# Secure and Private Source Coding with Private Key and Decoder Side Information

Onur Günlü[1], Rafael F. Schaefer[1], Holger Boche[2,3,4], and H. Vincent Poor[5]
[1]Chair of Communications Engineering and Security, University of Siegen, {onur.guenlue, rafael.schaefer}@uni-siegen.de
[2]Chair of Theoretical Information Technology, Technical University of Munich, boche@tum.de
[3]CASA: Cyber Security in the Age of Large-Scale Adversaries Exzellenzcluster, Ruhr-Universität Bochum
[4]BMBF Research Hub 6G-Life, Technical University of Munich
[5]Department of Electrical and Computer Engineering, Princeton University, poor@princeton.edu

***Abstract*—The problem of secure source coding with multiple terminals is extended by considering a remote source whose noisy measurements are the correlated random variables used for secure source reconstruction. The main additions to the problem include 1) all terminals noncausally observe a noisy measurement of the remote source; 2) a private key is available to all legitimate terminals; 3) the public communication link between the encoder and decoder is rate-limited; and 4) the secrecy leakage to the eavesdropper is measured with respect to the encoder input, whereas the privacy leakage is measured with respect to the remote source. Exact rate regions are characterized for a lossy source coding problem with a private key, remote source, and decoder side information under security, privacy, communication, and distortion constraints. By replacing the distortion constraint with a reliability constraint, we obtain the exact rate region also for the lossless case. Furthermore, the lossy rate region for scalar discrete-time Gaussian sources and measurement channels is established.**

## I. Introduction

Consider multiple terminals that observe correlated random sequences and wish to reconstruct these sequences at another terminal, called a decoder, by sending messages through noiseless communication links, i.e., the distributed source coding problem [1]. A sensor network, where each node observes a correlated random sequence that should be reconstructed at a distant node is a classic example for this problem [2, pp. 258]. Similarly, function computation problems in which a fusion center observes messages sent by other nodes to compute a function are closely related problems and can be used to model various recent applications [3], [4]. Since the messages sent over the communication links can be public, security constraints are imposed on these messages against an eavesdropper in the same network [5]. If all sent messages are available to the eavesdropper, then it is necessary to provide an advantage to the decoder over the eavesdropper to enable secure source coding. Providing side information, which is correlated with the sequences that should be reconstructed, to the decoder can provide such an advantage over the eavesdropper that can also have side information, as in [6]–[8]. Allowing the eavesdropper to access only a strict subset of all messages is also a method to enable secure distributed source coding, considered in [9]–[11]; see also [12] in which a similar method is applied to enable secure remote source reconstruction. Similarly, also a private key that is shared by legitimate terminals and hidden from the eavesdropper can provide such an advantage, as in [13], [14].

Source coding models in the literature commonly assume that dependent multi-letter random variables are available and should be compressed. For secret-key agreement [15], [16] and secure function computation problems [17], [18], which are instances of the source coding with side information problem [19, Section IV-B], the correlation between these multi-letter random variables is posited in [20], [21] to stem from an underlying ground truth that is a remote source such that its noisy measurements are these dependent random variables. Such a remote source allows to model the cause of correlation in a network, so we also posit 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 [22, pp. 78] and [23, Fig. 9]. Furthermore, in the chief executive officer (CEO) problem [24], there is a remote source whose noisy measurements are encoded such that a decoder can reconstruct the remote source by using the encoder outputs. Our model is different from the model in the CEO problem, since in our model the decoder aims to recover encoder observations rather than the remote source that is considered mainly to describe the cause of correlation between encoder observations. Thus, we define the *secrecy leakage* as the amount of information leaked to an eavesdropper about encoder observations. Since the remote source is common for all observations in the same network, we impose a *privacy leakage* constraint on the remote source because each encoder output observed by an eavesdropper leaks information about unused encoder observations, which might later cause secrecy leakage when the unused encoder observations are employed [25]–[27]; see [28]–[30] for joint secrecy and joint privacy constraints imposed due to multiple uses of the same source.

We characterize the rate region for a lossy secure and private source coding problem with one private key, remote source, encoder, decoder, eavesdropper, and eavesdropper and decoder side information. Requiring reliable source reconstruction, we characterize the rate region also for the lossless case. A Gaussian remote source and independent additive Gaussian noise measurement channels are considered to establish their lossy rate region under squared error distortion.

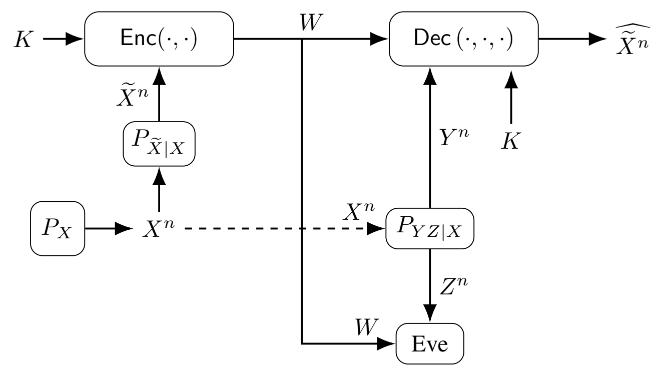


Fig. 1. Source coding with noisy measurements $(\widetilde{X}^n, Y^n)$ of a remote source $X^n$ and with a uniform private key $K$ under privacy, secrecy, communication, and distortion constraints.

## II. System Model

We consider the lossy source coding model with one encoder, one decoder, and an eavesdropper (Eve), depicted in Fig. 1. The encoder $\mathsf{Enc}(\cdot,\cdot)$ observes a noisy measurement $\widetilde{X}^n$ of an i.i.d. remote source $X^n \sim P_X^n$ through a memoryless channel $P_{\widetilde{X}|X}$ in addition to a private key $K \in [1:2^{nR_0}]$. The encoder output is an index $W$ that is sent over a link with limited communication rate. The decoder $\mathsf{Dec}(\cdot,\cdot,\cdot)$ observes the index $W$, as well as the private key $K$ and another noisy measurement $Y^n$ of the same remote source $X^n$ through another memoryless channel $P_{YZ|X}$ in order to estimate $\widetilde{X}^n$ as $\widehat{\widetilde{X}^n}$. The other noisy output $Z^n$ of $P_{YZ|X}$ is observed by Eve in addition to the index $W$. 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.

We next define the rate region for the lossy secure and private source coding problem defined above.

**Definition 1.** A *lossy* tuple $(R_\mathrm{w}, R_\mathrm{s}, R_\ell, D) \in \mathbb{R}^4_{\geq 0}$ is achievable, given a private key with rate $R_0 \geq 0$, if for any $\delta > 0$ there exist $n \geq 1$, an encoder, and a decoder such that

$$\log|\mathcal{W}| \leq n(R_\mathrm{w}+\delta) \quad \text{(storage)} \quad (1)$$
$$I(\widetilde{X}^n; W|Z^n) \leq n(R_\mathrm{s}+\delta) \quad \text{(secrecy)} \quad (2)$$
$$I(X^n; W|Z^n) \leq n(R_\ell+\delta) \quad \text{(privacy)} \quad (3)$$
$$\mathbb{E}\Big[d\Big(\widetilde{X}^n, \widehat{\widetilde{X}^n}(Y^n, W, K)\Big)\Big] \leq D+\delta \quad \text{(distortion)} \quad (4)$$

where $d(\widetilde{x}^n, \widehat{\widetilde{x}^n}) = \frac{1}{n}\sum_{i=1}^n d(\widetilde{x}_i, \widehat{\widetilde{x}}_i)$ is a per-letter bounded distortion metric. The *lossy* secure and private source coding region $\mathcal{R}_\mathrm{D}$ is the closure of the set of all achievable lossy tuples. ◊

Note that in (2) and (3) we consider conditional mutual information terms to take account of unavoidable privacy and secrecy leakages due to Eve's side information; see also [21], [31]. Furthermore, considering conditional mutual information terms rather than corresponding conditional entropy terms, the latter of which is used in [6], [14], [32]–[34], to characterize the secrecy and privacy leakages simplifies our analysis.

We next define the rate region for the lossless secure and private source coding problem.

**Definition 2.** A *lossless* tuple $(R_\mathrm{w}, R_\mathrm{s}, R_\ell) \in \mathbb{R}^3_{\geq 0}$ is achievable, given a private key with rate $R_0 \geq 0$, if for any $\delta > 0$ there exist $n \geq 1$, an encoder, and a decoder such that we have (1)-(3) and

$$\Pr\Big[\widetilde{X}^n \neq \widehat{\widetilde{X}^n}(Y^n, W, K)\Big] \leq \delta \quad \text{(reliability)}. \quad (5)$$

The *lossless* secure and private source coding region $\mathcal{R}$ is the closure of the set of all achievable lossless tuples. ◊

## III. Secure and Private Source Coding Regions

### A. Lossy Source Coding

The lossy secure and and private source coding region $\mathcal{R}_\mathrm{D}$ is characterized below; see Section V for its proof.

Define $[a]^- = \min\{a, 0\}$ for $a \in \mathbb{R}$ and denote

$$R' = [I(U;Z|V,Q) - I(U;Y|V,Q)]^-. \quad (6)$$

**Theorem 1.** *For given $P_X$, $P_{\widetilde{X}|X}$, $P_{YZ|X}$, and $R_0$, the region $\mathcal{R}_D$ is the set of all rate tuples $(R_w, R_s, R_\ell, D)$ satisfying*

$$R_w \geq I(U;\widetilde{X}|Y) \quad (7)$$

*and if $R_0 < I(U;\widetilde{X}|Y,V)$, then*

$$R_s \geq I(U;\widetilde{X}|Z) + R' - R_0 \quad (8)$$
$$R_\ell \geq I(U;X|Z) + R' - R_0 \quad (9)$$

*if $I(U;\widetilde{X}|Y,V) \leq R_0 < I(U;\widetilde{X}|Y)$, then*

$$R_s \geq I(V;\widetilde{X}|Z) \quad (10)$$
$$R_\ell \geq I(V;X|Z) \quad (11)$$

*if $R_0 \geq I(U;\widetilde{X}|Y)$, then*

$$R_s \geq 0 \quad (12)$$
$$R_\ell \geq 0 \quad (13)$$

*for some*

$$P_{QVU\widetilde{X}XYZ} = P_{Q|V}P_{V|U}P_{U|\widetilde{X}}P_{\widetilde{X}|X}P_XP_{YZ|X} \quad (14)$$

*such that $\mathbb{E}\big[d(\widetilde{X}, \widehat{\widetilde{X}}(U,Y))\big] \leq D$ for some reconstruction function $\widehat{\widetilde{X}}(U,Y)$. The region $\mathcal{R}_D$ is convexified by using the time-sharing random variable $Q$, required due to the $[\cdot]^-$ operation. One can limit the cardinalities to $|\mathcal{Q}| \leq 2$, $|\mathcal{V}| \leq |\widetilde{\mathcal{X}}|+3$, and $|\mathcal{U}| \leq (|\widetilde{\mathcal{X}}|+3)^2$.*

We remark that (12) and (13) show that one can simultaneously achieve *strong secrecy* and *strong privacy*, i.e., the conditional mutual information terms in (2) and (3), respectively, are negligible, by using a large private key $K$, which is a result missing in some recent works on secure source coding with private key.

### B. Lossless Source Coding

The lossless secure and and private source coding region $\mathcal{R}$ is characterized next; see below for a proof sketch.

Denote

$$R'' = [I(\widetilde{X};Z|V,Q) - I(\widetilde{X};Y|V,Q)]^{-}. \quad (15)$$

**Lemma 1.** *For given $P_X$, $P_{\widetilde{X}|X}$, $P_{YZ|X}$, and $R_0$, the region $\mathcal{R}$ is the set of all rate tuples $(R_w, R_s, R_\ell)$ satisfying*

$$R_w \geq H(\widetilde{X}|Y) \quad (16)$$

*and if $R_0 < H(\widetilde{X}|Y,V)$, then*

$$R_s \geq H(\widetilde{X}|Z) + R'' - R_0 \quad (17)$$

$$R_\ell \geq I(\widetilde{X};X|Z) + R'' - R_0 \quad (18)$$

*if $H(\widetilde{X}|Y,V) \leq R_0 < H(\widetilde{X}|Y)$, then*

$$R_s \geq I(V;\widetilde{X}|Z) \quad (19)$$

$$R_\ell \geq I(V;X|Z) \quad (20)$$

*if $R_0 \geq H(\widetilde{X}|Y)$, then*

$$R_s \geq 0 \quad (21)$$

$$R_\ell \geq 0 \quad (22)$$

*for some*

$$P_{QV\widetilde{X}XYZ} = P_{Q|V}P_{V|\widetilde{X}}P_{\widetilde{X}|X}P_X P_{YZ|X}. \quad (23)$$

*One can limit the cardinalities to $|\mathcal{Q}| \leq 2$ and $|\mathcal{V}| \leq |\widetilde{\mathcal{X}}| + 2$.*

*Proof Sketch:* The proof for the lossless region $\mathcal{R}$ follows from the proof for the lossy region $\mathcal{R}_{\text{D}}$, given in Theorem 1 above, by choosing $U = \widetilde{X}$ such that we have the reconstruction function $\widehat{\widetilde{X}}(\widetilde{X}, Y) = \widetilde{X}$, so we achieve $D = 0$. Thus, the reliability constraint in (5) is satisfied because $d(\cdot,\cdot)$ is a distortion metric. ■

## IV. Gaussian Sources and Channels

We evaluate the lossy rate region for a Gaussian example with squared error distortion by finding the optimal auxiliary random variable in the corresponding rate region. Consider a special lossy source coding case in which (i) there is no private key; (ii) the eavesdropper's channel observation $Z^n$ is less noisy than the decoder's channel observation $Y^n$ such that we obtain a lossy source coding region with a single auxiliary random variable that should be optimized.

We next define less noisy channels, considering $P_{YZ|X}$.

**Definition 3** ([35]). $Z$ (or eavesdropper) is *less noisy* than $Y$ (or decoder) if

$$I(L;Z) \geq I(L;Y) \quad (24)$$

holds for any random variable $L$ such that $L - X - (Y,Z)$ form a Markov chain. ◊

**Corollary 1.** *For given $P_X$, $P_{\widetilde{X}|X}$, $P_{YZ|X}$, and $R_0 = 0$, the region $\mathcal{R}_D$ when the eavesdropper is less noisy than the decoder is the set of all rate tuples $(R_w, R_s, R_\ell, D)$ satisfying*

$$R_w \geq I(U;\widetilde{X}|Y) = I(U;\widetilde{X}) - I(U;Y) \quad (25)$$

$$R_s \geq I(U;\widetilde{X}|Z) = I(U;\widetilde{X}) - I(U;Z) \quad (26)$$

$$R_\ell \geq I(U;X|Z) = I(U;X) - I(U;Z) \quad (27)$$

*for some*

$$P_{U\widetilde{X}XYZ} = P_{U|\widetilde{X}}P_{\widetilde{X}|X}P_X P_{YZ|X} \quad (28)$$

*such that $\mathbb{E}\big[d\big(\widetilde{X}, \widehat{\widetilde{X}}(U,Y)\big)\big] \leq D$ for some reconstruction function $\widehat{\widetilde{X}}(U,Y)$. One can limit the cardinality to $|\mathcal{U}|\leq |\widetilde{\mathcal{X}}|+3$.*

*Proof Sketch:* The proof for Corollary 1 follows from the proof for Theorem 1 by considering the bounds in (7)-(9) since $R_0 = 0$. Furthermore, $R'$ defined in (6) is 0 for the less noisy condition considered, which follows because $(Q,V) - U - X - (Y,Z)$ form a Markov chain. ■

Suppose the following scalar discrete-time Gaussian source and channel model for the lossy source coding problem depicted in Fig. 1

$$X = \rho_x \widetilde{X} + N_x \quad (29)$$

$$Y = \rho_y X + N_y \quad (30)$$

$$Z = \rho_z X + N_z \quad (31)$$

where we have the remote source $X \sim \mathcal{N}(0,1)$, fixed correlation coefficients $\rho_x, \rho_y, \rho_z \in (-1,1)$, and additive Gaussian noise random variables $N_x \sim \mathcal{N}(0, 1-\rho_x^2)$, $N_y \sim \mathcal{N}(0, 1-\rho_y^2)$, $N_z \sim \mathcal{N}(0, 1-\rho_z^2)$ such that $(\widetilde{X}, N_x, N_y, N_z)$ are mutually independent, and we consider the squared error distortion, i.e., $d(\widetilde{x}, \widehat{\widetilde{x}}) = (\widetilde{x} - \widehat{\widetilde{x}})^2$. We remark that (29) is an inverse measurement channel $P_{X|\widetilde{X}}$ that is a weighted sum of two independent Gaussian random variables, imposed to be able to apply the conditional entropy power inequality (EPI) [36, Lemma II]; see [20, Theorem 3] and [37, Section V] for binary symmetric inverse channel assumptions imposed to apply Mrs. Gerber's lemma [38]. Suppose $|\rho_z| > |\rho_y|$ such that $Y$ is stochastically-degraded than $Z$ since then there exists a random variable $\widetilde{Y}$ such that $P_{\widetilde{Y}|X} = P_{Y|X}$ and $P_{\widetilde{Y}Z|X} = P_{Z|X}P_{\widetilde{Y}|Z}$ [39, Lemma 6], so $Z$ is also less noisy than $Y$ since less noisy channels constitute a strict superset of the set of stochastically-degraded channels and both channel sets consider only the conditional marginal probability distributions [2, pp. 121].

We next take the liberty to use the lossy rate region in Corollary 1, characterized for discrete memoryless channels, for the model in (29)-(31). This is common in the literature since there is a discretization procedure to extend the achievability proof to well-behaved continuous-alphabet random variables and the converse proof applies to arbitrary random variables; see [2, Remark 3.8]. For Gaussian sources and channels, we use differential entropy and eliminate the cardinality bound on the auxiliary random variable. The lossy source coding region for the model in (29)-(31) without a private key is given below.

**Lemma 2.** *For the model in (29)-(31) such that $|\rho_z| > |\rho_y|$ and $R_0 = 0$, the region $\mathcal{R}_D$ with squared error distortion is the set of all rate tuples $(R_w, R_s, R_\ell, D)$ satisfying, for $0 < \alpha \leq 1$,*

$$R_w \geq \frac{1}{2}\log\Big(\frac{1-\rho_x^2\rho_y^2(1-\alpha)}{\alpha}\Big) \tag{32}$$

$$R_s \geq \frac{1}{2}\log\Big(\frac{1-\rho_x^2\rho_z^2(1-\alpha)}{\alpha}\Big) \tag{33}$$

$$R_\ell \geq \frac{1}{2}\log\Big(\frac{1-\rho_x^2\rho_z^2(1-\alpha)}{1-\rho_x^2(1-\alpha)}\Big) \tag{34}$$

$$D \geq \frac{\alpha(1-\rho_x^2\rho_y^2)}{1-\rho_x^2\rho_y^2(1-\alpha)}. \tag{35}$$

*Proof Sketch:* For the achievability proof, let $U \sim \mathcal{N}(0, 1-\alpha)$ and $\Theta \sim \mathcal{N}(0,\alpha)$, as in [40, Eq. (34)] and [41, Appendix B], be independent random variables for some $0 < \alpha \leq 1$ such that $\widetilde{X} = U + \Theta$ and $U - \widetilde{X} - X - (Y,Z)$ form a Markov chain. Choose the reconstruction function $\widehat{\widetilde{X}}(U,Y)$ as the minimum mean square error (MMSE) estimator, and given any fixed $D > 0$ auxiliary random variables are chosen such that the distortion constraint is satisfied. We then have for the squared error distortion

$$D = \mathbb{E}\Big[\big(\widetilde{X} - \widehat{\widetilde{X}}(U,Y)\big)^2\Big] \overset{(a)}{=} \frac{1}{2\pi e}e^{2h(\widetilde{X}|U,Y)} \tag{36}$$

where equality in $(a)$ is achieved because $\widetilde{X}$ is Gaussian and the reconstruction function is the MMSE estimator [42, Theorem 8.6.6]. Define the covariance matrix of the vector random variable $[\widetilde{X}, U, Y]$ as $\mathbf{K}_{\widetilde{X}UY}$ and of $[U,Y]$ as $\mathbf{K}_{UY}$, respectively. We then have

$$\begin{aligned} h(\widetilde{X}|U,Y) &= h(\widetilde{X},U,Y) - h(U,Y)\\ &= \frac{1}{2}\log\left(2\pi e\frac{\det(\mathbf{K}_{\widetilde{X}UY})}{\det(\mathbf{K}_{UY})}\right) \end{aligned} \tag{37}$$

where $\det(\cdot)$ is the determinant of a matrix; see also [12, Section F]. Combining (36) and (37), and calculating the determinants, we obtain

$$D = \frac{\alpha(1-\rho_x^2\rho_y^2)}{1-\rho_x^2\rho_y^2(1-\alpha)}. \tag{38}$$

One can also show that

$$I(U;\widetilde{X}) = h(\widetilde{X}) - h(\widetilde{X}|U) = \frac{1}{2}\log\Big(\frac{1}{\alpha}\Big) \tag{39}$$

$$I(U;X) = h(X) - h(X|U) = \frac{1}{2}\log\Big(\frac{1}{1-\rho_x^2(1-\alpha)}\Big) \tag{40}$$

$$I(U;Y) = h(Y) - h(Y|U) = \frac{1}{2}\log\Big(\frac{1}{1-\rho_x^2\rho_y^2(1-\alpha)}\Big) \tag{41}$$

$$I(U;Z) = h(Z) - h(Z|U) = \frac{1}{2}\log\Big(\frac{1}{1-\rho_x^2\rho_z^2(1-\alpha)}\Big). \tag{42}$$

Thus, by calculating (25)-(27), the achievability proof follows.

For the converse proof, one can first show that

$$I(U;\widetilde{X}) - I(U;Y) = h(Y|U) - h(\widetilde{X}|U) \tag{43}$$

$$I(U;\widetilde{X}) - I(U;Z) = h(Z|U) - h(\widetilde{X}|U) \tag{44}$$

$$I(U;X) - I(U;Z) = h(Z|U) - h(X|U) \tag{45}$$

which follow since $h(\widetilde{X}) = h(X) = h(Y) = h(Z)$. Suppose

$$h(\widetilde{X}|U) = \frac{1}{2}\log(2\pi e\alpha) \tag{46}$$

for any $0 < \alpha \leq 1$ that represents the unique variance of a Gaussian random variable; see [20, Lemma 2] for a similar result applied to binary random variables. Thus, by applying the conditional EPI, we obtain

$$\begin{aligned} e^{2h(Y|U)} &\overset{(a)}{=} e^{2h(\rho_x\rho_y\widetilde{X}|U)} + e^{2h(\rho_y N_x + N_y)}\\ &= 2\pi e\big(\rho_x^2\rho_y^2\alpha + \rho_y^2(1-\rho_x^2) + 1 - \rho_y^2\big)\\ &= 2\pi e\big(1 - \rho_x^2\rho_y^2(1-\alpha)\big) \end{aligned} \tag{47}$$

where $(a)$ follows because $U - \widetilde{X} - (N_x, N_y)$ form a Markov chain and $(N_x, N_y)$ are independent of $\widetilde{X}$, so $(N_x, N_y)$ are independent of $U$, and equality is satisfied since, given $U$, $\rho_x\rho_y\widetilde{X}$ and $(\rho_y N_x + N_y)$ are conditionally independent and they are Gaussian random variables, as imposed in (46) above; see [20, Lemma 1 and Eq. (28)] for a similar result applied to binary random variables by extending Mrs. Gerber's lemma. Similarly, we have

$$e^{2h(Z|U)} = 2\pi e\big(1 - \rho_x^2\rho_z^2(1-\alpha)\big) \tag{48}$$

which follows by replacing $(Y, \rho_y, N_y)$ with $(Z, \rho_z, N_z)$ in (47), respectively, because the channel $P_{Y|U}$ can be mapped to $P_{Z|U}$ with these changes due to (29)-(31) and the Markov chain $U - \widetilde{X} - X - (Y,Z)$. Furthermore, we have

$$\begin{aligned} e^{2h(X|U)} &\overset{(a)}{=} e^{2h(\rho_x\widetilde{X}|U)} + e^{2h(N_x)}\\ &= 2\pi e\big(\rho_x^2\alpha + 1 - \rho_x^2\big)\\ &= 2\pi e\big(1 - \rho_x^2(1-\alpha)\big) \end{aligned} \tag{49}$$

where $(a)$ follows because $N_x$ is independent of $U$, and equality is achieved since, given $U$, $\rho_x\widetilde{X}$ and $N_x$ are conditionally independent and are Gaussian random variables. Therefore, by applying (43)-(49) to (25)-(27), the converse proof for (32)-(34) follows.

Next, consider

$$\begin{aligned} &h(\widetilde{X}|U,Y) = -I(U;\widetilde{X}|Y) + h(\widetilde{X}|Y)\\ &\overset{(a)}{=} -h(Y|U) + h(\widetilde{X}|U) + h(Y|\widetilde{X})\\ &\overset{(b)}{=} \frac{1}{2}\log\Big(\frac{\alpha}{1-\rho_x^2\rho_y^2(1-\alpha)}\Big) + h(\rho_x\rho_y\widetilde{X} + \rho_y N_x + N_y|\widetilde{X})\\ &\overset{(c)}{=} \frac{1}{2}\log\Big(\frac{\alpha}{1-\rho_x^2\rho_y^2(1-\alpha)}\Big) + h(\rho_y N_x + N_y)\\ &= \frac{1}{2}\log\Big(2\pi e\frac{\alpha(\rho_y^2(1-\rho_x^2) + (1-\rho_y^2))}{1-\rho_x^2\rho_y^2(1-\alpha)}\Big)\\ &= \frac{1}{2}\log\Big(2\pi e\frac{\alpha(1-\rho_x^2\rho_y^2)}{1-\rho_x^2\rho_y^2(1-\alpha)}\Big) \end{aligned} \tag{50}$$

where $(a)$ follows by (25) and (43), and since $h(Y) = h(\widetilde{X})$, $(b)$ follows by (46) and (47), and $(c)$ follows because $(N_x, N_y)$ are independent of $\widetilde{X}$. Furthermore, for any random variable

$\widetilde{X}$ and reconstruction function $\widehat{\widetilde{X}}(U,Y)$, we have [42, Theorem 8.6.6]

$$\mathbb{E}\Big[\big(\widetilde{X}-\widehat{\widetilde{X}}(U,Y)\big)^2\Big] \geq \frac{1}{2\pi e}e^{2h(\widetilde{X}|U,Y)}. \qquad (51)$$

Combining the distortion constraint given in Corollary 1 with (50) and (51), the converse proof for (35) follows. ∎

## V. Proof for Theorem 1

### A. Achievability Proof for Theorem 1

*Proof Sketch:* We leverage the output statistics of random binning (OSRB) method [16], [43], [44] for the achievability proof by following the steps described in [45, Section 1.6].

Let $(V^n,U^n,\widetilde{X}^n,X^n,Y^n,Z^n)$ be i.i.d. according to $P_{VU\widetilde{X}XYZ}$ that can be obtained from (14) by fixing $P_{U|\widetilde{X}}$ and $P_{V|U}$ such that $\mathbb{E}[d(\widetilde{X},\widehat{\widetilde{X}})]\leq(D+\epsilon)$ for any $\epsilon>0$. To each $v^n$ assign two random bin indices $F_{\text{v}}\in[1:2^{n\widetilde{R}_{\text{v}}}]$ and $W_{\text{v}}\in[1:2^{nR_{\text{v}}}]$. Furthermore, to each $u^n$ assign three random bin indices $F_{\text{u}}\in[1:2^{n\widetilde{R}_{\text{u}}}]$, $W_{\text{u}}\in[1:2^{nR_{\text{u}}}]$, and $K_{\text{u}}\in[1:\ 2^{nR_0}]$, where $R_0$ is the private key rate defined in Section II. The public indices $F=(F_{\text{v}},F_{\text{u}})$ represent the choice of a source encoder and decoder pair. Furthermore, we impose that the messages sent by the source encoder $\mathsf{Enc}(\cdot,\cdot)$ to the source decoder $\mathsf{Dec}(\cdot,\cdot,\cdot)$ are

$$W=(W_{\text{v}},W_{\text{u}},K+K_{\text{u}}) \qquad (52)$$

where the summation with the private key is in modulo- $2^{nR_0}$, i.e., one-time padding.

The public index $F_{\text{v}}$ is almost independent of $(\widetilde{X}^n,X^n,Y^n,Z^n)$ if we have [43, Theorem 1]

$$\widetilde{R}_{\text{v}}<H(V|\widetilde{X},X,Y,Z)\overset{(a)}{=}H(V|\widetilde{X}) \qquad (53)$$

where $(a)$ follows since $(X,Y,Z)-\widetilde{X}-V$ form a Markov chain. The constraint in (53) suggests that the expected value, taken over the random bin assignments, of the variational distance between the joint probability distributions $\text{Unif}[1:2^{n\widetilde{R}_{\text{v}}}]\cdot P_{\widetilde{X}^n}$ and $P_{F_{\text{v}}\widetilde{X}^n}$ vanishes when $n\to\infty$. Moreover, the public index $F_{\text{u}}$ is almost independent of $(V^n,\widetilde{X}^n,X^n,Y^n,Z^n)$ if we have

$$\widetilde{R}_{\text{u}}<H(U|V,\widetilde{X},X,Y,Z)\overset{(a)}{=}H(U|V,\widetilde{X}) \qquad (54)$$

where $(a)$ follows from the Markov chain $(X,Y,Z)-\widetilde{X}-(U,V)$.

Using a Slepian-Wolf (SW) [1] decoder that observes $(Y^n,F_{\text{v}},W_{\text{v}})$, one can reliably estimate $V^n$ if we have [43, Lemma 1]

$$\widetilde{R}_{\text{v}}+R_{\text{v}}>H(V|Y) \qquad (55)$$

since then the expected error probability, taken over random bin assignments, vanishes when $n\to\infty$. Furthermore, one can reliably estimate $U^n$ by using a SW decoder that observes $(K,V^n,Y^n,F_{\text{u}},W_{\text{u}},K+K_{\text{u}})$ if we have

$$R_0+\widetilde{R}_{\text{u}}+R_{\text{u}}>H(U|V,Y). \qquad (56)$$

To satisfy (53)-(56), for any $\epsilon>0$ we fix

$$\widetilde{R}_{\text{v}}=H(V|\widetilde{X})-\epsilon \qquad (57)$$
$$R_{\text{v}}=I(V;\widetilde{X})-I(V;Y)+2\epsilon \qquad (58)$$
$$\widetilde{R}_{\text{u}}=H(U|V,\widetilde{X})-\epsilon \qquad (59)$$
$$R_0+R_{\text{u}}=I(U;\widetilde{X}|V)-I(U;Y|V)+2\epsilon. \qquad (60)$$

Since all tuples $(v^n,u^n,\widetilde{x}^n,x^n,y^n,z^n)$ are in the jointly typical set with high probability, by the typical average lemma [2, pp. 26], the distortion constraint (4) is satisfied.

**Communication Rate**: (58) and (60) result in a communication (storage) rate of

$$R_{\text{w}}=R_0+R_{\text{v}}+R_{\text{u}}\overset{(a)}{=}I(U;\widetilde{X}|Y)+4\epsilon \qquad (61)$$

where $(a)$ follows since $V-U-\widetilde{X}-Y$ form a Markov chain.

**Privacy Leakage Rate**: Since the private key $K$ is uniformly distributed and is independent of source and channel random variables, we can consider the following virtual scenario to calculate the leakage. We first assume for the virtual scenario that there is no private key such that the encoder output for the virtual scenario is

$$\overline{W}=(W_{\text{v}},W_{\text{u}},K_{\text{u}}). \qquad (62)$$

We calculate the leakage for the virtual scenario. Then, given the mentioned properties of the private key and due to the one-time padding step in (52), we can subtract $H(K)=nR_0$ from the leakage calculated for the virtual scenario to obtain the leakage for the original problem, which follows from the sum of (59) and (60) if $\epsilon\to0$ when $n\to\infty$. Thus, we have the privacy leakage

$$\begin{aligned}&I(X^n;W,F|Z^n)=I(X^n;\overline{W},F|Z^n)-nR_0\\&\overset{(a)}{=}H(\overline{W},F|Z^n)-H(\overline{W},F|X^n)-nR_0\\&\overset{(b)}{=}H(\overline{W},F|Z^n)-H(U^n,V^n|X^n)\\&\qquad+H(V^n|\overline{W},F,X^n)+H(U^n|V^n,\overline{W},F,X^n)-nR_0\\&\overset{(c)}{\leq}H(\overline{W},F|Z^n)-nH(U,V|X)+2n\epsilon_n-nR_0\end{aligned} \qquad (63)$$

where $(a)$ follows because $(\overline{W},F)-X^n-Z^n$ form a Markov chain, $(b)$ follows since $(U^n,V^n)$ determine $(F_{\text{u}},W_{\text{u}},K_{\text{u}},F_{\text{v}},W_{\text{v}})$, and $(c)$ follows since $(U^n,V^n,X^n)$ is i.i.d. and for some $\epsilon_n>0$ such that $\epsilon_n\to0$ when $n\to\infty$ because $(F_{\text{v}},W_{\text{v}},X^n)$ can reliably recover $V^n$ by (55) because of the Markov chain $V^n-X^n-Y^n$ and, similarly, $(F_{\text{u}},W_{\text{u}},K_{\text{u}},V^n,X^n)$ can reliably recover $U^n$ by (56) because of $H(U|V,Y)\geq H(U|V,X)$ that is proved in [21, Eq. (55)] for the Markov chain $(V,U)-X-Y$.

Next, we consider the term $H(\overline{W},F|Z^n)$ in (63) and provide single letter bounds on it by applying the six different decodability results given in [21, Section V-A] that are applied to an entirely similar conditional entropy term in [21, Eq. (54)] that measures the uncertainty in indices conditioned on an i.i.d. multi-letter random variable. Thus, combining the six decodability results in [21, Section V-A] with (63) we obtain

$$I(X^n;W,F|Z^n) \leq n\big([I(U;Z|V) - I(U;Y|V) + \epsilon]^- + I(U;X|Z) + 3\epsilon_n - R_0\big). \quad (64)$$

We remark that (60) implicitly assumes that the private key rate $R_0$ is less than $(I(U;\widetilde{X}|V) - I(U;Y|V) + 2\epsilon) = (I(U;\widetilde{X}|Y,V) + 2\epsilon)$, where the equality follows from the Markov chain $(V,U) - \widetilde{X} - Y$. The communication rate results are not affected by this assumption since $\widetilde{X}^n$ should be reconstructed by the decoder. However, if the private key rate $R_0$ is greater than or equal to $(I(U;\widetilde{X}|Y,V) + 2\epsilon)$, then we can remove the bin index $K_{\text{u}}$ from the code construction above and apply one-time padding to the bin index $W_{\text{u}}$ such that we have the encoder output

$$\overline{\overline{\overline{W}}} = (W_{\text{v}}, W_{\text{u}} + K) \quad (65)$$

where the summation with the private key is in modulo-$2^{nR_{\text{u}}} = 2^{n(I(U;\widetilde{X}|Y,V)+2\epsilon)}$. Thus, one then does not leak any information about $W_{\text{u}}$ to the eavesdropper because of the one-time padding step in (65). We then have the privacy leakage

$$\begin{aligned} I(X^n;\overline{\overline{\overline{W}}},F|Z^n) &= I(X^n;W_{\text{v}},F|Z^n)\\ &\overset{(a)}{\leq} H(X^n|Z^n) - H(X^n|Z^n,W_{\text{v}},F_{\text{v}}) + \epsilon_n'\\ &\overset{(b)}{\leq} H(X^n|Z^n) - H(X^n|Z^n,V^n) + \epsilon_n'\\ &\overset{(c)}{=} nI(V;X|Z) + \epsilon_n' \end{aligned} \quad (66)$$

where $(a)$ follows for some $\epsilon_n'$ such that $\epsilon_n' \to 0$ when $n \to \infty$ since by (54) $F_{\text{u}}$ is almost independent of $(V^n, X^n, Z^n)$; see also [46, Theorem 1], $(b)$ follows since $V^n$ determines $(F_{\text{v}}, W_{\text{v}})$, and $(c)$ follows because $(X^n, Z^n, V^n)$ are i.i.d.

Note that we can reduce the privacy leakage given in (66) if $R_0 \geq (I(U;\widetilde{X}) - I(U;Y) + 4\epsilon) = (I(U;\widetilde{X}|Y) + 4\epsilon)$, where the equality follows from the Markov chain $U - \widetilde{X} - Y$, since then we can apply one-time padding to both bin indices $W_{\text{v}}$ and $W_{\text{u}}$ with the sum rate

$$\begin{aligned} &R_{\text{v}} + R_{\text{u}}\\ &\overset{(a)}{=} I(V;\widetilde{X}) - I(V;Y) + 2\epsilon + I(U;\widetilde{X}|V) - I(U;Y|V) + 2\epsilon\\ &\overset{(b)}{=} I(U;\widetilde{X}) - I(U;Y) + 4\epsilon \end{aligned} \quad (67)$$

where $(a)$ follows by (58) and (60), and $(b)$ follows from the Markov chain $V - U - \widetilde{X} - Y$. Thus, one then does not leak any information about $(W_{\text{v}}, W_{\text{u}})$ to the eavesdropper because of the one-time padding step, so we then obtain the privacy leakage of

$$\begin{aligned} I(X^n;F|Z^n) &= I(X^n;F_{\text{v}}|Z^n) + I(X^n;F_{\text{u}}|Z^n,F_{\text{v}})\\ &\overset{(a)}{\leq} 2\epsilon_n' \end{aligned} \quad (68)$$

where $(a)$ follows since by (53) $F_{\text{v}}$ is almost independent of $(X^n, Z^n)$ and by (54) $F_{\text{u}}$ is almost independent of $(V^n, X^n, Z^n)$.

**Secrecy Leakage Rate**: Similar to the privacy leakage analysis above, we first consider the virtual scenario with the encoder output given in (62), and then calculate the leakage for the original problem by subtracting $H(K) = nR_0$ from the leakage calculated for the virtual scenario. Thus, we obtain

$$\begin{aligned} &I(\widetilde{X}^n;W,F|Z^n) = I(\widetilde{X}^n;\overline{W},F|Z^n) - nR_0\\ &\overset{(a)}{=} H(\overline{W},F|Z^n) - H(\overline{W},F|\widetilde{X}^n) - nR_0\\ &\overset{(b)}{=} H(\overline{W},F|Z^n) - H(U^n,V^n|\widetilde{X}^n)\\ &\qquad + H(V^n|\overline{W},F,\widetilde{X}^n) + H(U^n|V^n,\overline{W},F,\widetilde{X}^n)\\ &\overset{(c)}{\leq} H(\overline{W},F|Z^n) - nH(U,V|\widetilde{X}) + 2n\epsilon_n' - nR_0\\ &\overset{(d)}{\leq} n\big([I(U;Z|V) - I(U;Y|V) + \epsilon]^-\\ &\qquad + I(U;\widetilde{X}|Z) + 3\epsilon_n' - R_0\big) \end{aligned} \quad (69)$$

where $(a)$ follows from the Markov chain $(\overline{W},F) - \widetilde{X}^n - Z^n$, $(b)$ follows since $(U^n, V^n)$ determine $(\overline{W}, F)$, $(c)$ follows because $(V^n, U^n, \widetilde{X}^n)$ are i.i.d. and because $(F_{\text{v}}, W_{\text{v}}, \widetilde{X}^n)$ can reliably recover $V^n$ by (55) due to the Markov chain $V^n - \widetilde{X}^n - Y^n$ and, similarly, $(F_{\text{u}}, W_{\text{u}}, K_{\text{u}}, V^n, \widetilde{X}^n)$ can reliably recover $U^n$ by (56) due to $H(U|V,Y) \geq H(U|V,\widetilde{X})$ that can be proved as in [21, Eq. (55)] for the Markov chain $(V,U) - \widetilde{X} - Y$, and $(d)$ follows by applying the six decodability results in [21, Section V-A] that are applied to (63) with the final result in (64) by replacing $X$ with $\widetilde{X}$.

Similar to the privacy leakage analysis above if we have $R_0 \geq (I(U;\widetilde{X}|Y,V) + 2\epsilon)$, then we can eliminate $K_{\text{u}}$ and apply one-time padding as in (65) such that no information about $W_{\text{u}}$ is leaked to the eavesdropper and we have

$$\begin{aligned} I(\widetilde{X}^n;\overline{\overline{\overline{W}}},F|Z^n) &= I(\widetilde{X}^n;W_{\text{v}},F|Z^n)\\ &\overset{(a)}{\leq} H(\widetilde{X}^n|Z^n) - H(\widetilde{X}^n|Z^n,W_{\text{v}},F_{\text{v}}) + \epsilon_n'\\ &\overset{(b)}{\leq} H(\widetilde{X}^n|Z^n) - H(\widetilde{X}^n|Z^n,V^n) + \epsilon_n'\\ &\overset{(c)}{=} nI(V;\widetilde{X}|Z) + \epsilon_n' \end{aligned} \quad (70)$$

where $(a)$ follows because by (54) $F_{\text{u}}$ is almost independent of $(V^n, \widetilde{X}^n, Z^n)$, $(b)$ follows since $V^n$ determines $(F_{\text{v}}, W_{\text{v}})$, and $(c)$ follows because $(\widetilde{X}^n, Z^n, V^n)$ are i.i.d.

If $R_0 \geq (I(U;\widetilde{X}|Y) + 4\epsilon)$, we can apply one-time padding to hide $(W_{\text{v}}, W_{\text{u}})$, as in the privacy leakage analysis above. We then have the secrecy leakage of

$$\begin{aligned} I(\widetilde{X}^n;F|Z^n) &= I(\widetilde{X}^n;F_{\text{v}}|Z^n) + I(\widetilde{X}^n;F_{\text{u}}|Z^n,F_{\text{v}})\\ &\overset{(a)}{\leq} 2\epsilon_n' \end{aligned} \quad (71)$$

where $(a)$ follows since by (53) $F_{\text{v}}$ is almost independent of $(\widetilde{X}^n, Z^n)$ and by (54) $F_{\text{u}}$ is almost independent of $(V^n, \widetilde{X}^n, Z^n)$.

Suppose the public indices $F$ are generated uniformly at random, and the encoder generates $(V^n, U^n)$ according to $P_{V^nU^n|\widetilde{X}^nF_{\text{v}}F_{\text{u}}}$ that can be obtained from the proposed binning scheme above to compute the bins $W_{\text{v}}$ from $V^n$ and $W_{\text{u}}$

from $U^n$, respectively. Such a procedure results in a joint probability distribution almost equal to $P_{VU\widetilde{X}XYZ}$ fixed above [45, Section 1.6]. Note that the privacy and secrecy leakage metrics above are expectations over all possible public index realizations $F = f$. Therefore, using a time-sharing random variable $Q$ for convexification and applying the selection lemma [47, Lemma 2.2] to each decodability case separately, the achievability for Theorem 1 follows by choosing an $\epsilon > 0$ such that $\epsilon \to 0$ when $n \to \infty$. ■

### *B. Converse Proof for Theorem 1*

*Proof Sketch:* Assume that for some $\delta_n > 0$ and $n \geq 1$, there exist an encoder and a decoder such that (1)-(4) are satisfied for some tuple $(R_{\text{w}}, R_{\text{s}}, R_\ell, D)$ given a private key with rate $R_0$.

Define $V_i \triangleq (W, Y_{i+1}^n, Z^{i-1})$ and $U_i \triangleq (W, Y_{i+1}^n, Z^{i-1}, X^{i-1}, K)$ that satisfy the Markov chain $V_i - U_i - \widetilde{X}_i - X_i - (Y_i, Z_i)$ by definition of the source statistics. We have

$$\begin{aligned} D + \delta_n &\overset{(a)}{\geq} \mathbb{E}\Big[d\Big(\widetilde{X}^n, \widehat{\widetilde{X}^n}(Y^n, W, K)\Big)\Big] \\ &\overset{(b)}{\geq} \mathbb{E}\Big[d\Big(\widetilde{X}^n, \widehat{\widetilde{X}^n}(Y^n, W, K, X^{i-1}, Z^{i-1})\Big)\Big] \\ &\overset{(c)}{=} \mathbb{E}\Big[d\Big(\widetilde{X}^n, \widehat{\widetilde{X}^n}(Y_i^n, W, K, X^{i-1}, Z^{i-1})\Big)\Big] \\ &\overset{(d)}{=} \frac{1}{n}\sum_{i=1}^n \mathbb{E}\Big[d\Big(\widetilde{X}_i, \widehat{\widetilde{X}_i}(U_i, Y_i)\Big)\Big] \end{aligned} \tag{72}$$

where $(a)$ follows by (4), $(b)$ follows since providing more information to the reconstruction function does not increase expected distortion, $(c)$ follows from the Markov chain

$$Y^{i-1} - (Y_i^n, X^{i-1}, Z^{i-1}, W, K) - \widetilde{X}^n \tag{73}$$

and $(d)$ follows from the definition of $U_i$.

**Communication Rate**: For any $R_0 \geq 0$, we have

$$\begin{aligned} &n(R_{\text{w}} + \delta_n) \overset{(a)}{\geq} \log|\mathcal{W}| \\ &\geq H(W|Y_i^n, Z^{i-1}, K) - H(W|Y_i^n, Z^{i-1}, K, \widetilde{X}^n) \\ &= \sum_{i=1}^n I(W; \widetilde{X}_i|\widetilde{X}^{i-1}, Y_{i+1}^n, Z^{i-1}, K, Y_i) \\ &\overset{(b)}{=} \sum_{i=1}^n I(\widetilde{X}^{i-1}, Y_{i+1}^n, Z^{i-1}, K, W; \widetilde{X}_i|Y_i) \\ &\overset{(c)}{\geq} \sum_{i=1}^n I(X^{i-1}, Y_{i+1}^n, Z^{i-1}, K, W; \widetilde{X}_i|Y_i) \\ &\overset{(d)}{=} \sum_{i=1}^n I(U_i; \widetilde{X}_i|Y_i) \end{aligned} \tag{74}$$

where $(a)$ follows by (1), $(b)$ follows because $(\widetilde{X}_i, Y_i)$ are independent of $(\widetilde{X}^{i-1}, Y_{i+1}^n, Z^{i-1}, K)$, $(c)$ follows by applying the data processing inequality to the Markov chain

$$X^{i-1} - (\widetilde{X}^{i-1}, Y_i^n, Z^{i-1}, K, W) - \widetilde{X}_i \tag{75}$$

and $(d)$ follows from the definition of $U_i$.

**Privacy Leakage Rate**: We obtain

$$\begin{aligned} &n(R_\ell + \delta_n) \\ &\overset{(a)}{\geq} [I(W; Y^n) - I(W; Z^n)] + [I(W; X^n) - I(W; Y^n)] \\ &\overset{(b)}{=} [I(W; Y^n) - I(W; Z^n)] \\ &\qquad + I(W; X^n|K) - I(K; X^n|W) \\ &\qquad - I(W; Y^n|K) + I(K; Y^n|W) \\ &\overset{(c)}{=} [I(W; Y^n) - I(W; Z^n)] \\ &\qquad + [I(W; X^n|K) - I(W; Y^n|K)] - I(K; X^n|W, Y^n) \\ &\geq \sum_{i=1}^n \Big[I(W; Y_i|Y_{i+1}^n) - I(W; Z_i|Z^{i-1})\Big] \\ &\qquad + \sum_{i=1}^n \Big[I(W; X_i|X^{i-1}, K) - I(W; Y_i|Y_{i+1}^n, K)\Big] - H(K) \\ &\overset{(d)}{=} \sum_{i=1}^n \Big[I(W; Y_i|Y_{i+1}^n, Z^{i-1}) - I(W; Z_i|Z^{i-1}, Y_{i+1}^n) - R_0\Big] \\ &\qquad + \sum_{i=1}^n \Big[I(W; X_i|X^{i-1}, Y_{i+1}^n, K) \\ &\qquad\qquad - I(W; Y_i|Y_{i+1}^n, X^{i-1}, K)\Big] \\ &\overset{(e)}{=} \sum_{i=1}^n \Big[I(W; Y_i|Y_{i+1}^n, Z^{i-1}) - I(W; Z_i|Z^{i-1}, Y_{i+1}^n) - R_0\Big] \\ &\qquad + \sum_{i=1}^n \Big[I(W; X_i|X^{i-1}, Y_{i+1}^n, Z^{i-1}, K) \\ &\qquad\qquad - I(W; Y_i|Y_{i+1}^n, X^{i-1}, Z^{i-1}, K)\Big] \\ &\overset{(f)}{=} \sum_{i=1}^n \Big[I(W, Y_{i+1}^n, Z^{i-1}; Y_i) - I(W, Z^{i-1}, Y_{i+1}^n; Z_i) - R_0\Big] \\ &\qquad + \sum_{i=1}^n \Big[I(W, X^{i-1}, Y_{i+1}^n, Z^{i-1}, K; X_i) \\ &\qquad\qquad - I(W, Y_{i+1}^n, X^{i-1}, Z^{i-1}, K; Y_i)\Big] \\ &\overset{(g)}{=} \sum_{i=1}^n \Big[I(V_i; Y_i) - I(V_i; Z_i) - R_0 \\ &\qquad + I(U_i, V_i; X_i) - I(U_i, V_i; Y_i)\Big] \\ &= \sum_{i=1}^n \Big[-I(U_i, V_i; Z_i) - R_0 + I(U_i, V_i; X_i) \\ &\qquad + (I(U_i; Z_i|V_i) - I(U_i; Y_i|V_i))\Big] \\ &\overset{(h)}{\geq} \sum_{i=1}^n \Big[I(U_i; X_i|Z_i) - R_0 \\ &\qquad + [I(U_i; Z_i|V_i) - I(U_i; Y_i|V_i)]^-\Big] \end{aligned} \tag{76}$$

where $(a)$ follows by (3) and from the Markov chain $W -$

$X^n - Z^n$, $(b)$ follows since $K$ is independent of $(X^n, Y^n)$, $(c)$ follows from the Markov chain $(W, K) - X^n - Y^n$, $(d)$ follows because $H(K) = nR_0$ and from Csiszár's sum identity [48], $(e)$ follows from the Markov chains

$$Z^{i-1} - (X^{i-1}, Y_{i+1}^n, K) - (X_i, W) \tag{77}$$
$$Z^{i-1} - (X^{i-1}, Y_{i+1}^n, K) - (Y_i, W) \tag{78}$$

$(f)$ follows because $(X^n, Y^n, Z^n)$ are i.i.d. and $K$ is independent of $(X^n, Y^n, Z^n)$, $(g)$ follows from the definitions of $V_i$ and $U_i$, and $(h)$ follows from the Markov chain $V_i - U_i - X_i - Z_i$.

Next, we provide the matching converse for the privacy leakage rate in (66), which is achieved when $R_0 \geq I(U; \widetilde{X}|Y, V)$. We have

$$\begin{aligned}
n(R_\ell + \delta_n) &\overset{(a)}{\geq} H(X^n|Z^n) - H(X^n|Z^n, W) \\
&\overset{(b)}{=} H(X^n|Z^n) - \sum_{i=1}^{n} H(X_i|Z_i, Z^{i-1}, X_{i+1}^n, W, Y_{i+1}^n) \\
&\overset{(c)}{=} H(X^n|Z^n) - \sum_{i=1}^{n} H(X_i|Z_i, V_i, X_{i+1}^n) \\
&\overset{(d)}{\geq} \sum_{i=1}^{n} [H(X_i|Z_i) - H(X_i|Z_i, V_i)] \\
&= \sum_{i=1}^{n} I(V_i; X_i|Z_i)
\end{aligned} \tag{79}$$

where $(a)$ follows by (3), $(b)$ follows from the Markov chain

$$(Z_{i+1}^n, Y_{i+1}^n) - (X_{i+1}^n, W, Z^i) - X_i \tag{80}$$

$(c)$ follows from the definition of $V_i$, and $(d)$ follows because $(X^n, Z^n)$ are i.i.d.

We remark that the matching converse for the privacy leakage rate in (68), achieved when $R_0 \geq I(U; \widetilde{X}|Y)$, follows from the fact that conditional mutual information is non-negative.

**Secrecy Leakage Rate**: We have

$$\begin{aligned}
&n(R_\mathrm{s} + \delta_n) \\
&\overset{(a)}{\geq} [I(W; Y^n) - I(W; Z^n)] + [I(W; \widetilde{X}^n) - I(W; Y^n)] \\
&\overset{(b)}{=} [I(W; Y^n) - I(W; Z^n)] \\
&\qquad + I(W; \widetilde{X}^n|K) - I(K; \widetilde{X}^n|W) \\
&\qquad - I(W; Y^n|K) + I(K; Y^n|W) \\
&\overset{(c)}{=} [I(W; Y^n) - I(W; Z^n)] \\
&\qquad + [I(W; \widetilde{X}^n|K) - I(W; Y^n|K)] - I(K; \widetilde{X}^n|W, Y^n) \\
&\overset{(d)}{\geq} \sum_{i=1}^{n} \Big[ I(W; Y_i|Y_{i+1}^n) - I(W; Z_i|Z^{i-1}) \Big] \\
&\qquad + I(W; \widetilde{X}^n|Y^n, K) - H(K) \\
&\overset{(e)}{=} \sum_{i=1}^{n} \Big[ I(W; Y_i|Y_{i+1}^n, Z^{i-1}) - I(W; Z_i|Z^{i-1}, Y_{i+1}^n) - R_0 \Big] \\
&\qquad + nH(\widetilde{X}|Y) - \sum_{i=1}^{n} H(\widetilde{X}_i|Y_i, Y_{i+1}^n, W, K, \widetilde{X}^{i-1}) \\
&\overset{(f)}{\geq} \sum_{i=1}^{n} \Big[ I(W, Y_{i+1}^n, Z^{i-1}; Y_i) - I(W, Z^{i-1}, Y_{i+1}^n; Z_i) - R_0 \Big] \\
&\qquad + nH(\widetilde{X}|Y) - \sum_{i=1}^{n} H(\widetilde{X}_i|Y_i, Y_{i+1}^n, W, K, X^{i-1}, Z^{i-1}) \\
&\overset{(g)}{=} \sum_{i=1}^{n} \Big[ I(V_i; Y_i) - I(V_i; Z_i) - R_0 \Big] \\
&\qquad + nH(\widetilde{X}|Y) - \sum_{i=1}^{n} H(\widetilde{X}_i|Y_i, U_i, V_i) \\
&\overset{(h)}{=} \sum_{i=1}^{n} \Big[ I(V_i; Y_i) - I(V_i; Z_i) - R_0 \Big] \\
&\qquad + \sum_{i=1}^{n} \Big[ I(U_i, V_i; \widetilde{X}_i) - I(U_i, V_i; Y_i) \Big] \\
&= \sum_{i=1}^{n} \Bigg[ - I(U_i, V_i; Z_i) - R_0 + I(U_i, V_i; \widetilde{X}_i) \\
&\qquad\qquad + (I(U_i; Z_i|V_i) - I(U_i; Y_i|V_i)) \Bigg] \\
&\overset{(i)}{\geq} \sum_{i=1}^{n} \Big[ I(U_i; \widetilde{X}_i|Z_i) - R_0 \\
&\qquad\qquad + [I(U_i; Z_i|V_i) - I(U_i; Y_i|V_i)]^- \Big]
\end{aligned} \tag{81}$$

where $(a)$ follows by (2) and from the Markov chain $W - \widetilde{X}^n - Z^n$, $(b)$ follows because $K$ is independent of $(\widetilde{X}^n, Y^n)$, $(c)$ and $(d)$ follow from the Markov chain $(W, K) - \widetilde{X}^n - Y^n$, $(e)$ follows because $H(K) = nR_0$ and $(\widetilde{X}^n, Y^n)$ are i.i.d. and independent of $K$, and from the Csiszár's sum identity and the Markov chain

$$Y^{i-1} - (\widetilde{X}^{i-1}, W, K, Y_{i+1}^n, Y_i) - \widetilde{X}_i \tag{82}$$

$(f)$ follows since $(Y^n, Z^n)$ are i.i.d. and from the data processing inequality applied to the Markov chain

$$(X^{i-1}, Z^{i-1}) - (\widetilde{X}^{i-1}, W, K, Y_{i+1}^n, Y_i) - \widetilde{X}_i \tag{83}$$

$(g)$ follows from the definitions of $V_i$ and $U_i$, $(h)$ follows from the Markov chain $(V_i, U_i) - \widetilde{X}_i - Y_i$, and $(i)$ follows from the Markov chain $V_i - U_i - \widetilde{X}_i - Z_i$.

Next, the matching converse for the secrecy leakage rate in (70), achieved when $R_0 \geq I(U; \widetilde{X}|Y, V)$, is provided.

$$\begin{aligned}
n(R_\mathrm{s} + \delta_n) &\overset{(a)}{\geq} H(\widetilde{X}^n|Z^n) - H(\widetilde{X}^n|Z^n, W) \\
&\overset{(b)}{\geq} H(\widetilde{X}^n|Z^n) - \sum_{i=1}^{n} H(\widetilde{X}_i|Z_i, Z^{i-1}, \widetilde{X}_{i+1}^n, W, Y_{i+1}^n) \\
&\overset{(c)}{=} H(\widetilde{X}^n|Z^n) - \sum_{i=1}^{n} H(\widetilde{X}_i|Z_i, V_i, \widetilde{X}_{i+1}^n)
\end{aligned}$$

$$\overset{(d)}{\geq} \sum_{i=1}^{n}[H(\widetilde{X}_i|Z_i) - H(\widetilde{X}_i|Z_i, V_i)] = \sum_{i=1}^{n} I(V_i; \widetilde{X}_i|Z_i) \quad (84)$$

where $(a)$ follows by (2), $(b)$ follows from the Markov chain

$$(Z_{i+1}^n, Y_{i+1}^n) - (\widetilde{X}_{i+1}^n, W, Z^i) - \widetilde{X}_i \quad (85)$$

$(c)$ follows from the definition of $V_i$, and $(d)$ follows because $(\widetilde{X}^n, Z^n)$ are i.i.d.

Similar to the privacy leakage analysis above, the matching converse for the secrecy leakage rate in (71), achieved when $R_0 \geq I(U; \widetilde{X}|Y)$, follows from the fact that conditional mutual information is non-negative. ■

Introduce a uniformly distributed time-sharing random variable $Q \sim \text{Unif}[1:n]$ that is independent of other random variables, and define $X = X_Q$, $\widetilde{X} = \widetilde{X}_Q$, $Y = Y_Q$, $Z = Z_Q$, $V = V_Q$, and $U = (U_Q, Q)$, so

$$(Q, V) - U - \widetilde{X} - X - (Y, Z) \quad (86)$$

form a Markov chain. The converse proof follows by letting $\delta_n \to 0$.

**Cardinality Bounds**: We use the support lemma [48, Lemma 15.4] for the cardinality bound proofs, which is a standard step, so we omit the proof.


## ACKNOWLEDGMENT

O. Günlü and R. F. Schaefer were supported in part by the German Federal Ministry of Education and Research (BMBF) under the Grant 16KIS1004. H. Boche was supported in part by the BMBF within the national initiative on 6G Communication Systems through the research hub 6G-life under the Grant 16KISK002 and within the national initiative on 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.


## 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. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge University Press, 2011.

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

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

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

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

[19] 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.

[20] 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.

[21] O. Günlü, M. Bloch, and R. F. Schaefer, "Secure multi-function computation with private remote sources," Sep. 2021, [Online]. Available: arxiv.org/abs/2106.09485.

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

[23] 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.

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

[25] 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.

[26] 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.

[27] 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.

[28] 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.

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

[31] 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.

[32] J. Villard and P. Piantanida, "Secure multiterminal source coding with side information at the eavesdropper," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3668–3692, June 2013.

[33] S. I. Bross, "Secure cooperative source-coding with side information at the eavesdropper," *IEEE Trans. Inf. Theory*, vol. 62, no. 8, pp. 4544–4558, Aug. 2016.

[34] E. Ekrem and S. Ulukus, "Secure lossy source coding with side information," in *Proc. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, Sep. 2011, pp. 1098–1105.

[35] J. Körner and K. Marton, "Comparison of two noisy channels," *Topics Inf. Theory*, pp. 411–423, Aug. 1977.

[36] P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 279–280, Mar. 1974.

[37] O. Günlü, R. F. Schaefer, and H. V. Poor, "Biometric and physical identifiers with correlated noise for controllable private authentication," July 2020, [Online]. Available: arxiv.org/abs/2001.00847.

[38] A. D. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications: Part I," *IEEE Trans. Inf. Theory*, vol. 19, no. 6, pp. 769–772, Nov. 1973.

[39] S. Watanabe and Y. Oohama, "Secret key agreement from correlated Gaussian sources by rate limited public communication," *IEICE Trans. Fundam. Electron., Commun. Comp. Sci.*, vol. 93, no. 11, pp. 1976–1983, Nov. 2010.

[40] F. M. Willems and T. Ignatenko, "Quantization effects in biometric systems," in *Proc. Inf. Theory Appl. Workshop*, La Jolla, CA, Feb. 2009, pp. 372–379.

[41] V. Yachongka, H. Yagi, and Y. Oohama, "Secret key-based authentication with passive eavesdropper for scalar Gaussian sources," Feb. 2022, [Online]. Available: arxiv.org/abs/2202.10018.

[42] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: John Wiley & Sons, 2012.

[43] 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.

[44] 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.

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

[46] T. Holenstein and R. Renner, "On the randomness of independent experiments," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1865–1871, Apr. 2011.

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

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