entropy

MDPI

Article

Private Key and Decoder Side Information for Secure and Private Source Coding †

Onur Günlü [1,*], Rafael F. Schaefer [2,3], Holger Boche [4,5,6,7] and Harold Vincent Poor [8]

1 Information Coding Division, Linköping University, 58183 Linköping, Sweden
2 Chair of Information Theory and Machine Learning, Technische Universität Dresden, 01062 Dresden, Germany
3 BMBF Research Hub 6G-life, Technische Universität Dresden, 01062 Dresden, Germany
4 Lehrstuhl für Theoretische Informationstechnik, TUM School of Computation, Information and Technology, Technical University of Munich, 80333 Munich, Germany
5 CASA: Cyber Security in the Age of Large-Scale Adversaries Exzellenzcluster, Ruhr-Universität Bochum, 44780 Bochum, Germany
6 BMBF Research Hub 6G-life, Technical University of Munich, 80333 Munich, Germany
7 Munich Center for Quantum Science and Technology (MCQST), Schellingstr. 4, 80799 Munich, Germany
8 Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544, USA
* Correspondence: onur.gunlu@liu.se
† This paper is an extended version of our paper that will be published in the 2022 IEEE Information Theory Workshop, Mumbai, India, November 2022.

Abstract: We extend the problem of secure source coding by considering a remote source whose noisy measurements are correlated random variables used for secure source reconstruction. The main additions to the problem are as follows: (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 for the lossless case as well. Furthermore, the lossy rate region for scalar discrete-time Gaussian sources and measurement channels is established. An achievable lossy rate region that can be numerically computed is also provided for binary-input multiple additive discrete-time Gaussian noise measurement channels.

Keywords: information theoretic security; secure source coding; remote source; private key; side information

Citation: Günlü, O.; Schaefer, R.F.; Boche, H.; Poor, H.V. Private Key and Decoder Side Information for Secure and Private Source Coding. *Entropy* **2022**, *24*, 1716. https://doi.org/10.3390/e24121716

Academic Editor: T. Aaron Gulliver

Received: 18 October 2022
Accepted: 18 November 2022
Published: 24 November 2022

1. 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 needs to be reconstructed at a distant node is a classic example of this problem [2] (p. 258). Similarly, function computation problems in which a fusion center observes messages sent by other nodes to compute a function are closely related problems that can be used to model various recent applications [3,4]. Since messages sent over 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, it is necessary to provide an advantage to the decoder over the eavesdropper to enable secure source coding. Providing side information that 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 for the eavesdropper to access only a strict subset of all messages is also a method to enable secure distributed source coding, which was considered in [9–11]; see also [12], in which a similar method was applied to enable secure remote source reconstruction. Similarly, a private key that is shared by legitimate terminals and hidden from the eavesdropper can also provide such an advantage, as in [13,14].

Source coding models in the literature commonly assume that dependent multiletter 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 the side information problem [19] (Section IV-B), the correlation between these multiletter random variables was 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 one 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] (p. 78) and [23] (Figure 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 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.

1.1. Summary of Contributions

We extend the lossless and lossy source coding rate region analyses by considering a remote source that should be kept private, decoder and eavesdropper side information, and a private key shared by the encoder and decoder. Considering that one encoder provides insights with enough richness to extend the results to multiple encoders [31], in this work, we consider the single encoder case. A summary of the main contributions is as follows.

- We characterize the lossy secure and private source coding region when noisy measurements of a remote source are observed by all terminals, and there is one private key available.
- Requiring reliable source reconstruction, we also characterize the rate region for the lossless secure and private source coding problem.
- A Gaussian remote source and independent additive Gaussian noise measurement channels are considered to establish their lossy rate region under squared error distortion.
- We provide an achievable lossy secure and private source coding region for a binary remote source and its measurements through additive Gaussian noise channels, which includes computable differential entropy terms.

1.2. Organization

This paper is organized as follows. In Section 2, we introduce the lossless and lossy secure and private source coding problems with decoder and eavesdropper side information and a private key under storage, secrecy, privacy, and reliability or distortion constraints. In Section 3, we characterize the rate regions for the introduced problems, which include three parts that correspond to different private key rate regimes. In Section 4, we evaluate

the lossy rate region for Gaussian sources and channels with squared error distortion. In Section 5, we consider a binary modulated remote source measured through additive Gaussian noise channels and provide an inner bound for the lossy rate region with Hamming distortion. In Section 6, we provide the proof for the lossy secure and and private source coding region.

1.3. Notation

Uppercase X represents random variables and lowercase x their realizations from a set $\mathcal{X}$, denoted by calligraphic letters. A discrete random variable X has probability distribution P_X and a continuous random variable X probability density function (pdf) p_X. A subscript i denotes the position of a variable in a length-n sequence $X^n = X_1, X_2, \ldots, X_i, \ldots, X_n$. Boldface uppercase $\mathbf{X} = [X_1, X_2, \ldots]^T$ represent vector random variables, where T denotes the transpose. $[1:m]$ denotes the set $\{1, 2, \ldots, m\}$ for an integer $m \geq 1$. Define $[a]^- = \min\{a, 0\}$ for $a \in \mathbb{R}$. Function $H_b(x) = -x \log x - (1-x)\log(1-x)$ is the binary entropy function, where logarithms are to the base 2. A binary symmetric channel (BSC) with crossover probability ϵ is denoted by BSC(ϵ). $X \sim \text{Bern}(\beta)$ with $\mathcal{X} = \{0, 1\}$ is a binary random variable with $\Pr[X = 1] = \beta$. The $*$-operator represents $p * q = (1 - 2q)p + q$. Function $Q(\cdot)$ denotes the complementary cumulative distribution function of the standard Gaussian distribution. The function $\text{sgn}(\cdot)$ represents the signum function.

2. System Model

We consider the lossy source coding model with one encoder, one decoder, and an eavesdropper (Eve), depicted in Figure 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. Decoder $\mathsf{Dec}(\cdot, \cdot, \cdot)$ observes index W, 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 index W. Assume 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.

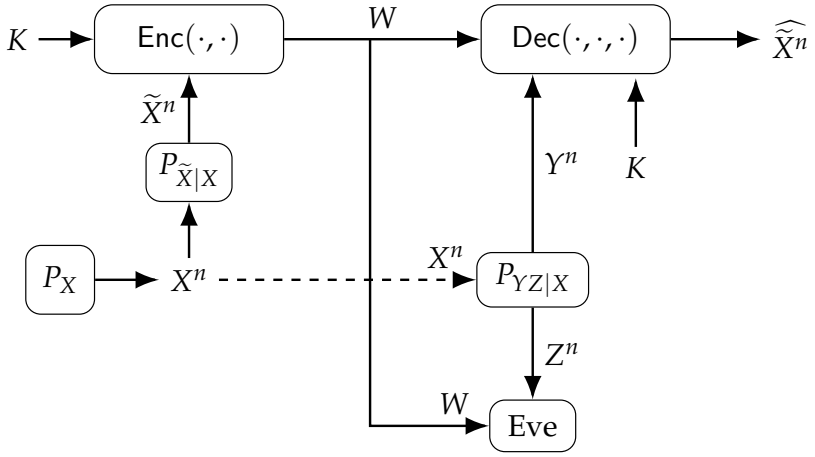

Figure 1. Source coding with noisy measurements $(\widetilde{X}^n, Y^n)$ of a remote source X^n and with a uniform private key K under privacy, secrecy, communication, and distortion constraints.

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

$$\frac{1}{n} \log |\mathcal{W}| \leq R_{\mathrm{w}} + \delta \qquad \text{(storage)} \tag{1}$$

$$\frac{1}{n} I(\widetilde{X}^n; W | Z^n) \leq R_{\mathrm{s}} + \delta \qquad \text{(secrecy)} \tag{2}$$

$$\frac{1}{n} I(X^n; W | Z^n) \leq R_\ell + \delta \qquad \text{(privacy)} \tag{3}$$

$$\mathbb{E}\Big[d\Big(\widetilde{X}^n, \widehat{\widetilde{X}^n}(Y^n, W, K) \Big) \Big] \leq D + \delta \qquad \text{(distortion)} \tag{4}$$

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. $\Diamond$

In (2) and (3), we consider conditional mutual information terms to take account of unavoidable secrecy and privacy leakages due to Eve's side information, i.e., $I(\widetilde{X}^n; Z^n)$ and $I(X^n; Z^n)$, respectively; see also [21,32]. Furthermore, we consider conditional mutual information terms rather than corresponding conditional entropy terms, the latter of which is used in [6,14,33–35], 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

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

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

3. Secure and Private Source Coding Regions

3.1. Lossy Source Coding

The lossy secure and and private source coding region $\mathcal{R}_{\mathrm{D}}$ is characterized below; see Section 6 for its proof.

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

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_ℓ, D) satisfying*

$$R_w \geq I(U; \widetilde{X} | Y) \tag{6}$$

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

$$R_s \geq I(U; \widetilde{X} | Z) + R' - R_0 \tag{7}$$

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

where we have

$$R' = [I(U; Z | V, Q) - I(U; Y | V, Q)]^- \tag{9}$$

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

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

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

$$R_s \geq 0 \tag{12}$$
$$R_\ell \geq 0 \tag{13}$$

for some

$$P_{QVU\tilde{X}XYZ} = P_{Q|V}P_{V|U}P_{U|\tilde{X}}P_{\tilde{X}|X}P_X P_{YZ|X} \tag{14}$$

such that $\mathbb{E}\big[d\big(\widetilde{X},\hat{\widetilde{X}}(U,Y)\big)\big] \leq D$ *for some reconstruction function* $\hat{\widetilde{X}}(U,Y)$. *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* $|\mathcal{Q}| \leq 2$, $|\mathcal{V}| \leq |\widetilde{X}| + 3$, *and* $|\mathcal{U}| \leq (|\widetilde{X}| + 3)^2$.

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 a private key.

3.2. Lossless Source Coding

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

Proposition 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_ℓ) *satisfying*

$$R_w \geq H(\widetilde{X}|Y) \tag{15}$$

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

$$R_s \geq H(\widetilde{X}|Z) + R'' - R_0 \tag{16}$$
$$R_\ell \geq I(\widetilde{X};X|Z) + R'' - R_0 \tag{17}$$

where we have

$$R'' = [I(\widetilde{X};Z|V,Q) - I(\widetilde{X};Y|V,Q)]^- \tag{18}$$

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

$$R_s \geq I(V;\widetilde{X}|Z) \tag{19}$$
$$R_\ell \geq I(V;X|Z) \tag{20}$$

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

$$R_s \geq 0 \tag{21}$$
$$R_\ell \geq 0 \tag{22}$$

for some

$$P_{QV\tilde{X}XYZ} = P_{Q|V}P_{V|\tilde{X}}P_{\tilde{X}|X}P_X P_{YZ|X}. \tag{23}$$

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 $\mathcal{R}_{\mathrm{D}}$, given in Theorem 1 above, by choosing $U = \widetilde{X}$, such that we have 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. □

4. Gaussian Sources and Additive Gaussian Noise 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 ([36]). Z (or eavesdropper) is *less noisy* than Y (or decoder) if

$$I(L;Z) \geq I(L;Y) \tag{24}$$

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_ℓ, D) *satisfying*

$$R_w \geq I(U;\widetilde{X}|Y) = I(U;\widetilde{X}) - I(U;Y) \tag{25}$$
$$R_s \geq I(U;\widetilde{X}|Z) = I(U;\widetilde{X}) - I(U;Z) \tag{26}$$
$$R_\ell \geq I(U;X|Z) = I(U;X) - I(U;Z) \tag{27}$$

for some

$$P_{U\widetilde{X}XYZ} = P_{U|\widetilde{X}} P_{\widetilde{X}|X} P_X P_{YZ|X} \tag{28}$$

such that $\mathbb{E}\left[d\left(\widetilde{X}, \widehat{\widetilde{X}}(U,Y)\right)\right] \leq D$ *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 (6)–(8) since $R_0 = 0$. Furthermore, R' defined in (9) 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 Figure 1

$$X = \rho_x \widetilde{X} + N_x \tag{29}$$
$$Y = \rho_y X + N_y \tag{30}$$
$$Z = \rho_z X + N_z \tag{31}$$

where we have 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

$$N_x \sim \mathcal{N}(0, 1-\rho_x^2) \tag{32}$$

$$N_y \sim \mathcal{N}(0, 1-\rho_y^2) \tag{33}$$

$$N_z \sim \mathcal{N}(0, 1-\rho_z^2) \tag{34}$$

such that $(\widetilde{X}, N_x, N_y, N_z)$ are mutually independent, and we consider the squared error distortion, i.e., $d(\widetilde{x}, \widehat{\widetilde{x}}) = (\widetilde{x} - \widehat{\widetilde{x}})^2$. Note 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) [37] (Lemma II); see [20] (Theorem 3) and [38] (Section V) for binary symmetric inverse channel assumptions imposed to apply Mrs. Gerber's lemma [39]. Suppose $|\rho_z| > |\rho_y|$, such that Y is less stochastically degraded than Z, since then there exists a random variable $\widetilde{Y}$ such that $P_{\widetilde{Y}|X} = P_{Y|X}$ and $P_{\widetilde{Y}Z|X} = P_{Z|X} P_{\widetilde{Y}|Z}$ [40] (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] (p. 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.

Proposition 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_ℓ, D) satisfying, for $\alpha \in (0,1]$,*

$$R_w \geq \frac{1}{2} \log \left(\frac{1 - \rho_x^2 \rho_y^2 (1-\alpha)}{\alpha} \right) \tag{35}$$

$$R_s \geq \frac{1}{2} \log \left(\frac{1 - \rho_x^2 \rho_z^2 (1-\alpha)}{\alpha} \right) \tag{36}$$

$$R_\ell \geq \frac{1}{2} \log \left(\frac{1 - \rho_x^2 \rho_z^2 (1-\alpha)}{1 - \rho_x^2 (1-\alpha)} \right) \tag{37}$$

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

Proof Sketch. For the achievability proof, let $U \sim \mathcal{N}(0, 1-\alpha)$ and $\Theta \sim \mathcal{N}(0, \alpha)$, as in [41] ([Equation (32)]) and [42] (Appendix B), be independent random variables for some $\alpha \in (0,1]$ 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,

$$D = \mathbb{E}\left[\left(\widetilde{X} - \widehat{\widetilde{X}}(U, Y) \right)^2 \right] \overset{(a)}{=} \frac{1}{2\pi e} e^{2h(\widetilde{X}|U,Y)} \tag{39}$$

where equality in (a) is achieved because $\widetilde{X}$ is Gaussian and the reconstruction function is the MMSE estimator [43] (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

$$\begin{aligned} h(\widetilde{X}|U,Y) &= h(\widetilde{X},U,Y) - h(U,Y) \\ &= \frac{1}{2}\log\left(2\pi e \frac{\det(\mathbf{K}_{\widetilde{X}UY})}{\det(\mathbf{K}_{UY})}\right) \end{aligned} \tag{40}$$

where $\det(\cdot)$ is the determinant of a matrix; see also [12] (Section F). Combining (39) and (40), and calculating the determinants, we obtain

$$D = \frac{\alpha(1-\rho_x^2\rho_y^2)}{1-\rho_x^2\rho_y^2(1-\alpha)}. \tag{41}$$

One can also show that

$$I(U;\widetilde{X}) = h(\widetilde{X}) - h(\widetilde{X}|U) = \frac{1}{2}\log\left(\frac{1}{\alpha}\right) \tag{42}$$

$$I(U;X) = h(X) - h(X|U) = \frac{1}{2}\log\left(\frac{1}{1-\rho_x^2(1-\alpha)}\right) \tag{43}$$

$$I(U;Y) = h(Y) - h(Y|U) = \frac{1}{2}\log\left(\frac{1}{1-\rho_x^2\rho_y^2(1-\alpha)}\right) \tag{44}$$

$$I(U;Z) = h(Z) - h(Z|U) = \frac{1}{2}\log\left(\frac{1}{1-\rho_x^2\rho_z^2(1-\alpha)}\right). \tag{45}$$

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

For the converse proof, one can first show that

$$I(U;\widetilde{X}) - I(U;Y) = h(Y|U) - h(\widetilde{X}|U) \tag{46}$$

$$I(U;\widetilde{X}) - I(U;Z) = h(Z|U) - h(\widetilde{X}|U) \tag{47}$$

$$I(U;X) - I(U;Z) = h(Z|U) - h(X|U) \tag{48}$$

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

$$h(\widetilde{X}|U) = \frac{1}{2}\log(2\pi e\alpha) \tag{49}$$

for any $\alpha \in (0,1]$ 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

$$\begin{aligned} e^{2h(Y|U)} &\overset{(a)}{=} e^{2h(\rho_x\rho_y\widetilde{X}|U)} + e^{2h(\rho_y N_x + N_y)} \\ &= 2\pi e\left(\rho_x^2\rho_y^2\alpha + \rho_y^2(1-\rho_x^2) + 1 - \rho_y^2\right) \\ &= 2\pi e\left(1-\rho_x^2\rho_y^2(1-\alpha)\right) \end{aligned} \tag{50}$$

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 (49) above; see [20] (Lemma 1 and Equation (28)) for a similar result applied to binary random variables by extending Mrs. Gerber's lemma. Similarly, we have

$$e^{2h(Z|U)} = 2\pi e\left(1-\rho_x^2\rho_z^2(1-\alpha)\right) \tag{51}$$

which follows by replacing (Y, ρ_y, N_y) with (Z, ρ_z, N_z) in (50), 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 relation $U - \widetilde{X} - X - (Y, Z)$. Furthermore, we have

$$\begin{aligned} e^{2h(X|U)} &\overset{(a)}{=} e^{2h(\rho_x \widetilde{X}|U)} + e^{2h(N_x)} \\ &= 2\pi e\big(\rho_x^2 \alpha + 1 - \rho_x^2\big) \\ &= 2\pi e\big(1 - \rho_x^2(1-\alpha)\big) \end{aligned} \tag{52}$$

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 (46)–(52) to (25)–(27), the converse proof for (35)–(37) follows.

Next, consider

$$\begin{aligned} h(\widetilde{X}|U,Y) &= -I(U;\widetilde{X}|Y) + h(\widetilde{X}|Y) \\ &\overset{(a)}{=} -h(Y|U) + h(\widetilde{X}|U) + h(Y|\widetilde{X}) \\ &\overset{(b)}{=} \frac{1}{2}\log\Big(\frac{\alpha}{1-\rho_x^2\rho_y^2(1-\alpha)}\Big) + h(\rho_x\rho_y\widetilde{X} + \rho_y N_x + N_y|\widetilde{X}) \\ &\overset{(c)}{=} \frac{1}{2}\log\Big(\frac{\alpha}{1-\rho_x^2\rho_y^2(1-\alpha)}\Big) + h(\rho_y N_x + N_y) \\ &= \frac{1}{2}\log\Big(2\pi e \frac{\alpha(\rho_y^2(1-\rho_x^2) + (1-\rho_y^2))}{1-\rho_x^2\rho_y^2(1-\alpha)}\Big) \\ &= \frac{1}{2}\log\Big(2\pi e \frac{\alpha(1-\rho_x^2\rho_y^2)}{1-\rho_x^2\rho_y^2(1-\alpha)}\Big) \end{aligned} \tag{53}$$

where (a) follows by (25) and (46), and since $h(Y) = h(\widetilde{X})$, (b) follows by (49) and (50), 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 [43] (Theorem 8.6.6)

$$\mathbb{E}\Big[\big(\widetilde{X} - \widehat{\widetilde{X}}(U,Y)\big)^2\Big] \geq \frac{1}{2\pi e} e^{2h(\widetilde{X}|U,Y)}. \tag{54}$$

Combining the distortion constraint given in Corollary 1 with (53) and (54), the converse proof for (38) follows. □

5. Multiple Binary-input Additive Gaussian Noise Channels

Consider next a binary remote source $X \in \{-1, 1\}$ and its binary noisy measurement $\widetilde{X} \in \{-1, 1\}$ observed by the encoder, which represents a practical setting with binary quantizations. For instance, a static random-access memory (SRAM) start-up output at a nominal temperature is a binary value obtained by quantizing sums of Gaussian random variables [28,44]. Suppose the noisy channel $P_{YZ|X}$ outputs consist of a single discrete-time additive Gaussian noise channel output Y observed by the decoder and two independent discrete-time additive Gaussian noise channel outputs $\mathbf{Z} = [Z_1, Z_2]^T$ observed by the eavesdropper, in which the eavesdropper obtains more information by measuring the remote source twice. Furthermore, assume that X is uniformly distributed, the binary channel $P_{\widetilde{X}|X}$ is symmetric such that $\Pr[\widetilde{X} \neq X] = p$ for $p \in [0, 1]$, and we also have

$$Y = \rho_y X + N_y \tag{55}$$

$$\mathbf{Z} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \rho_z X \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} N_{z_1} \\ N_{z_2} \end{bmatrix} \tag{56}$$

where we have fixed correlation coefficients $\rho_y, \rho_z \in (-1,1)$ and additive Gaussian noise random variables

$$N_y \sim \mathcal{N}(0, 1-\rho_y^2) \tag{57}$$

$$N_{z_1} \sim \mathcal{N}(0, 1-\rho_z^2) \tag{58}$$

$$N_{z_2} \sim \mathcal{N}(0, 1-\rho_z^2) \tag{59}$$

such that $(X, N_y, N_{z_1}, N_{z_2})$ are mutually independent. Consider the Hamming distortion, i.e., $d(\widetilde{x}, \widehat{\widetilde{x}}) = \mathbb{1}\{\widetilde{x} \neq \widehat{\widetilde{x}}\}$. Impose the condition $|\rho_z| > |\rho_y|$ such that Z_1 and Z_2 are less noisy than Y, so $\mathbf{Z}$ is also less noisy than Y, which follows by applying similar steps as being applied in Section 4. Thus, for $R_0 = 0$, the region $\mathcal{R}_{\mathrm{D}}$ characterized in Corollary 1 is also valid for such binary-input additive Gaussian noise channels when one replaces Z with $\mathbf{Z}$. A computable achievable lossy secure and private source coding region for such channels is given next.

Proposition 3. *For the setting with multiple binary-input additive Gaussian noise channels, defined above, such that $|\rho_z| > |\rho_y|$ and $R_0 = 0$, the region $\mathcal{R}_D$ with Hamming distortion includes the set of all rate tuples (R_w, R_s, R_ℓ, D) satisfying, for an independent random variable $C \sim Bern(p * q)$ with any $q \in [0, 0.5]$ and for any $\lambda \in [0,1]$,*

$$R_w \geq \lambda\Big(1 - H_b(q) - h\big(\rho_y X + N_y\big) + h\big(\rho_y(1-2C) + N_y\big)\Big) \tag{60}$$

$$R_s \geq \lambda\left(1 - H_b(q) - h\left(\begin{bmatrix}\rho_z X + N_{z_1}\\ \rho_z X + N_{z_2}\end{bmatrix}\right) + h\left(\begin{bmatrix}\rho_z(1-2C) + N_{z_1}\\ \rho_z(1-2C) + N_{z_2}\end{bmatrix}\right)\right) \tag{61}$$

$$R_\ell \geq \lambda\left(1 - H_b(p*q) - h\left(\begin{bmatrix}\rho_z X + N_{z_1}\\ \rho_z X + N_{z_2}\end{bmatrix}\right) + h\left(\begin{bmatrix}\rho_z(1-2C) + N_{z_1}\\ \rho_z(1-2C) + N_{z_2}\end{bmatrix}\right)\right) \tag{62}$$

$$D \geq \lambda q + (1-\lambda)\left(p * Q\left(\frac{\rho_y}{\sqrt{1-\rho_y^2}}\right)\right) \tag{63}$$

where random variable $Y = \big(\rho_y X + N_y\big)$ has pdf

$$\frac{1}{2}\frac{\left(e^{-\frac{(y+\rho_y)^2}{2(1-\rho_y^2)}} + e^{-\frac{(y-\rho_y)^2}{2(1-\rho_y^2)}}\right)}{\sqrt{2\pi(1-\rho_y^2)}} \tag{64}$$

the random variable $\bar{Y} = \big(\rho_y(1-2C) + N_y\big)$ has pdf

$$(p*q)\frac{e^{-\frac{(\bar{y}+\rho_y)^2}{2(1-\rho_y^2)}}}{\sqrt{2\pi(1-\rho_y^2)}} + (1-(p*q))\frac{e^{-\frac{(\bar{y}-\rho_y)^2}{2(1-\rho_y^2)}}}{\sqrt{2\pi(1-\rho_y^2)}} \tag{65}$$

the vector random variable $\begin{bmatrix}Z_1\\ Z_2\end{bmatrix} = \left(\begin{bmatrix}\rho_z X + N_{z_1}\\ \rho_z X + N_{z_2}\end{bmatrix}\right)$ has joint pdf

$$\frac{1}{2}\frac{\left(e^{-\frac{\left((z_1+\rho_z)^2+(z_2+\rho_z)^2\right)}{2(1-\rho_z^2)}} + e^{-\frac{\left((z_1-\rho_z)^2+(z_2-\rho_z)^2\right)}{2(1-\rho_z^2)}}\right)}{2\pi(1-\rho_z^2)} \tag{66}$$

and the vector random variable $\begin{bmatrix}\bar{Z}_1\\ \bar{Z}_2\end{bmatrix} = \left(\begin{bmatrix}\rho_z(1-2C)+N_{z_1}\\ \rho_z(1-2C)+N_{z_2}\end{bmatrix}\right)$ *has joint pdf*

$$(p*q)\frac{e^{-\frac{\left((\bar{z}_1+\rho_z)^2+(\bar{z}_2+\rho_z)^2\right)}{2(1-\rho_z^2)}}}{2\pi(1-\rho_z^2)}+(1-(p*q))\frac{e^{-\frac{\left((\bar{z}_1-\rho_z)^2+(\bar{z}_2-\rho_z)^2\right)}{2(1-\rho_z^2)}}}{2\pi(1-\rho_z^2)}. \tag{67}$$

Proof. We first evaluate (25)–(27) by choosing a binary uniformly distributed U and a channel $P_{\widetilde{X}|U}$ such that $\Pr[\widetilde{X}\neq U]=q$ for any $q\in[0,0.5]$. We have

$$I(U;\widetilde{X}) = H(\widetilde{X}) - H(\widetilde{X}|U) \overset{(a)}{=} 1-H_b(q) \tag{68}$$

$$I(U;X) = H(X) - H(X|U) \overset{(b)}{=} 1-H_b(p*q) \tag{69}$$

where (a) and (b) follow by relabeling the input and output symbols to represent the channels $P_{\widetilde{X}|U}$ and $P_{X|\widetilde{X}}$ as BSC(q) and BSC(p), respectively, which follows since entropy is preserved under a bijective mapping for discrete random variables. For relabeled symbols, the channel $P_{X|U}$ is a BSC$(p*q)$ since it is a concatenation of two BSCs, so denote the independent random noise component in this channel as $C\sim$ Bern$(p*q)$. Then, we obtain

$$h(Y|U) = h(\rho_y X+N_y|U) \overset{(a)}{=} h(\rho_y(1-2C)+N_y) = h(\bar{Y}) \tag{70}$$

where (a) follows since symbols $\{-1,1\}$ correspond to the antipodal modulation of binary symbols, and since (C,N_y,U) are mutually independent. One can compute (70) numerically by using the pdf

$$p_{\bar{Y}}(\bar{y}) = \sum_{c=0}^{1} P_C(c)p_{\bar{Y}|C}(\bar{y}|c) = (p*q)\frac{e^{-\frac{(\bar{y}+\rho_y)^2}{2(1-\rho_y^2)}}}{\sqrt{2\pi(1-\rho_y^2)}}+(1-(p*q))\frac{e^{-\frac{(\bar{y}-\rho_y)^2}{2(1-\rho_y^2)}}}{\sqrt{2\pi(1-\rho_y^2)}}. \tag{71}$$

Similarly, we can compute

$$h(Y) = h(\rho_y X+N_y) \tag{72}$$

numerically by using the pdf

$$p_Y(y) = \sum_{x\in\{-1,1\}} P_X(x)p_{Y|X}(y|x) = \frac{1}{2}\frac{\left(e^{-\frac{(y+\rho_y)^2}{2(1-\rho_y^2)}}+e^{-\frac{(y-\rho_y)^2}{2(1-\rho_y^2)}}\right)}{\sqrt{2\pi(1-\rho_y^2)}}. \tag{73}$$

Next, consider

$$h(\mathbf{Z}|U) = h\left(\left(\rho_z X\begin{bmatrix}1\\1\end{bmatrix}+\begin{bmatrix}N_{z_1}\\N_{z_2}\end{bmatrix}\right)\middle|U\right) \overset{(a)}{=} h\left(\begin{bmatrix}\rho_z(1-2C)+N_{z_1}\\ \rho_z(1-2C)+N_{z_2}\end{bmatrix}\right) = h\left(\begin{bmatrix}\bar{Z}_1\\ \bar{Z}_2\end{bmatrix}\right) \tag{74}$$

where (a) follows since (C,N_{z_1},N_{z_2},U) are mutually independent. Denote

$$\bar{\mathbf{Z}} = [\bar{Z}_1,\bar{Z}_2]^T. \tag{75}$$

We can compute (74) numerically by using the joint pdf

$$p_{\bar{\mathbf{Z}}}(\bar{\mathbf{z}}) = p_{\bar{Z}_1\bar{Z}_2}(\bar{z}_1,\bar{z}_2) = \sum_{c=0}^{1} P_C(c)\, p_{\bar{Z}_1\bar{Z}_2|C}(\bar{z}_1,\bar{z}_2|c)$$
$$= (p*q)\frac{e^{-\frac{\left((\bar{z}_1+\rho_z)^2+(\bar{z}_2+\rho_z)^2\right)}{2(1-\rho_z^2)}}}{2\pi(1-\rho_z^2)} + (1-(p*q))\frac{e^{-\frac{\left((\bar{z}_1-\rho_z)^2+(\bar{z}_2-\rho_z)^2\right)}{2(1-\rho_z^2)}}}{2\pi(1-\rho_z^2)} \tag{76}$$

which follows since $\bar{\mathbf{Z}}|C$ is a jointly Gaussian vector random variable with independent components $\bar{Z}_1|C$ and $\bar{Z}_2|C$, since every scalar linear combination of the components is Gaussian; see [45] (Theorem 1). Similarly, we can compute

$$h(\mathbf{Z}) = h\left(\begin{bmatrix}\rho_z X + N_{z_1}\\ \rho_z X + N_{z_2}\end{bmatrix}\right) \tag{77}$$

numerically by using the joint pdf

$$p_{\mathbf{Z}}(\mathbf{z}) = p_{Z_1Z_2}(z_1,z_2) = \sum_{x\in\{-1,1\}} P_X(x)\, p_{Z_1Z_2|X}(z_1,z_2|x)$$
$$= \frac{1}{2}\frac{\left(e^{-\frac{\left((z_1+\rho_z)^2+(z_2+\rho_z)^2\right)}{2(1-\rho_z^2)}} + e^{-\frac{\left((z_1-\rho_z)^2+(z_2-\rho_z)^2\right)}{2(1-\rho_z^2)}}\right)}{2\pi(1-\rho_z^2)}. \tag{78}$$

Now, we consider the expected distortion. First, choose the reconstruction function

$$\hat{\widetilde{X}}_1(U,Y) = U \tag{79}$$

for the binary uniformly distributed U and the channel $P_{\widetilde{X}|U}$ such that $\Pr[\widetilde{X} \neq U] = q$ for any $q \in [0, 0.5]$, as considered above. For this reconstruction function and choices of U and $P_{\widetilde{X}|U}$, we obtain the expected distortion

$$\mathbb{E}\left[d\left(\widetilde{X}, \hat{\widetilde{X}}_1(U,Y)\right)\right] = q. \tag{80}$$

Second, choose the reconstruction function

$$\hat{\widetilde{X}}_2(U,Y) = \operatorname{sgn}(Y) \tag{81}$$

and consider U. We then obtain

$$\mathbb{E}\left[d\left(\widetilde{X}, \hat{\widetilde{X}}_2(U,Y)\right)\right] = p*Q\left(\frac{\rho_y}{\sqrt{1-\rho_y^2}}\right) \tag{82}$$

which follows since the channel $P_{\operatorname{sgn}(Y)|\widetilde{X}}$ can be considered as a concatenation of two BSCs with crossover probabilities p and $Q\left(\frac{\rho_y}{\sqrt{1-\rho_y^2}}\right)$, where the former follows since $\Pr[\widetilde{X} \neq X] = p$ and the latter because $X \in \{-1,1\}$ and

$$\Pr[X \neq \operatorname{sgn}(Y)] = \Pr[X \neq \operatorname{sgn}(\rho_y X + N_y)] = \Pr[N_y > \rho_y]. \tag{83}$$

Therefore, the proof for the achievable lossy secure and private source coding region follows by combining (68)–(70), (72), (74), (77), (80), and (82) by applying time sharing, with time-

sharing parameter $\lambda \in [0,1]$, between the two reconstruction functions in (79) and (81) with corresponding U and $P_{\widetilde{X}|U}$, since for constant U the terms in (25)–(27) are zero. □

Remark 1. *The proof of Proposition 3 follows similar steps as those in [46] (Section II) and it seems that the achievable lossy secure and private source coding region given in Proposition 3 is optimal. Considering (R_w, R_s, R_ℓ), one can apply Mrs. Gerber's lemma to show that the choice of U such that $P_{\widetilde{X}|U}$ is a BSC(q) after relabeling the input and output symbols is optimal, since Mrs. Gerber's lemma is valid for all binary-input symmetric memoryless channels with discrete or continuous outputs [47]. This result follows because convexity is preserved; see also [48] (Appendix B) for an alternative proof of convexity preservation for independent BSC measurements. However, it is not entirely clear how to prove that the sign operation used for estimation suffices for the rate region.*

6. Proof for Theorem 1

6.1. Achievability Proof for Theorem 1

Proof Sketch. We leverage the output statistics of random binning (OSRB) method [16,49,50] for the achievability proof by following the steps described in [51] (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 $F_{\mathrm{v}} \in [1 : 2^{n\widetilde{R}_{\mathrm{v}}}]$ and $W_{\mathrm{v}} \in [1 : 2^{nR_{\mathrm{v}}}]$. Furthermore, to each u^n assign three random bin indices $F_{\mathrm{u}} \in [1 : 2^{n\widetilde{R}_{\mathrm{u}}}]$, $W_{\mathrm{u}} \in [1 : 2^{nR_{\mathrm{u}}}]$, and $K_{\mathrm{u}} \in [1 : 2^{nR_0}]$, where R_0 is the private key rate defined in Section 2. Public indices $F = (F_{\mathrm{v}}, F_{\mathrm{u}})$ 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

$$W = (W_{\mathrm{v}}, W_{\mathrm{u}}, K + K_{\mathrm{u}}) \tag{84}$$

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

The public index F_{v} is almost independent of $(\widetilde{X}^n, X^n, Y^n, Z^n)$ if we have [49] (Theorem 1)

$$\widetilde{R}_{\mathrm{v}} < H(V|\widetilde{X}, X, Y, Z) \overset{(a)}{=} H(V|\widetilde{X}) \tag{85}$$

where (a) follows since $(X, Y, Z) - \widetilde{X} - V$ form a Markov chain. The constraint in (85) suggests that the expected value, taken over the random bin assignments, of the variational distance between the joint probability distributions $\mathrm{Unif}[1{:}2^{n\widetilde{R}_{\mathrm{v}}}] \cdot P_{\widetilde{X}^n}$ and $P_{F_{\mathrm{v}}\widetilde{X}^n}$ vanishes when $n \to \infty$. Moreover, the public index F_{u} is almost independent of $(V^n, \widetilde{X}^n, X^n, Y^n, Z^n)$ if we have

$$\widetilde{R}_{\mathrm{u}} < H(U|V, \widetilde{X}, X, Y, Z) \overset{(a)}{=} H(U|V, \widetilde{X}) \tag{86}$$

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

Using a Slepian–Wolf (SW) [1] decoder that observes $(Y^n, F_{\mathrm{v}}, W_{\mathrm{v}})$, one can reliably estimate V^n if we have [49] (Lemma 1)

$$\widetilde{R}_{\mathrm{v}} + R_{\mathrm{v}} > H(V|Y) \tag{87}$$

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 $(K, V^n, Y^n, F_{\mathrm{u}}, W_{\mathrm{u}}, K + K_{\mathrm{u}})$ if we have

$$R_0 + \widetilde{R}_{\mathrm{u}} + R_{\mathrm{u}} > H(U|V, Y). \tag{88}$$

To satisfy (85)–(88), for any $\epsilon > 0$ we fix

$$\widetilde{R}_{\mathrm{v}} = H(V|\widetilde{X}) - \epsilon \tag{89}$$

$$R_{\mathrm{v}} = I(V;\widetilde{X}) - I(V;Y) + 2\epsilon \tag{90}$$

$$\widetilde{R}_{\mathrm{u}} = H(U|V,\widetilde{X}) - \epsilon \tag{91}$$

$$R_0 + R_{\mathrm{u}} = I(U;\widetilde{X}|V) - I(U;Y|V) + 2\epsilon. \tag{92}$$

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

Communication Rate: (90) and (92) result in a communication (storage) rate of

$$R_{\mathrm{w}} = R_0 + R_{\mathrm{v}} + R_{\mathrm{u}} \overset{(a)}{=} I(U;\widetilde{X}|Y) + 4\epsilon \tag{93}$$

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

Privacy Leakage Rate: Since 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

$$\overline{W} = (W_{\mathrm{v}}, W_{\mathrm{u}}, K_{\mathrm{u}}). \tag{94}$$

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 (84), 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 (91) and (92) if $\epsilon \to 0$ when $n \to \infty$. Thus, we have the privacy leakage

$$\begin{aligned}
I(X^n; W, F|Z^n) &= I(X^n; \overline{W}, F|Z^n) - nR_0 \\
&\overset{(a)}{=} H(\overline{W}, F|Z^n) - H(\overline{W}, F|X^n) - nR_0 \\
&\overset{(b)}{=} H(\overline{W}, F|Z^n) - H(U^n, V^n|X^n) + H(V^n|\overline{W}, F, X^n) + H(U^n|V^n, \overline{W}, F, X^n) - nR_0 \\
&\overset{(c)}{\leq} H(\overline{W}, F|Z^n) - nH(U, V|X) + 2n\epsilon_n - nR_0
\end{aligned} \tag{95}$$

where (a) follows because $(\overline{W}, F) - X^n - Z^n$ form a Markov chain, (b) follows since (U^n, V^n) determine $(F_{\mathrm{u}}, W_{\mathrm{u}}, K_{\mathrm{u}}, F_{\mathrm{v}}, W_{\mathrm{v}})$, and (c) follows since (U^n, V^n, X^n) is i.i.d. and for some $\epsilon_n > 0$ such that $\epsilon_n \to 0$ when $n \to \infty$ because $(F_{\mathrm{v}}, W_{\mathrm{v}}, X^n)$ can reliably recover V^n by (87) because of the Markov chain relation $V^n - X^n - Y^n$ and, similarly, $(F_{\mathrm{u}}, W_{\mathrm{u}}, K_{\mathrm{u}}, V^n, X^n)$ can reliably recover U^n by (88) because of $H(U|V,Y) \geq H(U|V,X)$ that is proved in [21] (Equation (55)) for the Markov chain relation $(V,U) - X - Y$.

Next, we consider the term $H(\overline{W}, F|Z^n)$ in (95) 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] (Equation (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 (95) we obtain

$$I(X^n; W, F|Z^n) \leq n\big([I(U;Z|V) - I(U;Y|V) + \epsilon]^- + I(U;X|Z) + 3\epsilon_n - R_0\big). \tag{96}$$

The equation (92) implicitly assumes that 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 relation $(V,U) - \widetilde{X} - Y$. 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_{u} from the

code construction above and apply one-time padding to the bin index W_{u}, such that we have the encoder output

$$\overline{\overline{W}} = (W_{\mathrm{v}}, W_{\mathrm{u}} + K) \tag{97}$$

where the summation with the private key is in modulo- $2^{nR_{\mathrm{u}}} = 2^{n(I(U;\widetilde{X}|Y,V)+2\epsilon)}$. Thus, one then does not leak any information about W_{u} to the eavesdropper because of the one-time padding step in (97). We then have privacy leakage

$$\begin{aligned} I(X^n; \overline{\overline{W}}, F|Z^n) &= I(X^n; W_{\mathrm{v}}, F|Z^n) \\ &\overset{(a)}{\leq} H(X^n|Z^n) - H(X^n|Z^n, W_{\mathrm{v}}, F_{\mathrm{v}}) + \epsilon'_n \\ &\overset{(b)}{\leq} H(X^n|Z^n) - H(X^n|Z^n, V^n) + \epsilon'_n \\ &\overset{(c)}{=} nI(V;X|Z) + \epsilon'_n \end{aligned} \tag{98}$$

where (a) follows for some ϵ'_n such that $\epsilon'_n \to 0$ when $n \to \infty$ since by (86) F_{u} is almost independent of (V^n, X^n, Z^n); see also [52] (Theorem 1), (b) follows since V^n determines $(F_{\mathrm{v}}, W_{\mathrm{v}})$, and (c) follows because (X^n, Z^n, V^n) are i.i.d.

Note we can reduce the privacy leakage given in (98) 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 relation $U - \widetilde{X} - Y$, since then we can apply one-time padding to both bin indices W_{v} and W_{u} with the sum rate

$$\begin{aligned} R_{\mathrm{v}} + R_{\mathrm{u}} &\overset{(a)}{=} I(V;\widetilde{X}) - I(V;Y) + 2\epsilon + I(U;\widetilde{X}|V) - I(U;Y|V) + 2\epsilon \\ &\overset{(b)}{=} I(U;\widetilde{X}) - I(U;Y) + 4\epsilon \end{aligned} \tag{99}$$

where (a) follows by (90) and (92), and (b) follows from the Markov chain relation $V - U - \widetilde{X} - Y$. Thus, one then does not leak any information about $(W_{\mathrm{v}}, W_{\mathrm{u}})$ to the eavesdropper because of the one-time padding step, so we then obtain the privacy leakage of

$$I(X^n; F|Z^n) = I(X^n; F_{\mathrm{v}}|Z^n) + I(X^n; F_{\mathrm{u}}|Z^n, F_{\mathrm{v}}) \overset{(a)}{\leq} 2\epsilon'_n \tag{100}$$

where (a) follows since by (85) F_{v} is almost independent of (X^n, Z^n) and by (86) F_{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 (94), 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

$$\begin{aligned} I(\widetilde{X}^n; W, F|Z^n) &= I(\widetilde{X}^n; \overline{W}, F|Z^n) - nR_0 \\ &\overset{(a)}{=} H(\overline{W}, F|Z^n) - H(\overline{W}, F|\widetilde{X}^n) - nR_0 \\ &\overset{(b)}{=} H(\overline{W}, F|Z^n) - H(U^n, V^n|\widetilde{X}^n) + H(V^n|\overline{W}, F, \widetilde{X}^n) + H(U^n|V^n, \overline{W}, F, \widetilde{X}^n) \\ &\overset{(c)}{\leq} H(\overline{W}, F|Z^n) - nH(U,V|\widetilde{X}) + 2n\epsilon'_n - nR_0 \\ &\overset{(d)}{\leq} n\big([I(U;Z|V) - I(U;Y|V) + \epsilon]^- + I(U;\widetilde{X}|Z) + 3\epsilon'_n - R_0\big) \end{aligned} \tag{101}$$

where (a) follows from the Markov chain relation $(\overline{W}, F) - \widetilde{X}^n - Z^n$, (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 $(F_{\mathrm{v}}, W_{\mathrm{v}}, \widetilde{X}^n)$ can reliably recover V^n by (87) due to the Markov chain relation $V^n - \widetilde{X}^n - Y^n$ and, similarly, $(F_{\mathrm{u}}, W_{\mathrm{u}}, K_{\mathrm{u}}, V^n, \widetilde{X}^n)$ can reliably recover U^n by (88) due to $H(U|V,Y) \geq H(U|V,\widetilde{X})$ that can

be proved as in [21] (Equation (55)) for the Markov chain relation $(V,U)-\widetilde{X}-Y$, and (d) follows by applying the six decodability results in [21] (Section V-A) that are applied to (95) with the final result in (96) 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_{u} and apply one-time padding as in (97), such that no information about W_{u} is leaked to the eavesdropper, we have

$$\begin{aligned} I(\widetilde{X}^n;\overline{\overline{W}},F|Z^n) &= I(\widetilde{X}^n;W_{\mathrm{v}},F|Z^n) \\ &\overset{(a)}{\leq} H(\widetilde{X}^n|Z^n)-H(\widetilde{X}^n|Z^n,W_{\mathrm{v}},F_{\mathrm{v}})+\epsilon_n' \\ &\overset{(b)}{\leq} H(\widetilde{X}^n|Z^n)-H(\widetilde{X}^n|Z^n,V^n)+\epsilon_n' \\ &\overset{(c)}{=} nI(V;\widetilde{X}|Z)+\epsilon_n' \end{aligned} \tag{102}$$

where (a) follows because by (86) F_{u} is almost independent of $(V^n,\widetilde{X}^n,Z^n)$, (b) follows since V^n determines $(F_{\mathrm{v}},W_{\mathrm{v}})$, 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 $(W_{\mathrm{v}},W_{\mathrm{u}})$, as in the privacy leakage analysis above. We then have the secrecy leakage of

$$I(\widetilde{X}^n;F|Z^n) = I(\widetilde{X}^n;F_{\mathrm{v}}|Z^n)+I(\widetilde{X}^n;F_{\mathrm{u}}|Z^n,F_{\mathrm{v}}) \overset{(a)}{\leq} 2\epsilon_n' \tag{103}$$

where (a) follows since by (85) F_{v} is almost independent of $(\widetilde{X}^n,Z^n)$ and by (86) F_{u} is almost independent of $(V^n,\widetilde{X}^n,Z^n)$.

Suppose that public indices F are generated uniformly at random, and the encoder generates (V^n,U^n) according to $P_{V^nU^n|\widetilde{X}^nF_{\mathrm{v}}F_{\mathrm{u}}}$ that can be obtained from the proposed binning scheme above to compute the bins W_{v} from V^n and W_{u} from U^n, respectively. Such a procedure results in a joint probability distribution almost equal to $P_{VU\widetilde{X}XYZ}$ fixed above [51] (Section 1.6). 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 [53] (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 $n\to\infty$. □

6.2. 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 $(R_{\mathrm{w}},R_{\mathrm{s}},R_\ell,D)$ 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 relation $V_i-U_i-\widetilde{X}_i-X_i-(Y_i,Z_i)$ by definition of the source statistics. We have

$$\begin{aligned} D+\delta_n &\overset{(a)}{\geq} \mathbb{E}\Big[d\Big(\widetilde{X}^n,\widehat{\widetilde{X}}^n(Y^n,W,K)\Big)\Big] \\ &\overset{(b)}{\geq} \mathbb{E}\Big[d\Big(\widetilde{X}^n,\widehat{\widetilde{X}}^n(Y^n,W,K,X^{i-1},Z^{i-1})\Big)\Big] \\ &\overset{(c)}{=} \mathbb{E}\Big[d\Big(\widetilde{X}^n,\widehat{\widetilde{X}}^n(Y_i^n,W,K,X^{i-1},Z^{i-1})\Big)\Big] \\ &\overset{(d)}{=} \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\Big[d\Big(\widetilde{X}_i,\widehat{\widetilde{X}}_i(U_i,Y_i)\Big)\Big] \end{aligned} \tag{104}$$

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 relation

$$Y^{i-1}-(Y_i^n,X^{i-1},Z^{i-1},W,K)-\widetilde{X}^n \tag{105}$$

and (d) follows from the definition of U_i.

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

$$\begin{aligned} n(R_{\mathrm{w}} + \delta_n) &\overset{(a)}{\geq} \log |\mathcal{W}| \\ &\geq H(W|Y^n, K) - H(W|Y^n, K, \widetilde{X}^n) \qquad (106) \\ &\overset{(b)}{=} \sum_{i=1}^{n} I(W; \widetilde{X}_i | \widetilde{X}^{i-1}, Y_{i+1}^n, Z^{i-1}, K, Y_i) \qquad (107) \\ &\overset{(c)}{=} \sum_{i=1}^{n} I(\widetilde{X}^{i-1}, Y_{i+1}^n, Z^{i-1}, K, W; \widetilde{X}_i | Y_i) \\ &\overset{(d)}{\geq} \sum_{i=1}^{n} I(X^{i-1}, Y_{i+1}^n, Z^{i-1}, K, W; \widetilde{X}_i | Y_i) \\ &\overset{(e)}{=} \sum_{i=1}^{n} I(U_i; \widetilde{X}_i | Y_i) \qquad (108) \end{aligned}$$

where (a) follows by (1), (b) follows from the Markov chain relation

$$(Y^{i-1}, X^{i-1}, Z^{i-1}) - (\widetilde{X}^{i-1}, Y_i^n, K) - (\widetilde{X}_i, W) \qquad (109)$$

(c) follows because $(\widetilde{X}_i, Y_i)$ are independent of $(\widetilde{X}^{i-1}, Y_{i+1}^n, Z^{i-1}, K)$, (d) follows by applying the data processing inequality to the Markov chain relation in (109), and (e) follows from the definition of U_i.

Privacy Leakage Rate: We obtain

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

$$\begin{aligned}
&\overset{(g)}{=} \sum_{i=1}^{n} \Big[I(V_i;Y_i) - I(V_i;Z_i) - R_0 + I(U_i,V_i;X_i) - I(U_i,V_i;Y_i) \Big] \\
&= \sum_{i=1}^{n} \Big[-I(U_i,V_i;Z_i) - R_0 + I(U_i,V_i;X_i) + (I(U_i;Z_i|V_i) - I(U_i;Y_i|V_i)) \Big] \\
&\overset{(h)}{\geq} \sum_{i=1}^{n} \Big[I(U_i;X_i|Z_i) - R_0 + [I(U_i;Z_i|V_i) - I(U_i;Y_i|V_i)]^{-} \Big] \qquad (110)
\end{aligned}$$

where (a) follows by (3) and from the Markov chain relation $W - X^n - Z^n$, (b) follows since K is independent of (X^n, Y^n), (c) follows from the Markov chain relation $(W,K) - X^n - Y^n$, (d) follows because $H(K) = nR_0$ and from Csiszár's sum identity [54], (e) follows from the Markov chain relations

$$Z^{i-1} - (X^{i-1}, Y_{i+1}^n, K) - (X_i, W) \qquad (111)$$

$$Z^{i-1} - (X^{i-1}, Y_{i+1}^n, K) - (Y_i, W) \qquad (112)$$

(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 relation $V_i - U_i - X_i - Z_i$.

Next, we provide the matching converse for the privacy leakage rate in (98), which is achieved when $R_0 \geq I(U;\widetilde{X}|Y,V)$. We have

$$\begin{aligned}
n(R_\ell + \delta_n) &\overset{(a)}{\geq} H(X^n|Z^n) - H(X^n|Z^n,W) \\
&\overset{(b)}{=} H(X^n|Z^n) - \sum_{i=1}^{n} H(X_i|Z_i, Z^{i-1}, X_{i+1}^n, W, Y_{i+1}^n) \\
&\overset{(c)}{=} H(X^n|Z^n) - \sum_{i=1}^{n} H(X_i|Z_i, V_i, X_{i+1}^n) \\
&\overset{(d)}{\geq} \sum_{i=1}^{n} [H(X_i|Z_i) - H(X_i|Z_i,V_i)] \\
&= \sum_{i=1}^{n} I(V_i;X_i|Z_i) \qquad (113)
\end{aligned}$$

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

$$(Z_{i+1}^n, Y_{i+1}^n) - (X_{i+1}^n, W, Z^i) - X_i \qquad (114)$$

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

The matching converse for the privacy leakage rate in (100), 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

$$\begin{aligned}
&n(R_s + \delta_n) \\
&\overset{(a)}{\geq} [I(W;Y^n) - I(W;Z^n)] + [I(W;\widetilde{X}^n) - I(W;Y^n)] \\
&\overset{(b)}{=} [I(W;Y^n) - I(W;Z^n)] + I(W;\widetilde{X}^n|K) - I(K;\widetilde{X}^n|W) - I(W;Y^n|K) + I(K;Y^n|W) \\
&\overset{(c)}{=} [I(W;Y^n) - I(W;Z^n)] + [I(W;\widetilde{X}^n|K) - I(W;Y^n|K)] - I(K;\widetilde{X}^n|W,Y^n) \\
&\overset{(d)}{\geq} \sum_{i=1}^{n} \Big[I(W;Y_i|Y_{i+1}^n) - I(W;Z_i|Z^{i-1}) \Big] + I(W;\widetilde{X}^n|Y^n,K) - H(K)
\end{aligned}$$

$$
\begin{aligned}
&\overset{(e)}{=} \sum_{i=1}^{n} \Big[I(W;Y_i|Y_{i+1}^n, Z^{i-1}) - I(W;Z_i|Z^{i-1}, Y_{i+1}^n) - R_0 \Big] \\
&\quad + nH(\widetilde{X}|Y) - \sum_{i=1}^{n} H(\widetilde{X}_i|Y_i, Y_{i+1}^n, W, K, \widetilde{X}^{i-1}) \\
&\overset{(f)}{\geq} \sum_{i=1}^{n} \Big[I(W, Y_{i+1}^n, Z^{i-1}; Y_i) - I(W, Z^{i-1}, Y_{i+1}^n; Z_i) - R_0 \Big] \\
&\quad + nH(\widetilde{X}|Y) - \sum_{i=1}^{n} H(\widetilde{X}_i|Y_i, Y_{i+1}^n, W, K, X^{i-1}, Z^{i-1}) \\
&\overset{(g)}{=} \sum_{i=1}^{n} \Big[I(V_i;Y_i) - I(V_i;Z_i) - R_0 \Big] + nH(\widetilde{X}|Y) - \sum_{i=1}^{n} H(\widetilde{X}_i|Y_i, U_i, V_i) \\
&\overset{(h)}{=} \sum_{i=1}^{n} \Big[I(V_i;Y_i) - I(V_i;Z_i) - R_0 \Big] + \sum_{i=1}^{n} \Big[I(U_i, V_i; \widetilde{X}_i) - I(U_i, V_i; Y_i) \Big] \\
&= \sum_{i=1}^{n} \Big[- I(U_i, V_i; Z_i) - R_0 + I(U_i, V_i; \widetilde{X}_i) + (I(U_i; Z_i|V_i) - I(U_i; Y_i|V_i)) \Big] \\
&\overset{(i)}{\geq} \sum_{i=1}^{n} \Big[I(U_i; \widetilde{X}_i|Z_i) - R_0 + [I(U_i; Z_i|V_i) - I(U_i; Y_i|V_i)]^- \Big]
\end{aligned} \tag{115}
$$

where (a) follows by (2) and from the Markov chain relation $W - \widetilde{X}^n - Z^n$, (b) follows because K is independent of $(\widetilde{X}^n, Y^n)$, (c) and (d) follow from the Markov chain relation $(W, K) - \widetilde{X}^n - Y^n$, (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 relation

$$
Y^{i-1} - (\widetilde{X}^{i-1}, W, K, Y_{i+1}^n, Y_i) - \widetilde{X}_i \tag{116}
$$

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

$$
(X^{i-1}, Z^{i-1}) - (\widetilde{X}^{i-1}, W, K, Y_{i+1}^n, Y_i) - \widetilde{X}_i \tag{117}
$$

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

Next, the matching converse for the secrecy leakage rate in (102), achieved when $R_0 \geq I(U; \widetilde{X}|Y, V)$, is provided.

$$
\begin{aligned}
n(R_s + \delta_n) &\overset{(a)}{\geq} H(\widetilde{X}^n|Z^n) - H(\widetilde{X}^n|Z^n, W) \\
&\overset{(b)}{\geq} H(\widetilde{X}^n|Z^n) - \sum_{i=1}^{n} H(\widetilde{X}_i|Z_i, Z^{i-1}, \widetilde{X}_{i+1}^n, W, Y_{i+1}^n) \\
&\overset{(c)}{=} H(\widetilde{X}^n|Z^n) - \sum_{i=1}^{n} H(\widetilde{X}_i|Z_i, V_i, \widetilde{X}_{i+1}^n) \\
&\overset{(d)}{\geq} \sum_{i=1}^{n} [H(\widetilde{X}_i|Z_i) - H(\widetilde{X}_i|Z_i, V_i)] = \sum_{i=1}^{n} I(V_i; \widetilde{X}_i|Z_i)
\end{aligned} \tag{118}
$$

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

$$
(Z_{i+1}^n, Y_{i+1}^n) - (\widetilde{X}_{i+1}^n, W, Z^i) - \widetilde{X}_i \tag{119}
$$

(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 (103), 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 $Q \sim \mathrm{Unif}[1:n]$ 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

$$(Q, V) - U - \widetilde{X} - X - (Y, Z) \tag{120}$$

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

Cardinality Bounds: We use the support lemma [54] (Lemma 15.4) for the cardinality bound proofs, which is a standard step, so we omit the proof.

Author Contributions: Conceptualization, O.G., R.F.S., H.B. and H.V.P.; Methodology, O.G. and H.V.P.; Software, H.B.; Validation, R.F.S.; Formal analysis, O.G., R.F.S., H.B. and H.V.P.; Resources, H.B.; Data curation, O.G. and R.F.S.; Writing—original draft, O.G.; Writing—review & editing, R.F.S., H.B. and H.V.P.; Project administration, R.F.S. and H.V.P.; Funding acquisition, R.F.S. and H.B. All authors have read and agreed to the published version of the manuscript.

Funding: O. Günlü was supported by the ZENITH Research and Career Development Fund and the ELLIIT funding endowed by the Swedish government. R. F. Schaefer was supported in part by the German Federal Ministry of Education and Research (BMBF) within the national initiative for Post-Shannon Communication (NewCom) under grant no. 16KIS1004 and the National Initiative for 6G Communication Systems through the Research Hub 6G-life under grant no. 16KISK001K. H. Boche was supported in part by the BMBF within the National Initiative for 6G Communication Systems through the Research Hub 6G-life under grant no. 16KISK002 and within the national initiative for Information Theory for Post Quantum Crypto "Quantum Token Theory and Applications—QTOK" under grant no. 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 grant no. CCF-1908308.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Slepian, D.; Wolf, J. Noiseless coding of correlated information sources. *IEEE Trans. Inf. Theory* **1973**, *19*, 471–480. [CrossRef]
2. Gamal, A.E.; Kim, Y.H. *Network Information Theory*; Cambridge University Press: Cambridge, UK, 2011.
3. Orlitsky, A.; Roche, J.R. Coding for computing. *IEEE Trans. Inf. Theory* **2001**, *47*, 903–917. [CrossRef]
4. Günlü, O. Function computation under privacy, secrecy, distortion, and communication constraints. *Entropy* **2022**, *24*, 110. [CrossRef] [PubMed]
5. Prabhakaran, V.; Ramchandran, K. On secure distributed source coding. In Proceedings of the 2007 IEEE Information Theory Workshop, Solstrand, Norway, 1–6 July 2007; pp. 442–447.
6. Gündüz, D.; Erkip, E.; Poor, H.V. Secure lossless compression with side information. In Proceedings of the 2008 IEEE Information Theory Workshop, Porto, Portugal, 5–9 May 2008; pp. 169–173.
7. Tandon, R.; Ulukus, S.; Ramchandran, K. Secure source coding with a helper. *IEEE Trans. Inf. Theory* **2013**, *59*, 2178–2187. [CrossRef]
8. Gündüz, D.; Erkip, E.; Poor, H.V. Lossless compression with security constraints. In Proceedings of the 2008 IEEE Information Theory Workshop, Porto, Portugal, 5–9 May 2008; pp. 111–115.
9. Luh, W.; Kundur, D. Distributed secret sharing for discrete memoryless networks. *IEEE Trans. Inf. Forensics Secur.* **2008**, *3*, 1–7. [CrossRef]
10. Kittichokechai, K.; Chia, Y.K.; Oechtering, T.J.; Skoglund, M.; Weissman, T. Secure source coding with a public helper. *IEEE Trans. Inf. Theory* **2016**, *62*, 3930–3949. [CrossRef]
11. Salimi, S.; Salmasizadeh, M.; Aref, M.R. Generalised secure distributed source coding with side information. *IET Commun.* **2010**, *4*, 2262–2272. [CrossRef]
12. Naghibi, F.; Salimi, S.; Skoglund, M. The CEO problem with secrecy constraints. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 1234–1249. [CrossRef]

13. Yamamoto, H. Coding theorems for Shannon's cipher system with correlated source outputs, and common information. *IEEE Trans. Inf. Theory* **1994**, *40*, 85–95. [CrossRef]
14. Ghourchian, H.; Stavrou, P.A.; Oechtering, T.J.; Skoglund, M. Secure source coding with side-information at decoder and shared key at encoder and decoder. In Proceedings of the 2021 IEEE Information Theory Workshop (ITW) 2021, Virtual, 17–21 October 2021; pp. 1–6.
15. Maurer, U.M. Secret key agreement by public discussion from common information. *IEEE Trans. Inf. Theory* **1993**, *39*, 2733–2742. [CrossRef]
16. Ahlswede, R.; Csiszár, I. Common randomness in information theory and cryptography—Part I: Secret sharing. *IEEE Trans. Inf. Theory* **1993**, *39*, 1121–1132. [CrossRef]
17. Yao, A.C. Protocols for secure computations. In Proceedings of the 3rd Annual Symposium on Foundations of Computer Science (SFCS 1982), Chicago, IL, USA, 3–5 November 1982; pp. 160–164.
18. Yao, A.C. How to generate and exchange secrets. In Proceedings of the 3rd Annual Symposium on Foundations of Computer Science (SFCS 1982), Chicago, IL, USA, 3–5 November 1982; pp. 162–167.
19. Bloch, M.; Günlü, O.; Yener, A.; Oggier, F.; Poor, H.V.; Sankar, L.; Schaefer, R.F. An overview of information-theoretic security and privacy: Metrics, limits and applications. *IEEE J. Sel. Areas Inf. Theory* **2021**, *2*, 5–22. [CrossRef]
20. Günlü, O.; Kramer, G. Privacy, secrecy, and storage with multiple noisy measurements of identifiers. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2872–2883. [CrossRef]
21. Günlü, O.; Bloch, M.; Schaefer, R.F. Secure multi-function computation with private remote sources. *arXiv* **2021**, arXiv:2106.09485.
22. Berger, T. *Rate Distortion Theory: A Mathematical Basis for Data Compression*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1971.
23. Permuter, H.; Weissman, T. Source coding with a side information "Vending Machine". *IEEE Trans. Inf. Theory* **2011**, *57*, 4530–4544. [CrossRef]
24. Berger, T.; Zhang, Z.; Viswanathan, H. The CEO problem. *IEEE Trans. Inf. Theory* **1996**, *42*, 887–902. [CrossRef]
25. Günlü, O. Key Agreement with Physical Unclonable Functions and Biometric Identifiers. Ph.D. Thesis, Technical University of Munich, Munich, Germany , February 2019.
26. Ignatenko, T.; Willems, F.M.J. Biometric systems: Privacy and secrecy aspects. *IEEE Trans. Inf. Forensics Secur.* **2009**, *4*, 956–973. [CrossRef]
27. Lai, L.; Ho, S.W.; Poor, H.V. Privacy-security trade-offs in biometric security systems - Part I: Single use case. *IEEE Trans. Inf. Forensics Secur.* **2011**, *6*, 122–139. [CrossRef]
28. Kusters, L.; Günlü, O.; Willems, F.M. Zero secrecy leakage for multiple enrollments of physical unclonable functions. In Proceedings of the 2018 Symposium on Information Theory and Signal Processing in the Benelux, Enschede, The Netherlands, 31 May–1 June 2018; pp. 119–127.
29. Lai, L.; Ho, S.W.; Poor, H.V. Privacy-security trade-offs in biometric security systems—Part II: Multiple use case. *IEEE Trans. Inf. Forensics Secur.* **2011**, *6*, 140–151. [CrossRef]
30. Günlü, O. Multi-Entity and Multi-Enrollment Key Agreement with Correlated Noise. *IEEE Trans. Inf. Forensics Secur.* **2021**, *16*, 1190–1202. [CrossRef]
31. Günlü, O.; Schaefer, R.F.; Boche, H.; Poor, H.V. Secure and private source coding with private key and decoder side information. *arXiv* **2022**, arXiv:2205.05068.
32. Tu, W.; Lai, L. On function computation with privacy and secrecy constraints. *IEEE Trans. Inf. Theory* **2019**, *65*, 6716–6733. [CrossRef]
33. Villard, J.; Piantanida, P. Secure multiterminal source coding with side information at the eavesdropper. *IEEE Trans. Inf. Theory* **2013**, *59*, 3668–3692. [CrossRef]
34. Bross, S.I. Secure cooperative source-coding with side information at the eavesdropper. *IEEE Trans. Inf. Theory* **2016**, *62*, 4544–4558. [CrossRef]
35. Ekrem, E.; Ulukus, S. Secure lossy source coding with side information. In Proceedings of the 2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 28–30 September 2011; pp. 1098–1105.
36. Körner, J.; Marton, K. Comparison of two noisy channels. *Topics Inf. Theory* **1977**, 411–423.
37. Bergmans, P. A simple converse for broadcast channels with additive white Gaussian noise (Corresp.). *IEEE Trans. Inf. Theory* **1974**, *20*, 279–280. [CrossRef]
38. Günlü, O.; Schaefer, R.F.; Poor, H.V. Biometric and Physical Identifiers with Correlated Noise for Controllable Private Authentication. *arXiv* **2020**, arXiv:2001.00847.
39. Wyner, A.D.; Ziv, J. A theorem on the entropy of certain binary sequences and applications: Part I. *IEEE Trans. Inf. Theory* **1973**, *19*, 769–772. [CrossRef]
40. Watanabe, S.; Oohama, Y. Secret key agreement from correlated Gaussian sources by rate limited public communication. *IEICE Trans. Fundam. Electron., Commun. Comp. Sci.* **2010**, *93*, 1976–1983. [CrossRef]
41. Willems, F.M.; Ignatenko, T. Quantization effects in biometric systems. In Proceedings of the 2009 Information Theory and Applications Workshop, San Diego, CA, USA, 27 January–1 February 2009; pp. 372–379.
42. Yachongka, V.; Yagi, H.; Oohama, Y. Secret key-based authentication with passive eavesdropper for scalar Gaussian sources. *arXiv* **2022**, arXiv:2202.10018.
43. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2012.

44. Maes, R. An accurate probabilistic reliability model for silicon PUFs. In *International Conference on Cryptographic Hardware and Embedded Systems*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 73–89.
45. Anantharam, V. *Lecture Notes in Stochastic Estimation and Control: Jointly Gaussian Random Variables*; University California Berkeley: Berkeley, CA, USA, 2007.
46. Wyner, A.; Ziv, J. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Inf. Theory* **1976**, *22*, 1–10. [CrossRef]
47. Chayat, N.; Shamai, S. Extension of an entropy property for binary input memoryless symmetric channels. *IEEE Trans. Inf. Theory* **1989**, *35*, 1077–1079. [CrossRef]
48. Günlü, O.; Kramer, G.; Skórski, M. Privacy and secrecy with multiple measurements of physical and biometric identifiers. In Proceedings of the 2015 IEEE Conference on Communications and Network Security (CNS), Florence, Italy, 28–30 September 2015; pp. 89–94.
49. Yassaee, M.H.; Aref, M.R.; Gohari, A. Achievability proof via output statistics of random binning. *IEEE Trans. Inf. Theory* **2014**, *60*, 6760–6786. [CrossRef]
50. Renes, J.M.; Renner, R. Noisy channel coding via privacy amplification and information reconciliation. *IEEE Trans. Inf. Theory* **2011**, *57*, 7377–7385. [CrossRef]
51. Bloch, M. *Lecture Notes in Information-Theoretic Security*; Georgia Institute of Technology: Atlanta, GA, USA, 2018.
52. Holenstein, T.; Renner, R. On the randomness of independent experiments. *IEEE Trans. Inf. Theory* **2011**, *57*, 1865–1871. [CrossRef]
53. Bloch, M.; Barros, J. *Physical-Layer Security*; Cambridge University Press: Cambridge, UK, 2011.
54. Csiszár, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2011.