Secrecy by Design With Applications to Privacy and Compression

Yanina Y. Shkel, *Member, IEEE*, Rick S. Blum, *Fellow, IEEE*, and H. Vincent Poor, *Life Fellow, IEEE*

***Abstract*—Secrecy by design is examined as an approach to information-theoretic secrecy. The main idea behind this approach is to design an information processing system from the ground up to be perfectly secure with respect to an explicit secrecy constraint. The principal technical contributions are decomposition bounds that allow the representation of a random variable X as a deterministic function of (S, Z), where S is a given fixed random variable and Z is constructed to be independent of S. Using the problems of privacy and lossless compression as examples, the utility cost of applying secrecy by design is investigated. Privacy is studied in the setting of the privacy funnel function previously introduced in the literature and new bounds for the regime of zero information leakage are derived. For the problem of lossless compression, it is shown that strong information-theoretic guarantees can be achieved using a reduced secret key size and a quantifiable penalty on the compression rate. The fundamental limits for both problems are characterized with matching lower and upper bounds when the secret S is a deterministic function of the information source X.**

***Index Terms*—Data compression, privacy, information security, information entropy, random variables.**

I. Introduction

WE INTRODUCE an approach to partial secrecy which we call *secrecy by design* and investigate the utility cost of this approach by applying it to two information processing problems: data privacy and lossless data compression. The idea behind secrecy by design is to identify a (possibly random) function of the data that must be secure from an eavesdropper, and to assure perfect secrecy for this function during the design of the information processing system. For example, for the problem of privacy this is done during the design of the privacy-assuring mapping, while for the problem of compression this is done during the design of the compressor. Since perfect information-theoretic secrecy is prohibitive for many information processing systems [1], developing practical and theoretically sound approaches to partial secrecy is of paramount importance.

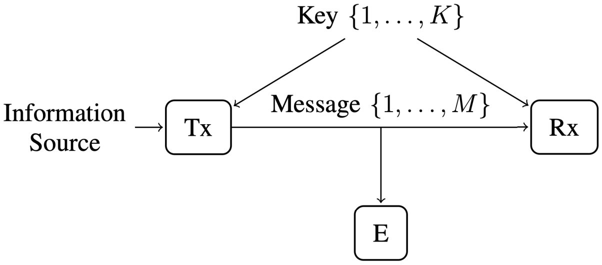

Fig. 1. Shannon cipher system. The transmitter (Tx) and the receiver (Rx) communicate a message over an unsecured channel where this message could be intercepted by the eavesdropper (E). The advantage that the transmitter and the receiver have over the eavesdropper is a shared secret key.

A. Perfect Secrecy

In his seminal work "Communication theory of secrecy systems", Shannon introduced a notion of perfect secrecy as complete statistical independence between publicly available data and private data [1]. Specifically, [1] analyzed a communication system, see Figure 1, in which one of M messages is transmitted over a communication channel 'in the clear'. The legitimate transmitter and receiver have an advantage over the eavesdropper in the form of a shared secret key, which is modeled by a random variable supported on $\{1, \ldots, K\}$. Perfect secrecy, in this case, is when the cipher text available to the eavesdropper is statistically independent from the true message being transmitted between the two trusted parties. A fundamental result of [1] is that perfect secrecy is possible if and only if $K \geq M$; that is, the shared secret key is at least the same length as the message.

Perfect secrecy via complete statistical independence is an intuitive and appealing notion of secrecy since it insures that the eavesdropper will not be able to make any inference about the information source. While perfect secrecy is expensive for communication systems, it is evidently completely prohibitive in other information processing settings. Consider a setting, see for example [2], [3], where an owner of a private database wishes to publish a sanitized version of the database for public use. The role of the eavesdropper is now played by a (not necessarily malicious) data analyst of the publicly available data who may wish to use it in unintended ways.

Manuscript received October 8, 2018; revised June 28, 2020; accepted November 7, 2020. Date of publication November 30, 2020; date of current version January 21, 2021. This work was supported in part by the U.S. National Science Foundation under Grant CNS-1702555, Grant ECCS-1744129, Grant CCF 0939370, and Grant CCF-1908308. This article was presented in part at the 2019 IEEE International Symposium on Information Theory and in part at the 2018 Conference on Information Sciences and Systems (CISS). *(Corresponding author: Yanina Y. Shkel.)*

Yanina Y. Shkel is with the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland (e-mail: yanina.shkel@epfl.ch).

Rick S. Blum is with the Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA 18015 USA (e-mail: rblum@lehigh.edu).

H. Vincent Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08540 USA (e-mail: poor@princeton.edu).

Communicated by S. Watanabe, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2020.3041224

Specifically, the database may contain records $X_1, \ldots, X_n$ which are kept private, and only sanitized or obscured versions $Z_1, \ldots, Z_n$ are released. Perfect secrecy in the sense of [1] would mean complete statistical independence between the private data $X_1, \ldots, X_n$ and the sanitized data $Z_1, \ldots, Z_n$. This would certainly ensure privacy, but would also render the publicly available data useless for any meaningful task.

B. Partial Secrecy

Since perfect secrecy is impractical for many information processing systems, it is necessary to relax the secrecy requirement to partial secrecy. A number of approaches to partial secrecy have been developed. In the setting of communication, cryptographic approaches assume a computationally limited eavesdropper and rely on computational hardness of certain problems. For example, the well-known RSA algorithm is based on the practical difficulty of factoring the product of two large prime numbers [4].

In a more general setting, approaches to partial secrecy focus on limiting the amount of information that is revealed to the eavesdropper. Traditional approaches to information theoretic secrecy focus on mechanisms that minimize measures of information leakage. For example, the original measure of leakage proposed in [1] is equivocation, or entropy of the source given the message; equivocation has been subsequently used as the default measure of partial secrecy in information-theoretic security [5]–[8]. Other approaches to partial secrecy — motivated by guessing, maximum information leakage, and rate-distortion [9]–[18] — have also been proposed. The problem of partial secrecy in the setting of database privacy has been widely studied in the computer science literature; approaches such as differential privacy [2], [3], (ϵ, δ)-differential privacy [19], and local differential privacy [20]–[22] can also be viewed through the lens of information leakage [23]–[27].

The leakage-based approach to partial secrecy has a number of notable advantages. First, it is a relaxation of perfect secrecy in a sense that zero leakage implies perfect secrecy for all of the above measures. Secondly, it is convenient to assign a number to the secrecy level of a system and to use that number for evaluation and design. Finally, measures like differential privacy have been shown to have desirable properties such as composeability across sequential and parallel usages, as well as graceful degradation in the presence of correlated data [3].

Criticisms of the leakage approach come down to two questions: *operational interpretability* and *utility cost*. For example, equivocation and rate-distortion approaches have been criticized as secrecy measures since their operational implications are not always clear [16], [28]. Likewise, differential privacy has been shown to be too restrictive for many applications, and not restrictive enough for some [29]. In general, one could argue that different measures of leakage are appropriate for different applications; however, this still leaves open the question of how to select the best measure of leakage given a specific application.

C. Secrecy by Design

In this work we propose another path towards relaxing perfect secrecy which we call *secrecy by design*.[1] The main idea behind this approach is to design an information processing system from the ground up to be perfectly secure with respect to an explicit secrecy constraint. For example, suppose X represents the private data or message and Z represents the publicly available data or message. If the system designer can identify some (possibly random) function of X — denoted by S — that completely and precisely encapsulates the secrecy needs of the system, the aim of system design should be to perfectly secure S. Mathematically, we model this by requiring that Z is statistically independent from S. That is, secrecy by design imposes the constraint

$$I(S; Z) = 0 \tag{1}$$

instead of the perfect secrecy constraint

$$I(X; Z) = 0. \tag{2}$$

An immediate appeal of secrecy by design is that it makes the secrecy guarantee for the system explicit and easy to interpret: no statistical inference about S could be made by the eavesdropper.

Our overarching motivation is the Internet of Things (IoT) setting in which the information processing components are resource constrained, have stringent delay requirements, and may be expected to last for decades after installation. Additive degradation in privacy offered by many leakage approaches may not be strict enough for the IoT setting, while limited computational and memory resources require secrecy solutions to be specially tailored to each application. Consider, for example, smart meters for monitoring in-home electricity use dynamics. A wide adaptation of this technology would offer great economic and environmental benefits, but they are also known to pose serious threats to individual privacy by leaking sensitive information like home occupancy patterns [30].

Potential use cases for secrecy by design are modeled by the following examples.

Example 1 (N-User Process): Let $\mathcal{S} = \{1, \ldots, N\}$ be the index of one of N users and $\mathcal{A}$ be an arbitrary finite alphabet over which some process in the system is taking place. Suppose that every user has its own probability distribution P_s over $\mathcal{A}$. We consider two versions of this setting.

First, suppose that $\mathcal{X} = \mathcal{S} \times \mathcal{A}$. For example, $x = (s, a)$ could designate a job that user $s \in \mathcal{S}$ requested from a list of possible jobs $\mathcal{A}$. Suppose that

$$P_X(s, a) = P_s(a). \tag{3}$$

In this case the data consists of the job being requested and the metadata communicating the identity of the user. Next, suppose that $\mathcal{X} = \mathcal{A}$ and

$$P_X(a) = \sum_{s \in \mathcal{S}} P_s(a) P_S(s). \tag{4}$$

[1] We borrow this terminology from the software engineering community where the principle of "security by design" means that the software has been designed from the foundation to be secure.

That is, the data consists of the job being requested without the metadata; however, the distribution of the data still depends on the identity of the user. Observe that in this case S is a random function of X.

The secret S in Example 1 could be more than just an identity of an individual. It could be any combination of features in the data that are well-known to reveal an individual's identity. It could also include other sensitive features: for example, a presence or absence of a disease in a medical history of an individual, or features which the data analyst is not allowed to discriminate against, such as race or gender. Mathematically, Example 1 captures a very general setting and subsumes most later examples in this work.

Example 2 (Exponential Family): Suppose $X \sim P_\beta$ where P_β belongs to an exponential family which is parametrized by Θ and the secret S is the sufficient statistic for this family. That is

$$P_\beta(x) \ = \exp\left(\beta^T \sigma(x) - \psi(\beta)\right) r(x) \tag{5}$$

for some $\beta \in \Theta$ and $S = \sigma(X)$.

Example 2 models the situation where the eavesdropper needs to be prevented from inferring the parameter $\beta \in \Theta$ in the data. Note that in this case S is a deterministic function of X.

To the best of our knowledge, this work is the first to focus on secrecy by design as a principled approach to information theoretic secrecy. However, there have been a number of relevant works that study optimal secrecy mechanism that target a function of data. Most notably, recent work on perfect privacy [31]–[33] investigates the feasibility of the secrecy by design principle outlined above in the context of the privacy funnel function. Other examples include perfect function secrecy in a rate-distortion setting [34], [35], and securing correlated random variables for compression using equivocation as a leakage measure [36]. Finally, maximal leakage [15], [16] is defined as the worst-case refinement in knowledge of some random transformation of data. It identifies securing a function of the data as the goal of partial secrecy; however, it does not target a specific function, but rather focuses on decreasing the leakage across all functions simultaneously.

In this work we investigate the utility cost of applying secrecy by design to two information processing problems: privacy and compression. Privacy is studied in the setting of the privacy funnel function and new bounds for the regime of zero information leakage are derived. For the problem of lossless compression, it is shown that strong information-theoretic guarantees can be achieved using a reduced secret key size and a quantifiable penalty on the compression rate. Our lower and upper bounds match for both problems when the secret S is a deterministic function of the information source X.

The rest of this paper is structured as follows. We conclude this section by introducing the information theoretic measures needed to state our results. In Section II we give an overview of our results. In Section III we provide decomposition bounds that allow us to represent a random variable X as a function of the secret S and a public random variable Z which is independent of S. We discuss applications of our decomposition bounds to privacy in Section IV and to compression in Section V. We end with concluding remarks in Section VI.

D. Notation and Preliminaries

Given a random variable Y jointly distributed with X, information and conditional information are given by[2]

$$\imath_X(x) \ = \log \frac{1}{P_X(x)} \tag{6}$$

and

$$\imath_{X|Y}(x|y) = \log \frac{1}{P_{X|Y}(x|y)} \tag{7}$$

respectively. The entropy and conditional entropy are given by

$$H(X) \ = \ \mathbb{E}\left[\imath_X(X)\right], \tag{8}$$

$$H(X|Y) \ = \ \mathbb{E}\left[\imath_{X|Y}(X|Y)\right], \tag{9}$$

respectively. Moreover,

$$H(X|Y=y) = \mathbb{E}\left[\imath_{X|Y}(X|Y)|Y=y\right]. \tag{10}$$

It follows that X and Y are independent if and only if

$$\imath_{X|Y}(x|y) = \imath_X(x), \tag{11}$$

for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Finally, recall that conditioning reduces entropy and

$$H(X) \geq H(X|Y) \tag{12}$$

always holds. However, $H(X|Y=y)$ could be greater than or smaller than $H(X)$; see [37].

Mutual information between X and Y is given by

$$I(X;Y) = H(X) - H(X|Y) \tag{13}$$

whenever the entropy of X is finite. It could be defined more generally [37]. We use the notation

$$X \leftrightarrow Y \leftrightarrow Z \tag{14}$$

to denote that

$$I(X;Z|Y) = 0. \tag{15}$$

In other words, $X \leftrightarrow Y \leftrightarrow Z$ form a Markov chain. Finally, we use

$$h(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} \tag{16}$$

to denote the binary entropy function.

II. Overview of the Results

In this work we focus on secrecy by design in the one-shot setting where a single realization of the information source X and a correlated secret S with a known joint distribution P_{XS} is given. We study mechanisms for constructing public information Z that is independent of S. Suppose S is a deterministic function of X, say $S = \mathsf{f_s}(X)$ for some secrecy function

$$\mathsf{f_s} : \mathcal{X} \rightarrow \mathcal{S}. \tag{17}$$

Then, we say Z is $\mathsf{f_s}$-secure. In general, we say that such Z is S-secure or simply secure when S is understood from context.

[2] Throughout this paper all logarithm and exponential functions are assumed to have base two.

A. Secure Decomposition Bounds

We begin by answering the following question. Given a random variable X and an arbitrary correlated random variable S, is it always possible to decompose X into S and an S-secure Z while capturing all (or most) of the information about X?

Example 3: Let $X = A^n = \{0,1\}^n$ and $S = A_1^k$. That is, the secret information is the first k bits of a^n. If the A_i's are independent and identically distributed (i.i.d) the answer to the question above is trivially "yes" with $Z = A_{k+1}^n$.

Example 3 is well-behaved in a sense that X and S interact in a manageable way and it is possible to find a secure decomposition by inspection. Consider a more general setting of the N-user process given in Example 1. In this case it is not enough to simply hide the value of $s \in \mathcal{S}$ since there is a statistical dependence between X and S.

Our first result, Lemma 1 in Section III, shows that it is always possible to find a decomposition for arbitrary correlated random variables X and S provided that some additional randomness is used. Moreover, the decomposition (S, Z) captures all information about X in a sense that X can be perfectly reconstructed from (S, Z); that is, X is a deterministic function of (S, Z). Lemma 1 is also known as the *Functional Representation Lemma* [38] and has been proved independently in [39]–[41] where it was shown that Z has finite support whenever X and S have finite support. Since it is a conceptual and a technical building block of many of our results, we present the full proof of Lemma 1.

Once we establish that a secure decomposition (S, Z) of X always exists, we study other properties of secure decompositions. For example, for the problem of compression the property of interest is the entropy of Z. We extend the analysis of Lemma 1 to the case when $\mathcal{X}$ is countably infinite and show in Lemma 2 that there exists an S-secure Z such that

$$H(Z) \leq \sum_{s \in \mathcal{S}} H(X|S = s). \tag{18}$$

We show a lower bound on the entropy of Z,

$$H(Z) \geq \max_{s \in \mathcal{S}} H(X|S = s) \tag{19}$$

in Lemma 3. We conclude Section III with Theorems 1-3 which give bounds on the information spectrum of Z in the style of [42]. In particular, Theorem 1 is enough to reconcile the gap between (18) and (19), and to show that (19) is the asymptotically optimal bound.

The rest of this work deals with applications of our decomposition bounds to problems of privacy and lossless compression. For the problem of privacy, X is decomposed into (S, Z) and Z is shared publicly, while S is kept private, see Figure 2. In the setting of the privacy funnel function the goal is to maximize the mutual information between X and Z. For the problem of compression, the first stage of our coding strategy uses the shared secret key to encode S with a one-time pad. The second stage reconciles the remaining uncertainty about X by compressing Z using any traditional strategy, see Figure 3. We show that the constructions in Figures 2 and 3 achieve the fundamental limits of privacy and lossless compression with matching converse bounds when S is a deterministic function of X.

B. Applications to Privacy

In Section IV we apply our decomposition bounds to the privacy funnel function originally introduced in [43]. The general privacy funnel function is given by

$$G_I(t, P_{XS}) = \inf\{I(S; Z) : I(X; Z) \geq t, S \leftrightarrow X \leftrightarrow Z\}. \tag{20}$$

In [31]–[33], special attention is paid to the perfect privacy[3] regime in which $I(S; Z) = 0$. Formally, this is done by studying the function

$$\mathsf{g}_0(P_{XS}) = \sup_{\substack{Z: I(S;Z)=0 \\ S \leftrightarrow X \leftrightarrow Z}} I(X; Z). \tag{21}$$

A distinguishing feature of the privacy funnel approach is that it imposes the Markov chain condition,

$$S \leftrightarrow X \leftrightarrow Z \tag{22}$$

on the S-secure construction of Z. In other words, the privacy-assuring mapping is required to be a function of the information source X only, and not of the secret S. In many applications of interest the value of S may actually be known and privacy-assuring mechanisms that incorporate this knowledge may lead to better system performance than the ones which are blind to the value of S. In order to address this setting, we define a secret-dependent perfect privacy function:

$$\mathsf{h}_0(P_{XS}) = \sup_{Z: I(S;Z)=0} I(X; Z). \tag{23}$$

It follows from (21) and (23) that

$$\mathsf{h}_0(P_{XS}) \geq \mathsf{g}_0(P_{XS}). \tag{24}$$

We show in Example 6 that the gap between $\mathsf{h}_0(P_{XS})$ and $\mathsf{g}_0(P_{XS})$ could be quite large in general.

The decomposition bounds in Section III are not immediately applicable to (21) since the Markov chain (22) is not enforced by Lemma 1. Nevertheless, we are able to leverage these bounds to make strong statements about the relationship between (21) and (23). Of particular interest is Theorem 7 which shows that

$$\mathsf{g}_0(P_{XS}) \leq \mathsf{h}_0(P_{XS}) \leq H(X|S) \tag{25}$$

and states the necessary and sufficient conditions for both inequalities in (25) to be equalities. These conditions hold, for example, when S is a deterministic function of X; they hold more generally and we discuss this in more detail in Section IV.

Finally, we remark that in the privacy funnel setting, mutual information is used as a proxy for utility. A motivation behind this is that $I(X; Z)$ shows up as a relevant measure of statistical dependence between X and Z in many information

[3] Note that [31]–[33] use perfect privacy to refer to complete statistical independence between Z and S. Compare this to perfect secrecy which we use to refer to perfect statistical independence between Z and X, as is done in [1].

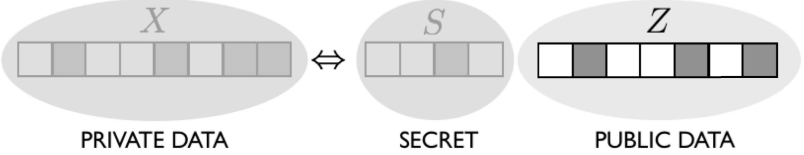

Fig. 2. Decomposition-based coding strategy for privacy. An information source X is represented as (S, Z) where S is the predetermined secret and Z is independent of S. Z is shared publicly while S and X are kept private. The goal is to maximize the mutual information between X and Z.

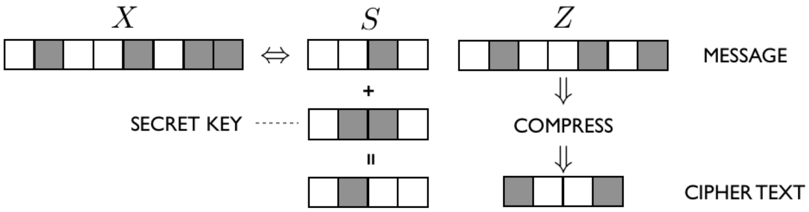

Fig. 3. Two-part coding strategy for secure compression. First, an information source X is represented as (S, Z) where S is the part of data that needs to be secret, and Z is independent of S. Secondly, S is encrypted using a shared secret key with one-time-pad encryption, while Z is compressed using any regular compression strategy. For variable-length lossless compression we require that X can be reconstructed from (S, Z) without error, while for fixed-length compression a small probability of error is allowed.

processing problems; this makes it a good target to test out the feasibility of perfect privacy. A more concrete operational motivation for maximizing mutual information is that it is equivalent to minimizing the logarithmic loss (log-loss) for a predictor of X based on Z, see [44]–[47]. From the perspective of prediction, (25) is equivalent to saying that the smallest log-loss achievable in secure prediction is $I(X; S)$. In other words, this is the smallest amount of information that needs to be hidden from a data analyst to keep S perfectly secure. We discuss this in more detail in Section IV as well.

C. Applications to Compression

In Section V we apply our decomposition bounds to study the problems of *variable-length* and *fixed-length* compression. The privacy system studied in Section IV assumes that the data analyst and the eavesdropper are the same entity. Because of this, it is generally impossible to completely disclose X in an S-secure way. Secure compression, on the other hand, happens over the Shannon cipher system, see Figure 1. In this setting the decompressor and the eavesdropper are two separate entities. By using a shared secret key between the two legitimate parties it becomes possible to losslessly communicate the value of X while still keeping the system S-secure. The utility cost of such lossless communication will be our primary focus and the main quantity of interest will be the compression rate.

We begin with variable-length compression. Traditional lossless compression is achieved via variable-length coding by representing the more likely realizations with shorter sequences, and the less likely realizations with longer sequences. When perfect secrecy is desired, variable-length compression is meaningless since the length of the compressed sequence would give away something about the sequence. As it turns out, when only partial secrecy is desired, variable-length compression is again possible with suitably designed codes.

We show in Theorem 9 that for any S-secure compressor the compression length of Z is bounded from below by

$$\max_{s \in \mathcal{S}} H(X|S = s). \tag{26}$$

Theorem 9 is a restatement of Lemma 3 with a small caveat of accounting for the shared secret between the compressor and the decompressor. This result makes no assumption on the distribution of the secret key or the amount of shared secret key available to the legitimate parties. It also makes no assumptions about the relationship between X and S or on whether S is available at the time of encoding.

The rest of Section V will emphasize the case when the secret key is equiprobable on $\{1, \ldots, K\}$ and when S is a deterministic function of X. We also take a perspective that the shared secret key is a limited resource and focus on the minimum required secret key needed for a given secure compression task. Not surprisingly, the smallest amount of secret key needed will be $K = |S|$ and in that case the fundamental limit of variable-length compression behaves like[4]

$$\log |S| + \max_{s \in S} H(X|S = s). \tag{27}$$

Observe that when S is a deterministic function of X, the fundamental limit of traditional compression can be written as

$$H(X) = H(S) + H(X|S). \tag{28}$$

The utility penalty for secrecy by design becomes readily apparent when we compare (27) with the right hand side of (28). This corresponds to the two stages of this encoding strategy in Figure 3. The partial secrecy constraint is directly responsible for the penalty observed in the first term, while in order to construct a Z that is independent from S we need local randomness at the compressor; and, this is the source of the penalty observed in the second term in (27).

We will also study almost-lossless fixed-length compression. In this setting the compressor and the decompressor are restricted to a fixed compression budget, but are allowed a small probability of error. Allowing a small probability of

[4] When we say "behaves like" we mean that the asymptotic compression limits are given by (27). In addition, the non-asymptotic codeword lengths are approximated by $\log |\mathcal{S}| + \max_{s \in \mathcal{S}} \imath_{X|S}(x|s)$ in the same way that the nonasymptotic codeword lengths for traditional compression are approximated by $\imath_X(x)$.

compression error dramatically improves the performance of secure compressors. That is, the second term in (27) behaves like $H(X|S)$, an average rather than worst case over $s \in S$. Such an improvement in performance may seem enticing, however it is worth highlighting that this holds only if no reconciliation of error will take place after the initial transmission of the message. A system that would apply these single-shot results, and then attempt to reconcile the compression error in an S-secure way is, in fact, a variable-length system and thus (27) would still the be the relevant bound.

Our bounds on variable-length secure compression demonstrate that there is a fundamental tension between compression and secrecy by design. This tension is not usually observed in information-theoretic security where secrecy is measured by equivocation. Most notably, in [36] Yamamoto studies a similar problem of compressing an information source X while securing a correlated secret S and finds that the rate of compression remains $H(X)$ under any positive leakage constraints on S.

III. Bounds on Secure Decomposition

In this section we investigate decompositions bounds that allow us to represent a random variable X as a deterministic function of (S, Z), where S is a given fixed random variable and Z is constructed to be independent of S. We establish that such decomposition always exists in Lemma 1. We then proceed to extend the analysis of Lemma 1 to the case when the alphabet of X is countably infinite, as well as provide a lower bound on the entropy of Z. We end the section by deriving more refined bounds on the entropy of Z using information spectrum techniques.

A. Functional Representation Lemma

We begin with the Functional Representation Lemma which can be found in [38] and was independently derived in [39]–[41]. While it has been previously used as an auxiliary result to analyze rate-regions in network information theory problems, it will be a basic building block in many of our results.

Lemma 1 (Functional Representation Lemma): Let X and S be two jointly distributed random variables supported on a finite or a countably infinite alphabet $\mathcal{X}$ and a finite alphabet $\mathcal{S}$, respectively. There exists a random variable Z on $\mathcal{Z}$ such that S and Z are independent, that is

$$I(S; Z) = 0, \tag{29}$$

X is a deterministic function of S and Z, that is

$$H(X|Z, S) = 0, \tag{30}$$

and the support of Z is bounded by

$$|\mathcal{Z}| \leq |\mathcal{S}|(|\mathcal{X}| - 1) + 1. \tag{31}$$

Proof: The main idea in the proof is to construct a Markov chain

$$(X, S) \leftrightarrow U \leftrightarrow Z \tag{32}$$

where U is a uniform random variable on $(0, 1)$ and is independent of S. Z will be defined via a suitable quantization of U; this will ensure that (29) holds. We describe the construction for finite, as well as for countably infinite alphabets.

First, to construct such a U assume without loss of generality that $\mathcal{X} = \{1, \dots, |\mathcal{X}|\}$ if $\mathcal{X}$ is finite and $\mathcal{X} = \{1, 2, \dots\}$ if $\mathcal{X}$ is countably infinite. Let $\mathcal{X}_s = \{x : P_{X|S}(x|s) > 0\}$. For each $s \in \mathcal{S}$ and $x \in \mathcal{X}_s$ define

$$k_{x,s} = F_{X|S}(x|s) \tag{33}$$

where

$$F_{X|Y}(x|s) = \mathbb{P}[X \leq x | S = s] \tag{34}$$

is the cumulative distribution function (CDF) of X given S and define $k_{0,s} = 0$ for all $s \in S$. For a given realization (X, S) define a random variable $\tilde{X}$ as

$$\tilde{X} = \max\{x \in \mathcal{X}_S \cup \{0\} : x < X\}. \tag{35}$$

Then

$$U = U[k_{\tilde{X},S}, k_{X,S}) \tag{36}$$

where $U[a, \tilde{a})$ denotes a uniform random variable on $[a, \tilde{a})$. It is straightforward to check that

$$F_{U|S}(u|s) = F_U(u) \ = \begin{cases} u, & u \in [0, 1) \\ 0, & \text{otherwise} \end{cases} \tag{37}$$

and therefore U and S are independent.

Next, the alphabet of Z is given by

$$\mathcal{Z} = \bigcup_{s \in \mathcal{S}} \bigcup_{x \in \mathcal{X}} \{k_{x,s}\} \tag{38}$$

and is a countable subset of the unit interval. When $\mathcal{X}$ is finite, (31) follows directly from (38). We define

$$Z = \min\{z \in \mathcal{Z} : U \leq z\}. \tag{39}$$

Since Z is obtained by the quantization of U that is independent of S, Z and S are also independent and (29) holds.

Observe that by this construction Z may take on the value $k_{X,S}$ or it may take on some other value $\tilde{z} \in (k_{\tilde{X},S}, k_{X,S})$ if such a $\tilde{z} \in Z$ exists. In other words, this construction guarantees that, for a fixed $s \in \mathcal{S}$, each $x \in \mathcal{X}_s$ is mapped to a disjoint segment of the unit interval. More explicitly, the function

$$\mathsf{g}(s, z) = \arg\min_{x \in \mathcal{X}_s} \{k_{x,s} \colon z \leq k_{x,s}\} \tag{40}$$

would recover x for a given (s, z) and thus (30) holds. □

Remark 1: When S is a deterministic function of X, and $|\mathcal{X}|$ is finite, (31) could be tightened to

$$|\mathcal{Z}| \leq |\mathcal{X}| - |\mathcal{S}| + 1. \tag{41}$$

This is because for each x there is a unique $s \in \mathcal{S}$ such that $x \in \mathcal{X}_s$. In other words, $\mathcal{X}_s$ partition $\mathcal{X}$ into $|\mathcal{S}|$ disjoint subsets.

Lemma 1 establishes an existence of a secure decomposition for a given (X, S). Recall the well-known trick of obtaining an arbitrary continuous random variable from a uniform random variable by applying the inverse CDF. The construction of the uniform random variable U based on the conditional distributions $P_{X|S}(\cdot|s)$ could be viewed as a reverse application

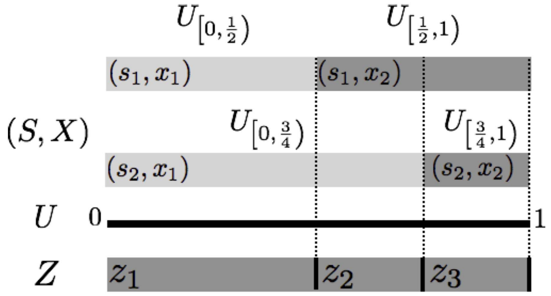

Fig. 4. An illustration of the S-secure construction from Lemma 1 for the joint distribution P_{XS} in Example 4.

of this trick. In the present case the distributions $P_{X|S}(\cdot|s)$ are discrete; therefore we randomize along the jumps in the CDF.

Example 4: Let X and S be finite random variables supported on $\mathcal{X} = \{x_1, x_2\}$ and $\mathcal{S} = \{s_1, s_2\}$ with the conditional distribution $P_{X|S}$ given by

$$P_{X|S}(x|s) = \begin{cases} \frac{1}{2}, & s = s_1 \\ \frac{3}{4}, & (s,x) = (s_2, x_1) \\ \frac{1}{4}, & (s,x) = (s_2, x_2) \end{cases} \tag{42}$$

According to Lemma 1, an S-secure Z supported on $\mathcal{Z} = \{z_1, z_2, z_3\}$ exists and is given by

$$P_{Z|XS}(z_1|x,s) = \begin{cases} 1, & (s,x) = (s_1, x_1) \\ \frac{2}{3}, & (s,x) = (s_2, x_1) \\ 0, & \text{otherwise} \end{cases} \tag{43}$$

$$P_{Z|XS}(z_2|x,s) = \begin{cases} \frac{1}{3}, & (s,x) = (s_2, x_1) \\ \frac{1}{2}, & (s,x) = (s_1, x_2) \\ 0, & \text{otherwise} \end{cases} \tag{44}$$

$$P_{Z|XS}(z_3|x,s) = \begin{cases} \frac{1}{2}, & (s,x) = (s_1, x_2) \\ 1, & (s,x) = (s_2, x_2) \\ 0, & \text{otherwise} \end{cases} \tag{45}$$

The construction of Z is illustrated in Figure 4.

B. Bounds on the Entropy of Z

The analysis in Lemma 1 could be extended to obtain a bound on $H(Z)$ which is particularly useful when $\mathcal{X}$ is countably infinite. This is done in the next lemma.

Lemma 2: Let X and S be two jointly distributed random variables supported on a finite or a countably infinite alphabet $\mathcal{X}$ and a finite alphabet $\mathcal{S}$, respectively. There exists a random variable Z on $\mathcal{Z}$ that satisfies (29), (30), and

$$H(Z) \le \sum_{s \in \mathcal{S}} H(X|S = s). \tag{46}$$

The next lemma gives a simple but conceptually important lower bound on the entropy of secure decompositions.

Lemma 3: Let (X, S) be jointly distributed random variables. If random the variable Z satisfies (29) and (30) then

$$H(Z) \ge \max_{s \in \mathcal{S}} H(X|S = s). \tag{47}$$

The proof of Lemma 2 is given in Appendix B. The proof of Lemma 3 is a special case of the proof of Lemma 6 and is given in Appendix D.

Example 5: Let X, S, and Z be as in Example 4. According to Lemma 3,

$$H(Z) \ge \max\left\{h\left(\frac{1}{2}\right), h\left(\frac{1}{4}\right)\right\} = h\left(\frac{1}{2}\right) = 1. \tag{48}$$

According to Lemmas 1 and 2,

$$H(Z) \le \min\left\{\log 3, h\left(\frac{1}{2}\right) + h\left(\frac{1}{4}\right)\right\} \le 1.585 \text{ bits}. \tag{49}$$

Then,

$$P_Z(s) = \begin{cases} \frac{1}{2} & z = z_1 \\ \frac{1}{4} & z \in \{z_2, z_3\} \end{cases} \tag{50}$$

and

$$H(Z) = 1.5 \text{ bits}. \tag{51}$$

C. Refinements via the Information Spectrum

Lemmas 2 and 3 give upper and lower bounds on the entropy of an S-secure Z in terms of the entropy of X conditioned on values of $s \in \mathcal{S}$. The next two theorems give an estimate on the entropy of Z using information spectrum techniques. That is, instead of directly bounding the entropy of Z, as in Lemmas 2 and 3, we derive bounds on the distribution of the information of Z. These bounds are more refined in a sense that they allow for better estimates of asymptotic fundamental limits, and are conceptually pleasing in a sense that they bound the information of Z in terms of the conditional information of X given S.

To state these theorems we define

$$\begin{aligned} \alpha(P_Y, L) = \mathbb{E}\left[1\left\{\imath_Y(Y) < \log L\right\} \exp\left(\imath_Y(Y) - \log L\right)\right] \\ + \mathbb{P}\left[\imath_Y(Y) > \log L\right] \end{aligned} \tag{52}$$

where P_Y is an arbitrary distribution and $L \ge 1$ is a real number.

Theorem 1: Let X and S be two jointly distributed random variables supported on a finite or a countably infinite alphabet $\mathcal{X}$ and a finite alphabet $\mathcal{S}$, respectively. Then, for any real $L \ge 1$ there exists a random variable Z satisfying (29), (30), and

$$\mathbb{P}\left[\imath_Z(Z) \; > \log L\right] \le \max_{s \in \mathcal{S}} \alpha\left(P_{X|S=s}, L\right) \tag{53}$$

$$\le \max_{s \in \mathcal{S}} \min_{\tau > 0} \left\{\mathbb{P}\left[\imath_{X|S}(X|S) > \log L - \tau | S = s\right] + \exp(-\tau)\right\}. \tag{54}$$

A detailed proof of Theorem 1 is given in the Appendix B. Informally, the proof of Theorem 1 is based on the idea of constructing a map from (X, S) to Z by breaking up the probabilities $P_{X|S}(x|s)$ into pieces of size $\frac{1}{L}$. The first term in (52) corresponds to the case when $P_{X|S}(x|s) > \frac{1}{L}$ and some amount of probability less than $\frac{1}{L}$ is left over while the second term in (52) corresponds to the case when this is not possible because $P_{X|S}(x|s) < \frac{1}{L}$.

Theorem 1 relates the distribution of $\imath_Z(Z)$ to the distribution of the worst-case (over $s \in \mathcal{S}$) distribution of $\imath_{X|S}(X|s)$.

Observe a nice correspondence between (54) and the lower bound given in the next theorem.

Theorem 2: Let X and S be jointly distributed random variables supported on a finite or a countably infinite alphabet $\mathcal{X}$ and a finite alphabet $\mathcal{S}$, respectively. If a random variable Z satisfies (29) and (30) then

$$\begin{aligned}\mathbb{P}\left[\imath_Z(Z) \right. & \left. > \log L\right] \\ & \geq \max_{s\in\mathcal{S},\tau>0} \left\{\mathbb{P}\left[\imath_{X|S}(X|S) > \log L + \tau | S = s\right] - \exp(-\tau)\right\}\end{aligned} \tag{55}$$

holds for any real $L \geq 1$.

Theorem 2 is proved in Appendix B. Theorems 1 and 2 could be used to show that, asymptotically, $H(Z)$ is $\max_{s\in\mathcal{S}} H(X|S = s)$.

The next theorem gives the distribution of Z in the case when exact lossless reconstruction of X is not required.

Theorem 3: Let (X, S) be jointly distributed random variables supported on a finite or a countably infinite alphabet $\mathcal{X}$ and a finite alphabet $\mathcal{S}$, respectively. Fix an integer $L \geq 1$. Then, there exists a random variable Z on $\mathcal{Z}$ satisfying (29) and

$$\mathbb{P}\left[\imath_{X|Z,S}(X|Z, S) > 0|S = s\right] \leq \alpha\left(P_{X|S=s}, L\right) \tag{56}$$

for all $s \in \mathcal{S}$. Moreover, Z is equiprobable on $\mathcal{Z} = \{1, \ldots, L\}$. That is, $H(Z) = \log L$.

Theorems 1 and 3 will be behind the achievability bounds in Section V. At the same time, we will derive corresponding lower bounds on the performance of secure compressors when S is a deterministic function of X.

IV. Applications to Privacy

In this section we explore the applications of the decomposition bounds from Section III to the problem of privacy in the setting of the privacy funnel function [43]. We focus on the perfect privacy functions $\mathsf{g}_0(P_{XS})$ and $\mathsf{h}_0(P_{XS})$ defined in (21) and (23), respectively. We begin by reviewing the necessary and sufficient conditions for $\mathsf{g}_0(P_{XS})$ and $\mathsf{h}_0(P_{XS})$ to be nonzero. We then show that $\mathsf{g}_0(P_{XS})$ and $\mathsf{h}_0(P_{XS})$ achieve the upper bound (25) whenever S is a deterministic function of X and characterize the necessary and sufficient conditions for the upper bound (25) to be achieved in general. We end the section by discussing the problem of perfect privacy from a perspective of secure prediction with log-loss.

A. On Feasibility of Perfect Privacy

The following result is shown in [48] (see also [31, Theorem 10]).

Theorem 4 ([31], [48]): Let X and S be two jointly distributed random variables. Then

$$\mathsf{g}_0(P_{XS}) > 0 \tag{57}$$

if and only if the rows of $[P_{S|X}(\cdot|x)]$ are linearly dependent.

According to Theorem 4, on the one hand, it is possible to achieve perfect privacy with non-trivial utility and unknown S whenever $|\mathcal{S}| < |\mathcal{X}|$ since in that case there will always be linearly dependent rows in the matrix $[P_{S|X}(\cdot|x)]$. On the other hand, if $|\mathcal{S}| \geq |\mathcal{X}|$ perfect privacy is only achievable if the joint distribution P_{XS} has a very particular structure.

Example 6 (Erasure Channel): Let $\mathcal{X} = \{0, 1\}$ and $\mathcal{S} = \{0, e, 1\}$. Suppose that (X, S) are jointly distributed according to

$$P_{XS}(x, s) = \begin{cases} \frac{1}{2} - \frac{\epsilon}{2}, & x = s \\ \frac{\epsilon}{2}, & s = e \\ 0, & \text{otherwise} \end{cases} \tag{58}$$

for some $0 \leq \epsilon \leq 1$. Applying the construction in proof of Lemma 1 we obtain the following lower bound

$$\mathsf{h}_0(P_{XS}) \geq 1 - h\left(\frac{1}{2}(1 - \epsilon)\right). \tag{59}$$

The random variable Z that achieves this lower bound is supported on $\{0, 1\}$ and is produced by the joint distribution

$$P_{Z|XS}(z|x, s) = \begin{cases} \frac{1}{2}, & x = s \\ 1, & x = z, s = e \\ 0, & \text{otherwise.} \end{cases} \tag{60}$$

However,

$$[P_{S|X}(\cdot|x)] = \begin{bmatrix} 1-\epsilon & \epsilon & 0 \\ 0 & \epsilon & 1-\epsilon \end{bmatrix} \tag{61}$$

and according to Theorem 4

$$\mathsf{g}_0(P_{XS}) = 0. \tag{62}$$

As can be seen from Example 6, requiring the S-secure Z to satisfy the Markov chain condition $S \leftrightarrow X \leftrightarrow Z$ could impose a significant cost in terms of utility. In many applications of interest S may actually be known to the designer of the privacy-assuring mapping, and it is beneficial to understand the behavior of $\mathsf{h}_0(P_{XS})$.

We begin with a corollary of Lemma 1 that characterizes when non-trivial utility is possible under perfect privacy with known S.

Theorem 5: Let X and S be two jointly distributed random variables supported on finite or countably infinite $\mathcal{X}$ and finite $\mathcal{S}$, respectively. Then

$$\mathsf{h}_0(P_{XS}) > 0 \tag{63}$$

if and only if

$$H(X|S) > 0. \tag{64}$$

That is, X is not a deterministic function of S.

The proof of Theorem 5 is given in Appendix C. The same result was also recently shown in [33] by explicitly constructing a Z on $\mathcal{Z} = \{z_1, z_2\}$ in a manner similar to the proof of Theorem 4 given in [31].

The functions $\mathsf{g}_0(P_{XS})$ and $\mathsf{h}_0(P_{XS})$ can be viewed as two extremes of a more general privacy problem: at one extreme the value of S is completely known during the construction of the privacy preserving mapping, while at the other extreme it is completely unknown. The feasibility of perfect privacy in an intermediate regime when some side information about S is known is also characterized in [33]. Moreover, [33] shows that $\mathsf{g}_0(P_{XS})$ and $\mathsf{h}_0(P_{XS})$ could be computed by solving a standard linear program. In the remainder of the section we look at a different aspect of this problem and characterize the conditions under which the spectrum of privacy regimes defined by $\mathsf{g}_0(P_{XS})$ and $\mathsf{h}_0(P_{XS})$ collapses into one.

B. New Lower and Upper Bounds on Perfect Privacy

In order to further explore the relationship between $\mathsf{h}_0(P_{XS})$ and $\mathsf{g}_0(P_{XS})$, we begin with the following basic, but important, observation. Let X, S, and Z be arbitrary random variables. Then, by the chain rule

$$\begin{aligned} I(X,S;Z) &= I(X;Z) + I(S;Z|X) && (65)\\ &= I(S;Z) + I(X;Z|S) && (66)\\ &= I(S;Z) + H(X|S) - H(X|S,Z). && (67)\end{aligned}$$

Therefore, for any S-secure Z it holds that

$$I(X;Z) = H(X|S) - H(X|S,Z) - I(S;Z|X). \qquad (68)$$

Since (68) holds for any S-secure Z, $H(X|S)$ is an upper bound on $\mathsf{h}_0(P_{XS})$ and on $\mathsf{g}_0(P_{XS})$. This is formalized for $\mathsf{h}_0(P_{XS})$ in Theorem 7 and was shown for $\mathsf{g}_0(P_{XS})$ in [31]. Equation (68) could also be used to lower bound $\mathsf{h}_0(P_{XS})$, as is shown in the following Theorem.

Theorem 6: Let X and S be two jointly distributed random variables supported on finite $\mathcal{X}$ and $\mathcal{S}$, respectively. Then

$$\mathsf{h}_0(P_{XS}) \geq H(X|S) - H(S|X) = H(X) - H(S). \qquad (69)$$

The lower bound (69) is tight if and only if S is a deterministic function of X.

Theorem 6 is proven in Appendix C. When S is a deterministic function of X

$$\mathsf{g}_0(P_{XS}) = \mathsf{h}_0(P_{XS}) = H(X|S) \qquad (70)$$

since then the Markov chain $S \leftrightarrow X \leftrightarrow Z$ holds for any S-secure Z.

Example 7 (Empirical Mean): Consider a special case of Example 2 where X^n is i.i.d. Bernoulli(p) and

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i. \qquad (71)$$

S is a deterministic function of X^n and applying Theorem 6 with $X \leftarrow X^n$ and $S \leftarrow S_n$,

$$\mathsf{g}_0(P_{X^nS_n}) = \mathsf{h}_0(P_{X^nS_n}) = \sum_{k=0}^{n} p^k(1-p)^{n-k}\log\binom{n}{k}. \qquad (72)$$

Using basic techniques from the method of types and Stirling's approximation [37] it can be shown that

$$\lim_{n\to\infty}\frac{\mathsf{g}_0(P_{X^nS_n})}{n} = \lim_{n\to\infty}\frac{\mathsf{h}_0(P_{X^nS_n})}{n} = h(p). \qquad (73)$$

In other words, the amount of mutual information per bit between X^n and an S_n-secure Z asymptotically approaches $H(X)$ and, asymptotically, there is no penalty for perfect privacy in this example.

Theorem 7: Let X and S be two jointly distributed random variables supported on finite $\mathcal{X}$ and $\mathcal{S}$, respectively. Then

$$\mathsf{g}_0(P_{XS}) \leq \mathsf{h}_0(P_{XS}) \leq H(X|S). \qquad (74)$$

Moreover, the following are equivalent:

1) $\mathsf{g}_0(P_{XS}) = H(X|S)$,
2) $\mathsf{h}_0(P_{XS}) = \mathsf{g}_0(P_{XS})$,
3) $\mathsf{h}_0(P_{XS}) = H(X|S)$.

The proof of Theorem 7 is given in Appendix C. It is based on (68) which shows that the gap between the performance of a given S-secure Z and the upper bound $H(X|S)$ is captured by two terms: $I(Z;S|X)$ and $H(X|S,Z)$. The term $I(Z;S|X)$ is equal to zero whenever the Markov chain $S \leftrightarrow X \leftrightarrow Z$ holds; that is, Z is a valid candidate for a maximizer of $\mathsf{g}_0(P_{XS})$. The condition $H(X|S,Z) = 0$, on the other hand, is exactly the condition guaranteed by Lemma 1 and says that X must be a deterministic function of S and Z. We show in Lemma 5 in Appendix C that a valid maximizer for $\mathsf{h}_0(P_{XS})$ must also satisfy $H(X|S,Z) = 0$.

The necessary and sufficient conditions for equalities in (74) are satisfied whenever S is a deterministic function of X. There are two other cases for which these conditions are trivially satisfied: 1) X and S are independent and 2) X is a deterministic function of S. The next example shows that these trivial cases are not the only ones for which the conditions in Theorem 7 hold.

Example 8 (Reverse Erasure Channel): Let $\mathcal{X} = \{0, e, 1\}$ and $\mathcal{S} = \{0, 1\}$. Suppose that (X, S) are jointly distributed according to

$$P_{XS}(x,s) = \begin{cases} \frac{1}{2} - \frac{\epsilon}{2}, & x = s\\ \frac{\epsilon}{2}, & x = e\\ 0, & \text{otherwise}\end{cases} \qquad (75)$$

for some $0 \leq \epsilon \leq 1$. Then

$$\mathsf{g}_0(P_{XS}) = \mathsf{h}_0(P_{XS}) = h(\epsilon). \qquad (76)$$

The random variable Z on $\mathcal{Z} = \{z_1, z_2\}$ that achieves (76) is produced by the joint distribution

$$P_{Z|X}(z|x) = \begin{cases} 1, & z = z_1 \text{ and } x \in \{0,1\}\\ 1, & z = z_2 \text{ and } x = e\\ 0, & \text{otherwise.}\end{cases} \qquad (77)$$

Example 8 is also derived in [32, Lemma 27] where it is shown that $\mathsf{g}_0(P_{XS}) = H(X|S)$.

In general, there does not appear to be a better way to characterize necessary and sufficient conditions for $\mathsf{h}_0(P_{XS})$ and $\mathsf{g}_0(P_{XS})$ to achieve (74) other than the ones given in Theorem 7. As the next example demonstrates, it is possible to leverage the insights of Theorem 7 to obtain such conditions given additional assumptions on S or X.

Example 9 (Binary Secret):

Define

$$\mathcal{S}_x = \{s \in \mathcal{S} : P_{X|S}(x|s) > 0\}. \qquad (78)$$

$\mathsf{g}_0(P_{XS}) = H(X|S)$ only if for all $x \in \mathcal{X}$,

$$P_{X|S}(x|s_1) = P_{X|S}(x|s_2) \text{ for all } s_1, s_2 \in \mathcal{S}_x. \qquad (79)$$

However, suppose $|\mathcal{S}| = 2$. Then $\mathrm{g}_0(P_{XS}) = H(X|S)$ if and only if (79) holds for all $x \in \mathcal{X}$.

C. Prediction With Log-Loss

Finally, we reformulate the privacy funnel problem from the perspective of secure prediction with logarithmic loss [44]–[47]. A widely used loss function in learning theory, the log-loss is a natural measure of loss in settings where the reconstructions are allowed to be soft, i.e. the predictor outputs a distribution over possible values of X rather than a hard guess. The log-loss between $x \in \mathcal{X}$ and its reconstruction $\hat{P} \in \mathcal{P}(\mathcal{X})$ is given by

$$\ell(x, \hat{P}) = \log \frac{1}{\hat{P}(x)} \tag{80}$$

where $\mathcal{P}(\mathcal{X})$ is the set of all probability mass functions on $\mathcal{X}$. Let

$$\hat{P} \colon \mathcal{Z} \to \mathcal{P}(\mathcal{X}) \tag{81}$$

denote such a soft predictor of X given an S-secure Z supported on $\mathcal{Z}$. That is, for a given value of $z \in \mathcal{Z}$, $\hat{P}(\cdot|z)$ is a distribution on $\mathcal{X}$.

When the joint distribution P_{XZ} is known, deriving the optimal average-case predictor is straightforward:

$$\begin{aligned}
\mathbb{E}\left[\ell(X, \hat{P})\right] &= \mathbb{E}\left[\log \frac{1}{\hat{P}(X|Z)}\right] && (82)\\
&= \mathbb{E}\left[\log \frac{P_{X|Z}(X|Z)}{\hat{P}(X|Z)}\right] + H(X|Z) && (83)\\
&= D(P_{X|Z} \| \hat{P} | P_Z) + H(X|Z) && (84)\\
&\geq H(X|Z). && (85)
\end{aligned}$$

Equation (84) is minimized by $\hat{P} = P_{X|Z}$ and the smallest attainable log-loss for secure prediction of X based on Z is $H(X|Z)$. This is equivalent to maximizing the mutual information between X and Z, as is done in in (21) and (23), since

$$I(X; Z) = H(X) - H(X|Z). \tag{86}$$

In light of this discussion, Theorem 7 shows that

$$\begin{aligned}
&\inf_{\substack{Z: I(S;Z)=0\\ S \leftrightarrow X \leftrightarrow Z}} \inf_{\hat{P} \colon \mathcal{Z} \to \mathcal{P}(\mathcal{X})} \mathbb{E}\left[\ell(X, \hat{P})\right] && (87)\\
&\geq \inf_{Z: I(S;Z)=0} \inf_{\hat{P} \colon \mathcal{Z} \to \mathcal{P}(\mathcal{X})} \mathbb{E}\left[\ell(X, \hat{P})\right] \geq I(X; S). && (88)
\end{aligned}$$

In other words, the log-loss suffered by an S-secure predictor is lower bounded by the mutual information between the information source X and the secret S. This lower bound is attainable, for example, when S is a deterministic function of X.

V. Applications to Compression

We apply decomposition bounds from Section III to the problem of lossless compression. We begin with variable length compression where we require completely lossless reconstruction of the information source. We then relax the complete reconstruction requirement and study almost lossless fixed length compression. We present corresponding converse bounds for the case when S is a deterministic function of X and the shared secret key size is $K = |\mathcal{S}|$ for both compression settings. The bounds presented in this section are used to derive asymptotic fundamental limits in Appendix A.

A. Variable-Length Compression

A prefix-free variable-length code with a secret key of size K is a pair of mappings:

$$\begin{aligned}
&\text{Encoder: } \mathsf{c} \colon \mathcal{X} \times \{1, \ldots, K\} \to \{0,1\}^*\\
&\text{Decoder: } \mathsf{d} \colon \{0,1\}^* \times \{1, \ldots, K\} \to \mathcal{X}
\end{aligned}$$

where no codeword in the image of c is a prefix of any another codeword. Let U be equiprobable on $\{1, \ldots, K\}$ and independent of X. The variable-length code (c, d) is S-secure if

$$I(\mathsf{c}(X, U); S) = 0. \tag{89}$$

When S is a deterministic function of X, with $S = \mathsf{f}_\mathsf{s}(X)$, we will also call such a code f_s-secure. An S-secure code (c, d) is lossless if

$$\mathbb{P}[\mathsf{d}(\mathsf{c}(X, U), U) = X] = 1. \tag{90}$$

We assume that, given a fixed key value k, $\mathsf{c}(x, k)$ could be random, that is, the compressor has access to local randomness, in addition to the shared randomness represented by U.

Let $\ell(s)$ denote the length of a string $s \in \{0,1\}^*$. We say that (c, d) is an (l, K)-variable-length code if

$$\mathbb{E}[\ell(\mathsf{c}(X, k))] \leq l, \quad \forall k \in \{1, \ldots, K\}. \tag{91}$$

We say that (c, d) is an (l, K, ϵ)-variable-length source code if

$$\mathbb{P}[\ell(\mathsf{c}(X, k)) > l] \leq \epsilon, \quad \forall k \in \{1, \ldots, K\}. \tag{92}$$

The non-asymptotic fundamental limits of S-secure lossless compression are given by

$$\ell^*_{XS}(K) = \inf\{l : \exists\ S\text{-secure } (l, K)\text{-code}\} \tag{93}$$

and

$$\epsilon^*_{XS}(l, K) = \inf\{\epsilon : \exists\ S\text{-secure } (l, K, \epsilon)\text{-code}\}. \tag{94}$$

We begin with direct lower and upper bound on the average length of S-secure variable-length codes.

Theorem 8: Let X and S be two jointly distributed random variables supported on a finite or a countably infinite alphabet $\mathcal{X}$ and a finite alphabet $\mathcal{S}$, respectively. Then

$$\ell^*_{XS}(|\mathcal{S}|) \leq \sum_{s \in \mathcal{S}} H(X|S = s) + 1 + \lceil \log |\mathcal{S}| \rceil. \tag{95}$$

If X is finite then

$$\ell^*_{XS}(|\mathcal{S}|) \leq \lceil \log(|\mathcal{S}|)(|\mathcal{X}| - 1) + 1) \rceil + \lceil \log |\mathcal{S}| \rceil, \tag{96}$$

and if S is also a deterministic function of X then

$$\ell^*_{XS}(|\mathcal{S}|) \leq \lceil \log(|\mathcal{X}| - |\mathcal{S}| + 1) \rceil + \lceil \log |\mathcal{S}| \rceil. \tag{97}$$

Proof: The theorem follows from the two-part coding construction outlined in Section II and Lemmas 1 and 2. □

The next theorem gives a direct lower bound on the average length of S-secure variable-length codes.

Theorem 9: Let X and S be two jointly distributed random variables supported on a finite or a countably infinite alphabet $\mathcal{X}$ and a finite alphabet $\mathcal{S}$, respectively. Then

$$\ell^*_{XS}(K) \geq \max_{s\in\mathcal{S}} H(X|S=s) \tag{98}$$

holds for any $K \in \{1,2,\dots\}$.

If S is a deterministic function of X then for any $K \geq |\mathcal{S}|$,

$$\ell^*_{XS}(K) \geq \log|\mathcal{S}| \tag{99}$$

and $\mathsf{f_s}$-secure codes do not exist for $K < |\mathcal{S}|$.

Lower bound (98) is a corollary of Lemma 3 with a caveat of accounting for the shared secret key and (99) is a corollary of [1]. Theorem 9 is proved in Appendix D. We highlight that, although our primary focus is on the case $K = |\mathcal{S}|$, (98) holds for an arbitrary value of K and its proof does not use the fact that U is equiprobable on $\{1,\dots,K\}$. In general, Theorems 8 and 9 are sometimes asymptotically tight as is shown in the next example.

Example 10 (Empirical Mean): Let X^n and S be as in Example 7. Applying Theorems 8 and 9 with $X \leftarrow X^n$ and $S \leftarrow S$,

$$\begin{aligned}\max\left\{\max_{k\in\{0,n\}}\log\binom{n}{k}, \log(n+1)\right\} &\leq \ell^*_{XS}(n+1) \qquad (100)\\ &\leq \log(2^n-n)+\log(n+1). \qquad (101)\end{aligned}$$

It is easy to check that

$$\lim_{n\to\infty}\frac{\ell^*_{XS}(n+1)}{n} = 1. \tag{102}$$

An asymptotic compression of this source is impossible even under a partial secrecy constraint. However, partial secrecy is achieved with a reduced secret key size of $K = n+1$ as opposed to $K = 2^n$ that would be required for full secrecy.

Theorems 8 and 9 are not tight in general. However, as is the case in traditional compression [42], it is possible to get better asymptotic bounds on (93) by leveraging bounds on the probability of excess length fundamental limit (94). Such upper and lower bounds on (94) are developed next.

Theorem 10: Let X and S be two jointly distributed random variables supported on a finite or a countably infinite alphabet $\mathcal{X}$ and a finite alphabet $\mathcal{S}$, respectively. For any $L \in \{1,2,\dots\}$

$$\epsilon^*_{XS}(l,|\mathcal{S}|) \leq \max\left\{\max_{s\in S}\{\alpha(P_{X|S=s},L)\}, 1-\frac{2^\eta-1}{L}\right\} \tag{103}$$

where $\eta = l - \lceil\log|\mathcal{S}|\rceil$. Moreover,

$$\begin{aligned}\epsilon^*_{XS}(l,|\mathcal{S}|) &\leq \max_{s\in\mathcal{S}}\{\alpha(P_{X|S=s},2^\eta)\} \qquad (104)\\ &\leq \max_{s\in S}\min_{\tau>0}\left\{\mathbb{P}[\imath_{X|S}(X|S)\geq\eta-\tau|S=s]+2^{-\tau}\right\} \qquad (105)\end{aligned}$$

where $\eta = l - \lceil\log|\mathcal{S}|\rceil$.

Note the correspondence between (105) and the lower bound in the next Theorem.

Theorem 11: Let X and S be two jointly distributed random variables supported on finite or countably infinite alphabet $\mathcal{X}$ and finite alphabet $\mathcal{S}$, respectively. Let S be a deterministic function of X. Then

$$\max_{\tau>0, s\in\mathcal{S}}\left\{\mathbb{P}[\imath_{X|S}(X|S)\geq\eta+\tau|S=s]-2^{-\tau}\right\} \leq \epsilon^*_{XS}(l,|\mathcal{S}|) \tag{106}$$

where $\eta = l - \log|\mathcal{S}|$.

Theorems 10 and 11 are proved in Appendix D. Taking $|\mathcal{S}| = 1$ recovers traditional compression. Particularizing Theorem 11 to this case we obtain

$$\max_{\tau>0}\left\{\mathbb{P}[\imath_X(X)\geq l+\tau]-2^{-\tau}\right\} \leq \epsilon^*_{XS}(l,1) \tag{107}$$

which is exactly [42, Theorem 4].

We show in Appendix A that the asymptotic fundamental limit of lossless variable-length compression is

$$\log|\mathcal{S}| + \max_{s\in\mathcal{S}} H(X|S=s) \tag{108}$$

whenever $K = |\mathcal{S}|$ and S is a deterministic function of X.

B. Fixed-Length Compression

Next, we focus on the special case when S is a deterministic function of X and relax the requirement that compression must be completely lossless. A fixed-length source code with M codewords and a secret key of size K is a pair of mappings:

$$\begin{aligned}&\text{Encoder: } \mathsf{c}\colon \mathcal{X}\times\{1,\dots,K\}\to\{1,\dots,M\}\\ &\text{Decoder: } \mathsf{d}\colon \{1,\dots,M\}\times\{1,\dots,K\}\to\mathcal{X}.\end{aligned}$$

Suppose $S = \mathsf{f_s}(X)$. An $\mathsf{f_s}$-secure code (c,d) (see (88)) is an (M,K,ϵ)-almost-lossless code if

$$\begin{aligned}\mathbb{P}[\mathsf{d}(\mathsf{c}(X,k),k)\in\mathsf{f_s}^{-1}(S)] = 1 \text{ and} \qquad (109)\\ \mathbb{P}[\mathsf{d}(\mathsf{c}(X,k),k)\ \neq X]\leq\epsilon,\quad \forall k\in\{1,\dots,K\}. \qquad (110)\end{aligned}$$

In other words, the value of a secrecy function must be reconstructed exactly, but otherwise a small probability of error is allowed. We assume that $\mathsf{c}(x,k)$ could be random, that is, the compressor has access to local randomness, in addition to shared randomness represented by U.

The non-asymptotic fundamental limits of $\mathsf{f_s}$-secure almost-lossless compression are given by

$$\tilde{\epsilon}_{XS}(M,K) = \inf\{\epsilon\colon \exists\ \mathsf{f_s}\text{-secure } (M,K,\epsilon)\text{-code}\} \tag{111}$$

and

$$\tilde{M}_{XS}(\epsilon,K) = \inf\{M\colon \exists\ \mathsf{f_s}\text{-secure } (M,K,\epsilon)\text{-code}\}. \tag{112}$$

In this section we focus on bounding $\tilde{\epsilon}_{XS}(M,K)$, but the presented results could be used to bound $\tilde{M}_{XS}(\epsilon,K)$ by noting that

$$\tilde{M}_{XS}(\epsilon,K) = \inf\{M\colon \tilde{\epsilon}_{XS}(M,K)\leq\epsilon\}. \tag{113}$$

Theorem 12: Let X and S be two jointly distributed random variables supported on a finite or a countably infinite alphabet $\mathcal{X}$ and a finite alphabet $\mathcal{S}$, respectively. Then

$$\tilde{\epsilon}_{XS}(M, |\mathcal{S}|) \leq \min_{L \in \{1,2,\dots\}} \mathbb{E}\left[\max\left\{\alpha(P_{X|S=\bar{S}}, L), 1 - \frac{M'}{L}\right\}\right] \tag{114}$$

$$\leq \mathbb{E}\left[\alpha(P_{X|S=\bar{S}}, M')\right] \tag{115}$$

$$\leq \mathbb{E}\left[\min_{\tau>0}\left\{\mathbb{P}[\imath_{X|S}(X|S) \geq \log M' - \tau|S] + 2^{-\tau}\right\}\right] \tag{116}$$

$$\leq \min_{\tau>0}\left\{\mathbb{P}[\imath_{X|S}(X|S) \geq \log M' - \tau] + 2^{-\tau}\right\} \tag{117}$$

where $M' = \left\lfloor \frac{M}{|\mathcal{S}|} \right\rfloor$ and the expectations are take with respect to $\bar{S} \sim P_S$.

Note the correspondence between (116) and the lower bound in the next Theorem.

Theorem 13: Let X and S be two jointly distributed random variables supported on a finite or a countably infinite alphabet $\mathcal{X}$ and a finite alphabet $\mathcal{S}$, respectively. Let S be a deterministic function of X. Then

$$\max_{\tau>0}\left\{\mathbb{P}[\imath_{X|S}(X|S) \geq \log M' + \tau] - 2^{-\tau}\right\} \leq \tilde{\epsilon}_{XS}(M, |\mathcal{S}|) \tag{118}$$

where $M' = \frac{M}{|\mathcal{S}|}$.

It is observed in [42] that for traditional compression the probability of error in almost-lossless fixed-length coding and the probability of excess length in variable-length coding are related even in the single-shot setting. Taking $|\mathcal{S}| = 1$ recovers traditional fixed-length compression. Particularizing Theorem 13 to this case yields

$$\max_{\tau>0}\left\{\mathbb{P}[\imath_{X}(X) \geq \log M + \tau] - 2^{-\tau}\right\} \leq \tilde{\epsilon}_{XS}(M, 1) \tag{119}$$

which is exactly the application of [42, Theorem 4] to fixed-length compression.

This equivalence between traditional almost-lossless fixed-length and variable-length compression provides a general justification for the focus on fixed-length compression in the theoretical compression literature. Indeed, given an error event in the fixed-length setting, the remaining uncertainty about the information source could be reconciled by adding more bits to the codeword. Thus, any almost-lossless fixed-length code could be turned into lossless variable-length code. However, under secrecy by design this relationship breaks down. Compare Theorems 10 and 11 with Theorems 12 and 13. In the variable-length setting the bounds depend on the worst case distribution of the conditional information, while in the almost lossless fixed-length setting the same bounds are expressed in terms of the average distribution of the conditional information. To underscore this point we show in Appendix A that the asymptotic fundamental limit of almost-lossless fixed-length compression is

$$\log |\mathcal{S}| + H(X|S) \tag{120}$$

whenever $K = |\mathcal{S}|$. The argument of simply reconciling the compression error does not hold under secrecy by design, even in the asymptotic setting. This is because reconciling the compression error in the naive way leaks information about the secret S. On the other hand, reconciling the compression error in an S-secure way reduces to the variable-length setting.

C. Secrecy by Design Versus Leakage

In [36, Case 6] almost-lossless compression of X with a random S is studied under leakage (equivocation) constraints; the same problem with deterministic S is studied in [36, Case 7]. It is shown that for both cases the asymptotic fundamental limit of compression is $H(X)$ and the amount of the required shared secret key decreases linearly with the leakage constraint. Note that in the leakage setting of [36] it is possible to leverage fixed-length compression to achieve variable-length compression with an asymptotically negligible increase in leakage. Comparing [36, Case 7] to (108) demonstrates that there is a distinct compression penalty under secrecy by design when S is a deterministic function of X. Even when S is not a deterministic function of X, the lower bound (98) on the compression rate will be, in general, above $H(X)$.

A basic building block for S-secure compressors is Lemma 1 and the two part coding strategy, see Figure 3. The approach in [36] is similar in that a new notion of common information is introduced that allows to decompose X into S and an almost independent Z. In short, [36] allows some error in the reconstruction of X, as well as some dependence between S and Z. As the discussion above demonstrates, either of these relaxations is sufficient to eliminate the asymptotic compression penalty for secure compression. The stark contrast between (108) and [36] shows that, in general, a small leakage regime is not a good approximation for the zero leakage regime imposed by secrecy by design.

VI. Discussion

A. Contributions

We have introduced an approach to partial secrecy which we call secrecy by design. The fundamental idea behind secrecy by design is to start with an explicit partial secrecy requirement and to construct the information processing system from the ground up to satisfy this requirement. This is in sharp contrast to commonly used approaches such as equivocation and differential privacy where a measure of information leakage is proposed and used as a design guide. We have developed basic tools that allow us to apply the secrecy by design principle; this includes decomposition strategies that let us represent an information source X as a function of the secret S and an S-secure publicly shareable Z.

We have applied the secrecy by design framework to the problems of privacy and lossless compression. For the problem of privacy we studied the privacy funnel setting where the goal is to maximize mutual information between the private and public data. We highlighted that there are two extreme regimes: one where the designer of the privacy assuring mapping knows the secret S, and one where the mapping is designed without this knowledge. We have developed new lower and upper bounds on these regimes and characterized when the two regimes collapse to yield the same utility.

Finally, we connected the current privacy problem to prediction with logarithmic loss.

We also applied secrecy by design to compression over the Shannon cipher system. Although perfect secrecy requires a shared secret key that is as large as the message set, we have shown that it is possible to have strong partial secrecy guarantees with a reduced shared secret key size. However, the resulting secure compressors do incur a compression penalty. We have characterized this penalty and shown that it depends strongly on how the secrecy constraint interacts with the statistics of the information source.

B. Future Work

The secrecy by design framework could be applied to other information processing problems such as lossy compression, channel coding, and multi-terminal coding. There are also a number of broader questions that could be investigated in regards to secrecy by design; for example, one such question is the robustness of the secrecy guarantees to the small errors in the statistical model. The results presented in this paper assume that the statistical model for the information source is known at the time of code design. A natural question to ask then is how to construct secure compressors that are universal with respect to a family of statistical models. As it turns out, the secure variable-length codes studied in Section V are already universal in that they only depend on the conditional distribution $P_{X|S}$ and not on P_S. That is, consider the Empirical Mean in Example 10. In this case the fundamental limits (as well as the code used to achieve them) are independent of the parameter p. Although secure lossless compressors incur a penalty for partial secrecy, this penalty does make them universal with respect to a family of statistical models.

Appendix

A. Asymptotic Fundamental Limits for Compression

Let $X \sim P_X$ and $S = \mathsf{f_s}(X)$ where

$$\mathsf{f_s} \colon \mathcal{X} \to \mathcal{S}. \tag{121}$$

Given a single letter secrecy function (121), define an n-letter secrecy function

$$\mathsf{f_s^n} \colon \mathcal{X}^n \to \mathcal{S}^n \tag{122}$$

via

$$\mathsf{f_s^n}(x^n) = (\mathsf{f_s}(x_1), \ldots, \mathsf{f_s}(x_n)). \tag{123}$$

We assume that X^n is i.i.d. according to P_X, and therefore S^n is also i.i.d. . The asymptotic variable-length fundamental limit is given by

$$\mathcal{R}^*_{XS} = \lim_{n\to\infty} \frac{\ell^*_{X^nS^n}(|\mathcal{S}^n|)}{n}. \tag{124}$$

The asymptotic fixed-length fundamental limit is given by

$$\tilde{\mathcal{R}}_{XS} = \lim_{\epsilon\to 0}\lim_{n\to\infty} \frac{1}{n}\log \tilde{M}_{X^nS^n}(\epsilon, |\mathcal{S}^n|). \tag{125}$$

The next theorem characterizes the two fundamental limits for the product setting.

Theorem 14:

$$\mathcal{R}^*_{XS} = \log|\mathcal{S}| + \max_{s\in\mathcal{S}} H(X|S=s) \tag{126}$$

and

$$\tilde{\mathcal{R}}_{XS} = \log|\mathcal{S}| + H(X|S). \tag{127}$$

Equation (126) is a direct consequence of Theorems 10 and 11, while (127) is a a direct consequence of Theorems 12 and 13. A detailed proof of Theorem 14 is given in Appendix E.

B. Proofs for Section III

Lemma 2: The construction in the proof of Lemma 1 works for finite and countably infinite $\mathcal{X}$. The Lemma 1 construction guarantees that each $x \in \mathcal{X}_s$ is mapped to a disjoint segment of the unit interval and thus X is a deterministic function of (S, Z) given by (40). For each $s \in \mathcal{S}$ define

$$W_s = \mathsf{g}(s, Z) \tag{128}$$

and observe that W_s is distributed according to $P_{X|S=s}$. Moreover, by construction,

$$Z = \min_{s\in\mathcal{S}} k_{W_s,s}. \tag{129}$$

That is, Z is a deterministic function of the set $\{W_s\}_{s\in\mathcal{S}}$. Thus

$$\begin{aligned} H(Z) &\le H\left(\{W_s\}_{s\in\mathcal{S}}\right) & (130)\\ &\le \sum_{s\in\mathcal{S}} H(W_s) & (131)\\ &= \sum_{s\in\mathcal{S}} H(X|S=s) & (132)\end{aligned}$$

and this shows (46). □

Theorem 1: Assume without loss of generality that $\mathcal{X} = \{1, \ldots, |\mathcal{X}|\}$ if $\mathcal{X}$ is finite and $\mathcal{X} = \{1, 2, \ldots\}$, if $\mathcal{X}$ is countably infinite. Fix a real $L \ge 1$, and let $\mathcal{Z} = \{1, \ldots, \lceil L\rceil + |\mathcal{X}|\}$. Define

$$n_{x,s} = \lfloor P_{X|S}(x|s)L \rfloor \tag{133}$$

and note that

$$N_s = \sum_{x\in\mathcal{X}} n_{x,s} \le L, \quad \forall s \in \mathcal{S}. \tag{134}$$

For a fixed $s \in \mathcal{S}$ partition $\{1, \ldots, N_s\}$ into subsets $\mathcal{M}_x$ such that

$$|\mathcal{M}_x| = n_{x,s}, \quad \forall x \in \mathcal{X} \tag{135}$$

and construct an auxiliary Z' on $\mathcal{Z}$:

$$P_{Z'|XS}(z|x,s) = \begin{cases} \frac{1}{LP_{X|S}(x,s)}, & z \in \mathcal{M}_x \\ \delta_{x,s}, & z = \lceil L \rceil + x \\ 0, & \text{otherwise} \end{cases} \tag{136}$$

where

$$\delta_{x,s} = 1 - \frac{n_{x,s}}{LP_{X|S}(x,s)} \le \frac{1}{LP_{X|S}(x,s)}. \tag{137}$$

This transformation satisfies

$$P_{Z'|S}(z|s) = \frac{1}{L}, \quad z \le N = \min_{s \in \mathcal{S}} N_s. \tag{138}$$

Moreover, X is losslessly recoverable from Z', S.

Finally, Z is obtained by applying transformation from Lemma 1 to $X \leftarrow (Z', S)$ and $S \leftarrow S$. By construction it still holds that

$$P_Z(z) = P_{Z|S}(z|s) = \frac{1}{L}, \quad z \le N = \min_{s \in \mathcal{S}} N_s, \tag{139}$$

and (29) and (30) hold by Lemma 1.

To complete the proof we need to get a bound on $\mathbb{P}[Z \ge N]$:

$$\mathbb{P}[Z \ge N] = \mathbb{P}[Z \ge N | S = s'] = \max_{s \in \mathcal{S}} \mathbb{P}[Z \ge N_S | S = s]. \tag{140}$$

Then

$$\begin{aligned} \mathbb{P}[Z \ge N_S | S = s] &= \mathbb{P}[n_{X,S} = 0 | S = s] \mathbb{P}[Z \ge N_S, n_{X,S} = 0 | S = s] \\ &\quad + \mathbb{P}[n_{X,S} \ge 1 | S = s] \mathbb{P}[Z \ge N_S, n_{X,S} \ge 1 | S = s] && (141) \\ &\le \mathbb{P}[n_{X,S} = 0 | S = s] + \mathbb{P}[Z \ge N_S, n_{X,S} \ge 1 | S = s] && (142) \\ &\le \mathbb{P}[\imath_{X|S}(X|S) > \log L | S = s] \\ &\quad + \mathbb{E}[\exp(\imath_{X|S}(X|S) - \log L) 1\{\imath_{X|S}(X|S) < \log L\} | S = s]. && (143) \end{aligned}$$

The bound on the second term in (142) follow since

$$\begin{aligned} &\mathbb{P}[Z \ge N_S, n_{X,S} \ge 1 | S = s] \\ &= \sum_{(z,x) \in \mathcal{Z} \times \mathcal{X}} P_{Z|XS}(z|x,s) P_{X|S}(x|s) 1\{z \ge N_s\} 1\{n_{x,s} \ge 1\} && (144) \\ &\le \sum_{x \in \mathcal{X}} \delta_{x,s} P_{X|S}(x|s) 1\{n_{x,s} \ge 1\} && (145) \\ &= \sum_{x \in \mathcal{X}} P_{X|S}(x|s) \exp(\imath_{X|S}(x|s) - \log L) 1\{x \in \mathcal{F}_s\} && (146) \\ &= \mathbb{E}\left[\exp(\imath_{X|S}(X|S) - \log L) 1\{X \in \mathcal{F}_S\} | S = s\right] && (147) \end{aligned}$$

where $\mathcal{F}_s = \{x \colon \imath_{X|S}(x|s) < \log L\}$. This shows (53).

Finally, (54) is obtained by loosening (53). That is, fix any $\tau > 0$, then

$$\begin{aligned} &\mathbb{P}[\imath_{X|S}(X|S) > \log L | S = s] \\ &+ \mathbb{E}\left[\exp(\imath_{X|S}(X|S) - \log L) 1\{X \in \mathcal{F}_S\} | S = s\right] && (148) \\ &\le \mathbb{P}[\imath_{X|S}(X|S) > \log L | S = s] \\ &+ \mathbb{E}\left[\exp(\imath_{X|S}(X|S) - \log L) 1\{X \in \mathcal{F}_{S,\tau}\} | S = s\right] \\ &+ \mathbb{E}\left[1\{\log L - \tau \le \imath_{X|S}(X|S) \le \log L\} | S = s\right] && (149) \\ &\le \mathbb{P}[\imath_{X|S}(X|S) > \log L - \tau | S = s] \\ &+ \mathbb{E}\left[\exp(\imath_{X|S}(X|S) - \log L) 1\{X \in \mathcal{F}_{S,\tau}\} | S = s\right] && (150) \\ &\le \mathbb{P}[\imath_{X|S}(X|S) > \log L - \tau | S = s] + \exp(-\tau) && (151) \end{aligned}$$

where $\mathcal{F}_{s,\tau} = \{x \colon \imath_{X|S}(x|s) < \log L - \tau\}$. □

Theorem 2: Fix an arbitrary $s \in \mathcal{S}$, $\tau > 0$ and define

$$\mathcal{L} = \{x \in \mathcal{X} \colon \log L + \tau \le \imath_{X|S}(x,s) < \infty\} \tag{152}$$

$$\mathcal{C} = \{z \in \mathcal{Z} \colon \imath_Z(z) \le \log L\}. \tag{153}$$

Then

$$\begin{aligned} &\mathbb{P}[\imath_{X|S}(X|S) > \log L + \tau | S = s] = \mathbb{P}[X \in \mathcal{L} | S = s] && (154) \\ &= \mathbb{P}[(X,Z) \in \mathcal{L} \times \mathcal{C} | S = s] + \mathbb{P}[(X,Z) \in \mathcal{L} \times \mathcal{C}^c | S = s] && (155) \\ &\le \mathbb{P}[(X,Z) \in \mathcal{L} \times \mathcal{C} | S = s] + \mathbb{P}[Z \in \mathcal{C}^c | S = s] && (156) \\ &\le 2^{\log L} 2^{-\log L - \tau} + \mathbb{P}[Z \in \mathcal{C}^c | S = s] && (157) \\ &= 2^{-\tau} + \mathbb{P}[Z \in \mathcal{C}^c] && (158) \\ &= 2^{-\tau} + \mathbb{P}[Z > \log L] && (159) \end{aligned}$$

where (158) follows because Z and S are independent. Equation (157) follows since

$$|\mathcal{C}| \le L = 2^{\log L} \tag{160}$$

and for any fixed s and z there always exists a unique $x \in \mathcal{L}$ such that

$$P_{XZ|S}(x,z|s) > 0. \tag{161}$$

Moreover, for any such $x \in \mathcal{L}$

$$P_{XZ|S}(x,z|s) = P_{X|S}(x|s) P_{Z|XS}(z|x,s) \le 2^{-\log L - \tau}. \tag{162}$$

□

Theorem 3: We will ensure that (29) holds by constructing Z in such a way that

$$\mathbb{P}[Z = z | S = s] = \mathbb{P}[Z = z] = \frac{1}{L} \tag{163}$$

for all $s \in \mathcal{S}$ and all $z \in \mathcal{Z}$.

Let

$$n_{x,s} = \lfloor P_{X|S}(x|s) L \rfloor \tag{164}$$

and note that

$$N_s = \sum_{x \in \mathcal{X}} n_{x,s} \le L, \quad \forall s \in \mathcal{S}. \tag{165}$$

For a fixed $s \in \mathcal{S}$ partition $\{1, \ldots, N_s\}$ into subsets $\mathcal{M}_x$ such that

$$|\mathcal{M}_x| = n_{x,s}, \quad \forall x \in \mathcal{X}. \tag{166}$$

Then

$$P_{Z|XS}(z|x,s) = \begin{cases} \frac{1}{LP_{X|S}(x,s)}, & z \in \mathcal{M}_x \\ \frac{\delta_{x,s}}{L - N_s}, & z =\in \{N_s + 1, \ldots, L\} \\ 0, & \text{otherwise} \end{cases} \tag{167}$$

where

$$\delta_{x,s} = 1 - \frac{n_{x,s}}{LP_{X|S}(x,s)} \le \frac{1}{LP_{X|S}(x,s)}. \tag{168}$$

Note that the construction satisfies (163). Indeed, fix $s \in \mathcal{S}$ suppose $x \in \mathcal{M}_x$ for some $x \in \mathcal{X}$. Then

$$\begin{aligned} P_{Z|S}(z|s) &= P_{X|S}(x|s) P_{Z|XS}(z|x,s) && (169) \\ &= P_{X|S}(x|s) \frac{1}{LP_{X|S}(x|s)} = \frac{1}{L}. && (170) \end{aligned}$$

Suppose $z \in \{N_s+1, \ldots, L\}$. Then

$$\begin{aligned} P_{Z|S}(z|s) &= \sum_{x \in \mathcal{X}} P_{X|S}(x|s) \frac{\delta_{x,s}}{L - N_s} && (171) \\ &= \frac{1}{L-N_s} \sum_{x \in \mathcal{X}} \left(P_{X|S}(x|s) - \frac{n_{x,s}}{L} \right) && (172) \\ &= \frac{1}{L-N_s} \left(1 - \frac{N_s}{L} \right) = \frac{1}{L}. && (173) \end{aligned}$$

Finally,

$$\begin{aligned} \mathbb{P}\left[\imath_{X|ZS}(X|Z,S) > 0 | S = s \right] &\leq \mathbb{P}\left[Z \geq N_S | S = s \right] && (174) \\ &\leq \alpha(P_{X|S=s}, L). && (175) \end{aligned}$$

This follows by exactly the same argument as in proof of Theorem 1 since N_s are defined identically. □

C. Proofs for Section IV

Theorem 5: Suppose X is a deterministic function of S and that Z is S-secure. Then

$$I(X;Z) \leq I(X,S;Z) = I(S;Z) = 0. \quad (176)$$

Suppose X is not a deterministic function of S. Then, there must exist $\tilde{x} \in \mathcal{X}$ and $\tilde{s} \in \mathcal{S}$ such that $0 < P_{X|S}(\tilde{x}|\tilde{s}) < 1$. Let us assume without loss of generality that $\tilde{x} = 1$, and let $\tilde{z} = \min\{z \in \mathcal{Z}\}$. The particular construction used in the proof of Lemma 1 has the property that $P_{X|ZS}(\tilde{x}|\tilde{z}, s) = 1$ whenever $P_{X|S}(\tilde{x}|s) > 0$. Thus, $P_{X|ZS}(\tilde{x}|\tilde{z}, \tilde{s}) > P_{X|S}(\tilde{x}|\tilde{s}) > 0$. Moreover, $P_{X|ZS}(\tilde{x}|\tilde{z}, s) = 0$ whenever $P_{X|S}(\tilde{x}|s) = 0$. Putting this all together we obtain

$$\begin{aligned} P_{X|Z}(\tilde{x}|\tilde{z}) &= \sum_{s \in \mathcal{S}} P_{X|ZS}(\tilde{x}|\tilde{z}, s) P_S(s) && (177) \\ &> \sum_{s \in \mathcal{S}} P_{X|S}(\tilde{x}|s) P_S(s) = P_X(\tilde{x}) && (178) \end{aligned}$$

and Z is not independent from X. □

Theorem 6: Let Z be the random variable guaranteed by Lemma 1. Then

$$\begin{aligned} I(X;Z) &= H(X|S) - H(X|S,Z) - I(Z;S|X) && (179) \\ &= H(X|S) - I(Z;S|X) && (180) \\ &= H(X|S) - H(S|X) + H(S|X,Z) && (181) \\ &\geq H(X|S) - H(S|X) && (182) \end{aligned}$$

where (180) follows since Lemma 1 guarantees that X is a deterministic function of (S,Z). This proves the lower bound in (69).

To prove the if and only if part observe that (68) holds for any S-secure Z and so

$$\begin{aligned} I(X;Z) &\leq H(X|S) - I(Z;S|X) && (183) \\ &= H(X|S) - H(S|X) + H(S|X,Z) && (184) \\ &= H(X|S) - H(S|X) && (185) \end{aligned}$$

whenever S is a deterministic function of X.

Finally, suppose S is not a deterministic function of X. Then, there exists $\tilde{x} \in \mathcal{X}$ and $\tilde{s}_1, \tilde{s}_2 \in \mathcal{S}$ such that $P(\tilde{x}|\tilde{s}_1) > 0$, $P(\tilde{x}|\tilde{s}_2) > 0$, and $\tilde{s}_1 \neq \tilde{s}_2$. Assume without loss of generality that $\tilde{x} = 1$, and let $\tilde{z} = \min\{z \in \mathcal{Z}\}$. The particular construction used in the proof of Lemma 1 has the property that $P_{S|ZX}(\tilde{s}_1|\tilde{z}, \tilde{x}) > 0$ and $P_{S|ZX}(\tilde{s}_2|\tilde{z}, \tilde{x}) > 0$. This implies that $H(S|X,Z) > 0$ since S is not a deterministic function of (X,Z) and the inequality in (69) is strict. □

Lemma 4: Let X and S be two jointly distributed random variables supported on finite or countably infinite X and S, respectively. Suppose Z^* is S-secure and Z is (Z^*, S)-secure. Then $Z = (Z^*, Z)$ is S-secure.

Proof:

$$\begin{aligned} I(S;Z) &= I(S;Z^*, \tilde{Z}) && (186) \\ &= I(S;Z^*) + I(S;\tilde{Z}|Z^*) && (187) \\ &= H(\tilde{Z}|Z^*) - H(\tilde{Z}|Z^*, S) && (188) \\ &= H(\tilde{Z}) - H(\tilde{Z}) = 0. && (189) \end{aligned}$$

□

Lemma 5: Let X and S be two jointly distributed random variables supported on finite or countably infinite X and finite S, respectively. Suppose Z is an optimizer for $\mathsf{h}_0(P_{XS})$, then

$$H(X|S,Z) = 0. \quad (190)$$

Proof: Suppose Z^* is an optimizer for $\mathsf{h}_0(P_{XS})$ and that $H(X|S,Z^*) > 0$; that is, X is not a deterministic function of (S, Z^*). Let $\tilde{Z}$ be the random variable constructed according to Lemma 1 with $X \leftarrow X$ and $S \leftarrow (S, Z^*)$. According to Theorem 5

$$I(X;\tilde{Z}) > 0. \quad (191)$$

Let $Z = (Z^*, \tilde{Z})$ and observe that Z is S-secure by Lemma 4. Finally,

$$\begin{aligned} I(S;Z) &= I(X;Z^*, \tilde{Z}) && (192) \\ &= I(X;Z^*) + I(X;\tilde{Z}|Z^*) && (193) \\ &= I(X;Z^*) + I(X, Z^*;\tilde{Z}) && (194) \\ &\geq I(X;Z^*) + I(X;\tilde{Z}) && (195) \\ &> I(X;Z^*). && (196) \end{aligned}$$

This contradicts the optimality of Z^*. □

Theorem 7: The upper bound follows directly from (68) and 1) ⇒ 2) is a trivial consequence of (74). Moreover, it is shown in [33] that computing $\mathsf{g}_0(P_{XS})$ and $\mathsf{h}_0(P_{XS})$ could be reduced to a standard linear program and therefore $\mathsf{g}_0(P_{XS})$ and $\mathsf{h}_0(P_{XS})$ are always achieved by some Z.

2) ⇒ 3): Suppose $\mathsf{g}_0(P_{XS}) = \mathsf{h}_0(P_{XS})$ and let Z be an S-secure random variable that achieves $\mathsf{g}_0(P_{XS})$. Note that in this case the Markov chain $S \leftrightarrow X \leftrightarrow Z$ must hold and hence $I(Z;S|X) = 0$. By assumption this Z also achieves $\mathsf{h}_0(P_{XS})$ and by Lemma 5 it satisfies $H(X|Z,S) = 0$. It follows from (68) that (74) holds with equality.

3) ⇒ 1): Suppose $\mathsf{h}_0(P_{XS}) = H(X|S)$ and let Z be an S-secure random variable that achieves $\mathsf{h}_0(P_{XS})$. Note that in this case $I(Z;S|X) = 0$ and hence the Markov chain $S \leftrightarrow X \leftrightarrow Z$ must hold. Therefor this Z also achieves $\mathsf{g}_0(P_{XS})$ and it follows from (68) that (74) holds with equality. □

D. Proofs for Section V

Lemma 6: Let (X, S) be jointly distributed random variables, and let U be an arbitrary random variable independent of X and S. Let Z be a random variable on $\mathcal{Z}$ such that

$$I(S; Z) = 0 \tag{197}$$

and

$$H(X|Z, U, S) = 0. \tag{198}$$

Then

$$H(Z) \geq \max_{s\in\mathcal{S}} H(X|S = s). \tag{199}$$

Proof: Equation (197) implies that

$$H(Z) = H(Z|S) = H(Z|S = s) \quad \forall s \in \mathcal{S}. \tag{200}$$

Equation (198) implies that

$$H(X|Z, U, S = s) = 0 \quad \forall s \in \mathcal{S}. \tag{201}$$

Finally, using the chain rule

$$\begin{aligned} H(X, Z|U, S = s) &= H(Z|U, S = s) + H(X|Z, U, S = s) && (202)\\ &\leq H(Z|S = s) && (203)\\ &= H(Z) && (204) \end{aligned}$$

and

$$\begin{aligned} H(X|S = s) &= H(X|U, S = s) && (205)\\ &\leq H(X, Z|U, S = s) && (206)\\ &\leq H(Z). && (207) \end{aligned}$$

Since (207) holds for all $s \in \mathcal{S}$ it must hold for a maximum over $\mathcal{S}$. □

Theorem 9: Fix an f_s-secure lossless code (c, d) and let $Z = \mathsf{c}(X, U)$ denote the output of the compressor. The secrecy constraint (89) can be rewritten as

$$I(Z; S) = 0 \tag{208}$$

and therefore

$$P_{Z|S}(z|s) = P_Z(z) \tag{209}$$

for all s, z. As Shannon observed in [1], either $P_Z(z) = 0$ or for every $s \in \mathcal{S}$ there needs to be a $k \in \{1, \ldots, K\}$ such that $P_{Z|US}(z|k, s) > 0$. But, due to the lossless constraint, $P_{Z|US}(z|k, s) > 0$ for at most one $s \in \mathcal{S}$. It follows that $K \geq |\mathcal{S}|$.

Next, we show that

$$H(Z) \geq \max\left\{\max_{s\in\mathcal{S}} H(X|S = s), \log|\mathcal{S}|\right\} \tag{210}$$

which is sufficient for (98) and (99) to hold. First, observe that random variables X, S, U and Z satisfy the conditions of Lemma 6 and so

$$H(Z) \geq \max_{s\in\mathcal{S}} H(X|S = s). \tag{211}$$

Secondly, on the one hand, (209) implies that

$$\sum_{s\in\mathcal{S}} P_{Z|S}(z|s) = |\mathcal{S}|P_Z(z). \tag{212}$$

On the other hand,

$$\begin{aligned} \sum_{s\in\mathcal{S}} P_{Z|S}(z|s) &= \sum_{s\in\mathcal{S}}\sum_{k=1}^{K} P_{Z|US}(z|k, s)P_U(k) && (213)\\ &= \sum_{k=1}^{K} P_U(k) \sum_{s\in\mathcal{S}} P_{Z|US}(z|k, s) && (214)\\ &\leq \sum_{k=1}^{K} P_U(k) = 1 && (215) \end{aligned}$$

since for a lossless code only one y can encode to z for any k. Thus, we get

$$P_Z(z) \leq \frac{1}{|\mathcal{S}|}, \quad \forall z \in \mathcal{Z} \tag{216}$$

and

$$H(Z) \geq \log|\mathcal{S}|. \tag{217}$$

□

Theorem 10: We use a two-part code described in Section II by representing X as (S, Z) where $I(S; Z) = 0$, $H(X|Z, S) = 0$ and

$$\mathbb{P}[\imath_Z(z) \ > \log L] \leq \max_{s\in\mathcal{S}} \alpha(P_{X|S=s}, L). \tag{218}$$

Recall that the existence of such Z is guaranteed by Theorem 1.

We use $\lceil\log|\mathcal{S}|\rceil$ bits to encode S with a one-time-pad. Z is encoded with a variable-length code that minimizes the probability of exceeding code length $\eta = l - \lceil\log|\mathcal{S}|\rceil$, as in [42]. That is, we encode the first $2^\eta - 1$ most likely elements of Z with strings of length η, and the rest with an arbitrary longer strings.

Let

$$|z\colon \imath_Z(z) \leq \log L| = M \tag{219}$$

and note that $M \leq L$. If $M \leq 2^\eta - 1$ then

$$\mathbb{P}[\ell(\mathsf{c}(X, k)) > l] \leq \mathbb{P}[\imath_Z(z) \ > \log L]. \tag{220}$$

Otherwise, $P_Z(z) \geq \frac{1}{L}$ for the $2^\eta - 1$ most probable $z \in \mathcal{Z}$ and we obtain

$$\mathbb{P}[\ell(\mathsf{c}(X, k)) > l] \leq 1 - \frac{2^\eta - 1}{L} \tag{221}$$

which completes the proof of (103).

Equation (104) is shown by taking $L = 2^\eta$ and noting that $M > 2^\eta - 1$ only if $P_Z(z) = \frac{1}{L}$ for all $z \in \mathcal{Z}$. In that case, $M = 2^\eta$ and it is possible to represent all $z \in \mathcal{Z}$ with strings of length η. Finally, (105) is obtained by loosening (104). □

Theorem 11: Fix a lossless variable-length code (c, d) and let $Z = \mathsf{c}(X, U)$ be the output of the encoder with Z denoting the space of codewords and U equiprobable on $\{1,\ldots,K\}$ and independent of X. Fix an arbitrary $s \in \mathcal{S}$ and $\tau > 0$. Define

$$\mathcal{L} = \{x \in \mathsf{f}_\mathsf{s}^{-1}(s) \colon P_{X|S}(x|s) \le 2^{-\eta-\tau}\} \tag{222}$$

$$\mathcal{C} = \{z \in \mathcal{Z} \colon \ell(z) \le l\} \tag{223}$$

$$\mathcal{C}_{x,k} = \{z \in \mathcal{C} \colon P_{XZ|SU}(x,z|s,k) > 0\}. \tag{224}$$

Note that

$$\sum_{k=1}^{K}\sum_{x\in\mathcal{L}} |\mathcal{C}_{x,k}| = \sum_{k=1}^{K}\sum_{x\in\mathcal{L}}\sum_{z\in\mathcal{C}} 1\{z \in \mathcal{C}_{x,k}\} \tag{225}$$

$$= \sum_{z\in\mathcal{C}}\sum_{x\in\mathcal{L}}\sum_{k=1}^{K} 1\{z \in \mathcal{C}_{x,k}\} \tag{226}$$

$$\le \sum_{z\in\mathcal{C}} 1 = |\mathcal{C}|. \tag{227}$$

Equation (227) holds because the secrecy assumption and the fact that $K = |\mathcal{S}|$ imply that $1\{z \in \mathcal{C}_{x,k}\} = 1$ for a unique pair (x,k). That is, there has to be an element from every set $\mathsf{f}_\mathsf{s}^{-1}(s)$ mapping to each z thus an element from $\mathcal{L}$ maps to z for only one value of k. Because the compressor is lossless, at most one element of $\mathcal{L}$ maps to z. Then

$$\mathbb{P}[\imath_{X|S}(x|s) > \eta + \tau | S = s] = \mathbb{P}[X \in \mathcal{L} | S = s] \tag{228}$$

$$= \mathbb{P}[(X,Z) \in \mathcal{L}\times\mathcal{C} | S = s] + \mathbb{P}[(X,Z) \in \mathcal{L}\times\mathcal{C}^c | S = s] \tag{229}$$

$$\le \mathbb{P}[(X,Z) \in \mathcal{L}\times\mathcal{C} | S = s] + \mathbb{P}[Z \in \mathcal{C}^c | S = s] \tag{230}$$

$$\le \sum_{k=1}^{K} \mathbb{P}[U = k]\mathbb{P}[(X,Z) \in \mathcal{L}\times\mathcal{C} | S = s, U = k] + \mathbb{P}[Z \in \mathcal{C}^c | S = s] \tag{231}$$

$$= \frac{1}{K}\sum_{k=1}^{K} \mathbb{P}[(X,Z) \in \mathcal{L}\times\mathcal{C} | S = s, U = k] + \mathbb{P}[Z \in \mathcal{C}^c | S = s] \tag{232}$$

$$\le \frac{1}{K}\sum_{k=1}^{K}\sum_{x\in\mathcal{L}} 2^{-\eta-\tau}|\mathcal{C}_{x,k}| + \mathbb{P}[Z \in \mathcal{C}^c | S = s] \tag{233}$$

$$\le \frac{2^l 2^{-\eta-\tau}}{K} + \mathbb{P}[Z \in \mathcal{C}^c | S = s] \tag{234}$$

$$= 2^{-\tau} + +\mathbb{P}[Z \in \mathcal{C}^c] \tag{235}$$

$$= 2^{-\tau} + +\mathbb{P}[\ell(\mathsf{c}(X,U)) \ge l] \tag{236}$$

where (235) follows because Z and S are independent. □

Theorem 12: We use a two-part code described in Section II by representing X as (S,Z) where $I(S;Z) = 0$,

$$\mathbb{P}[\imath_{X|ZS}(X|Z,S) > 0 | S = s] \le \alpha(P_{X|S=s}, L) \tag{237}$$

and Z is equiprobable on $\mathcal{Z} = \{1,\ldots,L\}$. Recall that the existence of such Z is guaranteed by Theorem 3.

We use $\lceil \log|\mathcal{S}| \rceil$ bits to encode S with a one-time-pad. For a fixed $s \in \mathcal{S}$, Z is encoded with a fixed-length code that uses

$$N = \min\{M', L\} \tag{238}$$

messages in the following way: N of the elements of $\mathcal{Z}$ are encoded losslessly, while the remaining $|\mathcal{Z}| - N$ contribute to the probability of error. They can, for example, be mapped to an arbitrary message and ignored by the decoder.

To select the N elements in $\mathcal{Z}$ which will be encoded losslessly, let

$$\mathcal{X}_s = \{z \colon \mathbb{P}[\imath_{X|ZS}(X|z,S) > 0 | S = s] = 0\}. \tag{239}$$

In other words, $|\mathcal{X}_s|$ contains all $x \in \mathcal{X}$ that can be reconstructed without error for a given $z \in \mathcal{Z}$ and $s \in \mathcal{S}$.

If $|\mathcal{X}_s| \le N$ then all of $\mathcal{X}_s$ is encoded losslessly (as well as some $z \in \mathcal{X}_s^c$ which we ignore). Then

$$\mathbb{P}[\mathsf{d}(\mathsf{c}(X,k),k) \ne X] \le \mathbb{P}[\imath_{X|ZS}(X|Z,S) > 0 | S = s]. \tag{240}$$

If $|\mathcal{X}_s| > N$ than arbitrary N elements of $\mathcal{X}_s$ are encoded losslessly. Note that in this case $N = M'$ and we obtain

$$\mathbb{P}[\mathsf{d}(\mathsf{c}(X,k),k) \ne X] \le 1 - \frac{M'}{L}. \tag{241}$$

Taking the expectation over S completes the proof of (114).

Equation (115) follows by particularizing (114) with $M' = L$. Equation (116) follows by the same relaxation as in Theorem 10. □

Theorem 13: Fix a fixed-length code (c,d) and let $\mathcal{Z} = \{1,\ldots,M,e\}$. Define a random variable

$$Z = \begin{cases} \mathsf{c}(X,U), & \mathsf{d}(\mathsf{c}(X,U)) = X \\ e, & \text{otherwise.} \end{cases} \tag{242}$$

Fix an arbitrary $s \in \mathcal{S}$ and $\tau > 0$. Define

$$\mathcal{L} = \{x \in \mathsf{f}_\mathsf{s}^{-1}(s) \colon P_{X|S}(x|s) \le 2^{-\log M' - \tau}\} \tag{243}$$

$$\mathcal{C} = \{1,\ldots,M\} \tag{244}$$

$$\mathcal{C}_{x,k} = \{z \in \mathcal{C} \colon \mathsf{d}(\mathsf{c}(X,U)) = x\}. \tag{245}$$

Note that

$$\sum_{k=1}^{K}\sum_{x\in\mathcal{L}} |\mathcal{C}_{x,k}| = \sum_{k=1}^{K}\sum_{x\in\mathcal{L}}\sum_{z\in\mathcal{C}} 1\{z \in \mathcal{C}_{x,k}\} \tag{246}$$

$$= \sum_{z\in\mathcal{C}}\sum_{x\in\mathcal{L}}\sum_{k=1}^{K} 1\{z \in \mathcal{C}_{x,k}\} \tag{247}$$

$$\le \sum_{z\in\mathcal{C}} 1 = |\mathcal{C}|. \tag{248}$$

Equation (248) holds because of the secrecy assumption and the fact that $K = |\mathcal{S}|$. That is, there has to be an element from every $\mathsf{f}_\mathsf{s}^{-1}(s)$ mapping to each z, thus an element from $\mathcal{L}$ maps to z for only one value of k. Because compression is lossless, only one element of $\mathcal{L}$ maps to z. Then

$$\mathbb{P}[\imath_{X|S}(X|S) > \log M' + \tau | S = s] = \mathbb{P}[X \in \mathcal{L} | S = s] \tag{249}$$

$$= \mathbb{P}[(X,Z) \in \mathcal{L}\times\mathcal{C} | S = s] + \mathbb{P}[(X,Z) \in \mathcal{L}\times\mathcal{C}^c | S = s] \tag{250}$$

$$\le \mathbb{P}[(X,Z) \in \mathcal{L}\times\mathcal{C} | S = s] + \mathbb{P}[Z \in \mathcal{C}^c | S = s] \tag{251}$$

$$\le \sum_{k=1}^{K} \mathbb{P}[U = k]\mathbb{P}[(X,Z) \in \mathcal{L}\times\mathcal{C} | S = s, U = k] + \mathbb{P}[Z \in \mathcal{C}^c | S = s] \tag{252}$$

$$= \frac{1}{K}\sum_{k=1}^{K}]\mathbb{P}[(X,Z)\in\mathcal{L}\times\mathcal{C}|S=s,U=k]$$
$$+\mathbb{P}[Z\in\mathcal{C}^c|S=s] \tag{253}$$
$$\leq \frac{1}{K}\sum_{k=1}^{K}\sum_{x\in\mathcal{L}}2^{-\log M'-\tau}|\mathcal{C}_{x,k}|+\mathbb{P}[Z\in\mathcal{C}^c|S=s] \tag{254}$$
$$\leq \frac{2^l 2^{-\log M'-\tau}}{|\mathcal{S}|}+\mathbb{P}[Z\in\mathcal{C}^c|S=s]. \tag{255}$$

The second term in (255) the is exactly the conditional probability of error. Averaging over all $s\in\mathcal{S}$ gives the desired result. □

E. Proofs for Appendix A

The proof of Theorems 14 makes use of McDiarmid's inequality, see for example [49], applied to the conditional information.

Theorem 15 (McDiarmid's Inequality): Let $\{X_i\}_{i=1}^n$ be independent (not necessarily identically distributed) random variables taking values in some measurable space $\mathcal{X}$. Consider a random variable $U = \mathsf{f}(X^n)$, where $\mathsf{f}:\mathcal{X}^n\to\mathbb{R}$ is a measurable function satisfying the bounded difference assumption. That is,

$$\sup_{x_1,\ldots,x_n,\tilde{x}_i\in\mathcal{X}}|\mathsf{f}(x_1,\ldots,x_i,\ldots x_n)-\mathsf{f}(x_1,\ldots,\tilde{x}_i,\ldots x_n)|\leq d_i \tag{256}$$

for every $1\leq i\leq n$ where d_i are non negative real constants. Then, for every $r\geq 0$,

$$\mathbb{P}\left[|U-\mathbb{E}U|\geq r\right]\leq 2\exp\left(-\frac{2r^2}{\sum_{i=1}^n d_i^2}\right). \tag{257}$$

Theorem 14: When (X^n,S^n) are distributed i.i.d. according to P_{XS} the function

$$\mathsf{f}(x^n,s^n)=\imath_{X^n|S^n}(x^n|s^n) \tag{258}$$

satisfies the bounded difference assumption with

$$d_i = d = \max_{x\in\mathcal{X},s\in\mathcal{S}}\imath_{X|S}(x|s). \tag{259}$$

Applying McDiarmid's inequality we thus obtain

$$\mathbb{P}\left[\left|\imath_{X^n|S^n}(X^n|S^n)-\sum_{i=1}^n H(X|S=s_i)\right|\geq n\gamma\middle| S^n=s^n\right]$$
$$\leq 2\exp\left(-\frac{2\gamma n}{d^2}\right) \tag{260}$$

for any $s^n\in\mathcal{S}^n$ and $\gamma>0$. Moreover,

$$\mathbb{P}\left[\left|\imath_{X^n|S^n}(X^n|S^n)-nH(X|S)\right|\geq n\gamma\right]\leq 2\exp\left(-\frac{2\gamma n}{d^2}\right) \tag{261}$$

for any $\gamma>0$.

Fix any $\gamma>0$ and let

$$s^*=\arg\max_{s\in\mathcal{S}}H(X|S=s). \tag{262}$$

Setting

$$l_n=\lceil\log|\mathcal{S}^n|\rceil+n(H(X|S=s^*)+\gamma) \tag{263}$$

and applying Theorem 10 with $\tau=\frac{1}{2}n\gamma$ we have that

$$\epsilon^*_{X^nS^n}(l_n,|\mathcal{S}^n|)\leq 2\exp\left(-\frac{\gamma n}{d^2}\right)+2^{-\frac{1}{2}n\gamma}. \tag{264}$$

Indeed, for every $s^n\in\mathcal{S}^n$

$$\mathbb{P}\left[\imath_{X^n|S^n}(X^n|S^n)\geq nH(X|S=s^*)+\frac{n\gamma}{2}\middle| S^n=s^n\right]$$
$$\leq\mathbb{P}\left[\imath_{X^n|S^n}(X^n|S^n)\geq H(X^n|S^n=s^n)+\frac{n\gamma}{2}\middle| S^n=s^n\right] \tag{265}$$
$$\leq 2\exp\left(-\frac{\gamma n}{d^2}\right). \tag{266}$$

On the other hand, letting

$$l_n=n(\log|\mathcal{S}|+\max_{s\in\mathcal{S}}H(X|S=s)-\gamma) \tag{267}$$

and applying Theorem 11 with $\tau=\frac{1}{2}n\gamma$ we have that

$$\epsilon^*_{X^nS^n}(l_n,|\mathcal{S}^n|)\geq 1-2\exp\left(-\frac{\gamma n}{d^2}\right)+2^{-\frac{1}{2}n\gamma}. \tag{268}$$

Note that

$$\mathbb{P}\left[\imath_{X^n|S^n}(X^n|s^{*n})<nH(X|S=s^*)-\frac{n\gamma}{2}\middle| S^n=s^{*n}\right]$$
$$\leq 2\exp\left(-\frac{2\gamma n}{d^2}\right) \tag{269}$$

follows from (260) and therefore

$$\mathbb{P}\left[\imath_{X^n|S^n}(X^n|s^{*n})\geq nH(X|S=s^*)-\frac{n\gamma}{2}\middle| S^n=s^{*n}\right]$$
$$\leq 1-2\exp\left(-\frac{2\gamma n}{d^2}\right). \tag{270}$$

Putting this together we obtain

$$\lim_{n\to\infty}\frac{\ell^*_{X^nS^n}(|\mathcal{S}^n|)}{n}\leq\log|\mathcal{S}|+\max_{s\in\mathcal{S}}H(X|S=s)+\gamma$$
$$+\lim_{n\to\infty}\left(2\exp\left(-\frac{\gamma}{d^2}n\right)(\log|\mathcal{X}|+\log|\mathcal{S}|)\right) \tag{271}$$
$$=\log|\mathcal{S}|+\max_{s\in\mathcal{S}}H(X|S=s)+\gamma \tag{272}$$

and

$$\lim_{n\to\infty}\frac{\ell^*_{X^nS^n}(|\mathcal{S}^n|)}{n}\geq\log|\mathcal{S}|+\max_{s\in\mathcal{S}}H(X|S=s)-\gamma. \tag{273}$$

Noting that (272) and (273) hold for all $\gamma>0$ gives (126). Equation (127) follows similarly by applying (27) to Theorems 12 and 13. □

Acknowledgment

The authors would like to thank the Associate Editor Dr. Shun Watanabe and the anonymous reviewer for their careful and insightful assessment of the manuscript. Yanina Y. Shkel would like to thank Robert Gallager for helpful discussions and Young-Han Kim for pointing out the prior examples of the Functional Representation Lemma in the information-theoretic literature.

References

[1] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, no. 4, pp. 656–715, Oct. 1949.

[2] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. 3rd Conf. Theory Cryptogr.*, Berlin, Germany: Springer, 2006, pp. 265–284, doi: 10.1007/11681878_14.

[3] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, nos. 3–4, pp. 211–407, 2013, doi: 10.1561/0400000042.

[4] O. Goldreich, *Foundations of Cryptography*, vol. 1. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[5] R. Shaefer, H. Boche, A. Khisti, and V. Pood, *Information Theoretic Security and Privacy of Information Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[6] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.

[7] D. Gündüz, E. Erkip, and H. V. Poor, "Source coding under secrecy constraints," in *Securing Wireless Communications at the Physical Layer*, R. Liu and W. Trappe, Eds. New York, NY, USA: Springer, 2010, ch. 8, pp. 173–200.

[8] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "Utility-privacy tradeoffs in databases: An information-theoretic approach," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 6, pp. 838–852, Jun. 2013.

[9] H. Yamamoto, "A rate-distortion problem for a communication system with a secondary decoder to be hindered," *IEEE Trans. Inf. Theory*, vol. 34, no. 4, pp. 835–842, Jul. 1988.

[10] C. Schieler and P. Cuff, "Rate-distortion theory for secrecy systems," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7584–7605, Dec. 2014.

[11] C. Schieler and P. Cuff, "The Henchman problem: Measuring secrecy by the minimum distortion in a list," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3436–3450, Jun. 2016.

[12] Y. Hayashi and H. Yamamoto, "Coding theorems for the Shannon cipher system with a guessing wiretapper and correlated source outputs," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2808–2817, Jun. 2008.

[13] N. Merhav and E. Arikan, "The Shannon cipher system with a guessing wiretapper," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1860–1866, Sep. 1999.

[14] N. Merhav, "A large-deviations notion of perfect secrecy," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 506–508, Feb. 2003.

[15] I. Issa, A. B. Wagner, and S. Kamath, "An operational approach to information leakage," *IEEE Trans. Inf. Theory*, vol. 66, no. 3, pp. 1625–1657, Mar. 2020.

[16] I. Issa and A. B. Wagner, "Measuring secrecy by the probability of a successful guess," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3783–3803, Jun. 2017.

[17] S. Asoodeh, M. Diaz, F. Alajaji, and T. Linder, "Estimation efficiency under privacy constraints," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1512–1534, Mar. 2019.

[18] J. Liao, O. Kosut, L. Sankar, and F. du Pin Calmon, "Tunable measures for information leakage and applications to privacy-utility tradeoffs," *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 8043–8066, Dec. 2019.

[19] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Advances in Cryptology—Eurocrypt*, S. Vaudenay, Ed. Berlin, Germany: Springer, 2006, pp. 486–503.

[20] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *Proc. IEEE 54th Annu. Symp. Found. Comput. Sci.* Washington, DC, USA: IEEE Computer Society, Oct. 2013, pp. 429–438, doi: 10.1109/FOCS.2013.53.

[21] A. Evfimievski, J. Gehrke, and R. Srikant, "Limiting privacy breaches in privacy preserving data mining," in *Proc. 22nd ACM SIGMOD-SIGACT-SIGART Symp. Princ. Database Syst.—PODS*, 2003, pp. 211–222, doi: 10.1145/773153.773174.

[22] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, "What can we learn privately?" *SIAM J. Comput.*, vol. 40, no. 3, pp. 793–826, Jan. 2011, doi: 10.1137/090756090.

[23] P. Cuff and L. Yu, "Differential privacy as a mutual information constraint," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Oct. 2016, pp. 43–54, doi: 10.1145/2976749.2978308.

[24] P. Kairouz, S. Oh, and P. Viswanath, "The composition theorem for differential privacy," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 4037–4049, Jun. 2017.

[25] G. Barthe and F. Olmedo, "Beyond differential privacy: Composition theorems and relational logic for f-divergences between probabilistic programs," in *Automata, Languages, and Programming*, F. V. Fomin, R. Freivalds, M. Kwiatkowska, and D. Peleg, Eds. Berlin, Germany: Springer, 2013, pp. 49–60.

[26] V. Feldman, I. Mironov, K. Talwar, and A. Thakurta, "Privacy amplification by iteration," in *Proc. IEEE 59th Annu. Symp. Found. Comput. Sci. (FOCS)*, Oct. 2018, pp. 521–532.

[27] S. Asoodeh, M. Diaz, and F. P. Calmon, "Privacy amplification of iterative algorithms via contraction coefficients," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2020, pp. 1–10.

[28] J. L. Massey, "Guessing and entropy," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 1994, p. 204.

[29] S. E. Fienberg, A. Rinaldo, and X. Yang, "Differential privacy and the risk-utility tradeoff for multi-dimensional contingency tables," in *Privacy in Statistical Databases*, J. Domingo-Ferrer and E. Magkos, Eds. Berlin, Germany: Springer, 2010, pp. 187–199.

[30] M. Jin, R. Jia, and C. J. Spanos, "Virtual occupancy sensing: Using smart meters to indicate your presence," *IEEE Trans. Mobile Comput.*, vol. 16, no. 11, pp. 3264–3277, Nov. 2017.

[31] F. D. P. Calmon, A. Makhdoumi, M. Medard, M. Varia, M. Christiansen, and K. R. Duffy, "Principal inertia components and applications," *IEEE Trans. Inf. Theory*, vol. 63, no. 8, pp. 5011–5038, Aug. 2017.

[32] S. Asoodeh, M. Diaz, F. Alajaji, and T. Linder, "Information extraction under privacy constraints," *Information*, vol. 7, no. 1, p. 15, Mar. 2016. [Online]. Available: https://www.mdpi.com/2078-2489/7/1/15

[33] B. Rassouli and D. Gunduz, "On perfect privacy," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 2551–2555.

[34] G. Bassi, M. Skoglund, and P. Piantanida, "Lossy communication subject to statistical parameter privacy," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 1031–1035.

[35] G. Bassi and M. Skoglund, "On the mutual information of two Boolean functions, with application to privacy," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2019, pp. 1197–1201.

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

[37] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[38] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge University Press, 2012.

[39] B. Hajek and M. Pursley, "Evaluation of an achievable rate region for the broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 1, pp. 36–46, Jan. 1979.

[40] F. Willems and E. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 3, pp. 313–327, May 1985.

[41] Y. Y. Shkel, R. S. Blum, and H. V. Poor, "Secure lossless compression," in *Proc. 52nd Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2018, pp. 1–6.

[42] I. Kontoyiannis and S. Verdú, "Optimal lossless data compression: Non-asymptotics and asymptotics," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 777–795, Feb. 2014.

[43] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Medard, "From the information bottleneck to the privacy funnel," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Nov. 2014, pp. 501–505.

[44] N. Merhav and M. Feder, "Universal prediction," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2124–2147, Oct. 1998.

[45] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 740–761, Jan. 2014.

[46] Y. Y. Shkel and S. Verdu, "A single-shot approach to lossy source coding under logarithmic loss," *IEEE Trans. Inf. Theory*, vol. 64, no. 1, pp. 129–147, Jan. 2018.

[47] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*. New York, NY, USA: Cambridge Univ. Press, 2006.

[48] T. Berger and R. W. Yeung, "Multiterminal source encoding with encoder breakdown," *IEEE Trans. Inf. Theory*, vol. 35, no. 2, pp. 237–244, Mar. 1989.

[49] M. Raginsky and I. Sason, "Concentration of measure inequalities in information theory, communications, and coding," *Found. Trends Commun. Inf. Theory*, vol. 10, nos. 1–2, pp. 1–246, 2013, doi: 10.1561/0100000064.

Yanina Y. Shkel (Member, IEEE) received the B.S. degrees in mathematics and in computer science and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Wisconsin-Madison in May 2005, August 2010, and December 2014, respectively.

Before joining the Graduate School, she worked as a Database Developer at Morningstar, Inc., and in 2013, she spent her time at 3M Corporate Research, as an Intern. From 2014 to 2019, she was a Post-Doctoral Researcher with Princeton University and the University of Illinois at Urbana–Champaign. She is currently a Scientist with the École Polytechnique Fédérale de Lausanne (EPFL), where she is affiliated with the Information Processing Group (IPG). Her research interests include theoretical aspects of data science, information, learning, coding theory, statistics, and cryptography, with a particular focus on applications to privacy and secrecy. She was a recipient of the NSF Center for Science of Information (CSoI) Postdoctoral Fellowship.

Rick S. Blum (Fellow, IEEE) received the B.S. degree in electrical engineering from Pennsylvania State University in 1984, and the M.S. and Ph.D. degrees in electrical engineering from the University of Pennsylvania in 1987 and 1991, respectively. From 1984 to 1991, he was a member of technical staff at General Electric Aerospace, Valley Forge, PA, USA, and he graduated from GE's Advanced Course in Engineering. Since 1991, he has been with the Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA, USA, where he is currently a Professor and holds the Robert W. Wieseman Endowed Professorship in electrical engineering. His research interests include signal processing for security, smart grid, communications, sensor networking, radar, and sensor processing. He was on the editorial board for the *Journal of Advances in Information Fusion* of the International Society of Information Fusion. He was an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING and IEEE COMMUNICATIONS LETTERS. He has edited Special Issues on IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, and IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He was a member of the SAM Technical Committee (TC) of the IEEE Signal Processing Society. He was a member of the Signal Processing for Communications TC of the IEEE Signal Processing Society, and is a member of the Communications Theory TC of the IEEE Communication Society. He was on the Awards Committee of the IEEE Communication Society. He served, two terms, as an IEEE Signal Processing Society Distinguished Lecturer. He is the IEEE Third Millennium Medal Winner, the Eleanor and Joseph F. Libsch Research Award Winner, a member of Eta Kappa Nu and Sigma Xi, and holds several patents. He was awarded the ONR Young Investigator Award in 1997 and the NSF Research Initiation Award in 1992. His IEEE Fellow Citation "for scientific contributions to detection, data fusion, and signal processing with multiple sensors" acknowledges contributions to the field of sensor networking.

H. Vincent Poor (Life Fellow, IEEE) received the Ph.D. degree in electrical engineering and computer science from Princeton University in 1977.

From 1977 to 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990, he has been on the faculty at Princeton, where he is the Michael Henry Strater University Professor of Electrical Engineering. From 2006 to 2016, he served as the Dean of Princeton's School of Engineering and Applied Science. He has also held visiting appointments at several other institutions, most recently at Berkeley and Cambridge. His research interests include the areas of information theory, machine learning, and network science, and their applications in wireless networks, energy systems, and related fields. Among his publications in these areas is the forthcoming book *Advanced Data Analytics for Power Systems* (Cambridge University Press, 2021).

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, and is a foreign member of the Chinese Academy of Sciences, the Royal Society, and other national and international academies. Recent recognition of his work includes the 2017 IEEE Alexander Graham Bell Medal and a D.Eng. *honoris causa* from the University of Waterloo, awarded in 2019.