Secure and Private Distributed Source Coding with Private Keys and Decoder Side Information

Onur Günlü, *Member, IEEE,* Rafael F. Schaefer, *Senior Member, IEEE,* Holger Boche, *Fellow, IEEE,* and H. Vincent Poor, *Life Fellow, IEEE*

***Abstract*—The distributed source coding problem is extended by positing that noisy measurements of a remote source are the correlated random variables that should be reconstructed at another terminal. We consider a secure and private distributed lossy source coding problem with two encoders and one decoder such that (i) all terminals noncausally observe a noisy measurement of the remote source; (ii) a private key is available to each legitimate encoder and all private keys are available to the decoder; (iii) rate-limited noiseless communication links are available between each encoder and the decoder; (iv) the amount of information leakage to an eavesdropper about the correlated random variables is defined as *secrecy* leakage, and *privacy* leakage is measured with respect to the remote source; and (v) two passive attack scenarios are considered, where a strong eavesdropper can access both communication links and a weak eavesdropper can choose only one of the links to access. Inner and outer bounds on the rate regions defined under secrecy, privacy, communication, and distortion constraints are derived for both passive attack scenarios. When one or both sources should be reconstructed reliably, the rate region bounds are simplified.**

***Index Terms*—Secure and private distributed source coding, remote source, rate-limited public communication, weak eavesdropper, passive attack.**

I. Introduction

A fundamental problem with numerous recent applications is the compression of correlated random sequences observed by multiple terminals such that another terminal, called a decoder, can reconstruct these sequences by using the compressed messages, i.e., the distributed source coding problem [1]. Smart grids with sensors that measure correlated sequences, such as voltage levels, that should be transmitted to a distant node is an example application of this problem. Similarly, function computation problems, where a fusion center observes compressed messages to compute a function of the correlated sequences, are closely related to distributed source coding problems [2], [3], and security constraints are generally imposed on both problems since the communication links can be public [4]. If all transmitted messages are available to an eavesdropper in the same network, then it is necessary to provide the decoder an advantage over the eavesdropper to enable secure source coding or function computation. Decoder side information that is correlated with the random sequences to be reconstructed provides such an advantage [5]–[7]. Limiting the access of the eavesdropper to a strict subset of all transmitted messages also enables secure distributed source coding, considered in [8]–[11]. Similarly, a private key shared only by the legitimate terminals also helps to hide the source sequences [12], [13].

The posit that a ground truth is the reason for correlations between the random sequences used in a distributed source coding problem makes the models used more realistic. For secure function computation [14], [15] and secret-key agreement [16], [17] problems, which can be considered as instances of the source coding with side information problem [18, Section IV- B], the correlations are posited in [3], [19] to stem from a remote source such that its noisy versions are these dependent sequences. We similarly assume that there is a remote source whose noisy measurements are used in the source coding problems discussed below, which is similar to the models in [20, Fig. 9] and [21, pp. 78]. Furthermore, in the chief executive officer (CEO) problem [22] the aim is to reconstruct a remote (or hidden) source at a decoder by using the messages transmitted by multiple encoders that observe noisy measurements of the same remote source. Our problem is different from the CEO problem, because in our model the decoder aims to recover encoder observations rather than the remote source. Therefore, we define the amount of information leakage to an eavesdropper about the encoder observations as the *secrecy leakage*, whereas the leakage about the remote source is the *privacy leakage* since the remote source is common for all encoder observations [23]–[25]. We consider two encoders, which requires different analysis from [26] with a single encoder, that observe different noisy measurements of the same remote source, and we impose joint secrecy and joint privacy constraints on the distributed source coding problems considered; see [27]–[30] for such joint constraints imposed for secret key agreement problems.

O. Günlü and R. F. Schaefer were supported in part by the German Federal Ministry of Education and Research (BMBF) within the national initiative for Post-Shannon Communication (NewCom) under the Grant 16KIS1004. H. Boche was supported in part by the BMBF within the national initiative for 6G Communication Systems through the Research Hub 6G-life under the Grant 16KISK002 and within the national initiative for Information Theory for Post Quantum Crypto "Quantum Token Theory and Applications - QTOK" under the Grant 16KISQ037K, which has received additional funding from the German Research Foundation (DFG) within Germany's Excellence Strategy EXC-2092 CASA-390781972. H. V. Poor was supported in part by the U.S. National Science Foundation (NSF) under the Grant CCF-1908308.

O. Günlü and R. F. Schaefer are with the Chair of Communications Engineering and Security, University of Siegen, 57076 Siegen, Germany (Email: {onur.guenlue, rafael.schaefer}@uni-siegen.de).

H. Boche is with the Chair of Theoretical Information Technology, Technical University of Munich, 80333 Munich, Germany; CASA: Cyber Security in the Age of Large-Scale Adversaries Exzellenzcluster, Ruhr-Universität Bochum, 44780 Bochum, Germany; and BMBF Research Hub 6G-Life, Technical University of Munich, 80333 Munich, Germany (Email: boche@tum.de).

H. V. Poor is with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544-1019, U.S.A. (Email: poor@princeton.edu).

A. Summary of Contributions

Consider a distributed source coding problem with two encoders, one decoder, and one eavesdropper, in which private keys are available to the legitimate terminals and all terminals observe a noisy measurement of a remote source. We impose distortion and joint secrecy constraints on the reconstructed source sequences, a joint privacy constraint on the remote source, and communication rate constraints on the communication links between encoders and the decoder to establish inner and outer bounds on the resulting rate regions. A summary of our main contributions is as follows.

- We derive inner and outer bounds on the rate region for the secure and private distributed lossy source coding problem with two encoders and private keys, in which an eavesdropper can access all messages transmitted by encoders. The bounds differ only in the Markov chain conditions imposed on the auxiliary random variables. The measurement channel model we consider corresponds to a physically degraded broadcast channel (BC), which is an extension of previous source models that include the classic model with noiseless encoder measurements, i.e., a semi-deterministic BC. The terms in the rate region bounds are shown to be different for low, middle, and high private key-rate regimes. Furthermore, we show that a time-sharing random variable enlarges the rate regions for the low private key-rate regime.
- We next consider that the eavesdropper is weak and can choose only one of the communication links to access the transmitted message, which changes the analyses and bounds for the joint secrecy- and joint privacy-leakage rates at low and middle private key-rate regimes.
- All inner and outer bounds for the distributed lossy source coding problems are shown to be straightforwardly extended to the corresponding lossless settings. Similarly, our bounds recover previous results in the literature, including the secure and private source coding results with a single encoder.

B. Organization

In Section II, we introduce two secure and private distributed lossy source coding problems with a remote source. In Section III, we provide inner and outer bounds, for which different Markov chain conditions are imposed, on the rate regions against a strong eavesdropper. In Section IV, we provide inner and outer bounds on the rate regions against a weak eavesdropper and also discuss how to simplify the bounds for the secure and private distributed lossless source coding problems against a strong or weak eavesdropper. In Sections V and VI, we prove the inner and outer bounds on the rate regions against a strong eavesdropper and a weak eavesdropper, respectively. In Section VII, we conclude the paper.

C. Notation

Upper case letters X represent random variables, lower case letters x their realizations, and calligraphic letters $\mathcal{X}$ the set of realizations such that X has probability distribution P_X. A subscript i denotes the position of a variable in a length-n sequence $X^n = X_1, X_2, \ldots, X_i, \ldots, X_n$. $[1:m]$ denotes the set $\{1, 2, \ldots, m\}$ for an integer $m \geq 1$. Define $[a]^- = \min\{a, 0\}$ for $a \in \mathbb{R}$.

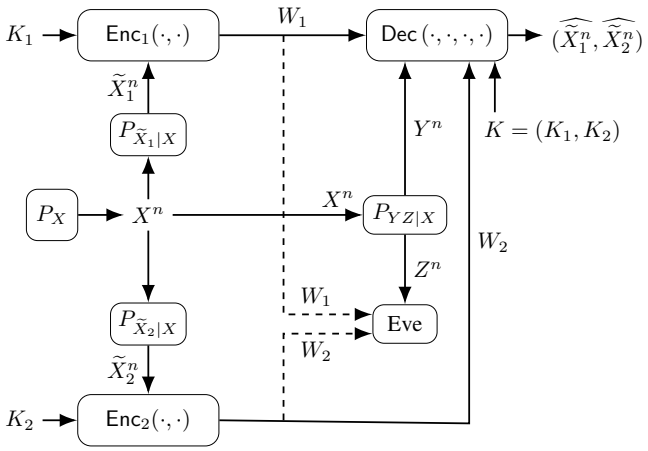

Fig. 1. A distributed source coding problem with two noisy measurements $(\widetilde{X}_1^n, \widetilde{X}_2^n)$ of a remote source X^n, private keys (K_1, K_2), and noncausal decoder side information Y^n under privacy, secrecy, communication, and distortion (or reliability) constraints. We consider the following attack scenarios: a strong Eve observes (W_1, W_2, Z^n), and a weak Eve can choose to observe only one of the indices in addition to the side information Z^n but her choice is not known by the legitimate terminals.

II. System Models

We consider the distributed lossy source coding model with two legitimate encoders, one legitimate decoder, and an eavesdropper (Eve), as depicted in Fig. 1. Two encoders $\mathsf{Enc}_1(\cdot,\cdot)$ and $\mathsf{Enc}_2(\cdot,\cdot)$ observe the noisy measurements $\widetilde{X}_1^n$ and $\widetilde{X}_2^n$ of an independent and identically distributed (i.i.d.) remote source $X^n \sim P_X^n$ through memoryless channels $P_{\widetilde{X}_1|X}$ and $P_{\widetilde{X}_2|X}$ in addition to private keys $K_1 \in [1 : 2^{nR_0}]$ and $K_2 \in [1 : 2^{nR_0}]$ for $R_0 \geq 0$, respectively. Encoder outputs are two indices W_1 and W_2 that are sent over separate links with communication-rate constraints. The decoder $\mathsf{Dec}(\cdot,\cdot,\cdot,\cdot)$ observes both indices, as well as the private keys $K = (K_1, K_2)$ and another noisy measurement Y^n of the same remote source X^n through a memoryless channel $P_{YZ|X}$ in order to reconstruct $\widetilde{X}_1^n$ and $\widetilde{X}_2^n$. The other noisy output Z^n of $P_{YZ|X}$ is observed by Eve in addition to a non-empty subset of the set of indices $\{W_1, W_2\}$, i.e., for a *strong* attack scenario Eve observes (W_1, W_2) and for a *weak* attack scenario Eve can choose W_1 or W_2 but her choice is not known by the legitimate terminals. Suppose K is uniformly distributed, hidden from Eve, and independent of the source output and its noisy measurements.

The source and measurement alphabets are finite sets.

Consider the distributed lossy source coding model illustrated in Fig. 1 such that Eve observes (W_1, W_2), i.e., strong Eve. The corresponding rate region is defined as follows.

Definition 1. A *distributed lossy* tuple $(R_\mathrm{s}, R_\ell, R_{\mathrm{w},1}, R_{\mathrm{w},2}, D_1, D_2) \in \mathbb{R}_{\geq 0}^6$ is achievable against a *strong* Eve, given two corresponding private keys each with

rate $R_0 \geq 0$, if for any $\delta > 0$ there exist $n \geq 1$, two encoders, and one decoder such that

$$\frac{1}{n} I(\widetilde{X}_1^n, \widetilde{X}_2^n; W_1, W_2 | Z^n) \leq R_{\mathrm{s}} + \delta \qquad \text{(secrecy)} \tag{1}$$

$$\frac{1}{n} I(X^n; W_1, W_2 | Z^n) \leq R_\ell + \delta \qquad \text{(privacy)} \tag{2}$$

$$\frac{1}{n} \log \left| \mathcal{W}_j \right| \leq R_{\mathrm{w},j} + \delta \qquad \text{for } j=1,2 \text{ (storages)} \tag{3}$$

$$\mathbb{E}\Big[d\Big(\widetilde{X}_j^n, \widehat{\widetilde{X}_j^n} \Big) \Big] \leq D_j + \delta \qquad \text{for } j=1,2 \text{ (distortions)} \tag{4}$$

where $\widehat{\widetilde{X}_j^n}$ is a function of (W_1, W_2, Y^n, K) for $j = 1, 2$ and $d(\widetilde{x}^n, \widehat{\widetilde{x}^n}) = \frac{1}{n} \sum_{i=1}^n d(\widetilde{x}_i, \widehat{\widetilde{x}}_i)$ is a per-letter bounded distortion metric. The secure and private *distributed lossy* source coding region $\mathcal{R}_{\text{D,strong}}$ is the closure of the set of all achievable distributed lossy tuples against a strong Eve. $\Diamond$

Consider next the distributed lossy source coding model illustrated in Fig. 1 such that Eve observes now either W_1 or W_2, i.e., weak Eve. The corresponding rate region is defined as follows.

Definition 2. A *distributed lossy* tuple $(R_{\mathrm{s}}, R_\ell, R_{\mathrm{w},1}, R_{\mathrm{w},2}, D_1, D_2) \in \mathbb{R}^6_{\geq 0}$ is achievable against a *weak* Eve, given two corresponding private keys each with rate $R_0 \geq 0$, if for any $\delta > 0$ there exist $n \geq 1$, two encoders, and one decoder such that (3), (4), and

$$\frac{1}{n} I(\widetilde{X}_1^n, \widetilde{X}_2^n; W_j | Z^n) \leq R_{\mathrm{s}} + \delta \qquad \forall j = 1, 2 \,(\text{secrecy}) \tag{5}$$

$$\frac{1}{n} I(X^n; W_j | Z^n) \leq R_\ell + \delta \qquad \forall j = 1, 2 \,(\text{privacy}). \tag{6}$$

The secure and private *distributed lossy* source coding region $\mathcal{R}_{\text{D,weak}}$ is the closure of the set of all achievable distributed lossy tuples against a weak Eve. $\Diamond$

III. Secure and Private Distributed Source Coding Against Strong Eavesdropper

We next provide inner and outer bounds on the rate region $\mathcal{R}_{\text{D,strong}}$; see Section V for their proofs.

Denote

$$j' = (3 - j) \tag{7}$$

and

$$R^* = \Big[I(U_1, U_2; Z | V_1, V_2, Q) - I(U_1, U_2; Y | V_1, V_2, Q) \Big]^- \tag{8}$$

$$\overline{R_0} = \max_{j=1,2} \; I(U_j; \widetilde{X}_j | V_1, V_2, Y) \tag{9}$$

$$\overline{\overline{R_0}} = \max_{j=1,2} \Big(I(V_j; \widetilde{X}_j | Y) + I(U_j; \widetilde{X}_j | V_1, V_2, Y) \Big). \tag{10}$$

Lemma 1 (Inner Bound)**.** *The region $\mathcal{R}_{\text{D,strong}}$ includes the union over all $P_{QV_1V_2U_1U_2|\widetilde{X}_1\widetilde{X}_2}$ of the rate tuples $(R_s, R_\ell, R_{w,1}, R_{w,2}, D_1, D_2)$ such that*

$$R_{w,1} \geq I(V_1; \widetilde{X}_1 | V_2, Y) + I(U_1; \widetilde{X}_1 | U_2, V_1, Y) \tag{11}$$

$$R_{w,2} \geq I(V_2; \widetilde{X}_2 | V_1, Y) + I(U_2; \widetilde{X}_2 | U_1, V_2, Y) \tag{12}$$

$$\begin{aligned} R_{w,1} + R_{w,2} \geq\; & I(U_2; \widetilde{X}_2 | U_1, V_2, Y) + I(U_1; \widetilde{X}_1 | V_1, V_2, Y) \\ & + I(V_2; \widetilde{X}_2 | V_1, Y) + I(V_1; \widetilde{X}_1 | Y) \end{aligned} \tag{13}$$

and if $R_0 < \overline{R_0}$, then

$$R_s \geq I(U_1, U_2; \widetilde{X}_1, \widetilde{X}_2 | Z) + R^* - 2R_0 \tag{14}$$

$$R_\ell \geq I(U_1, U_2; X | Z) + R^* - 2R_0 \tag{15}$$

if $\overline{R_0} \leq R_0 < \overline{\overline{R_0}}$, then

$$R_s \geq I(V_1, V_2; \widetilde{X}_1, \widetilde{X}_2 | Z) \tag{16}$$

$$R_\ell \geq I(V_1, V_2; X | Z) \tag{17}$$

and if $R_0 \geq \overline{\overline{R_0}}$, then

$$R_s \geq 0 \tag{18}$$

$$R_\ell \geq 0 \tag{19}$$

and we have

$$\begin{aligned} & P_{QV_1V_2U_1U_2\widetilde{X}_1\widetilde{X}_2XYZ} \\ & = P_{Q|V_1V_2} P_{V_1|U_1} P_{U_1|\widetilde{X}_1} P_{\widetilde{X}_1|X} P_{V_2|U_2} P_{U_2|\widetilde{X}_2} P_{\widetilde{X}_2|X} P_X P_{YZ|X} \end{aligned} \tag{20}$$

such that

$$D_j \geq \mathbb{E}\Big[d(\widetilde{X}_j, \widehat{\widetilde{X}_j}(U_1, U_2, Y)) \Big] \qquad \text{for } j=1,2 \tag{21}$$

for some reconstruction function $\widehat{\widetilde{X}_j}(U_1, U_2, Y)$.

Lemma 2 (Outer Bound)**.** *The region $\mathcal{R}_{\text{D,strong}}$ is included in the union of the rate tuples $(R_s, R_\ell, R_{w,1}, R_{w,2}, D_1, D_2)$ in (13) and*

$$\begin{aligned} R_{w,1} \geq\; & I(V_1; \widetilde{X}_1 | V_2, Y) + I(U_1; \widetilde{X}_1 | V_1, U_2, Y) \\ & - I(V_1; V_2 | \widetilde{X}_1, Y) - I(U_1; U_2 | \widetilde{X}_1, Y, V_1) \end{aligned} \tag{22}$$

$$\begin{aligned} R_{w,2} \geq\; & I(V_2; \widetilde{X}_2 | V_1, Y) + I(U_2; \widetilde{X}_2 | U_1, V_2, Y) \\ & - I(V_1; V_2 | \widetilde{X}_2, Y) - I(U_1; U_2 | \widetilde{X}_2, Y, V_2) \end{aligned} \tag{23}$$

as well as if $R_0 < \overline{R_0}$, then (14) and (15); if $\overline{R_0} \leq R_0 < \overline{\overline{R_0}}$, then (16) and (17); and if $R_0 \geq \overline{\overline{R_0}}$, then (18) and (19), where unions are over all $P_{QV_1V_2U_1U_2|\widetilde{X}_1\widetilde{X}_2}$ such that we have (21) for some reconstruction function $\widehat{\widetilde{X}_j}(U_1, U_2, Y)$ and

$$(Q, V_j) - U_j - \widetilde{X}_j - X - (\widetilde{X}_{j'}, Y, Z) \tag{24}$$

form Markov chains for $j = 1, 2$. One can limit the cardinalities to $|\mathcal{V}_j| \leq |\widetilde{\mathcal{X}}_j| + 6$, $|\mathcal{U}_j| \leq (|\widetilde{\mathcal{X}}_j| + 6)^2$, and $|\mathcal{Q}| \leq 2$.

Secure and private distributed lossy source coding rate region bounds given in Lemmas 1 and 2 do not match in general since the set of joint probability distributions $P_{QV_1V_2U_1U_2\widetilde{X}_1\widetilde{X}_2XYZ}$ that satisfy (20) is not equal to the set that satisfies the Markov chain conditions in (24). We remark that the negative terms in (22) and (23) are zero if one imposes (20), and that the time-sharing random variable Q enlarges the rate region above; see [31], [32] for secure and private function computation rate region bounds that do not match due to similar reasons and that are also enlarged and convexified by using a time-sharing random variable.

IV. Secure and Private Distributed Source Coding Against Weak Eavesdropper

Now, we provide inner and outer bounds on the rate region $\mathcal{R}_{\text{D,weak}}$; see Section VI for proof sketches.

Denote

$$R^{**} = \Big[I(U_j; Z|V_j, Q) - I(U_j; Y, V_{j'}|V_j, Q)\Big]^{-}. \tag{25}$$

Lemma 3 (Inner Bound)**.** *The region $\mathcal{R}_{\text{D,weak}}$ includes the union over all $P_{QV_1V_2U_1U_2|\widetilde{X}_1\widetilde{X}_2}$ of the rate tuples $(R_s, R_\ell, R_{w,1}, R_{w,2}, D_1, D_2)$ such that (11)-(13) and if $R_0 < \overline{R_0}$, then*

$$R_s \geq \max_{j=1,2} \Big(I(U_j; \widetilde{X}_j|Z) + R^{**} - R_0\Big) \tag{26}$$

$$R_\ell \geq \max_{j=1,2} \Big(I(U_j; X|Z) + R^{**} - R_0\Big) \tag{27}$$

if $\overline{R_0} \leq R_0 < \overline{\overline{R_0}}$, then

$$R_s \geq \max_{j=1,2} I(V_j; \widetilde{X}_j|Z) \tag{28}$$

$$R_\ell \geq \max_{j=1,2} I(V_j; X|Z) \tag{29}$$

and if $R_0 \geq \overline{\overline{R_0}}$, then (18) and (19), where $P_{QV_1V_2U_1U_2\widetilde{X}_1\widetilde{X}_2XYZ}$ is equal to (20) such that we have (21) for some reconstruction function $\widehat{\widetilde{X}}_j(U_1, U_2, Y)$.

Lemma 4 (Outer Bound)**.** *The region $\mathcal{R}_{\text{D,weak}}$ is included in the union of the rate tuples $(R_s, R_\ell, R_{w,1}, R_{w,2}, D_1, D_2)$ in (13), (22), and (23), as well as if $R_0 < \overline{R_0}$, then (26) and (27); if $\overline{R_0} \leq R_0 < \overline{\overline{R_0}}$, then (28) and (29); and if $R_0 \geq \overline{R_0}$, then (18) and (19), where unions are over all $P_{QV_1V_2U_1U_2|\widetilde{X}_1\widetilde{X}_2}$ such that we have (21) for some reconstruction function $\widehat{\widetilde{X}}_j(U_1, U_2, Y)$ and (24) form Markov chains for $j = 1, 2$. One can limit the cardinalities to $|\mathcal{V}_j| \leq |\widetilde{\mathcal{X}}_j| + 6$, $|\mathcal{U}_j| \leq (|\widetilde{\mathcal{X}}_j| + 6)^2$, and $|\mathcal{Q}| \leq 2$.*

One can show that if $\widetilde{X}_j^n$ should be reconstructed reliably, i.e., replace (4) with

$$\Pr\Big[\{\widehat{\widetilde{X}_1^n} \neq \widetilde{X}_1^n\} \bigcup \{\widehat{\widetilde{X}_2^n} \neq \widetilde{X}_2^n\}\Big] \leq \delta \tag{30}$$

to consider secure and private distributed lossless source coding, then the bounds given in Lemmas 1-4 can be simplified by assigning $U_j = \widetilde{X}_j$ such that $\widehat{\widetilde{X}}_j(\widetilde{X}_j, U_{j'}, Y) = \widetilde{X}_j$ and we achieve $D_j = 0$, which satisfies the reliability constraint since $d(\cdot, \cdot)$ is a distortion metric. Simplified inner and outer bounds for the lossless settings do not match in general due to different Markov chain conditions imposed, which is similar to Lemmas 1-4. We remark that computation of partially-invertible and invertible functions [33] is entirely similar to lossless source reconstruction and similar simplification steps can be applied to function computation rate region bounds for such functions. In [31], the auxiliary random variables V_1 and V_2 are chosen to be constant for such special function computation settings, which results in achievable regions that are suboptimal when secrecy or privacy constraint is imposed. Furthermore, since there are two auxiliary random variables to be optimized for each encoder in Lemmas 1-4, evaluating the rate region bounds for secure and private distributed source coding problems with two encoders is significantly more involved than evaluating the rate regions for such problems with a single encoder; see [26] for a Gaussian example for the latter.

V. Proofs of Lemmas 1 and 2

A. Proof of Lemma 1 (Inner Bound)

Proof Sketch: We use the output statistics of random binning (OSRB) method [17], [34], [35] for the achievability proof by following the steps described in [36, Section 1.6]. The code construction below proposed for low private-key rates, i.e., $R_0 < \overline{R_0}$, is similar to the construction in [31, Section 5.1], but we apply additional binning steps to leverage the private keys to reduce secrecy- and privacy-leakage rates. For middle and high private key rates, we change the code construction to further reduce the secrecy- and privacy-leakage rates.

Let $(V_1^n, V_2^n, U_1^n, U_2^n, \widetilde{X}_1^n, \widetilde{X}_2^n, X^n, Y^n, Z^n)$ be i.i.d. according to $P_{V_1V_2U_1U_2\widetilde{X}_1\widetilde{X}_2XYZ}$ that can be obtained by fixing probabilities $P_{V_1|U_1}$, $P_{U_1|\widetilde{X}_1}$, $P_{V_2|U_2}$, and $P_{U_2|\widetilde{X}_2}$ in (20) such that $\mathbb{E}[d(\widetilde{X}_j, \widehat{\widetilde{X}_j})] \leq (D_j + \epsilon)$ for $j = 1, 2$ and any $\epsilon > 0$. To each v_1^n assign two random bin indices $F_{\mathsf{v}_1} \in [1 : 2^{n\widetilde{R}_{\mathsf{v}_1}}]$ and $W_{\mathsf{v}_1} \in [1 : 2^{nR_{\mathsf{v}_1}}]$. To each u_1^n assign three random bin indices $F_{\mathsf{u}_1} \in [1 : 2^{n\widetilde{R}_{\mathsf{u}_1}}]$, $W_{\mathsf{u}_1} \in [1 : 2^{nR_{\mathsf{u}_1}}]$, and $K_{\mathsf{u}_1} \in [1 : 2^{nR_0}]$, where R_0 is the private key rate defined in Section II. Similarly, random indices $(F_{\mathsf{v}_2}, W_{\mathsf{v}_2})$ and $(F_{\mathsf{u}_2}, W_{\mathsf{u}_2}, K_{\mathsf{u}_2})$ are assigned to each v_2^n and u_2^n, respectively, where $F_{\mathsf{v}_2} \in [1 : 2^{n\widetilde{R}_{\mathsf{v}_2}}]$, $W_{\mathsf{v}_2} \in [1 : 2^{nR_{\mathsf{v}_2}}]$, $F_{\mathsf{u}_2} \in [1 : 2^{n\widetilde{R}_{\mathsf{u}_2}}]$, $W_{\mathsf{u}_2} \in [1 : 2^{nR_{\mathsf{u}_2}}]$, and $K_{\mathsf{u}_2} \in [1 : 2^{nR_0}]$.

The public indices $F_1 = (F_{\mathsf{v}_1}, F_{\mathsf{u}_1})$ and $F_2 = (F_{\mathsf{v}_2}, F_{\mathsf{u}_2})$ represent the choice of the source encoders and the source decoder. Furthermore, we impose that the messages sent by the source encoders $\mathsf{Enc}_1(\cdot, \cdot)$ and $\mathsf{Enc}_2(\cdot, \cdot)$ to the source decoder $\mathsf{Dec}(\cdot, \cdot, \cdot, \cdot)$ are

$$W_1 = (W_{\mathsf{v}_1}, W_{\mathsf{u}_1}, K_1 + K_{\mathsf{u}_1}) \tag{31}$$

$$W_2 = (W_{\mathsf{v}_2}, W_{\mathsf{u}_2}, K_2 + K_{\mathsf{u}_2}) \tag{32}$$

where the summations with the private keys are in modulo-2^{nR_0}, i.e., one-time padding.

We impose the following decoding order

1) using $(Y^n, F_{\mathsf{v}_1}, W_{\mathsf{v}_1})$, the decoder estimates V_1^n as $\widehat{V}_1^n$;
2) using $(Y^n, \widehat{V}_1^n, F_{\mathsf{v}_2}, W_{\mathsf{v}_2})$, the decoder estimates V_2^n as $\widehat{V}_2^n$;
3) using $(Y^n, K_1, \widehat{V}_1^n, \widehat{V}_2^n, F_{\mathsf{u}_1}, W_{\mathsf{u}_1}, K_1{+}K_{\mathsf{u}_1})$, the decoder estimates U_1^n as $\widehat{U}_1^n$;
4) using $(Y^n, K_2, \widehat{V}_1^n, \widehat{V}_2^n, \widehat{U}_1^n, F_{\mathsf{u}_2}, W_{\mathsf{u}_2}, K_2{+}K_{\mathsf{u}_2})$, the decoder estimates U_2^n as $\widehat{U}_2^n$.

Swapping the indices 1 and 2 above, one can achieve another corner point in the achievable rate region. Due to symmetry, the analysis for the swapped decoding order is entirely similar to the analysis for the decoding order given above, so we consider only the latter.

The public index F_{v_1} is almost independent of $(\widetilde{X}_1^n, \widetilde{X}_2^n, X^n, Y^n, Z^n)$ if we have [35, Theorem 1]

$$\widetilde{R}_{\mathsf{v}_1} < H(V_1|\widetilde{X}_1, \widetilde{X}_2, X, Y, Z) \overset{(a)}{=} H(V_1|\widetilde{X}_1) \tag{33}$$

where (a) follows since $(\widetilde{X}_2, X, Y, Z) - \widetilde{X}_1 - V_1$ form a Markov chain. The constraint in (33) suggests that the expected value, which is taken over the random bin assignments, of the variational distance between the joint probability distributions $\text{Unif}[1 : 2^{n\widetilde{R}_{\text{v}_1}}] \cdot P_{\widetilde{X}_1^n}$ and $P_{F_{\text{v}_1}\widetilde{X}_1^n}$ vanishes when $n \to \infty$. Moreover, the public index F_{u_1} is almost independent of $(V_1^n, \widetilde{X}_1^n, \widetilde{X}_2^n, X^n, Y^n, Z^n)$ if

$$\widetilde{R}_{\text{u}_1} < H(U_1|V_1, \widetilde{X}_1, \widetilde{X}_2, X, Y, Z) \overset{(a)}{=} H(U_1|V_1, \widetilde{X}_1) \quad (34)$$

where (a) follows from the Markov chain condition $(\widetilde{X}_2, X, Y, Z) - (\widetilde{X}_1, V_1) - U_1$. Similarly, F_{v_2} is almost independent of $(\widetilde{X}_1^n, \widetilde{X}_2^n, X^n, Y^n, Z^n)$ if we have

$$\widetilde{R}_{\text{v}_2} < H(V_2|\widetilde{X}_1, \widetilde{X}_2, X, Y, Z) \overset{(a)}{=} H(V_2|\widetilde{X}_2) \quad (35)$$

where (a) follows since $(\widetilde{X}_1, X, Y, Z) - \widetilde{X}_2 - V_2$ form a Markov chain, and F_{u_2} is almost independent of $(V_2^n, \widetilde{X}_1^n, \widetilde{X}_2^n, X^n, Y^n, Z^n)$ if

$$\widetilde{R}_{\text{u}_2} < H(U_2|V_2, \widetilde{X}_1, \widetilde{X}_2, X, Y, Z) \overset{(a)}{=} H(U_2|V_2, \widetilde{X}_2) \quad (36)$$

where (a) follows from the Markov chain condition $(\widetilde{X}_1, X, Y, Z) - (\widetilde{X}_2, V_2) - U_2$.

Using a Slepian-Wolf (SW) [1] decoder that observes $(Y^n, F_{\text{v}_1}, W_{\text{v}_1})$, one can reliably estimate V_1^n if we have [35, Lemma 1]

$$\widetilde{R}_{\text{v}_1} + R_{\text{v}_1} > H(V_1|Y) \quad (37)$$

since then the expected error probability taken over random bin assignments vanishes when $n \to \infty$. Similarly, Step 2 estimation, given above as the second step in the imposed decoding order, is reliable if we have

$$\widetilde{R}_{\text{v}_2} + R_{\text{v}_2} > H(V_2|V_1, Y). \quad (38)$$

Moreover, one can reliably estimate U_1^n from $(Y^n, K_1, V_1^n, V_2^n, F_{\text{u}_1}, W_{\text{u}_1}, K_1 + K_{u_1})$ if we have

$$R_0 + \widetilde{R}_{\text{u}_1} + R_{\text{u}_1} > H(U_1|V_1, V_2, Y). \quad (39)$$

Similarly, Step 4 estimation is reliable if we have

$$\begin{aligned} &R_0 + \widetilde{R}_{\text{u}_2} + R_{\text{u}_2} \\ &\quad > H(U_2|V_1, V_2, U_1, Y) \overset{(a)}{=} H(U_2|V_2, U_1, Y) \end{aligned} \quad (40)$$

where (a) follows from the Markov chain condition $U_2 - (U_1, V_2, Y) - V_1$.

To satisfy (33)-(40), for any $\epsilon > 0$ we fix

$$\widetilde{R}_{\text{v}_1} = H(V_1|\widetilde{X}_1) - \epsilon \quad (41)$$

$$R_{\text{v}_1} = I(V_1; \widetilde{X}_1) - I(V_1; Y) + 2\epsilon \quad (42)$$

$$\widetilde{R}_{\text{u}_1} = H(U_1|V_1, \widetilde{X}_1) - \epsilon \quad (43)$$

$$R_0 + R_{\text{u}_1} = I(U_1; \widetilde{X}_1|V_1) - I(U_1; V_2, Y|V_1) + 2\epsilon \quad (44)$$

$$\widetilde{R}_{\text{v}_2} = H(V_2|\widetilde{X}_2) - \epsilon \quad (45)$$

$$R_{\text{v}_2} = I(V_2; \widetilde{X}_2) - I(V_2; V_1, Y) + 2\epsilon \quad (46)$$

$$\widetilde{R}_{\text{u}_2} = H(U_2|V_2, \widetilde{X}_2) - \epsilon \quad (47)$$

$$R_0 + R_{\text{u}_2} = I(U_2; \widetilde{X}_2|V_2) - I(U_2; U_1, Y|V_2) + 2\epsilon. \quad (48)$$

Communication Rates: (42) and (44) result in a communication (storage) rate of

$$\begin{aligned} R_{\text{w}_1} &= R_0 + R_{\text{v}_1} + R_{\text{u}_1} \\ &\overset{(a)}{=} I(V_1; \widetilde{X}_1|Y) + H(U_1|V_1, V_2, Y) - H(U_1|V_1, \widetilde{X}_1) + 4\epsilon \\ &\overset{(b)}{=} I(V_1; \widetilde{X}_1|Y) + I(U_1; \widetilde{X}_1|V_1, V_2, Y) + 4\epsilon \end{aligned} \quad (49)$$

where (a) follows since $V_1 - \widetilde{X}_1 - Y$ form a Markov chain and (b) follows since $U_1 - (V_1, \widetilde{X}_1) - (V_2, Y)$ form a Markov chain. Similarly, (46) and (48) result in a communication rate of

$$\begin{aligned} R_{\text{w}_2} &= R_0 + R_{\text{v}_2} + R_{\text{u}_2} \\ &\overset{(a)}{=} I(V_2; \widetilde{X}_2|V_1, Y) + H(U_2|U_1, V_2, Y) \\ &\qquad - H(U_2|V_2, \widetilde{X}_2) + 4\epsilon \\ &\overset{(b)}{=} I(V_2; \widetilde{X}_2|V_1, Y) + I(U_2; \widetilde{X}_2|U_1, V_2, Y) + 4\epsilon \end{aligned} \quad (50)$$

where (a) follows since $V_2 - \widetilde{X}_2 - (V_1, Y)$ form a Markov chain and (b) follows from the Markov chain condition $U_2 - (V_2, \widetilde{X}_2) - (U_1, Y)$. By swapping the indices 1 and 2 in the decoding order given above, one can achieve the other corner point with

$$R'_{\text{w}_1} = I(V_1; \widetilde{X}_1|V_2, Y) + I(U_1; \widetilde{X}_1|U_2, V_1, Y) + 4\epsilon \quad (51)$$

$$R'_{\text{w}_2} = I(V_2; \widetilde{X}_2|Y) + I(U_2; \widetilde{X}_2|V_1, V_2, Y) + 4\epsilon. \quad (52)$$

Privacy Leakage: Since 1) the private keys $K = (K_1, K_2)$ are independent of the source and channel random variables; 2) random encoders given above are constructed independently; 3) K is uniformly distributed; and 4) all random bin indices can be shown to be mutually independent for any $\epsilon > 0$ such that $\epsilon \to 0$ when $n \to \infty$, we can consider the following virtual scenario to calculate the leakage. Suppose first that there is no private key such that the encoder outputs for the virtual scenario are

$$\overline{W}_1 = (W_{\text{v}_1}, W_{\text{u}_1}, K_{\text{u}_1}) \quad (53)$$

$$\overline{W}_2 = (W_{\text{v}_2}, W_{\text{u}_2}, K_{\text{u}_2}). \quad (54)$$

We first calculate the leakage for the virtual scenario. Then, given the mentioned four properties and due to the one-time padding steps in (31) and (32), we can subtract $H(K) = 2nR_0$ from the leakage calculated for the virtual scenario to obtain the leakage for the original problem. Thus, we have the privacy leakage

$$\begin{aligned} &I(X^n; W_1, W_2, F_1, F_2|Z^n) \\ &= I(X^n; \overline{W}_1, \overline{W}_2, F_1, F_2|Z^n) - 2nR_0 \\ &\overset{(a)}{=} H(\overline{W}_1, \overline{W}_2, F_1, F_2|Z^n) - H(\overline{W}_1, \overline{W}_2, F_1, F_2|X^n) - 2nR_0 \\ &\overset{(b)}{=} H(\overline{W}_1, \overline{W}_2, F_1, F_2|Z^n) \\ &\qquad - H(U_1^n, U_2^n, V_1^n, V_2^n|X^n) \\ &\qquad + H(V_1^n|\overline{W}_1, \overline{W}_2, F_1, F_2, X^n) \\ &\qquad + H(V_2^n|V_1^n, \overline{W}_1, \overline{W}_2, F_1, F_2, X^n) \\ &\qquad + H(U_1^n|V_1^n, V_2^n, \overline{W}_1, \overline{W}_2, F_1, F_2, X^n) \\ &\qquad + H(U_2^n|U_1^n, V_1^n, V_2^n, \overline{W}_1, \overline{W}_2, F_1, F_2, X^n) - 2nR_0 \end{aligned}$$

$$\begin{aligned}
&\overset{(c)}{\leq} H(\overline{W}_1,\overline{W}_2,F_1,F_2|Z^n)\\
&\quad - H(U_1^n,U_2^n,V_1^n,V_2^n|X^n) + 4n\epsilon_n' - 2nR_0\\
&\overset{(d)}{=} H(\overline{W}_1,\overline{W}_2,F_1,F_2|Z^n)\\
&\quad - nH(U_1,U_2,V_1,V_2|X) + 4n\epsilon_n' - 2nR_0 \qquad (55)
\end{aligned}$$

where
(a) follows because $(\overline{W}_1,\overline{W}_2,F_1,F_2) - X^n - Z^n$ form a Markov chain;
(b) follows since $(U_1^n,U_2^n,V_1^n,V_2^n)$ determine $(F_{u_1},W_{u_1},K_{u_1},F_{v_1},W_{v_1},F_{u_2},W_{u_2},K_{u_2},F_{v_2},W_{v_2})$;
(c) follows for some $\epsilon_n' > 0$ such that $\epsilon_n' \to 0$ when $n \to \infty$ since (F_{v_1},W_{v_1},X^n) can reliably recover V_1^n by (37) and, similarly, $(F_{v_2},W_{v_2},V_1^n,X^n)$ can reliably recover V_2^n by (38) both due to the Markov chain condition

$$(V_1^n,V_2^n) - (U_1^n,U_2^n) - (\widetilde{X}_1^n,\widetilde{X}_2^n) - X^n - Y^n. \qquad (56)$$

Moreover, because of the Markov chain condition in (56) $(F_{u_1},W_{u_1},K_{u_1},V_1^n,V_2^n,X^n)$ can reliably recover U_1^n by (39) and, similarly, $(F_{u_2},W_{u_2},K_{u_2},U_1^n,V_1^n,V_2^n,X^n)$ can reliably recover U_2^n by (40), which follows from $H(U_2|U_1,V_1,V_2,Y) \geq H(U_2|U_1,V_1,V_2,X)$ that can be obtained as

$$\begin{aligned}
&H(U_2|U_1,V_1,V_2,Y) - H(U_2|U_1,V_1,V_2,X)\\
&= I(U_2;U_1,V_1,V_2,X) - I(U_2;U_1,V_1,V_2,Y)\\
&\geq I(U_2;U_1,V_1,V_2,X) - I(U_2;U_1,V_1,V_2,X,Y) \overset{(a)}{=} 0 \qquad (57)
\end{aligned}$$

where (a) follows from the Markov chain condition $(U_1,U_2,V_1,V_2) - X - Y$;
(d) follows since $(U_1^n,U_2^n,V_1^n,V_2^n,X^n)$ are i.i.d.

Next, we consider the term $H(\overline{W}_1,\overline{W}_2,F_1,F_2|Z^n)$ in (55) and provide single letter bounds on it, which requires to analyze numerous decoding cases. In [3, Section V-A], six different decodability cases are analyzed for two decoding steps. We can leverage these results by combining the decoding order Steps 1 and 2 to treat (V_1^n,V_2^n) jointly, as well as combining Steps 3 and 4 to treat (U_1^n,U_2^n) jointly; see also [31, Section 5.1]. Thus, by replacing V^n with (V_1^n,V_2^n) and U^n with (U_1^n,U_2^n) in [3, Section V-A], respectively, we can apply the results of the six decodability analyses to (55). Furthermore, the subtracted term in [3, Eq. (54)] can be mapped to the second term in (55) by applying the same replacements. Thus, combining the replaced decodability analyses with (55), we obtain

$$\begin{aligned}
&I(X^n;W_1,W_2,F_1,F_2|Z^n)\\
&\leq n\big([I(U_1,U_2;Z|V_1,V_2) - I(U_1,U_2;Y|V_1,V_2) + \epsilon]^-\\
&\qquad + I(U_1,U_2;X|Z) - 2R_0 + 5\epsilon_n'\big). \qquad (58)
\end{aligned}$$

We remark that (44) and (48) implicitly assume that $R_0 < (\min\{I(U_1;\widetilde{X}_1|V_1,V_2,Y), I(U_2;\widetilde{X}_2|V_2,U_1,Y)\} + 2\epsilon)$, since the private keys K_1 and K_2 are used to apply one-time padding to one bin of U_1^n and U_2^n, respectively. The communication rate results are not affected by this assumption since both $\widetilde{X}_1$ and $\widetilde{X}_2$ should be reconstructed by the decoder. However, the leakage analysis changes if $R_0 \geq (\max\{I(U_1;\widetilde{X}_1|V_1,V_2,Y), I(U_2;\widetilde{X}_2|V_2,U_1,Y)\} + 2\epsilon)$, because then the private keys can be used to apply one-time padding to the single bins of U_1^n and U_2^n with encoder outputs

$$\overline{\overline{W}}_1 = (W_{\mathrm{v}_1}, W_{\mathrm{u}_1} + K_1) \qquad (59)$$

$$\overline{\overline{W}}_2 = (W_{\mathrm{v}_2}, W_{\mathrm{u}_2} + K_2) \qquad (60)$$

where the bins with indices K_{u_1} and K_{u_2} are removed from the code construction. We then have the privacy leakage

$$\begin{aligned}
&I(X^n;\overline{\overline{W}}_1,\overline{\overline{W}}_2,F_1,F_2|Z^n)\\
&= H(X^n|Z^n) - H(X^n|Z^n,\overline{\overline{W}}_1,\overline{\overline{W}}_2,F_1,F_2)\\
&\overset{(a)}{=} H(X^n|Z^n) - H(X^n|Z^n,W_{v_1},W_{v_2},F_1,F_2)\\
&\overset{(b)}{\leq} H(X^n|Z^n) - H(X^n|Z^n,V_1^n,V_2^n) + 2\epsilon_n'\\
&\overset{(c)}{=} nH(X|Z) - nH(X|Z,V_1,V_2) + 2\epsilon_n'\\
&= nI(V_1,V_2;X|Z) + 2\epsilon_n' \qquad (61)
\end{aligned}$$

where (a) follows since $(W_{\mathrm{u}_1} + K_1)$ is independent of $(U_1^n,U_2^n,V_1^n,V_2^n,\widetilde{X}_1^n,\widetilde{X}_2^n,X^n,Y^n,Z^n)$ due to the one-time padding with a uniform and independent private key and, similarly, $(W_{\mathrm{u}_2} + K_2)$ is also independent of $(U_1^n,U_2^n,V_1^n,V_2^n,\widetilde{X}_1^n,\widetilde{X}_2^n,X^n,Y^n,Z^n)$, (b) follows since (V_1^n,V_2^n) determine $(W_{v_1},W_{v_2},F_{v_1},F_{v_2})$ and for some $\epsilon_n' > 0$ such that $\epsilon_n' \to 0$ when $n \to \infty$ because by (34) F_{u_1} is almost independent of $(V_1^n,\widetilde{X}_1^n,\widetilde{X}_2^n,X^n,Z^n)$ and by (36) F_{u_2} is almost independent of $(V_2^n,\widetilde{X}_1^n,\widetilde{X}_2^n,X^n,Z^n)$, respectively, and (c) follows because (V_1^n,V_2^n,X^n,Z^n) are i.i.d.

Consider next that R_0 is greater than or equal to the maximum of $(I(V_1;\widetilde{X}_1|Y) + I(U_1;\widetilde{X}_1|V_1,V_2,Y) + 4\epsilon)$ and $(I(V_2;\widetilde{X}_2|V_1,Y) + I(U_2;\widetilde{X}_2|U_1,V_2,Y) + 4\epsilon)$, given in (49) and (50), respectively. Then, j-th encoder can apply one-time padding to both W_{v_j} and W_{u_j} for $j = 1,2$, so no information is leaked to the eavesdropper about $(W_{\mathrm{v}_1},W_{\mathrm{u}_1},W_{\mathrm{v}_2},W_{\mathrm{u}_2})$. Therefore, we obtain the privacy leakage of

$$\begin{aligned}
&I(X^n;F_1,F_2|Z^n) \overset{(a)}{\leq} I(X^n;F_1|Z^n) + I(X^n;F_2|Z^n) + 2\epsilon_n'\\
&= I(X^n;F_{\mathrm{v}_1}|Z^n) + I(X^n;F_{\mathrm{u}_1}|Z^n,F_{\mathrm{v}_1})\\
&\qquad + I(X^n;F_{\mathrm{v}_2}|Z^n) + I(X^n;F_{\mathrm{u}_2}|Z^n,F_{\mathrm{v}_2}) + 2\epsilon_n'\\
&\overset{(b)}{\leq} 6\epsilon_n' \qquad (62)
\end{aligned}$$

where (a) follows by (35) and (36) since F_2 is almost independent of (F_1,X^n,Z^n) due to the Markov chain condition $F_1 - (\widetilde{X}_1^n,X^n,Z^n) - F_2$ and (b) follows similarly from the corresponding almost independence results by applying (33)-(36).

Secrecy Leakage: Similar to the privacy-leakage analysis above, suppose first the virtual scenario with encoder outputs given in (53) and (54), and then calculate the leakage for the original problem by subtracting $H(K) = 2nR_0$ from the leakage calculated for the virtual scenario. Thus, we obtain

$$\begin{aligned}
&I(\widetilde{X}_1^n,\widetilde{X}_2^n;W_1,W_2,F_1,F_2|Z^n)\\
&= I(\widetilde{X}_1^n,\widetilde{X}_2^n;\overline{W}_1,\overline{W}_2,F_1,F_2|Z^n) - 2nR_0
\end{aligned}$$

$$
\begin{aligned}
&\overset{(a)}{=} H(\overline{W}_1,\overline{W}_2,F_1,F_2|Z^n)\\
&\quad - H(\overline{W}_1,\overline{W}_2,F_1,F_2|\widetilde{X}_1^n,\widetilde{X}_2^n) - 2nR_0\\
&\overset{(b)}{=} H(\overline{W}_1,\overline{W}_2,F_1,F_2|Z^n) - H(U_1^n,U_2^n,V_1^n,V_2^n|\widetilde{X}_1^n,\widetilde{X}_2^n)\\
&\quad + H(V_1^n|\overline{W}_1,\overline{W}_2,F_1,F_2,\widetilde{X}_1^n,\widetilde{X}_2^n)\\
&\quad + H(V_2^n|V_1^n,\overline{W}_1,\overline{W}_2,F_1,F_2,\widetilde{X}_1^n,\widetilde{X}_2^n)\\
&\quad + H(U_1^n|V_1^n,V_2^n,\overline{W}_1,\overline{W}_2,F_1,F_2,\widetilde{X}_1^n,\widetilde{X}_2^n)\\
&\quad + H(U_2^n|U_1^n,V_1^n,V_2^n,\overline{W}_1,\overline{W}_2,F_1,F_2,\widetilde{X}_1^n,\widetilde{X}_2^n) - 2nR_0\\
&\overset{(c)}{\leq} H(\overline{W}_1,\overline{W}_2,F_1,F_2|Z^n) - nH(U_1,U_2,V_1,V_2|\widetilde{X}_1,\widetilde{X}_2)\\
&\quad + 4n\epsilon_n' - 2nR_0
\end{aligned}
\tag{63}
$$

where (a) follows because $(\overline{W}_1,\overline{W}_2,F_1,F_2)-(\widetilde{X}_1^n,\widetilde{X}_2^n)-Z^n$ form a Markov chain, (b) follows since $(U_1^n,U_2^n,V_1^n,V_2^n)$ determine $(\overline{W}_1,\overline{W}_2,F_1,F_2)$, and (c) follows since $(V_1^n,V_2^n,U_1^n,U_2^n,\widetilde{X}_1^n,\widetilde{X}_2^n)$ are i.i.d. and because $(F_{v_1},W_{v_1},\widetilde{X}_1^n,\widetilde{X}_2^n)$ can reliably recover V_1^n by (37) and, similarly, $(F_{v_2},W_{v_2},V_1^n,\widetilde{X}_1^n,\widetilde{X}_2^n)$ can reliably recover V_2^n by (38) both due to the Markov chain condition $(V_1^n,V_2^n)-(\widetilde{X}_1^n,\widetilde{X}_2^n)-Y^n$. Recoverability of (U_1^n,U_2^n) follows similarly by (39) and (40). The terms in (63) can be obtained from the terms in (55) by replacing X with $(\widetilde{X}_1,\widetilde{X}_2)$. Thus, we apply the results of the decodability analyses from [3, Section V-A], applied to (55) above, also to (63) such that by replacing X with $(\widetilde{X}_1,\widetilde{X}_2)$ in (58), we have

$$
\begin{aligned}
&I(\widetilde{X}_1^n,\widetilde{X}_2^n;W_1,W_2,F_1,F_2|Z^n)\\
&\leq n\big([I(U_1,U_2;Z|V_1,V_2) - I(U_1,U_2;Y|V_1,V_2)+\epsilon]^-\\
&\qquad + I(U_1,U_2;\widetilde{X}_1,\widetilde{X}_2|Z) - 2R_0 + 5\epsilon_n'\big).
\end{aligned}
\tag{64}
$$

Furthermore, if we have

$$
R_0 \geq \max\{I(U_1;\widetilde{X}_1|V_1,V_2,Y),\ I(U_2;\widetilde{X}_2|V_2,U_1,Y)\} + 2\epsilon
$$

then by applying entirely similar steps as in (61) after replacing X^n with $(\widetilde{X}_1^n,\widetilde{X}_2^n)$, we obtain the secrecy leakage

$$
\begin{aligned}
&I(\widetilde{X}_1^n,\widetilde{X}_2^n;W_1,W_2,F_1,F_2|Z^n)\\
&\qquad \leq nI(V_1,V_2;\widetilde{X}_1,\widetilde{X}_2|Z).
\end{aligned}
\tag{65}
$$

Similarly, after replacing X^n with $(\widetilde{X}_1^n,\widetilde{X}_2^n)$ in (62), one can upper bound the secrecy leakage for the high private key-rate case, i.e., R_0 is greater than or equal to the maximum of $(I(V_1;\widetilde{X}_1|Y)+I(U_1;\widetilde{X}_1|V_1,V_2,Y)+4\epsilon)$ and $(I(V_2;\widetilde{X}_2|V_1,Y)+I(U_2;\widetilde{X}_2|U_1,V_2,Y)+4\epsilon)$, by $6\epsilon_n'$ for some $\epsilon_n'>0$ such that $\epsilon_n'\to 0$ when $n\to\infty$.

Suppose public indices (F_1,F_2) are generated uniformly at random and encoders generate $(V_1^n,V_2^n,U_1^n,U_2^n)$ according to $P_{V_1^nV_2^nU_1^nU_2^n|\widetilde{X}_1^nF_1\widetilde{X}_2^nF_2}$ that can be obtained from the binning scheme applied above. Such a procedure effectuates a joint probability distribution that is almost equal to $P_{V_1V_2U_1U_2\widetilde{X}_1\widetilde{X}_2XYZ}$ one can obtain from the fixed distribution given in (20) [36, Section 1.6]. Since both leakage metrics are expectations over all possible realizations $(F_1=f_1,F_2=f_2)$, using the selection lemma [37, Lemma 2.2] we prove Lemma 1 by choosing an $\epsilon>0$ such that $\epsilon\to 0$ when $n\to\infty$. The time-sharing random variable Q, for which we impose $P_{QV_1V_2}=P_QP_{V_1|Q}P_{V_2|Q}$, enlarges the rate region via convexification. ∎

B. Proof of Lemma 2 (Outer Bound)

Proof Sketch: Suppose for some $n\geq 1$ and $\delta_n>0$, there exist two encoders and a decoder such that (1)-(4) are satisfied against a strong eavesdropper for some tuple $(R_{\mathrm{s}},R_\ell,R_{\mathrm{w},1},R_{\mathrm{w},2},D_1,D_2)$ given two private keys each with rate R_0. Define

$$
V_{j,i} \triangleq (W_j,Y_{i+1}^n,Z^{i-1}) \tag{66}
$$

$$
U_{j,i} \triangleq (W_j,Y_{i+1}^n,Z^{i-1},X^{i-1},K_j) \tag{67}
$$

that satisfy the Markov chain condition

$$
V_{j,i}-U_{j,i}-\widetilde{X}_{j,i}-X_i-(\widetilde{X}_{j',i},Y_i,Z_i). \tag{68}
$$

For $j=1,2$, we obtain

$$
\begin{aligned}
D_j+\delta_n &\overset{(a)}{\geq} \mathbb{E}\Big[d\Big(\widetilde{X}_j^n,\widehat{\widetilde{X}_j^n}(W_1,W_2,Y^n,K)\Big)\Big]\\
&\geq \mathbb{E}\Big[d\Big(\widetilde{X}_j^n,\widehat{\widetilde{X}_j^n}(W_1,W_2,Y^n,K,X^{i-1},Z^{i-1})\Big)\Big]\\
&\overset{(b)}{=} \mathbb{E}\Big[d\Big(\widetilde{X}_j^n,\widehat{\widetilde{X}_j^n}(W_1,W_2,Y_i^n,K,X^{i-1},Z^{i-1})\Big)\Big]\\
&\overset{(c)}{=} \frac{1}{n}\sum_{i=1}^n \mathbb{E}\Big[d\Big(\widetilde{X}_{j,i},\widehat{\widetilde{X}_{j,i}}(U_{1,i},U_{2,i},Y_i)\Big)\Big]
\end{aligned}
\tag{69}
$$

where (a) follows by (4), (b) follows from the Markov chain condition

$$
Y^{i-1}-(Y_i^n,X^{i-1},Z^{i-1},W_1,W_2,K)-\widetilde{X}_j^n \tag{70}
$$

and (c) follows from the definition of $U_{j,i}$.

Communication Rates: We have

$$
\begin{aligned}
n(R_{\mathrm{w}_j}+\delta_n) &\overset{(a)}{\geq} \log|\mathcal{W}_j|\\
&\geq H(W_j|Y^n,K_j) - H(W_j|\widetilde{X}_j^n,Y^n,K_j)\\
&\overset{(b)}{=} H(\widetilde{X}_j^n|Y^n) - \sum_{i=1}^n H(\widetilde{X}_{j,i}|\widetilde{X}_j^{i-1},W_j,Y^n,K_j)\\
&\overset{(c)}{=} H(\widetilde{X}_j^n|Y^n) - \sum_{i=1}^n H(\widetilde{X}_{j,i}|\widetilde{X}_j^{i-1},W_j,Y_{i+1}^n,Y_i,K_j)\\
&\overset{(d)}{\geq} H(\widetilde{X}_j^n|Y^n) - \sum_{i=1}^n H(\widetilde{X}_{j,i}|X^{i-1},Z^{i-1},W_j,Y_{i+1}^n,Y_i,K_j)\\
&\overset{(e)}{=} \sum_{i=1}^n I(U_{j,i};\widetilde{X}_{j,i}|Y_i)\\
&\overset{(f)}{=} \sum_{i=1}^n [I(V_{j,i};\widetilde{X}_{j,i}|Y_i) + I(U_{j,i};\widetilde{X}_{j,i}|Y_i,V_{j,i})]\\
&= \sum_{i=1}^n \Big[I(V_{j,i};\widetilde{X}_{j,i},V_{j',i}|Y_i) - I(V_{j,i};V_{j',i}|\widetilde{X}_{j,i},Y_i)\\
&\qquad + I(U_{j,i};\widetilde{X}_{j,i},U_{j',i}|Y_i,V_{j,i})\\
&\qquad - I(U_{j,i};U_{j',i}|\widetilde{X}_{j,i},Y_i,V_{j,i})\Big]
\end{aligned}
$$

$$\geq \sum_{i=1}^{n} \Big[I(V_{j,i}; \widetilde{X}_{j,i}|V_{j',i}, Y_i) - I(V_{j,i}; V_{j',i}|\widetilde{X}_{j,i}, Y_i) + I(U_{j,i}; \widetilde{X}_{j,i}|V_{j,i}, U_{j',i}, Y_i) - I(U_{j,i}; U_{j',i}|\widetilde{X}_{j,i}, Y_i, V_{j,i}) \Big] \quad (71)$$

where (a) follows by (3), (b) follows since K_j is independent of $(\widetilde{X}_j^n, Y^n)$, (c) follows from the Markov chain condition

$$Y^{i-1} - (\widetilde{X}_j^{i-1}, W_j, Y_{i+1}^n, Y_i, K_j) - \widetilde{X}_{j,i} \quad (72)$$

(d) follows by applying the data processing inequality to the Markov chain

$$(X^{i-1}, Z^{i-1}) - (\widetilde{X}_j^{i-1}, W_j, Y_{i+1}^n, Y_i, K_j) - \widetilde{X}_{j,i} \quad (73)$$

(e) follows since $(\widetilde{X}^n, Y^n)$ are i.i.d. and from the definition of $U_{j,i}$, and (f) follows from the Markov chain condition in (68).

Next, we bound the sum-rate as

$$\begin{aligned}
&n(R_{\mathrm{w}_1} + \delta_n) + n(R_{\mathrm{w}_2} + \delta_n) \\
&\overset{(a)}{\geq} \log(|\mathcal{W}_1| \cdot |\mathcal{W}_2|) \\
&\geq H(W_1, W_2|Y^n, K) - H(W_1, W_2|\widetilde{X}_1^n, \widetilde{X}_2^n, Y^n, K) \\
&\overset{(b)}{=} H(\widetilde{X}_1^n, \widetilde{X}_2^n|Y^n) - \sum_{i=1}^{n} H(\widetilde{X}_{1,i}, \widetilde{X}_{2,i}|Y_i^n, W_1, W_2, K, \widetilde{X}_1^{i-1}, \widetilde{X}_2^{i-1}) \\
&\overset{(c)}{\geq} \sum_{i=1}^{n} \Big[H(\widetilde{X}_{1,i}, \widetilde{X}_{2,i}|Y_i) - H(\widetilde{X}_{1,i}, \widetilde{X}_{2,i}|Y_i^n, W_1, W_2, K, X^{i-1}, Z^{i-1}) \Big] \\
&\overset{(d)}{=} \sum_{i=1}^{n} I(U_{1,i}, U_{2,i}; \widetilde{X}_{1,i}, \widetilde{X}_{2,i}|Y_i) \\
&\overset{(e)}{=} \sum_{i=1}^{n} \Big[I(U_{1,i}, U_{2,i}; \widetilde{X}_{1,i}, \widetilde{X}_{2,i}|V_{1,i}, V_{2,i}, Y_i) + I(V_{1,i}, V_{2,i}; \widetilde{X}_{1,i}, \widetilde{X}_{2,i}|Y_i) \Big] \\
&\overset{(f)}{=} \sum_{i=1}^{n} \Big[I(U_{1,i}; \widetilde{X}_{1,i}, \widetilde{X}_{2,i}|V_{1,i}, V_{2,i}, Y_i) + I(U_{2,i}; \widetilde{X}_{1,i}, \widetilde{X}_{2,i}|U_{1,i}, V_{2,i}, Y_i) \\
&\qquad + I(V_{1,i}; \widetilde{X}_{1,i}, \widetilde{X}_{2,i}|Y_i) + I(V_{2,i}; \widetilde{X}_{1,i}, \widetilde{X}_{2,i}|V_{1,i}, Y_i) \Big] \\
&\geq \sum_{i=1}^{n} \Big[I(U_{1,i}; \widetilde{X}_{1,i}|V_{1,i}, V_{2,i}, Y_i) + I(U_{2,i}; \widetilde{X}_{2,i}|U_{1,i}, V_{2,i}, Y_i) \\
&\qquad + I(V_{1,i}; \widetilde{X}_{1,i}|Y_i) + I(V_{2,i}; \widetilde{X}_{2,i}|V_{1,i}, Y_i) \Big] \quad (74)
\end{aligned}$$

where (a) follows by (3), (b) follows since K is independent of $(\widetilde{X}_1^n, \widetilde{X}_2^n, Y^n)$ and from the Markov chain condition

$$Y^{i-1} - (W_1, W_2, Y_i^n, K, \widetilde{X}_1^{i-1}, \widetilde{X}_2^{i-1}) - (\widetilde{X}_{1,i}, \widetilde{X}_{2,i}) \quad (75)$$

(c) follows because $(\widetilde{X}_1^n, \widetilde{X}_2^n, Y^n)$ are i.i.d. and from the data processing inequality applied to the Markov chain

$$(X^{i-1}, Z^{i-1}) - (\widetilde{X}_1^{i-1}, \widetilde{X}_2^{i-1}, W_1, W_2, Y_i^n, K) - (\widetilde{X}_{1,i}, \widetilde{X}_{2,i}) \quad (76)$$

(d) follows from the definitions of $U_{1,i}$ and $U_{2,i}$, (e) follows from the Markov chain condition

$$(V_{1,i}, V_{2,i}) - (U_{1,i}, U_{2,i}) - (\widetilde{X}_{1,i}, \widetilde{X}_{2,i}) - Y_i \quad (77)$$

and (f) follows from the Markov chain condition

$$V_{1,i} - (U_{1,i}, Y_i, V_{2,i}) - (U_{2,i}, \widetilde{X}_{1,i}, \widetilde{X}_{2,i}). \quad (78)$$

Privacy Leakage: We have

$$\begin{aligned}
&n(R_\ell + \delta_n) \\
&\overset{(a)}{\geq} [I(W_1, W_2; Y^n) - I(W_1, W_2; Z^n)] + [I(W_1, W_2; X^n) - I(W_1, W_2; Y^n)] \\
&\overset{(b)}{=} [I(W_1, W_2; Y^n) - I(W_1, W_2; Z^n)] + I(W_1, W_2; X^n|K) - I(K; X^n|W_1, W_2) \\
&\qquad - I(W_1, W_2; Y^n|K) + I(K; Y^n|W_1, W_2) \\
&\overset{(c)}{=} [I(W_1, W_2; Y^n) - I(W_1, W_2; Z^n)] + [I(W_1, W_2; X^n|K) - I(W_1, W_2; Y^n|K)] \\
&\qquad - I(K; X^n|W_1, W_2, Y^n) \\
&\overset{(d)}{\geq} \sum_{i=1}^{n} \Big[I(W_1, W_2; Y_i|Y_{i+1}^n, Z^{i-1}) - I(W_1, W_2; Z_i|Z^{i-1}, Y_{i+1}^n) \Big] \\
&\quad + \sum_{i=1}^{n} \Big[I(W_1, W_2; X_i|X^{i-1}, Y_{i+1}^n, K) - I(W_1, W_2; Y_i|Y_{i+1}^n, X^{i-1}, K) \Big] - H(K) \\
&\overset{(e)}{=} \sum_{i=1}^{n} \Big[I(W_1, W_2; Y_i|Y_{i+1}^n, Z^{i-1}) - I(W_1, W_2; Z_i|Z^{i-1}, Y_{i+1}^n) \Big] \\
&\quad + \sum_{i=1}^{n} \Big[I(W_1, W_2; X_i|X^{i-1}, Y_{i+1}^n, Z^{i-1}, K) - I(W_1, W_2; Y_i|Y_{i+1}^n, X^{i-1}, Z^{i-1}, K) - 2R_0 \Big] \\
&\overset{(f)}{=} \sum_{i=1}^{n} \Big[I(W_1, W_2, Y_{i+1}^n, Z^{i-1}; Y_i) - I(W_1, W_2, Z^{i-1}, Y_{i+1}^n; Z_i) \Big] \\
&\quad + \sum_{i=1}^{n} \Big[I(W_1, W_2, X^{i-1}, Y_{i+1}^n, Z^{i-1}, K; X_i) - I(W_1, W_2, Y_{i+1}^n, X^{i-1}, Z^{i-1}, K; Y_i) - 2R_0 \Big] \\
&\overset{(g)}{=} \sum_{i=1}^{n} \Big[I(V_{1,i}, V_{2,i}; Y_i) - I(V_{1,i}, V_{2,i}; Z_i) + I(U_{1,i}, U_{2,i} V_{1,i}, V_{2,i}; X_i) \\
&\qquad - I(U_{1,i}, U_{2,i}, V_{1,i}, V_{2,i}; Y_i) - 2R_0 \Big]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \Big[- I(U_{1,i}, U_{2,i}, V_{1,i}, V_{2,i}; Z_i) \\
&\qquad + I(U_{1,i}, U_{2,i}, V_{1,i}, V_{2,i}; X_i) \\
&\qquad + I(U_{1,i}, U_{2,i}; Z_i | V_{1,i}, V_{2,i}) \\
&\qquad - I(U_{1,i}, U_{2,i}; Y_i | V_{1,i}, V_{2,i}) - 2R_0 \Big] \\
&\overset{(h)}{\geq} \sum_{i=1}^{n} \Bigg[I(U_{1,i}, U_{2,i}; X_i | Z_i) - 2R_0 \\
&\qquad + \Big[I(U_{1,i}, U_{2,i}; Z_i | V_{1,i}, V_{2,i}) \\
&\qquad\qquad - I(U_{1,i}, U_{2,i}; Y_i | V_{1,i}, V_{2,i}) \Big]^{-} \Bigg] \qquad (79)
\end{aligned}
$$

where (a) follows by (2) and from the Markov chain condition $(W_1, W_2) - X^n - Z^n$, (b) follows since K is independent of (X^n, Y^n), (c) follows from the Markov chain condition $(W_1, W_2, K) - X^n - Y^n$, (d) follows from Csiszár's sum identity [38], (e) follows from the Markov chain condition

$$(X_i, Y_i, W_1, W_2) - (X^{i-1}, Y_{i+1}^n, K) - Z^{i-1} \qquad (80)$$

(f) follows since K is independent of (X^n, Y^n, Z^n) that are i.i.d., (g) follows from the definitions of $U_{j,i}$ and $V_{j,i}$, and (h) follows from the Markov chain condition

$$(V_{1,i}, V_{2,i}) - (U_{1,i}, U_{2,i}) - X_i - Z_i. \qquad (81)$$

We next consider the case $R_0 \geq \overline{R_0}$, defined in (9), and provide the corresponding outer bound for the achieved privacy-leakage rate in (61). We obtain

$$
\begin{aligned}
n(R_\ell + \delta_n) &\overset{(a)}{\geq} H(X^n|Z^n) - H(X^n|Z^n, W_1, W_2) \\
&\overset{(b)}{=} \sum_{i=1}^{n} [H(X_i|Z_i) - H(X_i|Z^i, W_1, W_2, X_{i+1}^n, Y_{i+1}^n)] \\
&\overset{(c)}{=} \sum_{i=1}^{n} [H(X_i|Z_i) - H(X_i|Z_i, V_{1,i}, V_{2,i}, X_{i+1}^n)] \\
&\geq \sum_{i=1}^{n} I(V_{1,i}, V_{2,i}; X_i|Z_i) \qquad (82)
\end{aligned}
$$

where (a) follows by (2), (b) follows because (X^n, Z^n) are i.i.d. and from the Markov chain condition

$$X_i - (X_{i+1}^n, W_1, W_2, Z^i) - (Z_{i+1}^n, Y_{i+1}^n) \qquad (83)$$

and (c) follows from the definitions of $V_{1,i}$ and $V_{2,i}$. Furthermore, the corresponding outer bound for the high private key-rate case, i.e., $R_0 \geq \overline{\overline{R_0}}$ defined in (10), with the privacy-leakage result in (62) follows since conditional mutual information is non-negative.

Secrecy Leakage: We obtain

$$
\begin{aligned}
&n(R_{\rm s} + \delta_n) \\
&\overset{(a)}{\geq} [I(W_1, W_2; Y^n) - I(W_1, W_2; Z^n)] \\
&\qquad + [I(W_1, W_2; \widetilde{X}_1^n, \widetilde{X}_2^n) - I(W_1, W_2; Y^n)] \\
&\overset{(b)}{=} [I(W_1, W_2; Y^n) - I(W_1, W_2; Z^n)] \\
&\qquad + I(W_1, W_2; \widetilde{X}_1^n, \widetilde{X}_2^n | K) - I(K; \widetilde{X}_1^n, \widetilde{X}_2^n | W_1, W_2) \\
&\qquad - I(W_1, W_2; Y^n | K) + I(K; Y^n | W_1, W_2) \\
&\overset{(c)}{=} [I(W_1, W_2; Y^n) - I(W_1, W_2; Z^n)] \\
&\qquad + I(W_1, W_2; \widetilde{X}_1^n, \widetilde{X}_2^n | K, Y^n) \\
&\qquad - I(K; \widetilde{X}_1^n, \widetilde{X}_2^n | W_1, W_2, Y^n) \\
&\overset{(d)}{\geq} \sum_{i=1}^{n} \Big[I(W_1, W_2; Y_i | Y_{i+1}^n, Z^{i-1}) \\
&\qquad\qquad - I(W_1, W_2; Z_i | Z^{i-1}, Y_{i+1}^n) \Big] \\
&\qquad + H(\widetilde{X}_1^n, \widetilde{X}_2^n | K, Y^n) - H(K) \\
&\qquad - \sum_{i=1}^{n} H(\widetilde{X}_{1,i}, \widetilde{X}_{2,i} | \widetilde{X}_1^{i-1}, \widetilde{X}_2^{i-1}, W_1, W_2, Y_i^n, K) \\
&\overset{(e)}{\geq} \sum_{i=1}^{n} \Big[I(W_1, W_2, Y_{i+1}^n, Z^{i-1}; Y_i) \\
&\qquad\qquad - I(W_1, W_2, Z^{i-1}, Y_{i+1}^n; Z_i) \Big] \\
&\qquad + \sum_{i=1}^{n} H(\widetilde{X}_{1,i}, \widetilde{X}_{2,i} | Y_i) - 2nR_0 \\
&\qquad - \sum_{i=1}^{n} H(\widetilde{X}_{1,i}, \widetilde{X}_{2,i} | X^{i-1}, Z^{i-1}, W_1, W_2, Y_i^n, K) \\
&\overset{(f)}{=} \sum_{i=1}^{n} \Big[I(V_{1,i}, V_{2,i}; Y_i) - I(V_{1,i}, V_{2,i}; Z_i) - 2R_0 \\
&\qquad\qquad + I(U_{1,i}, U_{2,i}, V_{1,i}, V_{2,i}; \widetilde{X}_{1,i}, \widetilde{X}_{2,i} | Y_i) \Big] \\
&\overset{(g)}{=} \sum_{i=1}^{n} \Big[I(V_{1,i}, V_{2,i}; Y_i) - I(V_{1,i}, V_{2,i}; Z_i) - 2R_0 \\
&\qquad\qquad + I(U_{1,i}, U_{2,i}, V_{1,i}, V_{2,i}; \widetilde{X}_{1,i}, \widetilde{X}_{2,i}) \\
&\qquad\qquad - I(U_{1,i}, U_{2,i}, V_{1,i}, V_{2,i}; Y_i) \Big] \\
&= \sum_{i=1}^{n} \Big[- I(U_{1,i}, U_{2,i}, V_{1,i}, V_{2,i}; Z_i) - 2R_0 \\
&\qquad\qquad + I(U_{1,i} U_{2,i}, V_{1,i}, V_{2,i}; \widetilde{X}_{1,i}, \widetilde{X}_{2,i}) \\
&\qquad\qquad + I(U_{1,i}, U_{2,i}; Z_i | V_{1,i}, V_{2,i}) \\
&\qquad\qquad - I(U_{1,i}, U_{2,i}; Y_i | V_{1,i}, V_{2,i}) \Big] \\
&\overset{(h)}{\geq} \sum_{i=1}^{n} \Bigg[I(U_{1,i}, U_{2,i}; \widetilde{X}_{1,i}, \widetilde{X}_{2,i} | Z_i) - 2R_0 \\
&\qquad + \Big[I(U_{1,i}, U_{2,i}; Z_i | V_{1,i}, V_{2,i}) \\
&\qquad\qquad - I(U_{1,i}, U_{2,i}; Y_i | V_{1,i}, V_{2,i}) \Big]^{-} \Bigg] \qquad (84)
\end{aligned}
$$

where (a) follows by (1) and from the Markov chain condition $(W_1, W_2) - (\widetilde{X}_1^n, \widetilde{X}_2^n) - Z^n$, (b) follows because K is independent of $(\widetilde{X}_1^n, \widetilde{X}_2^n, Y^n)$, (c) follows from the Markov chain condition

$$(W_1, W_2, K) - (\widetilde{X}_1^n, \widetilde{X}_2^n) - Y^n \qquad (85)$$

(d) follows from the Csiszár's sum identity and because (75) form a Markov chain, (e) follows because $(\widetilde{X}_1^n, \widetilde{X}_2^n, Y^n, Z^n)$ are i.i.d., K is independent of $(\widetilde{X}_1^n, \widetilde{X}_2^n, Y^n)$, and from the data processing inequality applied to the Markov chain

condition in (76), (f) follows from the definitions of $V_{j,i}$ and $U_{j,i}$, (g) follows from the Markov chain condition in (77), and (h) follows from the Markov chain condition

$$(V_{1,i},V_{2,i})-(U_{1,i},U_{2,i})-(\widetilde{X}_{1,i},\widetilde{X}_{2,i})-Z_i. \tag{86}$$

Now, we consider the case $R_0 \geq \overline{R}_0$ and prove the corresponding outer bound for the secrecy-leakage rate in (65). We obtain

$$\begin{aligned}
n(R_\ell+\delta_n) &\overset{(a)}{\geq} H(\widetilde{X}_1^n,\widetilde{X}_2^n|Z^n)-H(\widetilde{X}_1^n,\widetilde{X}_2^n|Z^n,W_1,W_2)\\
&\overset{(b)}{=} \sum_{i=1}^n \Big[H(\widetilde{X}_{1,i},\widetilde{X}_{2,i}|Z_i)\\
&\qquad -H(\widetilde{X}_{1,i},\widetilde{X}_{2,i}|Z^i,W_1,W_2,\widetilde{X}_{1,i+1}^n,\widetilde{X}_{2,i+1}^n,Y_{i+1}^n)\Big]\\
&\overset{(c)}{=} \sum_{i=1}^n \Big[H(\widetilde{X}_{1,i},\widetilde{X}_{2,i}|Z_i)\\
&\qquad -H(\widetilde{X}_{1,i},\widetilde{X}_{2,i}|Z_i,V_{1,i},V_{2,i},\widetilde{X}_{1,i+1}^n,\widetilde{X}_{2,i+1}^n)]\\
&\geq \sum_{i=1}^n I(V_{1,i},V_{2,i};\widetilde{X}_{1,i},\widetilde{X}_{2,i}|Z_i)
\end{aligned} \tag{87}$$

where (a) follows by (1), (b) follows because $(\widetilde{X}_1^n,\widetilde{X}_2^n,Z^n)$ are i.i.d. and from the Markov chain condition

$$(\widetilde{X}_{1,i},\widetilde{X}_{2,i})-(\widetilde{X}_{1,i+1}^n,\widetilde{X}_{2,i+1}^n,W_1,W_2,Z^i)-(Z_{i+1}^n,Y_{i+1}^n) \tag{88}$$

and (c) follows from the definitions of $V_{1,i}$ and $V_{2,i}$. Moreover, the corresponding outer bound for the case $R_0 \geq \overline{\overline{R}}_0$ follows because conditional mutual information is non-negative.

Introduce a uniformly distributed time-sharing random variable $Q \sim \text{Unif}[1:n]$ independent of other random variables, and define $X=X_Q$, $\widetilde{X}_j=\widetilde{X}_{j,Q}$, $Y=Y_Q$, $Z=Z_Q$, $V_j=V_{j,Q}$, and $U_j=(U_{j,Q},Q)$, so

$$(Q,V_j)-U_j-\widetilde{X}_j-X-(\widetilde{X}_{j'},Y,Z) \tag{89}$$

form Markov chains for $j=1,2$. The proof of the outer bound follows by letting $\delta_n \to 0$ and using the support lemma [38, Lemma 15.4] for the cardinality bounds. ∎

VI. Proofs of Lemmas 3 and 4

Proof Sketches: The proof of the inner bound for the weak eavesdropper setting follows by using the same random binning steps used in the proof of Lemma 1, so the same communication rates are achieved.

Using the definitions given in (53) and (54), and applying similar steps to (55) we have the secrecy leakage

$$\begin{aligned}
&I(\widetilde{X}_1^n,\widetilde{X}_2^n;W_j,F_j,F_{j'}|Z^n)\\
&=I(\widetilde{X}_1^n,\widetilde{X}_2^n;\overline{W}_j,F_j|Z^n)\\
&\qquad +I(\widetilde{X}_1^n,\widetilde{X}_2^n;F_{j'}|\overline{W}_j,F_j,Z^n)-nR_0\\
&\overset{(a)}{\leq} H(\overline{W}_j,F_j|Z^n)-H(\overline{W}_j,F_j|\widetilde{X}_j^n)+\epsilon_n'-nR_0\\
&\overset{(b)}{=} H(\overline{W}_j,F_j|Z^n)-H(U_j^n,V_j^n|\widetilde{X}_j^n)+H(V_j^n|\overline{W}_j,F_j,\widetilde{X}_j^n)\\
&\qquad +H(U_j^n|\overline{W}_j,F_j,\widetilde{X}_j^n,V_j^n)+\epsilon_n'-nR_0\\
&\overset{(c)}{\leq} H(\overline{W}_j,F_j|Z^n)-nH(U_j,V_j|\widetilde{X}_j)+2n\epsilon_n'+\epsilon_n'-nR_0
\end{aligned} \tag{90}$$

where (a) follows for some $\epsilon_n'>0$ such that $\epsilon_n' \to 0$ when $n\to\infty$ since $F_{j'}$ is almost independent of $(\widetilde{X}_1^n,\widetilde{X}_2^n,Z^n)$ by (33)-(36) and because

$$F_{j'}-(\widetilde{X}_1^n,\widetilde{X}_2^n,Z^n)-(F_j,\overline{W}_j) \tag{91}$$

$$(\overline{W}_j,F_j)-\widetilde{X}_j^n-(Z^n,\widetilde{X}_{j'}^n) \tag{92}$$

form Markov chains, (b) follows since (U_j^n,V_j^n) determine $(\overline{W}_j,F_j)$, and (c) follows because $(\overline{W}_j,F_j,\widetilde{X}_j^n)$ can recover V_j^n by (37)-(38) and $(V_j^n,\overline{W}_j,F_j,\widetilde{X}_j^n)$ can recover U_j^n by (39)-(40), respectively, which follow since we have

$$\begin{aligned}
H(V_j|\widetilde{X}_j) &\overset{(c.1)}{=} H(V_j|\widetilde{X}_j,V_{j'},Y)\\
&\leq H(V_j|V_{j'},Y)\leq H(V_j|Y)
\end{aligned} \tag{93}$$

and

$$\begin{aligned}
H(U_j|V_j,\widetilde{X}_j) &\overset{(c.2)}{=} H(U_j|V_j,\widetilde{X}_j,V_{j'},U_{j'},Y)\\
&\leq H(U_j|U_{j'},V_j,V_{j'},Y)\leq H(U_j|V_j,V_{j'},Y)
\end{aligned} \tag{94}$$

where $(c.1)$ follows from the Markov chain condition $V_j-\widetilde{X}_j-(V_{j'},Y)$ and $(c.2)$ follows from the Markov chain condition $V_j-U_j-\widetilde{X}_j-(V_{j'},U_{j'},Y)$.

Next, we provide a single letter upper bound on the term $H(\overline{W}_j,F_j|Z^n)$ in (90) by applying the results in [3, Section V - A] that consider six different decodability cases, which are applied also to (63). However, for the weak eavesdropper setting, the conditional entropy term measures only the j-th encoder's leakage, unlike in (63) where the leakage is with respect to both encoders. Therefore, we have to adapt our analysis to the decoding order imposed. If $j=1$, we should then consider a decoder that observes not only Y^n but (Y^n,V_2^n) to measure the leakage about (U_1^n,V_1^n). Furthermore, if $j=2$, we should then consider a decoder that observes (Y^n,V_1^n,U_1^n) to measure the leakage about (U_2^n,V_2^n). Since the information leakage cannot decrease when the decoder obtains less information, we can apply the results in [3, Section V-A], as being applied in [26, Eq. (69)] to the secure and private source coding model with one encoder, by replacing Y with $(Y,V_{j'})$, so by (90) we obtain

$$\begin{aligned}
&I(\widetilde{X}_1^n,\widetilde{X}_2^n;W_j,F_j,F_{j'}|Z^n)\\
&\leq n\Big([I(U_j;Z|V_j)-I(U_j;Y,V_{j'}|V_j)+\epsilon]^-\\
&\qquad +I(U_j;\widetilde{X}_j|Z)-R_0+3\epsilon_n'\Big)+\epsilon_n'.
\end{aligned} \tag{95}$$

By replacing $(\widetilde{X}_1^n,\widetilde{X}_2^n)$ with X^n and applying entirely similar steps, one can show that we have the privacy leakage

$$\begin{aligned}
&I(X^n;W_j,F_j,F_{j'}|Z^n)\\
&\leq n\Big([I(U_j;Z|V_j)-I(U_j;Y,V_{j'}|V_j)+\epsilon]^-\\
&\qquad +I(U_j;X|Z)-R_0+3\epsilon_n'\Big)+\epsilon_n'.
\end{aligned} \tag{96}$$

If the private key rate is such that $R_0 \geq (\max\{I(U_1;\widetilde{X}_1|V_1,V_2,Y),I(U_2;\widetilde{X}_2|V_2,U_1,Y)\} + 2\epsilon)$,

by using the definitions in (59) and (60) we then have the secrecy leakage

$$\begin{aligned}
&I(\widetilde{X}_1^n, \widetilde{X}_2^n; W_j, F_j, F_{j'}|Z^n)\\
&= I(\widetilde{X}_1^n, \widetilde{X}_2^n; \overline{\overline{W}}_j, F_j|Z^n) + I(\widetilde{X}_1^n, \widetilde{X}_2^n; F_{j'}|\overline{\overline{W}}_j, F_j, Z^n)\\
&\overset{(a)}{\leq} I(\widetilde{X}_j^n; \overline{\overline{W}}_j, F_j|Z^n) + \epsilon'_n\\
&\overset{(b)}{\leq} nH(\widetilde{X}_j|Z) - H(\widetilde{X}_j^n|Z^n, V_j^n, W_{v_j}, F_j) + \epsilon'_n\\
&\overset{(c)}{\leq} nI(V_j; \widetilde{X}_j|Z) + 2\epsilon'_n
\end{aligned} \tag{97}$$

where (a) follows because $F_{j'}$ is almost independent of $(\widetilde{X}_1^n, \widetilde{X}_2^n, Z^n)$ by (33)-(36) and since (91) and (92) form Markov chains after replacing $\overline{W}_j$ with $\overline{\overline{W}}_j$, (b) follows because $(W_{\mathrm{u}_j} + K_j)$ is independent of $(V_j^n, \widetilde{X}_1^n, \widetilde{X}_2^n, Z^n)$ due to the one-time padding with a uniform and independent private key, and (c) follows because V_j^n determines (W_{v_j}, F_{v_j}) and because by (34) and (36) F_{u_j} is almost independent of (V_j^n, X^n, Z^n). Similarly, by replacing $(\widetilde{X}_1^n, \widetilde{X}_2^n)$ with X^n, one can show that we then have the privacy leakage

$$I(X^n; W_j, F_j, F_{j'}|Z^n) \leq nI(V_j; X|Z) + 2\epsilon'_n. \tag{98}$$

Furthermore, for the high private key-rate regime, we can apply entirely similar steps as in (62) to show that negligible privacy and secrecy leakages are achieved.

The proof of the outer bound for the communication rates and distortion follows from the proof of Lemma 2. Furthermore, for the secrecy- and privacy-leakage outer bound terms, one can lower bound the term $-I(U_{j,i}; Y_i|V_{j,i})$ in [26, Eqs. (76)(h) and (81)(i)] by the term $-I(U_{j,i}; Y_i, V_{j',i}|V_{j,i})$, so we omit the proof. ∎

VII. Conclusion

The classic secure distributed source coding problems were extended by considering a remote source, where two noisy measurements of the remote source output are reconstructed at a decoder that observes (1) private keys shared with legitimate terminals; (2) correlated side information that is also a noisy measurement of the remote source output; and (3) public indices sent by the legitimate terminals through rate-limited communication links. We considered two passive attack scenarios, in which either a weak eavesdropper can choose which communication link to access or a strong eavesdropper can access both links. We derived inner and outer bounds for these problems when the eavesdropper observes also another noisy measurement of the remote source output, i.e., eavesdropper has correlated side information. The private key rate was shown to affect the secrecy- and privacy-leakage rate terms significantly, which are different for low, middle, and high key-rate regimes. Furthermore, we showed that even if a weak eavesdropper can access only a single communication link, the secrecy- and privacy-leakage rate terms depend on the joint probability distribution for the low private key-rate regime.

In future work, we will impose symmetry on the measurement channels to reduce sufficient limits on the cardinalities of the auxiliary random variables, which might allow to compute the boundary tuples.

References

[1] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, July 1973.

[2] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, Mar. 2001.

[3] O. Günlü, M. Bloch, and R. F. Schaefer, "Secure multi-function computation with private remote sources," Mar. 2022, [Online]. Available: arxiv.org/abs/2106.09485.

[4] V. Prabhakaran and K. Ramchandran, "On secure distributed source coding," in *Proc. IEEE Inf. Theory Workshop*, Tahoe City, CA, Sep. 2007, pp. 442–447.

[5] D. Gündüz, E. Erkip, and H. V. Poor, "Secure lossless compression with side information," in *Proc. IEEE Inf. Theory Workshop*, Porto, Portugal, May 2008, pp. 169–173.

[6] R. Tandon, S. Ulukus, and K. Ramchandran, "Secure source coding with a helper," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2178–2187, Apr. 2013.

[7] D. Gündüz, E. Erkip, and H. V. Poor, "Lossless compression with security constraints," in *Proc. IEEE Int. Symp. Inf. Theory*, Toronto, ON, Canada, July 2008, pp. 111–115.

[8] W. Luh and D. Kundur, "Distributed secret sharing for discrete memoryless networks," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 3, pp. 1–7, Sep. 2008.

[9] K. Kittichokechai, Y.-K. Chia, T. J. Oechtering, M. Skoglund, and T. Weissman, "Secure source coding with a public helper," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3930–3949, July 2016.

[10] S. Salimi, M. Salmasizadeh, and M. R. Aref, "Generalised secure distributed source coding with side information," *IET Commun.*, vol. 4, no. 18, pp. 2262–2272, Dec. 2010.

[11] F. Naghibi, S. Salimi, and M. Skoglund, "The CEO problem with secrecy constraints," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 6, pp. 1234–1249, June 2015.

[12] H. Yamamoto, "Coding theorems for Shannon's cipher system with correlated source outputs, and common information," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 85–95, Jan. 1994.

[13] H. Ghourchian, P. A. Stavrou, T. J. Oechtering, and M. Skoglund, "Secure source coding with side-information at decoder and shared key at encoder and decoder," in *Proc. IEEE Inf. Theory Workshop*, Kanazawa, Japan, Oct. 2021, pp. 1–6.

[14] A. C. Yao, "Protocols for secure computations," in *Proc. IEEE Symp. Foundations Comp. Sci.*, Chicago, IL, Nov. 1982, pp. 160–164.

[15] ——, "How to generate and exchange secrets," in *Proc. IEEE Symp. Foundations Comp. Sci.*, Toronto, ON, Canada, Oct. 1986, pp. 162–167.

[16] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 2733–2742, May 1993.

[17] 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, July 1993.

[18] M. Bloch *et al.*, "An overview of information-theoretic security and privacy: Metrics, limits and applications," *IEEE J. Sel. Areas Inf. Theory*, vol. 2, no. 1, pp. 5–22, Mar. 2021.

[19] O. Günlü and G. Kramer, "Privacy, secrecy, and storage with multiple noisy measurements of identifiers," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 11, pp. 2872–2883, Nov. 2018.

[20] H. Permuter and T. Weissman, "Source coding with a side information "Vending Machine"," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4530–4544, July 2011.

[21] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[22] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[23] O. Günlü, "Key agreement with physical unclonable functions and biometric identifiers," Ph.D. dissertation, TU Munich, Germany, Nov. 2018, published by Dr.-Hut Verlag in Feb. 2019.

[24] 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.

[25] L. Lai, S.-W. Ho, and H. V. Poor, "Privacy-security trade-offs in biometric security systems - Part I: Single use case," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 1, pp. 122–139, Mar. 2011.

[26] O. Günlü, R. F. Schaefer, H. Boche, and H. V. Poor, "Secure and private source coding with private key and decoder side information," Aug. 2022, [Online]. Available: arxiv.org/abs/2205.05068.

[27] O. Günlü, "Multi-entity and multi-enrollment key agreement with correlated noise," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1190–1202, 2021.

[28] L. Kusters and F. M. J. Willems, "Multiple observations for secret-key binding with SRAM PUFs," *Entropy*, vol. 23, no. 5, May 2021.

[29] L. Lai, S. W. Ho, and H. V. Poor, "Privacy-security trade-offs in biometric security systems - Part II: Multiple use case," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 1, pp. 140–151, Mar. 2011.

[30] L. Kusters, O. Günlü, and F. M. Willems, "Zero secrecy leakage for multiple enrollments of physical unclonable functions," in *Proc. Symp. Inf. Theory Sign. Process. Benelux*, Twente, Netherlands, May-June 2018, pp. 119–127.

[31] O. Günlü, "Function computation under privacy, secrecy, distortion, and communication constraints," *Entropy*, vol. 24, no. 1, June 2022.

[32] W. Tu and L. Lai, "On function computation with privacy and secrecy constraints," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6716–6733, Oct. 2019.

[33] M. Sefidgaran and A. Tchamkerten, "Computing a function of correlated sources: A rate region," in *IEEE Int. Symp. Inf. Theory*, St. Petersburg, Russia, July-Aug. 2011, pp. 1856–1860.

[34] J. M. Renes and R. Renner, "Noisy channel coding via privacy amplification and information reconciliation," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7377–7385, Nov. 2011.

[35] M. H. Yassaee, M. R. Aref, and A. Gohari, "Achievability proof via output statistics of random binning," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6760–6786, Nov. 2014.

[36] M. Bloch, *Lecture Notes in Information-Theoretic Security*. Atlanta, GA: Georgia Inst. Technol., July 2018.

[37] M. Bloch and J. Barros, *Physical-layer Security*. Cambridge, U.K.: Cambridge University Press, 2011.

[38] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge, U.K.: Cambridge University Press, 2011.