Privacy–Security Trade-Offs in Biometric Security Systems—Part I: Single Use Case

Lifeng Lai, *Member, IEEE*, Siu-Wai Ho, *Member, IEEE*, and H. Vincent Poor, *Fellow, IEEE*

***Abstract*—This is the first part of a two-part paper on the information theoretic study of biometric security systems. In this paper, the design of *single-use* biometric security systems is analyzed from an information theoretic perspective. A fundamental trade-off between privacy, measured by the normalized equivocation rate of the biometric measurements, and security, measured by the rate of the key generated from the biometric measurements, is identified. The privacy–security region, which characterizes the above-noted trade-off, is derived for this case. The scenario in which an attacker of the system has side information is then considered. Inner and outer bounds on the privacy–security region are derived in this case. Finally, biometric security systems with perfect privacy are studied, which is shown to be possible if and only if common randomness can be generated from two biometric measurements.**

***Index Terms*—Biometric, information theoretic security, perfect privacy, privacy–security trade-off, side information.**

I. Introduction

BIOMETRIC security systems have widespread applications. One typical example is a biometric authentication system, in which users' identities are verified by their biometric characteristics. Biometric characteristics are unique and do not change dramatically over time. The employment of biometric systems relieves the burden of selecting, memorizing, and protecting passwords.

There are usually two stages in a biometric authentication system: an enrollment stage and a release stage. In the enrollment stage, biometric characteristics, such as fingerprints, are sampled. The biometric measurements themselves or a transformation of the biometric measurements are stored in the database. In the release stage, the biometric characteristics are sampled again. The newly sampled biometric measurements are then used for authentication. Due to measurement noise or other factors such as aging or injury, two measurements of the same biometric characteristics will not produce the same result. Another issue in a biometric system is *privacy*. Biometric characteristics are stored in a certain form in the database, which leads to potential privacy leakage. For example, it has been shown that it is possible to recover fingerprints from minutiae points stored in the database [3]. Unlike passwords, biometric characteristics cannot be changed. Hence, if the database is compromised, irreversible identity theft is possible.

In recent years, there has been increasing research interest in addressing these issues. A number of interesting approaches have been proposed. For example, a secure sketch approach was studied in [4]–[6]. In the secure sketch approach, one stores a hash of the biometric information along with certain helper data that assists the recovery of biometric information from noisy observations during the release stage. Using results from error control coding, [7]–[9] developed practical coding schemes for the secure sketch approach. The security weaknesses of the secure sketch approach were studied in [10]. The cancelable biometric scheme was proposed in [11], in which an irreversible transformation of the biometric measurements is stored in the database [12]. Furthermore, the fuzzy vault scheme, in which keys are extracted from the biometric information and then used to encrypt secret information in the database, has been studied in [13]–[17]. The information theoretic analysis of these schemes can be found in [18]–[21]. The use of a cryptographic approach to protect the biometric template is studied in [22]. Developments in this area are summarized in [23] and [24]. Based on an information theoretic perspective, the basic idea of these approaches is to generate a secret key and helper data during an initial enrollment stage. A hash of the key is stored in the database for authentication purposes. The helper data is stored in the database. In the release stage, by combining the noisy measurements with the helper data, one can recover the key which is passed through the hash function and compared with the value stored in the database. The helper data can be viewed as the syndrome of an error correcting code, and the effects of noise can be mitigated by such error correction. The existing approaches focus on maximizing the rate of the key that can be recovered successfully from the noisy measurements. This approach is motivated by the fact that in an authentication system, the ability of an attacker to guess a correct value of the key and illegally gain access to the system is related to the rate of the key. From an information theoretic perspective, these existing approaches can be modelled as a problem of generating a secret key from common randomness [25]–[27], and hence the largest rate of the key can be characterized [20].[1] On the other hand, although

Manuscript received April 15, 2010; revised November 23, 2010; accepted December 04, 2010. Date of publication December 10, 2010; date of current version February 16, 2011. This work was supported in part by the National Science Foundation under Grant CCF-07-28208, CNS-09-05398 and Grant CCF-10-16671 and in part by the Australian Research Council under an Australian Postdoctoral Fellowship. Some of the results in this paper were presented at the Forty-Sixth Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, September 23–26, 2008, and the 3rd IAPR/IEEE International Conference on Biometrics, Sassari, Italy, June 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Arun Ross.

L. Lai is with the Department of Systems Engineering, University of Arkansas, Little Rock, AR 72204 USA (e-mail: lxlai@ualr.edu).

S.-W. Ho is with the Institute for Telecommunications Research, University of South Australia, Adelaide SA 5095, Australia (e-mail: SiuWai.Ho@unisa.edu.au).

H. V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Digital Object Identifier 10.1109/TIFS.2010.2098872

[1]In the general case in which the attacker also has access to correlated observations, the largest key rate has not been fully characterized.

the biometric measurements are not stored in the database in plain form, the helper data still contains information about the biometric measurements. It has been shown that in the existing approaches, the mutual information between the biometric measurements and the data stored in the database is $nH(X \mid Y)$ [7], where n is the length of the biometric measurement, X and Y are the biometric measurements taken during enrollment and release stages, respectively, and $H(\cdot \mid \cdot)$ denotes conditional entropy.

While the existing approaches maximize the key rate, they do not address the privacy issue adequately. In practice, the protection of the biometric measurements themselves is at least as important as maximizing the key rate. To increase the difficulty of the attacker to gain access to the system illegally, we would like to make the key rate as large as possible. On the other hand, to preserve the privacy, we need to ensure that information leakage about the biometric measurements themselves is as small as possible. One question naturally arises: can we maximize the security level while simultaneously minimizing the information leakage? This is the focus of this paper.

To illustrate the idea, consider the following match-on-card application. During the enrolment stage, a user's biometric characteristics are measured to generate a template. Suppose a biometric cryptosystem is used to protect the template [24]. Then helper data is stored on a card and a key is generated. During the authentication process, the user's biometric characteristics are measured again to generate a key with the assistance of the helper data. The authentication result relies on whether the same key can be generated. One of the security threats in this system is that the user's card can be dropped or stolen by adversaries. If this happens, the privacy of the user's biometric characteristics and the security of the key will be in danger. Therefore, it is important to investigate the effects of stolen helper data.

In this paper, by establishing an information theoretic foundation for single use biometric security systems,[2] we study the fundamental trade-off among security (the rate of the generated key), privacy (conditional entropy of the biometric measurements given the helper data), and key protection (conditional entropy of the key given the helper data) in any biometric security system. We note that whether or not it is preferable to generate secret keys from biometric measurements is subject to debate in the security community. Our work does not address this issue. Instead, we focus on characterizing the fundamental limits of biometric systems if such a practice is adopted. In order to increase the security level in an authentication system (e.g., a smaller false acceptance rate), a key with a larger rate is always preferred. As we will see, this will inevitably sacrifice the privacy or key protection in a system. More specifically, we first rigorously formulate different trade-offs in biometric security systems. We then identify and characterize these fundamental trade-offs for two different scenarios. In this paper, we assume that the attacker can access the database that contains the helper data but is not able to modify the entries. The scenario in which the attacker can modify the entries in the database is an interesting topic for future study.

In the first scenario, the trade-off between privacy and security is studied under the requirement that the helper data has arbitrarily small correlation with the key, i.e., perfect key protection. We consider both key-binding and key-generating biometric cryptosystems [24], which are also know as randomized and nonrandomized systems, respectively. In each system, we characterize the security–privacy trade-off. Furthermore, we propose schemes that fully achieve any particular point on the trade-off curve. We show that the performance of the existing approach is one particular point on the derived trade-off curve. We further show that the randomized and nonrandomized systems are equivalent in terms of privacy–security trade-off. After that, we study a generalized situation in which an attacker has side information about the biometric measurements. Both randomized and nonrandomized systems are considered. Inner and outer bounds on the privacy–security region are derived for these situations. These bounds are shown to match under certain conditions of interest.

In the second scenario, we study the ultimate goal that the helper data has an arbitrarily small correlation with the biometric characteristics, i.e., perfect privacy. We show this possibility when the key protection is not perfect and common randomness is shared between the biometric characteristics obtained during the enrollment and release stages. The trade-off between security and key protection is studied for both randomized and nonrandomized systems.

A line of related work is the key generation problem with rate constraint considered in [29]. Our work, on the other hand, can be viewed as a key generation problem with privacy constraints. It will become clear in the sequel that in the scenario considered in Section III-A, the key generation problem with privacy constraints can be converted to a key generation problem with rate constraints. And hence in this scenario, we can borrow results from [29] to obtain the optimal privacy–security trade-off. On the other hand, for other scenarios considered in the paper, the constraint on the privacy leakage is different from and is more involved than the rate constraint. We note that the privacy constraint adopted here is motivated by applications in biometric security systems, which have widespread applications. We also note that the scenario considered in Section III, which we presented in [1], also appeared independently and concurrently in [30] and [31] in a slightly different form. Compared with the work in [30] and [31], we also consider the scenario in which the attacker has side-information and the scenario with perfect privacy. On the other hand, [31] also considers several scenarios that are not considered in the current work. This paper focuses only on deriving theoretical bounds. These fundamental bounds illuminate the feasible system requirements. Also, they provide insights that are useful for the design of practical schemes.

The rest of the paper is organized as follows. In Section II, we introduce our system model and notation. Section III is devoted to the perfect key protection scenario. The situation in which the attack has side information is analyzed in Section IV. Next, we discuss the perfect privacy scenario in Section V. Finally, in Section VI, we offer some concluding remarks.

II. Model

We denote the biometric measurements taken during the enrollment stage by X^n and the biometric measurements taken during the release stage by Y^n. Here, we assume that X^n and Y^n are sequences with length n taking values from n-fold

[2]The scenario in which the biometric information is used in several locations is discussed in the second part of this two-part paper [28].

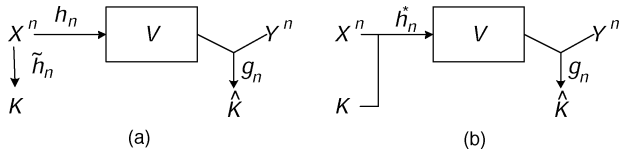


Fig. 1. Two different approaches for generating the key in biometric authentication systems: (a) nonrandomized approach; (b) randomized approach.

product sets $\mathcal{X}^n$ and $\mathcal{Y}^n$, respectively. Specific models for the distribution of biometric measurements can be found, for example, in [7], [8], and [32]. Further assumptions on statistical properties of X^n and Y^n will be given in Sections III and V as needed.

During the enrollment stage both a key K, ranging over $\mathcal{K}$, and helper data V, ranging over $\mathcal{V}$, are generated. In order to avoid technical difficulties that arise when the alphabets grow too fast [33], we assume without loss of practical interest that

$$\log|\mathcal{K}| = O(n) \tag{1}$$

and

$$\log|\mathcal{V}| = O(n). \tag{2}$$

The key K is used for authentication. The helper data V is stored in the database to assist the recovery of the key from the noisy measurements Y^n during the release stage. Regarding the generation of key K, we consider two types of systems that are widely used, namely nonrandomized systems [4], [5], [7] and randomized systems [14], [34]. In nonrandomized systems, as shown in Fig. 1(a), both V and K are generated from X^n by functions h_n and $\tilde{h}_n$, respectively, so that $V = h_n(X^n)$ and $K = \tilde{h}_n(X^n)$. In randomized systems, the key K, which is independent of X^n and Y^n in this case, is randomly generated during the enrollment stage. Then V is generated from the randomly chosen key K and the biometric measurements X^n by a function h_n^* so that $V = h_n^*(X^n, K)$. The randomized system is illustrated in Fig. 1(b). The randomized system and nonrandomized system are equivalent to key-binding and key-generating biometric cryptosystems, respectively [24].

During the release stage, by providing the noisy measurement Y^n and data stored in the database V, we generate an estimate $\hat{K}$ of the key. Let g_n be the recovery function, and thus $\hat{K} = g_n(Y^n, V)$. In order to allow access for legitimate users, we require an arbitrarily small error probability during the key recovery stage.

The following two performance metrics are important. The first one is the rate of the key generated, i.e., $n^{-1}H(K)$. We use the term "secrecy" to denote this quantity. As mentioned in the introduction, we would like to make this quantity large, since it is related to the difficulty of the attacker gaining access to the system. The second quantity of interest is the normalized equivocation level $H(X^n \,|\, V)/H(X^n)$. We use the term "privacy" to denote this term, and would like to make this term large.

A. Perfect Key Protection Systems

Although real biometric data do not obey independent and identically distributed (i.i.d.) or even ergodic statistics, biometric data can be transformed into binary vectors that are approximately i.i.d. Bernoulli (1/2) using various algorithms, e.g., the one in [32]. In this scenario, we further assume X^n and Y^n are generated according to a joint distribution (the case of sources with memory is studied in the perfect privacy system considered in Section V)

$$P_{X^nY^n}(x^n, y^n) = \prod_{i=1}^{n} P_{XY}(x_i, y_i). \tag{3}$$

We first consider perfect key protection systems, in which we require that V does not contain any information about the generated key. More specifically we require that for any $\epsilon > 0$, $n^{-1}I(K;V) \le \epsilon$ for sufficiently large n. As mentioned before, the difficulty of an attacker illegally accessing the system is related to the rate of the generated key, and hence we measure the security level of the system by $n^{-1}H(K)$. The privacy of the biometric measurements is defined as the normalized equivocation rate $H(X^n \,|\, V)/H(X^n)$. The larger this quantity, the greater the degree of privacy of the biometric measurements. Suppose this quantity can be made arbitrarily close to 1. Since X^n is generated from a stationary and memoryless source, $H(X^n \,|\, V)/H(X^n) \to 1$ is equivalent to $n^{-1}I(X^n;V) \to 0$ [35, Th. 4]. In this case, the adversary will suffer the maximum average symbol error probability and maximum block error probability even though the adversary knows V and has the best decoding function [35, Th. 3].

Definition 1 (Perfect Key Protection System): In a perfect key protection biometric authentication system, a privacy–security pair (Δ_P, R) is said to be achievable if, for each $\epsilon > 0$, there exist an integer n, coding functions, namely h_n and $\tilde{h}_n$ in nonrandomized systems (i.e., $K = \tilde{h}_n(X^n)$, $V = h_n(X^n)$) and h_n^* in randomized systems (i.e., $V = h_n^*(X^n, K)$), and a decoding function, namely g_n (i.e., $\hat{K} = g_n(V, Y^n)$), satisfying the following conditions:

$$n^{-1}H(K) \ge R \tag{4}$$

$$H(X^n \,|\, V)/H(X^n) \ge \Delta_P \tag{5}$$

$$n^{-1}I(V;K) \le \epsilon \tag{6}$$

and

$$\mathbb{P}[K \ne \hat{K}] \le \epsilon. \tag{7}$$

Another situation of interest is that in which, besides the data V stored in the database, an attacker of the system has side-information about the biometric characteristics. This models the situation in which the attacker obtains side-information from other sources, such as biometric characteristics stored in other databases or biometric characteristics from the relatives of the user. We denote the side observation at the attacker by Z^n, ranging in the set $\mathcal{Z}^n$, and assume that it is correlated with (X^n, Y^n). Following the definition in (3), we assume

$$P_{X^nY^nZ^n}(x^n, y^n, z^n) = \prod_{i=1}^{n} P_{XYZ}(x_i, y_i, z_i). \tag{8}$$

Since the attacker knows both V and Z^n, the privacy level is now measured as $H(X^n \,|\, VZ^n)/H(X^n)$, and the generated key is required to be independent of V and Z^n.

Definition 2 (Side-Information at Attacker): In a biometric system with side-information Z^n available to the attacker, a privacy–security pair (Δ_P, R) is said to be achievable if, for each $\epsilon > 0$, there exist an integer n, coding functions, namely h_n and $\tilde{h}_n$ in nonrandomized systems (i.e., $K = \tilde{h}_n(X^n)$, $V = h_n(X^n)$) and h_n^* in randomized systems

(i.e., $V = h_n^*(X^n, K)$), and a decoding function, namely g_n (i.e., $\hat{K} = g_n(V, Y^n)$), satisfying conditions (4)–(7) and

$$H(X^n \,|\, VZ^n)/H(X^n) \geq \Delta_P \tag{9}$$

and

$$n^{-1} I(VZ^n; K) \leq \epsilon. \tag{10}$$

B. Perfect Privacy System

In a perfect privacy system, we require that the data stored in the database does not leak any information about biometric measurements; that is, for each $\epsilon > 0$, we require $I(X^n; V) \leq \epsilon$ for sufficiently large n. At the same time, we generalize the requirement on the generated key, that is to allow $I(V; K)$ to range from 0 to $H(K)$. Of course, the smaller $I(V; K)$ the better. We measure the performance of a perfect privacy system by 1) the rate of the generated key $n^{-1}H(K)$, and 2) the normalized equivocation of the generated key $H(K \,|\, V)/H(K)$. If $H(K \,|\, V)/H(K) = 1$, we have $I(V; K) = 0$.

Here, we drop the i.i.d. assumption made in (3). In other words, we assume X^n and Y^n are two correlated random processes that are generated from sources with memory. This assumption is justified by the fact that the real biometric data tend not to be i.i.d. or ergodic [32]. Indeed, the intrinsic memory in X^n will be seen as a critical resource which makes a perfect privacy system possible.

Definition 3 (Perfect Privacy System): In a perfect privacy biometric security system, a rate-equivocation pair (R, Δ_s) is achievable if, for each $\epsilon > 0$, there exist an integer n, coding functions, namely h_n and $\tilde{h}_n$ in nonrandomized systems (i.e., $K = \tilde{h}_n(X^n)$, $V = h_n(X^n)$) and h_n^* in randomized systems (i.e., $V = h_n^*(X^n, K)$), and a decoding function, namely g_n (i.e., $\hat{K} = g_n(V, Y^n)$), satisfying the following conditions:

$$n^{-1} H(K) \geq R \tag{11}$$

$$I(X^n; V) \leq \epsilon \tag{12}$$

$$H(K \,|\, V)/H(K) \geq \Delta_s \tag{13}$$

and

$$\mathbb{P}[K \neq \hat{K}] \leq \epsilon. \tag{14}$$

III. Perfect Key Protection Case

In this section, we study perfect key protection systems, in which data stored in the database contains limited information about the generated key. Our goal is to characterize the relationship between the key size and information leakage about the biometric measurements.

A. Nonrandomized System

As discussed in Section II, in a nonrandomized system, both the key K and data V are generated from the biometric measurements X^n. Some existing schemes, for example, the secure sketch approach of [4] and [5] and the coding approach in [7], belong to this category.

This scenario can be converted to the problem of key generation with rate constraints considered in [29]. More specifically, the following problem is considered[3] in [29]. Two terminals, a source and a destination, possess correlated observations X^n and Y^n, respectively. These observations are generated according to a joint distribution $P_{X^n,Y^n}(x^n, y^n) = \prod_{i=1}^n P_{XY}(x_i, y_i)$. The source is allowed to send a message $f(X^n)$ to the destination with a rate constraint $(1/n)\log \|f\| \leq R_1$ over a noiseless public channel. After the transmission, the source generates a key $K = K(X^n)$, and the destination also generates a key $\hat{K} = \hat{K}(f, Y^n)$. The requirements are $\Pr\{K \neq \hat{K}\} \leq \epsilon$ for an arbitrarily small ϵ, and $(1/n)I(K; f) \leq \epsilon$. These requirements mean that the keys generated at the source and the destination are the same with high probability, and the message $f(X^n)$ transmitted over the public channel does not leak too much information about the generated keys. A number R is called achievable, if there exist a function f and random variables K and $\hat{K}$ satisfying the above mentioned conditions and $(1/n)H(K) \geq R - \delta$. Reference [29] has provided the following characterization of a set of achievable key rates as a function of the rate constraint R_1.

Theorem 1 ([29]): Within rate constraint R_1, a rate R is achievable if it satisfies the following condition:

$$\begin{aligned} & R = I(U; Y) \\ \text{s.t.} \quad & I(U; X) - I(U; Y) \leq R_1 \end{aligned}$$

for some auxiliary random variables U such that (U, X, Y) satisfies the Markov chain relationship $U \to X \to Y$.

Proof: Please refer to [29] for details of the proof. ■

One can now establish the connection between our scenario and the problem considered in [29] by setting $V = f(X^n)$ and converting the privacy constraint (5) to an equivalent rate constraint. From (5), we have $I(V; X^n) \leq (1 - \Delta_P)H(X^n) = n(1 - \Delta_P)H(X)$. Since in the nonrandomized scenario, V is a function of X^n, we have $H(V) \leq (1 - \Delta_P)H(X^n) = n(1 - \Delta_P)H(X)$. Without loss of generality, we can require V to be uniformly distributed, and hence we further have $(1/n)\log \|\mathcal{V}\| \leq (1 - \Delta_P)H(X)$. As a result, a Δ_P constraint is equivalent to a $(1 - \Delta_P)H(X)$ rate constraint. Thus, using Theorem 1, we have the following result.

Definition 4: $\mathcal{C}_N$ is the set of the privacy–security pairs (Δ_P, R) satisfying the following conditions:

$$\Delta_P \leq 1 - \frac{I(U; X) - I(U; Y)}{H(X)} \tag{15}$$

and

$$R \leq I(U; Y) \tag{16}$$

for some auxiliary random variable U such that (U, X, Y) satisfies the Markov chain condition $U \to X \to Y$.

Proposition 1: Any privacy–security pair (Δ_P, R) is achievable by a nonrandomized approach if and only if $(\Delta_P, R) \in \mathcal{C}_N$.

Remark 1: To maximize the rate of the key, we should set $U = X$. The rate of the key is then $I(X; Y)$. Correspondingly, the privacy level is $1 - H(X \,|\, Y)/H(X)$. This recovers the existing results of [18], [19], and [20].

Remark 2: In order to achieve both perfect privacy and perfect key protection, the auxiliary random variable U in (15)

[3]More general models are considered in [29]. Here, we cite only the model that is directly related to the problem under consideration.

should be chosen such that $I(U;X) = I(U;Y)$. The maximal rate achievable is then

$$\max_{U} I(U;Y) \quad \text{s.t.} \quad U \to X \to Y \quad \text{and} \quad I(U;X) = I(U;Y). \tag{17}$$

B. Randomized Approach

In randomized systems, during the enrollment stage, users have the freedom to choose the values of the keys but they are not required to remember them. For example, the fuzzy vault scheme studied in [14] and [34] belongs to this category. Here, the key K can be viewed as a source of additional randomness. The theorem below characterizes the performance of the randomized approach. The basic idea of the achievablity scheme is as follows. We first use the scheme in the nonrandomized approach to generate a key J, choosing from a set $\mathcal{J}$ with size $|\mathcal{J}|$. Then for a uniformly generated key K from a set $\mathcal{K}$, we store $J \oplus K$ in the database, along with other information required to be stored in the nonrandomized approach. Here $\oplus$ denotes mod-$|\mathcal{J}|$ addition. If we set $\mathcal{K} = \mathcal{J}$, $J \oplus K$ will be approximately uniformly (these terms will be made rigorous in the proof) distributed over $\mathcal{J}$, and is independent of other random variables of interest. Hence, this additional information stored in the database will not provide any information about the generated key and biometric measurements. In the release stage, we first obtain an estimate $\hat{J}$ of J using the same scheme as that of the nonrandomized approach. We then recover K via $J \oplus K \oplus \hat{J}$. Since $\hat{J} = J$ with high probability, $\hat{K}$ is equal to K with high probability. We show in the converse that the performance of the above mentioned scheme is optimal.

Let $\mathcal{C}_R$ denote the set of the privacy–security pairs (Δ_P, R) that are achievable by a randomized approach. We then have the following result.

Theorem 2: $\mathcal{C}_R = \mathcal{C}_N$.

Proof: Here we show that for any auxiliary random variable U with $U \to X \to Y$, and any $\epsilon_1 > 0$, the pair (Δ_P, R) with

$$\Delta_P = 1 - \frac{I(U;X) - I(U;Y)}{H(X)} - \epsilon_1$$

and

$$R = I(U;Y) - \epsilon_1 \tag{18}$$

is achievable. That is, any pair in the region $\mathcal{C}_N$ is achievable. The proof of the converse is presented in Appendix A.

For a given joint distribution $P_{UXY}(u,x,y) = P_{U|X}(u|x)P_{XY}(xy)$, we use the following scheme.

1) Code construction. Fix $\gamma > 0$, $\eta > 0$, and $\xi > 0$. Randomly select $M = 2^{n(I(U;X)+\gamma)}$ sequences U^n from $T^n_{[U],\xi|\mathcal{X}|}$, and divide them into $2^{n(I(U;X)-I(U;Y)+\gamma+\eta)}$ bins so that each bin contains $2^{n(I(U;Y)-\eta)}$ typical sequences. Following the notation in [36], $T^n_{[U],\xi|\mathcal{X}|}$ denotes the set of strongly typical U-sequences. We use L to denote the bin index, and J to denote the index of the sequence within each bin. Denote the set of these M sequences by $\mathcal{M}$. From the construction above, we can see that each sequence $u^n \in \mathcal{M}$ is uniquely identified by two indices $(l(u^n), j(u^n))$.

2) Enrollment stage. For each $x^n \in \mathcal{X}^n$, we associate a sequence $u^n \in \mathcal{M}$ with it by the following procedure. First, we find a list of sequences in $\mathcal{M}$ that are jointly typical with x^n. If there are more than one sequence in the list, we set u^n to be the one with the smallest index (we first compare the bin indices and if there is a tie, we then compare the index within the bin). If no such sequence exists, we set u^n to be the sequence with index $(l = 1, j = 1)$. Using this procedure, we associate every $x^n \in \mathcal{X}^n$ with a sequence $u^n \in \mathcal{M}$. Now, we randomly generate a key $K = k$ from the set $\mathcal{K} = \{1, \ldots, 2^{n(I(U;Y)-\eta)}\}$ with a uniform distribution. We then store the bin index $l(u^n)$ and $j(u^n) \oplus k$ in the database, in which $j(u^n)$ denotes the index of u^n in bin $l(u^n)$. Here $\oplus$ denotes mod-$2^{n(I(U;Y)-\eta)}$ addition. Hence, in this particular scheme $V = (L, J \oplus K)$. Also, we have

$$n^{-1} \log |\mathcal{K}| \le I(U;Y) - \eta. \tag{19}$$

3) Release stage. With the noisy measurement y^n, and the data stored in the database $(l, j \oplus k)$, we obtain an estimate $\hat{k}$ of k using the following procedure. We first look for a list of sequences in bin l that are jointly typical with y^n. Then, we obtain an estimate $\hat{u}^n$ of u^n as follows: 1) if there is only one sequence in the list, we set $\hat{u}^n$ equal to this sequence; 2) if there are more than one sequence in the list, we randomly choose one sequence from the list and set $\hat{u}^n$ equal to this sequence; and 3) if the list is empty, we set $\hat{u}^n$ to be the first sequence in bin l. Hence, for each $y^n \in \mathcal{Y}^n$, we have one $\hat{u}^n$ associated with it. We then set $\hat{k} = \hat{j} \oplus (j \oplus k)$.

4) Error probability analysis follows from standard techniques [37], and is omitted for this and other theorems in the paper for the sake of compactness.

5) Rate analysis. Since in our scheme, K is generated from $\mathcal{K} = \{1, \ldots, 2^{n(I(U;Y)-\eta)}\}$ with a uniform distribution, the rate of the key is

$$R = I(U;Y) - \eta. \tag{20}$$

6) Security analysis. For any u^n with $l(u^n) \neq 1$ and $j(u^n) \neq 1$, we have

$$\mathbb{P}[U^n = u^n] \le \sum_{x^n \in T^n_{[X|U],\xi}(u^n)} P^n_X(x^n) \le 2^{-n(I(U;X)-\zeta)}$$

in which ζ is a function of ξ, and approaches zero as ξ decreases.

Thus,

$$\begin{aligned} H(U^n) &= \sum_{u^n \in \mathcal{M}} -\mathbb{P}[U^n = u^n] \log(\mathbb{P}[U^n = u^n]) \\ &\ge \sum_{u^n \in \mathcal{M}} \mathbb{P}[U^n = u^n] n(I(U;X) - \zeta) \\ &= n(I(U;X) - \zeta). \end{aligned} \tag{21}$$

We also have

$$H(L) \le n(I(U;X) - I(U;Y) + \gamma + \eta) \tag{22}$$

since the value of l ranges from 1 to $2^{n(I(U;X)-I(U;Y)+\gamma+\eta)}$.

From the fact that $H(U^n) = H(L,J) = H(L) + H(J \mid L)$, we have

$$\begin{aligned} H(J) &\geq H(J \mid L) = H(U^n) - H(L) \\ &\geq n(I(U;Y) - \zeta - \gamma - \eta) \end{aligned} \tag{23}$$

in which we have used (21) and (22).

Thus we have

$$\begin{aligned} & n^{-1} I(K;V) \\ &\quad = n^{-1} I(K; L, K \oplus J) \\ &\quad = n^{-1}(H(L, K \oplus J) - H(L, K \oplus J \mid K)) \\ &\quad = n^{-1}(H(L) + H(K \oplus J \mid L) - H(L \mid K) \\ &\qquad - H(K \oplus J \mid K, L)) \\ &\quad = n^{-1}(H(K \oplus J \mid L) - H(J \mid L)) \\ &\quad \leq I(U;Y) - \eta) - (I(U;Y) - \zeta - \gamma - \eta) \\ &\quad \leq \gamma + \zeta \end{aligned} \tag{24}$$

where we have used (23) and the fact that the value of $k \oplus j$ ranges from 1 to $2^{n(I(U;Y)-\eta)}$.

7) Privacy analysis.

We can write

$$\begin{aligned} & H(X^n \mid L, J \oplus K) \\ &\quad = H(X^n, U^n \mid L, J \oplus K) - H(U^n \mid X^n, L, J \oplus K) \\ &\quad = H(U^n \mid L, J \oplus K) + H(X^n \mid U^n, L, J \oplus K) \\ &\qquad - H(U^n \mid X^n, L, J \oplus K) \\ &\quad \overset{(a)}{=} H(L, J \mid L, J \oplus K) + H(X^n \mid U^n, L, J \oplus K) \\ &\quad \overset{(b)}{\geq} nI(U;Y) - 2n(\zeta + \gamma + \eta) \\ &\qquad + H(X^n \mid U^n, L, J \oplus K) \\ &\quad \overset{(c)}{\geq} nI(U;Y) + H(X^n \mid U^n, J \oplus K) - 2n(\zeta + \gamma + \eta) \\ &\quad = nI(U;Y) + H(X^n \mid U^n) - I(X^n; J \oplus K \mid U^n) \\ &\qquad - 2n(\zeta + \gamma + \eta) \\ &\quad \overset{(d)}{=} nI(U;Y) + H(X^n) - H(U^n) - 2n(\zeta + \gamma + \eta) \\ &\quad \overset{(e)}{\geq} nI(U;Y) + nH(X) - nI(U;X) - n\gamma \\ &\qquad - 2n(\zeta + \gamma + \eta). \end{aligned} \tag{25}$$

Here (a) is due to the fact that there is a one-to-one correspondence between U^n and (L,J), and $H(U^n \mid X^n, L, J \oplus K) = 0$, since in our scheme U^n is a function of X^n; (b) is due to the fact that $H(L, J \mid L, J \oplus K) = H(J \mid L, J \oplus K) = H(J \mid L) - I(J; J \oplus K) \geq nI(U;Y) - 2n(\zeta + \gamma + \eta)$, due to (23) and the fact that $I(J; J \oplus K) \leq n(\zeta + \gamma + \eta)$, which can be easily shown; (c) is due to the fact that L is a function of U^n; (d) is due to the fact that $H(X^n \mid U^n) = H(X^n, U^n) - H(U^n) = H(X^n) - H(U^n)$, since in our scheme U^n is a function of X^n, and the fact that $I(X^n; J \oplus K \mid U^n) = H(J \oplus K \mid J, L) - H(J \oplus K \mid X^n, J, L) = 0$; and (e) is due to the fact that U^n takes at most $2^{n(I(U;X)+\gamma)}$ different values in our codebook.

On defining $\epsilon_1 = \max\{(\gamma + 2(\zeta+\gamma+\eta))/H(X), \eta, \gamma+\zeta\}$, from (19) (key size requirement), (20) (rate requirement), (24) (security requirement), and (25) (privacy requirement), we have that any pair (Δ_P, R) with

$$\begin{aligned} \Delta_P &= \frac{H(X^n \mid V)}{H(X^n)} \\ &\geq 1 - \frac{I(U;X) - I(U;Y)}{H(X)} - \epsilon_1 \end{aligned}$$

and

$$R \geq I(U;Y) - \epsilon_1 \tag{26}$$

is achieved by the presented scheme. The proof of the achievability part is thus complete. ■

Remark 3: Since $\mathcal{C}_N = \mathcal{C}_R$, we see that randomization does not increase the region. But one advantage of this randomized approach is that the system is revocable, meaning that different keys can be generated using the same scheme.

IV. Side-Information at the Attacker

In this section, we consider a situation in which, besides the data V stored in the database, the attacker has side-information about the biometric characteristics. This models the situation in which the attacker obtains side-information from other sources, such as biometric characteristics stored in other databases or biometric characteristics from relatives of the user.

A. Nonrandomized Approach

We first consider the nonrandomized approach, in which both V and K are functions of the biometric measurements X^n, i.e., $V = h_n(X^n)$ and $K = \tilde{h}_n(X^n)$.

We begin with a scheme that provides an inner bound on the set of all achievable privacy–security pairs. The basic idea is based on that of Proposition 1. We first generate a compressed version U^n of X^n, and then perform source coding with side information (U^n as the source sequence at the source coding encoder, and Y^n as the side information present at the decoder). That is we divide U^ns into bins, and store the bin index in the database. In Proposition 1, we set the key value as the index of U^n in each bin. Now since the attacker has additional information, the key rate should be reduced accordingly in order to guarantee that the attacker does not obtain any information about the generated key. We fulfill this goal by further partitioning each bin into subsets. We set the key value as the subset index. Using ideas from the analysis of the wiretap channel [38], it can be shown that there exists a partition such that even with the side information at the attacker and bin index, the attacker will not be able to infer too much information about the generated key (in this case, the key is the subset index). We then characterize the privacy leakage of this scheme. With the bin index and noisy information Y^n, we can recover U^n, and then recover the key by looking at the subset index of the recovered sequence U^n. Using information inequalities, we also provide an upper bound on the performance achievable by any scheme.

Theorem 3: Let $C_{s,\text{in}}$ be the set of (Δ_P, R) satisfying the following conditions:

$$\begin{aligned} \Delta_P \leq 1 & \\ - \frac{I(X;UZ) - I(U;Y \mid W) + I(U;Z \mid W)}{H(X)} & \end{aligned}$$

and

$$R \le I(U;Y\,|\,W) - I(U;Z\,|\,W) \tag{27}$$

and $C_{s,\text{out}}$ be the set of (Δ_P, R) satisfying the following conditions:

$$\Delta_P \le 1 - \frac{I(X;UZ) - I(U;Y) + I(U;Z\,|\,W)}{H(X)}$$

and

$$R \le I(U;Y\,|\,W) - I(U;Z\,|\,W) \tag{28}$$

in which W and U are auxiliary random variables such that (W,U,X,Y,Z) satisfies the following Markov chain condition $W \to U \to X \to (Y,Z)$.

Any pair in $C_{s,\text{in}}$ is achievable, while any pair outside of $C_{s,\text{out}}$ is not achievable.

Proof: Here we show that for any auxiliary random variables W and U with $W \to U \to X \to YZ$, and any $\epsilon^* > 0$, any pair (Δ_P, R) satisfying

$$\begin{aligned}\Delta_P &= 1 - \frac{I(X;UZ) - I(U;Y\,|\,W) + I(U;Z\,|\,W)}{H(X)} \\ &\quad - \epsilon^* \\ R &= I(U;Y\,|\,W) - I(U;Z\,|\,W) - \epsilon^*\end{aligned} \tag{29}$$

is achievable. That is, any pair in the region $\mathcal{C}_{s,\text{in}}$ is achievable. The proof the converse is presented in Appendix B.

Fix a joint distribution $P_{WUXYZ}(wuxyz) = P_W(w)P_{U\,|\,W}(u\,|\,w)P_{X\,|\,U}(x\,|\,u)P_{YZ\,|\,X}(yz\,|\,x)$, and use the following scheme.

1) Code construction. We fix ϕ, γ, η, δ, and ν to be positive real numbers. Randomly select a set $\Lambda_W \subset T^n_{[W]\delta}$ of typical sequences w^n with size $|\Lambda_W| = 2^{n(I(X;W)+\phi)}$. We arbitrarily order the sequences in Λ_W, and give an index, ranging from 1 to $2^{n(I(X;W)+\phi)}$, to each sequence. We also denote the sequence with index 1 by w_1^n. For each $w^n \in \Lambda_W$, randomly select a set $\Lambda_U(w^n) \subset T^n_{[U\,|\,W]\delta'}(w^n)$ of sequences u^n with size $|\Lambda_U(w^n)| = 2^{n(I(X;U\,|\,W)+\gamma)}$. For each set $\Lambda_U(w^n)$, we divide these sequences into $2^{n(I(U;X\,|\,W)-I(U;Y\,|\,W)+\gamma+\eta)}$ bins so that each bin contains $2^{n(I(U;Y\,|\,W)-\eta)}$ typical sequences. We further divide each bin into $2^{n(I(U;Y\,|\,W)-I(U;Z\,|\,W)-\eta+\nu)}$ subsets so that each subset contains $2^{n(I(U;Z\,|\,W)-\nu)}$ typical sequences. We use Q as the bin index, K as the index of the subset within each bin, and L as the index of the sequence within each subset. Then each sequence U^n can be uniquely identified by three indices (Q,K,L) and W^n.

2) Enrollment stage. For each $x^n \in \mathcal{X}^n$, we associate a u^n sequence with it using the following procedure. First, we find a sequence $w^n \in \Lambda_W$ such that (w^n, x^n) is jointly typical. If there is more than one sequence, we select w^n to be the one with the smallest index. If no such sequence exists, we choose w_1^n. After finding w^n, we make a list of $u^n \in \Lambda_U(w^n)$ such that (u^n, x^n) is jointly typical. If the list has more than one sequence, we select the one with the smallest index and associate it with x^n (we first compare Q; if there is a tie, then we compare K; if there is still a tie, then we compare L). If the list is empty, we set u^n as the sequence with index $(q=1, k=1, l=1)$ in $\Lambda_U(w_1^n)$, and associate it with x^n. After this procedure, each $x^n \in \mathcal{X}^n$ has a u^n associated with it. We set the key value to be the subset index k in which the sequence u^n falls. Hence, in this scheme $\mathcal{K} = \{1, \ldots, 2^{n(I(U;Y\,|\,W)-I(U;Z\,|\,W)-\eta+\nu)}\}$, and thus

$$n^{-1}\log|\mathcal{K}| \le I(U;Y\,|\,W) - I(U;Z\,|\,W) - \eta + \nu. \tag{30}$$

We store w^n and the bin index q in the database. Therefore, in this particular scheme $V = (W^n, Q)$.

3) Release stage. With the noisy measurement y^n, and the data stored in the database (w^n, q), we obtain an estimate $\hat{k}$ of k using the following procedure. We first make a list of sequences in bin q of $\Lambda_U(w^n)$ that are jointly typical with y^n. Then, we obtain an estimate $\hat{u}^n$ of u^n as follows: 1) if there is only one sequence in the list, we set $\hat{u}^n$ equal to this sequence; 2) if there are more than one sequence in the list, we randomly choose one sequence from the list and set $\hat{u}^n$ equal to this sequence; and 3) if the list is empty, we set $\hat{u}^n$ to be the first sequence in bin q of $\Lambda_U(w^n)$. Hence, for each $y^n \in \mathcal{Y}^n$, we have a $\hat{u}^n$ associated with it. We then obtain an estimate of the key $\hat{k}$, by setting it as the subset index of $\hat{u}^n$ in the bin q of $\Lambda_U(w^n)$.

4) Rate analysis. For any u^n that is not the first sequence in $\Lambda_U(w_1^n)$, we have

$$\begin{aligned}\mathbb{P}[U^n = u^n] &\le \sum_{x^n \in T^n_{[X\,|\,UW],\xi}(u^n, w^n)} P_X^n(x^n) \\ &\le 2^{(-n(I(WU;X)+\varsigma_\xi))}\end{aligned} \tag{31}$$

in which ς_ξ is a function of ξ, and goes to zero as ξ decreases.

Hence, there exists a $\zeta > 0$, which is again a function of ξ and goes to zero as ξ decreases, such that

$$n^{-1}H(U^n) \ge I(WU;X) - \zeta. \tag{32}$$

On the other hand, $H(W^n) \le n(I(X;W)+\phi)$, since the codebook contains only $2^{n(I(X;W)+\phi)}$ different w^ns. Similarly, we have $H(Q) \le n(I(U;X\,|\,W) - I(U;Y\,|\,W) + \gamma + \eta)$, and $H(L) \le n(I(U;Z\,|\,W) - \nu)$. Thus, we have

$$\begin{aligned}R &= n^{-1}H(K) \\ &\ge n^{-1}H(K\,|\,W^n, Q, L) \\ &= n^{-1}(H(W^n, K, Q, L) - H(W^n, Q, L)) \\ &\overset{(a)}{=} n^{-1}(H(U^n) - H(W^n, Q, L)) \\ &\ge n^{-1}(H(U^n) - H(W^n) - H(Q) - H(L)) \\ &\ge I(WU;X) - \zeta - (I(X;W) + \phi) \\ &\quad - (I(U;X\,|\,W) - I(U;Y\,|\,W) + \gamma + \eta) \\ &\quad - (I(U;Z\,|\,W) - \nu) \\ &= I(U;Y\,|\,W) - I(U;Z\,|\,W) \\ &\quad - \zeta - \phi - \gamma - \eta + \nu\end{aligned} \tag{33}$$

in which (a) is due to the fact that there is a one-to-one correspondence between U^n and (W^n, Q, K, L). From (33), it follows that the rate of the key is larger than $I(U;Y\,|\,W) - I(U;Z\,|\,W) - \epsilon$ for a suitable parameter ϵ.

6) Security analysis. In the following, we bound $I(K;VZ^n)$. First, we have

$$\begin{aligned}
&H(K\,|\,VZ^n)\\
&= H(K,V,Z^n) - H(VZ^n)\\
&= H(K,V,Z^n,U^n)\\
&\quad - H(U^n\,|\,K,V,Z^n) - H(VZ^n)\\
&= H(K,V,U^n) + H(Z^n\,|\,U^n,K,V)\\
&\quad - H(U^n\,|\,K,V,Z^n) - H(VZ^n)\\
&\overset{(a)}{=} H(U^n) + H(Z^n\,|\,U^n)\\
&\quad - H(U^n\,|\,K,V,Z^n) - H(V)\\
&\quad - H(Z^n\,|\,V)\\
&\geq H(U^n) + H(Z^n\,|\,U^n)\\
&\quad - H(U^n\,|\,K,V,Z^n) - H(V)\\
&\quad - H(Z^n\,|\,W^n)\\
&\overset{(b)}{\geq} n(I(X;UW)-\zeta) + n(H(Z\,|\,U)-\varsigma) - n\delta_n\\
&\quad - n(I(X;W) + I(U;X\,|\,W)\\
&\quad - I(U;Y\,|\,W) + \phi + \gamma + \eta)\\
&\quad - n(H(Z\,|\,W) + \epsilon_n)\\
&\overset{(c)}{=} n(I(U;Y\,|\,W) - I(U;Z\,|\,W) - \epsilon). \qquad (34)
\end{aligned}$$

Here (a) is due to the fact that K and V are functions of U^n, and (b) is due to the following facts: 1) $H(U^n) \geq n(I(WU;X) - \zeta)$, which was shown in (32); 2) $H(U^n\,|\,K,V,Z^n) \leq n\delta_n$, which will be shown in Lemma 1 of Appendix C; 3) $H(V) \leq H(W^n)+H(Q) \leq n(I(X;W)+I(U;X\,|\,W)-I(U;Y\,|\,W)+\phi+\gamma+\eta)$; 4) $H(Z^n\,|\,W^n) \leq n(H(Z\,|\,W)+\epsilon_1)$ with ϵ_1 goes to 0 as n increases, which will be shown in the Lemma 2 of Appendix C; and 5) $H(Z^n\,|\,U^n) \geq n(H(Z\,|\,U)-\varsigma)$ which can be shown similarly as in Lemma 3 of Appendix C. In (c), we define $\epsilon = \zeta + \varsigma + \delta_n + \epsilon_n + \phi + \gamma + \eta$.

Thus

$$\begin{aligned}
&n^{-1}I(K;VZ^n)\\
&\quad = n^{-1}(H(K) - H(K\,|\,VZ^n))\\
&\quad \overset{(a)}{\leq} I(U;Y\,|\,W) - I(U;Z\,|\,W) - \eta + \nu\\
&\qquad - (I(U;Y\,|\,W) - I(U;Z\,|\,W) - \epsilon)\\
&\quad = \nu - \eta + \epsilon
\end{aligned}$$

in which (a) follows from (34) and the fact that the value of K ranges from 1 to $2^{n(I(U;Y\,|\,W)-I(U;Z\,|\,W)-\eta+\nu)}$.

7) Privacy analysis

We have

$$\begin{aligned}
&H(X^n\,|\,V,Z^n)\\
&\quad = H(X^n,U^n\,|\,V,Z^n) - H(U^n\,|\,V,X^n,Z^n)\\
&\quad = H(U^n\,|\,V,Z^n) + H(X^n\,|\,U^n,V,Z^n)\\
&\qquad - H(U^n\,|\,X^n,V,Z^n)\\
&\quad \overset{(a)}{\geq} n(I(U;Y\,|\,W) - I(U;Z\,|\,W) - \epsilon)\\
&\qquad + H(X^n\,|\,U^n,V,Z^n)\\
&\quad \overset{(b)}{=} n(I(U;Y\,|\,W) - I(U;Z\,|\,W))\\
&\qquad + H(X^n\,|\,U^n,Z^n,W^n) - n\epsilon\\
&\quad \overset{(c)}{\geq} n((1-\delta_n)H(X) - ((1-\delta_n)I(X;UZW)\\
&\qquad - (I(U;Y\,|\,W) - I(U;Z\,|\,W))) - 2n\delta_n - n\epsilon\\
&\quad \overset{(d)}{=} n((1-\delta_n)H(X) - ((1-\delta_n)I(X;UZ)\\
&\qquad - I(U;Y\,|\,W) + I(U;Z\,|\,W))) - 2n\delta_n - n\epsilon. \qquad (35)
\end{aligned}$$

Here, (a) is due to 1) the inequality $H(U^n\,|\,VZ^n) \geq H(K\,|\,VZ^n)$ and (34) and 2) the fact that U^n is a function of X^n in our scheme; (b) is due to the fact that $V = (W^n, Q)$ where Q is a function of U^n; (c) is due to Lemma 3 of Appendix C; and (d) is due to the Markov chain relationship $W \to U \to X$.

On defining $\epsilon^* = (2 + H(X))\delta_n + \epsilon$, from (30) (key size requirement), (33) (rate requirement), (34) (security requirement), and (35) (privacy requirement), we have that any pair (Δ_P, R) satisfying

$$\begin{aligned}
\Delta_P &\geq 1 - \frac{I(X;UZ) - I(U;Y\,|\,W) + I(U;Z\,|\,W)}{H(X)}\\
&\quad - \epsilon^*\\
R &\geq I(U;Y\,|\,W) - I(U;Z\,|\,W) - \epsilon^* \qquad (36)
\end{aligned}$$

is achieved by the presented scheme. The proof of the achievability part is thus complete.

Remark 4: In general, these two bounds do not match. If the attacker does not have side information, that is $\mathcal{Z} = \Phi$, then the lower bound does match the upper-bound. Furthermore, the result recovers that of Theorem 1, since, if $\mathcal{Z} = \Phi$, the lower bound becomes

$$\Delta_P \leq 1 - \frac{I(X;U) - I(U;Y\,|\,W)}{H(X)} \qquad (37)$$

and

$$R \leq I(U;Y\,|\,W) \qquad (38)$$

and the upper bound becomes

$$\Delta_P \leq 1 - \frac{I(X;U) - I(U;Y)}{H(X)} \qquad (39)$$

and

$$R \leq I(U;Y\,|\,W). \qquad (40)$$

Since $W \to U \to Y$, we have $I(U;Y\,|\,W) \leq I(U;Y)$, in which the equality can be achieved by setting W to be a constant. Thus, choosing W as a constant maximizes both R and Δ_P simultaneously in both the lower and upper-bounds. Furthermore, when we choose W to be a constant, these two bounds match.

B. Randomized Approach

As in Section III-B, during the enrollment stage, the key K is randomly generated and is independent of X^n. The helper data V is a function of K and X^n; that is $V = h_n^*(K, X^n)$.

An achievable region is described by the following scheme. The basic idea is to first generate a key J, choosing from a set $\mathcal{J}$ with size $|\mathcal{J}|$, using the scheme in the proof of Theorem 3. Then for a uniformly generated key K from a set $\mathcal{K}$, we store $J \oplus K$ in the database, along with other information required

to be stored in Theorem 3. Here $\oplus$ denotes mod-$|\mathcal{J}|$ addition. If we set $\mathcal{K} = \mathcal{J}$, $J \oplus K$ will be approximately uniformly (these terms will be made rigorous in the proof) distributed over $\mathcal{J}$, and is independent of other random variables of interest. Hence, this additional information stored in the database will not provide any information about the generated key and biometric measurements. In the release stage, we first obtain an estimate $\hat{J}$ of J using the same scheme as that of Theorem 3. We then recover K via $J \oplus K \oplus \hat{J}$. Since $\hat{J} = J$ with high probability, $\hat{K}$ is equal to K with high probability. Using information theoretic inequalities, we also provide an upper-bound on the achievable privacy–security pairs.

Theorem 4: Let $C_{sr,\text{in}}$ be the set of (Δ_P, R) pairs satisfying the following conditions:

$$\Delta_P \leq 1 - \frac{I(X;UZ) - I(U;Y\,|\,W) + I(U;Z\,|\,W)}{H(X)}$$

and

$$R \leq I(U;Y\,|\,W) - I(U;Z\,|\,W) \tag{41}$$

and let $C_{sr,\text{out}}$ be the set of (Δ_P, R) pair satisfying the following conditions:

$$\Delta_P \leq 1 - \frac{I(X;Z\,|\,U) - I(U;Y\,|\,W) + I(U;Z\,|\,W)}{H(X)}$$

and

$$R \leq I(U;Y\,|\,W) - I(U;Z\,|\,W) \tag{42}$$

in which W and U are auxiliary random variables such that (W,U,X,Y,Z) satisfies the following Markov chain condition $W \to U \to X \to (Y,Z)$.

Then any pair in $C_{sr,\text{in}}$ is achievable, while any pair outside of $C_{sr,\text{out}}$ is not achievable.

Proof: Here we show that for any auxiliary random variable W and U with $W \to U \to X \to YZ$, and any $\epsilon^* > 0$, any pair (Δ_P, R) satisfying

$$\Delta_P = 1 - \frac{I(X;UZ) - I(U;Y\,|\,W) + I(U;Z\,|\,W)}{H(X)} - \epsilon^*$$
$$R = I(U;Y\,|\,W) - I(U;Z\,|\,W) - \epsilon^* \tag{43}$$

is achievable. That is, any pair in the region $\mathcal{C}_{sr,\text{in}}$ is achievable. The proof of the converse is presented in Appendix D.

Fix a joint distribution $P_{WUXYZ}(wuxyz) = P_W(w)P_{U\,|\,W}(u\,|\,w)P_{X\,|\,U}(x\,|\,u)P_{YZ\,|\,X}(yz\,|\,x)$, and use the following scheme.

1) Code Construction: Fix ϕ, γ, η, δ, and ν to be positive real numbers. Randomly select a set $\Lambda_W \subset T^n_{[W]\delta}$ of typical sequences w^n with size $|\Lambda_W| = 2^{n(I(X;W)+\phi)}$. We arbitrarily order the sequences in Λ_W, and give an index ranging from 1 to $2^{n(I(X;W)+\phi)}$ to each sequence. We also denote the sequence with index 1 by w_1^n. For each $w^n \in \Lambda_W$, randomly select a set $\Lambda_U(w^n) \subset T^n_{[U\,|\,W]\delta'}(w^n)$ of sequences u^n with size $|\Lambda_U(w^n)| = 2^{n(I(X;U\,|\,W)+\gamma)}$. For each set $\Lambda_U(w^n)$ we divide the sequences into $2^{n(I(U;X\,|\,W)-I(U;Y\,|\,W)+\gamma+\eta)}$ bins so that each bin contains $2^{n(I(U;Y\,|\,W)-\eta)}$ typical sequences. We further divide each bin into $2^{n(I(U;Y\,|\,W)-I(U;Z\,|\,W)-\eta+\nu)}$ subsets so that each subset contains $2^{n(I(U;Z\,|\,W)-\nu)}$ typical sequences. We use Q as the bin index, J as the index of the subset within each bin, and L as the index of the sequence within each subset. Then each sequence U^n can be uniquely identified by three indices (Q,J,L) and W^n.

2) Enrollment Stage: For each $x^n \in \mathcal{X}^n$, we associate a u^n sequence with it using the following procedure. First, we find a sequence $w^n \in \Lambda_W$ such that (w^n, x^n) is jointly typical. If there is more than one sequence, we select w^n to be the one with the smallest index. If no such sequence exists, we choose w_1^n. After finding w^n, we find a list of $u^n \in \Lambda_U(w^n)$ such that (u^n, x^n) is jointly typical. If the list has more than one sequence, we select the one with the smallest index and associate it with x^n (we first compare Q; if there is a tie, then we compare J; if there is still a tie, then we compare L). If the list is empty, we set u^n as the sequence with index $(q=1, j=1, l=1)$ in $\Lambda_U(w_1^n)$, and associate it with x^n. After this procedure, each $x^n \in \mathcal{X}^n$ has a u^n associated with it. We now randomly generate a key K from the set $\mathcal{K} = \{1, \ldots, 2^{n(I(U;Y\,|\,W)-I(U;Z\,|\,W)-\eta+\nu)}\}$ with a uniform distribution. We store W^n, the bin index Q and $J \oplus K$ in the database. Here $\oplus$ denotes mod-$2^{n(I(U;Y\,|\,W)-I(U;Z\,|\,W)-\eta+\nu)}$ addition. Hence, in this particular scenario $V = (W^n, Q, J \oplus K)$.

3) Release Stage: With the noisy measurement y^n, and the data stored in the database $(w^n, q, j \oplus k)$, we obtain an estimate $\hat{k}$ of k using the following procedure. We first look for a list of sequences in bin q of $\Lambda_U(w^n)$ that are jointly typical with y^n. Then, we obtain an estimate $\hat{u}^n$ of u^n as follows: 1) if there is only one sequence in the list, we set $\hat{u}^n$ equal to this sequence; 2) if there is more than one sequence in the list, we randomly choose one sequence from the list and set $\hat{u}^n$ equal to this sequence; and 3) if the list is empty, we set $\hat{u}^n$ to be the first sequence in bin q of $\Lambda_U(w^n)$. Hence, for each $y^n \in \mathcal{Y}^n$, we have one $\hat{u}^n$ associated with it. We then obtain an estimate of the key $\hat{k}$, namely $\hat{k} = \hat{j} \oplus (j \oplus k)$, in which $\hat{j}$ is the subset index of $\hat{u}^n$ in the bin q of $\Lambda_U(w^n)$.

The error probability, rate, security, and privacy analysis follow similarly to those in the proof of Theorem 3, and we omit the analysis of these quantities for the sake of compactness. ■

V. Perfect Privacy for Biometric Measurements

In this section, we consider the perfect privacy case, in which we require that the mutual information between the data stored in the database and biometric measurements be arbitrarily small. This models the situation in which privacy is of primary concern. In the following, we show a close relationship between perfect privacy and common random processes. In this section, the results apply to finite n. Also, we assume X^n is generated from a source with memory since real biometric data are not well modeled as having i.i.d. statistics [32]. Therefore, the following definition is different from and more general than the common randomness for discrete memoryless stationary information sources as shown in [29], [36, p. 402], and [39].

Definition 5: For two random processes X^n and Y^n, there exists a common random process between them with entropy rate not less than α if for each $\eta > 0$, there exist n and functions ψ_n of X^n and ϕ_n of Y^n such that

$$\mathbb{P}[\psi_n(X^n) \neq \phi_n(Y^n)] \leq \eta \tag{44}$$

and

$$n^{-1}H(\psi_n(X^n)) \geq \alpha - \eta. \tag{45}$$

This definition says that if X^n and Y^n have a common random process with entropy rate α, then one can generate two random variables: $\psi_n(X^n)$ solely based on X^n and $\phi_n(Y^n)$ solely based on Y^n, with the property that each of these two random variables has entropy $n\alpha$ and is equal to the other one with high probability. Now, the common random process and the perfect privacy case are connected in the following theorem.

Theorem 5: A privacy-rate pair (R, Δ_s) is achievable with perfect privacy if, and only if, there exists a common random process between X^n and Y^n with entropy rate not less than $R\Delta_s$.

Proof: A proof of the necessity of common randomness for perfect privacy is given in Appendix E. Here, we prove the sufficiency. For any $\eta > 0$, there exist $\psi_n(X^n)$ and $\phi_n(Y^n)$ such that (44) and (45) are satisfied. Let $\alpha = n^{-1}H(\psi_n(X^n))$ so that $\alpha \geq R\Delta_S - \eta$. If $R \leq \alpha$, let $L = \psi_n(X^n)$, $L' = \phi_n(Y^n)$, $\hat{K} = L' \oplus V$, and $V = L \oplus K$, where K is randomly generated such that

$$n^{-1}H(K) = n^{-1}H(L) = \alpha \geq R$$

and K is independent of X^n and Y^n. In this case, $(H(K\,|\,V))/(H(K)) = 1 \geq \Delta_s$. If $R > \alpha$, then let $\beta = R - \alpha$. In this case, let $K = (K_1, K_2)$ be independent of X^n and Y^n and randomly generated such that $n^{-1}H(K_1) = \beta$ and

$$n^{-1}H(K_2) = n^{-1}H(L) = \alpha.$$

Here, K_1 is independent of K_2. Let $V = (K_1, L \oplus K_2)$ and $\hat{K} = (K_1, L' \oplus L \oplus K_2)$. Then $n^{-1}H(K) = \alpha + \beta = R$ and

$$\frac{H(K\,|\,V)}{H(K)} = \frac{n^{-1}H(\psi_n(X^n))}{n^{-1}H(K)} \geq \frac{R\Delta_s - \eta}{R} = \Delta_s - \frac{\eta}{R}. \tag{46}$$

In both cases, it is obvious that $I(X^n; V) = 0$ and $\mathbb{P}[K \neq \hat{K}] \leq \mathbb{P}[\psi_n(X^n) \neq \phi_n(Y^n)] \leq \eta$. Since $\eta > 0$ is arbitrary, the privacy-rate pair (R, Δ_s) is achievable. ∎

Together with Theorem 5, we have shown that a privacy-rate pair (R, Δ_s) is achievable if and only if there exists a common random process between X^n and Y^n with entropy rate not less than $R\Delta_s$.

Now consider nonrandomized systems, in which $H(K\,|\,X^n) = 0$ since K is a function of X^n. Thus in this case, $I(X^n; V) \leq \epsilon$ implies that $I(K; V) \leq \epsilon$. Since $\epsilon > 0$ is arbitrary, $\Delta_s = 1$. Hence in nonrandomized systems, perfect privacy means perfect key protection. By letting $K = \psi_n(X^n)$, $\hat{K} = \phi_n(Y^n)$, and V be a constant, we can prove that privacy-rate pairs $(R, 1)$ for all R are achievable by a nonrandomized system if there exists a common random process between X^n and Y^n with entropy rate not less than R. Thus the achievable privacy-rate region of nonrandomized systems is a proper subset of that for randomized systems. In other words, the randomized approach provides the flexibility to determine a system that has perfect privacy but not perfect secrecy.

Our results reveal that the possibility of building a biometric security system with perfect privacy depends on whether common information between two biometric measurements can be generated. The common randomness for discrete memoryless stationary information sources has been studied in [29], [36, p. 402], and [39]. When $P_{X^nY^n}$ satisfies (3), the common randomness is 0 if P_{XY} is indecomposable [36, p. 403]. This result has been extended to nonstationary independent sources [40]. Indeed, the results may be extendable to stationary ergodic sources according to the proofs in [39]. Therefore, if the biometric measurements are converted to i.i.d. or ergodic sequences, it is unlikely that one can build a system with perfect privacy. Fortunately, real biometric measurements do not follow i.i.d. or ergodic statistics [32]. Developing techniques for generating common randomness from two biometric measurements is thus of interest in the development of biometric security systems allowing perfect privacy.

VI. Conclusion

The design of single-use biometric security systems has been studied under a privacy–security trade-off framework. Two different scenarios, in which the attacker either has side-information about the biometric measurements or not, have been considered. In the scenario for which the attacker does not have side-information, we have considered two cases of perfect key protection and perfect privacy. In both cases, the complete privacy–security region has been identified. More specifically, an upper-bound on the privacy–security pair achievable by any scheme has been derived. Moreover, a scheme has been proposed to achieve this upper bound. For the scenario in which the attacker has side-information about the biometric measurements, inner and outer bounds on the privacy–security region have been derived. We have also shown the close relationship between perfect privacy and common randomness between two biometric measurements. The possibility of building a biometric security system with perfect privacy relies on whether common randomness can be generated from two biometric measurements.

The extension of these ideas to the design of biometric security systems in which biometric information is used in several locations is discussed in the second part of this two-part paper [28]. Several other interesting questions arise from our work as well. For example, designing practical codes that achieve the derived theoretical bounds is a natural next step. Moreover, deriving tighter bounds for the side-information case and the characterization of the performance for finite block lengths are of interest. In addition, the study of more advanced attacker models in which the attacker can modify the entries in the database is important.

Appendix A
Proof of Theorem 2

We now show the converse result that $\mathcal{C}_N$ is exactly the privacy–security region. To do so, we let (Δ_P, R) be a privacy–security pair achieved by using encoding functions h_n^* and decoding function g_n. Then $V = h_n^*(X^n, K)$ and $\mathbb{P}[K \neq g_n(Y^n)] \leq \epsilon$. In the following, we will show that there exists a random variable U with $U \to X \to Y$, such that

$$\Delta_P \leq 1 - \frac{I(U;X) - I(U;Y)}{H(X)} + \epsilon_n \tag{47}$$

and

$$R \le I(U;Y) + \epsilon_n \tag{48}$$

in which ϵ_n approaches 0 as n increases. That is $(\Delta_P, R) \in \mathcal{C}_N$.

First, from the conditions $\mathbb{P}[K \neq g_n(Y^n, V)] \le \epsilon$, (1), and Fano's inequality, we know that there exists a sequence of δ_n that approaches 0 as n increases, such that

$$H(K \,|\, Y^n, V) \le n\delta_n. \tag{49}$$

Now, we bound the privacy leakage as follows:

$$\begin{aligned} H(X^n \,|\, V) &\le H(X^n, K \,|\, V) \\ &= H(X^n, K) - I(X^n, K; V) \\ &= H(X^n) + H(K) - H(V) \\ &\quad + H(V \,|\, X^n, K). \end{aligned} \tag{50}$$

Now

$$\begin{aligned} 0 &\ge H(V \,|\, Y^n) - H(V) \\ &= H(V, K \,|\, Y^n) - H(K \,|\, V, Y^n) - H(V) \\ &\ge H(V, K \,|\, Y^n) - H(V) - n\delta_n \\ &= H(Y^n \,|\, V, K) + H(V, K) - H(Y^n) - H(V) - n\delta_n \\ &= H(Y^n \,|\, V, K) + H(K \,|\, V) - H(Y^n) - n\delta_n \\ &= H(Y^n \,|\, V, K) + H(K) - H(Y^n) - I(K; V) - n\delta_n \\ &\ge H(Y^n \,|\, V, K) + H(K) - H(Y^n) - n\delta_n - n\epsilon \end{aligned} \tag{51}$$

due to the requirement that $I(K; V) \le n\epsilon$.

Thus, subtracting (51) from (50), we have

$$\begin{aligned} H(X^n \,|\, V) \le H(X^n) + H(Y^n) - H(Y^n \,|\, V, K) \\ - H(V) + n\delta_n + n\epsilon. \end{aligned} \tag{52}$$

We also have

$$\begin{aligned} H(V) &\ge I(V; X^n, K) \\ &= H(X^n, K) - H(X^n, K \,|\, V) \\ &= H(X^n, K) - H(K \,|\, V) - H(X^n \,|\, K, V) \\ &\ge H(X^n) - H(X^n \,|\, K, V) \\ &= \sum_{i=1}^{n} \{H(X_i) - H(X_i \,|\, X^{i-1}, K, V)\} \\ &= \sum_{i=1}^{n} \{H(X_i) - H(X_i \,|\, X^{i-1}, Y^{i-1}, K, V)\} \end{aligned}$$

which is due to the fact that $Y^{i-1} \to (X^{i-1}, K, V) \to X_i$. To show this Markov chain relationship, we first note that $Y^{i-1} \to (X^{i-1}, K) \to (X^n, K, X_i)$, from which we have $Y^{i-1} \to (X^{i-1}, K) \to (X^n, K, X_i) \to (V, X_i)$, because $V = f(X^n, K)$. Now, we have $Y^{i-1} \to (X^{i-1}, K) \to (V, X_i)$, which leads to $Y^{i-1} \to (X^{i-1}, K, V) \to X_i$.

We continue as follows:

$$\begin{aligned} H(V) &\ge \sum_{i=1}^{n} \{H(X_i) - H(X_i \,|\, X^{i-1}, Y^{i-1}, K, V)\} \\ &\ge \sum_{i=1}^{n} \{H(X_i) - H(X_i \,|\, Y^{i-1}, K, V)\} \\ &= \sum_{i=1}^{n} I(Y^{i-1}, K, V; X_i). \end{aligned} \tag{53}$$

Hence

$$\begin{aligned} &H(X^n \,|\, V) \\ &\le H(X^n) + H(Y^n) - H(Y^n \,|\, V, K) - H(V) \\ &\quad + n\delta_n + n\epsilon \\ &\overset{(a)}{\le} \sum_{i=1}^{n} \{H(X_i) + H(Y_i) - H(Y_i \,|\, V, K, Y^{i-1}) \\ &\qquad - I(Y^{i-1}, K, V; X_i)\} + n\delta_n + n\epsilon \\ &= \sum_{i=1}^{n} \{H(X_i) + I(V, K, Y^{i-1}; Y_i) \\ &\qquad - I(Y^{i-1}, K, V; X_i)\} + n\delta_n + n\epsilon \\ &\overset{(b)}{=} \sum_{i=1}^{n} \{H(X_i) + I(U_i; Y_i) - I(U_i; X_i)\} \\ &\quad + n\delta_n + n\epsilon. \end{aligned} \tag{54}$$

Here, in (a), we have used (53), and in (b), we have set $U_i = (Y^{i-1}, K, V)$.

Moreover, we have

$$\begin{aligned} H(K) &= I(K; VY^n) + H(K \,|\, VY^n) \\ &\overset{(a)}{\le} I(K; VY^n) + n\delta_n \\ &= I(K; V) + I(K; Y^n \,|\, V) + n\delta_n \\ &\le n\epsilon + n\delta_n + \sum_{i=1}^{n} I(K; Y_i \,|\, Y^{i-1} V) \\ &\le n\epsilon + n\delta_n + \sum_{i=1}^{n} I(K, Y^{i-1}, V; Y_i) \\ &= n\epsilon + n\delta_n + \sum_{i=1}^{n} I(U_i; Y_i) \end{aligned} \tag{55}$$

in which (a) is due to (49).

Now, by introducing a random variable T uniformly distributed over the set $\{1, \ldots, n\}$, and setting $U = (U_T, T)$, $X = X_T$, $Y = Y_T$, and $Z = Z_T$, we get the desired result by following the standard single-letter characterization technique.

Appendix B
Proof of Theorem 3

Here we show that $C_{s,\text{out}}$ is an upper-bound on the privacy–security pair achieved by any scheme. To do this, we let (Δ_P, R) be a privacy–security pair achieved by using encoding functions h_n and $\tilde{h}_n$ and decoding function g_n. That is $V = h_n(X^n)$, $K = \tilde{h}_n(X^n)$, and $\mathbb{P}[K \neq g_n(Y^n)] \le \epsilon$. In the following, we will show that there exist random variables W and U with $W \to U \to X \to (Y, Z)$, such that

$$\begin{aligned} \Delta_P \le 1 - \frac{I(X; UZ) - I(U; Y) + I(U; Z \,|\, W)}{H(X)} \\ + \epsilon_n \end{aligned}$$

and

$$R \le I(U; Y \,|\, W) - I(U; Z \,|\, W) + \epsilon_n \tag{56}$$

in which ϵ_n approaches 0 as n increases. That is $(\Delta_P, R) \in \mathcal{C}_{s,\text{out}}$.

Similarly to (49), we have

$$H(K\,|\,Y^n,V)\le n\delta_n. \tag{57}$$

We proceed as follows:

$$\begin{aligned}
&H(K)\\
&\quad= H(K\,|\,VZ^n)+I(K;VZ^n)\\
&\quad\overset{(a)}{\le} H(K\,|\,VZ^n)-H(K\,|\,VY^n)+n\delta_n+n\epsilon\\
&\quad= I(K;Y^n\,|\,V)-I(K;Z^n\,|\,V)+n\delta_n+n\epsilon\\
&\quad\overset{(b)}{=}\sum_{i=1}^{n}\big[I\left(K;Y_i\,\big|VY^{i-1}Z_{i+1}^n\right)\\
&\qquad\qquad-I\left(K;Z_i\,\big|VY^{i-1}Z_{i+1}^n\right)\big]+n\delta_n+n\epsilon\\
&\quad=\sum_{i=1}^{n}\big[I\left(KVY^{i-1}Z_{i+1}^n;Y_i\,\big|VY^{i-1}Z_{i+1}^n\right)\\
&\qquad\qquad-I\left(KVY^{i-1}Z_{i+1}^n;Z_i\big|\,VY^{i-1}Z_{i+1}^n\right)\big]\\
&\qquad+n\delta_n+n\epsilon\\
&\quad=\sum_{i=1}^{n}[I(U_i;Y_i\,|\,W_i)-I(U_i;Z_i\,|\,W_i)]+n\delta_n+n\epsilon
\end{aligned} \tag{58}$$

in which we have defined

$$W_i=\left(V,Y^{i-1},Z_{i+1}^n\right),\quad U_i=\left(K,V,Y^{i-1},Z_{i+1}^n\right). \tag{59}$$

Here (a) follows from (57) and the requirement that $I(K;VZ^n)\le n\epsilon$, and (b) can be obtained by using [41, Lemma 7].

In the following, we bound $H(X^n\,|\,VZ^n)$:

$$\begin{aligned}
&H(X^n\,|\,VZ^n)\\
&\quad=H(X^n)-I(X^n;VZ^n)\\
&\quad=H(X^n)-H(V)+H(V\,|\,X^n)-I(X^n;Z^n\,|\,V)\\
&\quad=H(X^n)-H(V)-H(Z^n\,|\,V)+H(Z^n\,|\,X^nV)
\end{aligned}$$

because V is a function of X^n.

We continue as follows:

$$\begin{aligned}
&H(X^n\,|\,VZ^n)\\
&=H(X^n)-H(V)-H(Z^n\,|\,V)+H(Z^n\,|\,X^nV)\\
&=\sum_{i=1}^{n}\left\{H(X_i)-H\left(Z_i\,\big|VZ_{i+1}^n\right)\right.\\
&\qquad\left.+H\left(Z_i\,\big|VX^nZ_{i+1}^n\right)\right\}-H(V)\\
&\le\sum_{i=1}^{n}\left\{H(X_i)-H\left(Z_i\,\big|VY^{i-1}Z_{i+1}^n\right)\right.\\
&\qquad\left.+H\left(Z_i\,\big|VX^nZ_{i+1}^n\right)\right\}-H(V)\\
&\overset{(a)}{=}\sum_{i=1}^{n}\left\{H(X_i)-H\left(Z_i\,\big|VY^{i-1}Z_{i+1}^n\right)\right.\\
&\qquad\left.+H\left(Z_i\,\big|VKX^nZ_{i+1}^n\right)\right\}-H(V)\\
&\overset{(b)}{=}\sum_{i=1}^{n}\left\{H(X_i)-H\left(Z_i\,\big|VY^{i-1}Z_{i+1}^n\right)\right.\\
&\qquad\left.+H\left(Z_i\,\big|VKX^nY^{i-1}Z_{i+1}^n\right)\right\}-H(V)\\
&\le\sum_{i=1}^{n}\left\{H(X_i)-H\left(Z_i\,\big|VY^{i-1}Z_{i+1}^n\right)\right.\\
&\qquad\left.+H\left(Z_i\,\big|VKX_iY^{i-1}Z_{i+1}^n\right)\right\}-H(V)\\
&\le\sum_{i=1}^{n}\left\{H(X_i)-I\left(X_iK;Z_i\,\big|VY^{i-1}Z_{i+1}^n\right)\right\}-H(V)\\
&=\sum_{i=1}^{n}\left\{H(X_i)-I\left(X_iKVY^{i-1}Z_{i+1}^n;Z_i\,\big|VY^{i-1}Z_{i+1}^n\right)\right\}\\
&\quad-H(V)
\end{aligned} \tag{60}$$

in which (a) is due to the fact that K is a function of X^n, and (b) is due to the Markov chain relationship

$$Y^{i-1}\to VKX^nZ_{i+1}^n\to Z_i. \tag{61}$$

To show this, we have that $Y^{i-1}Z_{i+1}^n\to X^n\to Z_i$, which leads to $Y^{i-1}\to Z_{i+1}^nVKX^n\to Z_i$, since VK is a function of X^n.

Now,

$$\begin{aligned}
H(V)&\ge H(V\,|\,Y^n)=H(VK\,|\,Y^n)-H(K\,|\,VY^n)\\
&\ge H(VK\,|\,Y^n)-n\delta_n\\
&=H(VK\,|\,Y^n)-H(VK\,|\,X^n)-n\delta_n\\
&=I(VK;X^n)-I(VK;Y^n)-n\delta_n\\
&=\sum_{i=1}^{n}\{I\left(VK;X_i\,\big|X_{i+1}^nY^{i-1}\right)\\
&\qquad-I\left(VK;Y_i\,\big|X_{i+1}^nY^{i-1}\right)\}-n\delta_n
\end{aligned} \tag{62}$$

in which we have used [41, Lemma 7].

We continue as follows:

$$\begin{aligned}
&H(V)\\
&\ge\sum_{i=1}^{n}\left\{I\left(VKX_{i+1}^nY^{i-1};X_i\right)\right.\\
&\qquad\left.-I\left(VKX_{i+1}^nY^{i-1};Y_i\right)\right\}\\
&\quad-n\delta_n\\
&\overset{(a)}{=}\sum_{i=1}^{n}\left\{I\left(VKX_{i+1}^nY^{i-1}Z_{i+1}^n;X_i\right)\right.\\
&\qquad\left.-I\left(VKX_{i+1}^nY^{i-1}Z_{i+1}^n;Y_i\right)\right\}-n\delta_n\\
&=\sum_{i=1}^{n}\left\{I\left(VKZ_{i+1}^nY^{i-1};X_i\right)\right.\\
&\qquad\left.-I\left(VKZ_{i+1}^nY^{i-1};Y_i\right)\right\}\\
&\quad+\sum_{i=1}^{n}\left\{I\left(X_{i+1}^n;X_i\big|\,VKY^{i-1}Z_{i+1}^n\right)\right.\\
&\qquad\left.-I\left(X_{i+1}^n;Y_i\big|\,VKY^{i-1}Z_{i+1}^n\right)\right\}-n\delta_n\\
&\overset{(b)}{\ge}\sum_{i=1}^{n}\left\{I\left(VKZ_{i+1}^nY^{i-1};X_i\right)\right.\\
&\qquad\left.-I\left(VKZ_{i+1}^nY^{i-1};Y_i\right)\right\}\\
&\quad-n\delta_n
\end{aligned} \tag{63}$$

in which (a) is due to the Markov chain relationship $Z_{i+1}^n \to VKX_{i+1}^n Y^{i-1} \to X_i Y_i$, which can be shown similarly to (61), and (b) is due to the fact that

$$I\left(X_{i+1}^n; X_i \middle| VKY^{i-1}Z_{i+1}^n\right) = I\left(X_{i+1}^n; X_i Y_i \middle| VKY^{i-1}Z_{i+1}^n\right)$$

which is due to the Markov chain relationship $X_{i+1}^n \to VKY^{i-1}Z_{i+1}^n X_i \to Y_i$. Combining (60) with (63), we have

$$\begin{aligned}
H(X^n \,|\, VZ^n) &\le \sum_{i=1}^n \big\{H(X_i) \\
&\qquad - I\left(X_i KVY^{i-1}Z_{i+1}^n; Z_i \middle| VY^{i-1}Z_{i+1}^n\right)\big\} \\
&\quad - H(V) + n\delta_n \\
&\le \sum_{i=1}^n \big\{H(X_i) \\
&\qquad - I\left(X_i KVY^{i-1}Z_{i+1}^n; Z_i \middle| VY^{i-1}Z_{i+1}^n\right) \\
&\qquad - I\left(VKZ_{i+1}^n Y^{i-1}; X_i\right) \\
&\qquad + I\left(VKZ_{i+1}^n Y^{i-1}; Y_i\right)\big\} + n\delta_n \\
&= \sum_{i=1}^n \{H(X_i) - I(X_i U_i; Z_i \,|\, W_i) - I(U_i; X_i) \\
&\qquad + I(U_i; Y_i)\} + n\delta_n \\
&= \sum_{i=1}^n \{H(X_i) - I(U_i; Z_i \,|\, W_i) - I(X_i; Z_i \,|\, W_i U_i) \\
&\qquad - I(U_i; X_i) + I(U_i; Y_i)\} + n\delta_n \\
&\overset{(a)}{=} \sum_{i=1}^n \{H(X_i) - I(U_i; Z_i \,|\, W_i) - I(X_i; Z_i \,|\, U_i) \\
&\qquad - I(U_i; X_i) + I(U_i; Y_i)\} + n\delta_n \\
&= \sum_{i=1}^n \{H(X_i) - I(X_i; U_i Z_i) - I(U_i; Z_i \,|\, W_i) \\
&\qquad + I(U_i; Y_i)\} + n\delta_n. \qquad (64)
\end{aligned}$$

Here W_i and U_i are defined in (59) and (a) is due to the Markov chain condition $W \to U \to X \to (Y, Z)$.

Now, by introducing a random variable T uniformly distributed over the set $\{1, \ldots, n\}$, and setting $W = (W_T, T)$, $U = (U_T, T)$, $X = X_T$, $Y = Y_T$, and $Z = Z_T$, we get the desired result by following the standard single-letter characterization technique.

Appendix C
Lemmas for the Proof of Theorem 3

In this section, we state and prove several lemmas used in Appendix B (the proof of Theorem 3).

Lemma 1: For the coding scheme in Theorem 3, we can write $H(U^n \,|\, K, V, Z^n) \le n\delta_n$, where $\delta_n \to 0$ as n increases.

Proof: With knowledge of k and v, the attacker can obtain an estimate $\tilde{u}^n$ of u^n by looking for a sequence in the subset k, bin i of $\Lambda_U(w^n)$ that is jointly typical with z^n. Based on an error probability analysis similar to that given in the proof of Theorem 1, one can show that the probability that $\tilde{U}^n \neq U^n$ goes to zero as n increases. Thus, using Fano's inequality, we have $H(U^n \,|\, K, V, Z^n) \le n\delta_n$ for a suitable choice of δ_n, which approaches 0 as n increases. ∎

Lemma 2: For the coding scheme in Theorem 3, we have

$$H(Z^n \,|\, W^n) \le nH(Z \,|\, W) + n\epsilon_n$$

in which ϵ_n goes to zero as n increases.

Proof: For each w^n, we define $\hat{z}^n$ as follows:

$$\hat{z}^n = \begin{cases} z^n, & \text{if } z^n \in T_{[Z\,|\,W]\delta}^n(w^n) \\ z_t^n, & \text{if } z^n \notin T_{[Z\,|\,W]\delta}^n(w^n) \end{cases} \qquad (65)$$

in which z_t^n is an arbitrary sequence in $\mathcal{Z}^n$.

We have

$$\begin{aligned}
H(Z^n \,|\, W^n) &\le H(Z^n, \hat{Z}_n \,|\, W^n) \\
&= H(Z^n \,|\, \hat{Z}^n, W^n) + H(\hat{Z}^n \,|\, W^n) \\
&\le H(Z^n \,|\, \hat{Z}^n) + H(\hat{Z}^n \,|\, W^n). \qquad (66)
\end{aligned}$$

From the Markov lemma [37], (Z^n, W^n) are jointly typical with high probability. Hence $Z^n = \hat{Z}^n$ with high probability, and thus we have

$$H(Z^n \,|\, \hat{Z}^n) \le n\epsilon_n^{'}$$

for a suitable choice of $\epsilon_n^{'}$ that approaches 0 as n increases, due to Fano's inequality [37].

At the same time, for any $w^n \in \Lambda_W \subset T_{[W]\delta}^n$, we have

$$H(\hat{Z}^n \,|\, w^n) \le \log |T_{[Z|W]\delta}^n(w^n)| \le n(H(Z \,|\, W) + \epsilon_n^{''})$$

for a suitable choice of $\epsilon_n^{''}$ that approaches 0 as n increases [36].

Hence

$$\begin{aligned}
H(Z^n \,|\, W^n) &\le H(Z^n \,|\, \hat{Z}^n) + H(\hat{Z}^n \,|\, W^n) \\
&\le n\epsilon_n + \sum_{w^n \in \Lambda_W} \mathbb{P}(W^n = w^n) H(\hat{Z}^n \,|\, W^n = w^n) \\
&\le n\epsilon_n^{'} + n(H(Z \,|\, W) + \epsilon_n^{''}). \qquad (67)
\end{aligned}$$

On defining $\epsilon_n = \epsilon_n^{'} + \epsilon_n^{''}$, which approaches zero as n increases, the claim is proved. ∎

Lemma 3: For any $\epsilon > 0$, there exists a sufficiently large n such that $H(X^n \,|\, U^n, Z^n, W^n) \ge (1 - \epsilon)nH(X \,|\, UZW) - 2n\epsilon$.

Proof: Consider

$$\begin{aligned}
H(X^n \,|\, U^n, Z^n, W^n) &\ge - \sum_{(x^n, u^n, z^n, w^n) \in T_{[XUZW]\epsilon}^n} P_{X^n, U^n, Z^n, W^n} \\
&\quad \times (x^n, u^n, z^n, w^n) \\
&\quad \times \log P_{X^n \,|\, U^n, Z^n, W^n}(x^n \,|\, u^n, z^n, w^n) \\
&\ge \sum_{(x^n, u^n, z^n, w^n) \in T_{[XUZW]\epsilon}^n} P_{X^n, U^n, Z^n, W^n}(x^n, u^n, z^n, w^n) \\
&\quad \times n(H(X \,|\, UZW) - 2\epsilon) \\
&= \mathbb{P}\left[(X^n, W^n, U^n, Z^n) \in T_{[XWUZ]\epsilon}^n\right]
\end{aligned}$$

$$\begin{aligned}
&\times n(H(X\,|\,UZW)-2\epsilon)\\
&\overset{(a)}{\geq}(1-\epsilon)n(H(X\,|\,UZW)-2\epsilon)\\
&\geq(1-\epsilon)nH(X\,|\,UZW)-2n\epsilon. \qquad (68)
\end{aligned}$$

Here for each $\epsilon>0$, (a) is true for sufficiently large n [37]. ■

Appendix D
Proof of Theorem 4

Here we show that $C_{sr,\text{out}}$ is an upper-bound on the privacy–security pair achieved by any scheme. To do this, we let (Δ_P,R) be a privacy–security pair achieved by using encoding functions h_n and $\tilde{h}_n$, and decoding function g_n. That is $V=h_n(X^n)$, $K=\tilde{h}_n(X^n)$, $n^{-1}\log|\mathcal{K}|\leq R+\epsilon$, and $\mathbb{P}[K\neq g_n(Y^n)]\leq\epsilon$. In the following, we will show that there exist random variables W and U with $W\to U\to X\to(Y,Z)$, such that

$$\begin{aligned}
\Delta_P&\leq 1-\frac{I(X;Z\,|\,U)-I(U;Y)+I(U;Z\,|\,W)}{H(X)}+\epsilon_n\\
R&\leq I(U;Y\,|\,W)-I(U;Z\,|\,W)+\epsilon_n \qquad (69)
\end{aligned}$$

in which ϵ_n approaches zero as n increases. That is $(\Delta_P,R)\in\mathcal{C}_{sr,\text{out}}$.

Again, similar to (49), we have $H(K\,|\,Y^n,V)\leq n\delta_n$.

We first bound the privacy leakage, as follows:

$$\begin{aligned}
&H(X^n\,|\,VZ^n)\\
&\quad\leq H(X^n,K\,|\,VZ^n)\\
&\quad=H(X^n,K)-I(X^n,K;VZ^n)\\
&\quad=H(X^n)+H(K)-I(X^n,K;V)\\
&\qquad-I(X^n,K;Z^n\,|\,V)\\
&\quad=H(X^n)+H(K)-H(V)+H(V\,|\,X^n,K)\\
&\qquad-H(Z^n\,|\,V)+H(Z^n\,|\,X^n,K,V)\\
&\quad=H(X^n)-H(Z^n\,|\,V)+H(Z^n\,|\,X^n,K,V)\\
&\qquad+H(K)-H(V)
\end{aligned}$$

since V is a function of (X^n,K). We continue:

$$\begin{aligned}
&H(X^n\,|\,VZ^n)\\
&\leq\sum_{i=1}^n\{H(X_i)-H(Z_i\,|\,Z_{i+1}^n,V)\\
&\qquad+H(Z_i\,|\,Z_{i+1}^n,X^n,K,V)\}+H(K)\\
&\leq\sum_{i=1}^n\{H(X_i)-H(Z_i\,|\,Z_{i+1}^n,V,Y^{i-1},K)\\
&\qquad+H(Z_i\,|\,Z_{i+1}^n,X^n,K,V)\}+H(K)\\
&\overset{(a)}{\leq}\sum_{i=1}^n\left\{H(X_i)-H\left(Z_i\middle|Z_{i+1}^n,V,Y^{i-1},K\right)\right.\\
&\qquad\left.+H\left(Z_i\middle|Z_{i+1}^n,Y^{i-1},X^n,K,V\right)\right\}+H(K)\\
&\leq\sum_{i=1}^n\left\{H(X_i)-H\left(Z_i\middle|Z_{i+1}^n,V,Y^{i-1},K\right)\right.\\
&\qquad\left.+H\left(Z_i\middle|X_i,Z_{i+1}^n,Y^{i-1},K,V\right)\right\}+H(K)\\
&\leq\sum_{i=1}^n\left\{H(X_i)-I(X_i;Z_i\middle|Z_{i+1}^n,V,Y^{i-1},K)\right\}\\
&\qquad+H(K).
\end{aligned}$$

In the derivation above, (a) is due to the Markov chain relationship $Y^{i-1}\to(Z_{i+1}^n,X^n,K,V)\to Z_i$, which can be easily shown.

At the same time, we have

$$\begin{aligned}
H(K)&=H(K\,|\,VZ^n)+I(K;VZ^n)\\
&\overset{(a)}{\leq}H(K\,|\,VZ^n)-H(K\,|\,VY^n)+n\delta_n+n\epsilon\\
&=I(K;Y^n\,|\,V)-I(K;Z^n\,|\,V)+n\delta_n+n\epsilon\\
&\overset{(b)}{=}\sum_{i=1}^n\left[I\left(K;Y_i\middle|VY^{i-1}Z_{i+1}^n\right)\right.\\
&\qquad\left.-I\left(K;Z_i\middle|VY^{i-1}Z_{i+1}^n\right)\right]+n\delta_n+n\epsilon\\
&=\sum_{i=1}^n\left[I\left(KVY^{i-1}Z_{i+1}^n;Y_i\middle|VY^{i-1}Z_{i+1}^n\right)\right.\\
&\qquad\left.-I\left(KVY^{i-1}Z_{i+1}^n;Z_i\middle|VY^{i-1}Z_{i+1}^n\right)\right]\\
&\quad+n\delta_n+n\epsilon\\
&=\sum_{i=1}^n[I(U_i;Y_i\,|\,W_i)-I(U_i;Z_i\,|\,W_i)]\\
&\quad+n\delta_n+n\epsilon
\end{aligned}$$

in which we define

$$W_i=(V,Y^{i-1},Z_{i+1}^n),\quad U_i=\left(K,V,Y^{i-1},Z_{i+1}^n\right). \qquad (70)$$

Here (a) follows from Fano's inequality and the requirement that $I(K;VZ^n)\leq n\epsilon$, and (b) can be obtained by using [41, Lemma 7].

Now, by introducing a random variable T uniformly distributed over the set $\{1,\ldots,n\}$, and setting $W=(W_T,T)$, $U=(U_T,T)$, $X=X_T$, $Y=Y_T$, and $Z=Z_T$, we get the desired result by following the standard single-letter characterization technique.

Appendix E
Proof of Necessity in Theorem 5

Consider a coding scheme (h_n^*,g_n,P_K) which achieves the pair (R,Δ_s) so that (11)–(14) are satisfied. For simplicity, we assume $P_V(v)>0$ for all v. We will show that a common random process exists between X^n and Y^n with entropy rate $R\Delta_s$. We show this by explicitly constructing the functions $\psi_n(X^n)$ and $\phi_n(Y^n)$ from the coding scheme (h_n^*,g_n,P_K).

Note that the joint distribution of X^n, K, and V is given by

$$P_{X^n,K,V}(x^n,k,v)=P_{X^n}(x^n)P_K(k)\mathbf{1}\{v=h_n^*(x^n,k)\}.$$

Let

$$\begin{aligned}
f_n(x^n,v)=&\operatorname*{argmax}_{k^*}\sum_{y^n}P_{Y^n\,|\,X^n}(y^n\,|\,x^n)\\
&\times\mathbf{1}\{k^*=g_n(y^n,v)\} \qquad (71)
\end{aligned}$$

and

$$K^* = f_n(X^n, V).$$

Then

$$\mathbb{P}[K \neq K^*] \leq \mathbb{P}[K \neq \hat{K}] \leq \epsilon \tag{72}$$

where the last inequality follows from (14).

Let $\tilde{V}$ be an auxiliary random variable that has the same distribution as V but is independent of all the above named variables. Let $\tilde{K} = f_n(X^n, \tilde{V})$. Then

$$P_{\tilde{K}\tilde{V}}(kv) = \sum_{x^n : f_n(x^n, v) = k} P_{X^n}(x^n) P_V(v)$$

and

$$P_{K^*V}(kv) = \sum_{x^n : f_n(x^n, v) = k} P_{X^nV}(x^n v).$$

Therefore,

$$\begin{aligned}
&\sum_{kv} |P_{K^*V}(kv) - P_{\tilde{K}\tilde{V}}(kv)| \\
&\quad = \sum_{kv} \left| \sum_{x^n : f_n(x^n, v) = k} P_{X^nV}(x^n v) \right. \\
&\qquad \left. - \sum_{x^n : f_n(x^n, v) = k} P_{X^n}(x^n) P_V(v) \right| \\
&\quad \leq \sum_{kv} \sum_{x^n : f_n(x^n, v) = k} |P_{X^nV}(x^n v) - P_{X^n}(x^n) P_V(v)| \\
&\quad = \sum_{x^n v} |P_{X^nV}(x^n v) - P_{X^n}(x^n) P_V(v)| && (73) \\
&\quad \leq \sqrt{2 I(X^n; V) \ln 2} && (74) \\
&\quad \leq \sqrt{2\epsilon \ln 2} && (75)
\end{aligned}$$

where (74) follows from Pinsker's inequality [37] and (75) follows from (12). Hence,

$$\begin{aligned}
&n^{-1} |H(K^* \,|\, V) - H(\hat{K}|\tilde{V})| \\
&\quad = n^{-1} |H(K^* V) - H(\hat{K}\tilde{V})| && (76) \\
&\quad \leq n^{-1} \left(h\left(\frac{\sqrt{2\epsilon \ln 2}}{2} \right) + \frac{\sqrt{2\epsilon \ln 2}}{2} \log(|\mathcal{K}||\mathcal{V}|) \right) && (77) \\
&\quad \leq h\left(\frac{\sqrt{2\epsilon \ln 2}}{2} \right) + \frac{\sqrt{2\epsilon \ln 2}}{2n} \log(|\mathcal{K}||\mathcal{V}|) && (78)
\end{aligned}$$

where (77) follows from (75) and [42, Theorem 7]. Therefore,

$$\begin{aligned}
&n^{-1} H(\tilde{K} \,|\, \tilde{V}) \\
&\quad \geq n^{-1} H(K^* \,|\, V) - h\left(\frac{\sqrt{2\epsilon \ln 2}}{2} \right) \\
&\qquad - \frac{\sqrt{2\epsilon \ln 2}}{2n} \log(|\mathcal{K}||\mathcal{V}|) \\
&\quad \geq n^{-1} I(K; K^* \,|\, V) - h\left(\frac{\sqrt{2\epsilon \ln 2}}{2} \right) \\
&\qquad - \frac{\sqrt{2\epsilon \ln 2}}{2n} \log(|\mathcal{K}||\mathcal{V}|) \\
&\quad \geq n^{-1} (H(K \,|\, V) - H(K \,|\, K^*)) - h\left(\frac{\sqrt{2\epsilon \ln 2}}{2} \right) \\
&\qquad - \frac{\sqrt{2\epsilon \ln 2}}{2n} \log(|\mathcal{K}||\mathcal{V}|) \\
&\quad \geq n^{-1} H(K \,|\, V) - h\left(\frac{\sqrt{2\epsilon \ln 2}}{2} \right) \\
&\qquad - \frac{\sqrt{2\epsilon \ln 2}}{2n} \log(|\mathcal{K}||\mathcal{V}|) - n^{-1}\delta && (79)
\end{aligned}$$

where

$$\delta = h(\epsilon) + \epsilon \log |\mathcal{K}|$$

and the last inequality follows from (72) and Fano's inequality [37]. For any $\mu > 0$, let

$$\begin{aligned}
\Gamma_\mu = \{ & v : n^{-1} H(\tilde{K} \,|\, \tilde{V} = v) \\
& \geq n^{-1} H(\tilde{K} \,|\, \tilde{V}) - \mu \}
\end{aligned} \tag{80}$$

and

$$\begin{aligned}
v^* = \operatorname*{argmin}_{v^* \in \Gamma_\mu} & \sum_{x^n y^n} P_{X^nY^n}(x^n y^n) \\
& \times \mathbf{1}\{ f_n(x^n, v^*) \neq g_n(y^n, v^*) \}.
\end{aligned}$$

Now we consider a pair of random variables $(f_n(X^n, v^*), g_n(Y^n, v^*))$. Note that

$$\begin{aligned}
&n^{-1} H(f_n(X^n, v^*)) \\
&\quad = n^{-1} H(\tilde{K} \,|\, \tilde{V} = v^*) \\
&\quad \geq n^{-1} H(\tilde{K} \,|\, \tilde{V}) - \mu \\
&\quad \geq n^{-1} H(K \,|\, V) - h\left(\frac{\sqrt{2\epsilon \ln 2}}{2} \right) && (81) \\
&\qquad - \frac{\sqrt{2\epsilon \ln 2}}{2n} \log(|\mathcal{K}||\mathcal{V}|) - \mu - n^{-1}\delta && (82)
\end{aligned}$$

where (81) follows from the fact that $v^* \in \Gamma_\mu$, and (82) follows from (79).

Since $Y^n \to X^n \to V$ and $I(X^n; V) \leq \epsilon$, we have $I(X^n Y^n; V) \leq \epsilon$. Together with Pinsker's inequality, we have

$$\begin{aligned}
&\sqrt{2\epsilon \ln 2} \\
&\quad \geq \sum_{x^n y^n v} |P_{X^nY^n}(x^n y^n) P_V(v) - P_{X^nY^nV}(x^n y^n v)| \\
&\quad \geq \sum_v P_V(v) \sum_{x^n y^n} |P_{X^nY^n}(x^n y^n) \\
&\qquad - P_{X^nY^n \,|\, V}(x^n y^n \,|\, v)|
\end{aligned}$$

$$\begin{aligned}
&\geq \sum_v P_V(v) \sum_{x^n y^n} |P_{X^nY^n}(x^n y^n) \\
&\qquad - P_{X^nY^n\,|\,V}(x^n y^n\,|\,v)| \\
&\quad \times \mathbf{1}\{f_n(x^n, v) \neq g_n(y^n, v)\} \\
&\geq \sum_v P_V(v) \sum_{x^n y^n} (P_{X^nY^n}(x^n y^n) \\
&- P_{X^nY^n\,|\,V}(x^n y^n\,|\,v)) \\
&\quad \times \mathbf{1}\{f_n(x^n, v) \neq g_n(y^n, v)\}. \qquad (83)
\end{aligned}$$

From (71), we can see that

$$\begin{aligned}
&\mathbb{P}[K^* = \hat{K}\,|\,X^n = x^n, V = v] \\
&\quad = \sum_{y^n} P_{Y^n\,|\,X^n}(y^n\,|\,x^n)\mathbf{1}\{f_n(x^n, v) = g_n(y^n, v)\} \\
&\quad \geq \sum_{y^n} P_{Y^n\,|\,X^n}(y^n\,|\,x^n)\mathbf{1}\{k = g_n(y^n, v)\} \\
&\quad = \mathbb{P}[K = \hat{K}\,|\,X^n = x^n, V = v, K = k]. \qquad (84)
\end{aligned}$$

Therefore, $\mathbb{P}[K^* = \hat{K}] \geq \mathbb{P}[K = \hat{K}]$. Together with the requirement that $\mathbb{P}[K \neq \hat{K}] \leq \epsilon$, we get

$$\begin{aligned}
\epsilon &\geq \mathbb{P}[K \neq \hat{K}] \\
&\geq \mathbb{P}[K^* \neq \hat{K}] \\
&= \sum_v P_V(v) \sum_{x^n y^n} P_{X^nY^n\,|\,V}(x^n y^n\,|\,v) \\
&\quad \times \mathbf{1}\{f_n(x^n, v) \neq g_n(y^n, v)\}. \qquad (85)
\end{aligned}$$

Together with (83), we have

$$\begin{aligned}
\sqrt{2\epsilon \ln 2} + \epsilon &\geq \sum_v P_V(v) \sum_{x^n y^n} P_{X^nY^n}(x^n y^n) \\
&\quad \times \mathbf{1}\{f_n(x^n, v) \neq g_n(y^n, v)\} \\
&\geq \sum_{v \in \Gamma_\mu} P_V(v) \sum_{x^n y^n} P_{X^nY^n}(x^n y^n) \\
&\quad \times \mathbf{1}\{f_n(x^n, v) \neq g_n(y^n, v)\} \\
&\geq \left(\sum_{v \in \Gamma_\mu} P_V(v) \right) \sum_{x^n y^n} P_{X^nY^n}(x^n y^n) \\
&\quad \times \mathbf{1}\{f_n(x^n, v^*) \neq g_n(y^n, v^*)\}. \qquad (86)
\end{aligned}$$

The proof can be completed if we can find a lower bound on $(\sum_{v \in \Gamma_\mu} P_V(v))$. Note that

$$\begin{aligned}
n^{-1} H(\tilde{K}\,|\,\tilde{V}) &= n^{-1} \sum_{v \in \Gamma_\mu} P_V(v) H(\tilde{K}\,|\,\tilde{V} = v) \\
&\quad + n^{-1} \sum_{v \in \Gamma_\mu^c} P_V(v) H(\tilde{K}\,|\,\tilde{V} = v) \\
&< n^{-1} \sum_{v \in \Gamma_\mu} P_V(v) H(X^n) \\
&\quad + \sum_{v \in \Gamma_\mu^c} P_V(v)(n^{-1} H(\tilde{K}\,|\,\tilde{V}) - \mu).
\end{aligned}$$

After rearranging the terms, we obtain

$$\begin{aligned}
\sum_{v \in \Gamma_\mu} P_V(v) &\geq \frac{\mu}{n^{-1} H(X^n) - n^{-1} H(\tilde{K}\,|\,\tilde{V}) + \mu} \\
&\geq \frac{\mu}{n^{-1} H(X^n) + \mu} \\
&\geq \frac{\mu}{\log |\mathcal{X}| + \mu}. \qquad (87)
\end{aligned}$$

Together with (86), we have

$$\begin{aligned}
&\sum_{xy} P_{X^nY^n}(x^n y^n)\mathbf{1}\{f_n(x^n, v^*) \neq g_n(y^n, v^*)\} \\
&\qquad \leq \frac{(\sqrt{2\epsilon \ln 2} + \epsilon)(\log |\mathcal{X}| + \mu)}{\mu}. \qquad (88)
\end{aligned}$$

Finally, for any $\eta > 0$, we take

$$\mu = \frac{\eta}{2}.$$

Due to (1) and (2), there exists a sufficiently small ϵ such that

$$\begin{aligned}
\max \Bigg\{ h\left(\frac{\sqrt{2\epsilon \ln 2}}{2}\right) + \frac{\sqrt{2\epsilon \ln 2}}{2n} \log(|\mathcal{K}||\mathcal{V}|) + n^{-1}\delta, \\
\frac{(\sqrt{2\epsilon \ln 2} + \epsilon)(\log |\mathcal{X}| + \mu)}{\mu} \Bigg\} \leq \frac{\eta}{2} \qquad (89)
\end{aligned}$$

even though $|\mathcal{K}|$ and $|\mathcal{V}|$ may be increasing as ϵ decreasing. At the same time, $n^{-1}\delta \to 0$ as $\epsilon \to 0$.

Finally, let $\psi_n(X^n) = f_n(X^n, v^*)$ and $\phi_n(Y^n) = g_n(Y^n, v^*)$. Then

$$\begin{aligned}
n^{-1} H(\psi_n(X^n)) &= n^{-1} H(f_n(X^n, v^*)) \\
&\geq n^{-1} H(K\,|\,V) - \eta \qquad (90) \\
&\geq R\Delta_s - \eta \qquad (91)
\end{aligned}$$

where (90) follows from (82), and (91) follows from (11) and (13). At the same time, the relationship $\mathbb{P}[\psi_n(X^n) \neq \phi_n(Y^n)] = \mathbb{P}[f_n(X^n, v^*) \neq g_n(Y^n, v^*)] \leq \eta$ follows from (88) and (89).

Thus, we have successfully constructed $\psi_n(X^n)$ and $\phi_n(Y^n)$, and hence there is a common random process between X^n and Y^n with entropy rate $R\Delta_s$.

REFERENCES

[1] L. Lai, S.-W. Ho, and H. V. Poor, "Privacy-security tradeoffs in biometric security systems," in *Proc. 46th Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Sep. 23–26, 2008.

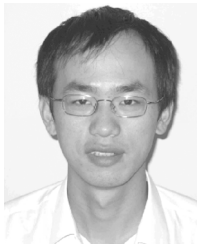
[2] L. Lai, S.-W. Ho, and H. V. Poor, "An information theoretic framework for biometric security systems," in *Proc. Int. Conf. Biometrics*, Sassari, Italy, Jun. 2009.

[3] A. Ross, J. Shah, and A. K. Jain, "From template to image: Reconstructing fingerprints from minutiae points," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 544–560, Apr. 2007.

[4] Y. Sutcu, Q. Li, and N. Memon, "Protecting biometric templates with sketch: Theory and practice," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 503–512, Sep. 2007.

[5] Q. Li, Y. Sutcu, and N. Memon, "Secure sketch for biometric templates," in *Proc. Asiacrypt*, Shanghai, China, Dec. 2006, pp. 99–113.

[6] Y. Sutcu, Q. Li, and N. Memon, "Secure biometric templates from fingerprint-face features," in *Proc. IEEE Computer Society Workshop on Biometrics*, Minneapolis, MN, Jun. 2007.
[7] S. Draper, A. Khisti, E. Martinian, A. Vetro, and J. Yedidia, "Using distributed source coding to secure fingerprint biometrics," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Honolulu, HI, Apr. 2007, pp. 129–132.
[8] Y. Sutcu, S. Rane, J. S. Yedidia, S. Draper, and A. Vetro, "Feature transformation of biometric templates for secure biometric systems based on error correcting codes," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Anchorage, AK, June 2008.
[9] G. Davida, Y. Frankel, and B. Matt, "On enabling secure applications through off-line biometric identifications," in *Proc. IEEE Symp. Security and Privacy*, Oakland, CA, May 1998, pp. 148–157.
[10] K. Simoens, P. Tuyls, and B. Preneel, "Privacy weaknesses in biometric sketches," in *Proc. IEEE Int. Symp. Security and Privacy*, May 2009, pp. 188–203.
[11] J. Bringer, H. Chabannea, and B. Kindarji, "The best of both worlds: Applying secure sketches to cancelable biometrics," *Sci. Comput. Program.*, vol. 74, pp. 43–51, 2008.
[12] N. Ratha, S. Chikkerur, J. Connell, and R. Bolle, "Generating cancelable fingerprint templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 561–572, Apr. 2007.
[13] X. Boyen, "Reusable cryptographic fuzzy extractors," in *Proc. ACM Conf. Computer and Communications Security*, New York, 2004, pp. 82–91, ACM Press.
[14] A. Juels and M. Sudan, "A fuzzy vault scheme," in *Proc. IEEE Int. Symp. Inform. Theory*, Lausanne, Switzerland, Jun./Jul. 2002, pp. 293–297.
[15] U. Uludag, S. Pankanti, and A. K. Jain, "Fuzzy vault for fingerprints," *Audio- and Video-Based Biometric Person Authentication*, pp. 310–319, Jul. 2005.
[16] U. Uludag and A. K. Jain, "Securing fingerprint template: Fuzzy vault with helper data," in *Proc. IEEE Workshop Privacy Research in Vision*, New York, 2006.
[17] Y. Dodis, L. Reyzin, and A. Smith, "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data," in *Proc. Advances in Cryptology (EUROCRYPT)*, Interlaken, Switzerland, May 2004, pp. 523–540.
[18] G. Cohen and G. Zemor, "The wire-tap channel applied to biometrics," in *Proc. IEEE Int. Symp. Inf. Theory and Its Applications*, Parma, Italy, Oct. 2004.
[19] T. Ignatenko and F. M. J. Willems, "On privacy in secure biometrics authentication systems," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Honolulu, HI, Apr. 2007, pp. 121–124.
[20] P. Tuyls and J. Goseling, "Capacity and examples of template-protecting biometric authentication systems," in *Biometric Authentication*. Berlin, Germany: Springer, 2004, pp. 158–170.
[21] A. Nagar, K. Nandakumar, and A. K. Jain, "Biometric template transformation: A security analysis," in *Proc. SPIE, Electronic Imaging, Media Forensics and Security*, San Jose, Jan. 2010.
[22] M. Upmanyu, A. M. Namboodiri, K. Srinathan, and C. V. Jawahar, "Efficient biometric verification in encrypted domain," in *Proc. Int. Conf. Biometrics*, Sassari, Italy, 2009.
[23] P. Tuyls and J. Goseling, *Biometric Authentication*. Berlin, Germany: Springer, 2004.
[24] A. K. Jain, K. Nandakumar, and A. Nagar, "Biometric template security," *EURASIP J. Advances Signal Process.*, pp. 1–17, Jan. 2008.
[25] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 733–742, May 1993.
[26] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography, Part I: Secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, Jul. 1993.
[27] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography, Part II: CR capacity," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 225–240, Jan. 1998.
[28] L. Lai, S.-W. Ho, and H. V. Poor, "Privacy–security trade-offs in biometric security systems—Part II: Multiple uses case," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 1, Mar. 2011.
[29] I. Csiszár and P. Narayan, "Common randomness and secret key generation with a helper," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 344–366, Mar. 2000.
[30] T. Ignatenko and F. M. J. Willems, "Privacy leakage in biometric secrecy systems," in *Proc. Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Sep. 2008.
[31] T. Ignatenko and F. M. J. Willems, "Biometric systems: Privacy and secrecy aspects," *IEEE Trans. Inf. Forensics Security*, vol. 4, no. 4, pp. 956–973, Dec. 2009.
[32] Y. Sutcu, S. Rane, J. S. Yedidia, S. Draper, and A. Vetro, "Feature extraction for a Slepian-Wolf biometric system using LDPC codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Toronto, Canada, Jul. 2008.
[33] S.-W. Ho and R. W. Yeung, "On the discontinuity of the Shannon information measures," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5362–5374, Dec. 2009.
[34] W. J. Scheirer and T. E. Boult, "Cracking fuzzy vaults and biometric encryption," in *Proc. IEEE Biometrics Symp.*, Baltimore, MD, Sep. 2007, pp. 1–6.
[35] S.-W. Ho, "On the interplay between shannon's information measures and reliability criteria," in *Proc. IEEE Int. Symp. Inf. Theory*, Seoul, South Keora, Jun./Jul. 2009, pp. 154–158.
[36] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
[37] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[38] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, pp. 1355–1387, Oct. 1975.
[39] P. Gács and J. Körner, "Common information is far less than mutual information," *Probl. Control Inf. Theory*, vol. 2, pp. 149–162, 1973.
[40] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM J. Appl. Math.*, vol. 28, no. 1, pp. 100–113, Jan. 1975.
[41] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, May 1978.
[42] S.-W. Ho and R. W. Yeung, "The interplay between entropy and variational distance," in *Proc. IEEE Int. Symp. Inf. Theory*, Nice, France, Jul. 2007, pp. 491–495.



Lifeng Lai (M'07) received the B.E. and M.E. degrees from Zhejiang University, Hangzhou, China, in 2001 and 2004, respectively, and the Ph.D. degree from the Ohio State University at Columbus, OH, in 2007.

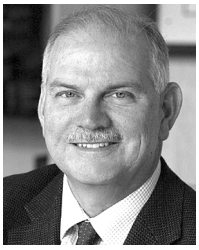
He was a postdoctoral research associate at Princeton University, Princeton, NJ, from 2007 to 2009. Since Aug. 2009, he has been an assistant professor at the University of Arkansas, Little Rock. His current research interests include network information theory, information theoretic security, statistical analysis of wireless networks, and biometric security systems.

Dr. Lai was a Distinguished University Fellow of the Ohio State University from 2004 to 2007. He coauthored a paper that received the Best Paper Award from the IEEE Global Communications Conference, 2008.

Siu-Wai Ho (S'05–M'07) was born in Hong Kong. He received the B.Eng., M.Phil., and Ph.D. degrees in information engineering from The Chinese University of Hong Kong in 2000, 2003, and 2006, respectively.

During 2006–2008, he was a Postdoctoral Research Fellow in the Department of Electrical Engineering, Princeton University, Princeton, NJ. Since 2009, he has been a Research Fellow at the Institute for Telecommunications Research (ITR), University of South Australia (UniSA), Adelaide, Australia, where he holds the ITR Director's Fellowship. His current research interests include Shannon theory, data communications and recording systems, and biometric security systems.

Dr. Ho was a recipient of the Croucher Foundation Fellowship for 2006–2008, the 2008 Young Scientist Award from the Hong Kong Institution of Science, UniSA Research SA Fellowship for 2010–2013, and the Australian Research Council Australian Postdoctoral Fellowship for 2010–2013.



H. Vincent Poor (S'72–M'77–SM'82–F'87) received the Ph.D. degree in electrical engineering and computer science from Princeton University, Princeton, NJ, in 1977.

From 1977 to 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990 he has been on the faculty at Princeton, where he is the Dean of Engineering and Applied Science, and the Michael Henry Strater University Professor of Electrical Engineering. His research interests are in the areas of stochastic analysis, statistical signal processing and information theory, and their applications in wireless networks and related fields. Among his publications in these areas are *Quickest Detection* (Cambridge University Press, 2009), coauthored with Olympia Hadjiliadis, and *Information Theoretic Security* (Now Publishers, 2009), coauthored with Yingbin Liang and Shlomo Shamai.

Dr. Poor is a member of the National Academy of Engineering, a Fellow of the American Academy of Arts and Sciences, and an International Fellow of the Royal Academy of Engineering (U.K.). He is also a Fellow of the Institute of Mathematical Statistics, the Optical Society of America, and other organizations. In 1990, he served as President of the IEEE Information Theory Society, from 2004 to 2007 as the Editor-in-Chief of the IEEE TRANSACTIONS ON INFORMATION THEORY, and in 2009 as General Cochair of the IEEE International Symposium on Information Theory, held in Seoul, South Korea. He received a Guggenheim Fellowship in 2002 and the IEEE Education Medal in 2005. Recent recognition of his work includes the 2009 Edwin Howard Armstrong Achievement Award of the IEEE Communications Society, the 2010 IET Ambrose Fleming Medal for Achievement in Communications, and the 2011 IEEE Eric E. Sumner Award.