary from gaining access or stealing the user's identity in the future.

II. Prior Art and Our Contribution

To address the security and privacy vulnerabilities of biometric systems, secure biometric schemes have been proposed. These fall under two main categories, viz., fuzzy commitment and secure sketche schemes.

In fuzzy commitment, a random vector is combined with the user's enrollment biometric via a commitment function. The output of the commitment function is stored at the access control or identity verification station as helper data.

Authentication or identity verification is accomplished by means of a decommitment function, which takes as its inputs the stored data and the user's probe biometric and recovers the random vector. To verify whether the random vector has been recovered exactly, its cryptographic hash is also stored at the access control device. This stored hash must match the hash of the recovered vector for access to be granted. Fuzzy commitment schemes can be efficiently constructed using error correcting codes (ECC), and indeed, we use the ECC-based fuzzy commitment scheme in our theoretical analysis. There is a rich literature on the principles and methods of fuzzy commitment, especially [1], [2], [3], [4] and an equivalent framework called a fuzzy extractor [5].

In secure sketch-based schemes, the user provides their biometric at enrollment, from which a signal called a "sketch" is derived and stored on the access control device. By itself the sketch provides very little information about the enrollment biometric. However, when combined with a probe biometric from the legitimate user, the enrollment biometric can be recovered. From an information theoretic point of view, this is equivalent to Slepian-Wolf decoding [6]: the helper data is used to decode the enrollment biometric using the probe biometric as side information. As in fuzzy commitment, a cryptographic hash of the enrollment biometric is used to verify that the recovery was successful. Principles of secure sketch-based biometric schemes and various methods of implementing such schemes have been reported in [7], [8], [9], [10].

It should be clear from the preceding discussion that both fuzzy commitment and secure sketche schemes enable biometric authentication by utilizing common randomness available at enrollment and authentication. Indeed, many works, notably [11], [5], [12], have referred to the conceptual equivalence of the two schemes. In most previous work, the equivalence has been drawn by viewing both biometric systems as encoder-decoder pairs (codecs) with a certain achievable tradeoff between secret key rate and the probability of accurate authentication. This methodology of analyzing the key rate versus equivocation rate tradeoff has also been adopted in recent information-theoretic studies in secure biometrics conducted in [13], [14].

In this paper, we take a different approach. We present a generalized secure biometrics framework for which, in addition to the usual metrics such as false accept rate (FAR) and false reject rate (FRR), we emphasize the successful attack rate (SAR) and the information leaked about the user's biometric when the system is compromised. Starting from the generalized framework, we derive a fuzzy commitment-based scheme and a secure sketch-based scheme based on error correcting codes. We present an information theoretic analysis of each of these schemes, comparing the information leakage and error exponents. In particular, for ECC-based implementations, we show that, for a given error correcting code, the two realizations are identical in terms of error exponents and information leakage. Additionally, for both fuzzy commitment and secure sketch, we explicitly consider both keyless and two-factor schemes (biometric in conjunction with a smart card) and analyze their revocability and reusability, especially in cases where more than one biometric system used by an individual is compromised. It is shown that the two-factor variants of fuzzy commitment and secure sketch are resistant to such linkage attacks.

The remainder of this paper is organized as follows: We set up a generalized framework for secure biometrics and mathematically define the design objectives in section III. Using this framework, fuzzy commitment-based and secure sketch-based secure biometric schemes are described in Section IV and Section V respectively. For each of these schemes, considering the keyless and two-factor variants, we determine the FAR, FRR and SAR in Section VI. Section VII contains an analysis of the information leakage under various kinds of attack and and examines the revocability of the biometric template for both fuzzy commitments and secure sketches. Section VIII concludes the paper.

III. A Generalized Secure Biometrics Framework

We now describe a generalized framework within which it is possible to analyze secure biometrics systems. In particular, we present an abstract model of a secure biometric system, we enumerate system design objectives, and we characterize these objectives using information theory.

A. Model of a Secure Biometric System

Consider the generalized secure biometric system in Fig. 1 which consists of encoding and decoding modules that manipulate feature vectors extracted from human biometric traits. In the treatment below, all feature vectors and secret keys consist of binary elements. The generalization to higher alphabets is straightforward.

Feature Vectors: At enrollment, the user provides a biometric measurement, from which is extracted an enrollment feature vector $\mathbf{A} := (A_1, \ldots, A_n)$. This vector is used to generate the secure template that is stored on the access control device. For authentication, the user provides a biometric measurement, from which is extracted a probe feature vector $\mathbf{B} := (B_1, \ldots, B_n)$ using the same feature extraction algorithm as before. This vector is used by the decoding module of the access control device to verify the user's identity. In general, bits extracted from biometric measurements are neither independent nor identically distributed. However, it is possible to design feature transformation algorithms that convert biometric readings into vectors of i.i.d. Bernoulli-0.5 bits [10]. However, due to measurement noise, each biometric reading and binary conversion may produce bit errors. Therefore, we assume that A_i, B_i are i.i.d. samples of a doubly symmetric binary source with crossover probability $p < 0.5$.

Enrollment: The (potentially randomized) encoding function $F(\cdot)$ takes the enrollment feature vector $\mathbf{A}$ as input and produces as outputs $\mathbf{S}$, which is stored on the access control device, and (optionally) a key vector $\mathbf{K}$, which is returned to the user. Thus, $(\mathbf{S}, \mathbf{K}) = F(\mathbf{A})$. The encoding function is governed by the conditional distribution $P_{\mathbf{S},\mathbf{K}|\mathbf{A}}$. Depending

upon the physical realization of the system, the user may be required to carry the key $\mathbf{K}$ on a smart card. These systems are called *two-factor* systems as both the key and the stored data are needed for authentication. Systems where $\mathbf{K}$ is null are called *keyless* systems and do not require the use of a smart card.

Authentication: To perform biometric authentication, a legitimate user provides the probe feature vector $\mathbf{B}$ and the key $\mathbf{K}$. An adversary, on the other hand, provides a stolen or artificially synthesized biometric feature vector $\mathbf{C}$ and a stolen or artificially synthesized key $\mathbf{J}$. The presence of the legitimate user or the adversary is indicated by the unknown binary parameter θ. Let $(\mathbf{D}, \mathbf{L})$ denote the feature vector, key pair that is provided during the authentication step, that is,

$$(\mathbf{D}, \mathbf{L}) := \begin{cases} (\mathbf{B}, \mathbf{K}), & \text{if } \theta = 1, \\ (\mathbf{C}, \mathbf{J}), & \text{if } \theta = 0. \end{cases}$$

The authentication decision is computed by the decoding function as $\hat{\theta} = g(\mathbf{D}, \mathbf{L}, \mathbf{S})$. In keyless systems, the procedure is similar with $\mathbf{K}$, $\mathbf{J}$, and $\mathbf{L}$ removed from the above description.

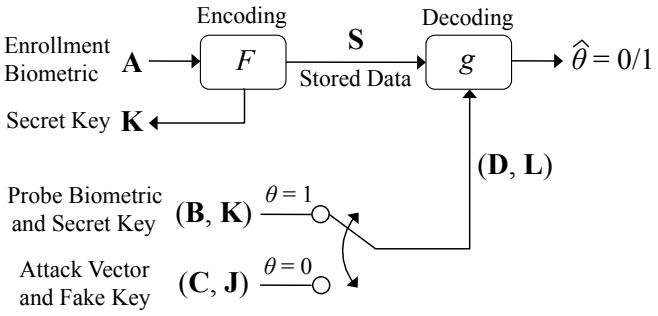

Fig. 1. Generalized framework for secure biometrics. This framework encompasses both fuzzy commitment-based and secure sketch-based realizations. For keyless realizations, $\mathbf{K}$ is null. For two-factor realizations, $\mathbf{K}$ is a secret key output by the randomized encoding function. Given the probe biometric and, in two-factor realizations, a secret key, the decoder solves a hypothesis testing problem.

B. System Design Objectives

The objective is to design a biometric authentication system through a choice of $F(\cdot)$ and $g(\cdot)$ that

- maximizes the authentication (verification) accuracy, and
- minimizes the data storage and key length requirements for a given accuracy.

However, the problem is complicated by the possibility that the enrollment biometric feature vector $\mathbf{A}$, the stored data $\mathbf{S}$, the key $\mathbf{K}$, or any combination thereof could be compromised by an adversary. An adversary with access to this information could not only potentially undermine the authentication integrity of the system, but may also be able to extract information about the underlying biometric via the feature vector $\mathbf{A}$. The system should be robust in these scenarios. Specifically, the system design should additionally aim to:

- minimize information leakage from compromised data,
- preserve authentication integrity after data exposure,
- enable revocability of compromised enrollments, and
- resist linkage attacks, that is, satisfy the above requirements even when multiple parallel systems are compromised.

Next, we quantify the above design objectives. The data storage requirement is given by the number of bits need to represent $\mathbf{S}$, which is $\log_2 |\mathcal{S}|$ bits. The key length requirement is given by the number of bits need to represent $\mathbf{K}$ which is $\log_2 |\mathcal{K}|$ bits. The authentication accuracy is measured by the achievable tradeoff between the probability of missed detection and probability of false detection. The probability of missed detection, also called the false reject rate (FRR), is given by

$$P_m = \Pr\left[\hat{\theta} = 0 | \theta = 1\right],$$

which depends only on the known and fixed statistics of $(\mathbf{A}, \mathbf{B}, \mathbf{K})$ and the specification of the system, $F(\cdot)$ and $g(\cdot)$. Since the aim of the adversary is to gain unauthorized access by defeating the system, the *baseline* probability of false detection, also called the false accept rate (FAR) is taken to be the worst case probability of false detection across all attack vector and key distributions. This is given by

$$P_f := \max_{p_{\mathbf{C},\mathbf{J}}} \Pr\left[\hat{\theta} = 1 | \theta = 0\right],$$

such that $(\mathbf{C}, \mathbf{J})$ are independent of $(\mathbf{A}, \mathbf{B}, \mathbf{K}, \mathbf{S})$. This definition is very general; if cryptographic functions are used in the biometric system, then additional assumptions on the adversary (e.g., computationally bounded) can be captured by limiting the class of admissible distributions to be maximized over.

To measure the information leaked about the enrollment biometric feature vector $\mathbf{A}$ in the various scenarios of data exposure, i.e., when either the stored data $\mathbf{S}$, the secret key $\mathbf{K}$, or both are compromised, we use the mutual informations $I(\mathbf{A}; \mathbf{S})$, $I(\mathbf{A}; \mathbf{K})$, and $I(\mathbf{A}; \mathbf{S}, \mathbf{K})$. These measures quantify information leakage in an information theoretic sense. Capturing the notion of information leakage for a computationally bounded adversary would require a different measure.

In the event of data exposure, the probability of false detection could increase. We also consider the possible scenarios where an adversary has access to $\mathbf{A}$ as well. For $\mathbf{V}$ equal some combination of $\mathbf{A}$, $\mathbf{S}$, and $\mathbf{K}$, the probability of false detection against an adversary with access to $\mathbf{V}$ is computed as

$$P_a(\mathbf{V}) := \max_{p_{\mathbf{C},\mathbf{J}|\mathbf{V}}} \Pr\left[\hat{\theta} = 1 | \theta = 0\right],$$

capturing that the $\mathbf{C}, \mathbf{J}$ may be generated from $\mathbf{V}$. To distinguish $P_a(\mathbf{V})$ from the FAR, it will be referred to as the successful attack rate (SAR). The SAR captures the probability of false detection when an adversary is *enhanced* with knowledge of $\mathbf{V}$. Note that in any keyless or two-factor system, knowledge of the stored data $\mathbf{S}$ drastically improves the ability of the adversary to gain access. We characterize this later in Theorem 4.

Ideally, in the two-factor systems, each factor, the feature vector $\mathbf{A}$, and the key $\mathbf{K}$, should be useless on its own, i.e.,

with respect to the ability of an adversary to gain access. This motivates the following definition. We say that a system is *two-factor secure* if $P_a(\mathbf{A}) = P_a(\mathbf{K}) = P_f$.

Quantifying revocability and robustness to linkage attacks requires analysis of system performance in the context of multiple, parallel enrollments. Let $(\mathbf{S}_1, \mathbf{K}_1), (\mathbf{S}_2, \mathbf{K}_2), \ldots, (\mathbf{S}_u, \mathbf{K}_u)$ denote u conditionally independent enrollments for a given biometric feature vector $\mathbf{A}$, that is,

$$P_{\mathbf{S}^u, \mathbf{K}^u|\mathbf{A}}(s^u, k^u|a) = \prod_{i=1}^{u} P_{\mathbf{S},\mathbf{K}|\mathbf{A}}(s_i, k_i|a).$$

Note that for a system corresponding to one of these enrollments, the FRR,

$$P_m(i) := \Pr\big[g(\mathbf{B}, \mathbf{S}_i, \mathbf{K}_i) = 0\big],$$

and FAR,

$$P_f(i) := \max_{p_{\mathbf{C},\mathbf{J}}} \Pr\big[g(\mathbf{C}, \mathbf{S}_i, \mathbf{J}) = 1\big],$$

are unchanged, that is, $P_m(i) = P_m$ and $P_f(i) = P_f$, since the presence of parallel enrollments does not change the marginal statistics. However, the presence of parallel enrollments necessitates the generalization of the SAR definition to model the possibility of an adversary attacking a particular enrollment while having compromised that enrollment and/or other enrollments. For a random variable $\mathbf{V}$ that is a subset of $\{\mathbf{A}, \mathbf{S}_1, \ldots, \mathbf{S}_u, \mathbf{K}_1, \ldots, \mathbf{K}_u\}$, the SAR against system i by an adversary that is enhanced with knowledge of $\mathbf{V}$ is given by

$$P_a(i, \mathbf{V}) = \max_{p_{\mathbf{C},\mathbf{J}|\mathbf{V}}} \Pr\big[g(\mathbf{C}, \mathbf{S}_i, \mathbf{J}) = 1\big].$$

When multiple parallel enrollments are compromised, the natural extension for measuring the information leakage is $I(\mathbf{A}; \mathbf{V})$, where $\mathbf{V}$ is a subset of $\{\mathbf{A}, \mathbf{S}_1, \ldots, \mathbf{S}_u, \mathbf{K}_1, \ldots, \mathbf{K}_u\}$. For clarity, we may also write $\mathbf{V}$ as equal to $\{\mathbf{V}_1, \ldots, \mathbf{V}_u\}$, where $\mathbf{V}_i$ is either $\mathbf{S}_i$, $\mathbf{K}_i$, $(\mathbf{S}_i, \mathbf{K}_i)$ or null, representing the compromised data at each enrollment.

For conditionally independent parallel enrollments of any keyless or two-factor system, the total information leakage is less than the sum of the information leakage from each compromised enrollment, that is, for any $\{\mathbf{V}_1, \ldots, \mathbf{V}_u\}$,

$$I(\mathbf{A}; \mathbf{V}_1, \ldots, \mathbf{V}_u) \leq \sum_{i=1}^{u} I(\mathbf{A}; \mathbf{V}_i).$$

This property is a consequence of Lemma 1 and captures the notion that compromises across multiple enrollments does not amplify information leakage beyond an additive sense. The following definition requires the stronger property that the information leaked when multiple enrollments are compromised be no more than the most information leaked by any individual compromised enrollment.

Definition 1 *A system is* Resistant to Linking Attacks *if for any* $\{\mathbf{V}_1, \ldots, \mathbf{V}_u\}$*, where* $\mathbf{V}_i$ *is either* $\mathbf{S}_i$*,* $\mathbf{K}_i$*,* $(\mathbf{S}_i, \mathbf{K}_i)$ *or null, the following inequality holds,*

$$I(\mathbf{A}; \mathbf{V}_1, \ldots, \mathbf{V}_u) = \max_{i \in \{1,\ldots,u\}} I(\mathbf{A}; \mathbf{V}_i).$$

Note that this definition implies that the information leakage across multiple compromised enrollments meets the lower bound. A consequence of this property is that, if an adversary's sole objective is to maximize the information gathered about $\mathbf{A}$, then once the adversary has fully compromised one system nothing is gained by compromising others.

Since compromising the stored information $\mathbf{S}$ implies a drastic increase in the ability of an adversary to gain access, a system should allow for a graceful recovery from such an exposure. A revocable system should allow parallel enrollments to maintain authentication integrity even when other parallel enrollments have been compromised. The following definition captures this property for various patterns of exposure.

Definition 2 *A system offers* Revocability against Multiple Exposures *if for any disjoint subsets* $M, N \subset \{1, \ldots, u\}$*,* $I(\mathbf{A}; \{\mathbf{S}_i\}_{i\in M}, \{\mathbf{K}_j\}_{j\in N}) = 0$ *and for* $k \notin M$*,* $P_a(k, (\{\mathbf{S}_i\}_{i\in M}, \{\mathbf{K}_j\}_{j\in N})) = P_f$*, and also for any subsets* $M', N' \subset \{1, \ldots, u\}$ *and* $k \notin M' \cup N'$*,* $P_a(k, (\mathbf{A}, \{\mathbf{S}_i\}_{i\in M'}, \{\mathbf{K}_j\}_{j\in N'})) = P_f$*.*

For a system satisfying this revocability definition, an adversary that compromises multiple enrollments, but has only either the key or stored data of each enrollment, does not gain any information about $\mathbf{A}$ nor can improve his ability to gain access to a system corresponding to an enrollment where he only has the key. Thus, if it is known that either the stored data or key of particular enrollment has been compromised, the other unexposed value can be destroyed to nullify that enrollment, while other parallel enrollments and new enrollments remain unaffected. As another consequence of the definition, even if the feature vector $\mathbf{A}$ and several enrollments have been arbitrarily compromised (i.e. either the key, stored data, or both have been exposed), the adversary still cannot improve his ability to gain access to system where neither the key nor the stored data has been exposed. Note that this property also implies two-factor security.

IV. Fuzzy Commitment Systems based on ECC

As outlined in Section II, a fuzzy commitment scheme binds a random vector to the user's enrollment biometric $\mathbf{A}$ to produce the length-n stored data vector $\mathbf{S}$, cf. Figure 1. Subsequently we exclusively consider fuzzy commitment schemes wherein the random vector corresponds to a uniformly selected codeword of a binary $[n, k]$ linear error correcting code. We use $\mathbf{G}$ to denote the code's $k \times n$ generator matrix and $\mathbf{H}$ to denote the code's $m \times n$ parity check matrix with $m = (n-k)$. The stored data is $\mathbf{S} = \mathbf{A} \oplus \mathbf{G}^{\mathrm{T}}\mathbf{Z}$ where $\mathbf{Z}$ is an i.i.d. Bernoulli(0.5) random sequence and $\oplus$ is the binary XOR operation.

At authentication the system has $\mathbf{S}$ and the pair $(\mathbf{D}, \mathbf{L})$. When the legitimate user is trying to gain access ($\theta = 1$ in Figure 1), $\mathbf{D} = \mathbf{B}$. A legitimate probe vector $\mathbf{B}$ is related

to the enrollment $\mathbf{A}$ as $\mathbf{B} = \mathbf{A} \oplus \mathbf{N}$ where $\mathbf{N}$ is a length-n i.i.d. Bernoulli(p) sequence with $\Pr[N_i = 1] = p$ for all i. We present two variants of our system, a keyless variant (where there is no $\mathbf{L}$ or, formally, $\mathbf{L}$ is some constant) and a two-factor variant which enables revocability.

A. Keyless System

Enrollment: The enrollment procedure first generates an independent i.i.d. Bernoulli(0.5) sequence $\mathbf{Z} := (Z_1, \ldots, Z_k)$. The stored data is then computed as

$$\mathbf{S} = \mathbf{A} \oplus \mathbf{G}^{\mathrm{T}}\mathbf{Z},$$

and the smartcard key $\mathbf{K}$ is null.

Authentication: The authentication procedure first performs syndrome decoding to recover

$$\hat{\mathbf{W}} = \underset{\mathbf{W}:\mathbf{HW}=\mathbf{H}(\mathbf{D}\oplus\mathbf{S})}{\arg\min} d(\mathbf{W}),$$

where $d(\cdot)$ is the Hamming weight. This is operationally equivalent to the optimal channel decoding of $\mathbf{G}^{\mathrm{T}}\mathbf{Z}$ corrupted by $\mathbf{A} \oplus \mathbf{D}$, where $\hat{\mathbf{W}}$ is the corresponding optimal recovery of $\mathbf{A} \oplus \mathbf{D}$. The authentication decision is made via the following threshold test,

$$\hat{\theta} = d(\hat{\mathbf{W}}) \underset{0}{\overset{1}{\lessgtr}} \tau n,$$

which accepts or rejects based on the closeness of the probe biometric feature vector to the enrollment biometric feature vector.

Note that, unlike in the introduction to fuzzy commitment schemes in Section II, we do not store at enrollment, nor check at authentication, a cryptographic hash of the random vector ($\mathbf{Z}$ in our system). This is because our focus is on information theoretic security and the security of a cryptographic hash is only computational. Hence, a cryptographic hash cannot be used as part of an information theoretically secure system and we instead rely on the threshold test described above.

B. Two-Factor System

Enrollment: The enrollment is similar to the keyless system, except that additionally, the key $\mathbf{K} := (K_1, \ldots, K_n)$ is generated as an independent i.i.d. Bernoulli(0.5) sequence. The stored data is masked by the key,

$$\mathbf{S} = \mathbf{A} \oplus \mathbf{G}^{\mathrm{T}}\mathbf{Z} \oplus \mathbf{K}.$$

Authentication: Likewise, the authentication procedure first performs the optimal recovery of $\mathbf{A} \oplus \mathbf{D}$ via

$$\hat{\mathbf{W}} = \underset{\mathbf{W}:\mathbf{HW}=\mathbf{H}(\mathbf{D}\oplus\mathbf{S}\oplus\mathbf{L})}{\arg\min} d(\mathbf{W}),$$

and accepts or rejects based on

$$\hat{\theta} = d(\hat{\mathbf{W}}) \underset{0}{\overset{1}{\lessgtr}} \tau n.$$

V. Secure Sketch Systems based on ECC

We now introduce the family of secure sketch systems studied in this paper. While, as was the case for fuzzy commitment, there are other ways to develop a secure sketch, we concentrate on secure sketch systems based on error correcting codes. Again we use linear error correcting codes and, mimicking the notation of Section IV, we denote by $\mathbf{H}$ the $m \times n$ parity check matrix of a binary $[n, k]$ linear error correcting codes ($m = n - k$). In our secure sketch systems the stored data $\mathbf{S}$ will be the length-m syndrome of the enrollment biometric $\mathbf{A}$ calculated as $\mathbf{S} = \mathbf{HA}$. The syndrome indexes a coset of possible enrollment biometrics and the challenge at authentication is to identify the actual enrollment based on $\mathbf{S}$ and the pair $(\mathbf{D}, \mathbf{L})$. The relationship between $\mathbf{D}$ and $\mathbf{A}$ is exactly as in the fuzzy commitment model presented Section IV. Again we study both keyless and two-factor variants and, for the reasons already discussed, we do not employ a cryptographic hash.

A. Keyless System

Enrollment: The enrollment procedure stores the syndrome of enrollment biometric feature vector $\mathbf{A}$,

$$\mathbf{S} = F(\mathbf{A}) = \mathbf{HA},$$

and the smartcard key $\mathbf{K}$ is null.

Authentication: The authentication procedure first performs syndrome decoding to recover

$$\hat{\mathbf{W}} = \underset{\mathbf{W}:\mathbf{HW}=\mathbf{HD}\oplus\mathbf{S}}{\arg\min} d(\mathbf{W}),$$

which is the optimal recovery of $\mathbf{A} \oplus \mathbf{D}$. The authentication decision is made via the following threshold test,

$$\hat{\theta} = d(\hat{\mathbf{W}}) \underset{0}{\overset{1}{\lessgtr}} \tau n,$$

which accepts or rejects based on the closeness of the probe biometric feature vector to the enrollment biometric feature vector.

B. Two-Factor System

Enrollment: The enrollment is similar to the keyless system, except that additionally, the key $\mathbf{K} := (K_1, \ldots, K_m)$ is generated as an independent i.i.d. Bernoulli(0.5) sequence. The stored data is masked by the key,

$$\mathbf{S} = \mathbf{HA} \oplus \mathbf{K}.$$

Authentication: Likewise, the authentication procedure first performs the optimal recovery of $\mathbf{A} \oplus \mathbf{D}$ via

$$\hat{\mathbf{W}} = \underset{\mathbf{W}:\mathbf{HW}=\mathbf{HD}\oplus\mathbf{S}\oplus\mathbf{L}}{\arg\min} d(\mathbf{W})$$

and accepts or rejects based on

$$\hat{\theta} = d(\hat{\mathbf{W}}) \underset{0}{\overset{1}{\lessgtr}} \tau n.$$

VI. Authentication Accuracy Analysis

In this section, we analyze the FRR, FAR, and SAR of all of our systems. We will first analyze the FRR P_m and the FAR P_f for the keyless secure sketch system. We will

then argue that these quantities are the same for the two-factor secure sketch system and both variants of the fuzzy commitment system. Then, for each system and the various scenarios of data exposure, we will show that the SAR P_a is equal to either P_f or one.

A. Notation and Assumptions

In this section, we will use the binary entropy function defined by

$$h_b(p) = -p\log_2 p - (1-p)\log_2(1-p)$$

and the binary relative entropy (K-L divergence) function defined by

$$D(q\|p) = q\log_2\frac{q}{p} + (1-q)\log_2\frac{1-q}{1-p}$$

In the following analysis we make the following assumptions on the operating parameters of our systems. The threshold ratio is larger than p but smaller than 0.5, that is.

$$0.5 > \tau > p.$$

The coding rate of the error correcting code is within the channel capacity of a binary symmetric channel (BSC) with crossover probability τ (and hence also within the channel capacity of BSC with crossover probability p),

$$R = k/n < 1 - h_b(\tau) < 1 - h_b(p),$$

or equivalently

$$m/n > h_b(\tau) > h_b(p).$$

Since the channel codes are operating under capacity, we will assume that they have a positive error exponent $E(R) > 0$ and that the probability of decoding error when using these codes on a BSC with crossover probability p is bounded by

$$P_e \leq 2^{-nE(R)+o(n)}.$$

It is well known that there exist code constructions that support these assumptions [15]. The parity check matrix $\mathbf{H}$ is fixed and full-rank. Hence $\mathbf{HA}$ is i.i.d. Bernoulli(0.5).

B. Keyless Secure Sketch FRR Analysis

In this subsection, we bound the FRR and FAR of the keyless secure sketch systems.

Theorem 1 *For the Keyless Secure Sketch System, the FRR is bounded by*

$$P_m \leq 2^{-nD(\tau\|p)} + 2^{-nE(R)+o(n)},$$

where $E(R) > 0$ is the error exponent of the code used in the system.

Proof: The FRR is given by

$$P_m = \Pr\left[d(\hat{\mathbf{W}}) > \tau n\right],$$

where, since for the legitimate user $\mathbf{D} = \mathbf{B}$ and $\mathbf{L} = \mathbf{K}$,

$$\hat{\mathbf{W}} = \underset{\mathbf{W}:\mathbf{HW}=\mathbf{H}(\mathbf{A}\oplus\mathbf{B})}{\arg\min} d(\mathbf{W}).$$

The FRR can be bounded by

$$\begin{aligned} P_m &= \Pr\left[d(\hat{\mathbf{W}}) > \tau n, \hat{\mathbf{W}} = \mathbf{A}\oplus\mathbf{B}\right] \\ &\quad + \Pr\left[d(\hat{\mathbf{W}}) > \tau n, \hat{\mathbf{W}} \neq \mathbf{A}\oplus\mathbf{B}\right] \\ &\leq \Pr\left[d(\mathbf{A}\oplus\mathbf{B}) > \tau n\right] + \Pr\left[\hat{\mathbf{W}} \neq \mathbf{A}\oplus\mathbf{B}\right]. \end{aligned}$$

The decoding procedure to produce $\hat{\mathbf{W}}$ is operationally equivalent to the optimal syndrome decoding of $\mathbf{A}$ from the noisy version $\mathbf{B}$, since

$$\begin{aligned} \hat{\mathbf{W}} &= \underset{\mathbf{W}:\mathbf{HW}=\mathbf{H}(\mathbf{A}\oplus\mathbf{B})}{\arg\min} d(\mathbf{W}) \\ &= \mathbf{B} \oplus \underset{\mathbf{A}':\mathbf{HA}'=\mathbf{HA}}{\arg\min} d(\mathbf{A}'\oplus\mathbf{B}). \end{aligned}$$

Thus, the probability that $\hat{\mathbf{W}}$ fails to recover $\mathbf{A}\oplus\mathbf{B}$ is equal to the probability of error of the code, which is bounded by

$$\Pr\left[\hat{\mathbf{W}} \neq \mathbf{A}\oplus\mathbf{B}\right] \leq 2^{-nE(R)+o(n)}.$$

The probability that $\mathbf{A}\oplus\mathbf{B}$ fails the threshold test can be bounded by the Chernoff-Hoeffding bound [16],

$$\Pr\left[d(\mathbf{A}\oplus\mathbf{B}) > \tau n\right] \leq 2^{-nD(\tau\|p)}.$$

Combining these two bounds yields the theorem. ■

Theorem 2 *For the Keyless Secure Sketch System, the FAR is bounded by*

$$P_f \leq 2^{-n\left(\frac{m}{n} - h_b(\tau)\right)}.$$

Proof: The FAR is given by

$$\begin{aligned} P_f &= \max_{P_{\mathbf{C}}} \Pr\left[\exists \mathbf{w} : d(\mathbf{w}) \leq \tau n, \mathbf{Hw} = \mathbf{H}(\mathbf{C}\oplus\mathbf{A})\right] \\ &= \Pr\left[\exists \mathbf{w} : d(\mathbf{w}) \leq \tau n, \mathbf{Hw} = \mathbf{HA}\right], \end{aligned}$$

since $\mathbf{A}$ is i.i.d. Bernoulli(0.5) and independent of $\mathbf{C}$, thus removing the effect of $P_{\mathbf{C}}$. To bound this expression, we first use [17, Lemma 8, Ch. 10] to bound the number of sequences $\mathbf{w}$ in $\{0,1\}^n$ with Hamming weight less than τn,

$$\begin{aligned} |\{\mathbf{w} : d(\mathbf{w}) \leq \tau n\}| &= \sum_{i=0}^{\tau n} |\{\mathbf{w} : d(\mathbf{w}) = i\}| \\ &= \sum_{i=0}^{\tau n} \binom{n}{i} \\ &\leq 2^{nh_b(\tau)}. \end{aligned}$$

Secondly, since $\mathbf{A}$ is i.i.d. Bernoulli(0.5) and $\mathbf{H}$ is full rank, $\mathbf{HA}$ is also i.i.d. Bernoulli(0.5). Thus, for any given sequence $\mathbf{w}$ in $\{0,1\}^n$,

$$\Pr\left[\mathbf{Hw} = \mathbf{HA}\right] = 2^{-m}.$$

Combining these results, we can bound FAR by

$$\begin{aligned} P_f &= \Pr\left[\exists \mathbf{w} : d(\mathbf{w}) \leq \tau n, \mathbf{H}\mathbf{w} = \mathbf{H}\mathbf{A}\right] \\ &= \Pr\left[\bigcup_{\mathbf{w}:d(\mathbf{w})\leq \tau n} \mathbf{H}\mathbf{w} = \mathbf{H}\mathbf{A}\right] \\ &\leq \sum_{\mathbf{w}:d(\mathbf{w})\leq \tau n} \Pr\left[\mathbf{H}\mathbf{w} = \mathbf{H}\mathbf{A}\right] \\ &\leq 2^{nh_b(\tau)-m} = 2^{-n\left(\frac{m}{n}-h_b(\tau)\right)}. \end{aligned}$$

■

Theorems 1 and 2 provide exponentially decaying upper bounds on the FRR and FAR, and hence also lower bounds on the exponents. In order to obtain these exponentially decaying bounds, the operating parameters must satisfy the previously listed assumptions, that is, $0.5 > \tau > p$ and $m/n > h_b(\tau)$.

C. FRR and FAR Analysis of the Other Systems

The decoding procedures of all four system variants are nearly identical. The authentication decision is determined by whether or not $\hat{\mathbf{W}}$, the lowest Hamming weight sequence in a given coset, has Hamming weight less than the threshold τn. The coset is specified by its corresponding syndrome, and the only difference between each variant is how this syndrome is computed as a function $q(\cdot)$ of $\mathbf{S}$, $\mathbf{D}$, and (for two-factor systems) $\mathbf{L}$. In the keyless secure sketch system, the syndrome is

$$\begin{aligned} q(\mathbf{S}, \mathbf{D}) &= \mathbf{H}\mathbf{D} \oplus \mathbf{S} \\ &= \mathbf{H}(\mathbf{A} \oplus \mathbf{D}). \end{aligned}$$

In the two-factor secure sketch system, the syndrome is

$$\begin{aligned} q(\mathbf{S}, \mathbf{D}, \mathbf{L}) &= \mathbf{H}\mathbf{D} \oplus \mathbf{S} \oplus \mathbf{L} \\ &= \mathbf{H}(\mathbf{A} \oplus \mathbf{D}) \oplus \mathbf{K} \oplus \mathbf{L}. \end{aligned}$$

In the keyless fuzzy commitment system, the syndrome is

$$\begin{aligned} q(\mathbf{S}, \mathbf{D}) &= \mathbf{H}(\mathbf{D} \oplus \mathbf{S}) \\ &= \mathbf{H}(\mathbf{A} \oplus \mathbf{D}) \oplus \mathbf{H}\mathbf{G}^{\mathrm{T}}\mathbf{Z} \\ &= \mathbf{H}(\mathbf{A} \oplus \mathbf{D}). \end{aligned}$$

In the two-factor fuzzy commitment system, the syndrome is

$$\begin{aligned} q(\mathbf{S}, \mathbf{D}, \mathbf{L}) &= \mathbf{H}(\mathbf{D} \oplus \mathbf{S} \oplus \mathbf{L}) \\ &= \mathbf{H}(\mathbf{A} \oplus \mathbf{D}) \oplus \mathbf{H}\mathbf{G}^{\mathrm{T}}\mathbf{Z} \oplus \mathbf{H}(\mathbf{K} \oplus \mathbf{L}) \\ &= \mathbf{H}(\mathbf{A} \oplus \mathbf{D}) \oplus \mathbf{H}(\mathbf{K} \oplus \mathbf{L}). \end{aligned}$$

The authentication decision $\hat{\theta}$ is only a function of the syndrome. The syndrome is only a function of $\mathbf{H}(\mathbf{A} \oplus \mathbf{D})$, and (for two-factor systems) $(\mathbf{K}, \mathbf{L})$.

In all four systems, during the authentication of the legitimate user, where $\mathbf{D} = \mathbf{B}$ and $\mathbf{L} = \mathbf{K}$, the syndrome computed is identical and equal to $\mathbf{H}(\mathbf{A} \oplus \mathbf{B})$. Thus, the distribution of $\hat{\theta}$ given $\theta = 1$ is identical across all of the systems, and hence the FRR performance must be the same.

In computing the FAR, the case of an attack by the adversary, the input vectors $(\mathbf{D}, \mathbf{L}) = (\mathbf{C}, \mathbf{J})$ can have an arbitrary distribution, but must be independent from $(\mathbf{A}, \mathbf{K})$. Thus, regardless of the distribution on $(\mathbf{C}, \mathbf{J})$, the syndrome in any of the four systems is i.i.d. Bernoulli(0.5), since $\mathbf{A}$ and $\mathbf{K}$ are i.i.d. Bernoulli(0.5). Since the syndromes are equal in distribution, the authentication decisions $\hat{\theta}$ are also equal in distribution across systems, and hence the FAR performance must be the same.

The conclusions of this subsection are summarized in the following theorem. We can also apply the bounds on the FRR and FAR of the keyless secure sketch system to all four systems.

Theorem 3 *The FRR and FAR is the same for all four systems, namely, the keyless secure sketch system, the two-factor secure sketch system, the keyless fuzzy commitment system, and the two-factor fuzzy commitment system.*

D. Successful Attack Rate Analysis

In any keyless or two-factor system, knowledge of the stored data $\mathbf{S}$ drastically improves the ability of the adversary to gain access. For all of our four systems, the SAR is equal to one for an adversary enhanced with the knowledge of $\mathbf{S}$, that is,

$$P_a(\mathbf{S}) = 1.$$

This is because an adversary with knowledge of $\mathbf{S}$ can gain access by choosing $\mathbf{C}$ (and also $\mathbf{J}$ in the two-factor systems) so that the decoding function will select a decoding coset containing a sequence with weight less than τn in order to gain access. In fact, this limitation is not unique to ECC-based systems as the following theorem shows.

Theorem 4 *For any given keyless or two-factor system, if for every $\mathbf{S} \in \mathcal{S}$, the condition that there exist $\mathbf{D}$ (and also $\mathbf{L}$ for two-factor systems) such that $g(\mathbf{D}, \mathbf{S}) = 1$ (or $g(\mathbf{D}, \mathbf{L}, \mathbf{S}) = 1$ for two-factor systems) is satisfied, then*

$$P_a(\mathbf{S}) = 1.$$

In general,

$$P_a(\mathbf{S}) \geq 1 - P_m.$$

Proof: If the condition is satisfied, then the adversary can always choose $\mathbf{C}$ (and also $\mathbf{J}$ in the two-factor systems) such that $\hat{\theta} = g(\mathbf{C}, \mathbf{S}) = 1$ (or $g(\mathbf{D}, \mathbf{L}, \mathbf{S}) = 1$ for two-factor systems) in order to gain access with probability one. Let $\mathcal{S}_a \subset \mathcal{S}$ denote the subset for which the condition is satisfied. If $\mathbf{S} \notin \mathcal{S}_a$, then $\hat{\theta} = 0$. Therefore, the FRR must be bounded by

$$P_m \geq \Pr\left[\mathbf{S} \notin \mathcal{S}_a\right].$$

Since the adversary can gain access when $\mathbf{S} \in \mathcal{S}_a$, the SAR is bounded by

$$P_a(\mathbf{S}) \geq \Pr\left[\mathbf{S} \in \mathcal{S}_a\right] \geq 1 - P_m.$$

■

In both of the keyless systems, an adversary with knowledge of $\mathbf{A}$ can select the syndrome, $\mathbf{H}(\mathbf{A} \oplus \mathbf{D})$, in order to select

a coset with a low-weight sequence and hence

$$P_a(\mathbf{A}) = 1.$$

Since $P_a(\mathbf{V}_1, \mathbf{V}_2) \geq P_a(\mathbf{V}_1)$, we also have

$$P_a(\mathbf{A}, \mathbf{S}) = 1.$$

The SAR performance of our two keyless systems is summarized in the following theorem.

Theorem 5 *For both the keyless secure sketch system and the keyless fuzzy commitment system, the SAR for various cases of data exposure are given by*

$$P_a(\mathbf{S}) = P_a(\mathbf{A}) = P_a(\mathbf{A}, \mathbf{S}) = 1.$$

In both of the two-factor systems, an adversary with knowledge of only $\mathbf{K}$, submits attack vectors $(\mathbf{C}, \mathbf{J})$ that are independent of $\mathbf{A}$. Hence, the distribution of the syndrome is still Bernoulli(0.5), as in the FAR analysis, and thus

$$P_a(\mathbf{K}) = P_f.$$

An adversary with knowledge of only $\mathbf{A}$, submits attack vectors $(\mathbf{C}, \mathbf{J})$ that are independent of $\mathbf{K}$. Hence again the distribution of the syndrome is still Bernoulli(0.5), and thus

$$P_a(\mathbf{A}) = P_f.$$

Knowledge of both $\mathbf{A}$ and $\mathbf{K}$ allows an adversary to arbitrarily choose the syndrome. Thus,

$$P_a(\mathbf{A}, \mathbf{K}) = 1.$$

The SAR performance of our two two-factor systems is summarized in the following theorem.

Theorem 6 *For both the two-factor secure sketch system and the two-factor fuzzy commitment system, the SAR for various cases of data exposure are given by*

$$P_a(\mathbf{K}) = P_a(\mathbf{A}) = P_f,$$

hence they are two-factor secure, and

$$\begin{aligned} 1 = P_a(\mathbf{S}) = P_a(\mathbf{A}, \mathbf{K}) = P_a(\mathbf{A}, \mathbf{S}) \\ = P_a(\mathbf{S}, \mathbf{K}) = P_a(\mathbf{A}, \mathbf{S}, \mathbf{K}). \end{aligned}$$

VII. Information Leakage and Revocability Properties

In this section we will analyze the information leaked about the enrollment biometric in the event of a compromise. We will also analyze the revocability and linkage attack resistance of our secure biometrics realizations.

A. Information Leakage

Theorem 7 *In our keyless systems, the information leakage of* $\mathbf{A}$ *from* $\mathbf{S}$ *is given by*

$$I(\mathbf{A}; \mathbf{S}) = m = n(1 - R) > 0.$$

Proof: In the keyless fuzzy commitment scheme,

$$\begin{aligned} I(\mathbf{A}; \mathbf{S}) &= H(\mathbf{S}) - H(\mathbf{S}|\mathbf{A}) \\ &= H(\mathbf{A} \oplus \mathbf{G}^T\mathbf{Z}) - H(\mathbf{A} \oplus \mathbf{G}^T\mathbf{Z}|\mathbf{A}) \\ &= H(\mathbf{A}) - H(\mathbf{G}^T\mathbf{Z}) \\ &= n - k = m, \end{aligned}$$

and in the keyless secure sketch scheme,

$$I(\mathbf{A}; \mathbf{S}) = H(\mathbf{S}) - H(\mathbf{S}|\mathbf{A}) = H(\mathbf{S}) = m.$$

■

Theorem 8 *In our two-factor systems, the information leakages of* $\mathbf{A}$ *from* $\mathbf{S}$*,* $\mathbf{K}$*, or* $(\mathbf{S}, \mathbf{K})$ *are given by*

$$\begin{aligned} I(\mathbf{A}; \mathbf{K}) &= 0, \\ I(\mathbf{A}; \mathbf{S}) &= 0, \\ I(\mathbf{A}; \mathbf{S}, \mathbf{K}) &= m = n(1 - R) > 0. \end{aligned}$$

Proof: In the two-factor fuzzy commitment scheme,

$$\begin{aligned} I(\mathbf{A}; \mathbf{K}) &= 0, \\ I(\mathbf{A}; \mathbf{S}) &= 0, \\ I(\mathbf{A}; \mathbf{S}, \mathbf{K}) &= H(\mathbf{A}) - H(\mathbf{A}|\mathbf{S}, \mathbf{K}) \\ &= H(\mathbf{A}) - H(\mathbf{A}|\mathbf{A} \oplus \mathbf{G}^T\mathbf{Z}) \\ &= n - k = m, \end{aligned}$$

and in the two-factor secure sketch scheme,

$$\begin{aligned} I(\mathbf{A}; \mathbf{K}) &= 0, \\ I(\mathbf{A}; \mathbf{S}) &= H(\mathbf{S}) - H(\mathbf{S}|\mathbf{A}), \\ &= H(\mathbf{HA} \oplus \mathbf{K}) - H(\mathbf{HA} \oplus \mathbf{K}|\mathbf{A}) \\ &= 0 \\ I(\mathbf{A}; \mathbf{S}, \mathbf{K}) &= H(\mathbf{S}, \mathbf{K}) - H(\mathbf{S}, \mathbf{K}|\mathbf{A}) \\ &= H(\mathbf{S}) + H(\mathbf{K}|\mathbf{S}) - H(\mathbf{K}|\mathbf{A}) - H(\mathbf{S}|\mathbf{A}, \mathbf{K}) \\ &= H(\mathbf{S}) + H(\mathbf{K}) - H(\mathbf{K}) - 0 \\ &= H(\mathbf{S}) = m. \end{aligned}$$

■

B. Linkage Resistance and Revocability

Now, we analyze an attack in which the adversary has compromised multiple access control or identity verification devices containing the user's stored templates.

Lemma 1 *For any given keyless or two-factor system, let* u *parallel enrollments be given. For any* $M \subset \{1, \ldots, u\}$ *and any* $\mathbf{V}_1, \ldots, \mathbf{V}_u$*, where* $\mathbf{V}_i$ *is either null or* $\mathbf{S}_i$ *or (for two-factor systems)* $\mathbf{K}_i$ *or* $(\mathbf{S}_i, \mathbf{K}_i)$*,*

$$I(\mathbf{A}; \mathbf{V}_1, \ldots, \mathbf{V}_u) \leq I\big(\mathbf{A}; \{\mathbf{V}_j\}_{j \in M}\big) + \sum_{i \notin M} I(\mathbf{A}; \mathbf{V}_i).$$

Proof:

$$\begin{aligned}
&I(\mathbf{A};\mathbf{V}_1,\ldots,\mathbf{V}_u)\\
&\quad= H(\mathbf{V}_1,\ldots,\mathbf{V}_u) - H(\mathbf{V}_1,\ldots,\mathbf{V}_u|\mathbf{A})\\
&\quad= H(\mathbf{V}_1,\ldots,\mathbf{V}_u) - H\big(\{\mathbf{V}_j\}_{j\in M}|\mathbf{A}\big) - \sum_{i\notin M} H(\mathbf{V}_i|\mathbf{A})\\
&\quad\le H\big(\{\mathbf{V}_j\}_{j\in M}\big) - H\big(\{\mathbf{V}_j\}_{j\in M}|\mathbf{A}\big)\\
&\qquad+ \sum_{i\notin M} H(\mathbf{V}_i) - H(\mathbf{V}_i|\mathbf{A})\\
&\quad= I\big(\mathbf{A};\{\mathbf{V}_j\}_{j\in M}\big) + \sum_{i\notin M} I(\mathbf{A};\mathbf{V}_i).
\end{aligned}$$

■

Theorem 9 *The two-factor systems are Resistant to Linkage Attacks.*

Proof: If none of the $\mathbf{V}_i = (\mathbf{S}_i, \mathbf{K}_i)$, then, by Lemma 1,

$$\begin{aligned}
I(\mathbf{A};\mathbf{V}_1,\ldots,\mathbf{V}_u) &\le \sum_{i=1}^{u} I(\mathbf{A};\mathbf{V}_i) = 0\\
&= \max_{i\in\{1,\ldots,u\}} I(\mathbf{A};\mathbf{V}_i),
\end{aligned}$$

since $I(\mathbf{A};\mathbf{S}_i) = I(\mathbf{A};\mathbf{K}_i) = 0$.

Otherwise, let the set $M \subset \{1,\ldots,u\}$ denote the locations where for $i \in M$, $\mathbf{V}_i = (\mathbf{S}_i, \mathbf{K}_i)$. By Lemma 1,

$$\begin{aligned}
&I(\mathbf{A};\mathbf{V}_1,\ldots,\mathbf{V}_u)\\
&\quad\le I\big(\mathbf{A};\{\mathbf{V}_j\}_{j\in M}\big) + \sum_{i\notin M} I(\mathbf{A};\mathbf{V}_i)\\
&\quad= I\big(\mathbf{A};\{\mathbf{S}_j,\mathbf{K}_j\}_{j\in M}\big),
\end{aligned}$$

since for $i \notin M$, $\mathbf{V}_i$ is either $\mathbf{S}_i$, $\mathbf{K}_i$ or null, and hence $I(\mathbf{A};\mathbf{V}_i) = 0$. For both of the two-factor systems, $\mathbf{A} - \mathbf{HA} - \{(\mathbf{S}_j,\mathbf{K}_j)\}_{j\in M}$ forms a Markov chain. Hence, by the data processing inequality,

$$\begin{aligned}
&I(\mathbf{A};\mathbf{V}_1,\ldots,\mathbf{V}_u)\\
&\quad\le I\big(\mathbf{A};\{\mathbf{S}_j,\mathbf{K}_j\}_{j\in M}\big)\\
&\quad\le I(\mathbf{A};\mathbf{HA}) = m = n(1-R)\\
&\quad= I(\mathbf{A};\mathbf{S},\mathbf{K}) = \max_{i\in\{1,\ldots,u\}} I(\mathbf{A};\mathbf{V}_i)
\end{aligned}$$

■

Theorem 10 *The two-factor systems are Revocable against Multiple Exposures.*

Proof: The information leakage condition is satisfied, as a consequence of Lemma 1 and the disjointedness of (M, N),

$$\begin{aligned}
&I(\mathbf{A};\{\mathbf{S}_i\}_{i\in M},\{\mathbf{K}_j\}_{j\in N})\\
&\quad\le \left[\sum_{i\in M} I(\mathbf{A};\mathbf{S}_i)\right] + \left[\sum_{j\in N} I(\mathbf{A};\mathbf{K}_j)\right]\\
&\quad= 0.
\end{aligned}$$

When the knowledge enhancing the adversary is $(\{\mathbf{S}_i\}_{i\in M},\{\mathbf{K}_j\}_{j\in N})$, the attack vectors $(\mathbf{C},\mathbf{J})$ must still be independent of $\mathbf{A}$. Hence, any attack by the adversary still results in a uniform distribution on the syndrome, and the SAR performance is the same as the FAR performance. Likewise, when the knowledge enhancing the adversary is $(\mathbf{A},\{\mathbf{S}_i\}_{i\in M'},\{\mathbf{K}_j\}_{j\in N'})$, the attack vectors $(\mathbf{C},\mathbf{J})$ must still be independent of $\mathbf{K}_k$. Thus, again any attack still results in a uniform distribution on the syndrome, and the SAR performance is the same as the FAR performance. ■

C. Nonidentical Enrollments

So far, when discussing multiple enrollments, we have assumed that the same code is used to generate each parallel enrollment. This results in enrollments that are identically and independently distributed given $\mathbf{A}$. Another possibility to consider is if different codes are used in each parallel enrollment. Each enrollment (for the two-factor secure sketch system) would be given by

$$\mathbf{S}_i = \mathbf{H}_i\mathbf{A} \oplus \mathbf{K}_i.$$

The nonidentical enrollments for the other systems are also similarly described with $\mathbf{H}$ replaced with $\mathbf{H}_i$. The enrollments would still be independent conditioned on $\mathbf{A}$ but would no longer be identical. Using different codes in each enrollment would mean that the Resistance to Linkage Attacks property might not necessarily be satisfied. The following theorem gives the information leakage for this setup.

Theorem 11 *Let u nonidentical enrollments be generated for the secure biometric systems considered in this paper.*

For the keyless systems, for any $\mathbf{V}_1,\ldots,\mathbf{V}_u$, where $\mathbf{V}_i$ is either null or $\mathbf{S}_i$, and $M \subset \{1,\ldots,u\}$ such that $i \in M$ if and only if $\mathbf{V}_i = \mathbf{S}_i$, then

$$I(\mathbf{A};\mathbf{V}_1,\ldots,\mathbf{V}_u) = \operatorname{rank}\big(\{\mathbf{H}_i\}_{i\in M}\big),$$

where $\operatorname{rank}\big(\{\mathbf{H}_i\}_{i\in M}\big)$ is the number of independent rows in set of matrices, $\{\mathbf{H}_i\}_{i\in M}$.

For the two-factor systems, for any $\mathbf{V}_1,\ldots,\mathbf{V}_u$, where $\mathbf{V}_i$ is either null, $\mathbf{S}_i$, $\mathbf{K}_i$ or $(\mathbf{S}_i,\mathbf{K}_i)$, and $M \subset \{1,\ldots,u\}$ such that $i \in M$ if and only if $\mathbf{V}_i = (\mathbf{S}_i,\mathbf{K}_i)$, then

$$I(\mathbf{A};\mathbf{V}_1,\ldots,\mathbf{V}_u) = \operatorname{rank}\big(\{\mathbf{H}_i\}_{i\in M}\big).$$

Proof: By Lemma 1, and since $\mathbf{A} - \{\mathbf{H}_j\mathbf{A}\}_{j\in M} - \{\mathbf{V}_j\}_{j\in M}$ form a Markov chain,

$$\begin{aligned}
&I(\mathbf{A};\mathbf{V}_1,\ldots,\mathbf{V}_u)\\
&\quad\le I\big(\mathbf{A};\{\mathbf{V}_j\}_{j\in M}\big) + \sum_{i\notin M} I(\mathbf{A};\mathbf{V}_i)\\
&\quad= I\big(\mathbf{A};\{\mathbf{V}_j\}_{j\in M}\big)\\
&\quad\le I\big(\mathbf{A};\{\mathbf{H}_j\mathbf{A}\}_{j\in M}\big)\\
&\quad= H(\mathbf{A}) - H\big(\mathbf{A}|\{\mathbf{H}_j\mathbf{A}\}_{j\in M}\big)\\
&\quad= n - \big(n - \operatorname{rank}\big(\{\mathbf{H}_i\}_{i\in M}\big)\big)\\
&\quad= \operatorname{rank}\big(\{\mathbf{H}_i\}_{i\in M}\big).
\end{aligned}$$

■

VIII. Conclusions

In this paper, we presented a generalized framework, within which it is possible to characterize the security and privacy of secure biometric systems. Specifically, these attributes are specified using the false reject rate, the false acceptance rate and the successful attack rate. Further, the framework allows us to examine the robustness of secure biometric systems to theft of the biometric template, i.e., to measure the information leakage when a biometric template is compromised, and must be revoked and replaced by a new template. The issue of revocability naturally led us to consider the possibility of simultaneous attacks on various biometric systems that utilize templates derived from a given user.

We conducted an information-theoretic analysis of the above properties of secure biometric systems, by looking at two popular realizations of secure biometrics, namely secure sketch, and fuzzy commitment-based schemes. We considered two variants of each scheme, the first being a keyless scheme and the second being a two-factor scheme in which the biometric system is augmented by a secret key held on a smart card. Our analysis shows that secure sketch-based schemes and fuzzy-commitment based schemes are equivalent with respect to the false reject rate, false accept rate, successful attack rate, and information leakage during partial or full compromise of biometric templates and smart-card keys.

For both secure sketches, and fuzzy commitment, compromising the stored data renders the biometric system information theoretically vulnerable to attacks, i.e., an attacker can gain access to the system with probability one, though he may not be able to recover the user's biometric sample. Thus security, in the form of authentication integrity is compromised with probability one, but the user can still retain positive information theoretic privacy. By incorporating a two-factor scheme using a one time pad as a secret key carried on a smart card, it is possible to prevent this security compromise, i.e., the successful attack rate is now no larger than the false acceptance rate of the system. This holds, also for the case of simultaneous attacks on multiple biometric systems, so long as no single system suffers from a theft of *both* the stored data as well as its smart card key.

Finally, the one distinction between secure sketches and fuzzy commitment is the data storage requirement needed to obtain a given tradeoff between the false reject rate and the false acceptance rate (or the successful attack rate). In an implementation based on error correcting codes (ECC), for example, fuzzy commitment-based scheme requires the stored data to be a codeword of the ECC, while secure sketch requires it to be a syndrome of the ECC. Depending on the rate of the ECC, and the number of enrolled individuals, this difference could have a significant impact on the storage requirements for the biometric database.

Regarding future work, an immediate extension is to the case of nonidentical enrollment data. In Section VII-C we explored the issues raised when different codes are used for different enrollments. Returning to our original motivation that no two enrollments are the same, we can model the system as having an underlying biometric, from which each (enrollment) measurement differs in a conditionally independent manner.

References

[1] G. Davida, Y. Frankel, and B. Matt, "On enabling secure applications through off-line biometric identification," in *IEEE Symp. on Security and Privacy*, 1998, pp. 148–157.

[2] A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in *Proc. ACM Conf. on Computer and Communications Security*, 1999, pp. 28–36.

[3] A. Juels and M. Sudan, "A fuzzy vault scheme," in *IEEE Intl. Symp. on Information Theory*, 2002.

[4] P. Tuyls, A. Akkermans, T. Kevenaar, G. Schrijen, A. Bazen, and R. Veldhuis, "Practical biometric authentication with template protection," in *Audio- and Video-Based Biometric Person Authentication, 5th International Conference*, vol. 3546. Hilton Rye Town, NY: Springer Lecture Notes in Computer Science, T. Kanade, A. Jain and N. Ratha eds., July 2005, pp. 436–446.

[5] Y. Dodis, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," in *Eurocrypt*, ser. LNCS, vol. 3027. Springer-Verlag, 2004, pp. 523–540.

[6] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Information Theory*, pp. 471–480, Jul. 1973.

[7] Q. Li, Y. Sutcu, and N. Memon, "Secure sketch for biometric templates," in *Asiacrypt*, Shanghai, China, December 2006.

[8] Y. Sutcu, Q. Li, and N. Memon, "Protecting biometric templates with sketch: Theory and practice," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 503–512, September 2007.

[9] S. Draper, A. Khisti, E. Martinian, A. Vetro, and J. Yedidia, "Secure storage of fingerprint biometrics using slepian-wolf codes," in *Information Theory and Applications Workshop in San Diego, CA*, 2007.

[10] Y. Sutcu, S. Rane, J. Yedidia, S. Draper, and A. Vetro, "Feature extraction for a slepian-wolf biometric system using ldpc codes," in *Proceedings of the IEEE International Symposium on Information Theory*, Toronto, Canada, July 2008.

[11] X. Boyen, "Reusable cryptographic fuzzy extractors," in *Proceedings of the 11th ACM conference on Computer and Communications Security*. ACM Press, 2004, pp. 82–91.

[12] K. Simoens, P. Tuyls, and B. Preneel, "Privacy Weakness in Biometric Sketches," in *IEEE Symposium on Security and Privacy*, Oakland, CA, may 2009.

[13] L. Lai, S. Ho, and H. V. Poor, "Privacy-security tradeoffs in biometric security systems," in *Proceedings of the Third international Conference on Advances in Biometrics*, vol. 5558. Alghero, Italy: Springer-Verlag Lecture Notes in Computer Science, June 2009, pp. 879–888.

[14] T. Ignatenko and F. Willems, "Biometric systems: Privacy and secrecy aspects," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 4, Dec 2009.

[15] R. Gallager, *Information Theory and Reliable Communication*. Wiley Publishing, 1968.

[16] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 13–30, March 1963.

[17] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error Correcting Codes*. North Holland Publishing Company, 1977.