Secure Multi-Function Computation with Private Remote Sources

Onur Günlü[1], Matthieu Bloch[2], and Rafael F. Schaefer[1]
[1]Chair of Communications Engineering and Security, University of Siegen, {onur.guenlue, rafael.schaefer}@uni-siegen.de
[2]School of Electrical and Computer Engineering, Georgia Institute of Technology, matthieu.bloch@ece.gatech.edu

***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 privacy and secrecy. Specifically, 1) the remote source should remain *private* from an eavesdropper and the fusion center, measured in terms of the information leaked about the remote source; 2) the function computed should remain *secret* from the eavesdropper, measured in terms of the information leaked about the arguments of the function, to ensure secrecy regardless of the exact function used. 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 differing only in the Markov chain conditions imposed are characterized.**

2021 IEEE International Symposium on Information Theory (ISIT) | 978-1-5386-8209-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISIT45174.2021.9518019

I. Introduction

The problem of distributed function computation consists in characterizing how multiple terminals that observe dependent random sequences can facilitate the computation of a function of their sequences at a fusion center by exchanging messages through public communication links [1], [2]. One application for which distributed function computation problem is relevant is network function virtualization [3] via, e.g., software defined networking. The use of distributed lossless source coding techniques [4] in such applications, may significantly reduce the public communication rate, called the *storage rate*, by allowing the fusion center to reconstruct the sequences observed by the terminals instead of communicating the exact sequences [5]. Furthermore, for certain function computations that only require the fusion center to recover a distorted version of the terminal sequences, distributed lossy source coding methods [6] further reduce the storage rate. Such reductions are crucial for next generation resource-limited networks, such as those formed by Internet-of-Things (IoT) devices that must aggregate sensor data and make decision using lightweight mechanisms [5], [7]–[9]; see [10]–[14] for extensions.

To capture emerging security concerns in IoT networks, one may include *secrecy* and *privacy* constraints in the distributed function computation problem. Secrecy requires the computed function outputs to be hidden from eavesdroppers [15] that have access to correlated observations and the exchanged public messages. Several variations of the secure function computation problem have been analyzed in the literature [16]–[22]. Privacy, in contrast, requires the source sequences observed at the terminals to remain partially hidden from eavesdroppers [23]. Operationally, the analysis of privacy leakage allows one to upper bound the secrecy leakage about a *future* function computed by the terminals using the same source sequences [24], [25]. In the present work, we extend [23] by imposing several 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, in the present work, we posit that there exists an underlying ground truth, called the *remote source*, of which terminals only observe noisy versions. Noisy measurements of a hidden source are generally modeled as measurements through broadcast channels (BCs) [26] to have a generic measurement model that allows noise components at different terminals to be correlated, as considered in [27], [28]. Such a hidden source model is proposed and motivated in [29] for authentication problems and in [27], [30] for key-agreement problems with a privacy constraint. The privacy constraints are therefore measured with respect to the remote source and therefore differ from the single privacy leakage constraint considered in [23], which is measured with respect to the random sequence of the transmitting terminal. As shown next, this leads to a different set of trade-offs between privacy leakage and storage rates.

These results are strict extensions of [23] as we consider a remote source. Our inner bound proofs use a different coding method from the ones used in the literature [23] to simplify the analysis and our outer bounds do not follow from previous results. 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, which has not been previously considered.

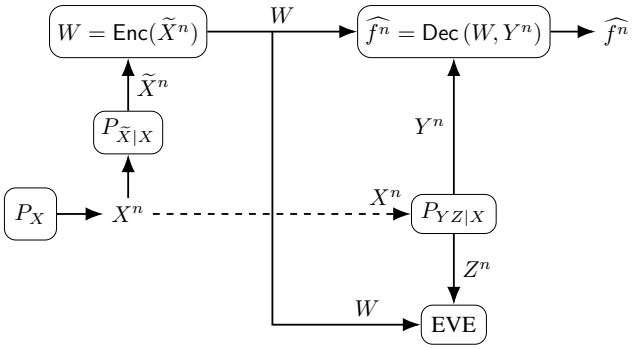

Fig. 1. Noisy measurements of a remote source used to compute a function securely and privately with the help of a public communication link.

II. Problem Definitions

A. Lossless and Lossy Single-Function Computation

Consider the function computation model in Fig. 1, where noisy measurements $(\widetilde{X}^n, Y^n)$ of a remote source X^n are inputs of a targeted function $f^n(\widetilde{X}^n, Y^n)$ such that

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

while the eavesdropper observes a correlated sequence Z^n and the public message W. The source $\mathcal{X}$ and measurement $\widetilde{\mathcal{X}}, \mathcal{Y}, \mathcal{Z}$ alphabets are finite sets. The encoder observes the noisy measurement $\widetilde{X}^n$ of the i.i.d. hidden source outputs X^n through a memoryless channel $P_{\widetilde{X}|X}$. The encoder computes the public message $W = \mathsf{Enc}(\widetilde{X}^n)$, which is sent over the public communication link. The decoder observes a noisy measurement Y^n of the hidden source X^n through a memoryless channel $P_{YZ|X}$ together with the public message W to estimate the targeted function $f^n(\widetilde{X}^n, Y^n)$ as $\widehat{f^n} = \mathsf{Dec}(W, Y^n)$. The eavesdropper (EVE) observes the output Z^n of the same memoryless channel and the public message W. We impose one secrecy and two privacy constraints, given in Definition 1 in addition to reliability (or distortion) and storage constraints, to the single-function computation problem depicted in Fig. 1 to characterize two rate regions, where lossless and lossy function computations satisfy, respectively, a reliability and a distortion constraint.

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 noiseless measurement case for the $\mathsf{Enc}(\cdot)$, for which $\widetilde{X}^n = X^n$, can be treated as a semi-deterministic BC.

Definition 1. A tuple $(R_{\mathrm{s}}, R_{\mathrm{w}}, R_{\ell,\mathrm{Dec}}, R_{\ell,\mathrm{Eve}})$ is *achievable* if, given $\delta > 0$, there exist $n \geq 1$, an encoder, and a decoder such that

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

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

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

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

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

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

The metric $I(f^n(\widetilde{X}^n, Y^n); W|Z^n)$ is a natural way to measure the information leakage to the eavesdropper who observes (W, Z^n) of the computed function $f^n(\cdot,\cdot)$, which is a proper secrecy-leakage metric since the function output is to be secured. 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 [23] represents the same secrecy-leakage metric as (3). Metrics in (5) and (6) measure the information leakage about the remote source to the decoder and eavesdropper, respectively, due to function computation.

The lossy single-function computation problem extends the lossless single-function computation model depicted in Fig. 1 by replacing the reliability constraint in (2) with an expected distortion constraint to allow a distorted reconstruction of the function $f(\cdot,\cdot)$.

Definition 2. A *lossy* tuple $(R_{\mathrm{s}}, R_{\mathrm{w}}, R_{\ell,\mathrm{Dec}}, R_{\ell,\mathrm{Eve}}, D)$ is *achievable* if, given $\delta > 0$, there exist $n \geq 1$, an encoder, and a decoder that satisfy (3)-(6) and

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

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 $\mathcal{R}_{\mathrm{D}}$ is the closure of the set of all achievable lossy distortion tuples.

B. Lossless and Lossy Multi-Function Computation

We extend the lossless single-function computation model by considering that the same hidden 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 $f_j^n(\widetilde{X}_j^n, Y_j^n) = \{f_j(\widetilde{X}_{j,i}, Y_{j,i})\}_{i=1}^n$ 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 $P_{Y_j Z_j|X}$. The eavesdropper observes $(Z_{[1:J]}^n, W_{[1:J]})$. This multi-function computation model is illustrated in Fig. 2 for $J = 2$. We next consider such multiple-function computations in the same network with *joint secrecy and privacy constraints* over all terminals. Lossless and lossy function computations are analyzed to provide inner and outer bounds for the multi-function rate regions.

Definition 3. A *multi-function* tuple $(R_{\mathrm{s}}, R_{\mathrm{w},[1:J]}, R_{\ell,\mathrm{Dec},[1:J]}, R_{\ell,\mathrm{Eve}})$ with j-th encoder measurements through $P_{\widetilde{X}_j|X}$ and j-th decoder measurements

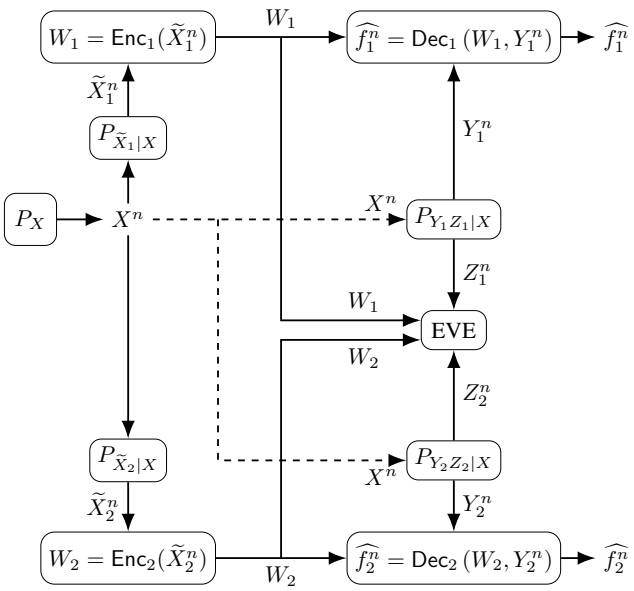

Fig. 2. Noisy measurements of the same hidden source used to compute $J = 2$ functions (via $2J = 4$ parties) securely and privately with the help of public communication links.

through $P_{Y_jZ_j|X}$ for all $j \in [1:J]$ is *achievable* if, given $\delta > 0$, there exist $n \geq 1$, and J encoder and decoder pairs such that

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

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

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

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

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

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

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

As above, the extension to the lossy case is obtained by allowing distorted function computations for multiple functions $f_j^n(\widetilde{X}_j^n, Y_j^n) = \{f_j(\widetilde{X}_{j,i}, Y_{j,i})\}_{i=1}^n$ computed from different measurements $(\widetilde{X}_j^n, Y_j^n)$ of the same hidden source X^n.

Definition 4. A *lossy multi-function* tuple $(R_{\mathrm{s}}, R_{\mathrm{w},[1:J]}, R_{\ell,\mathrm{Dec},[1:J]}, R_{\ell,\mathrm{Eve}}, D_{[1:J]})$ 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, given $\delta > 0$, there are some $n \geq 1$, 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 we have $d(f^n, \widehat{f^n}) = \frac{1}{n}\sum_{i=1}^n d(f_i, \widehat{f}_i)$. The *lossy multi-function* region $\mathcal{R}_{\mathrm{mf,D}}$ is the closure of the set of all achievable lossy distortion tuples.

III. Rate Regions

We first define 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 $U - \widetilde{X} - Y$ forms a Markov chain and $H(f(\widetilde{X}, Y)|U, Y) = 0$, i.e., (U, Y) determine $f(\widetilde{X}, Y)$.

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; see [31, Section V] for the complete proof and below for a proof sketch.

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

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

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

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

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

such that U is admissible and $(Q, V) - U - \widetilde{X} - X - (Y, Z)$ forms 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 $|\mathcal{Q}| \leq 2$, $|\mathcal{V}| \leq |\widetilde{X}| + 4$, and $|\mathcal{U}| \leq (|\widetilde{X}| + 4)^2$.

Proof Sketch: For the achievability proof, we use the *output statistics of random binning (OSRB)* method from [32] (see also [33]) that assigns random bin indices to auxiliary sequences $U^n = u^n$, where U^n is admissible, and $V^n = v^n$ such that the reliability constraint in (2) is satisfied. Using the OSRB method consecutively, six different recoverability cases that indicate whether it is possible obtain single-letter terms are analyzed. All six cases are bounded by the same mutual information terms. A time-sharing random variable Q is used to convexify the rate region. The converse proof follows by using standard properties of the Shannon entropy in addition to a single-letterization step from [28, Lemma 2] to prove the admissibility of U. ∎

In [23], some lower bounds include terms with the maximization operator $[\cdot]^+$. One can show that the rate regions in [23] 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 in Theorems 1-4.

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 (14)-(17) and*

$$D \geq \mathbb{E}[d(f(\widetilde{X}, Y), g(U, Y))] \tag{18}$$

for some function $g(\cdot,\cdot)$ such that $(Q,V)-U-\widetilde{X}-X-(Y,Z)$ forms 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 $\mathbb{E}[d(f(\widetilde{X},Y)), g(U,Y)] \leq D+\epsilon_n$, 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 [34, pp. 26], the distortion constraint (18) is satisfied. The converse proof of Theorem 2 follows from the converse proof of Theorem 1 by replacing the admissibility step in [31, Eq. (81)] with the steps

$$\begin{aligned}
D+\delta_n &\geq \mathbb{E}\Big[d\left(f^n(\widetilde{X}^n, Y^n), \widehat{f^n}(W, Y^n)\right)\Big] \\
&= \frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^{n} d\left(f_i(\widetilde{X}_i, Y_i), \widehat{f}_i(W, Y^n)\right)\Big] \\
&\overset{(a)}{\geq} \frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^{n} d\left(f(\widetilde{X}_i, Y_i), g(W, Y^n, X^{i-1}, Z^{i-1}, i)\right)\Big] \\
&\overset{(b)}{=} \frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^{n} d\left(f(\widetilde{X}_i, Y_i), g(W, Y_i^n, X^{i-1}, Z^{i-1}, i)\right)\Big] \\
&\overset{(c)}{=} \frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^{n} d\left(f(\widetilde{X}_i, Y_i), g(U_i, i, Y_i)\right)\Big]
\end{aligned} \tag{19}$$

where (a) follows since there exists a function $g(\cdot,\cdot)$ 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 $f(\widetilde{X}_i, Y_i)$ for all $i \in [1:n]$, because $g(\cdot,\cdot)$ has additional inputs, (b) follows from the Markov chain in [31, Eq. (82)], and (c) follows from the definition of $U_i = (W, X^{i-1}, Y_{i+1}^n, Z^{i-1})$. The cardinality bounds follow by preserving the same terms as in Theorem 1 in addition to $g(U,Y)=g(U,V,Y)$, which follows from V–(U,Y)–$g(U,Y)$. The region $\mathcal{R}_D$ is also convexified by using a time-sharing random variable Q. ∎

All rate regions in [23, Section III] (and, naturally, all previous rate regions recovered by manipulating the regions in [23, 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 $\mathcal{R}_{\text{mf}}$ in Theorem 3; see [31, Section VI] for the complete proof and below for a proof sketch.

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

$$\begin{aligned}
R_s \geq\ & [I(U_{[1:J]}; Z_{[1:J]}|V_{[1:J]}, Q) - I(U_{[1:J]}; Y_{[1:J]}|V_{[1:J]}, Q)]^- \\
& + I(U_{[1:J]}; \widetilde{X}_{[1:J]}|Z_{[1:J]})
\end{aligned} \tag{20}$$

$$R_{w,j} \geq I(U_j; \widetilde{X}_j|Y_j), \qquad \forall j \in [1:J] \tag{21}$$

$$\sum_{j=1}^{J} R_{w,j} \geq I(U_{[1:J]}; \widetilde{X}_{[1:J]}|Y_{[1:J]}) \tag{22}$$

$$R_{\ell,Dec,j} \geq I(U_j; X|Y_j), \qquad \forall j \in [1:J] \tag{23}$$

$$\begin{aligned}
R_{\ell,Eve} \geq\ & [I(U_{[1:J]}; Z_{[1:J]}|V_{[1:J]}, Q) - I(U_{[1:J]}; Y_{[1:J]}|V_{[1:J]}, Q)]^- \\
& + I(U_{[1:J]}; X|Z_{[1:J]})
\end{aligned} \tag{24}$$

where $P_{QV_{[1:J]}U_{[1:J]}\widetilde{X}_{[1:J]}XY_{[1:J]}Z_{[1:J]}}$ should be equal to

$$P_Q P_X \prod_{j=1}^{J} P_{V_j|U_j} P_{U_j|\widetilde{X}_j} P_{\widetilde{X}_j|X} P_{Y_j Z_j|X}. \tag{25}$$

(Outer Bound): *An outer bound for the multi-function region $\mathcal{R}_{mf}$ is the union of the rate tuples in (20) - (24) over all $P_{U_j|\widetilde{X}_j}$ and $P_{V_j|U_j}$ such that U_j is admissible and $(Q,V_j)-U_j-\widetilde{X}_j-X-(Y_j,Z_j)$ forms 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 $j \in [1:J]$.*

Proof Sketch: The inner bound proof follows by using the OSRB method for each encoder-decoder pair. An additional virtual joint encoder is considered to jointly tackle sets of random variables observed by different encoders in secrecy and privacy analyses. The outer bound proof follows by using standard properties of the Shannon entropy. ∎

Remark 2. *Inner and outer bounds differ because outer bounds define rate regions for the Markov chains $(Q,V_j)-U_j-\widetilde{X}_j-X-(Y_j,Z_j)$ for all $j \in [1:J]$, which are larger than the rate regions defined by inner bounds that satisfy (25).*

D. Lossy Multi-Function Computation

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

Theorem 4. (Inner Bound): *An achievable lossy multi-function region is the union over all $P_{U_j|\widetilde{X}_j}$ and $P_{V_j|U_j}$ 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 (20)-(24) and*

$$D_j \geq \mathbb{E}[d(f_j(\widetilde{X}_j, Y_j), g_j(U_j, Y_j))] \qquad \forall j \in [1:J] \tag{26}$$

for a set of functions $\{g_j(\cdot,\cdot)\}_{j=1}^J$ and where (25) 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 (20)-(24) and (26) over all $P_{U_j|\widetilde{X}_j}$ and $P_{V_j|U_j}$ such that $(Q,V_j)-U_j-\widetilde{X}_j-X-(Y_j,Z_j)$ forms a Markov chain for all $j \in [1:J]$. One*

can limit the cardinalities to $|\mathcal{Q}| \leq 2$, $|\mathcal{V}_j| \leq |\widetilde{\mathcal{X}}_j| + 6$, *and* $|\mathcal{U}_j| \leq (|\widetilde{\mathcal{X}}_j| + 6)^2$ *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 $\mathbb{E}[d(f_j(\widetilde{X}_j, Y_j)), g_j(U_j, Y_j)] \leq D_j + \epsilon_n$ 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 (26) are satisfied. The outer bound proof follows from the converse proof of Theorem 3 with the replacement of the admissibility step in [31, Eq. (108)] with the steps given in (19) for random variables and functions with the indices $j = 1, 2, \ldots, J$. ■

IV. Information Bottleneck Example

Consider a function computation scenario in a network where the evaluation of the rate region is an information bottleneck problem with a remote source. Consider the lossy single-function computation problem and suppose $X - Y - Z$ forms a Markov chain. We obtain the following rate region, which 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 for the Markov chain* $X - Y - Z$ *is the set of all tuples* $(R_s, R_w, R_{\ell,Dec}, R_{\ell,Eve}, D)$ *satisfying*

$$R_s \geq I(U; \widetilde{X}|Y) = I(U; \widetilde{X}) - I(U; Y) \tag{27}$$

$$R_w \geq I(U; \widetilde{X}|Y) = I(U; \widetilde{X}) - I(U; Y) \tag{28}$$

$$R_{\ell,Dec} \geq I(U; X|Y) = I(U; X) - I(U; Y) \tag{29}$$

$$R_{\ell,Eve} \geq I(U; X|Y) = I(U; X) - I(U; Y) \tag{30}$$

$$D \geq \mathbb{E}[d(f(\widetilde{X}, Y), g(U, Y))] \tag{31}$$

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

The proof of Corollary 1 follows by applying steps identical to the proof of [23, Corollary 3] to Theorem 2. 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 possible $P_{U|\widetilde{X}}$. This problem is an information bottleneck problem [35], [36]. The optimal function $g^*(\cdot, \cdot)$ that minimizes the lower bound in (31), depends on the realization $U = u$. If the distortion metric $d(\cdot, \cdot)$ is the Hamming distance, the optimal function $g^*(u, y)$ for all $(u, y) \in \mathcal{U} \times \mathcal{Y}$ is $g^*(u, y) = \arg\max_f P_{F|UY}(f|u, y)$ [23, Eq. (26)], 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 its inverse channel $P_{X|\widetilde{X}}$ is a binary symmetric channel (BSC) with crossover probability p, i.e., BSC(p), for any $0 \leq p \leq 0.5$. Furthermore, consider

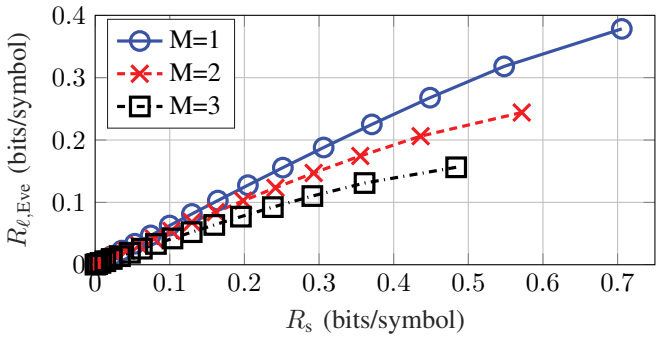

Fig. 3. Secrecy-leakage vs. privacyEve-leakage rate projection of the boundary tuples $(R_\text{s}, R_\text{w}, R_{\ell,\text{Dec}}, R_{\ell,\text{Eve}}, D)$ for $p = 0.06$ and for the number of independent BSC measurements at the decoder $M = 1, 2, 3$.

a measurement channel $P_{Y|X}$ for the decoder $\mathsf{Dec}(\cdot)$ that is a binary input symmetric output channel [37, p. 21], which can be decomposed into a mixture of binary subchannels as defined in [38, Section III-B] [39]. We remark that the rate region defined in Corollary 1 by (27)-(31) 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., $P_{YZ|X} = P_{Z|Y} P_{Y|X}$. Define $H_b(x) = -x \log x - (1-x) \log(1-x)$ as the binary entropy function and $H_b^{-1}(\cdot)$ as its inverse with range $[0, 0.5]$. Since $P_{\widetilde{X}XYZ}$ is fixed, to solve the information bottleneck problem given above the optimal auxiliary random variable U for these channels is such that $P_{\widetilde{X}|U}$ is a BSC with crossover probability $\widetilde{p} = (H_b^{-1}(H(X|U)) - p)/(1-2p)$ [25, Theorem 3].

Suppose $P_X \sim$ Bernoulli(0.5), $P_{\widetilde{X}|X} \sim$ BSC$(p = 0.06)$, and $P_{Y|X}$ is $M > 1$ independent BSCs each with crossover probability 0.15, which satisfies the assumptions listed above. Using auxiliary random variables $\sim$ BSC$(\widetilde{p})$, 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 $\widetilde{p}$, 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 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.

Acknowledgment

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.

REFERENCES

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

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

[3] 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.

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

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

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

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

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

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

[19] 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.

[20] 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.

[21] 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.

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

[23] 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.

[24] 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.

[25] 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.

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

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] O. Günlü, M. Bloch, and R. F. Schaefer, "Secure multi-function computations with private remote sources," May 2021, [Online]. Available: arxiv.org.

[32] 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.

[33] 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.

[34] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge University Press, 2011.

[35] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," Apr. 2000, [Online]. Available: arxiv.org/abs/physics/0004057.

[36] 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.

[37] R. G. Gallager, *Low-Density Parity-Check Codes*. Cambridge, MA: M.I.T. Press, 1963.

[38] 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.

[39] 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.