Multi-Key Generation over a Cellular Model with a Helper [1] [2]

Huishuai Zhang, Yingbin Liang, Lifeng Lai, Shlomo Shamai (Shitz) [3]

Abstract

The problem of simultaneously generating multiple keys for a cellular source model with a helper is investigated. In the model considered, there are four terminals, $\mathcal{X}_0$, $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{X}_3$, each of which observes one component of a vector source. Terminal $\mathcal{X}_0$ wishes to generate two secret keys K_1 and K_2 respectively with terminals $\mathcal{X}_1$ and $\mathcal{X}_2$ under the help of terminal $\mathcal{X}_3$. All terminals are allowed to communicate over a public channel. An eavesdropper is assumed to have access to the public discussion. Both symmetric and asymmetric key generations are considered. In *symmetric* key generation models, model 1a (with a trusted helper) requires that the two keys are concealed from the eavesdropper, and model 1b (with an untrusted helper) further requires that the two keys are concealed from the helper in addition to the eavesdropper. The *asymmetric* key generation models 2a and 2b are the same as symmetric key generation models 1a and 1b, respectively, except that the key K_2 is further required to be concealed from terminal $\mathcal{X}_1$. For all models studied, the key capacity region is established by designing a unified achievable strategy to achieve the cut-set outer bounds.

1 Introduction

In Shannon's secrecy system, it is essential that distinct terminals share a common secret key, which can be exploited to achieve secure communications. In [1–3], it has been shown that such a secret key can be established between two remote terminals if each terminal has access to a component of a vector source sequence and the two terminals can communicate with each other via a public channel. The generated key can be kept secure from the eavesdropper, which has full access to the public channel. The key capacity for the two-terminal source model has been established in these studies. Since then, various models of key generation

[1]This paper was presented in part at the Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, Nov. 2014

[2]The work of H. Zhang and Y. Liang was supported by a National Science Foundation CAREER Award under Grant CCF-10-26565 and by the National Science Foundation under Grant CNS-11-16932. The work of L. Lai was supported by a National Science Foundation CAREER Award under Grant CCF-13-18980 and by the National Science Foundation under Grant CNS-13-21223. The work of S. Shamai was supported by the Israel Science Foundation (ISF), and the European Commission in the framework of the Network of Excellence in Wireless COMmunications NEWCOM#.

[3]H. Zhang and Y. Liang are with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY 13244 USA (email: {hzhan23,yliang06}@syr.edu). L. Lai is with the Department of Electrical and Computer Engineering, Worcester Polytechnic Institute, Worcester, MA 01609 USA (email: llai@wpi.edu). S. Shamai is with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Technion city, Haifa 32000, Israel (email:sshlomo@ee.technion.ac.il).

with public discussion have been studied. In [4], the model with two terminals and a helper was studied and the key capacity was characterized when there is no rate constraint on the public discussion. Later, in [5], the problem of secret key generation among multiple (more than two) terminals was studied and the key capacity was established when there is no constraint on the public discussion. In [6], the same problem was revisited, and the scenario with only a subset of terminals being allowed to talk publicly was studied and the capacity was characterized. In [7], an alternative form of the secret key capacity for the multi-terminal source model was characterized in terms of *mutual dependence*, which can be interpreted as generalization of mutual information to more than two variables.

All the studies described above considered the problem of generating a single secret key. More recently, there is increasing interest in the problem of simultaneously generating multiple secret/private keys over source models. The model of generating a secret key and a private key among three terminals was first studied in [8], in which three terminals observe correlated source outputs, and all three terminals wish to agree on a common secret key to be kept secure from an eavesdropper while two designated terminals wish to agree on a private key to be kept secure from both the third terminal and the eavesdropper. In [8], an outer bound on the key capacity region was provided and was shown to be achievable under a certain condition. More recently, in [9,10], it was further shown that the outer bound established in [8] is also achievable for the remaining cases of this model by employing a random binning and joint decoding scheme. Hence, the key capacity region for this model is established in general. Another so-called cellular source model was recently studied in [11, 12], in which one (base station) terminal wishes to generate independent keys respectively with a number of (mobile) terminals. In [11], the key sum capacity was established, and inner and outer bounds on the key capacity region were derived for the special case of two-key generation with an helper. In [12], the key capacity region was established for a two-key generation model with an additional requirement that one key should also be kept secure from the other mobile terminal.

To the best of our knowledge, so far the multi-key capacity region is characterized only for three-terminal systems. It is thus of interest to investigate whether it is possible to characterize the multi-key capacity region for systems with more than three terminals. Such exploration needs to address two challenging issues. (1) Although the key capacity region for the three-terminal model [8, 10, 12] was shown to be equal to the cut-set bound, it is not clear at the outset whether the cut-set bound can still be achieved for models with different secrecy requirements and for four-terminal models with an additional helper. In general, cut-set bound is less likely achievable as the system gets more complicated. (2) For the three-terminal model, there are three cuts for generating two keys, and schemes can be designed to achieve corner points of the cut-set bound for each case of the source distribution. However, for four-terminal models with a helper, there are six cuts for generating two keys, and these six cuts yield eight possible cases of the cut-set bound due to different source distributions. It is not clear whether there exists a unified design of schemes to achieve the cut-set bound for all cases.

Our contribution in this paper lies in establishing key capacity regions for four source models (see Figures 1, 2, 3, and 4) of generating a pair of keys, and our results provide

affirmative answers to both of the above issues. In all models, there are four terminals, and each terminal observes one component of a correlated vector source. Terminals $\mathcal{X}_0$ and $\mathcal{X}_1$ wish to agree on a key K_1, and terminals $\mathcal{X}_0$ and $\mathcal{X}_2$ wish to agree on another independent key K_2. The four terminals are allowed to communicate over a pubic channel, and an eavesdropper is assumed to have access to the public discussion without ambiguity. The four models differentiate from each other due to secrecy constraints. Models 1a and 1b address *symmetric key generation*, in which secrecy requirements for two keys are the same. Model 1a (*with a trusted helper*) requires that the two keys are concealed from the eavesdropper, and model 1b (*with an untrusted helper*) further requires that the two keys are concealed from terminal $\mathcal{X}_3$ in addition to the eavesdropper. In the untrusted helper case, we assume that the helper is curious but honest, i.e., the helper attempts to infer the information about the generated keys but still follows the protocol. Models 2a and 2b (*with a trusted and an untrusted helper*) have the same secrecy requirements for the two keys as models 1a and 1b, respectively, except that the key K_2 is further required to be concealed from terminal $\mathcal{X}_1$ for both models. Thus, models 2a and 2b address *asymmetric key generation*.

For all of the above four models, we establish the cut-set bound to be the key capacity region by showing that the cut-set bound is indeed achievable. Furthermore, we construct a unified achievable strategy to achieve the corner points of the cut-set bound corresponding to all cases of source distributions. The schemes to achieve different cases vary only in the rate at which each terminal reveals information to public. Thus, the achievability proof is significantly simplified. More specifically, the achievable strategy is based on random binning and joint decoding. Given such a unified strategy, we derive the Slepian-Wolf conditions that guarantee correct key agreement and derive sufficient conditions that guarantee the secrecy requirements. Then for each individual case, it is sufficient to verify the public transmission rates of terminals satisfy the derived Slepian-Wolf conditions and secrecy conditions, which can be performed easily.

The remainder of the paper is organized as follows. Section 2 introduces the system models. Section 3 provides our main results on the key capacity region for four models and describes intuition to design key generation schemes. Section 4 concludes this paper with comments. Appendix sections provide the detailed technical proofs for our main results.

2 System Models

Consider system models (see Figures 1, 2, 3, and 4) with four distinct terminals $\mathcal{X}_i$, $i = 0, \cdots, 3$, each of which observes one of the four components of a discrete memoryless vector source generated based on a joint distribution $P_{X_0X_1X_2X_3}$. Here, for simplicity, we also use $\mathcal{X}_i$ to denote the finite alphabet set from which the random variable X_i takes values. Terminal $\mathcal{X}_i$ observes n independently and identically distributed (i.i.d.) repetitions of X_i, denoted by X_i^n. The four terminals can communicate interactively via a public channel with no rate constraints. The public channel is noiseless in the sense that all four terminals and an eavesdropper can access the public discussion without ambiguity. We assume that the eavesdropper does not observe any further information such as source sequences. Without

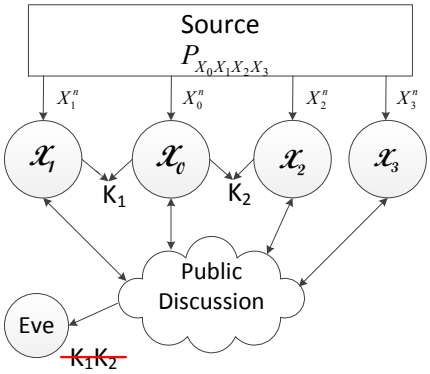

Figure 1: Model 1a. Symmetric key generation with a trusted helper

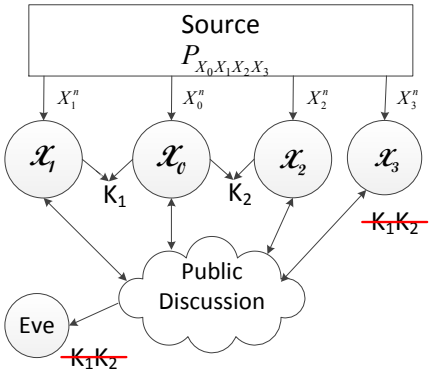

Figure 2: Model 1b. Symmetric key generation with an untrusted helper

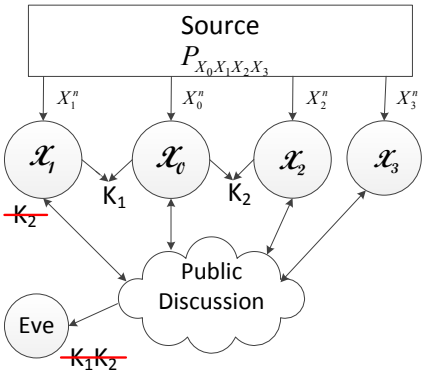

Figure 3: Model 2a. Asymmetric key generation with a trusted helper

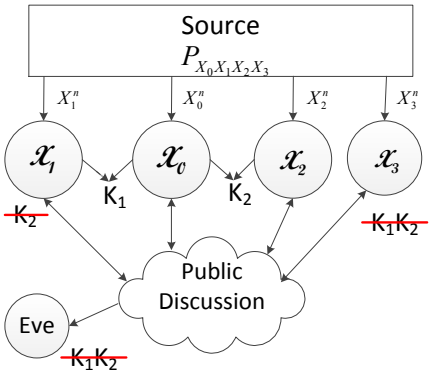

Figure 4: Model 2b. Asymmetric key generation with an untrusted helper

loss of generality, we assume that the four terminals take turns to transmit for r rounds over $4r$ consecutive slots. We use $4r$ random variables $F_1, \ldots, F_{4r}$ to denote these transmissions, where F_t denotes the transmission in time slot t for $1 \leq t \leq 4r$. The transmission F_t is a deterministic function of its own observation $X_{t \bmod 4}^n$ and all previous transmissions $F_{[1,t-1]} = (F_1, \ldots, F_{t-1})$. We use $\mathbf{F}$ to denote all transmissions in $4r$ rounds, i.e., $\mathbf{F} = (F_1, \ldots, F_{4r})$.

In this paper, we study four models, and in all models the base station terminal $\mathcal{X}_0$ wishes to agree on keys K_1 and K_2 with mobile terminals $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively, under the help of terminal $\mathcal{X}_3$. The models differentiate from each other due to secrecy requirements. Models 1a and 1b address *symmetric key generation*, in which secrecy requirements for two keys are the same. Model 1a (*with a trusted helper*) requires that the two keys are concealed from the eavesdropper, and model 1b (*with an untrusted helper*) further requires that the two keys are concealed from terminal $\mathcal{X}_3$ in addition to the eavesdropper. Models 2a and 2b (*with a trusted and an untrusted helper*) have the same secrecy requirements for the two

keys as models 1a and 1b, respectively, except that the key K_2 is further required to be concealed from terminal $\mathcal{X}_1$ for both models. Thus, models 2a and 2b address *asymmetric key generation*.

We next introduce mathematical conditions that a key pair (K_1, K_2) should satisfy. A random variable U is said to be ϵ-*recoverable* from another random variable V, if there exists a function f such that

$$\Pr\{U \neq f(V)\} < \epsilon. \tag{1}$$

For all models studied in this paper, the key K_1 is required to be ϵ-*recoverable* at terminals $\mathcal{X}_0$ and $\mathcal{X}_1$ with the public transmission $\mathbf{F}$, i.e., it can be ϵ-*recoverable* from $(X_0^n, \mathbf{F})$ and $(X_1^n, \mathbf{F})$, respectively, and the key K_2 is required to be ϵ-*recoverable* at terminals $\mathcal{X}_0$ and $\mathcal{X}_2$ with public transmission $\mathbf{F}$, i.e., it can be ϵ-*recoverable* from $(X_0^n, \mathbf{F})$ and $(X_2^n, \mathbf{F})$, respectively. Furthermore, K_1 and K_2 should satisfy the uniformity conditions:

$$\frac{1}{n} H(K_1) \geq \frac{1}{n} \log |\mathcal{K}_1| - \epsilon, \tag{2}$$

$$\frac{1}{n} H(K_2) \geq \frac{1}{n} \log |\mathcal{K}_2| - \epsilon, \tag{3}$$

where $\mathcal{K}_1$ and $\mathcal{K}_2$ denote the alphabets of the random variables K_1 and K_2, respectively.

For symmetric key generation with a trusted helper (model 1a) and with an untrusted helper (model 1b), K_1 and K_2 are required to respectively satisfy the secrecy conditions

$$\frac{1}{n} I(K_1 K_2; \mathbf{F}) < \epsilon, \tag{4}$$

and

$$\frac{1}{n} I(K_1 K_2; X_3^n \mathbf{F}) < \epsilon. \tag{5}$$

For asymmetric key generation with a trusted helper (model 2a) and with an untrusted helper (model 2b), K_1 and K_2 are required to respectively satisfy the secrecy conditions

$$\frac{1}{n} I(K_1; \mathbf{F}) < \epsilon, \quad \frac{1}{n} I(K_2; \mathbf{F}, X_1^n) < \epsilon. \tag{6}$$

and

$$\frac{1}{n} I(K_1; \mathbf{F}, X_3^n) < \epsilon, \quad \frac{1}{n} I(K_2; \mathbf{F}, X_1^n, X_3^n) < \epsilon. \tag{7}$$

We note that the parameter ϵ in (2)-(7) can be arbitrarily small as the sequence length n gets sufficiently large.

In this paper our focus is on weak secrecy requirements as given in (4)-(7). Our results can be strengthened to satisfy strong secrecy (with $1/n$ factor removed in (4)-(7)) without loss of performance by applying the idea in [13].

Definition 1. *For each of the above models, a rate pair (R_1, R_2) is said to be achievable if for every $\epsilon > 0$, $\delta > 0$, and for sufficiently large n, a key pair (K_1, K_2) can be generated to satisfy*

$$\frac{1}{n}H(K_1) > R_1 - \delta, \quad \frac{1}{n}H(K_2) > R_2 - \delta \tag{8}$$

and (4), (2), *and* (3) *for symmetric key generation with a trusted helper, to satisfy* (8) (5), (2), *and* (3) *for symmetric key generation with an untrusted helper, to satisfy* (8), (6), (2), (3) *for asymmetric key generation with a trusted helper, and to satisfy* (8), (7), (2), (3) *for asymmetric key generation with an untrusted helper.*

Our goal is to characterize the *key capacity region* that contains all achievable rate pairs (R_1, R_2) for all models.

3 Main Results

In this section, we provide characterizations of the key capacity region for the four models described in Section 2. We also provide intuitive understanding of these regions and key generation schemes for achieving these regions with the detailed technical proofs relegated to appendices.

3.1 Symmetric Key Generation with a Trusted Helper

In this subsection, we study the problem of symmetric key generation with a trusted helper, in which the generated two keys are required to be secure only from the eavesdropper. To assist the presentation, we introduce the following notations:

$$R_A := \min\{I(X_1X_3; X_0X_2), I(X_1; X_0X_2X_3)\}; \tag{9a}$$
$$R_B := \min\{I(X_0X_1; X_2X_3), I(X_2; X_0X_1X_3)\}; \tag{9b}$$
$$R_C := \min\{I(X_0; X_1X_2X_3), I(X_0X_3; X_1X_2)\}. \tag{9c}$$

The following theorem characterizes the key capacity region of this model.

Theorem 1. *The key capacity region for symmetric key generation with a trusted helper contains rate pairs (R_1, R_2) satisfying the following inequalities:*

$$R_1 \le R_A; \tag{10}$$
$$R_2 \le R_B; \tag{11}$$
$$R_1 + R_2 \le R_C. \tag{12}$$

Proof. See Appendix A. □

Since the secrecy constraints on K_1 and K_2 are symmetric, the bounds on R_1 and R_2 are also symmetric. These bounds can be intuitively understood as cut-set bounds. In particular, the upper bound on R_1 in (10) is due to two cuts separating $\mathcal{X}_0$ and $\mathcal{X}_1$ for generating K_1 (two more bounds on R_1 due to the other two cuts separating $\mathcal{X}_0$ and $\mathcal{X}_1$ become redundant due to the sum rate bound (12)). The upper bound on R_2 in (11) is due to two cuts separating $\mathcal{X}_0$ and $\mathcal{X}_2$ for generating K_2 (two more bounds on R_2 due to the other two cuts separating $\mathcal{X}_0$ and $\mathcal{X}_1$ become redundant due to the sum rate bound (12)). The sum rate bound (12) is due to the two cuts separating $\mathcal{X}_0$ and $(\mathcal{X}_1, \mathcal{X}_2)$ for generating the two keys K_1 and K_2 simultaneously.

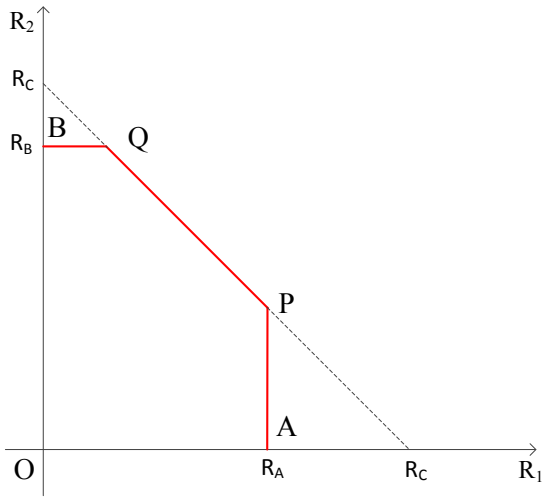

Figure 5: The key capacity region for symmetric key generation with a trusted helper.

The structure of the key capacity region is illustrated in Fig. 5 as the pentagon O-A-P-Q-B-O. We next describe the idea of constructing an achievable scheme to achieve the key capacity region. It suffices to show the achievability of the points P and Q. Since the secrecy constraints on K_1 and K_2 are symmetric, it is sufficient to show that the corner point P is achievable, and then the achievability of the point Q follows by symmetry. We assume that $R_A < R_C$, because otherwise the point P would collapse to the point A, which is shown to be achievable by the previous work [5]. The key rate pair of the point P is given by $R_1 = R_A$ and $R_2 = R_C - R_A$. Corresponding to different source distributions, each of R_A and R_C can take one of the two mutual information terms given in (9a) and (9c), respectively. Hence, the coordinates of the point P can take four forms, i.e., case 1 with $R_A = I(X_1; X_0X_2X_3)$ and $R_C = I(X_0; X_1X_2X_3)$, case 2 with $R_A = I(X_1; X_0X_2X_3)$ and $R_C = I(X_0X_3; X_1X_2)$, case 3 with $R_A = I(X_1X_3; X_0X_2)$ and $R_C = I(X_0; X_1X_2X_3)$, and case 4 with $R_A = I(X_1X_3; X_0X_2)$ and $R_C = I(X_0X_3; X_1X_2)$.

We construct a unified scheme to achieve the rate point P for all cases. In our unified scheme, terminals $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{X}_3$ reveal enough information to public so that terminal $\mathcal{X}_0$ can recover X_1^n, X_2^n and X_3^n. Then, K_1 is generated by terminals $\mathcal{X}_0$ and $\mathcal{X}_1$ based on X_1^n, and K_2 is generated by terminals $\mathcal{X}_0$ and $\mathcal{X}_2$ based on X_2^n. The schemes for the four cases are

different only in the rate at which each terminal reveals information to public. Let $\tilde{R}_1, \tilde{R}_2$ and $\tilde{R}_3$ denote the rates at which terminals $\mathcal{X}_1, \mathcal{X}_2$ and $\mathcal{X}_3$ reveal information to public, respectively. In the following, we explain the main idea of the achievable scheme for each case.

Case 1. The key rate pair of the point P is given by $(I(X_1; X_0X_2X_3), H(X_2X_3|X_1) - H(X_2X_3|X_0))$. Since the rate of K_1 needs to be maximized, terminal $\mathcal{X}_1$ should reveal as little information as possible. Hence, terminals $\mathcal{X}_2$ and $\mathcal{X}_3$ first reveal information at the rate $\tilde{R}_2 + \tilde{R}_3 = H(X_2X_3|X_0)$ so that terminal $\mathcal{X}_1$ needs to release only at the rate $\tilde{R}_1 = H(X_1|X_0X_2X_3)$ in order for $\mathcal{X}_0$ to recover X_1^n. Thus, the rate of K_1 can be as large as $I(X_1; X_0X_2X_3)$. Since the generation of K_1 already uses up information contained in X_1^n, K_2 can be generated only based on information contained in X_2^n and X_3^n given X_1^n. Thus, R_2 can be as large as $H(X_2X_3|X_1) - H(X_2X_3|X_0)$, where the subtraction is due to public discussion at the rate $\tilde{R}_2 + \tilde{R}_3 = H(X_2X_3|X_0)$.

Case 2. The key rate pair of the point P is given by $(I(X_1; X_0X_2X_3), H(X_2|X_1) - H(X_2|X_0X_3))$. The argument for R_1 is the same as that for case 1. In case 2, $R_C = I(X_0X_3; X_1X_2)$, which implies that $I(X_0X_3; X_1X_2) \leq I(X_0; X_1X_2X_3)$. This further implies $I(X_1X_2; X_3) \leq I(X_0; X_3)$. Thus, in order for terminal $\mathcal{X}_0$ to recover X_2^n and X_3^n, it is more efficient to let $\mathcal{X}_3$ reveal information first at the rate $\tilde{R}_3 = H(X_3|X_0)$, and then let $\mathcal{X}_2$ reveal information at the rate $\tilde{R}_2 = H(X_2|X_0X_3)$. Since in this case $\mathcal{X}_2$ does not recover X_3^n, the key rate $R_2 = H(X_2|X_1) - H(X_2|X_0X_3)$.

Case 3. The key rate pair of the point P is given by $(I(X_0X_2; X_1X_3), H(X_2|X_1X_3) - H(X_2|X_0))$. The case conditions $R_A = I(X_0X_2; X_1X_3)$ and $R_C = I(X_0; X_1X_2X_3)$ implies that X_0 and X_2 have high correlation, and X_1 and X_3 have high correlation. Thus, a natural idea for public discussion is to let terminal $\mathcal{X}_0$ recover X_2^n and let $\mathcal{X}_1$ recover X_3^n first, and then generate K_1 between terminals $\mathcal{X}_0$ and $\mathcal{X}_1$ to achieve $R_1 = I(X_0X_2; X_1X_3)$. Since X_1^n and X_3^n have been fully used for generating K_1, then K_2 can achieve the rate $R_2 = H(X_2|X_1X_3) - H(X_2|X_0)$, where the subtraction is due to the public transmission of terminal $\mathcal{X}_2$ to let $\mathcal{X}_0$ recover X_2^n.

Case 4. This case does not exist because of the contradiction induced by setting $R_A = I(X_0X_2; X_1X_3)$ and $R_C = I(X_0X_3; X_1X_2)$ (see (70), (71) in the proof for details).

3.2 Symmetric Key Generation with an Untrusted Helper

In this subsection, we study the problem of symmetric key generation with an untrusted helper, in which the two generated keys are required to be secure from both the eavesdropper and the helper terminal $\mathcal{X}_3$. The following theorem characterizes the key capacity region of this model.

Theorem 2. *The key capacity region for symmetric key generation with an untrusted helper*

contains rate pairs (R_1, R_2) *satisfying the following inequalities:*

$$R_1 < I(X_1; X_0X_2|X_3), \tag{13}$$
$$R_2 < I(X_2; X_0X_1|X_3), \tag{14}$$
$$R_1 + R_2 < I(X_0; X_1X_2|X_3). \tag{15}$$

Proof. See Appendix B. □

These bounds can also be intuitively understood as cut-set bounds. In particular, the upper bound on R_1 in (13) is due to the cut separating $\mathcal{X}_0$ and $\mathcal{X}_1$ for generating K_1 (one more bound due to the other cut separating $\mathcal{X}_0$ and $\mathcal{X}_1$ is redundant due to the sum rate bound (15)). The upper bound on R_2 in (14) is due to the cut separating $\mathcal{X}_0$ and $\mathcal{X}_2$ for generating K_2 (one more bound due to the other cut separating $\mathcal{X}_0$ and $\mathcal{X}_1$ is redundant due to the sum rate bound (15)). The sum rate bound (15) is due to the cut separating $\mathcal{X}_0$ and $(\mathcal{X}_1, \mathcal{X}_2)$ for generating two keys K_1 and K_2 simultaneously. All the bounds (13)-(15) are conditioned on X_3 because both K_1 and K_2 are required to be secure from terminal $\mathcal{X}_3$.

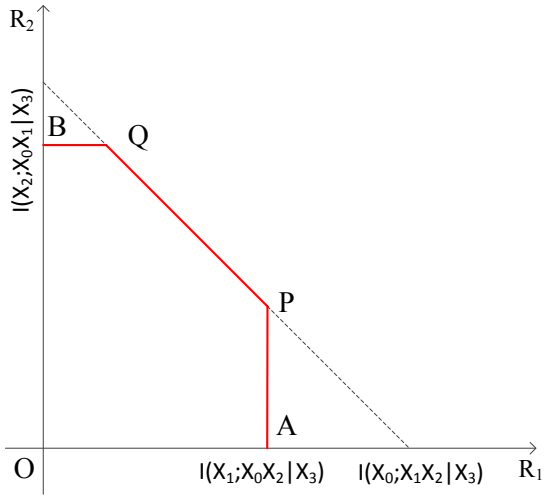

Figure 6: The key capacity region for symmetric key generation with an untrusted helper

Comparing Theorem 2 with Theorem 1, it is clear that the key capacity region for symmetric key generation with an untrusted helper is contained in the key capacity region for symmetric key generation with a trusted helper, because it always holds that $I(X_1; X_0X_2|X_3) \leq R_A$, $I(X_2; X_0X_1|X_3) \leq R_B$ and $I(X_0; X_1X_2|X_3) \leq R_C$. This is reasonable due to the additional requirement for the keys to be concealed from the helper when the helper is untrusted.

The structure of the key capacity region is illustrated in Fig. 6 as the pentagon O-A-P-Q-B-O. In order to justify the achievability of the region, by symmetry, it is sufficient to show the achievability of the point P in Fig. 6. The rate pair at point P is given by $(I(X_1; X_0X_2|X_3), H(X_2|X_1X_3) - H(X_2|X_0X_3))$. The idea to achieve the point P follows the

unified strategy described in Section 3.1, i.e., public discussion first guarantees that terminal $\mathcal{X}_0$ recovers X_1^n, X_2^n and X_3^n correctly, and then K_1 is generated by terminals $\mathcal{X}_0$ and $\mathcal{X}_1$ based on X_1^n and K_2 is generated by terminals $\mathcal{X}_0$ and $\mathcal{X}_2$ based on X_2^n. Since the generated keys should be concealed from the helper terminal $\mathcal{X}_3$, X_3^n cannot be used as random resource for generating the keys, although the helper can still participate the public discussion to assist the recovery of source sequences.

More specifically, since the rate of K_1 needs to be maximized, terminal $\mathcal{X}_1$ should reveal as little information as possible. Hence, terminals $\mathcal{X}_2$ and $\mathcal{X}_3$ first reveal information at the rate $\tilde{R}_2 + \tilde{R}_3 = H(X_2X_3|X_0)$ so that terminal $\mathcal{X}_1$ needs to release only at the rate $\tilde{R}_1 = H(X_1|X_0X_2X_3)$ in order for $\mathcal{X}_0$ to recover X_1^n. Thus, the rate of K_1 can be as large as $I(X_1; X_0X_2|X_3)$. Since $\mathcal{X}_3$ is an untrusted helper, it can reveal information at any rate. Hence, $\tilde{R}_2$ can be set to be $H(X_2|X_0X_3)$. Consequently, $R_2 = H(X_2|X_1X_3) - H(X_2|X_0X_3)$. Since the generation of K_1 already uses up information contained in X_1^n, K_2 can be generated only based on information contained in X_2^n given X_1^n and X_3^n.

3.3 Asymmetric Key Generation with a Trusted Helper

In this subsection, we study the problem of asymmetric key generation with a trusted helper, in which the two generated keys are required to be kept secure from the eavesdropper, and furthermore, the key K_2 is also required to be kept secure from terminal $\mathcal{X}_1$. For concise presentation, we introduce the following notations:

$$R_A := \min\{I(X_1X_3; X_0X_2), I(X_1; X_0X_2X_3)\}; \tag{16a}$$
$$R'_B := \min\{I(X_0; X_2X_3|X_1), I(X_2; X_0X_3|X_1)\}; \tag{16b}$$
$$R_C := \min\{I(X_0; X_1X_2X_3), I(X_0X_3; X_1X_2)\}. \tag{16c}$$

We note that the expressions of R_A and R_C remain unchanged comparing with those in Section 3.1, but R'_B is different from R_B by having X_1 in the conditioning.

The following theorem characterizes the key capacity region of the model of interest.

Theorem 3. *The key capacity region for asymmetric key generation with a trusted helper contains rate pairs (R_1, R_2) satisfying the following inequalities:*

$$R_1 \le R_A; \tag{17}$$
$$R_2 \le R'_B; \tag{18}$$
$$R_1 + R_2 \le R_C. \tag{19}$$

Proof. See Appendix C. □

We note that if X_3^n is independent of (X_0^n, X_1^n, X_2^n), then the helper terminal is not able to help in the key generation. In such a case, the key capacity region characterized in Theorem 3 reduces to that established in [12] for the same model without a helper.

The bounds in Theorem 3 can be intuitively understood as cut-set bounds. In particular, the upper bound on R_1 in (17) is due to the two cuts separating $\mathcal{X}_0$ and $\mathcal{X}_1$ for generating K_1 (two more bounds on R_1 due to the other two cuts separating $\mathcal{X}_0$ and $\mathcal{X}_1$ are redundant due to the sum rate bound (19)). The sum rate bound (19) is due to the cut separating $\mathcal{X}_0$ and $(\mathcal{X}_1, \mathcal{X}_2)$ for generating two keys K_1 and K_2 simultaneously. The upper bound on R_2 in (18) is due to the cut separating $\mathcal{X}_0$ and $\mathcal{X}_2$ for generating K_2, and the conditioning on X_1 is due to the requirement that K_2 be concealed from $\mathcal{X}_1$.

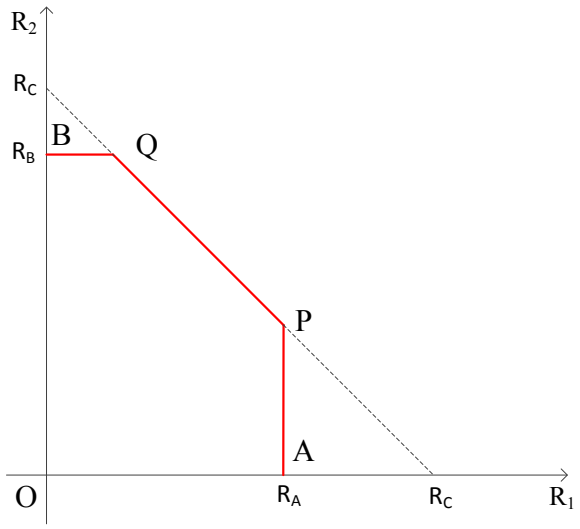

Figure 7: The key capacity region for asymmetric key generation with a trusted helper

The key capacity region is illustrated in Fig. 7 as the pentagon O-A-P-Q-B-O. We next describe our idea of constructing an achievable scheme to achieve the key capacity region. Since the secrecy requirements for the two keys are different, we need to design achievable schemes to achieve the points P and Q separately. Corresponding to different source distributions, the rate pairs of the points P and Q can take different forms. Interestingly, the same unified strategy described for the previous models can achieve the points P and Q for all cases. Namely, terminals $\mathcal{X}_1, \mathcal{X}_2$ and $\mathcal{X}_3$ reveal information to public such that terminal $\mathcal{X}_0$ can recover (X_1^n, X_2^n, X_3^n) correctly. Then K_1 is generated based on X_1^n and K_2 is generated based on X_2^n. The schemes for different cases vary only in the rate at which each terminal reveals information to public. Here, we still use $\tilde{R}_1, \tilde{R}_2$ and $\tilde{R}_3$ to denote the rates at which terminals $\mathcal{X}_1, \mathcal{X}_2$ and $\mathcal{X}_3$ reveal information to public, respectively.

For the point P, it can be observed that its rate coordinates are exactly the same as those in Section 3.1, and can be shown to be achievable by the same schemes designed in Section 3.1 for symmetric key generation with a trusted helper. This is because both models should reach the same maximum rate R_1 of K_1 due to the same secrecy requirement for K_1. Furthermore, since both models should exhaust all random resource in X_1^n for generating K_1, K_2 should be generated from random resource independent from X_1^n even if it is not required to be concealed from terminal $\mathcal{X}_1$ in symmetric key generation. Thus, the two models also

have the same rate R_2 at the point P.

For the point Q, the rate coordinates are given by $(R_C - R'_B, R'_B)$. Corresponding to different source distributions, each of R_C and R'_B can take one of the two mutual information terms given in (19) and (18), respectively. Hence, the coordinates of the point Q can take four forms, i.e., case 1 with $R_C = I(X_0; X_1X_2X_3)$ and $R'_B = I(X_0; X_2X_3|X_1)$; case 2 with $R_C = I(X_0; X_1X_2X_3)$ and $R'_B = I(X_2; X_0X_3|X_1)$; case 3 with $R_C = I(X_0X_3; X_1X_2)$ and $R'_B = I(X_0; X_2X_3|X_1)$; and case 4 with $R_C = I(X_0X_3; X_1X_2)$ and $R'_B = I(X_2; X_0X_3|X_1)$. In the following, we explain the idea of the achievable scheme for each case and relegate the detailed proof to Appendix C. We note that achievable designs for the following cases are different due to the source distributions that determine these cases.

Case 1. The key rate pair of the point Q is given by $(I(X_0; X_1), I(X_0; X_2X_3|X_1))$. In order to generate K_1 at the rate $I(X_0; X_1)$, terminal $\mathcal{X}_1$ can reveal information at the rate $\tilde{R}_1 = H(X_1|X_0)$. Then terminal $\mathcal{X}_0$ can recover X_1^n correctly. In order for terminal $\mathcal{X}_0$ to further recover (X_2^n, X_3^n), terminals $\mathcal{X}_2$ and X_3 jointly release information at the sum rate $\tilde{R}_2 + \tilde{R}_3 = H(X_2X_3|X_0X_1)$. Since the resource to generate K_2 should be contained in (X_2^n, X_3^n) given X_1^n, the key rate R_2 should satisfy $R_2 = H(X_2X_3|X_1) - \tilde{R}_2 - \tilde{R}_3$ which yields $R_2 = I(X_0; X_2X_3|X_1)$.

Case 2. The key rate pair of the point Q is given by $(I(X_0; X_1X_3) - I(X_2; X_3|X_1), I(X_2; X_0X_3|X_1))$. Terminals $\mathcal{X}_1$ and $\mathcal{X}_3$ first jointly reveal information at the sum rate $\tilde{R}_1 + \tilde{R}_3 = H(X_1X_3|X_0)$. Then terminal $\mathcal{X}_0$ recovers X_1^n and X_3^n correctly. Terminal $\mathcal{X}_2$ needs to release information only at the rate $\tilde{R}_2 = H(X_2|X_0X_1X_3)$ and then X_2^n can be successfully recovered at $\mathcal{X}_0$. Thus, K_2 can be generated at the rate $H(X_2|X_1) - H(X_2|X_0X_1X_3)$, which yields $R_2 = I(X_2; X_0X_3|X_1)$. In order to generate K_1 at the rate $R_1 = I(X_0; X_1X_3) - I(X_2; X_3|X_1)$, $\tilde{R}_1$ should be chosen to be $\tilde{R}_1 = H(X_1X_3|X_0) - H(X_3|X_1X_2)$, and then $\tilde{R}_3 = H(X_3|X_1X_2)$.

Case 3. This case does not exist because of the contradiction induced by setting $R_C = I(X_0X_3; X_1X_2)$ and $R'_B = I(X_0; X_2X_3|X_1)$ (see (117) and (118) in the proof for details).

Case 4. The key rate pair of the point Q is given by $(I(X_1; X_0X_3), I(X_2; X_0X_3|X_1))$. Still, terminals $\mathcal{X}_1$ and $\mathcal{X}_3$ first reveal information at the sum rate $\tilde{R}_1 + \tilde{R}_3 = H(X_1X_3|X_0)$. Due to the case conditions, X_3 has high correlation with X_0 and hence let $\mathcal{X}_3$ reveal its information at the rate $\tilde{R}_3 = H(X_3|X_0)$. Then $\mathcal{X}_1$ reveals its information at the rate $\tilde{R}_1 = H(X_1|X_0X_3)$, and finally $\mathcal{X}_2$ reveals its information at the rate $\tilde{R}_2 = H(X_2|X_0X_1X_3)$ in order for $\mathcal{X}_0$ to recover (X_1^n, X_2^n, X_3^n). Thus, the key rate $R_2 = I(X_0X_3; X_2|X_1)$, and the rate $R_1 = I(X_0X_3; X_1)$.

We note that for asymmetric key generation with a trusted helper, the point Q achieves the same sum rate as that for symmetric key generation with a trusted helper. This is because although the rate of K_2 decreases in the asymmetric model due to the additional secrecy requirement for K_2 to be concealed from $\mathcal{X}_1$, the random resource contained in X_1^n can still be used for generating K_1 so that there is no loss in the sum rate.

3.4 Asymmetric Key Generation with an Untrusted Helper

In this subsection, we study the problem of asymmetric key generation with an untrusted helper, in which the two generated keys are required to be kept secure from both the eavesdropper and the helper terminal $\mathcal{X}_3$, and furthermore, the key K_2 is required to be kept secure from terminal $\mathcal{X}_1$. The following theorem characterizes the key capacity region for this model.

Theorem 4. *The key capacity region for asymmetric key generation with an untrusted helper contains rate pairs (R_1, R_2) satisfying the following inequalities:*

$$R_1 < I(X_1; X_0X_2|X_3), \tag{20}$$
$$R_2 < I(X_2; X_0|X_1X_3), \tag{21}$$
$$R_1 + R_2 < I(X_0; X_1X_2|X_3). \tag{22}$$

Proof. See Appendix D. □

These bounds can also be intuitively understood as cut-set bounds. In particular, the upper bound on R_1 in (20) is due to the cut separating $\mathcal{X}_0$ and $\mathcal{X}_1$ for generating K_1 (one more bound due tothe other cut separating $\mathcal{X}_0$ and $\mathcal{X}_1$ is redundant due to the sum rate bound (22)). The upper bound on R_2 in (21) is due to the cut separating $\mathcal{X}_0$ and $\mathcal{X}_2$ for generating K_2, and the conditioning on X_1 is due to the requirement that K_2 be concealed from $\mathcal{X}_1$. The sum rate bound (22) is due to the cut separating $\mathcal{X}_0$ and $(\mathcal{X}_1, \mathcal{X}_2)$ for generating two keys K_1 and K_2 simultaneously. All these bounds (20)-(22) are conditioned on X_3 because both K_1 and K_2 are required to be kept secure from terminal $\mathcal{X}_3$.

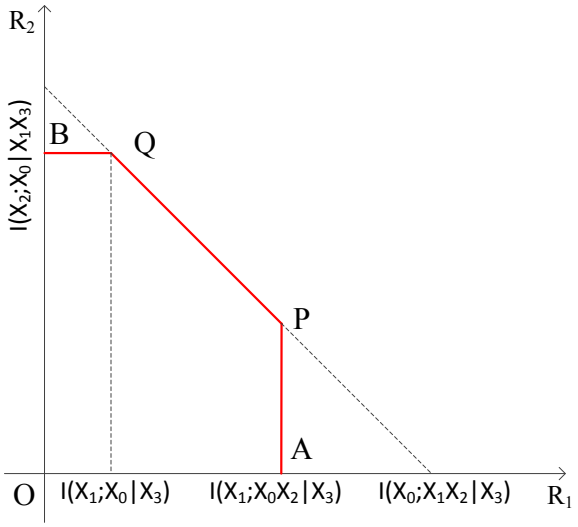

Figure 8: The key capacity region for asymmetric key generation with an untrusted helper.

Similarly to the symmetric case, comparing Theorem 4 with Theorem 3, it is clear that the key capacity region for asymmetric key generation with an untrusted helper is contained in that for the asymmetric key generation with a trusted helper, because it always holds that $I(X_1; X_0X_2|X_3) \leq R_A$, $I(X_2; X_0|X_1X_3) \leq R'_B$ and $I(X_0; X_1X_2|X_3) \leq R_C$. This is anticipated due to the additional requirement for the keys to be concealed from the helper.

The key capacity region is illustrated in Fig. 8 as the pentagon O-A-P-Q-B-O. It can be shown that the point P can be achieved by the same scheme as that for achieving the point P in symmetric key generation with an untrusted helper in Section 3.2. We next describe the idea to achieve the point Q. The rate pair of the point Q is given by $(I(X_0; X_1|X_3), I(X_2; X_0|X_1X_3))$. Terminals $\mathcal{X}_1$ and $\mathcal{X}_3$ first reveal information at the rate $\tilde{R}_1+\tilde{R}_3 = H(X_1X_3|X_0)$ so that terminal $\mathcal{X}_0$ can recover X_1^n and X_3^n correctly. Then terminal $\mathcal{X}_2$ needs to release only at the rate $\tilde{R}_2 = H(X_2|X_0X_1X_3)$ in order for $\mathcal{X}_0$ to recover X_2^n. Thus, the rate of K_2 can be as large as $I(X_2; X_0|X_1X_3)$. Since $\mathcal{X}_3$ is an untrusted helper, it can reveal information at any rate. Hence, $\tilde{R}_1$ can be set to be $H(X_1|X_0X_3)$. Consequently, $R_1 = H(X_1|X_3) - H(X_1|X_0X_3) = I(X_0; X_1|X_3)$.

4 Conclusion

In this paper, we have studied the problem of generating a pair of keys for a cellular source model with a helper. We have established the full key capacity region for four models with different secrecy requirements. The models studied here consist of four terminals, which are more complicated to analyze than three-terminal models studied previously, because the cut-set outer bound takes more cases due to different source distributions. Instead of designing a specific achievable scheme for each case one by one, we have developed a unified strategy, which achieves corner points for all cases, and hence significantly reduces the complexity of the achievability proof. It will be of future interest to generalize the studies here to cellular models with more than two mobile terminals, in which each terminal wishes to generate a key with the base station terminal. In such a case, a unified strategy is desirable to facilitate feasible analysis. As another direction, it will be interesting to study this type of multiple key generation problems with rate constraints on the public discussion. For such a case, previous studies of the source model with the helper subject to finite rate constraints in [4] and of vector Gaussian source model with public discussion subject to finite rate constraints in [14] provide useful techniques.

Appendix

A Proof of Theorem 1

Proof of Converse. First, if we need only to generate K_1 , the model reduces to the secret key generation problem studied in [5]. The key capacity is shown to be $\min\{R_A, R_C\}$, which provides an upper bound (10) on R_1. Similarly, if we dedicate to generate K_2, the key capacity is shown to be $\min\{R_B, R_C\}$ in [5], which provides an upper bound (11) on R_2. For the sum rate bound, we consider an enhanced model by replacing terminals $\mathcal{X}_1$ and $\mathcal{X}_2$ with a super terminal $\mathcal{X}_s$ that observes both X_1^n and X_2^n. The secret key rate between $\mathcal{X}_0$ and $\mathcal{X}_s$ is upper bounded by R_C as shown in [5].

Proof of Achievability. We design an achievable scheme to achieve the key capacity region plotted in Fig. 5 as the pentagon O-A-P-Q-B-O, where the coordinates of the points A and B are $(\min\{R_A, R_C\}, 0)$ and $(0, \min\{R_B, R_C\})$, respectively. The corner point A is achieved by letting $\mathcal{X}_2$ be a dedicated helper to generate K_1 following the omniscience scheme in [5]. Similarly, the corner point B is achieved by letting $\mathcal{X}_1$ be a dedicated helper to generate K_2 as shown in [5]. We note that the point P would collapse to the point A if $R_C \leq R_A$ and the point Q would collapse to the point B if $R_C \leq R_B$. It is thus sufficient to show that the points P and Q are achievable whenever they are different from the points A and B, respectively. Then the entire pentagon can be achieved by time sharing.

We note that since the secrecy constraints on K_1 and K_2 are symmetric, it is sufficient to show that the corner point P in Fig. 5 is achievable and the achievability of the point Q follows by symmetry. We assume that $R_A < R_C$, because otherwise the point P would collapse to the point A and has been justified to be achievable. The key rate pair of the point P can take different forms due to different source distributions. In the following, we first describe a unified scheme that is applicable to all cases, and then study each case one by one. In general, our scheme is based on random binning and joint decoding.

Codebook Generation: At terminal $\mathcal{X}_1$, randomly and independently assign a bin index f to each sequence $x_1^n \in \mathcal{X}_1^n$, where $f \in [1 : 2^{n\tilde{R}_1}]$. We use $f(x_1^n)$ to denote the bin index of the sequence x_1^n, and use $B_1(f)$ to denote the bin indexed by f. Then randomly and independently assign a sub-bin index ϕ to each sequence in each nonempty bin $B_1(f)$, where $\phi \in [1 : 2^{nR_1}]$. We further use $B_1(f, \phi)$ to denote the sub-bin indexed by ϕ within the bin $B_1(f)$.

At terminal $\mathcal{X}_2$, randomly and independently assign a bin index g to each sequence $x_2^n \in \mathcal{X}_2^n$, where $g \in [1 : 2^{n\tilde{R}_2}]$. We use $g(x_2^n)$ to denote the bin index of the sequence x_2^n, and use $B_2(g)$ to denote the bin indexed by g. Then randomly and independently assign a sub-bin index ψ to each sequence in each nonempty bin $B_2(g)$, where $\psi \in [1 : 2^{nR_2}]$. We further use $B_2(g, \psi)$ to denote the sub-bin indexed by ψ within the bin $B_2(g)$.

At terminal $\mathcal{X}_3$, randomly and independently assign a bin index l to each sequence $x_3^n \in \mathcal{X}_3^n$, where $l \in [1 : 2^{n\tilde{R}_3}]$. We use $l(x_3^n)$ to denote the bin index of the sequence x_3^n, and use $B_3(l)$ to denote the bin indexed by l.

The codebook is revealed to all parties, i.e., terminals $\mathcal{X}_0, \mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3$ and the eavesdropper.

Encoding and Transmission: Given a sequence x_1^n, terminal $\mathcal{X}_1$ finds the index pair (f, ϕ) such that $x_1^n \in B_1(f, \phi)$. Then it reveals the index $f = f(x_1^n)$ over the public channel to all parties.

Given a sequence x_2^n, terminal $\mathcal{X}_2$ finds the index pair (g, ψ) such that $x_2^n \in B_2(g, \psi)$. Then it reveals the index $g = g(x_2^n)$ over the public channel to all parties.

Given a sequence x_3^n, terminal $\mathcal{X}_3$ finds the index l such that $x_3^n \in B_3(l)$. Then it reveals the index $l = l(x_3^n)$ over the public channel to all parties.

Decoding: The decoding scheme is based on joint typicality.

Terminal $\mathcal{X}_0$, given x_0^n and the bin indexes f, g and l, claims $\hat{x}_1^n$, $\hat{x}_2^n$ and $\hat{x}_3^n$ are observations of terminals $\mathcal{X}_1, \mathcal{X}_2$ and $\mathcal{X}_3$, respectively, if there exists a unique tuple of sequences $(\hat{x}_1^n, \hat{x}_2^n, \hat{x}_3^n)$ such that $\hat{x}_1^n \in B_1(f)$, $\hat{x}_2^n \in B_2(g)$, $\hat{x}_3^n \in B_3(l)$, and $(x_0^n, \hat{x}_1^n, \hat{x}_2^n, \hat{x}_3^n) \in T_\epsilon^{(n)}(P_{X_0X_1X_2X_3})$.

Based on Slepian-Wolf coding theorem [5, 15], the decoding error can be arbitrarily small if the rates $\tilde{R}_1, \tilde{R}_2$ and $\tilde{R}_3$ satisfy the following Slepian-Wolf conditions:

$$\tilde{R}_1 > H(X_1|X_0X_2X_3), \tag{23}$$

$$\tilde{R}_2 > H(X_2|X_0X_1X_3), \tag{24}$$

$$\tilde{R}_3 > H(X_3|X_0X_1X_2), \tag{25}$$

$$\tilde{R}_1 + \tilde{R}_2 > H(X_1X_2|X_0X_3), \tag{26}$$

$$\tilde{R}_1 + \tilde{R}_3 > H(X_1X_3|X_0X_2), \tag{27}$$

$$\tilde{R}_2 + \tilde{R}_3 > H(X_2X_3|X_0X_1), \tag{28}$$

$$\tilde{R}_1 + \tilde{R}_2 + \tilde{R}_3 > H(X_1X_2X_3|X_0). \tag{29}$$

Key Generation: Terminal $\mathcal{X}_1$ sets $K_1 = \phi(X_1^n)$. Terminal $\mathcal{X}_2$ sets $K_2 = \psi(X_2^n)$. Terminal $\mathcal{X}_0$ sets $\hat{K}_1 = \phi(\hat{X}_1^n)$ and $\hat{K}_2 = \psi(\hat{X}_2^n)$. If the decoding error vanishes asymptotically (i.e., (23)-(29) are satisfied), we have

$$\Pr\{K_1 = \hat{K}_1\} > 1 - \epsilon, \tag{30}$$

$$\Pr\{K_2 = \hat{K}_2\} > 1 - \epsilon. \tag{31}$$

Secrecy: We derive the sufficient conditions for achieving the secrecy requirement (4). Then these sufficient conditions need to be verified for each of the four cases later on.

We evaluate the key leakage rates averaged over the random codebook ensemble. Let $f := f(X_1^n)$, $g := g(X_2^n)$ and $l := l(X_3^n)$. Then it is clear that $\mathbf{F} = \{f, g, l\}$. We further let $\phi := \phi(X_1^n)$ and $\psi := \psi(X_2^n)$. Hence, $K_1 = \phi$ and $K_2 = \psi$. We first derive

$$\begin{aligned} I(K_1K_2; \mathbf{F}|\mathcal{C}) &= I(\phi, \psi; f, g, l|\mathcal{C}) \\ &= I(\psi; f, g, l|\mathcal{C}) + I(\phi; f, g, l|\psi, \mathcal{C}) \\ &\le I(\psi; g|\mathcal{C}) + I(\psi; f, l|g, \mathcal{C}) + I(\phi; f, g, \psi, l|\mathcal{C}) \\ &\le I(\psi; g|\mathcal{C}) + I(g, \psi; f, l|\mathcal{C}) + I(\phi; f|\mathcal{C}) + I(\phi; g, \psi, l|f, \mathcal{C}) \\ &\le I(\psi; g|\mathcal{C}) + I(g, \psi; f, l|\mathcal{C}) + I(\phi; f|\mathcal{C}) + I(f, \phi; g, \psi, l|\mathcal{C}). \end{aligned} \tag{32}$$

We next consider each of the four terms in (32). It can be shown as in [10, Appendix A] that if

$$\tilde{R}_1 + R_1 < H(X_1) - 2\delta(\epsilon), \tag{33}$$

then

$$\frac{1}{n} I(\phi; f|\mathcal{C}) < \delta(\epsilon); \tag{34}$$

and if

$$\tilde{R}_2 + R_2 < H(X_2) - 2\delta(\epsilon), \tag{35}$$

then

$$\frac{1}{n} I(\psi; g|\mathcal{C}) < \delta(\epsilon). \tag{36}$$

In order to bound the second term in (32), we have the following derivation:

$$\begin{aligned}
&I(g, \psi; f, l|\mathcal{C}) \\
&\quad \le I(X_2^n, g, \psi; f, l|\mathcal{C}) \\
&\quad = I(X_2^n; f, l|\mathcal{C}) \\
&\quad = I(X_2^n; X_1^n, X_3^n|\mathcal{C}) - I(X_2^n; X_1^n, X_3^n|f, l, \mathcal{C}) \\
&\quad = H(X_1^n, X_3^n|\mathcal{C}) - H(X_1^n, X_3^n|X_2^n, \mathcal{C}) - H(X_1^n, X_3^n|f, l, \mathcal{C}) + H(X_1^n, X_3^n|X_2^n, f, l, \mathcal{C}) \\
&\quad \le H(X_1^n, X_3^n|\mathcal{C}) - H(X_1^n, X_3^n|X_2^n, \mathcal{C}) - [H(X_1^n, X_3^n|\mathcal{C}) - n\tilde{R}_1 - n\tilde{R}_3] \\
&\qquad + H(X_1^n, X_3^n|X_2^n, f, l, \mathcal{C}) \\
&\quad \le n[\tilde{R}_1 + \tilde{R}_3 - H(X_1X_3|X_2)] + H(X_1^n, X_3^n|X_2^n, f, l, \mathcal{C}).
\end{aligned}$$

It can be shown as in [10, Appendix A] that if

$$\tilde{R}_1 + \tilde{R}_3 \le H(X_1X_3|X_2) - 2\delta(\epsilon), \tag{37}$$

then

$$\limsup_{n\to\infty} \frac{1}{n} H(X_1^n, X_3^n|X_2^n, f, l, \mathcal{C}) < H(X_1X_3|X_2) - \tilde{R}_1 - \tilde{R}_3 + \delta(\epsilon). \tag{38}$$

Consequently,

$$\frac{1}{n} I(g, \psi; f, l|\mathcal{C}) < \delta(\epsilon). \tag{39}$$

Next we consider the last term in (32) as follows:

$$\begin{aligned}
&I(f, \phi; g, \psi, l|\mathcal{C}) \le I(X_1^n; g, \psi, l|\mathcal{C}) \\
&= I(X_1^n; X_2^n, X_3^n, g, \psi, l|\mathcal{C}) - I(X_1^n; X_2^n, X_3^n|g, \psi, l, \mathcal{C}) \\
&= I(X_1^n; X_2^n, X_3^n|\mathcal{C}) - I(X_1^n; X_2^n, X_3^n|g, \psi, l, \mathcal{C}) \\
&= H(X_2^n, X_3^n|\mathcal{C}) - H(X_2^n, X_3^n|X_1^n, \mathcal{C}) - H(X_2^n, X_3^n|g, \psi, l, \mathcal{C}) + H(X_2^n, X_3^n|X_1^n, g, \psi, l, \mathcal{C}) \\
&\le n[\tilde{R}_2 + R_2 + \tilde{R}_3 - H(X_2X_3|X_1)] + H(X_2^n, X_3^n|X_1^n, g, \psi, l, \mathcal{C}).
\end{aligned} \tag{40}$$

It can be shown that if

$$\tilde{R}_2 + R_2 + \tilde{R}_3 < H(X_2X_3|X_1) - 2\delta(\epsilon), \tag{41}$$

then

$$\limsup_{n\to\infty}\frac{1}{n}H(X_2^n, X_3^n|X_1^n, g, \psi, l, \mathcal{C}) < H(X_2X_3|X_1) - \tilde{R}_2 - R_2 - \tilde{R}_3 + \delta(\epsilon).$$

Consequently,

$$\frac{1}{n}I(f, \phi; g, \psi, l|\mathcal{C}) < \delta(\epsilon). \tag{42}$$

Therefore, (33), (35), (37) and (41) are sufficient conditions that guarantee the secrecy requirement (4).

Uniformity: Uniformity of keys is due to properties of random binning and typicality.

We next show the achievability of the point P with rate coordinates $R_1 = R_A$ and $R_2 = R_C - R_A$. Corresponding to different source distributions, each of R_A and R_C can take one of the two mutual information terms given in (9a) and (9c), respectively. Hence, the coordinates of the point P can take four forms, i.e., case 1 with $R_A = I(X_1; X_0X_2X_3)$ and $R_C = I(X_0; X_1X_2X_3)$, case 2 with $R_A = I(X_1; X_0X_2X_3)$ and $R_C = I(X_0X_3; X_1X_2)$, case 3 with $R_A = I(X_0X_2; X_1X_3)$ and $R_C = I(X_0; X_1X_2X_3)$, and case 4 with $R_A = I(X_0X_2; X_1X_3)$ and $R_C = I(X_0X_3; X_1X_2)$.

For each case, it is sufficient to set the rates $\tilde{R}_1$, R_1, $\tilde{R}_2$, R_2 and $\tilde{R}_3$ to satisfy the Slepian-Wolf conditions (23)-(29) for guaranteeing correct key agreement and to satisfy the sufficient conditions (33), (35), (37) and (41) for guaranteeing secrecy.

Case 1: $R_A = I(X_1; X_0X_2X_3)$ and $R_C = I(X_0; X_1X_2X_3)$, which imply

$$H(X_3|X_0X_2) < H(X_3|X_1), \tag{43}$$
$$H(X_3|X_1X_2) < H(X_3|X_0). \tag{44}$$

Moreover, $R_A < R_C$ implies

$$H(X_2X_3|X_0) < H(X_2X_3|X_1). \tag{45}$$

The rate pair at the point P is given by $(I(X_1; X_0X_2X_3), H(X_2X_3|X_1) - H(X_2X_3|X_0))$. To achieve this rate pair, we set the binning rates in the achievable strategy as follows:

$$\tilde{R}_1 = H(X_1|X_0X_2X_3) + \epsilon, \tag{46}$$
$$R_1 = I(X_1; X_0X_2X_3) - 2\delta(\epsilon) - 2\epsilon, \tag{47}$$
$$\tilde{R}_2 = H(X_2X_3|X_0) - \tilde{R}_3 + \epsilon, \tag{48}$$
$$R_2 = H(X_2X_3|X_1) - H(X_2X_3|X_0) - 4\delta(\epsilon) - 3\epsilon, \tag{49}$$
$$\tilde{R}_3 = \min\{H(X_3|X_2), H(X_3|X_0)\} - 2\delta(\epsilon) - 2\epsilon. \tag{50}$$

It can be verified that the above rates satisfy the Slepian-Wolf conditions (23)-(29) if $\tilde{R}_3 > H(X_3|X_0X_1X_2)$. Otherwise, either the Markov chain $X_3 - X_0 - X_1X_2$ or the Markov chain $X_3 - X_2 - X_0X_1$ holds and thus the rate pair can be easily achieved without the

helper's assistance. It can also be verified that the above rates (46)-(50) satisfy the sufficient conditions (33), (35), (37) and (41) for secrecy.

Case 2: $R_A = I(X_1; X_0X_2X_3)$ and $R_C = I(X_0X_3; X_1X_2)$, which imply the following two inequalities:

$$H(X_3|X_0X_2) < H(X_3|X_1), \tag{51}$$
$$H(X_3|X_0) < H(X_3|X_1X_2). \tag{52}$$

Here, $R_A < R_C$ is equivalent to

$$H(X_2|X_0X_3) < H(X_2|X_1). \tag{53}$$

The key rate pair at the point P in this case is given by $(I(X_1; X_0X_2X_3), H(X_2|X_1) - H(X_2|X_0X_3))$. To achieve this rate pair, we set the binning rates in the achievable strategy as follows:

$$\tilde{R}_1 = H(X_1|X_0X_2X_3) + \epsilon, \tag{54}$$
$$R_1 = I(X_1; X_0X_2X_3) - 2\delta(\epsilon) - 2\epsilon, \tag{55}$$
$$\tilde{R}_2 = H(X_2|X_0X_3) + \epsilon, \tag{56}$$
$$R_2 = H(X_2|X_1) - H(X_2|X_0X_3) - 2\delta(\epsilon) - 2\epsilon, \tag{57}$$
$$\tilde{R}_3 = H(X_3|X_0) + \epsilon. \tag{58}$$

It can be verified that the above rates satisfy the Slepian-Wolf conditions (23)-(29) and the sufficient conditions (33), (35), (37) and (41) for secrecy. In particular, (37) holds due to (52) as follows:

$$\begin{aligned}\tilde{R}_1 + \tilde{R}_3 &= H(X_1|X_0X_2X_3) + H(X_3|X_0) + 2\epsilon \\ &< H(X_1|X_2X_3) + H(X_3|X_1X_2) - 2\delta(\epsilon) \\ &< H(X_1X_3|X_2) - 2\delta(\epsilon),\end{aligned}$$

and (41) holds due to (52) as follows:

$$\begin{aligned}\tilde{R}_2 + R_2 + \tilde{R}_3 &= H(X_3|X_0) + H(X_2|X_1) - 2\delta(\epsilon) - \epsilon \\ &< H(X_3|X_1X_2) + H(X_2|X_1) - 2\delta(\epsilon) \\ &< H(X_2X_3|X_1) - 2\delta(\epsilon).\end{aligned}$$

Case 3: $R_A = I(X_0X_2; X_1X_3)$ and $R_C = I(X_0; X_1X_2X_3)$, which imply the following two inequalities:

$$H(X_3|X_0X_2) > H(X_3|X_1), \tag{59}$$
$$H(X_3|X_0) > H(X_3|X_1X_2), \tag{60}$$

Here, $R_A < R_C$ is equivalent to

$$H(X_2|X_0) < H(X_2|X_1X_3). \tag{61}$$

The rate pair of the point P is given by $(I(X_0X_2;X_1X_3), H(X_2|X_1X_3) - H(X_2|X_0))$. To achieve this rate pair, we set the binning rates as follows:

$$\tilde{R}_1 = H(X_1X_3|X_0X_2) - H(X_3|X_1) + \epsilon, \tag{62}$$
$$R_1 = I(X_0X_2;X_1X_3) - 2\delta(\epsilon) - 2\epsilon, \tag{63}$$
$$\tilde{R}_2 = H(X_2|X_0) + \epsilon, \tag{64}$$
$$R_2 = H(X_2|X_1X_3) - H(X_2|X_0) - 2\delta(\epsilon) - 2\epsilon, \tag{65}$$
$$\tilde{R}_3 = H(X_3|X_1) + \epsilon. \tag{66}$$

It can be verified that the above rates satisfy the Slepian-Wolf conditions (23)-(29). It is also easy to verify that the rate settings (62)-(66) satisfy the sufficient conditions (33), (35) and (41) for secrecy. Furthermore, the condition (37) is satisfied if

$$H(X_1X_3|X_0X_2) < H(X_1X_3|X_2). \tag{67}$$

Otherwise, the Markov chain $X_1X_3 - X_2 - X_0$ holds. To show the secrecy, we derive a new condition to replace (37) to guarantee that $I(g,\psi;f,l|C)$ is asymptotically small as follows.

$$\begin{aligned} I(g,\psi;f,l|\mathcal{C}) &\le I(g,\psi;X_1^n,X_3^n|\mathcal{C}) \\ &= I(X_2^n;X_1^n,X_3^n|\mathcal{C}) - I(X_2^n;X_1^n,X_3^n|g,\psi,\mathcal{C}) \\ &= H(X_2^n|\mathcal{C}) - H(X_2^n|X_1^n,X_3^n,\mathcal{C}) - H(X_2^n|g,\psi,\mathcal{C}) + H(X_2^n|X_1^n,X_3^n,g,\psi,\mathcal{C}) \\ &\le n[\tilde{R}_2 + R_2 - H(X_2|X_1X_3)] + H(X_2^n|X_1^n,X_3^n,g,\psi,\mathcal{C}). \end{aligned}$$

It can be shown that if

$$\tilde{R}_2 + R_2 \le H(X_2|X_1X_3) - 2\delta(\epsilon), \tag{68}$$

then

$$\limsup_{n\to\infty} \frac{1}{n} H(X_2^n|X_1^n,X_3^n,f,l,\mathcal{C}) < H(X_2|X_1X_3) - \tilde{R}_2 - R_2 + \delta(\epsilon). \tag{69}$$

Thus, $I(g,\psi;f,l|C) < \delta(\epsilon)$. It is clear that the rates given in (62)-(66) satisfy (68) and hence secrecy is guaranteed.

Case 4: $R_A = I(X_0X_2;X_1X_3)$ and $R_C = I(X_0X_3;X_1X_2)$, which imply the following two inequalities:

$$H(X_3|X_0X_2) > H(X_3|X_1), \tag{70}$$
$$H(X_3|X_1X_2) > H(X_3|X_0), \tag{71}$$

Then, we have $H(X_3|X_1) < H(X_3|X_0X_2) \le H(X_3|X_0) < H(X_3|X_1X_2)$, which yields a contradiction. Thus, this case does not exist.

B Proof of Theorem 2

Proof of Converse. First, if we need only to generate K_1, the model reduces to the private key generation problem studied in [5]. The key capacity is shown to be

$$R_1 = \min\{I(X_1;X_0X_2|X_3), I(X_0;X_1X_2|X_3)\},$$

which provides an outer bound (13) on R_1. Similarly, if we dedicated to generate K_2, the key capacity is shown to be $R_2 = \min\{I(X_2; X_0X_1|X_3), I(X_0; X_1X_2|X_3)\}$, which provides an outer bound (14) on R_2. For the sum rate bound, we consider an enhanced model by replacing terminals $\mathcal{X}_1$ and $\mathcal{X}_2$ with a super terminal $\mathcal{X}_s$ which observes both X_1^n and X_2^n. The rate of the private key between $\mathcal{X}_0$ and $\mathcal{X}_s$ concealed from terminal $\mathcal{X}_3$ is upper bounded by $I(X_0; X_1, X_2|X_3)$ as shown in [5], which yields the sum rate bound (15).

Proof of Achievability: The key capacity region is the pentagon O-A-P-Q-B-O as illustrated in Fig. 6, where the coordinates of the points A and B are given by $\min\{I(X_1; X_0X_2|X_3), I(X_0; X_1X_2|X_3)\}$ and $\min\{I(X_2; X_0X_1|X_3), I(X_0; X_1X_2|X_3)\}$, respectively. The corner point A can be achieved by letting $\mathcal{X}_2$ be a dedicated helper to generate K_1 following the omniscience scheme in [5]. The corner point B can be achieved by letting $\mathcal{X}_1$ be a dedicated helper to generate K_2 as shown in [5]. We note that the point P would collapse to the point A if $I(X_0; X_1X_2|X_3) \leq I(X_1; X_0X_2|X_3)$ and the point Q would collapse to the point B if $I(X_0; X_1X_2|X_3) \leq I(X_2; X_0X_1|X_3)$. Thus, it is sufficient to show that the corner points P and Q are achievable whenever they are different from the points A and B, respectively.

We note that since the secrecy requirements on K_1 and K_2 are symmetric, it is sufficient to show that the corner point P is achievable, and then the achievability of the point Q follows by symmetry. Furthermore, we assume that $I(X_1; X_0X_2|X_3) < I(X_0; X_1X_2|X_3)$, which implies

$$H(X_2|X_0X_3) < H(X_2|X_1X_3), \tag{72}$$

because otherwise the point P would collapse to the point A and has been justified to be achievable.

The rate pair at the point P is given by $(I(X_1; X_0X_2|X_3), H(X_2|X_1X_3) - H(X_2|X_0X_3))$. The idea to achieve the point P follows the same achievable strategy as in Appendix A. The steps of codebook generation, encoding and transmission, decoding and key generation are the same as the corresponding steps in Appendix A, and are omitted here. In particular, Slepian-Wolf conditions (23)-(29) also guarantee the correct key establishment here.

The secrecy requirement (5) here is different from that for symmetric key generation with a trusted helper. Hence, we next develop the sufficient conditions that guarantee (5) and then choose the binning rates to satisfy these sufficient conditions.

Secrecy: We evaluate the key leakage rates averaged over the random codebook ensemble. Let $f := f(X_1^n)$, $g := g(X_2^n)$ and $l := l(X_3^n)$. Then it is clear that $\mathbf{F} = \{f, g, l\}$. We further let $\phi := \phi(X_1^n)$ and $\psi := \psi(X_2^n)$. Hence, $K_1 = \phi$ and $K_2 = \psi$. We first derive

$$\begin{aligned}
&I(K_1, K_2; X_3^n, \mathbf{F}|\mathcal{C}) \\
&\quad = I(\phi, \psi; f, g, l, X_3^n|\mathcal{C}) \\
&\quad = I(\phi; f, g, X_3^n|\mathcal{C}) + I(\psi; f, g, X_3^n|\phi, \mathcal{C}) \\
&\quad \leq I(\phi; f|\mathcal{C}) + I(\phi, f; g, X_3^n|\mathcal{C}) + I(\psi; \phi, f, g, X_3^n|\mathcal{C}) \\
&\quad \leq I(\phi; f|\mathcal{C}) + I(\phi, f; g, X_3^n|\mathcal{C}) + I(\psi; g|\mathcal{C}) + I(\psi, g; \phi, f, X_3^n|\mathcal{C}) \\
&\quad \leq I(\phi; f|\mathcal{C}) + I(\phi, f; X_3^n|\mathcal{C}) + I(\phi, f; g|X_3^n, \mathcal{C}) \\
&\qquad + I(\psi; g|\mathcal{C}) + I(\psi, g; X_3^n|\mathcal{C}) + I(\psi, g; \phi, f|X_3^n, \mathcal{C}).
\end{aligned} \tag{73}$$

We consider each of the six terms in (73). Similarly to the techniques used in [10, Appendix A], it can be shown that if

$$\tilde{R}_1 + R_1 < H(X_1) - 2\delta(\epsilon), \tag{74}$$

then

$$\frac{1}{n} I(\phi; f|\mathcal{C}) < \delta(\epsilon); \tag{75}$$

and if

$$\tilde{R}_2 + R_2 < H(X_2) - 2\delta(\epsilon), \tag{76}$$

then

$$\frac{1}{n} I(\psi; g|\mathcal{C}) < \delta(\epsilon). \tag{77}$$

In order to bound the second term in (73), we have the following derivation:

$$\begin{aligned}
&I(\phi, f; X_3^n|\mathcal{C}) \\
&= I(X_1^n; X_3^n|\mathcal{C}) - I(X_1^n; X_3^n|\phi, f, \mathcal{C}) \\
&= H(X_1^n|\mathcal{C}) - H(X_1^n|X_3^n, \mathcal{C}) - H(X_1^n|\phi, f, \mathcal{C}) + H(X_1^n|\phi, f, X_3^n, \mathcal{C}) \\
&\leq n[\tilde{R}_1 + R_1 - H(X_1|X_3)] + H(X_1^n|\phi, f, X_3^n, \mathcal{C}).
\end{aligned}$$

It can be shown as in [10, Appendix A] that if

$$\tilde{R}_1 + R_1 \leq H(X_1|X_3) - 2\delta(\epsilon), \tag{78}$$

then

$$\limsup_{n\to\infty} \frac{1}{n} H(X_1^n|\phi, f, X_3^n, \mathcal{C}) < H(X_1|X_3) - \tilde{R}_1 - R_1 + \delta(\epsilon). \tag{79}$$

Consequently,

$$\frac{1}{n} I(\phi, f; X_3^n|\mathcal{C}) < \delta(\epsilon). \tag{80}$$

Similarly, it can be shown that if

$$\tilde{R}_2 + R_2 \leq H(X_2|X_3) - 2\delta(\epsilon), \tag{81}$$

then

$$\frac{1}{n} I(\psi, g; X_3^n|\mathcal{C}) < \delta(\epsilon). \tag{82}$$

By noting that $I(\phi, f; g|X_3^n, \mathcal{C}) \leq I(\psi, g; \phi, f|X_3^n, \mathcal{C})$, it is sufficient to bound the latter term:

$$\begin{aligned}
&I(\psi, g; \phi, f|X_3^n, \mathcal{C}) \\
&\leq I(\psi, g; X_1^n|X_3^n, \mathcal{C}) \\
&= I(X_2^n; X_1^n|X_3^n, \mathcal{C}) - I(X_2^n; X_1^n|g, \psi, X_3^n, \mathcal{C}) \\
&= H(X_2^n|X_3^n, \mathcal{C}) - H(X_2^n|X_1^n, X_3^n, \mathcal{C}) - H(X_2^n|g, \psi, X_3^n, \mathcal{C}) + H(X_2^n|g, \psi, X_1^n, X_3^n, \mathcal{C}) \\
&\leq n[\tilde{R}_2 + R_2 - H(X_2|X_1X_3)] + H(X_2^n|g, \psi, X_1^n, X_3^n, \mathcal{C}).
\end{aligned} \tag{83}$$

It can be shown as in [10, Appendix A] that if

$$\tilde{R}_2 + R_2 < H(X_2|X_1X_3) - 2\delta(\epsilon), \tag{84}$$

then

$$\limsup_{n\to\infty} \frac{1}{n} H(X_2^n|g, \psi, X_1^n, X_3^n, \mathcal{C}) < H(X_2|X_1X_3) - \tilde{R}_2 - R_2 + \delta(\epsilon).$$

Consequently,

$$\frac{1}{n} I(f, \phi; g, \psi|X_3^n, \mathcal{C}) < \delta(\epsilon). \tag{85}$$

Thus, (78) and (84) are sufficient conditions that guarantee the secrecy requirement (5).

The rate pair at point P is given by $(I(X_1; X_0X_2|X_3), H(X_2|X_1X_3) - H(X_2|X_0X_3))$. To achieve this rate pair, we set the binning rates as follows:

$$\tilde{R}_1 = H(X_1|X_0X_2X_3) + \epsilon, \tag{86}$$
$$R_1 = I(X_1; X_0X_2|X_3) - 2\delta(\epsilon) - 2\epsilon, \tag{87}$$
$$\tilde{R}_2 = H(X_2|X_0X_3) + \epsilon, \tag{88}$$
$$R_2 = H(X_2|X_1X_3) - H(X_2|X_0X_3) - 2\delta(\epsilon) - 2\epsilon, \tag{89}$$
$$\tilde{R}_3 = H(X_3|X_0) + \epsilon. \tag{90}$$

It is easy to verify that the above rates satisfy the Slepian-Wolf conditions (23)-(29), and the sufficient conditions (78) and (84) for secrecy.

C Proof of Theorem 3

Proof of Converse. First, if we need only to generate K_1, the model reduces to the secret key generation problem studied in [5]. The key capacity is shown to be $\min\{R_A, R_C\}$ which provides an upper bound on R_1 as given in (17). Next, if we dedicate to generate K_2, the model reduces to the private key generation problem over multiple terminals also studied in [5] as the private key model. The key capacity is shown to be R'_B which serves as an upper bound (18) on R_2. For the sum rate bound, we consider an enhanced model, which replaces terminals $\mathcal{X}_1$ and $\mathcal{X}_2$ with a super terminal $\mathcal{X}_s$ that observes both X_1^n and X_2^n. Then, the secret key rate between $\mathcal{X}_0$ and $\mathcal{X}_s$ is upper bounded by R_C as given in [5] for the secret key model, which yields the sum rate bound (19).

Proof of Achievability. The key capacity region is plotted in Fig. 7 as the pentagon O-A-P-Q-B-O, where the coordinates of the points A and B are $(\min\{R_A, R_C\}, 0)$ and $(0, R'_B)$, respectively. The corner point A is achieved by letting $\mathcal{X}_2$ be a dedicated helper to generate K_1 following [5]. The corner point B is achieved by letting $\mathcal{X}_1$ be a dedicated helper to generate K_2 following [5]. We note that the point P may collapse to the point A if $R_C \le R_A$, and the point Q may collapse to the point B if $R_C \le R_B$. It is thus sufficient to show the

achievability of the points P and Q whenever they are different from the points A and B, respectively. Then the entire pentagon can be achieved by time sharing.

For the point P, it can be observed that its rate coordinates are exactly the same as those in Section 3.1, and can be shown to be achievable by the same scheme designed in Appendix A for symmetric key generation with a trusted helper. The idea to achieve the point Q follows the same achievable strategy as in Appendix A. Hence, the steps of codebook generation, encoding and transmission, decoding and key generation are the same as the corresponding steps in Appendix A, and are omitted here. In particular, Slepian-Wolf conditions (23)-(29) guarantee the correct key establishment.

The secrecy requirement (6) here is different. Hence, we next derive sufficient conditions for achieving secrecy requirement (6). Then these sufficient conditions need to be verified for the point Q in all cases.

Secrecy: We evaluate the key leakage rates averaged over the random codebook ensemble. Let $f := f(X_1^n)$, $g := g(X_2^n)$ and $l := l(X_3^n)$. Then it is clear that $\mathbf{F} = \{f, g, l\}$. We further let $\phi := \phi(X_1^n)$ and $\psi := \psi(X_2^n)$. Hence, $K_1 = \phi$ and $K_2 = \psi$. We first derive

$$\begin{aligned} I(K_1; \mathbf{F}|\mathcal{C}) &= I(\phi; f, g, l|\mathcal{C}) \\ &= I(\phi; f|\mathcal{C}) + I(\phi; g, l|f, \mathcal{C}) \\ &\leq I(\phi; f|\mathcal{C}) + I(\phi, f; g, l|\mathcal{C}), \end{aligned} \tag{91}$$

and

$$\begin{aligned} I(K_2; \mathbf{F}, X_1^n|\mathcal{C}) &= I(\psi; f, g, l, X_1^n|\mathcal{C}) \\ &= I(\psi; g, l, X_1^n|\mathcal{C}) \\ &= I(\psi; g|\mathcal{C}) + I(\psi; l, X_1^n|g, \mathcal{C}) \\ &\leq I(\psi; g|\mathcal{C}) + I(\psi, g; l|\mathcal{C}) + I(\psi, g, l; X_1^n|\mathcal{C}). \end{aligned} \tag{92}$$

It can be shown as in [10, Appendix A] that if

$$\tilde{R}_1 + R_1 < H(X_1) - 2\delta(\epsilon), \tag{93}$$

then

$$\frac{1}{n} I(\phi; f|\mathcal{C}) < \delta(\epsilon); \tag{94}$$

and if

$$\tilde{R}_2 + R_2 < H(X_2) - 2\delta(\epsilon), \tag{95}$$

then

$$\frac{1}{n} I(\psi; g|\mathcal{C}) < \delta(\epsilon). \tag{96}$$

Furthermore,

$$\begin{aligned}
&I(g,\psi;l|\mathcal{C})\\
&\quad\le I(g,\psi,X_2^n;l|\mathcal{C})\\
&\quad= I(X_2^n;l,X_3^n|\mathcal{C}) - I(X_2^n;X_3^n|l,\mathcal{C})\\
&\quad= I(X_2^n;X_3^n|\mathcal{C}) - I(X_2^n;X_3^n|l,\mathcal{C})\\
&\quad= H(X_3^n|\mathcal{C}) - H(X_3^n|X_2^n,\mathcal{C}) - H(X_3^n|l,\mathcal{C}) + H(X_3^n|X_2^n,l,\mathcal{C})\\
&\quad\le H(X_3^n|\mathcal{C}) - H(X_3^n|X_2^n,\mathcal{C}) - [H(X_3^n|\mathcal{C}) - n\tilde{R}_3] + H(X_3^n|X_2^n,l,\mathcal{C})\\
&\quad= n(\tilde{R}_3 - H(X_3|X_2)) + H(X_3^n|X_2^n,l,\mathcal{C})
\end{aligned}$$

where we used the fact that $H(X_3^n|X_2^n,\mathcal{C}) = nH(X_3|X_2)$ due to the assumption of discrete memoryless source.

It can be shown as in [10, Appendix A] that if

$$\tilde{R}_3 \le H(X_3|X_2) - 2\delta(\epsilon), \tag{97}$$

then

$$\limsup_{n\to\infty} \frac{1}{n} H(X_3^n|X_2^n,l,\mathcal{C}) < H(X_3|X_2) - \tilde{R}_3 + \delta(\epsilon). \tag{98}$$

Consequently,

$$\frac{1}{n} I(g,\psi;l|\mathcal{C}) < \delta(\epsilon). \tag{99}$$

We next consider the last term in (92):

$$\begin{aligned}
&I(\psi,g,l;X_1^n|\mathcal{C})\\
&\quad= I(g,\psi,l,X_2^n,X_3^n;X_1^n|\mathcal{C}) - I(X_2^n,X_3^n;X_1^n|g,\psi,l,\mathcal{C})\\
&\quad= I(X_2^n,X_3^n;X_1^n|\mathcal{C}) - I(X_2^n,X_3^n;X_1^n|g,\psi,l,\mathcal{C})\\
&\quad\le n(\tilde{R}_2 + R_2 + \tilde{R}_3 - H(X_2X_3|X_1)) + H(X_2^nX_3^n|X_1^n,g,\psi,l,\mathcal{C}).
\end{aligned}$$

Similarly, it can be shown as in [10, Appendix A] that if

$$\tilde{R}_2 + R_2 + \tilde{R}_3 < H(X_2X_3|X_1) - 2\delta(\epsilon), \tag{100}$$

then

$$\limsup_{n\to\infty} \frac{1}{n} H(X_2^nX_3^n|X_1^n,g,\psi,l,\mathcal{C}) < H(X_2X_3|X_1) - \tilde{R}_2 - R_2 - \tilde{R}_3 + \delta(\epsilon).$$

Consequently,

$$\frac{1}{n} I(g,\psi,l;X_1^n|\mathcal{C}) < \delta(\epsilon). \tag{101}$$

Thus, (93), (95), (97) and (100) are sufficient conditions that guarantee the secrecy requirements in (6).

Uniformity: Uniformity of keys is due to properties of random binning and typicality.

We next show the achievability of the point Q, whose rate coordinates are given by $(R_C - R'_B, R'_B)$. Corresponding to different source distributions, each of R_C and R'_B can take one of the two mutual information terms given in (19) and (18), respectively. Hence, the coordinates of the point Q can take four forms, i.e., case 1 with $R_C = I(X_0; X_1X_2X_3)$ and $R'_B = I(X_0; X_2X_3|X_1)$; case 2 with $R_C = I(X_0; X_1X_2X_3)$ and $R'_B = I(X_2; X_0X_3|X_1)$; case 3 with $R_C = I(X_0X_3; X_1X_2)$ and $R'_B = I(X_0; X_2X_3|X_1)$; and case 4 with $R_C = I(X_0X_3; X_1X_2)$ and $R'_B = I(X_2; X_0X_3|X_1)$. For each case, it is sufficient to set the rates $\tilde{R}_1$, R_1, $\tilde{R}_2$, R_2 and $\tilde{R}_3$ to satisfy the Slepian-Wolf conditions (23)-(29) for guaranteeing correct key agreement and to satisfy the sufficient conditions (93), (95), (97) and (100) for guaranteeing secrecy.

Case 1: $R_C = I(X_0; X_1X_2X_3)$ and $R'_B = I(X_0; X_2X_3|X_1)$, which imply

$$H(X_3|X_1X_2) < H(X_3|X_0), \tag{102}$$
$$H(X_3|X_1X_2) < H(X_3|X_0X_1). \tag{103}$$

The rate pair at the point Q is given by $(I(X_0; X_1), I(X_0; X_2X_3|X_1))$. To achieve this rate pair, we set the binning rates in the achievable strategy as follows:

$$\tilde{R}_1 = H(X_1|X_0) + \epsilon, \tag{104}$$
$$R_1 = I(X_0; X_1) - 2\delta(\epsilon) - 2\epsilon, \tag{105}$$
$$\tilde{R}_2 = H(X_2X_3|X_0X_1) - H(X_3|X_1X_2) + 2\delta(\epsilon) + 2\epsilon, \tag{106}$$
$$R_2 = I(X_0; X_2X_3|X_1) - 4\delta(\epsilon) - 3\epsilon, \tag{107}$$
$$\tilde{R}_3 = H(X_3|X_1X_2) - 2\delta(\epsilon) - \epsilon. \tag{108}$$

It can be verified that the above rates satisfy the Slepian-Wolf conditions (23)-(29) if $\tilde{R}_3 > H(X_3|X_0X_1X_2)$. Otherwise, the rate pair can be easily achieved without the helper's assistance. It can also be verified that the above rates (104)-(108) satisfy the secrecy conditions (93), (95), (97) and (100).

Case 2: $R_C = I(X_0; X_1X_2X_3)$ and $R'_B = I(X_2; X_0X_3|X_1)$, which imply

$$H(X_3|X_1X_2) < H(X_3|X_0), \tag{109}$$
$$H(X_3|X_0X_1) < H(X_3|X_1X_2). \tag{110}$$

The rate pair at the point Q is given by $(I(X_0; X_1X_3) - I(X_2; X_3|X_1), I(X_2; X_0X_3|X_1))$. To achieve this rate pair, we set the binning rates in the achievable strategy as follows:

$$\tilde{R}_1 = H(X_1X_3|X_0) - H(X_3|X_1X_2) + \epsilon, \tag{111}$$
$$R_1 = I(X_0; X_1X_3) - I(X_2; X_3|X_1) - 2\delta(\epsilon) - 2\epsilon, \tag{112}$$
$$\tilde{R}_2 = H(X_2|X_0X_1X_3) + \epsilon, \tag{113}$$
$$R_2 = I(X_2; X_0X_3|X_1) - 4\delta(\epsilon) - 3\epsilon, \tag{114}$$
$$\tilde{R}_3 = H(X_3|X_1X_2) + \epsilon. \tag{115}$$

It can be verified that the above rates satisfy the Slepian-Wolf conditions (23)-(29). In particular, $\tilde{R}_1 > H(X_1|X_0X_2X_3)$ and $\tilde{R}_1 + \tilde{R}_2 > H(X_1X_2|X_0X_3)$ hold due to (109), and $\tilde{R}_2 + \tilde{R}_3 > H(X_2X_3|X_0X_1)$ holds due to (110).

It can also be verified that the above rates (111)-(115) satisfy the secrecy conditions (93), (95), (97) and (100). In particular, (97) holds under the following assumption

$$H(X_2|X_0X_1X_3) < H(X_2|X_1) - 2\delta(\epsilon) - 2\epsilon. \tag{116}$$

If the above assumption does not hold, then the Markov chain $X_2 - X_1 - X_0X_3$ holds, which implies $R_2 = I(X_2; X_0X_3|X_1) = 0$. The point Q coincides with the point B, which has been justified to be achievable.

Case 3: $R_C = I(X_0X_3; X_1X_2)$ and $R'_B = I(X_0; X_2X_3|X_1)$, which imply

$$H(X_3|X_0) \leq H(X_3|X_1X_2), \tag{117}$$
$$H(X_3|X_1X_2) \leq H(X_3|X_0X_1). \tag{118}$$

Then, we have $H(X_3|X_0) \leq H(X_3|X_1X_2) \leq H(X_3|X_0X_1)$, which yields contradiction. Thus, this case does not exist.

Case 4: $R_C = I(X_0X_3; X_1X_2)$ and $R'_B = I(X_2; X_0X_3|X_1)$, which imply

$$H(X_3|X_0) \leq H(X_3|X_1X_2), \tag{119}$$
$$H(X_3|X_0X_1) < H(X_3|X_1X_2). \tag{120}$$

The rate pair at the point Q is given by $(I(X_1; X_0X_3), I(X_2; X_0X_3|X_1))$. To achieve this rate pair, we set the binning rates in the achievable strategy as follows:

$$\tilde{R}_1 = H(X_1|X_0X_3) + \epsilon, \tag{121}$$
$$R_1 = I(X_1; X_0X_3) - 2\delta(\epsilon) - 2\epsilon, \tag{122}$$
$$\tilde{R}_2 = H(X_2|X_0X_1X_3) + \epsilon, \tag{123}$$
$$R_2 = I(X_2; X_0X_3|X_1) - 2\delta(\epsilon) - 2\epsilon, \tag{124}$$
$$\tilde{R}_3 = H(X_3|X_0) + \epsilon. \tag{125}$$

It is easy to verify that the above rates satisfy the Slepian-Wolf conditions (23)-(29) and the secrecy conditions (93), (95), (97) and (100).

D Proof of Theorem 4

Proof of Converse. First, if we need only to generate K_1, the model reduces to the private key generation problem over multiple terminals studied in [5]. The key capacity is shown to be $\min\{I(X_1; X_0X_2|X_3), I(X_0; X_1X_2|X_3)\}$, which serves as an upper bound (20) on R_1. Next, if we dedicate to generate K_2, it also reduces to the private key model studied in [5].

The key capacity is shown to be $I(X_2; X_0|X_1X_3)$, which serves an upper bound (21) on R_2. For the sum rate bound, we consider an enhanced model which replaces terminals $\mathcal{X}_1$ and $\mathcal{X}_2$ with a super terminal $\mathcal{X}_s$ that observes both X_1^n and X_2^n. Then, the private key rate between $\mathcal{X}_0$ and $\mathcal{X}_s$ is upper bounded by $I(X_0; X_1, X_2|X_3)$ due to the private key model in [5], which yields the sum rate bound (22).

Proof of Achievability. The key capacity region is illustrated in Fig. 8 as the pentagon O-A-P-Q-B-O, where the coordinates of the points A and B are $(\min\{I(X_0; X_1X_2|X_3), I(X_1; X_0X_2|X_3)\}, 0)$ and $(0, I(X_0; X_2|X_1X_3))$, respectively. The corner point A is achieved by letting $\mathcal{X}_2$ be a dedicated helper to generate K_1 following [5]. The corner point B is achieved by letting $\mathcal{X}_1$ be a dedicated helper to generate K_2 following [5]. We note that the point P would collapse to the point A if $I(X_0; X_1X_2|X_3) \leq I(X_1; X_0X_2|X_3)$, and the point Q would collapse to the point B if $I(X_0; X_1X_2|X_3) \leq I(X_2; X_0|X_1X_3)$. It is thus sufficient to show the achievability of the points P and Q whenever they are different from the points A and B, respectively. Then the entire pentagon can be achieved by time sharing.

The idea to achieve the points P and Q follows the same achievable strategy described in Appendix A. The steps of codebook generation, encoding and transmission, decoding and key generation are the same as those in Appendix A, and are omitted here. In particular, Slepian-Wolf conditions (23)-(29) guarantee the correct key establishment. Since the secrecy requirements given in (7) are different, we next develop the sufficient conditions that guarantee (7), and then choose the binning rates to satisfy these sufficient conditions.

Secrecy: We evaluate the key leakage rates averaged over the random codebook ensemble. Let $f := f(X_1^n)$, $g := g(X_2^n)$ and $l := l(X_3^n)$. Then it is clear that $\mathbf{F} = \{f, g, l\}$. We further let $\phi := \phi(X_1^n)$ and $\psi := \psi(X_2^n)$. Hence, $K_1 = \phi$ and $K_2 = \psi$. We first derive

$$\begin{aligned} I(K_1; X_3^n, \mathbf{F}|\mathcal{C}) &= I(\phi; f, g, l, X_3^n|\mathcal{C}) \\ &= I(\phi; f|\mathcal{C}) + I(\phi; g, X_3^n|f, \mathcal{C}) \\ &\leq I(\phi; f|\mathcal{C}) + I(\phi, f; g, X_3^n|\mathcal{C}) \\ &\leq I(\phi; f|\mathcal{C}) + I(\phi, f; X_3^n|\mathcal{C}) + I(\phi, f; g|X_3^n, \mathcal{C}), \end{aligned} \tag{126}$$

and

$$\begin{aligned} I(K_2; X_1^n, X_3^n, \mathbf{F}|\mathcal{C}) &= I(\psi; f, g, l, X_1^n, X_3^n|\mathcal{C}) \\ &= I(\psi; g, X_1^n, X_3^n|\mathcal{C}) \\ &\leq I(\psi; g|\mathcal{C}) + I(\psi, g; X_1^n, X_3^n|\mathcal{C}). \end{aligned} \tag{127}$$

We next bound each of the five terms in (126) and (127). It can be shown as in [10, Appendix A] that if

$$\tilde{R}_1 + R_1 < H(X_1) - 2\delta(\epsilon), \tag{128}$$

then

$$\frac{1}{n} I(\phi; f|\mathcal{C}) < \delta(\epsilon); \tag{129}$$

and if

$$\tilde{R}_2 + R_2 < H(X_2) - 2\delta(\epsilon), \tag{130}$$

then

$$\frac{1}{n}I(\psi; g|\mathcal{C}) < \delta(\epsilon). \tag{131}$$

To bound the second term in (126), we have

$$\begin{aligned}
&I(\phi, f; X_3^n|\mathcal{C}) \\
&= I(X_1^n; X_3^n|\mathcal{C}) - I(X_1^n; X_3^n|\phi, f, \mathcal{C}) \\
&= H(X_1^n|\mathcal{C}) - H(X_1^n|X_3^n, \mathcal{C}) - H(X_1^n|\phi, f, \mathcal{C}) + H(X_1^n|\phi, f, X_3^n, \mathcal{C}) \\
&\leq n[\tilde{R}_1 + R_1 - H(X_1|X_3)] + H(X_1^n|\phi, f, X_3^n, \mathcal{C}).
\end{aligned}$$

It can be shown as in [10, Appendix A] that if

$$\tilde{R}_1 + R_1 \leq H(X_1|X_3) - 2\delta(\epsilon), \tag{132}$$

then

$$\limsup_{n\to\infty} \frac{1}{n}H(X_1^n|\phi, X_3^n, \mathcal{C}) < H(X_1|X_3) - R_1 + \delta(\epsilon), \tag{133}$$

and consequently,

$$\frac{1}{n}I(\phi, f; X_3^n|\mathcal{C}) < \delta(\epsilon). \tag{134}$$

We observe that $I(\phi, f; g|X_3^n, \mathcal{C}) \leq I(X_1^n X_3^n; \psi, g|\mathcal{C})$ by simple calculation. Thus, it is sufficient to only bound the last term in (127):

$$\begin{aligned}
&I(\psi, g; X_1^n, X_3^n|\mathcal{C}) \\
&\quad = I(X_2^n; X_1^n, X_3^n|\mathcal{C}) - I(X_2^n; X_1^n, X_3^n|\psi, g, \mathcal{C}) \\
&\quad = H(X_2^n|\mathcal{C}) - H(X_2^n|X_1^n, X_3^n, \mathcal{C}) - H(X_2^n|\psi, g, \mathcal{C}) \\
&\qquad + H(X_2^n|g, X_1^n, X_3^n, \mathcal{C}) + H(X_2^n|\psi, g, X_1^n, X_3^n, \mathcal{C}) \\
&\quad \leq n[\tilde{R}_2 + R_2 - H(X_2|X_1X_3)] + H(X_2^n|\psi, g, X_1^n, X_3^n, \mathcal{C}).
\end{aligned} \tag{135}$$

Similarly, it can be shown as in [10, Appendix A] that if

$$\tilde{R}_2 + R_2 \leq H(X_2|X_1X_3) - 2\delta(\epsilon), \tag{136}$$

then

$$\limsup_{n\to\infty} \frac{1}{n}H(X_2^n|g, X_1^n, X_3^n, \mathcal{C}) < H(X_2|X_1X_3) - \tilde{R}_2 + \delta(\epsilon), \tag{137}$$

and consequently,

$$\frac{1}{n}I(\psi, g; X_1^n, X_3^n|\mathcal{C}) \leq \delta(\epsilon). \tag{138}$$

Hence, (132) and (136) are the sufficient conditions that guarantee the secrecy requirements given in (7).

We next show the achievability of the points P and Q. It can be shown that the binning rates chosen to achieve the point P in Appendix B for symmetric key generation with an

untrusted helper is applicable here to achieve the point P. To achieve the point Q, whose coordinates are given by $(I(X_0; X_1|X_3), I(X_0; X_2|X_1X_3))$, we set the binning rates in the achievable strategy as follows:

$$\tilde{R}_1 = H(X_1|X_0X_3) + \epsilon, \tag{139}$$
$$R_1 = I(X_0; X_1|X_3) - 2\delta(\epsilon) - 2\epsilon, \tag{140}$$
$$\tilde{R}_2 = H(X_2|X_0X_1X_3) + \epsilon, \tag{141}$$
$$R_2 = I(X_0; X_2|X_1X_3) - 2\delta(\epsilon) - 2\epsilon, \tag{142}$$
$$\tilde{R}_3 = H(X_3|X_0) + \epsilon. \tag{143}$$

It can be verified that the above rates satisfy the Slepian-Wolf conditions (23)-(29) and the secrecy conditions (132) and (136).

References

[1] U. M. Maurer. Secrete key agreement by public discussion based on common information. *IEEE Trans. Inform. Theory*, 39(5):733–742, May 1993.

[2] R. Ahlswede and I. Csiszár. Common randomness in information theory and cryptography-Part I: Secret sharing. *IEEE Trans. Inform. Theory*, 39(4):1121–1132, July 1993.

[3] R. Ahlswede and I. Csiszár. Common randomness in information theory and cryptography-Part II: CR capacity. *IEEE Trans. Inform. Theory*, 44(1):225–240, January 1998.

[4] I. Csiszár and P. Narayan. Common randomness and secret key generation with a helper. *IEEE Trans. Inform. Theory*, 46(2):344–366, March 2000.

[5] I. Csiszár and P. Narayan. Secrecy capacities for mulitple terminals. *IEEE Trans. Inform. Theory*, 50(12):3047–3061, December 2004.

[6] A. A. Gohari and V. Anantharam. Information-theoretic key agreement of multiple terminals-Part I. *IEEE Trans. Inform. Theory*, 56(8):3973–3996, Aug 2010.

[7] C. Chan and L. Zheng. Mutual dependence for secret key agreement. In *Proc. Conf. on Information Sciences and Systems (CISS)*, Princeton University, NJ, USA, March 2010.

[8] C. Ye and P. Narayan. The secret key-private key capacity region for three terminals. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Adelaide, Australia, September 2005.

[9] H. Zhang, L. Lai, Y. Liang, and H. Wang. The secret key-private key generation over three terminals: Capacity region. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Honolulu, HI, USA, June 2014.

[10] H. Zhang, L. Lai, Y. Liang, and H. Wang. The capacity region of the source-type model for secret key and private key generation. *IEEE Trans. Inform. Theory*, 60(10):6389–6398, October 2014.

[11] L. Lai and L. Huie. Simultaneously generating multiple keys in many to one networks. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Istanbul, Turkey, July 2013.

[12] H. Zhang, Y. Liang, and L. Lai. Key capacity region for a cellular source model. In *Proc. IEEE Information Theory Workshop (ITW)*, Hobart, Tasmania, Australia, November 2014.

[13] U. M. Maurer and S. Wolf. From weak to strong information-theoretic key agreement. In *Proc. IEEE Int. Symp. Information Theory (ISIT)*, page 18, Sorrento, Italy, June 2000.

[14] S. Wantanabe and Y. Oohama. Secret key agreement from vector gaussian sources by rate limited public communication. *IEEE Trans. Inform. Forensics and Security*, 6(3):541–550, September 2011.

[15] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Inform. Theory*, IT-19:471–480, 1973.