Private Remote Sources for Secure Multi-Function Computation

Onur Günlü, *Member, IEEE*, Matthieu Bloch, *Senior Member, IEEE*, and Rafael F. Schaefer, *Senior Member, IEEE*

***Abstract*—We consider a distributed function computation problem in which parties observing noisy versions of a remote source facilitate the computation of a function of their observations at a fusion center through public communication. The distributed function computation is subject to constraints, including not only reliability and storage but also secrecy and privacy. Specifically, 1) the function computed should remain *secret* from an eavesdropper observing the public communication and correlated observations, measured in terms of the information leaked about the arguments of the function, to ensure secrecy regardless of the exact function used; 2) the remote source should remain *private* from the eavesdropper and the fusion center, measured in terms of the information leaked about the remote source itself. We derive the exact rate regions for lossless and lossy single-function computation and illustrate the lossy single-function computation rate region for an information bottleneck example, in which the optimal auxiliary random variables are characterized for binary-input symmetric-output channels. We extend the approach to lossless and lossy asynchronous multiple-function computations with joint secrecy and privacy constraints, in which case inner and outer bounds for the rate regions that differ only in the Markov chain conditions imposed are characterized.**

***Index Terms*—secure multiple function computation, private remote source, lossy function computation, information bottleneck, rate-limited public communication.**

I. Introduction

Consider a scenario in which multiple terminals that observe dependent random sequences want to compute a function of their sequences by exchanging messages through public communication links [2], [3]. One application for which this distributed function computation problem is relevant is network function virtualization [4] via, e.g., software defined networking. It is not always necessary for the terminal computing the function, called *fusion center*, to observe the exact sequences [5]. This fact allows one to reduce the public communication rate, also called *storage rate*, required for reliable function computations by using, e.g., distributed lossless source coding techniques [6]. Furthermore, if the function to compute only requires recovering a distorted version of the original sequence, distributed lossy source coding methods [7] further reduce the amount of public storage. This is useful for resource-limited networks such as Internet-of-Things (IoT) devices that make aggregated decisions using lightweight mechanisms [5], [8]–[12]; see [13]–[17] for various extensions of the basic function computation problem with reliability and storage constraints.

Reliable function computation and small public storage constraints have also been combined with *secrecy* constraints, requiring that the computed function outputs be hidden from an eavesdropper [18]. In addition to the public messages exchanged between terminals, the eavesdropper is considered to have access to a random sequence correlated with other sequences. Various extensions of the basic secure function computation or distributed source coding problems have been analyzed in the literature [19]–[25]. Furthermore, a *privacy* constraint has been added in [26] to the problem. The main difference between *secrecy* and *privacy* is that secrecy leakage is measured with respect to the functions computed while privacy leakage is measured with respect to the source sequences themselves. A privacy leakage analysis provides an upper bound on the secrecy leakage of future function computations involving the terminals already participating in earlier function computations [27], [28]. This is because the information leaked about the sequence of a terminal might leak information about another function computed by using the same sequence. We extend [26] by considering separate privacy constraints on the source of the random sequence of the *transmitting terminal* that sends a public message to the fusion center.

A common assumption in the literature is that sequences observed by all terminals are distributed according to a joint probability distribution. However, the correlated random sequences observed by terminals in a network generally stem from a common source of information, e.g., some sensor location information transmitted through the network before the next function computation starts, distorted versions of which are distributed within the network. Thus, we posit that there exists a common true source, called the remote source, hidden from all terminals and of which the observed sequences are noisy versions. Such a remote source model allows a terminal to combine multiple observed sequences to obtain a single "higher quality" random sequence, which is similar to applying maximal ratio combining over an additive white Gaussian noise (AWGN) channel. This approach is

Manuscript received June 17, 2021; revised December 1, 2021; accepted June 2, 2022. This work has been supported in part by the German Research Foundation (DFG) under the Grant SCHA 1944/9-1 and in part by the National Science Foundation (NSF) under the Grant CCF 1955401. Parts of this work were presented at the IEEE International Symposium on Information Theory 2021 [1]. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Himanshu Tyagi (*Corresponding Author: Onur Günlü*).

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

M. Bloch is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 (email: matthieu.bloch@ece.gatech.edu).

thus useful to model the quality differences between random sequences observed by different terminals. If the function computation network is mistakenly modeled with a visible source model, the code construction designed for the assumed visible source model might result in unnoticed secrecy leakage and reduction in computation reliability, as illustrated in [28] for key agreement.

Noisy measurements of a hidden source are generally modeled as observations through broadcast channels (BCs) [29] to have a generic measurement model that allows noise components at different terminals to be correlated [30], [31]. Such a hidden source model is proposed and motivated in [32] for authentication problems and in [30], [33] for secret-key agreement problems with a privacy constraint. As we detail in Section II, such a hidden source model results in two different privacy leakage constraints measured with respect to the hidden source, which is different from the single privacy leakage constraint considered in [26] measured with respect to the random sequence observed by the transmitting terminal. Furthermore, the equivocation of the source is commonly used in the literature to measure the secrecy leakage, which results in rate bounds with conditional entropy terms. By replacing the equivocation with the mutual information terms, we obtain rate regions with simpler notation and easier interpretations.

We consider two function computation settings. The first setting imposes a reliable (*lossless*) computation of the function of interest and the other one allows a fixed level of distortion between the computed function and the actual function output (*lossy function computation*) [26]. These settings address different applications. For instance, the lossless function computation setting might model user/terminal identification, where the exact identifier recovery is necessary; in contrast, the lossy function computation setting might model user/terminal authentication, where a set of users whose computed functions are close to a pre-defined value are authenticated. We bound the error probability for the reliable function computation task for the lossless setting and the expected distortion for the lossy setting, respectively, which require different proof steps. We exactly characterize the rate regions for both settings when a single function is computed.

We further extend the function computation with privacy and secrecy problem by considering multiple function computations with *joint secrecy and privacy constraints* on all terminals involved in any function computation task. This extension allows one to measure the total amount of information leaked to an eavesdropper about all computed functions within a network. This extension also allows one to correctly characterize the privacy leakage to an eavesdropper, i.e., the amount of information about the hidden source leaked to an eavesdropper who might observe all public messages and all side information obtained during all (not necessarily synchronous) function computations within the same network. Multiple function computations with joint secrecy and privacy constraints are closely related to the multi-entity and multi-enrollment key agreement problems in [34], where the noisy measurements of the same hidden source are used for multiple key agreements. Both lossless and lossy function computation settings are analyzed to provide inner and outer bounds for the multi-function rate regions, for which only the imposed Markov chains differ.

A. Summary of Contributions

Our problem formulation introduces one secrecy and two privacy constraints, in addition to reliability (or distortion) and storage constraints, to the single function computation problem to characterize the resulting rate regions. These results are strict extensions of [26] as we consider a remote source common to all terminals with side information sequences that are noisy measurements of the remote source. Furthermore, we also consider multiple asynchronous function computations within the same network with joint secrecy and privacy constraints over all terminals involved in any function computation. A summary of the main contributions is as follows.

- We derive the rate region for lossless single-function computation with secrecy and privacy constraints. The remote source model we consider corresponds to a physically-degraded BC and when the transmitting terminal observes the remote (noiseless) source outputs, the model reduces to a semi-deterministic BC. Furthermore, we show that convexification with a time-sharing random variable is necessary.
- We next consider the lossless multi-function computations where a finite number J of functions are computed from different noisy measurements (observed by different terminals) of the same remote source asynchronously. We impose one secrecy and privacy constraints that consider the total leakage in the network, i.e., they are joint constraints for all parties involved in any function computation. We propose inner and outer bounds for the multi-function rate region that differ only in the Markov chain conditions imposed on the auxiliary random variables. The rate regions include both separate constraints for each terminal and joint constraints for all terminals.
- All inner and outer bounds for the lossless single- and multi-function computations are extended to the corresponding lossy settings. Similar to the lossless case, we characterize the lossy rate region for the single-function computation, and we provide inner and outer bounds for the multi-function computations that differ only in the Markov chains imposed.
- We evaluate the rate region for a lossy single-function computation problem, in which the measurement channel of the eavesdropper is physically-degraded compared to the measurement channel of the fusion center. We solve an information bottleneck problem to obtain the rate region boundary tuples.

B. Organization

In Section II, we introduce the lossless or lossy and single-function or multi-function computation problems with a remote source. In Section III, we present the rate regions for the lossless and lossy single-function computation in addition to inner and outer bounds with different Markov chains for the lossless and lossy multi-function computations for any finite number of functions. In Section IV, we solve an information

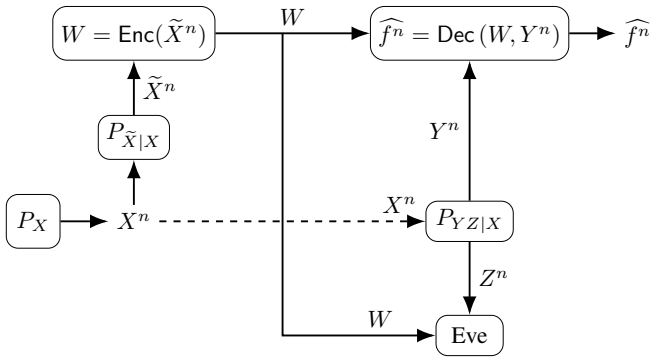

Fig. 1. Noisy measurements of a remote source used to compute a function securely and privately with the help of a public communication link.

bottleneck problem to illustrate the rate region for the lossy single-function computation problem. In Section V, we conclude the paper.

C. Notation

Upper case letters represent random variables and lower case letters their realizations. A superscript denotes a sequence of variables, e.g., $X^n = X_1, X_2, \ldots, X_i, \ldots, X_n$, and a subscript i denotes the position of a variable in a sequence. A random variable X has probability distribution P_X. Calligraphic letters such as $\mathcal{X}$ denote sets, set sizes are written as $|\mathcal{X}|$ and their complements as $\mathcal{X}^c$. $[1:J]$ denotes the set $\{1, 2, \ldots, J\}$ for an integer $J \geq 1$ and $[1:J]\backslash\{j\}$ denotes the set $\{1, 2, \ldots, j-1, j+1, \ldots, J\}$ for any $j \in [1:J]$. $H_b(x) = -x\log x - (1-x)\log(1-x)$ is the binary entropy function, where logarithms are to the base 2, and $H_b^{-1}(\cdot)$ denotes its inverse with range $[0, 0.5]$. A binary symmetric channel (BSC) with crossover probability p is denoted by BSC(p). $X \sim \text{Bern}(\alpha)$ is a binary random variable with $\Pr[X = 1] = \alpha$.

II. Problem Definitions

A. Lossless Single-Function Computation

Consider the function computation model illustrated in Fig. 1. Three terminals obtain noisy observations $\widetilde{X}^n, Y^n, Z^n$, respectively, of a single i.i.d. remote source X^n, through a memoryless channel with transition probability $p_{\widetilde{X}|X}p_{YZ|X}$. The source alphabet $\mathcal{X}$ and measurement alphabets $\widetilde{\mathcal{X}}, \mathcal{Y}, \mathcal{Z}$ are finite sets. The objective is for the terminal observing $\widetilde{X}^n$ to transmit a message $W = \mathsf{Enc}(\widetilde{X}^n)$ over a public channel and to enable the terminal observing Y^n to compute a function $f^n(\widetilde{X}^n, Y^n)$ such that

$$f^n(\widetilde{X}^n, Y^n) = \{f(\widetilde{X}_i, Y_i)\}_{i=1}^n. \tag{1}$$

The terminal observing Z^n and obtaining W through the public channel is treated as an eavesdropper (Eve).

Since $P_{\widetilde{X}XYZ}$ is fixed, the separate measurement channels $P_{\widetilde{X}|X}$ and $P_{YZ|X}$ in Fig. 1 can be modeled as a physically-degraded BC with transition probability $P_{XYZ|\widetilde{X}} = P_{X|\widetilde{X}}P_{YZ|X}$ and with fixed input probability distribution $P_{\widetilde{X}}$. For such a BC, the case of a noiseless measurement for which $\widetilde{X}^n = X^n$ can be treated as a semi-deterministic BC.

Definition 1. A tuple $(R_\text{s}, R_\text{w}, R_{\ell,\text{Dec}}, R_{\ell,\text{Eve}}) \in \mathbb{R}^4_{\geq 0}$ is *achievable* if, for any $\delta > 0$, there exist $n \geq 1$, an encoder, and a decoder such that

$$\Pr\left[f^n(\widetilde{X}^n, Y^n) \neq \widehat{f^n}\right] \leq \delta \qquad \text{(reliability)} \tag{2}$$

$$\frac{1}{n}I(\widetilde{X}^n, Y^n; W|Z^n) \leq R_\text{s} + \delta \qquad \text{(secrecy)} \tag{3}$$

$$\frac{1}{n}\log\left|\mathcal{W}\right| \leq R_\text{w} + \delta \qquad \text{(storage)} \tag{4}$$

$$\frac{1}{n}I(X^n; W|Y^n) \leq R_{\ell,\text{Dec}} + \delta \qquad \text{(privacyDec)} \tag{5}$$

$$\frac{1}{n}I(X^n; W|Z^n) \leq R_{\ell,\text{Eve}} + \delta \qquad \text{(privacyEve)}. \tag{6}$$

The region $\mathcal{R}$ is the closure of the set of all achievable tuples. $\lozenge$

Note that the metric $I(f^n(\widetilde{X}^n, Y^n); W|Z^n)$ might seem a more natural way to measure the information leakage to the eavesdropper who observes (W, Z^n) of the computed function $f^n(\cdot,\cdot)$. However, the analysis of this metric depends on the specific properties of the function $f(\cdot,\cdot)$. Since the data-processing inequality ensures that $I(f^n(\widetilde{X}^n, Y^n); W|Z^n) \leq I(\widetilde{X}^n, Y^n; W|Z^n)$ for all functions $f(\cdot,\cdot)$ with equality if $f(\cdot,\cdot)$ is a bijective mapping, we instead consider the metric in (3). The analysis then does not depend on the computed function $f(\cdot,\cdot)$ and provides a valid upper bound on the proper secrecy-leakage rate metric for any $f(\cdot,\cdot)$. Since $I(\widetilde{X}^n, Y^n; W|Z^n) = I(\widetilde{X}^n; W|Z^n)$ because of the Markov chain $W - \widetilde{X}^n - (Y^n, Z^n)$, the equivocation $H(\widetilde{X}^n|W, Z^n)$ considered in previous works [26] captures the same secrecy leakage as (3). Furthermore, the privacy leakage metrics in (5) and (6) measure the information leakage about the remote source to the decoder and eavesdropper, respectively, due to function computation because the same remote source would be measured if another function would be computed in the same network, as in Fig. 2 below; see also [28] for motivations to consider privacy leakage with respect to a remote source. We remark that in (3), (5), and (6), we consider conditional mutual information terms to take into consideration the unavoidable secrecy or privacy leakage due to side information available at the fusion center or eavesdropper.

B. Lossy Single-Function Computation

Consider again the single-function computation model depicted in Fig. 1 and replace the reliability constraint in (2) with an expected distortion constraint to allow a distorted reconstruction of the function $f(\cdot,\cdot)$. This defines the lossy single-function computation model, for which the notion of achievability is as follows.

Definition 2. A *lossy* tuple $(R_\text{s}, R_\text{w}, R_{\ell,\text{Dec}}, R_{\ell,\text{Eve}}, D) \in \mathbb{R}^5_{\geq 0}$ is *achievable* if, for any $\delta > 0$, there exist $n \geq 1$, an encoder, and a decoder that satisfy (3)-(6) and

$$\mathbb{E}\Big[d(f^n(\widetilde{X}^n, Y^n), \widehat{f^n})\Big] \leq D + \delta \tag{7}$$

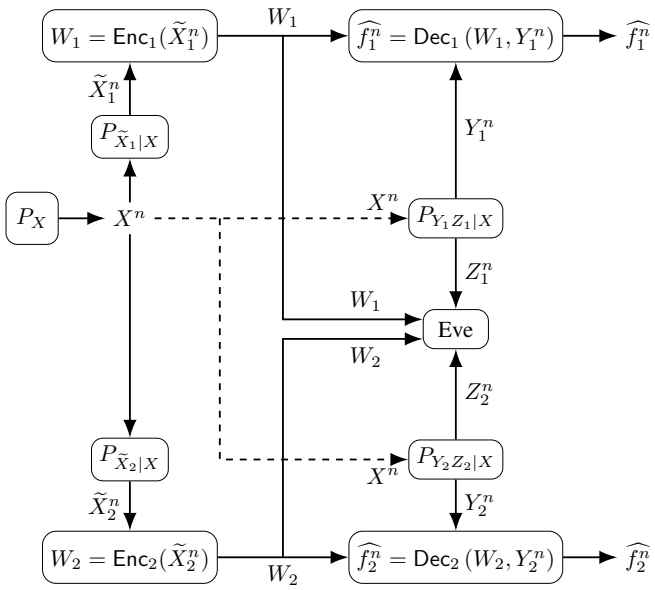

Fig. 2. Noisy measurements of the same remote source used to compute $J = 2$ functions (via $2J = 4$ parties) securely and privately with the help of public communication links.

where $d(f^n, \widehat{f^n}) = \frac{1}{n}\sum_{i=1}^n d(f_i, \widehat{f}_i)$ is a per-letter distortion metric. The *lossy* region $\mathcal{R}_{\text{D}}$ is the closure of the set of all achievable lossy tuples. $\Diamond$

C. Lossless Multi-Function Computation

We next extend the lossless single-function computation model by considering that the same remote source X^n is measured by multiple encoder and decoder pairs to compute different functions. Consider a finite number $J \geq 1$ of encoders $\mathsf{Enc}_j(\widetilde{X}_j) = W_j$, decoders $\mathsf{Dec}_j(W_j, Y_j^n) = \widehat{f_j^n}$, and functions $f_j^n(\widetilde{X}_j^n, Y_j^n) = \{f_j(\widetilde{X}_{i,j}, Y_{i,j})\}_{i=1}^n$ for $j \in [1:J]$, where $\widetilde{X}_j^n$ is measured through the channel $P_{\widetilde{X}_j|X}$ and (Y_j^n, Z_j^n) are measured through the BC $P_{Y_jZ_j|X}$. The eavesdropper observes $(Z_{[1:J]}^n, W_{[1:J]})$. This multi-function computation model is illustrated in Fig. 2 for $J = 2$.

Definition 3. A *multi-function* tuple $(R_{\text{s}}, R_{\text{w},[1:J]}, R_{\ell,\text{Dec},[1:J]}, R_{\ell,\text{Eve}}) \in \mathbb{R}_{\geq 0}^{2J+2}$ with j-th encoder measurements through $P_{\widetilde{X}_j|X}$ and j-th decoder measurements through $P_{Y_jZ_j|X}$ for all $j \in [1:J]$ is *achievable* if, for any $\delta > 0$, there exist $n \geq 1$, and J encoder and decoder pairs such that

$$\Pr\left[\bigcup_{j\in[1:J]} \left\{f_j^n(\widetilde{X}_j^n, Y_j^n) \neq \widehat{f_j^n}\right\}\right] \leq \delta \tag{8}$$

$$\frac{1}{n} I(\widetilde{X}_{[1:J]}^n, Y_{[1:J]}^n; W_{[1:J]}|Z_{[1:J]}^n) \leq R_{\text{s}} + \delta \tag{9}$$

$$\frac{1}{n}\log\left|\mathcal{W}_j\right| \leq R_{\text{w},j} + \delta, \qquad \forall j \in [1:J] \tag{10}$$

$$\frac{1}{n} I(X^n; W_j|Y_j^n) \leq R_{\ell,\text{Dec},j} + \delta, \qquad \forall j \in [1:J] \tag{11}$$

$$\frac{1}{n} I(X^n; W_{[1:J]}|Z_{[1:J]}^n) \leq R_{\ell,\text{Eve}} + \delta. \tag{12}$$

The *multi-function* region $\mathcal{R}_{\text{mf}}$ is the closure of the set of all achievable tuples. $\Diamond$

Remark 1. *The storage rate constraints in (10) and the corresponding privacy leakage rate constraints in (11) are J separate constraints. However, the reliability constraint in (8), the secrecy leakage constraint in (9), and the privacy leakage rate constraint in (12) are joint constraints that depend on the parameters of all J encoder-decoder pairs.*

D. Lossy Multi-Function Computation

Similar to Section II-B, we extend the model of Section II-C to allow distorted function computations for multiple functions $f_j^n(\widetilde{X}_j^n, Y_j^n) = \{f_j(\widetilde{X}_{i,j}, Y_{i,j})\}_{i=1}^n$ computed from different measurements $(\widetilde{X}_j^n, Y_j^n)$ of the same remote source X^n.

Definition 4. A *lossy multi-function* tuple $(R_{\text{s}}, R_{\text{w},[1:J]}, R_{\ell,\text{Dec},[1:J]}, R_{\ell,\text{Eve}}, D_{[1:J]}) \in \mathbb{R}_{\geq 0}^{3J+2}$ with j-th encoder measurements through $P_{\widetilde{X}_j|X}$ and j-th decoder measurements through $P_{Y_jZ_j|X}$ for all $j \in [1:J]$ is *achievable* if, for any $\delta > 0$, there exist $n \geq 1$, and J encoder and decoder pairs that satisfy (9)-(12) and

$$\mathbb{E}\Big[d(f_j^n(\widetilde{X}_j^n, Y_j^n), \widehat{f_j^n})\Big] \leq D_j + \delta, \qquad \forall j \in [1:J] \tag{13}$$

where $d(f^n, \widehat{f^n}) = \dfrac{1}{n}\displaystyle\sum_{i=1}^n d(f_i, \widehat{f}_i)$ is a per-letter distortion metric. The *lossy multi-function* region $\mathcal{R}_{\text{mf,D}}$ is the closure of the set of all achievable lossy multi-function tuples. $\Diamond$

III. Rate Regions

We first recall the notion of an *admissible random variable*, used in Theorems 1 and 3.

Definition 5 ([5])**.** A (vector) random variable U is admissible for a function $f(\widetilde{X}, Y)$ if $U - \widetilde{X} - Y$ form a Markov chain and $H(f(\widetilde{X}, Y)|U, Y) = 0$, i.e., (U, Y) determine $f(\widetilde{X}, Y)$.$\Diamond$

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

A. Lossless Single-Function Computation

We characterize the region $\mathcal{R}$ for the lossless single function computation problem in Theorem 1. The corresponding proof is detailed in Appendix A.

Theorem 1. *The region $\mathcal{R}$ is the set of all tuples $(R_s, R_w, R_{\ell,Dec}, R_{\ell,Eve})$ satisfying*

$$R_s \geq I(U; \widetilde{X}|Z) + \left[I(U; Z|V, Q) - I(U; Y|V, Q)\right]^- \tag{14}$$

$$R_w \geq I(U; \widetilde{X}|Y) \tag{15}$$

$$R_{\ell,Dec} \geq I(U; X|Y) \tag{16}$$

$$R_{\ell,Eve} \geq I(U; X|Z) + \left[I(U; Z|V, Q) - I(U; Y|V, Q)\right]^- \tag{17}$$

such that U is admissible for the function $f(\widetilde{X}, Y)$ and $(Q, V) - U - \widetilde{X} - X - (Y, Z)$ form a Markov chain. The region $\mathcal{R}$ is convexified by using the time-sharing random variable Q, which is required because of the $[\cdot]^-$ operation. One can limit the cardinalities of Q, V, and U to $|\mathcal{Q}| \leq 2$, $|\mathcal{V}| \leq |\widetilde{\mathcal{X}}| + 4$, and $|\mathcal{U}| \leq (|\widetilde{\mathcal{X}}| + 4)^2$.

In [26], some lower bounds on the rates in the rate regions include terms with the maximization operator $[\cdot]^+$. One can

show that the rate regions in [26] that include such lower bounds are not convex and can be enlarged by using a time-sharing random variable Q, as considered in this work.

B. Lossy Single-Function Computation

We next characterize the lossy region $\mathcal{R}_D$ for the lossy single function computation problem in Theorem 2.

Theorem 2. *The* lossy *region $\mathcal{R}_D$ is the set of all tuples $(R_s, R_w, R_{\ell,Dec}, R_{\ell,Eve}, D)$ satisfying*

$$R_s \geq I(U;\widetilde{X}|Z) + \big[I(U;Z|V,Q) - I(U;Y|V,Q)\big]^- \quad (18)$$

$$R_w \geq I(U;\widetilde{X}|Y) \quad (19)$$

$$R_{\ell,Dec} \geq I(U;X|Y) \quad (20)$$

$$R_{\ell,Eve} \geq I(U;X|Z) + \big[I(U;Z|V,Q) - I(U;Y|V,Q)\big]^- \quad (21)$$

$$D \geq \mathbb{E}[d(f(\widetilde{X},Y), g(U,Y))] \quad (22)$$

for some function $g(\cdot,\cdot)$ such that $(Q,V) - U - \widetilde{X} - X - (Y,Z)$ form a Markov chain. One can limit the cardinalities to $|\mathcal{Q}| \leq 2$, $|\mathcal{V}| \leq |\widetilde{X}| + 5$, and $|\mathcal{U}| \leq (|\widetilde{X}| + 5)^2$.

Proof Sketch. The achievability proof of Theorem 2 follows from the achievability proof of Theorem 1, except that U is not necessarily admissible, and with the addition that $P_{U|\widetilde{X}}$ and $P_{V|U}$ are chosen such that there exists a function $g(U,Y)$ that satisfies $g^n(U^n,Y^n) = \{g(U_i,Y_i)\}_{i=1}^n$ and $\mathbb{E}[d(f^n(\widetilde{X}^n,Y^n), g^n(U^n,Y^n))] \leq D + \epsilon_n$, where $\epsilon_n > 0$ such that $\epsilon_n \to 0$ when $n \to \infty$. Since all sequence tuples $(\widetilde{x}^n, y^n, u^n)$ are in the jointly typical set with high probability, by the typical average lemma [35, pp. 26], the distortion constraint (22) is satisfied. The converse proof follows from the converse proof of Theorem 1 by replacing the admissibility step in (82) with the steps

$$\begin{aligned}
D + \delta_n &\geq \mathbb{E}\Big[d\Big(f^n(\widetilde{X}^n,Y^n), \widehat{f^n}(W,Y^n)\Big)\Big] \\
&= \frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^n d\Big(f_i(\widetilde{X}_i,Y_i), \widehat{f}_i(W,Y^n)\Big)\Big] \\
&\overset{(a)}{\geq} \frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^n d\Big(f_i(\widetilde{X}_i,Y_i), g_i(W,Y^n,X^{i-1},Z^{i-1})\Big)\Big] \\
&\overset{(b)}{=} \frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^n d\Big(f_i(\widetilde{X}_i,Y_i), g_i(W,Y_i^n,X^{i-1},Z^{i-1})\Big)\Big] \\
&\overset{(c)}{=} \frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^n d\Big(f(\widetilde{X}_i,Y_i), g(U_i,Y_i)\Big)\Big] \quad (23)
\end{aligned}$$

where (a) follows since there exists a function $g_i(\cdot,\cdot)$ that results in a distortion smaller than or equal to the distortion obtained from $\widehat{f}_i(W,Y^n)$, where the distortion is measured with respect to $f_i(\widetilde{X}_i,Y_i)$ for all $i \in [1:n]$, because $g_i(\cdot,\cdot)$ has additional inputs, (b) follows from the Markov chain $Y^{i-1} - (X^{i-1},Z^{i-1},W,Y_i,Y_{i+1}^n) - f_i$ that can be proved by establishing conditional independence in the functional dependence graph of the problem by using the fundamental result in [36, Section 3.3], and (c) follows from the definition of $U_i \triangleq (W,X^{i-1},Y_{i+1}^n,Z^{i-1})$ given in Appendix A-B. The cardinality bounds follow by preserving the same probability and conditional entropy values as being preserved in Theorem 1 with the addition of preserving the value of $g(U,Y) = g(U,V,Y)$, following from the Markov chain $V - (U,Y) - g(U,Y)$. The region $\mathcal{R}_D$ is convexified by using a time-sharing random variable Q. □

All rate regions in [26, Section III] (and, naturally, all previous rate regions recovered by manipulating the regions in [26, Section III]) can be recovered from Theorems 1 and 2 by eliminating the remote source, i.e., assuming $\widetilde{X}^n = X^n$, and by rewriting the secrecy leakage constraint in (3) as an equivocation measure rather than a mutual information.

C. Lossless Multi-Function Computation

We provide inner and outer bounds for the multi-function region $\mathcal{R}_{\text{mf}}$ defined in Section II-C in Theorem 3. The corresponding proof is detailed in Appendix B.

Theorem 3. (Inner Bound): *An achievable multi-function region is the union over all P_Q, $P_{V_j|Q}$, $P_{U_j|V_j}$, and $P_{\widetilde{X}_j|U_j}$ of the rate tuples $(R_s, R_{w,[1:J]}, R_{\ell,Dec,[1:J]}, R_{\ell,Eve})$ such that U_j is admissible for the function $f_j(\widetilde{X}_j,Y_j)$ for all $j \in [1:J]$ and*

$$\begin{aligned}
R_s \geq \Big[&I(U_{[1:J]};Z_{[1:J]}|V_{[1:J]},Q) - I(U_{[1:J]};Y_{[1:J]}|V_{[1:J]},Q)\Big]^- \\
&+ I(U_{[1:J]};\widetilde{X}_{[1:J]}|Z_{[1:J]}) \quad (24)
\end{aligned}$$

$$R_{w,j} \geq I(U_j;\widetilde{X}_j|Y_j), \qquad \forall j \in [1:J] \quad (25)$$

$$\sum_{j=1}^J R_{w,j} \geq I(U_{[1:J]};\widetilde{X}_{[1:J]}|Y_{[1:J]}) \quad (26)$$

$$R_{\ell,Dec,j} \geq I(U_j;X|Y_j), \qquad \forall j \in [1:J] \quad (27)$$

$$\begin{aligned}
R_{\ell,Eve} \geq \Big[&I(U_{[1:J]};Z_{[1:J]}|V_{[1:J]},Q) - I(U_{[1:J]};Y_{[1:J]}|V_{[1:J]},Q)\Big]^- \\
&+ I(U_{[1:J]};X|Z_{[1:J]}) \quad (28)
\end{aligned}$$

where we have

$$\begin{aligned}
&P_{QV_{[1:J]}U_{[1:J]}\widetilde{X}_{[1:J]}XY_{[1:J]}Z_{[1:J]}} \\
&\quad = P_{Q|V_{[1:J]}}P_X\prod_{j=1}^J P_{V_j|U_j}P_{U_j|\widetilde{X}_j}P_{\widetilde{X}_j|X}P_{Y_jZ_j|X}. \quad (29)
\end{aligned}$$

(Outer Bound): *An outer bound for the multi-function region $\mathcal{R}_{mf}$ is the union of the rate tuples in (24)-(28) over all P_Q, $P_{V_j|Q}$, $P_{U_j|V_j}$, and $P_{\widetilde{X}_j|U_j}$ such that U_j is admissible for the function $f_j(\widetilde{X}_j,Y_j)$ and $(Q,V_j) - U_j - \widetilde{X}_j - X - (\widetilde{X}_{[1:J]\setminus j},Y_j,Z_j)$ form a Markov chain for all $j \in [1:J]$. One can limit the cardinalities to $|\mathcal{Q}| \leq 2$, $|\mathcal{V}_j| \leq |\widetilde{X}_j| + 5$, and $|\mathcal{U}_j| \leq (|\widetilde{X}_j| + 5)^2$ for all $j \in [1:J]$.*

Remark 2. *The inner and outer bounds differ because the outer bounds define rate regions for the Markov chains $(Q,V_j) - U_j - \widetilde{X}_j - X - (\widetilde{X}_{[1:J]\setminus j},Y_j,Z_j)$ for all $j \in [1:J]$, which are larger than the rate regions defined by the inner bounds that satisfy (29).*

D. Lossy Multi-Function Computation

We next give inner and outer bounds for the lossy multi-function region $\mathcal{R}_{\text{mf,D}}$, defined in Section II-D, in Theorem 4.

Theorem 4. (Inner Bound): *An achievable lossy multi-function region is the union over all P_Q, $P_{V_j|Q}$, $P_{U_j|V_j}$, and $P_{\widetilde{X}_j|U_j}$ for all $j \in [1:J]$ of the rate tuples $(R_s, R_{w,[1:J]}, R_{\ell,Dec,[1:J]}, R_{\ell,Eve}, D_{[1:J]})$ satisfying*

$$R_s \geq \Big[I(U_{[1:J]}; Z_{[1:J]}|V_{[1:J]}, Q) - I(U_{[1:J]}; Y_{[1:J]}|V_{[1:J]}, Q)\Big]^- + I(U_{[1:J]}; \widetilde{X}_{[1:J]}|Z_{[1:J]}) \quad (30)$$

$$R_{w,j} \geq I(U_j; \widetilde{X}_j|Y_j), \qquad \forall j \in [1:J] \quad (31)$$

$$\sum_{j=1}^{J} R_{w,j} \geq I(U_{[1:J]}; \widetilde{X}_{[1:J]}|Y_{[1:J]}) \quad (32)$$

$$R_{\ell,Dec,j} \geq I(U_j; X|Y_j), \qquad \forall j \in [1:J] \quad (33)$$

$$R_{\ell,Eve} \geq \Big[I(U_{[1:J]}; Z_{[1:J]}|V_{[1:J]}, Q) - I(U_{[1:J]}; Y_{[1:J]}|V_{[1:J]}, Q)\Big]^- + I(U_{[1:J]}; X|Z_{[1:J]}) \quad (34)$$

$$D_j \geq \mathbb{E}[d(f_j(\widetilde{X}_j, Y_j), g_j(U_j, Y_j))] \qquad \forall j \in [1:J] \quad (35)$$

for a set of functions $\{g_j(\cdot,\cdot)\}_{j=1}^J$ and where (29) is satisfied.

(Outer Bound): *An outer bound for the lossy multi-function region $\mathcal{R}_{mf,D}$ is the union of the rate tuples in (30) - (35) over all P_Q, $P_{V_j|Q}$, $P_{U_j|V_j}$, and $P_{\widetilde{X}_j|U_j}$ such that $(Q, V_j) - U_j - \widetilde{X}_j - X - (\widetilde{X}_{[1:J]\setminus j}, Y_j, Z_j)$ form a Markov chain for all $j \in [1:J]$. One can limit the cardinalities to $|\mathcal{Q}| \leq 2$, $|\mathcal{V}_j| \leq |\widetilde{X}_j| + 6$, and $|\mathcal{U}_j| \leq (|\widetilde{X}_j| + 6)^2$ for all $j \in [1:J]$.*

Proof Sketch. The inner bound proof of Theorem 4 follows from the achievability proof of Theorem 3, except that U_j's are not necessarily admissible, and with the addition that $P_{U_j|\widetilde{X}_j}$ and $P_{V_j|U_j}$ are chosen such that there exists a set of functions $\{g_j(U_j, Y_j)\}_{j=1}^J$ that satisfy $g_j^n(U_j^n, Y_j^n) = \{g_j(U_{i,j}, Y_{i,j})\}_{i=1}^n$ and $\mathbb{E}[d(f_j^n(\widetilde{X}_j^n, Y_j^n), g_j^n(U_j^n, Y_j^n))] \leq D_j + \epsilon_n$ for all $j \in [1:J]$, where $\epsilon_n > 0$ such that $\epsilon_n \to 0$ when $n \to \infty$. Since all sequence tuples $(\widetilde{x}_j^n, y_j^n, u_j^n)$ are in the jointly typical set with high probability for all $j \in [1:J]$, by the typical average lemma, the distortion constraints in (35) are satisfied. The outer bound proof of Theorem 4 follows from the converse proof of Theorem 3 with the replacement of the admissibility step in (110) with the steps given in (23) for random variables and functions with the indices $j = 1, 2, \ldots, J$. □

IV. Information Bottleneck Example

Consider the lossy single-function computation problem and suppose $X - Y - Z$ form a Markov chain. The characterization of the corresponding rate region requires one to maximize a mutual information term upper bounded by another mutual information term that should be minimized simultaneously, i.e., an information bottleneck.

Corollary 1. *The lossy region of Theorem 2 when $X - Y - Z$ form a Markov chain is the set of all tuples $(R_s, R_w, R_{\ell,Dec}, R_{\ell,Eve}, D)$ satisfying*

$$R_s \geq I(U; \widetilde{X}|Y) = I(U; \widetilde{X}) - I(U; Y) \quad (36)$$

$$R_w \geq I(U; \widetilde{X}|Y) = I(U; \widetilde{X}) - I(U; Y) \quad (37)$$

$$R_{\ell,Dec} \geq I(U; X|Y) = I(U; X) - I(U; Y) \quad (38)$$

$$R_{\ell,Eve} \geq I(U; X|Y) = I(U; X) - I(U; Y) \quad (39)$$

$$D \geq \mathbb{E}[d(f(\widetilde{X}, Y), g(U, Y))] \quad (40)$$

for some function $g(\cdot,\cdot)$ such that $U - \widetilde{X} - X - Y - Z$ form a Markov chain. One can limit the cardinality to $|\mathcal{U}| \leq |\widetilde{X}| + 2$.

The proof of Corollary 1 follows by applying steps identical to the proof of [26, Corollary 3] to Theorem 2, we thus omit it. The boundary points of the rate region defined in Corollary 1 can be obtained by maximizing $I(U;Y)$ and minimizing $I(U;\widetilde{X})$ simultaneously for a fixed $I(U;X)$ for all $P_{U|\widetilde{X}}$ such that $U - \widetilde{X} - X - Y - Z$ form a Markov chain. This problem is an information bottleneck problem [37], [38]. If the distortion metric $d(\cdot,\cdot)$ is chosen to be the Hamming distance, we then obtain the optimal function $g^*(u,y)$ for all $(u,y) \in \mathcal{U} \times \mathcal{Y}$ as [26, Eq. (26)]

$$g^*(u,y) = \arg\max_f P_{F|UY}(f|u,y) \quad (41)$$

where $f = f(\widetilde{x}, y)$ is a realization of the random function output F for any $(\widetilde{x}, y) \in \widetilde{\mathcal{X}} \times \mathcal{Y}$.

Consider a measurement channel $P_{\widetilde{X}|X}$ and source P_X for the encoder $\mathsf{Enc}(\cdot)$ such that the inverse channel $P_{X|\widetilde{X}}$ is a BSC(p) for any $0 \leq p \leq 0.5$. Furthermore, suppose the measurement channel $P_{Y|X}$ for the decoder $\mathsf{Dec}(\cdot)$ is a binary input symmetric output channel [39, p. 21], which can be decomposed into a mixture of binary subchannels as defined in [40, Section III-B] [41]. We remark that the rate region defined in Corollary 1 by (36)-(40) does not depend on the random variable Z. Therefore, the measurement channel for the eavesdropper does not affect the rate region as long as the measurement channel for the eavesdropper is physically-degraded as compared to the channel for the decoder $\mathsf{Dec}(\cdot)$, i.e., $P_{YZ|X} = P_{Z|Y} P_{Y|X}$. Since $P_{\widetilde{X}XYZ}$ is fixed, the optimal auxiliary random variable U is such that $P_{\widetilde{X}|U}$ is a BSC with crossover probability

$$\frac{H_b^{-1}(H(X|U)) - p}{1 - 2p} \quad (42)$$

which follows from [28, Theorem 3].

Suppose $P_X \sim \text{Bern}(0.5)$, $P_{\widetilde{X}|X} \sim \text{BSC}(p = 0.06)$, and assume that the measurement channel $P_{Y|X}$ consists of $M > 1$ independent BSCs each with crossover probability 0.15, which satisfies the assumptions listed above. Using auxiliary random variables satisfying (42), we depict the projections of $(R_{\text{s}}, R_{\text{w}}, R_{\ell,\text{Dec}}, R_{\ell,\text{Eve}}, D)$ boundary tuples onto the $(R_{\text{s}}, R_{\ell,\text{Eve}})$ plane in Fig. 3 for $M = 1, 2, 3$ independent BSC measurements by the decoder $\mathsf{Dec}(\cdot)$.

Fig. 3 suggests that given a boundary point achieved by a crossover probability calculated as in (42), any larger secrecy-leakage rate and any larger privacyEve-leakage rate are also achievable. Conversely, given such an achievable boundary point, no smaller secrecy-leakage rate and no smaller privacyEve-leakage rate is achievable. Furthermore, increasing the number M of measurements at the decoder significantly decreases the corresponding boundary point such that, e.g., when $M = 3$ measurements are used as compared to $M = 1$, the maximum secrecy-leakage rate decreases by approximately 31.45% and simultaneously the maximum privacy-leakage

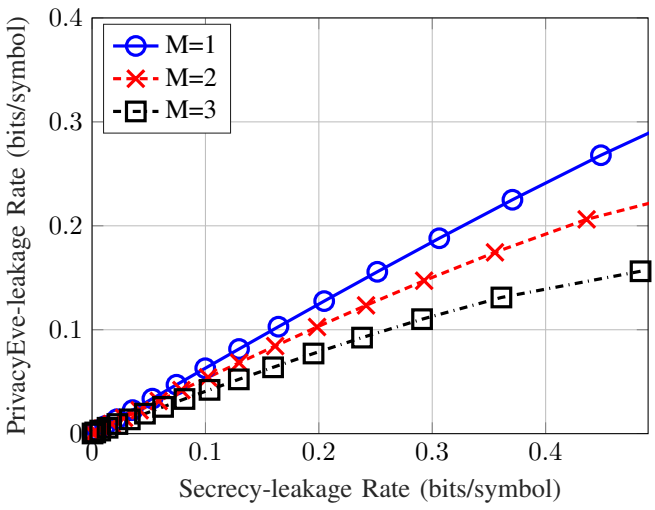

Fig. 3. Secrecy-leakage rate vs. privacyEve-leakage rate projection of the boundary tuples $(R_{\mathrm{s}}, R_{\mathrm{w}}, R_{\ell,\mathrm{Dec}}, R_{\ell,\mathrm{Eve}}, D)$ for $p = 0.06$ and for the number of independent BSC measurements at the decoder $M = 1, 2, 3$.

rate to the eavesdropper decreases by approximately 58.68%. These gains can be seen as multiplexing gains, in analogy to multiple antenna systems for wireless communications.

V. Conclusion

We derived the secrecy-storage-privacyDec-privacyEve(-distortion) regions for lossless and lossy single-function computations with a remote source. The remote source model allows to model multiple sequences observed by a single terminal as multiple noisy measurements of a hidden source, which allows to measure the diversity gains. The equivocation measure common in the literature was replaced with a mutual information metric, which resulted in simpler notation and easier interpretations. A new privacy metric was considered to bound the information leakage to a fusion center about the remote source sequence. Bounds for the storage and privacy leakage to the eavesdropper rates were shown to be different, unlike in the previous models. Inner and outer bounds for multiple asynchronous function computations within the same network were given to illustrate the effects of joint constraints for all terminals involved in any function computation. These bounds differ only in the Markov chain conditions imposed. We evaluated the rate region for a single-function computation problem by solving an information bottleneck problem for binary input symmetric output channels. In future work, we will consider multi-function computations with multiple transmitting terminals for each function computation and derive the rate regions for two-function computations with two transmitting terminals if a set of symmetry conditions are satisfied.

References

[1] O. Günlü, M. Bloch, and R. F. Schaefer, "Secure multi-function computation with private remote sources," in *IEEE Int. Symp. Inf. Theory*, Melbourne, Victoria, Australia, July 2021, pp. 1403–1408.

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

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

[4] R. Mijumbi, J. Serrat, J.-L. Gorricho, N. Bouten, F. De Turck, and R. Boutaba, "Network function virtualization: State-of-the-art and research challenges," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 236–262, Firstquarter 2016.

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

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

[7] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[8] I. Csiszár and P. Narayan, "Common randomness and secret key generation with a helper," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 344–366, Mar. 2000.

[9] O. Günlü, O. İşcan, V. Sidorenko, and G. Kramer, "Code constructions for physical unclonable functions and biometric secrecy systems," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 11, pp. 2848–2858, Nov. 2019.

[10] J. Ren, B. D. Boyle, G. Ku, S. Weber, and J. M. Walsh, "Overhead performance tradeoffs - A resource allocation perspective," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3243–3269, June 2016.

[11] O. Günlü and R. F. Schaefer, "An optimality summary: Secret key agreement with physical unclonable functions," *Entropy*, vol. 23, no. 1, Jan. 2021.

[12] M. Bloch, O. Günlü, A. Yener, F. Oggier, H. V. Poor, L. Sankar, and R. F. Schaefer, "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.

[13] N. Ma and P. Ishwar, "Some results on distributed source coding for interactive function computation," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6180–6195, Aug. 2011.

[14] M. Sefidgaran and A. Tchamkerten, "Distributed function computation over a rooted directed tree," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7135–7152, Dec. 2016.

[15] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498–3516, Sep. 2007.

[16] H. Kowshik and P. R. Kumar, "Optimal function computation in directed and undirected graphs," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3407–3418, Feb. 2012.

[17] S. Kannan and P. Viswanath, "Multi-session function computation and multicasting in undirected graphs," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 4, pp. 702–713, Mar. 2013.

[18] H. Tyagi, P. Narayan, and P. Gupta, "When is a function securely computable?" *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6337–6350, Oct. 2011.

[19] H. Tyagi and S. Watanabe, "A bound for multiparty secret key agreement and implications for a problem of secure computing," in *Int. Conf. Theory Appl. Crypt. Techn.*, Copenhagen, Denmark, May 2014, pp. 369–386.

[20] ——, "Converses for secret key agreement and secure computing," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4809–4827, July 2015.

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

[22] M. Goldenbaum, H. Boche, and H. V. Poor, "On secure computation over the binary modulo-2 adder multiple-access wiretap channel," in *IEEE Inf. Theory Workshop*, Cambridge, U.K., Sep. 2016, pp. 21–25.

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

[24] G. R. Kurri and V. M. Prabhakaran, "Secure computation to hide functions of inputs," in *IEEE Int. Symp. Inf. Theory*, Los Angeles, CA, June 2020, pp. 972–977.

[25] H. Tyagi, "Distributed function computation with confidentiality," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 4, pp. 691–701, Apr. 2013.

[26] 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.

[27] 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.

[28] 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.

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

[30] O. Günlü, R. F. Schaefer, and G. Kramer, "Private authentication with physical identifiers through broadcast channel measurements," in *IEEE Inf. Theory Workshop*, Visby, Sweden, Aug. 2019, pp. 1–5.
[31] 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.
[32] Y. Wang, S. Rane, S. C. Draper, and P. Ishwar, "A theoretical analysis of authentication, privacy, and reusability across secure biometric systems," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1825–1840, July 2012.
[33] O. Günlü, K. Kittichokechai, R. F. Schaefer, and G. Caire, "Controllable identifier measurements for private authentication with secret keys," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 8, pp. 1945–1959, Aug. 2018.
[34] O. Günlü, "Multi-entity and multi-enrollment key agreement with correlated noise," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1190–1202, 2021.
[35] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge University Press, 2011.
[36] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.
[37] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," in *Allerton Conf. Comm., Control, Comp.*, Monticello, IL, Sep. 1999, pp. 368–377.
[38] H. Witsenhausen and A. Wyner, "A conditional entropy bound for a pair of discrete random variables," *IEEE Trans. Inf. Theory*, vol. 21, no. 5, pp. 493–501, Sep. 1975.
[39] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA: M.I.T. Press, 1963.
[40] O. Günlü, G. Kramer, and M. Skórski, "Privacy and secrecy with multiple measurements of physical and biometric identifiers," in *IEEE Int. Conf. Commun. Netw. Security*, Florence, Italy, Sep. 2015, pp. 89–94.
[41] N. Chayat and S. Shamai, "Extension of an entropy property for binary input memoryless symmetric channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 5, pp. 1077–1079, Sep. 1989.
[42] 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.
[43] 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.
[44] M. Bloch, *Lecture Notes in Information-Theoretic Security*. Atlanta, GA: Georgia Inst. Technol., July 2018.
[45] T. Holenstein and R. Renner, "On the randomness of independent experiments," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1865–1871, Apr. 2011.
[46] M. Bloch and J. Barros, *Physical-layer Security*. Cambridge, U.K.: Cambridge University Press, 2011.
[47] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge, U.K.: Cambridge University Press, 2011.

Onur Günlü (S'10–M'18) received the B.Sc. degree (Highest Distinction) in Electrical and Electronics Engineering from Bilkent University, Turkey in 2011; M.Sc. (Highest Distinction) and Dr.-Ing. (Ph.D. equivalent) degrees in Communications Engineering both from the Technical University of Munich (TUM), Germany in October 2013 and November 2018, respectively. He was a Working Student in the Communication Systems division of Intel Mobile Communications (IMC), now Apple Inc., in Munich, Germany between November 2012 - March 2013. He later worked as a Research and Teaching Assistant at TUM between February 2014 - May 2019. Onur was a Visiting Researcher, among more than ten Research Stays at Top Universities and Companies, at TU Eindhoven, Netherlands during February 2018 - March 2018 to visit Prof. Frans M. J. Willems, and later at Georgia Institute of Technology, Atlanta, USA during February 2022 - March 2022 to visit Prof. Matthieu Bloch. He was a Research Associate & Dozent between June 2019 - September 2020 and a Research Group Leader & Project Leader & Dozent between October 2020 - March 2021 both at TU Berlin, Germany, and he has held the same academic titles at the Chair of Communications Engineering and Security at the University of Siegen, Germany since April 2021. Onur has been a BrainCityBerlin Ambassador since June 2020 and has recently been awarded the prestigious 2021 Johann-Philipp-Reis Award by the VDE Information Technology Society (ITG). He has been selected by the IEEE Communications Society as 2021 Exemplary Reviewer of the IEEE Transactions on Communications (TCOM). His research interests include information-theoretic privacy and security, coding theory, statistical signal processing for biometrics and physical unclonable functions (PUFs), private (federated) learning and function computation, secure terahertz (THz) communications, and doubly-exponential (secure) identification via channels. Among his publications is the recent book *Key Agreement with Physical Unclonable Functions and Biometric Identifiers* (Dr. Hut Verlag, 2019). He is currently an Associate Editor for the EURASIP JOURNAL ON WIRELESS COMMUNICATIONS AND NETWORKING and ENTROPY JOURNAL, a Reviewer Board Member for the COMPUTERS and INFORMATION JOURNALS, and was recently a Guest Editor for the IEEE JOURNAL ON SELECTED AREAS IN INFORMATION THEORY.

Matthieu Bloch is a Professor in the School of Electrical and Computer Engineering. He received the Engineering degree from Supélec, Gif-sur-Yvette, France, the M.S. degree in Electrical Engineering from the Georgia Institute of Technology, Atlanta, in 2003, the Ph.D. degree in Engineering Science from the Université de Franche-Comté, Besançon, France, in 2006, and the Ph.D. degree in Electrical Engineering from the Georgia Institute of Technology in 2008. In 2008-2009, he was a postdoctoral research associate at the University of Notre Dame, South Bend, IN. Since July 2009, Dr. Bloch has been on the faculty of the School of Electrical and Computer Engineering, and from 2009 to 2013 Dr. Bloch was based at Georgia Tech Lorraine. His research interests are in the areas of information theory, error-control coding, wireless communications, and cryptography. Dr. Bloch has served on the organizing committee of several international conferences; he was the chair of the Online Committee of the IEEE Information Theory Society from 2011 to 2014, an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2016 to 2019, and he has been on the Board of Governors of the IEEE Information Theory Society since 2016 and currently serves as the 1st Vice-President. He has been an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY since 2019. He is the co-recipient of the IEEE Communications Society and IEEE Information Theory Society 2011 Joint Paper Award and the co-author of the textbook *Physical-Layer Security: From Information Theory to Security Engineering* published by Cambridge University Press.

Rafael F. Schaefer (S'08–M'12–SM'17) received the Dipl.-Ing. degree in electrical engineering and computer science from the Technische Universität Berlin, Germany, in 2007, and the Dr.-Ing. degree in electrical engineering from the Technische Universität München, Germany, in 2012. From 2013 to 2015, he was a Post-Doctoral Research Fellow with Princeton University. From 2015 to 2020, he was an Assistant Professor with the Technische Universität Berlin. Since 2021, he has been a Full Professor with Universität Siegen. Among his publications is the recent book *Information Theoretic Security and Privacy of Information Systems* (Cambridge University Press, 2017). He was a recipient of the VDE Johann-Philipp-Reis Award in 2013. He received the best paper award of the German Information Technology Society (ITG-Preis) in 2016. He was one of the exemplary reviewers of the IEEE COMMUNICATION LETTERS in 2013. He is currently an Associate Editor of the IEEE TRANSACTIONS OF INFORMATION FORENSICS AND SECURITY and of the IEEE TRANSACTIONS OF COMMUNICATIONS. He is a Member of the IEEE Information Forensics and Security Technical Committee.

APPENDIX A
PROOF OF THEOREM 1

A. Achievability Proof of Theorem 1

Proof Sketch. We use the output statistics of random binning (OSRB) method, proposed in [42] (see also [43]) for strong secrecy by following steps in [44, Section 1.6]. This approach simplifies the analysis compared to previous proofs in the literature.

Fix $P_{U|\widetilde{X}}$ and $P_{V|U}$ such that U is admissible and let $(V^n, U^n, \widetilde{X}^n, X^n, Y^n, Z^n)$ be i.i.d. according to $P_{VU\widetilde{X}XYZ} = P_{V|U}P_{U|\widetilde{X}}P_{\widetilde{X}|X}P_XP_{YZ|X}$. We remark that since all n-letter random variables are i.i.d., U^n is also admissible.

Assign two random bin indices $(F_{\mathrm{v}}, W_{\mathrm{v}})$ to each v^n. Assume $F_{\mathrm{v}} \in [1 : 2^{n\widetilde{R}_{\mathrm{v}}}]$ and $W_{\mathrm{v}} \in [1 : 2^{nR_{\mathrm{v}}}]$. Similarly, assign two indices $(F_{\mathrm{u}}, W_{\mathrm{u}})$ to each u^n, where $F_{\mathrm{u}} \in [1 : 2^{n\widetilde{R}_{\mathrm{u}}}]$ and $W_{\mathrm{u}} \in [1 : 2^{nR_{\mathrm{u}}}]$. The public message is $W = (W_{\mathrm{v}}, W_{\mathrm{u}})$ and the indices $F = (F_{\mathrm{v}}, F_{\mathrm{u}})$ represent the public choice of encoder-decoder pairs.

Using a Slepian-Wolf (SW) [6] decoder, one can reliably estimate V^n from $(F_{\mathrm{v}}, W_{\mathrm{v}}, Y^n)$, such that the expected value of the error probability taken over the random bin assignments vanishes when $n \to \infty$, if we have [42, Lemma 1]

$$\widetilde{R}_{\mathrm{v}} + R_{\mathrm{v}} > H(V|Y). \tag{43}$$

Similarly, one can reliably estimate U^n from $(F_{\mathrm{u}}, W_{\mathrm{u}}, Y^n, V^n)$ by using a SW decoder if we have

$$\widetilde{R}_{\mathrm{u}} + R_{\mathrm{u}} > H(U|V, Y). \tag{44}$$

Thus, the reliability constraint in (2) is satisfied if (43) and (44) are satisfied.

The public index F_{v} is almost independent of $\widetilde{X}^n$, so it is almost independent of $(\widetilde{X}^n, X^n, Y^n, Z^n)$, if we have [42, Theorem 1]

$$\widetilde{R}_{\mathrm{v}} < H(V|\widetilde{X}) \tag{45}$$

since it results in the expected value, which is 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 $P_{F_{\mathrm{v}}\widetilde{X}^n}$ to vanish when $n \to \infty$. Furthermore, the public index F_{u} is almost independent of $(V^n, \widetilde{X}^n)$, so it is almost independent of $(V^n, \widetilde{X}^n, X^n, Y^n, Z^n)$, if we have

$$\widetilde{R}_{\mathrm{u}} < H(U|V, \widetilde{X}). \tag{46}$$

To satisfy the constraints (43)-(46), we fix the rates to

$$\widetilde{R}_{\mathrm{v}} = H(V|\widetilde{X}) - \epsilon \tag{47}$$

$$R_{\mathrm{v}} = I(V; \widetilde{X}) - I(V; Y) + 2\epsilon \tag{48}$$

$$\widetilde{R}_{\mathrm{u}} = H(U|V, \widetilde{X}) - \epsilon \tag{49}$$

$$R_{\mathrm{u}} = I(U; \widetilde{X}|V) - I(U; Y|V) + 2\epsilon \tag{50}$$

for any $\epsilon > 0$.

Storage (Public Message) Rate: (47)-(50) result in a storage (public message) rate R_{w} of

$$\begin{aligned} R_{\mathrm{w}} &= R_{\mathrm{v}} + R_{\mathrm{u}} = I(V, U; \widetilde{X}) - I(V, U; Y) + 4\epsilon \\ &\overset{(a)}{=} I(U; \widetilde{X}|Y) + 4\epsilon \end{aligned} \tag{51}$$

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

Privacy Leakage to the Decoder: We have

$$\begin{aligned} I(X^n; W, F|Y^n) &= I(X^n; W|F, Y^n) + I(X^n; F|Y^n) \\ &\overset{(a)}{\leq} H(X^n|Y^n) - H(X^n|W, F, V^n, U^n, Y^n) + 2\epsilon_n \\ &\overset{(b)}{=} H(X^n|Y^n) - H(X^n|U^n, Y^n) + 2\epsilon_n \\ &\overset{(c)}{=} nI(U; X|Y) + 2\epsilon_n \end{aligned} \tag{52}$$

where (a) follows for some $\epsilon_n > 0$ with $\epsilon_n \to 0$ when $n \to \infty$ because

$$\begin{aligned} I(X^n; F|Y^n) &= I(X^n; F_{\mathrm{v}}|Y^n) + I(X^n; F_{\mathrm{u}}|F_{\mathrm{v}}, Y^n) \\ &\leq 2\epsilon_n \end{aligned} \tag{53}$$

since by (45) F_v is almost independent of $(\widetilde{X}^n, X^n, Y^n, Z^n)$ and by (46) F_u is almost independent of $(V^n, \widetilde{X}^n, X^n, Y^n, Z^n)$ and because V^n determines F_v; see also [45, Theorem 1], (b) follows because V^n determines (F_v, W_v), U^n determines (F_u, W_u), and $V^n - U^n - (X^n, Y^n)$ form a Markov chain, and (c) follows because (X^n, U^n, Y^n) are i.i.d.

Privacy Leakage to the Eavesdropper: We have

$$\begin{aligned} I(X^n; W, F|Z^n) &\overset{(a)}{=} H(W, F|Z^n) - H(W, F|X^n) \\ &\overset{(b)}{=} H(W, F|Z^n) - H(W_u, F_u, V^n|X^n) \\ &\quad + H(V^n|W_v, F_v, W_u, F_u, X^n) \\ &\overset{(c)}{\leq} H(W, F|Z^n) - H(W_u, F_u, V^n|X^n) + n\epsilon'_n \\ &\overset{(d)}{=} H(W, F|Z^n) - H(U^n, V^n|X^n) \\ &\quad + H(U^n|W_u, F_u, V^n, X^n) + n\epsilon'_n \\ &\overset{(e)}{\leq} H(W, F|Z^n) - H(U^n, V^n|X^n) + 2n\epsilon'_n \\ &\overset{(f)}{=} H(W, F|Z^n) - nH(U, V|X) + 2n\epsilon'_n \end{aligned} \tag{54}$$

where (a) follows because $(W, F) - X^n - Z^n$ form a Markov chain, (b) follows since V^n determines (F_v, W_v), (c) follows for some $\epsilon'_n > 0$ such that $\epsilon'_n \to 0$ when $n \to \infty$ because (F_v, W_v, X^n) can reliably recover V^n due to the Markov chain $V^n - X^n - Y^n$ and by (43), (d) follows because U^n determines (F_u, W_u), (e) follows by (44) because (W_u, F_u, V^n, X^n) can reliably recover U^n due to the inequality $H(U|V, Y) \geq H(U|V, X)$ that follows from

$$\begin{aligned} H(U|V, Y) - H(U|V, X) &= I(U; V, X) - I(U; V, Y) \\ &\geq I(U; V, X) - I(U; V, Y, X) = 0 \end{aligned} \tag{55}$$

since $U - (V, X) - Y$ form a Markov chain, and (f) follows because (U^n, V^n, X^n) are i.i.d.

We need to analyze six different decodability cases to consider whether (F_v, W_v, Z^n) can recover V^n and whether (F_u, W_u, V^n, Z^n) or (F_u, W_u, Z^n) can recover U^n.

Case 1: Assume

$$0 \leq R_{\mathrm{v}} + \widetilde{R}_{\mathrm{v}} < H(V|Z), \tag{56}$$

$$0 \leq R_{\mathrm{u}} + \widetilde{R}_{\mathrm{u}} < H(U|V, Z) \tag{57}$$

so that (F_v, W_v) are almost independent of Z^n and are also almost mutually independent, and (F_u, W_u) are almost independent of (V^n, Z^n) and are also almost mutually independent. Using (54), we obtain

$$\begin{aligned}
&I(X^n; W, F|Z^n)\\
&\le H(W_v)+H(F_v)+H(W_u)+H(F_u)\\
&\quad -nH(U,V|X)+2n\epsilon'_n\\
&\le n(R_{\mathrm{v}}+\widetilde{R}_{\mathrm{v}}+R_{\mathrm{u}}+\widetilde{R}_{\mathrm{u}})-nH(U,V|X)+2n\epsilon'_n\\
&\overset{(a)}{=} n(I(U,V;X)-I(U,V;Y)+2\epsilon+2\epsilon'_n)\\
&\overset{(b)}{=} n(I(U;X)-I(U;Y|V)-I(V;Y)+2\epsilon+2\epsilon'_n)\\
&\overset{(c)}{\le} n(I(U;X)-I(U;Y|V)-I(V;Z)+\epsilon+2\epsilon'_n)\\
&\overset{(d)}{=} n(I(U;X)-[I(U;Y|V)-I(U;Z|V)]-I(U;Z)+\epsilon+2\epsilon'_n)\\
&\overset{(e)}{=} n(I(U;X|Z)+[I(U;Z|V)-I(U;Y|V)+\epsilon]^-+2\epsilon'_n) \quad (58)
\end{aligned}$$

where (a) follows by (47)-(50) and (b) follows from the Markov chain $V-U-X$, (c) follows by (47), (48), and (56) such that equality is achieved when $n\to\infty$, (d) follows from the Markov chain $V-U-Z$, and (e) follows from the Markov chain $U-X-Z$.

Case 2: Assume

$$0 \le R_{\mathrm{v}}+\widetilde{R}_{\mathrm{v}} < H(V|Z), \quad (59)$$

$$H(U|V,Z) < R_{\mathrm{u}}+\widetilde{R}_{\mathrm{u}} < H(U|Z) \quad (60)$$

so that (F_v, W_v) are almost independent of Z^n and are also almost mutually independent, and (F_u, W_u) are almost independent of Z^n and are also almost mutually independent; however, (F_u, W_u, V^n, Z^n) can reliably recover U^n. Using (54), we have

$$\begin{aligned}
&I(X^n; W, F|Z^n)\\
&\overset{(a)}{\le} H(U^n, V^n|Z^n)-nH(U,V|X)+2n\epsilon'_n\\
&\overset{(b)}{=} n(I(U;X|Z)+[I(U;Z|V)-I(U;Y|V)+\epsilon]^-+2\epsilon'_n) \quad (61)
\end{aligned}$$

where (a) follows because V^n determines (F_v, W_v) and U^n determines (F_u, W_u), and (b) follows from the Markov chain $V-U-X-Z$ and by (49), (50), and (60).

Case 3: Assume

$$0 \le R_{\mathrm{v}}+\widetilde{R}_{\mathrm{v}} < H(V|Z), \quad (62)$$

$$H(U|Z) < R_{\mathrm{u}}+\widetilde{R}_{\mathrm{u}} \quad (63)$$

so that (F_v, W_v) are almost independent of Z^n and are also almost mutually independent, and (F_u, W_u, Z^n) can reliably recover U^n. Using (54), we obtain

$$\begin{aligned}
&I(X^n; W, F|Z^n)\\
&\overset{(a)}{\le} H(U^n|Z^n)+H(W_v, F_v|U^n, Z^n)-nH(U,V|X)+2n\epsilon'_n\\
&\overset{(b)}{\le} H(U^n|Z^n)+H(V^n|U^n, Z^n)-nH(U,V|X)+2n\epsilon'_n\\
&\overset{(c)}{=} n(I(U;X|Z)+2\epsilon'_n)\\
&\overset{(d)}{=} n(I(U;X|Z)+[I(U;Z|V)-I(U;Y|V)+\epsilon]^-+2\epsilon'_n) \quad (64)
\end{aligned}$$

where (a) follows because U^n determines (F_u, W_u), (b) follows since V^n determines (F_v, W_v), (c) follows from the Markov chain $V-U-X-Z$ and because (V^n, U^n, X^n, Z^n) are i.i.d., and (d) follows by (49), (50), and (63).

Case 4: Assume

$$H(V|Z) < R_{\mathrm{v}}+\widetilde{R}_{\mathrm{v}}, \quad (65)$$

$$0 \le R_{\mathrm{u}}+\widetilde{R}_{\mathrm{u}} < H(U|V,Z) \quad (66)$$

so that (F_v, W_v, Z^n) can reliably recover V^n, and (F_u, W_u) are almost independent of (V^n, Z^n) and are also almost mutually independent. Using (54), we have

$$\begin{aligned}
&I(X^n; W, F|Z^n)\\
&\overset{(a)}{\le} H(V^n|Z^n)+H(W_u, F_u|W_v, F_v, Z^n)\\
&\quad -nH(U,V|X)+2n\epsilon'_n\\
&\le H(V^n|Z^n)+H(W_u)+H(F_u)-nH(U,V|X)+2n\epsilon'_n\\
&\le n(H(V|Z)+R_{\mathrm{u}}+\widetilde{R}_{\mathrm{u}}-H(U,V|X)+2\epsilon'_n)\\
&\overset{(b)}{=} n(H(V|Z)+H(U|V,Y)+\epsilon-H(U,V|X)+2\epsilon'_n)\\
&= n(I(U;X|V)-I(U;Y|V)+I(V;X)-I(V;Z)+2\epsilon'_n+\epsilon)\\
&\overset{(c)}{=} n(I(U;X)-I(U;Y|V)-I(V;Z)+2\epsilon'_n+\epsilon\\
&\overset{(d)}{=} n(I(U;X)-[I(U;Y|V)-I(U;Z|V)]-I(U;Z)+\epsilon+2\epsilon'_n)\\
&\overset{(e)}{=} n(I(U;X|Z)+[I(U;Z|V)-I(U;Y|V)+\epsilon]^-+2\epsilon'_n) \quad (67)
\end{aligned}$$

where (a) follows because V^n determines (F_v, W_v), (b) follows because (V^n, Z^n) are i.i.d. and by (49) and (50), (c) follows from the Markov chain $V-U-X$, (d) follows from the Markov chain $V-U-Z$, and (e) follows from the Markov chain $U-X-Z$.

Case 5: Assume

$$H(V|Z) < R_{\mathrm{v}}+\widetilde{R}_{\mathrm{v}}, \quad (68)$$

$$H(U|V,Z) < R_{\mathrm{u}}+\widetilde{R}_{\mathrm{u}} < H(U|Z) \quad (69)$$

so that (F_v, W_v, Z^n) can reliably recover V^n, and (F_u, W_u) are almost independent of Z^n and are also almost mutually independent; however, (F_u, W_u, V^n, Z^n) can reliably recover U^n. Using (54), we have

$$\begin{aligned}
&I(X^n; W, F|Z^n)\\
&\overset{(a)}{\le} H(V^n|Z^n)+H(W_u, F_u|W_v, F_v, Z^n)\\
&\quad -nH(U,V|X)+2n\epsilon'_n\\
&\overset{(b)}{\le} H(V^n|Z^n)+H(W_u, F_u|V^n, Z^n)\\
&\quad +H(V^n|W_v, F_v, Z^n)-nH(U,V|X)+2n\epsilon'_n\\
&\overset{(c)}{\le} H(V^n|Z^n)+H(U^n|V^n, Z^n)-nH(U,V|X)+3n\epsilon'_n\\
&\overset{(d)}{=} n(I(U;X|Z)+3\epsilon'_n)\\
&\overset{(e)}{=} n(I(U;X|Z)+[I(U;Z|V)-I(U;Y|V)+\epsilon]^-+3\epsilon'_n) \quad (70)
\end{aligned}$$

where (a) and (b) follow because V^n determines (F_v, W_v), (c) follows because U^n determines (F_u, W_u) and by (68), (d) follows because (V^n, Z^n, U^n) are i.i.d. and from the Markov chain $V-U-X-Z$, and (e) follows by (49), (50), and (69).

Case 6: Assume

$$H(V|Z) < R_{\mathrm{v}} + \widetilde{R}_{\mathrm{v}}, \tag{71}$$
$$H(U|Z) < R_{\mathrm{u}} + \widetilde{R}_{\mathrm{u}} \tag{72}$$

so that (F_v, W_v, Z^n) can reliably recover V^n, and (F_u, W_u, Z^n) can reliably recover U^n. Using (54), we obtain

$$\begin{aligned} &I(X^n; W, F|Z^n) \\ &\overset{(a)}{\leq} H(V^n, U^n|Z^n) - nH(U,V|X) + 2n\epsilon'_n \\ &\overset{(b)}{=} n(I(U;X|Z) + 2\epsilon'_n) \\ &\overset{(c)}{=} n(I(U;X|Z) + [I(U;Z|V) - I(U;Y|V) + \epsilon]^- + 2\epsilon'_n) \end{aligned} \tag{73}$$

where (a) follows because U^n determines (F_u, W_u) and V^n determines (F_v, W_v), (b) follows because (V^n, U^n, Z^n) are i.i.d. and from the Markov chain $V - U - X - Z$, and (c) follows by (49), (50), and (72).

Secrecy Leakage (to the Eavesdropper): Consider the secrecy leakage. We have

$$\begin{aligned} I(\widetilde{X}^n, Y^n; W, F|Z^n) &\overset{(a)}{=} H(W,F|Z^n) - H(W,F|\widetilde{X}^n) \\ &\overset{(b)}{\leq} H(W,F|Z^n) - H(W_u, F_u, V^n|\widetilde{X}^n) + n\epsilon'_n \\ &\overset{(c)}{\leq} H(W,F|Z^n) - nH(U,V|\widetilde{X}) + 2n\epsilon'_n \end{aligned} \tag{74}$$

where (a) follows from the Markov chain $(W,F) - \widetilde{X}^n - (Y^n, Z^n)$, (b) follows since $(W_v, F_v, \widetilde{X}^n)$ can reliably recover V^n due to the Markov chain $V^n - \widetilde{X}^n - Y^n$ and (43), and (c) follows by (44) since $(W_u, F_u, V^n, \widetilde{X}^n)$ can reliably recover U^n due to the inequality $H(U|V,Y) \geq H(U|V,\widetilde{X})$ that can be proved similarly as in (55), and because $(U^n, V^n, \widetilde{X}^n)$ are i.i.d.

Similar to the analysis of the privacy leakage to the eavesdropper, we need to analyze the same six decodability cases to consider whether (F_v, W_v, Z^n) can recover V^n and whether (F_u, W_u, V^n, Z^n) or (F_u, W_u, Z^n) can recover U^n. One can show that all steps applied in Cases 1-6 for the privacy leakage to the eavesdropper follow also for the Cases 1-6 for the secrecy leakage by replacing X with $\widetilde{X}$. We; therefore, list the results for Cases 1-6 as follows.

Case 1: We obtain for (56) and (57) that

$$\begin{aligned} &I(\widetilde{X}^n, Y^n; W, F|Z^n) \\ &\leq n(I(U;\widetilde{X}|Z) + [I(U;Z|V) - I(U;Y|V) + \epsilon]^- + 2\epsilon'_n). \end{aligned} \tag{75}$$

Case 2: We obtain for (59) and (60) that

$$\begin{aligned} &I(\widetilde{X}^n, Y^n; W, F|Z^n) \\ &\leq n(I(U;\widetilde{X}|Z) + [I(U;Z|V) - I(U;Y|V) + \epsilon]^- + 2\epsilon'_n). \end{aligned} \tag{76}$$

Case 3: We obtain for (62) and (63) that

$$\begin{aligned} &I(\widetilde{X}^n, Y^n; W, F|Z^n) \\ &\leq n(I(U;\widetilde{X}|Z) + [I(U;Z|V) - I(U;Y|V) + \epsilon]^- + 2\epsilon'_n). \end{aligned} \tag{77}$$

Case 4: We obtain for (65) and (66) that

$$\begin{aligned} &I(\widetilde{X}^n, Y^n; W, F|Z^n) \\ &\leq n(I(U;\widetilde{X}|Z) + [I(U;Z|V) - I(U;Y|V) + \epsilon]^- + 2\epsilon'_n). \end{aligned} \tag{78}$$

Case 5: We obtain for (68) and (69) that

$$\begin{aligned} &I(\widetilde{X}^n, Y^n; W, F|Z^n) \\ &\leq n(I(U;\widetilde{X}|Z) + [I(U;Z|V) - I(U;Y|V) + \epsilon]^- + 3\epsilon'_n). \end{aligned} \tag{79}$$

Case 6: We obtain for (71) and (72) that

$$\begin{aligned} &I(\widetilde{X}^n, Y^n; W, F|Z^n) \\ &\leq n(I(U;\widetilde{X}|Z) + [I(U;Z|V) - I(U;Y|V) + \epsilon]^- + 2\epsilon'_n). \end{aligned} \tag{80}$$

Now suppose the public indices F are generated uniformly at random. The encoder $\mathsf{Enc}(\cdot)$ generates (V^n, U^n) according to $P_{V^nU^n|\widetilde{X}^nF_{\mathrm{v}}F_{\mathrm{u}}}$ obtained from the binning scheme above to compute the bins W_{v} from V^n and W_{u} from U^n, respectively. This procedure induces a joint probability distribution that is almost equal to $P_{VU\widetilde{X}XYZ}$ fixed above [44, Section 1.6]. We remark that the privacy and secrecy leakage metrics considered above are expectations over all possible realizations $F = f$. Thus, using a time-sharing random variable Q and applying the selection lemma [46, Lemma 2.2] to each decodability case separately, these results prove the achievability for Theorem 1 by choosing an $\epsilon > 0$ such that $\epsilon \to 0$ when $n \to \infty$. □

B. Converse Proof of Theorem 1

Proof Sketch. Suppose for some $\delta_n > 0$ and $n \geq 1$, there exists a pair of encoders and decoders such that (2)-(6) are satisfied for some tuple $(R_{\mathrm{s}}, R_{\mathrm{w}}, R_{\ell,\mathrm{Dec}}, R_{\ell,\mathrm{Eve}})$.

Let $V_i \triangleq (W, Y_{i+1}^n, Z^{i-1})$ and $U_i \triangleq (W, X^{i-1}, Y_{i+1}^n, Z^{i-1})$, which satisfy the Markov chain $V_i - U_i - \widetilde{X}_i - X_i - (Y_i, Z_i)$ for all $i \in [1:n]$ by definition of the source statistics.

Admissibility of U: Define

$$\epsilon_n = \delta_n \log(|\widetilde{\mathcal{X}}||\mathcal{Y}|) + \frac{H_b(\delta_n)}{n} \tag{81}$$

where $H_b(\delta) = -\delta\log\delta - (1-\delta)\log(1-\delta)$ is the binary entropy function, so that $\epsilon_n \to 0$ if $\delta_n \to 0$. Using (2) and Fano's inequality, we obtain

$$\begin{aligned} n\epsilon_n &\geq H(f^n|\widehat{f^n}) \overset{(a)}{=} H(f^n|\bar{f}^n) \overset{(b)}{=} \sum_{i=1}^n H(f_i|\bar{f}_i) \\ &\geq \sum_{i=1}^n H(f_i|\bar{f}^n) \overset{(c)}{\geq} \sum_{i=1}^n H(f_i|W, Y^n) \\ &\geq \sum_{i=1}^n H(f_i|W, Y^n, X^{i-1}, Z^{i-1}) \\ &\overset{(d)}{=} \sum_{i=1}^n H(f_i|W, Y_{i+1}^n, X^{i-1}, Z^{i-1}, Y_i) \\ &\overset{(e)}{=} \sum_{i=1}^n H(f_i|U_i, Y_i) \end{aligned} \tag{82}$$

where (a) follows from [31, Lemma 2] so that when $n \to \infty$, there exists an i.i.d. random variable $\bar{f}^n$ such that $H(f^n|\widehat{f^n}) = H(f^n|\bar{f}^n)$ and $\widehat{f^n} - \bar{f}^n - (W, Y^n)$ form a Markov chain, (b) follows because $(f^n, \bar{f}^n)$ are i.i.d., (c) follows from the Markov chain $f^n - (W, Y^n) - \bar{f}^n$ and permits randomized decoding, (d) follows from the Markov chain for all $i \in [1:n]$

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

and (e) follows from the definition of U_i.

Storage (Public Message) Rate: We have

$$\begin{aligned}
&n(R_{\mathrm{w}} + \delta_n) \overset{(a)}{\geq} \log|\mathcal{W}| \geq H(W|Y^n) - H(W|\widetilde{X}^n, Y^n)\\
&= I(\widetilde{X}^n; W|Y^n) = H(\widetilde{X}^n|Y^n) - H(\widetilde{X}^n|W, Y^n)\\
&= H(\widetilde{X}^n|Y^n) - \sum_{i=1}^n H(\widetilde{X}_i|\widetilde{X}^{i-1}, W, Y^n)\\
&\overset{(b)}{=} H(\widetilde{X}^n|Y^n) - \sum_{i=1}^n H(\widetilde{X}_i|\widetilde{X}^{i-1}, W, Y_{i+1}^n, Y_i)\\
&\overset{(c)}{\geq} H(\widetilde{X}^n|Y^n) - \sum_{i=1}^n H(\widetilde{X}_i|X^{i-1}, Z^{i-1}, W, Y_{i+1}^n, Y_i)\\
&\overset{(d)}{=} nH(\widetilde{X}|Y) - \sum_{i=1}^n H(\widetilde{X}_i|U_i, Y_i) = \sum_{i=1}^n I(U_i; \widetilde{X}_i|Y_i)
\end{aligned} \tag{84}$$

where (a) follows by (4), (b) follows from the Markov chain for all $i \in [1:n]$

$$Y^{i-1} - (\widetilde{X}^{i-1}, W, Y_{i+1}^n, Y_i) - \widetilde{X}_i \tag{85}$$

(c) follows from the data processing inequality applied to the Markov chain for all $i \in [1:n]$

$$(X^{i-1}, Z^{i-1}) - (\widetilde{X}^{i-1}, W, Y_{i+1}^n, Y_i) - \widetilde{X}_i \tag{86}$$

and (d) follows from the definition of U_i.

Privacy Leakage to the Decoder: We obtain

$$\begin{aligned}
&n(R_{\ell,\mathrm{Dec}} + \delta_n) \overset{(a)}{\geq} H(W|Y^n) - H(W|X^n)\\
&= \sum_{i=1}^n \Big[I(W; X_i|X^{i-1}) - I(W; Y_i|Y_{i+1}^n)\Big]\\
&\overset{(b)}{=} \sum_{i=1}^n \Big[I(W; X_i|X^{i-1}, Y_{i+1}^n) - I(W; Y_i|Y_{i+1}^n, X^{i-1})\Big]\\
&\overset{(c)}{=} \sum_{i=1}^n \Big[I(W; X_i|X^{i-1}, Z^{i-1}, Y_{i+1}^n)\\
&\qquad - I(W; Y_i|Y_{i+1}^n, X^{i-1}, Z^{i-1})\Big]\\
&\overset{(d)}{=} \sum_{i=1}^n \Big[I(W, X^{i-1}, Z^{i-1}, Y_{i+1}^n; X_i)\\
&\qquad - I(W, Y_{i+1}^n, X^{i-1}, Z^{i-1}; Y_i)\Big]\\
&\overset{(e)}{=} \sum_{i=1}^n \Big[I(U_i; X_i) - I(U_i; Y_i)\Big] \overset{(f)}{=} \sum_{i=1}^n I(U_i; X_i|Y_i)
\end{aligned} \tag{87}$$

where (a) follows by (5) and from the Markov chain $W - X^n - Y^n$, (b) follows from Csiszár's sum identity [47], (c) follows from the Markov chains

$$Z^{i-1} - (X^{i-1}, Y_{i+1}^n) - (X_i, W) \tag{88}$$

$$Z^{i-1} - (X^{i-1}, Y_{i+1}^n) - (Y_i, W) \tag{89}$$

(d) follows because X^n is i.i.d. and the measurement channels are memoryless, (e) follows from the definition of U_i, and (f) follows from the Markov chain $U_i - X_i - Y_i$ for all $i \in [1:n]$.

Privacy Leakage to the Eavesdropper: We obtain

$$\begin{aligned}
&n(R_{\ell,\mathrm{Eve}} + \delta_n)\\
&\overset{(a)}{\geq} [H(W|Z^n) - H(W|Y^n)] + [H(W|Y^n) - H(W|X^n)]\\
&= \sum_{i=1}^n \Big[I(W; Y_i|Y_{i+1}^n) - I(W; Z_i|Z^{i-1})\Big]\\
&\qquad + \sum_{i=1}^n \Big[I(W; X_i|X^{i-1}) - I(W; Y_i|Y_{i+1}^n)\Big]\\
&\overset{(b)}{=} \sum_{i=1}^n \Big[I(W; Y_i|Y_{i+1}^n, Z^{i-1}) - I(W; Z_i|Z^{i-1}, Y_{i+1}^n)\Big]\\
&\qquad + \sum_{i=1}^n \Big[I(W; X_i|X^{i-1}, Y_{i+1}^n) - I(W; Y_i|Y_{i+1}^n, X^{i-1})\Big]\\
&\overset{(c)}{=} \sum_{i=1}^n \Big[I(W; Y_i|Y_{i+1}^n, Z^{i-1}) - I(W; Z_i|Z^{i-1}, Y_{i+1}^n)\Big]\\
&\qquad + \sum_{i=1}^n \Big[I(W; X_i|X^{i-1}, Y_{i+1}^n, Z^{i-1})\\
&\qquad\qquad - I(W; Y_i|Y_{i+1}^n, X^{i-1}, Z^{i-1})\Big]\\
&\overset{(d)}{=} \sum_{i=1}^n \Big[I(W, Y_{i+1}^n, Z^{i-1}; Y_i) - I(W, Z^{i-1}, Y_{i+1}^n; Z_i)\Big]\\
&\qquad + \sum_{i=1}^n \Big[I(W, X^{i-1}, Y_{i+1}^n, Z^{i-1}; X_i)\\
&\qquad\qquad - I(W, Y_{i+1}^n, X^{i-1}, Z^{i-1}; Y_i)\Big]\\
&\overset{(e)}{=} \sum_{i=1}^n \Big[I(V_i; Y_i) - I(V_i; Z_i)\\
&\qquad\qquad + I(U_i, V_i; X_i) - I(U_i, V_i; Y_i)\Big]\\
&= \sum_{i=1}^n \Big[- I(U_i, V_i; Z_i) + I(U_i, V_i; X_i)\\
&\qquad\qquad + (I(U_i; Z_i|V_i) - I(U_i; Y_i|V_i))\Big]\\
&\overset{(f)}{\geq} \sum_{i=1}^n \Big[I(U_i; X_i|Z_i) + [I(U_i; Z_i|V_i) - I(U_i; Y_i|V_i)]^-\Big]
\end{aligned} \tag{90}$$

where (a) follows by (6) and from the Markov chain $W - X^n - Z^n$, (b) follows from Csiszár's sum identity, (c) follows from the Markov chains in (88) and (89), (d) follows because X^n is i.i.d. and the measurement channels are memoryless,

(e) follows from the definitions of V_i and U_i, and (f) follows from the Markov chain $V_i - U_i - X_i - Z_i$ for all $i \in [1:n]$.

Secrecy Leakage (to the Eavesdropper): We have

$$\begin{aligned}
&n(R_{\mathrm{s}}+\delta_n)\\
&\overset{(a)}{\geq} [H(W|Z^n)-H(W|Y^n)]+[H(W|Y^n)-H(W|\widetilde{X}^n,Y^n)]\\
&\overset{(b)}{=} \sum_{i=1}^{n}\Big[I(W;Y_i|Y_{i+1}^n)-I(W;Z_i|Z^{i-1})\Big]\\
&\qquad+\Big[nH(\widetilde{X}|Y)-\sum_{i=1}^{n}H(\widetilde{X}_i|\widetilde{X}^{i-1},W,Y^n)\Big]\\
&\overset{(c)}{=} \sum_{i=1}^{n}\Big[I(W;Y_i|Y_{i+1}^n,Z^{i-1})-I(W;Z_i|Z^{i-1},Y_{i+1}^n)\Big]\\
&\qquad+\Big[nH(\widetilde{X}|Y)-\sum_{i=1}^{n}H(\widetilde{X}_i|\widetilde{X}^{i-1},W,Y_{i+1}^n,Y_i)\Big]\\
&\overset{(d)}{\geq} \sum_{i=1}^{n}\Big[I(W,Y_{i+1}^n,Z^{i-1};Y_i)-I(W,Z^{i-1},Y_{i+1}^n;Z_i)\Big]\\
&\qquad+\Big[nH(\widetilde{X}|Y)-\sum_{i=1}^{n}H(\widetilde{X}_i|X^{i-1},Z^{i-1},W,Y_{i+1}^n,Y_i)\Big]\\
&\overset{(e)}{=} \sum_{i=1}^{n}\Big[I(V_i;Y_i)-I(V_i;Z_i)+I(U_i,V_i;\widetilde{X}_i|Y_i)\Big]\\
&\overset{(f)}{=} \sum_{i=1}^{n}\Big[I(V_i;Y_i)-I(V_i;Z_i)+I(U_i,V_i;\widetilde{X}_i)-I(U_i,V_i;Y_i)\Big]\\
&= \sum_{i=1}^{n}\Bigg[-I(U_i,V_i;Z_i)+I(U_i,V_i;\widetilde{X}_i)\\
&\qquad\qquad+(I(U_i;Z_i|V_i)-I(U_i;Y_i|V_i))\Bigg]\\
&\overset{(g)}{\geq} \sum_{i=1}^{n}\Big[I(U_i;\widetilde{X}_i|Z_i)+[I(U_i;Z_i|V_i)-I(U_i;Y_i|V_i)]^-\Big] \quad (91)
\end{aligned}$$

where (a) follows by (3), (b) follows because $(\widetilde{X}^n,Y^n)$ are i.i.d., (c) follows from Csiszár's sum identity and the Markov chain in (85), (d) follows because X^n is i.i.d. and the measurement channels are memoryless, and from the data processing inequality applied to the Markov chain in (86), (e) follows from the definitions of V_i and U_i, (f) follows from the Markov chain $(U_i,V_i)-\widetilde{X}_i-Y_i$ for all $i\in[1:n]$, and (g) follows from the Markov chain $V_i-U_i-\widetilde{X}_i-Z_i$ for all $i\in[1:n]$.

Introduce a uniformly distributed time-sharing random variable $Q\sim\mathrm{Unif}[1:n]$ independent of other random variables. Define $X=X_Q$, $\widetilde{X}=\widetilde{X}_Q$, $Y=Y_Q$, $Z=Z_Q$, $V=V_Q$, $U=(U_Q,Q)$, and $f=f_Q$ so that $(Q,V)-U-\widetilde{X}-X-(Y,Z)$ form a Markov chain. The converse proof of Theorem 1 follows by letting $\delta_n\to 0$.

Cardinality Bounds: We use the support lemma [47, Lemma 15.4]; see also [35, Appendix C]. One can preserve $P_{\widetilde{X}}$ by using $|\widetilde{\mathcal{X}}|-1$ real-valued continuous functions. We have to preserve two expressions for the two cases such that $I(U;Z|V,Q=q)>I(U;Y|V,Q=q)$ and $I(U;Z|V,Q=q)\leq I(U;Y|V,Q=q)$ for all $q\in\mathcal{Q}$, so one can limit the cardinality $|\mathcal{Q}|$ of Q to $|\mathcal{Q}|\leq 2$. Furthermore, we have to preserve five more expressions, i.e., $H(\widetilde{X}|U,V,Z)$, $H(\widetilde{X}|U,V,Y)$, $H(X|U,V,Y)$, $H(X|U,V,Z)$, and $(I(U;Z|V)-I(U;Y|V))$. Thus, one can limit the cardinality $|\mathcal{V}|$ of V to $|\mathcal{V}|\leq|\widetilde{\mathcal{X}}|+4$. Similarly, in addition to the $|\widetilde{\mathcal{X}}|-1$ real-valued continuous functions, one should preserve the same five expressions for the auxiliary random variable U. To satisfy the Markov condition $(Q,V)-U-\widetilde{X}-X-(Y,Z)$, one can limit the cardinality $|\mathcal{U}|$ of U to $|\mathcal{U}|\leq(|\widetilde{\mathcal{X}}|+4)^2$. □

APPENDIX B
PROOF OF THEOREM 3

A. Achievability (Inner Bound) Proof of Theorem 3

The achievability proof follows by using the OSRB method, as described below.

Proof Sketch. Similar to Appendix A-A, fix $P_{U_j|\widetilde{X}_j}$ and $P_{V_j|U_j}$ such that U_j is admissible for the function $f_j(\widetilde{X}_j,Y_j)$ for all $j\in[1:J]$ and let $(V_{[1:J]}^n,U_{[1:J]}^n,\widetilde{X}_{[1:J]}^n,X^n,Y_{[1:J]}^n,Z_{[1:J]}^n)$ be i.i.d. according to (29). We remark that since all n-letter random variables are i.i.d., U_j^n is also admissible for all $j\in[1:J]$.

Assign two random bin indices $(F_{\mathrm{v},j},W_{\mathrm{v},j})$ to each v_j^n, and assume $F_{\mathrm{v},j}\in[1:2^{n\widetilde{R}_{\mathrm{v},j}}]$ and $W_{\mathrm{v},j}\in[1:2^{nR_{\mathrm{v},j}}]$ for all $j\in[1:J]$. Similarly, for all $j\in[1:J]$ assign two indices $(F_{\mathrm{u},j},W_{\mathrm{u},j})$ to each u_j^n, where $F_{\mathrm{u},j}\in[1:2^{n\widetilde{R}_{\mathrm{u},j}}]$ and $W_{\mathrm{u},j}\in[1:2^{nR_{\mathrm{u},j}}]$. The public message is $W_j=(W_{\mathrm{v},j},W_{\mathrm{u},j})$ and indices $F_j=(F_{\mathrm{v},j},F_{\mathrm{u},j})$ represent the public choice of encoder-decoder pairs for all $j\in[1:J]$.

For all $j\in[1:J]$, using a Slepian-Wolf (SW) decoder, one can reliably estimate V_j^n from $(F_{\mathrm{v},j},W_{\mathrm{v},j},Y_j^n)$ if we have

$$\widetilde{R}_{\mathrm{v},j}+R_{\mathrm{v},j}>H(V_j|Y_j) \quad (92)$$

and one can reliably estimate U_j^n from $(F_{\mathrm{u},j},W_{\mathrm{u},j},Y_j^n,V_j^n)$ by using a SW decoder if we have

$$\widetilde{R}_{\mathrm{u},j}+R_{\mathrm{u},j}>H(U_j|V_j,Y_j). \quad (93)$$

Thus, applying the union bound, we can show that the reliability constraint in (8) is satisfied if (92) and (93) are satisfied for all $j\in[1:J]$.

The public index $F_{\mathrm{v},j}$ is almost independent of $\widetilde{X}_j^n$, so it is almost independent of $(V_{[1:J]\setminus\{j\}}^n,U_{[1:J]\setminus\{j\}}^n,\widetilde{X}_{[1:J]}^n,X^n,Y_{[1:J]}^n,Z_{[1:J]}^n)$, if we have

$$\widetilde{R}_{\mathrm{v},j}<H(V_j|\widetilde{X}_j), \qquad \forall j\in[1:J]. \quad (94)$$

The public index $F_{\mathrm{u},j}$ is almost independent of $(V_j^n,\widetilde{X}_j^n)$, so it is almost independent of $(V_{[1:J]}^n,U_{[1:J]\setminus\{j\}}^n,\widetilde{X}_{[1:J]}^n,X^n,Y_{[1:J]}^n,Z_{[1:J]}^n)$, if we have

$$\widetilde{R}_{\mathrm{u},j}<H(U_j|V_j,\widetilde{X}_j), \qquad \forall j\in[1:J]. \quad (95)$$

To satisfy the constraints (92)-(95), similar to Appendix A, we fix the rates to

$$\begin{aligned}
\widetilde{R}_{\mathrm{v},j}&=H(V_j|\widetilde{X}_j)-\epsilon, && \forall j\in[1:J] \quad (96)\\
R_{\mathrm{v},j}&=I(V_j;\widetilde{X}_j)-I(V_j;Y_j)+2\epsilon, && \forall j\in[1:J] \quad (97)\\
\widetilde{R}_{\mathrm{u},j}&=H(U_j|V_j,\widetilde{X}_j)-\epsilon, && \forall j\in[1:J] \quad (98)\\
R_{\mathrm{u},j}&=I(U_j;\widetilde{X}_j|V_j)-I(U_j;Y_j|V_j)+2\epsilon, && \forall j\in[1:J] \quad (99)
\end{aligned}$$

for any $\epsilon > 0$.

Storage (Public Message) Rate: (96)-(99) result in a storage (public message) rate $R_{\mathrm{w},j}$ of

$$\begin{aligned} R_{\mathrm{w},j} &= R_{\mathrm{v},j} + R_{\mathrm{u},j} = I(V_j, U_j; \widetilde{X}_j) - I(V_j, U_j; Y_j) + 4\epsilon \\ &\overset{(a)}{=} I(U_j; \widetilde{X}_j | Y_j) + 4\epsilon, \qquad \forall j \in [1:J] \end{aligned} \tag{100}$$

where (a) follows because $V_j - U_j - \widetilde{X}_j - Y_j$ form a Markov chain for all $j \in [1:J]$.

Privacy Leakage to the Decoder: We have

$$I(X^n; W_j, F_j | Y_j^n) \overset{(a)}{\leq} nI(U_j; X|Y_j) + 2\epsilon_n, \ \forall j \in [1:J] \tag{101}$$

where (a) follows for some $\epsilon_n > 0$ with $\epsilon_n \to 0$ when $n \to \infty$ by applying the steps in (52).

Privacy Leakage to the Eavesdropper: Suppose an additional virtual joint encoder assigns $4J$ indices $(F_{v,[1:J]}, W_{v,[1:J]}, F_{u,[1:J]}, W_{u,[1:J]})$ to each realization tuple $(v_1^n, v_2^n, \ldots, v_J^n, u_1^n, u_2^n, \ldots, u_J^n) \in \mathcal{V}_1 \times \mathcal{V}_2 \times \ldots \times \mathcal{V}_J \times \mathcal{U}_1 \times \mathcal{U}_2 \times \ldots \times \mathcal{U}_J$ such that

$$\sum_{j=1}^{J} (\widetilde{R}_{\mathrm{v},j} + R_{\mathrm{v},j}) > H(V_{[1:J]} | Y_{[1:J]}), \tag{102}$$

$$\sum_{j=1}^{J} (\widetilde{R}_{\mathrm{u},j} + R_{\mathrm{u},j}) > H(U_{[1:J]} | V_{[1:J]}, Y_{[1:J]}). \tag{103}$$

Thus, $(W_{v,[1:J]}, F_{v,[1:J]}, Y_{[1:J]}^n)$ can reliably recover $V_{[1:J]}^n$ and $(V_{[1:J]}^n, W_{u,[1:J]}, F_{u,[1:J]}, Y_{[1:J]}^n)$ can reliably recover $U_{[1:J]}^n$. Therefore, we have for the total storage rate that

$$\begin{aligned} \sum_{j=1}^{J} R_{\mathrm{w},j} &= \sum_{j=1}^{J} (R_{\mathrm{v},j} + R_{\mathrm{u},j}) \\ &\overset{(a)}{\geq} I(U_{[1:J]}, V_{[1:J]}; \widetilde{X}_{[1:J]}) - I(U_{[1:J]}, V_{[1:J]}; Y_{[1:J]}) \\ &\overset{(b)}{=} I(U_{[1:J]}; \widetilde{X}_{[1:J]} | Y_{[1:J]}) \end{aligned} \tag{104}$$

where (a) follows by (102) and (103), and because (94) and (95) ensure that $(F_{v,[1:J]}, F_{u,[1:J]})$ are almost mutually independent of $\widetilde{X}_{[1:J]}^n$ since $\sum_{j=1}^{J} (\widetilde{R}_{v,j} + \widetilde{R}_{u,j}) < H(U_{[1:J]}, V_{[1:J]} | \widetilde{X}_{[1:J]})$ such that equality is achieved when $n \to \infty$ and (b) follows from the Markov chain $V_{[1:J]} - U_{[1:J]} - \widetilde{X}_{[1:J]} - Y_{[1:J]}$.

Consider the privacy leakage to the eavesdropper. We have

$$\begin{aligned} &I(X^n; W_{[1:J]}, F_{[1:J]} | Z_{[1:J]}^n) \\ &\overset{(a)}{=} H(W_{[1:J]}, F_{[1:J]} | Z_{[1:J]}^n) - H(W_{[1:J]}, F_{[1:J]} | X^n) \\ &\overset{(b)}{=} H(W_{[1:J]}, F_{[1:J]} | Z_{[1:J]}^n) - nH(U_{[1:J]}, V_{[1:J]} | X) \\ &\quad + \sum_{j=1}^{J} \Big[H(V_j^n | V_{[1:j-1]}^n, W_{[1:J]}, F_{[1:J]}, X^n) \\ &\qquad + H(U_j^n | U_{[1:j-1]}^n, V_{[1:J]}^n, W_{[1:J]}, F_{[1:J]}, X^n) \Big] \\ &\overset{(c)}{\leq} H(W_{[1:J]}, F_{[1:J]} | Z_{[1:J]}^n) \\ &\quad - nH(U_{[1:J]}, V_{[1:J]} | X) + 2Jn\epsilon_n' \end{aligned} \tag{105}$$

where (a) follows from the Markov chain $Z_{[1:J]}^n - X^n - (W_{[1:J]}, F_{[1:J]})$, (b) follows since U_j^n determines $(W_{u,j}, F_{u,j})$ and V_j^n determines $(W_{v,j}, F_{v,j})$ for all $j \in [1:J]$, and $(U_{[1:J]}^n, V_{[1:J]}^n, X^n)$ are i.i.d., and (c) follows for some $\epsilon_n' > 0$ such that $\epsilon_n' \to 0$ when $n \to \infty$ because $(F_{v,j}, W_{v,j} X^n)$ can reliably recover V_j^n due to the Markov chain $V_j^n - X^n - Y_j^n$ and (92), and because $(W_{u,j}, F_{u,j}, V_j^n, X^n)$ can reliably recover U_j^n due to the inequality $H(U_j|V_j, Y_j) \geq H(U_j|V_j, X)$, proved in (55), for all $j \in [1:J]$.

We consider the six decodability cases considered in Appendix A-A by replacing $[(R_v + \widetilde{R}_v), (R_u + \widetilde{R}_u)]$ with $\Big[\Big(\sum_{j=1}^{J}(R_{v,j} + \widetilde{R}_{v,j})\Big), \Big(\sum_{j=1}^{J}(R_{u,j} + \widetilde{R}_{u,j})\Big)\Big]$, respectively, and $[H(V|Z), H(U|V,Z), H(U|Z)]$ with $[H(V_{[1:J]}|Z_{[1:J]}), H(U_{[1:J]}|V_{[1:J]}, Z_{[1:J]}), H(U_{[1:J]}|Z_{[1:J]})]$, respectively. Using these replacements, applying the steps in (58),(61), (64), (67), (70), and (73) in combination with (105), and by choosing trivial rates that satisfy (102) and (103), one can show that

$$\begin{aligned} &I(X^n; W_{[1:J]}, F_{[1:J]} | Z_{[1:J]}^n) \\ &\leq n[I(U_{[1:J]}; Z_{[1:J]} | V_{[1:J]}) - I(U_{[1:J]}; Y_{[1:J]} | V_{[1:J]}) + \epsilon]^- \\ &\qquad + n(I(U_{[1:J]}; X | Z_{[1:J]}) + 3J\epsilon_n'). \end{aligned} \tag{106}$$

Secrecy Leakage (to the Eavesdropper): Consider the secrecy leakage. We have

$$\begin{aligned} &I(\widetilde{X}_{[1:J]}^n, Y_{[1:J]}^n; W_{[1:J]}, F_{[1:J]} | Z_{[1:J]}^n) \\ &\overset{(a)}{=} H(W_{[1:J]}, F_{[1:J]} | Z_{[1:J]}^n) - H(W_{[1:J]}, F_{[1:J]} | \widetilde{X}_{[1:J]}^n) \\ &\overset{(b)}{\leq} H(W_{[1:J]}, F_{[1:J]} | Z_{[1:J]}^n) \\ &\qquad - H(U_{[1:J]}^n, V_{[1:J]}^n | \widetilde{X}_{[1:J]}^n) + 2Jn\epsilon_n' \end{aligned} \tag{107}$$

where (a) follows from the Markov chain $(W_{[1:J]}, F_{[1:J]}) - \widetilde{X}_{[1:J]}^n - (Y_{[1:J]}^n, Z_{[1:J]}^n)$, (b) follows for some $\epsilon_n' > 0$ such that $\epsilon_n' \to 0$ when $n \to \infty$ because U_j^n determines $(W_{u,j}, F_{u,j})$ and V_j^n determines $(W_{v,j}, F_{v,j})$, and $(W_{v,j}, F_{v,j}, \widetilde{X}_j^n)$ can reliably recover V_j^n due to the Markov chain $V_j^n - \widetilde{X}_j^n - Y_j^n$ and (92), and similarly $(W_{u,j}, F_{u,j}, V_j^n, \widetilde{X}_j^n)$ can reliably recover U_j^n because $H(U_j|V_j, Y_j) \geq H(U_j|V_j, \widetilde{X}_j)$, which can be proved as in (55).

By using the same joint virtual encoder used for the privacy-leakage to the eavesdropper analysis above and replacing X by $\widetilde{X}_{[1:J]}$ in the analyses of (106), we obtain from (107) that

$$\begin{aligned} &I(\widetilde{X}_{[1:J]}^n, Y_{[1:J]}^n; W_{[1:J]}, F_{[1:J]} | Z_{[1:J]}^n) \\ &\leq n[I(U_{[1:J]}; Z_{[1:J]} | V_{[1:J]}) - I(U_{[1:J]}; Y_{[1:J]} | V_{[1:J]}) + \epsilon]^- \\ &\qquad + n(I(U_{[1:J]}; \widetilde{X}_{[1:J]} | Z_{[1:J]}) + 3J\epsilon_n'). \end{aligned} \tag{108}$$

Suppose the public indices $F_{[1:J]}$ are generated uniformly at random. The encoder $\mathsf{Enc}_j(\cdot)$ generates (V_j^n, U_j^n) according to $P_{V_j^n U_j^n | \widetilde{X}_j^n F_{\mathrm{v},j} F_{\mathrm{u},j}}$ obtained from the binning scheme above to compute the bins $W_{\mathrm{v},j}$ from V_j^n and $W_{\mathrm{u},j}$ from U_j^n for all $j \in [1:J]$. This procedure induces a joint probability distribution that is almost equal to $P_{V_{[1:J]} U_{[1:J]} \widetilde{X}_{[1:J]} X Y_{[1:J]} Z_{[1:J]}}$ fixed above [44, Section 1.6]. We remark that the privacy and secrecy leakage metrics considered above are expectations over all

possible realizations $F_{[1:J]} = f_{[1:J]}$. Thus, using a time-sharing random variable Q such that $P_{QV_{[1:J]}} = P_Q \prod_{j=1}^{J} P_{V_j|Q}$ and applying the selection lemma to each decodability case separately, these results prove the achievability for the rate tuples given in Theorem 3 by choosing an $\epsilon > 0$ such that $\epsilon \to 0$ when $n \to \infty$. □

B. Converse (Outer Bound) Proof of Theorem 3

Proof Sketch. Suppose for some $\delta_n > 0$ and $n \geq 1$, there exists a pair of encoders and decoders such that (8)-(12) are satisfied for some tuple $(R_{\text{s}}, R_{\text{w},[1:J]}, R_{\ell,\text{Dec},[1:J]}, R_{\ell,\text{Eve}})$.

Let $V_{i,j} \triangleq (W_j, Y_{i+1,j}^n, Z_j^{i-1})$ and $U_{i,j} \triangleq (W_j, X^{i-1}, Y_{i+1,j}^n, Z_j^{i-1})$, which satisfy the Markov chain $V_{i,j} - U_{i,j} - X_{i,j} - X_i - (\widetilde{X}_{i,[1:J]\setminus j}, Y_{i,j}, Z_{i,j})$ for all $i \in [1:n]$ and $j \in [1:J]$ by definition of the source statistics.

Admissibility of $\mathbf{U_j}$: Define

$$\epsilon_n = \max_{j\in[1:J]} \Big(\delta_{n,j} \log(|\widetilde{\mathcal{X}}_j||\mathcal{Y}_j|) + \frac{H_b(\delta_{n,j})}{n}\Big) \tag{109}$$

so that $\epsilon_n \to 0$ if $\max_{j\in[1:J]} \delta_{n,j} = \delta_n \to 0$. Applying the union bound to (8) and using Fano's inequality, we obtain

$$\begin{aligned} n\epsilon_n &\geq H(f_j^n|\widehat{f_j^n}) \\ &\overset{(a)}{\geq} \sum_{i=1}^{n} H(f_{i,j}|U_{i,j}, Y_{i,j}), \quad \forall j \in [1:J] \end{aligned} \tag{110}$$

where (a) follows applying the steps in (82) and from the definition of $U_{i,j}$.

Storage (Public Message) Rate: We have for all $j \in [1:J]$ that

$$n(R_{\text{w},j} + \delta_n) \overset{(a)}{\geq} \log|\mathcal{W}_j| \overset{(b)}{\geq} \sum_{i=1}^{n} I(U_{i,j}; \widetilde{X}_{i,j}|Y_{i,j}) \tag{111}$$

where (a) follows by (10) and (b) follows by applying the steps in (84) and from the definition of $U_{i,j}$.

Privacy Leakage to the Decoder: We obtain for all $j \in [1:J]$ that

$$\begin{aligned} n(R_{\ell,\text{Dec},j} + \delta_n) &\overset{(a)}{\geq} H(W_j|Y_j^n) - H(W_j|X^n) \\ &\overset{(b)}{\geq} \sum_{i=1}^{n} I(U_{i,j}; X_i|Y_{i,j}) \end{aligned} \tag{112}$$

where (a) follows by (11) and from the Markov chain $W_j - X^n - Y_j^n$ and (b) follows by applying the steps in (87) and from the definition of $U_{i,j}$.

Sum-Storage Rate: We have for all $j \in [1:J]$ that

$$\begin{aligned} n\sum_{j=1}^{J}(R_{\text{w},j} + \delta_n) &\overset{(a)}{\geq} \log\left|\prod_{j=1}^{J}|\mathcal{W}_j|\right| \\ &\geq H(W_{[1:J]}|Y_{[1:J]}^n) - H(W_{[1:J]}|\widetilde{X}_{[1:J]}^n, Y_{[1:J]}^n) \\ &= H(\widetilde{X}_{[1:J]}^n|Y_{[1:J]}^n) - \sum_{i=1}^{n} H(\widetilde{X}_{i,[1:J]}|\widetilde{X}_{[1:J]}^{i-1}, Y_{[1:J]}^n, W_{[1:J]}) \\ &\overset{(b)}{=} H(\widetilde{X}_{[1:J]}^n|Y_{[1:J]}^n) \\ &\quad - \sum_{i=1}^{n} H(\widetilde{X}_{i,[1:J]}|\widetilde{X}_{[1:J]}^{i-1}, Y_{i+1,[1:J]}^n, Y_{i,[1:J]}, W_{[1:J]}) \\ &\overset{(c)}{\geq} H(\widetilde{X}_{[1:J]}^n|Y_{[1:J]}^n) \\ &\quad - \sum_{i=1}^{n} H(\widetilde{X}_{i,[1:J]}|X_{[1:J]}^{i-1}, Z_{[1:J]}^{i-1}, Y_{i+1,[1:J]}^n, Y_{i,[1:J]}, W_{[1:J]}) \\ &\overset{(d)}{=} \sum_{i=1}^{n} I(U_{i,[1:J]}; \widetilde{X}_{i,[1:J]}|Y_{i,[1:J]}) \end{aligned} \tag{113}$$

where (a) follows by (10), (b) follows from the Markov chain for all $i \in [1:n]$

$$Y_{[1:J]}^{i-1} - (\widetilde{X}_{[1:J]}^{i-1}, W_{[1:J]}, Y_{i,[1:J]}^n) - \widetilde{X}_{i,[1:J]} \tag{114}$$

(c) follows from applying the data processing inequality to the Markov chain for all $i \in [1:n]$

$$(X^{i-1}, Z_{[1:J]}^{i-1}) - (\widetilde{X}_{[1:J]}^{i-1}, W_{[1:J]}, Y_{i,[1:J]}^n) - \widetilde{X}_{i,[1:J]} \tag{115}$$

and (d) follows because $(\widetilde{X}_{[1:J]}^n, Y_{[1:J]}^n)$ are i.i.d. and from the definition of $U_{i,j}$ for all $j \in [1:J]$.

Privacy Leakage to the Eavesdropper: We obtain

$$\begin{aligned} &n(R_{\ell,\text{Eve}} + \delta_n) \\ &\overset{(a)}{\geq} [H(W_{[1:J]}|Z_{[1:J]}^n) - H(W_{[1:J]}|Y_{[1:J]}^n)] \\ &\quad + [H(W_{[1:J]}|Y_{[1:J]}^n) - H(W_{[1:J]}|X^n)] \\ &\overset{(b)}{=} \sum_{i=1}^{n} \Big[I(W_{[1:J]}; Y_{i,[1:J]}|Y_{i+1,[1:J]}^n, Z_{[1:J]}^{i-1}) \\ &\qquad - I(W_{[1:J]}; Z_{i,[1:J]}|Z_{[1:J]}^{i-1}, Y_{i+1,[1:J]}^n)\Big] \\ &\quad + \sum_{i=1}^{n} \Big[I(W_{[1:J]}; X_i|X^{i-1}, Y_{i+1,[1:J]}^n) \\ &\qquad - I(W_{[1:J]}; Y_{i,[1:J]}|Y_{i+1,[1:J]}^n, X^{i-1})\Big] \\ &\overset{(c)}{=} \sum_{i=1}^{n} \Big[I(W_{[1:J]}; Y_{i,[1:J]}|Y_{i+1,[1:J]}^n, Z_{[1:J]}^{i-1}) \\ &\qquad - I(W_{[1:J]}; Z_{i,[1:J]}|Z_{[1:J]}^{i-1}, Y_{i+1,[1:J]}^n)\Big] \\ &\quad + \sum_{i=1}^{n} \Big[I(W_{[1:J]}; X_i|X^{i-1}, Y_{i+1,[1:J]}^n, Z_{[1:J]}^{i-1}) \\ &\qquad - I(W_{[1:J]}; Y_{i,[1:J]}|Y_{i+1,[1:J]}^n, X^{i-1}, Z_{[1:J]}^{i-1})\Big] \end{aligned}$$

$$
\begin{aligned}
n(R_{\mathrm{s}}+\delta_n) &\overset{(a)}{\geq} [H(W_{[1:J]}|Z^n_{[1:J]})-H(W_{[1:J]}|Y^n_{[1:J]})] + [H(W_{[1:J]}|Y^n_{[1:J]})-H(W_{[1:J]}|\widetilde{X}^n_{[1:J]},Y^n_{[1:J]})]\\
&\overset{(b)}{=} \sum_{i=1}^n \Big[I(W_{[1:J]};Y_{i,[1:J]}|Y^n_{i+1,[1:J]}) - I(W_{[1:J]};Z_{i,[1:J]}|Z^{i-1}_{[1:J]})\Big] + \Big[nH(\widetilde{X}_{[1:J]}|Y_{[1:J]}) - \sum_{i=1}^n H(\widetilde{X}_{i,[1:J]}|\widetilde{X}^{i-1}_{[1:J]},W_{[1:J]},Y^n_{[1:J]})\Big]\\
&\overset{(c)}{=} \sum_{i=1}^n \Big[I(W_{[1:J]};Y_{i,[1:J]}|Y^n_{i+1,[1:J]},Z^{i-1}_{[1:J]}) - I(W_{[1:J]};Z_{i,[1:J]}|Z^{i-1}_{[1:J]},Y^n_{i+1,[1:J]})\Big]\\
&\quad + \Big[nH(\widetilde{X}_{[1:J]}|Y_{[1:J]}) - \sum_{i=1}^n H(\widetilde{X}_{i,[1:J]}|\widetilde{X}^{i-1}_{[1:J]},W_{[1:J]},Y^n_{i+1,[1:J]},Y_{i,[1:J]})\Big]\\
&\overset{(d)}{\geq} \sum_{i=1}^n \Big[I(W_{[1:J]},Y^n_{i+1,[1:J]},Z^{i-1}_{[1:J]};Y_{i,[1:J]}) - I(W_{[1:J]},Z^{i-1}_{[1:J]},Y^n_{i+1,[1:J]};Z_{i,[1:J]})\Big]\\
&\quad + \Big[nH(\widetilde{X}_{[1:J]}|Y_{[1:J]}) - \sum_{i=1}^n H(\widetilde{X}_{i,[1:J]}|X^{i-1},Z^{i-1}_{[1:J]},W_{[1:J]},Y^n_{i+1,[1:J]},Y_{i,[1:J]})\Big]\\
&\overset{(e)}{=} \sum_{i=1}^n \Big[I(V_{i,[1:J]};Y_{i,[1:J]}) - I(V_{i,[1:J]};Z_{i,[1:J]}) + I(U_{i,[1:J]},V_{i,[1:J]};\widetilde{X}_{i,[1:J]}|Y_{i,[1:J]})\Big]\\
&\overset{(f)}{=} \sum_{i=1}^n \Big[I(V_{i,[1:J]};Y_{i,[1:J]}) - I(V_{i,[1:J]};Z_{i,[1:J]}) + I(U_{i,[1:J]},V_{i,[1:J]};\widetilde{X}_{i,[1:J]}) - I(U_{i,[1:J]},V_{i,[1:J]};Y_{i,[1:J]})\Big]\\
&= \sum_{i=1}^n \Big[-I(U_{i,[1:J]},V_{i,[1:J]};Z_{i,[1:J]}) + I(U_{i,[1:J]},V_{i,[1:J]};\widetilde{X}_{i,[1:J]}) + I(U_{i,[1:J]};Z_{i,[1:J]}|V_{i,[1:J]}) - I(U_{i,[1:J]};Y_{i,[1:J]}|V_{i,[1:J]})\Big]\\
&\overset{(g)}{\geq} \sum_{i=1}^n \Big[[I(U_{i,[1:J]};Z_{i,[1:J]}|V_{i,[1:J]}) - I(U_{i,[1:J]};Y_{i,[1:J]}|V_{i,[1:J]})]^- + I(U_{i,[1:J]};\widetilde{X}_{i,[1:J]}|Z_{i,[1:J]})\Big]
\end{aligned}
\tag{116}
$$

$$
\begin{aligned}
&\overset{(d)}{=} \sum_{i=1}^n \Big[I(W_{[1:J]},Y^n_{i+1,[1:J]},Z^{i-1}_{[1:J]};Y_{i,[1:J]})\\
&\qquad - I(W_{[1:J]},Z^{i-1}_{[1:J]},Y^n_{i+1,[1:J]};Z_{i,[1:J]})\Big]\\
&\quad + \sum_{i=1}^n \Big[I(W_{[1:J]},X^{i-1},Y^n_{i+1,[1:J]},Z^{i-1}_{[1:J]};X_i)\\
&\qquad - I(W_{[1:J]},Y^n_{i+1,[1:J]},X^{i-1},Z^{i-1}_{[1:J]};Y_{i,[1:J]})\Big]\\
&\overset{(e)}{=} \sum_{i=1}^n \Big[I(V_{i,[1:J]};Y_{i,[1:J]}) - I(V_{i,[1:J]};Z_{i,[1:J]})\\
&\qquad + I(U_{i,[1:J]},V_{i,[1:J]};X_i)\\
&\qquad - I(U_{i,[1:J]},V_{i,[1:J]};Y_{i,[1:J]})\Big]\\
&= \sum_{i=1}^n \Big[-I(U_{i,[1:J]},V_{i,[1:J]};Z_{i,[1:J]}) + I(U_{i,[1:J]},V_{i,[1:J]};X_i)\\
&\qquad + I(U_{i,[1:J]};Z_{i,[1:J]}|V_{i,[1:J]})\\
&\qquad - I(U_{i,[1:J]};Y_{i,[1:J]}|V_{i,[1:J]})\Big]\\
&\overset{(f)}{\geq} \sum_{i=1}^n \Big[I(U_{i,[1:J]};X_i|Z_{i,[1:J]})\\
&\qquad + \big[I(U_{i,[1:J]};Z_{i,[1:J]}|V_{i,[1:J]})\\
&\qquad\quad - I(U_{i,[1:J]};Y_{i,[1:J]}|V_{i,[1:J]})\big]^-\Big]
\end{aligned}
\tag{117}
$$

where (a) follows by (12) and from the Markov chain $W_{[1:J]} - X^n - Z^n_{[1:J]}$, (b) follows from Csiszár's sum identity, (c) follows from the Markov chains for all $i \in [1:n]$

$$Z^{i-1}_{[1:J]} - (X^{i-1}, Y^n_{i+1,[1:J]}) - (X_i, W_{[1:J]}) \tag{118}$$

$$Z^{i-1}_{[1:J]} - (X^{i-1}, Y^n_{i+1,[1:J]}) - (Y_{i,[1:J]}, W_{[1:J]}) \tag{119}$$

(d) follows because X^n is i.i.d. and the measurement channels are memoryless, (e) follows from the definitions of $V_{i,j}$ and $U_{i,j}$ for all $j \in [1:J]$, and (f) follows from the Markov chain $V_{i,[1:J]} - U_{i,[1:J]} - X_i - Z_{i,[1:J]}$ for all $i \in [1:n]$.

Secrecy Leakage (to the Eavesdropper): We have (116) given at the top of this page, where (a) follows by (9), (b) follows since $(\widetilde{X}^n_{[1:J]}, Y^n_{[1:J]})$ are i.i.d., (c) follows from Csiszár's sum identity and the Markov chain in (114), (d) follows because X^n is i.i.d. and the measurement channels are memoryless, and from the data processing inequality applied to the Markov chain in (115), (e) follows from the definitions of $V_{i,[1:J]}$ and $U_{i,[1:J]}$, (f) follows from the Markov chain $(U_{i,[1:J]}, V_{i,[1:J]}) - \widetilde{X}_{i,[1:J]} - Y_{i,[1:J]}$ for all $i \in [1:n]$, and (g) follows from the Markov chain $V_{i,[1:J]} - U_{i,[1:J]} - \widetilde{X}_{i,[1:J]} - Z_{i,[1:J]}$ for all $i \in [1:n]$.

Introduce a uniformly distributed time-sharing random variable $Q \sim \mathrm{Unif}[1:n]$ independent of other random variables. Define $X = X_Q$, $\widetilde{X}_j = \widetilde{X}_{Q,j}$, $Y_j = Y_{Q,j}$, $Z_j = Z_{Q,j}$, $V_j = V_{Q,j}$, $U_j = (U_{Q,j}, Q)$, and $f_j = f_{Q,j}$ so that $(Q, V_j) - U_j - \widetilde{X}_j - X - (\widetilde{X}_{[1:J]\setminus j}, Y_j, Z_j)$ form a Markov chain for all $j \in [1:J]$. The converse proof of Theorem 3 follows by letting $\delta_n \to 0$.

Cardinality Bounds follow by using the support lemma as in Appendix A-B. □