MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Feature Extraction for a Slepian-Wolf Biometric System Using LDPC Codes

Yagiz Sutcu, Shantanu Rane, Jonathan Yedidia, Stark Draper, Anthony Vetro

TR2008-036 August 2008

Abstract

We present an information-theoretically secure biometric storage system using graph-based error correcting codes in a Slepian-Wolf coding framework. Our architecture is motivated by the noisy nature of personal biometrics and the requirement to provide security without storing the true biometric at the device. The principal difficulty is that real biometric signals, such as fingerprints, do not obey the i.i.d. or ergodic statistics that are required for the underlying typicality properties in the Slepian-Wolf coding framework. To meet this challenge, we propose to transform the biometric data into binary feature vectors that are i.i.d. Bernoulli (0.5), independent across different users, and related within the same user through a BSC-p channel with small p less-than 0.5. Since this is a standard channel model for LDPC codes, the feature vectors are now suitable for LDPC syndrome coding. The syndromes serve as secure biometrics for access control. Experiments on a fingerprint database demonstrate that the system is information-theoretically secure, and achieves very low false accept rates and low reject rates.

IEEE International Symposium on Information Theory, 2008

Feature Extraction for a Slepian-Wolf Biometric System Using LDPC Codes

Yagiz Sutcu, Shantanu Rane, Jonathan S. Yedidia, Stark C. Draper, Anthony Vetro

***Abstract*— We present an information-theoretically secure biometric storage system using graph-based error correcting codes in a Slepian-Wolf coding framework. Our architecture is motivated by the noisy nature of personal biometrics and the requirement to provide security without storing the true biometric at the device. The principal difficulty is that real biometric signals, such as fingerprints, do not obey the i.i.d. or ergodic statistics that are required for the underlying typicality properties in the Slepian-Wolf coding framework. To meet this challenge, we propose to transform the biometric data into binary feature vectors that are i.i.d. Bernoulli(0.5), independent across different users, and related within the same user through a BSC-p channel with small $p < 0.5$. Since this is a standard channel model for LDPC codes, the feature vectors are now suitable for LDPC syndrome coding. The syndromes serve as secure biometrics for access control. Experiments on a fingerprint database demonstrate that the system is information-theoretically secure, and achieves very low false accept rates and low false reject rates.**

***Index Terms*— Biometric security, Slepian-Wolf coding, LDPC codes, feature transformation, fingerprint**

I. Introduction

Computer-verifiable biometrics, such as fingerprints and iris scans, provide an attractive alternative to classical access control solutions like passwords and identifying documents. Biometrics have the advantage that, unlike passwords, they do not have to be remembered and, unlike identifying documents, they are difficult to forge. However, they pose new challenges and create new security holes. A key characteristic is that each time a biometric is measured, the observation differs slightly. For example, in the case of fingerprints, the reading might change because of elastic deformations in the skin when placed on the sensor surface, dust or oil between finger and sensor, or a cut to the finger. Biometric authentication systems must be robust to such variations. Most biometric authentication systems deal with such variability by relying on pattern recognition. To perform recognition, the enrollment biometric is stored on the device for comparison with the probe biometric. This creates a security hole: An attacker who gains access to the device also gains access to the biometric. This is clearly a serious problem, made worse by the fact that an individual cannot generate new biometrics if the system is compromised.

Y. Sutcu is with the Electrical & Computer Engineering Department, Polytechnic University, {yagiz}@isis.poly.edu. S. Rane, J. S. Yedidia and A. Vetro are with Mitsubishi Electric Research Laboratories (MERL) {rane,yedidia,vetro}@merl.com. S. C. Draper is with the Electrical and Computer Engineering Department at the University of Wisconsin, Madison, {sdraper}@ece.wisc.edu. This work was performed when Y. Sutcu was an intern at MERL.

From an information theoretic perspective the secure biometric problem is a problem of "common randomness" [1]. Different parties observe correlated random variables (the enrollment and the probe) and then attempt to agree on a shared secret (the enrollment biometric). The tool used to extract the secret is a Slepian-Wolf code [2].

More specifically, error correction coding within the Slepian-Wolf framework has been proposed to deal with the joint problem of providing security against attackers while accounting for the inevitable variability of biometrics. On the one hand, the error correction capability of a channel code can accommodate the slight variation between multiple measurements of the same biometric [3], [4]. On the other hand, the check bits of the error correction code can perform much the same function as a cryptographic hash of a password on conventional access control systems. Just as a hacker cannot invert the hash and steal the password, he cannot just use the check bits to recover and steal the biometric. However, it has been found that schemes based on this principle [5], [6], [7] yield high false reject rates. One reason for this is that the statistical relationship between the enrollment biometric and probe is not accurately captured by the simple noise models assumed in the theoretical works [3], [4].

In references [8], [9], the shortcomings of the prior algebraic coding approaches were addressed by using graphical coding techniques. Graphical codes, e.g., LDPC codes, can closely approach the Shannon bound, and can therefore be much more powerful than algebraic coding techniques. The LDPC code graph was augmented with a second graph that described the complex "fingerprint channel" relating the enrollment biometric to the probe biometric. Syndromes generated via LDPC coding of the enrollment biometric were used as "secure" biometrics and stored on the device. During authentication, iterative decoding using belief propagation (BP) was performed across *both* graphs. Even though the "fingerprint channel" used was apparently a reasonable model of the variations between the enrollment and probe fingerprints, the performance of the overall decoding system was insufficient to obtain information-theoretic security. In other words, a standard LDPC code designed for a binary symmetric channel (BSC) was not an efficient code for the fingerprint channel model.

We propose here a different approach that still aims for a Slepian-Wolf system using LDPC codes. Rather than trying to incorporate the fingerprint channel into a factor graph modeling the entire system, we propose to transform the enrollment and probe biometrics into feature vectors that are related by a simpler channel model that is more suitable for LDPC coding. We explicitly design the feature set to be com-

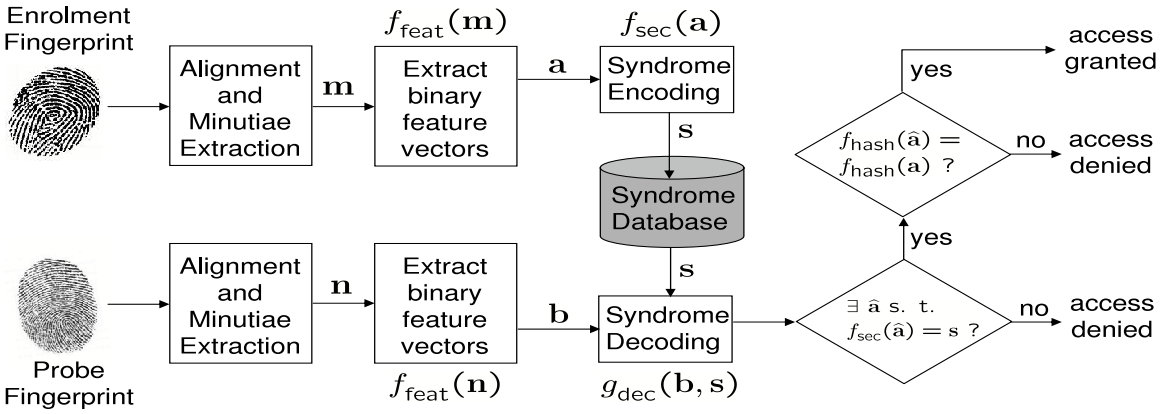


Fig. 1. Robust feature extraction is combined with syndrome coding to build a secure fingerprint biometrics system.

patible with code designs, syndrome encoding and syndrome decoding procedures that already exist. For a particular set of features with predetermined statistical properties, we are then able to utilize a LDPC code for a Binary Symmetric Channel (BSC) that matches the designed feature set. The construction of LDPC codes for the BSC and their associated syndrome encoding and decoding procedures are already well-understood and deeply explored topics. When the code is thus matched to the feature set, the resulting system can be shown to be information-theoretically secure.

This paper is organized as follows. In Section II, we describe a secure biometrics scheme which transforms fingerprint biometrics into feature vectors that are appropriate for LDPC syndrome coding. This section lists the desirable statistical properties of these feature vectors and provides an information-theoretic justification for the security of a syndrome code operating on these feature vectors. In Section III, the actual process of feature extraction is described. In Section IV, we extract feature vectors from a fingerprint database, evaluate them for security and robustness and investigate the performance of the overall distributed biometric coding scheme.

II. Secure Fingerprint Biometrics Scheme

The proposed scheme for secure fingerprint biometrics is shown in Fig. 1. The central idea is to generate binary feature vectors which are i.i.d. Bernoulli(0.5), independent across different users but different measurements of the same user are related by a binary symmetric channel with crossover probability p (BSC-p) where p is much smaller than 0.5. This is one of the standard channel models for LDPC codes and therefore standard LDPC codes can be used for Slepian-Wolf coding of the feature vectors. We emphasize that the feature transformation is made public and is *not* assumed to provide any security. Security is provided by the syndromes generated by the Slepian-Wolf coder.

A. Minutiae-Based Fingerprint Representation

A popular method for working with fingerprint data is to extract a set of "minutiae points" and to perform all subsequent operations on them. Minutiae points have been observed to be stable over many years. Each minutiae is a discontinuity in the ridge map of a fingerprint, characterized by a triplet (x, y, θ) representing its spatial location in two dimensions and the angular orientation. In the minutiae map $\mathbf{M}$ of a fingerprint, $\mathbf{M}(x, y) = \theta$ if there is a minutia point at (x, y) and $\mathbf{M}(x, y) = \emptyset$ (empty set) otherwise. A minutiae map may be considered as a joint quantization and feature extraction function which operates on the fingerprint image. Different fingerprints normally have different numbers of minutiae points.

B. Enrollment and Authentication Procedure

During enrollment, the user provides a fingerprint from which the system first determines a minutiae map $\mathbf{m}$. Next, a feature transformation function $f_{\text{feat}}(\cdot)$ maps the minutiae array into a binary feature vector $\mathbf{a} = f_{\text{feat}}(\mathbf{m})$. We consider that $\mathbf{m}$ is a realization of a random array $\mathbf{M}$ with some unknown distribution. The feature vector $\mathbf{a}$, obtained from this map is a realization of a binary random vector $\mathbf{A}$ of fixed preset length N, drawn according to some distribution $P_{\mathbf{A}}(\mathbf{a})$ on the finite set $\mathcal{A} = \{0, 1\}^N$. Individual bits of $\mathbf{A}$ are denoted by A_i with $i \in \mathcal{I} = \{1, 2, ..., N\}$. Next, a function $f_{\text{sec}}(\cdot)$ maps the binary feature vector into a secure biometric $\mathbf{s} = f_{\text{sec}}(\mathbf{a})$. In the proposed scheme, $f_{\text{sec}}(\cdot)$ is a syndrome encoding using a graph of an LDPC code $\mathbb{C}$. The access control system stores $\mathbf{s}$, $\mathbb{C}$ and a cryptographic hash of the binary feature vector $f_{\text{hash}}(\mathbf{a})$. It does not store $\mathbf{m}$ or $\mathbf{a}$ or the image of the original fingerprint.

During authentication, a user or attacker requests access by providing a probe fingerprint from which the authenticator obtains a minutiae map $\mathbf{n}$. Next, it transforms $\mathbf{n}$ into a probe feature vector $\mathbf{b} = f_{\text{feat}}(\mathbf{n})$. Now, the LDPC decoder assumes that the probe feature vector $\mathbf{b}$ is an error prone version of the enrollment feature vector $\mathbf{a}$. It combines the secure biometric $\mathbf{s}$ (syndrome) and the probe feature vector $\mathbf{b}$ and performs belief propagation decoding. The result of belief propagation is either an estimate $\widehat{\mathbf{a}}$ of enrollment feature vector $\mathbf{a}$, or a special symbol $\emptyset$ indicating decoder failure. Now, it is possible that $\widehat{\mathbf{a}} \neq \mathbf{a}$, yet $\widehat{\mathbf{a}}$ satisfies the syndrome $\mathbf{s}$. To protect against this possibility, and more importantly to protect against an attacker using a stolen set of syndromes to construct his own estimate $\widehat{\mathbf{a}}$ which satisfies the syndromes but is not the true biometric, access is granted if and only if $f_{\text{hash}}(\widehat{\mathbf{a}}) = f_{\text{hash}}(\mathbf{a})$.

C. Desirable Statistical Properties of Feature Vectors

Based on the requirements mentioned at the beginning of this section, it is desirable that the feature vectors possess the following properties:

1) A bit in a feature vector representation is equally likely to be a 0 or a 1. Thus, $Pr\{A_i = 0\} = Pr\{A_i = 1\} = 1/2$ and $H(A_i) = 1$ bit for all $i \in \mathcal{I}$.
2) Different bits in a given feature vector are independent of each other, so that a given bit provides no information about any other bit. Thus, the pairwise entropy $H(A_i, A_j) = H(A_i) + H(A_j) = 2$ bits for all $i \neq j$ where $i, j \in \mathcal{I}$.
3) Feature vectors $\mathbf{A}$ and $\mathbf{B}$ from different fingers are independent of each other, so that one person's feature vector provides no information about another person's feature vector. Thus, the pairwise entropy $H(A_i, B_j) = H(A_i) + H(B_j) = 2$ bits for all $i, j \in \mathcal{I}$.
4) Feature vectors $\mathbf{A}$ and $\mathbf{A}'$ obtained from different readings of the same finger are statistically related by a BSC-p. If p is small, it means that the feature vectors are robust to repeated noisy measurements with the same finger. Thus, $H(A'_i|A_i) = H(p)$ for all $i \in \mathcal{I}$.

D. Syndrome Coding of Feature Vectors

The feature extraction function $f_{\text{feat}}(\cdot)$ induces a distribution on the enrollment feature vector $\mathbf{A}$ and probe feature vector $\mathbf{B}$. Assume that $\mathbf{A}$ and $\mathbf{B}$ are jointly ergodic and, as noted earlier, they take values from a finite set $\mathcal{A} = \{0, 1\}^N$.

A Slepian-Wolf code [2] is a rate-R_{SW} random "binning" function that encodes a particular enrollment vector $\mathbf{A} = \mathbf{a}$ into the secured biometric $\mathbf{s}$. Specifically, we assign each to possible sequence $\mathbf{a} \in \mathcal{A}$ an integer s selected uniformly from $\{1, 2, ..., 2^{NR_{SW}}\}$. This index serves as the secure biometric $s = f_{\text{sec}}(f_{\text{feat}}(\mathbf{m}))$ derived from the given minutiae map $\mathbf{m}$. Each possible index $s \in \{1, 2, ..., 2^{NR_{SW}}\}$ indexes a set or "bin" of enrollment feature vectors, $\{\mathbf{a}|f_{\text{sec}}(\mathbf{a}) = s\}$. The secure biometric can be thought of as a scalar index s or its binary expansion $\mathbf{s}$, which is a uniformly distributed bit sequence of length NR_{SW}.

During authentication, the Slepian-Wolf decoder is provided with a particular probe feature vector $\mathbf{b}$, generated from a minutiae map $\mathbf{n}$ which claims to be from a particular enrolled user $\mathbf{a}$, for example. The decoder $g_{\text{dec}}(\mathbf{b}, \mathbf{s})$ searches for a vector $\widehat{\mathbf{a}} \in \mathcal{A}$ such that is $\widehat{\mathbf{a}}$ is jointly typical with $\mathbf{b}$ under the joint distribution $p_{\mathbf{a},\mathbf{n}}$ and is in bin $\mathbf{s}$, i.e., $f_{\text{sec}}(\widehat{\mathbf{a}}) = \mathbf{s}$. If a unique $\widehat{\mathbf{a}}$ is found, then the decoder outputs this result. Otherwise, an authentication failure is declared and the decoder returns $\emptyset$. According to the Slepian-Wolf Theorem [2], syndrome decoding will succeed with probability approaching 1 as N increases, provided that $R_{\text{SW}} > \frac{1}{N} H(\mathbf{A}|\mathbf{B})$.

Now, consider the probability that an attacker can estimate a feature vector $\mathbf{A}$ given the syndrome $\mathbf{S}$. By the asymptotic equipartition property (AEP) [10], under the mild technical condition of ergodicity, it can be shown that, conditioned on $\mathbf{S} = f_{\text{sec}}(\mathbf{A})$, $\mathbf{A}$ is uniformly distributed over the typical set of size $2^{H(\mathbf{A}|\mathbf{S})}$. Therefore, with high probability, the number of guesses required to identify $\mathbf{a}$ is $2^{H(\mathbf{A}|\mathbf{S})}$. But,

$$\begin{aligned} H(\mathbf{A}|\mathbf{S}) &= H(\mathbf{A}, \mathbf{S}) - H(\mathbf{S}) = H(\mathbf{A}) - H(\mathbf{S}) \\ &= H(\mathbf{A}) - NR_{\text{SW}} = N\left(H(A_i) - R_{\text{SW}}\right) \\ &= N(1 - R_{\text{SW}}) = NR_{\text{LDPC}} > 0 \end{aligned} \quad (1)$$

where the last two equalities follow from properties 1 and 2 in Section II-C, and R_{LDPC} is the rate of the LDPC code used. Thus, the higher the LDPC code rate, the smaller is the probability of successful attack conditioned on an observation of $\mathbf{S}$. Moreover, $H(\mathbf{A}|\mathbf{S}) > 0$ and hence $NR_{\text{SW}} < H(\mathbf{A})$ implies that the system has positive information-theoretic security for any LDPC code rate. This motivates the design of feature vectors with the aforementioned properties. To find the upper bound on the LDPC code rate, note that

$$\begin{aligned} NR_{\text{LDPC}} &= N(1 - R_{\text{SW}}) < N - H(\mathbf{A}|\mathbf{B}) \\ &= H(\mathbf{A}) - H(\mathbf{A}|\mathbf{B}) = I(\mathbf{A}; \mathbf{B}) \end{aligned} \quad (2)$$

which is the capacity of the channel between the enrollment and probe feature vectors.

In a syndrome-independent attack, the attacker guesses candidate vectors $\mathbf{B}$ from the typical set (of $\mathbf{B}$'s) independently of $\mathbf{A}$. For this attack to be successful, syndrome decoding should succeed with $\mathbf{B}$, i.e., $\mathbf{B}$ must be jointly typical with $\mathbf{A}$. Using the AEP for jointly typical sequences, this requires approximately $2^{I(\mathbf{A};\mathbf{B})}$ guesses. From (1) and (2), $I(\mathbf{A}; \mathbf{B}) > H(\mathbf{A}|\mathbf{S})$. Thus, a syndrome-independent attack is even more difficult than an attack in which the choice of $\mathbf{B}$ is conditioned on the syndrome $\mathbf{S}$.

III. Scheme for Obtaining Feature Vectors

To extract N bits from a minutiae map, it suffices to ask N "questions," each with a binary answer. A general framework to accomplish this is shown in Fig. 2. N operations are performed on the biometric to yield a non-binary feature representation which can then be converted to binary by thresholding. As an example, one can project the minutiae map onto N orthogonal basis vectors and quantize the positive projections to 1's and negative projections to 0's.

We define the operation as counting the number of minutiae points that fall in a randomly chosen cuboid in $X - Y - \Theta$ space, as shown in Fig. 2. To chose a cuboid, an origin is selected uniformly at random in $X - Y - \Theta$ space, and the dimensions along the three axes are also chosen at random.

Next, we define the threshold as the median of the number of minutiae points in the chosen cuboid, measured across the complete training set, a method used for face recognition in [11]. The threshold value may differ for each cuboid based on its position and volume. If the number of minutiae points in a randomly generated cuboid exceeds the threshold, then a 1-bit is appended to the feature vector, otherwise a 0-bit is appended. We consider the combined operation of (a) generating a cuboid and (b) thresholding as equivalent to posing a question with a binary answer. N such questions result in an N-bit feature vector.

The simplest way to generate feature vectors is to use the same questions for all users. In the sequel, we consider a more

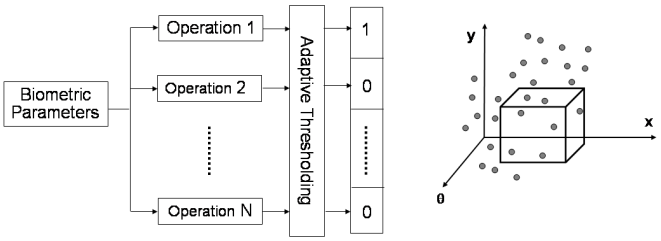


Fig. 2. N questions can be asked by performing N operations on the biometric followed by thresholding. In our scheme, the operation involves counting the minutiae points in a randomly generated cuboid.

advanced approach in which the questions are user-specific. The rationale behind using user-specific questions is that some questions are more robust (reliable) than others. In particular, a question is robust if the number of minutiae points in a cuboid is far removed from the median calculated over the entire dataset. Thus, even if there is spurious insertion or deletion of minutiae points when a noisy measurement of the same fingerprint is provided at a later time, the answer to the question (0 or 1) is less likely to change. On the other hand, if the number of minutiae points is close to the median, the 0 or 1 answer to that question is less reliable. Thus, more reliable questions result in a BSC-p intra-user channel with low p. Different users have a different set of robust questions, and we propose to use these while constructing the feature vector. We emphasize that for the purposes of security analysis, the set of questions used in the system is assumed public. An attacker who steals a set of syndromes and poses falsely as a user will be given the set of questions appropriate to that user. Our security analysis is not based in any way on the obscurity of the questions, but rather on the information-theoretic difficulty of recovering the biometric given only the stolen syndromes.

For a given user i, the *average* number of minutiae points $\bar{m}_{i,j}$ in a given cuboid $\mathcal{C}_j$ is calculated over repeated noisy measurements of the same fingerprint. Let m_j and σ_j be the median and standard deviation of the number of minutiae points in $\mathcal{C}_j$ over the dataset of all users. Then, let $\Delta_{i,j} = (\bar{m}_{i,j} - m_j)/\sigma_j$. The magnitude, $|\Delta_{i,j}|$ is directly proportional to the robustness of the question posed by cuboid $\mathcal{C}_j$ for user i. The sign of $\Delta_{i,j}$ determines whether the cuboid $\mathcal{C}_j$ should be placed into $\mathcal{L}_{0,i}$, a list of questions with a 0 answer for user i, or into $\mathcal{L}_{1,i}$, a list of questions with a 1 answer for user i. Both these lists are sorted in the decreasing order of $|\Delta_{i,j}|$. Now, a fair coin is flipped to choose between $\mathcal{L}_{0,i}$ and $\mathcal{L}_{1,i}$ and the question at the top of the chosen list is stored on the device. After N coin flips, approximately $N/2$ of the most robust questions from each list will be stored on the device. This process is repeated for each enrolled user i.

IV. Experimental Results

A. Data Set and Experimental Setup

In our experiments, we use a proprietary Mitsubishi fingerprint database which contains minutiae maps of 1035 fingers with 15 fingerprint samples taken from each finger. The average number of minutiae points in a single map is approximately 32. All fingerprints are pre-aligned.

B. Statistical Analysis

To measure the extent to which the desired target statistical properties in Section II-C are achieved, we examine the feature vectors obtained from the minutiae maps according to the method described in Section III. The N most robust questions were selected to generate the feature vectors, with N ranging from 50 to 350. Fig. 3 shows the statistical properties of the feature vectors for N=150. As shown in Fig. 3(a), the histogram of the average number of 1-bits in the feature vectors is clustered around $N/2 = 75$. Fig. 3(b) shows that the pair-wise entropy measured between bits of different users is very close to 2 bits, indicating that the bits are nearly pairwise independent.

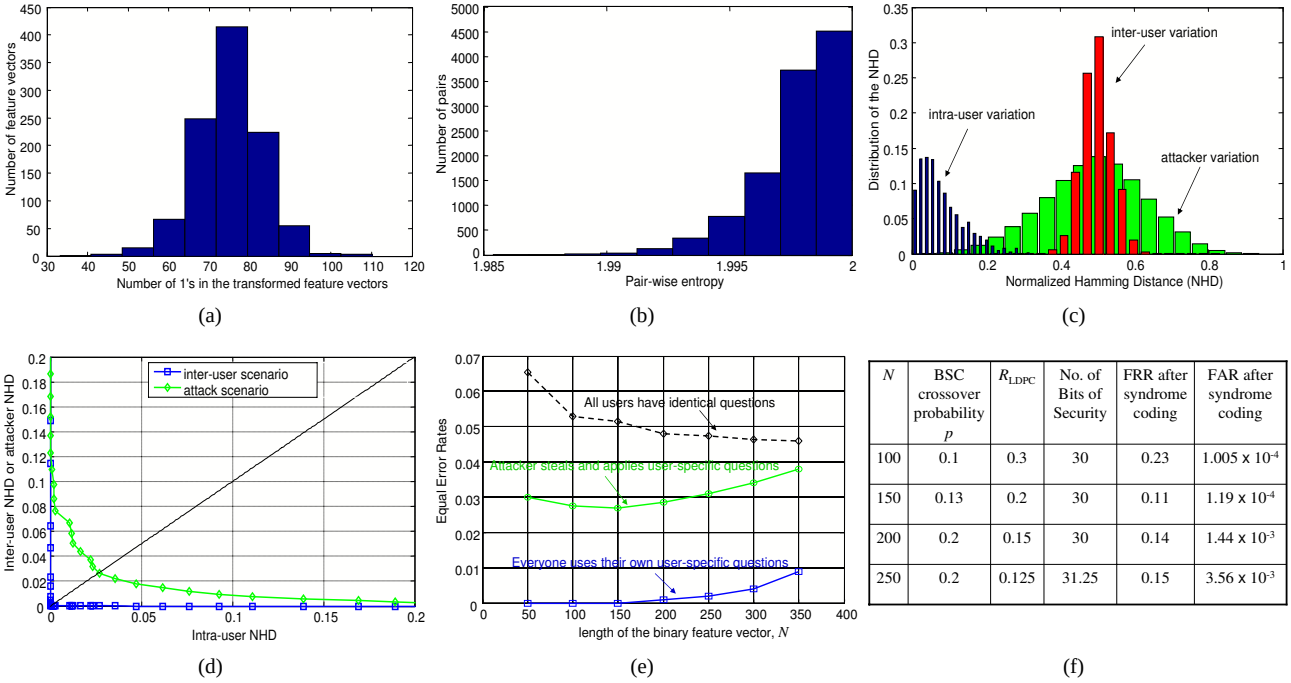
C. Security and Robustness

In order to measure the similarity or dissimilarity of two feature vectors, Normalized Hamming Distance (NHD) is used. The NHD between two feature vectors $\mathbf{a}$ and $\mathbf{b}$, each having length N, is calculated as follows:

$$\text{NHD}(\mathbf{a}, \mathbf{b}) = \frac{1}{N} \sum_{i=1}^{N} (a_i \oplus b_i)$$

where $\oplus$ is summation modulo 2. The plot in Fig. 3(c) contains three histograms: (1) The intra-user variation is the distribution of the average NHD measured pairwise over 15 samples of the same finger, (2) The inter-user variation is the distribution of the NHD averaged over all possible pairs of users, each with his own specific set of questions (3) The attacker variation is the NHD for the case in which an attacker attempts to identify himself as a given user i, while using a different fingerprint $j \neq i$, but while using the 150 robust questions of user i. There is a clean separation between the intra-user and inter-user NHD distributions, and a small overlap between the intra-user and attacker distributions. In Fig. 3(c), the attacker variation becomes relevant if the attacker gains access to the victim's questions. The inter-user variation is relevant if the attacker has not broken into the system, but is merely trying to pose as the victim without knowing the victim's specific questions. In a practical biometric system, the questions would not be publicized. So, most attackers will not have access to them and therefore, the inter-user variation will be relevant instead of the more conservative attacker variation.

To ascertain the effectiveness of the feature vectors, we plot the inter-user NHD against the intra-user NHD in Fig. 3(d) both for the case in which every user employs specific questions and for the case in which an attacker uses the questions stolen from the user being impersonated. One metric for evaluating plots such as Fig. 3(d) is the "Equal Error Rate (EER)", which is the point where the inter-user NHD and intra-user NHD are equal. A lower EER indicates a superior tradeoff. Fig. 3(e) plots the EER for various values of N. Observe that user-specific questions provide a significantly lower EER than using the same questions for all users irrespective of the robustness of the questions. Even if the attacker is provided with the user-specific questions, the resulting EER is lower than the case in which everybody has the same questions.



N	BSC crossover probability p	R_{LDPC}	No. of Bits of Security	FRR after syndrome coding	FAR after syndrome coding
100	0.1	0.3	30	0.23	1.005×10^{-4}
150	0.13	0.2	30	0.11	1.19×10^{-4}
200	0.2	0.15	30	0.14	1.44×10^{-3}
250	0.2	0.125	31.25	0.15	3.56×10^{-3}

(f)

Fig. 3. (a) Histogram of the number of ones in the feature vectors for N=150 is clustered around $N/2 = 75$. (b) The pairwise entropy measured across all pairs and all users is very close to 2 bits. (c) The Normalized Hamming distance between feature vectors shows clear separation within and across users. (d) The tradeoff between intra-user and inter-user separation is plotted by sweeping a threshold NHD in Fig. 3(c). For N=150, equal error rate (EER) is 0.027. (e) User-specific questions result in lower EER than common questions, even if the user-specific questions are given to the attacker. (f) Syndrome coding with an appropriate LDPC code provides an information-theoretically secure biometrics system with low FRR and extremely low FAR.

Based on the separation of intra-user and inter-user distributions, we expect that a syndrome code designed for a BSC-p, with appropriate $p < 0.5$ would authenticate almost all genuine users while rejecting almost all impostors. The table in Fig. 3(f) shows the False Reject Rate (FRR) and the False Accept Rate (FAR)[1] for syndrome coding with different values of N and p. These FAR and FRR values are measures of the security-robustness tradeoff of the distributed biometric coding system. The LDPC code rate is chosen so as to provide about 30 bits of security. This restriction on the LDPC code rate in turn places a restriction on how large p can be, especially for small N. Due to this restriction, the FRR is relatively large for $N = 100$. The lowest FRR is achieved for $N = 150$. As N increases, less robust questions need to be employed, so the statistical properties of the feature vectors diverge from those in Section II-C. Thus, the FRR increases again when N becomes too large.

V. Conclusions

Fingerprint minutiae maps have been transformed into binary feature vectors which are appropriate for LDPC coding. These feature vectors account for the location and orientation of the minutiae points and are robust to the variation in minutiae maps derived from repeated noisy measurements from the same finger. Syndromes obtained via LDPC coding of these feature vectors serve as secure biometrics. In addition to providing very low false accept rates and low false reject rates, the design of the feature vectors ensures that distributed biometric coding is information-theoretically secure. We expect the benefits of syndrome-compatible feature vectors to extend to richer modalities such as ridge information in the case of fingerprint biometrics. This is the focus of our current work.

[1]While determining the FAR, if an input feature vector $\widehat{\mathbf{a}}$ satisfies the syndrome, then we count it as a false accept case. This is a conservative FAR estimate because any $\widehat{\mathbf{a}}$ for which $f_{\text{hash}}(\widehat{\mathbf{a}}) \neq f_{\text{hash}}(\mathbf{a})$ is denied acceptance.

References

[1] R. Ahlswede and I. Csiszar, "Common Randomness in Information Theory and Cryptography I: Secret Sharing," *IEEE Trans. Information Theory*, vol. 39, no. 4, pp. 1121–1132, July 1993.

[2] D. Slepian and J. K. Wolf, "Noiseless Coding of Correlated Information Sources," *IEEE Trans. Information Theory*, pp. 471–480, July 1973.

[3] 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.

[4] A. Juels and M. Sudan, "A Fuzzy Vault Scheme," in *IEEE Intl. Symp. on Information Theory*, 2002.

[5] T. C. Clancy, N. Kiyavash, and D. J. Lin, "Secure Smartcard-based Fingerprint Authentication," in *ACM SIGMM workshop on biometrics methods and applications*, 2003.

[6] S. Yang and I. Verbauwhede, "Automatic Secure Fingerprint Verification System based on Fuzzy Vault Scheme," in *IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, 2005, pp. 609–612.

[7] U. Uludag and A. Jain, "Fuzzy Fingerprint Vault," in *Workshop on Biometrics: Challenges Arising from Theory to Practice*, Aug. 2004, pp. 13–16.

[8] 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.

[9] ——, "Using Distributed Source Coding to Secure Fingerprint Biometrics," in *IEEE Conf. on Acoustics, Speech and Signal Processing*, 2007.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[11] T. Kevenaar, G. Schrijen, M. V. der Veen, A. Akkermans, and F. Zuo, "Face Recognition with Renewable and Privacy Preserving Binary Templates," *Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, pp. 21–26, 2005.