Secrecy by design with applications to privacy and compression

Yanina Y. Shkel, Rick S. Blum, and H. Vincent Poor

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 principle technical contributions are decomposition bounds that allow 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 $S$. 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 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 $X$.

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-ensuring 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.

A. Perfect secrecy

In his seminal work on “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. 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 public 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]–[7]. Other approaches to partial secrecy—motivated by guessing, maximum information leakage, and rate-distortion [8]–[15]–have also been proposed. The problem of partial secrecy in the setting of database privacy has been widely studied in computer science literature; approaches like differential privacy [2], [3], $(\epsilon, \delta)$-differential privacy [16], and local differential privacy [17]–[19] can also be viewed through the lens of information leakage [20].

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 interoperability and utility cost. For example, equivocation and rate-distortion approaches have been criticized as secrecy measures since their operational implications are not always clear [15], [21]. Likewise, differential privacy has been shown to be too restrictive for many applications, and not restrictive enough for some [22]. 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

\[
I(S; Z) = 0
\]

instead of the perfect secrecy constraint

\[
I(X; Z) = 0.
\]

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 [23], [24].

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

**Example 1** (Exponential Family). Suppose \(X \sim P_\beta\) where \(P_\beta\) belongs to an exponential family which is parametrized by \(\theta\) and the secret \(S\) is the sufficient statistic for this family. That is

\[
P_\beta(x) = \exp(\beta^T \sigma(x) - \psi(\beta)) r(x)
\]

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

Example 1 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 \(X\).

**Example 2** (N-User Process). Let \(S = \{1, \ldots, N\}\) be the index of one of \(N\) users and \(A\) is 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 \(A\). We consider two versions of this setting.

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

\[
P_X(s, a) = P_s(a).
\]

In this case the data consists of the job being requested and the metadata communicating the identity of the use.

Next, suppose that \(X = A\) and

\[
P_X(a) = \sum_{s \in S} P_s(a) P_S(s).
\]

\(^1\)We borrow this terminology from 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 2 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, 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 2 captures a very general setting and subsumes most later examples in this work.

To the best of our knowledge, this work is the first to focus on secrecy by design as a principled approach to information-theoretical secrecy. However, there have been a number of relevant works that study optimal secrecy mechanism which target a function of data. Most notably, recent work on perfect secrecy [25]–[27] 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 [28], [29], and securing correlated random variables for compression using equivocation as a leakage measure [30]. Finally, maximal equivocation as a leakage measure [31].

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 $S$ is given. We study mechanisms for constructing public information $Z$ that 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$^2$

$$I_X(x) = \log \frac{1}{P_X(x)},$$

and

$$I_{X|Y}(x|y) = \log \frac{1}{P_{X|Y}(x|y)},$$

respectively. The entropy and conditional entropy are given by

$$H(X) = \mathbb{E} [I_X(X)],$$

were $H(X)$ denotes the binary entropy function.

$$H(X|Y) = \mathbb{E} [I_{X|Y}(X|Y)],$$

and

$$H(X|Y=y) = \mathbb{E} [I_{X|Y}(X|Y=y)].$$

respectively. Moreover,

$$H(X|Y = y) = \mathbb{E} [I_{X|Y}(X|y)|Y = y].$$

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

$$I_{X|Y}(x|y) = I_X(x),$$

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

$$H(X) \geq H(X|Y)$$

always holds. However, $H(X|Y = y)$ could be greater than or smaller than $H(X)$; see [31].

**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 $S = f_s(X)$ for some secrecy function

$$f_s : X \rightarrow S.$$}

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

**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 $S = A^n_k$. That is, the secret information is the first $k$ bits of $a^n$. If $A^n_k$’s are independent and identically distributed (i.i.d), the answer to the question above is trivially “yes” with $Z = A^n_{k+1}$.

$^2$Throughout this paper all logarithm and exponential functions are assumed to have base two.
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 2. In this case it is not enough to simply hide the value of $s \in S$ 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 [32] and has been proved independently in [33]–[35] 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 $X$ is countably infinite and show in Lemma 2 that there exists an $S$-secure $Z$ such that

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

We show a lower bound on the entropy of $Z$,

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

in Lemma 3. We conclude Section III with Theorems 1-3 which give bounds on the information spectrum of $Z$ in the style of [36]. 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 [37]. The general privacy funnel function is given by

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

In [25]–[27] special attention is paid to the perfect privacy regime in which $I(S; Z) = 0$. Formally, this is done by studying the function

$$g_0(P_{XS}) = \sup_{Z : I(S; Z) = 0, S \leftrightarrow X \leftrightarrow Z} I(X; Z).$$

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

$$S \leftrightarrow X \leftrightarrow Z,$$

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 new perfect privacy function:

$$h_0(P_{XS}) = \sup_{Z : I(S; Z) = 0} I(X; Z).$$

It follows from (21) and (23) that

$$h_0(P_{XS}) \geq g_0(P_{XS}).$$

We show in Example 6 that the gap between $h_0(P_{XS})$ and $g_0(P_{XS})$ 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

$$g_0(P_{XS}) \leq h_0(P_{XS}) \leq H(X|S)$$

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 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 [38]–[41]. 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 details in Section IV as well.

1Note that [25]–[27] 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].
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 rate of $Z$ is bounded from below by

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

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 weather $S$ 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 limits of variable-length compression behaves like

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

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

$$H(X) = H(S) + H(X|S).$$  \hfill (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).

\footnote{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 $\log |S| + \max_{s \in S} H(X|S = s)$ in the same way that the non-asymptotic codeword lengths for traditional compression are approximated by $H(X|S)$.}

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 reconstituted from $(S, Z)$ without error, while for fixed-length compression a small probability of error is allowed.
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 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 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 [30] 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 [32] and was independently derived in [33]–[35]. While it has been previously used as an auxiliary result to analyze rate-regions in network information theory, 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 finite alphabets $X$ and $S$, respectively. There exists a random variable $Z$ on $Z$ such that $S$ and $Z$ are independent, that is

$$I(S; Z) = 0, \quad (29)$$

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

$$H(X|Z, S) = 0, \quad (30)$$

and the support of $Z$ is bounded by

$$|Z| \leq |S|(|X| - 1) + 1. \quad (31)$$

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

$$(X, S) \leftrightarrow U \leftrightarrow Z \quad (32)$$

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 $X = \{1, \ldots, |X|\}$ if $X$ is finite and $X = \{1, 2, \ldots\}$ if $X$ is countably infinite. Let $\mathcal{X}_s = \{x : P_{X|S}(x|s) > 0\}$. For each $s \in S$ and $x \in \mathcal{X}_s$ define

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

where

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

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

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

Then

$$U = U(k_{\tilde{X},S}, k_{X,S})$$

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

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

and therefore $U$ and $S$ are independent.

Next, the alphabet of $Z$ is given by

$$Z = \bigcup_{s \in S} \bigcup_{x \in \mathcal{X}_s} \{k_{x,s}\} \quad (38)$$

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

$$Z = \min\{z \in Z : U \leq z\}. \quad (39)$$

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 $\tilde{z} \in Z$ exists. Thus, (30) holds since this construction guarantees that, for a fixed $s \in S$, each $x \in \mathcal{X}_s$ is mapped to a disjoint segment of the unit interval.

Remark 1. When $S$ is a deterministic function of $X$, (31) can be tightened to

$$|Z| \leq |X| - |S| + 1. \quad (40)$$

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
According to Lemma 1, an $S$-secure $Z$ 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 $\mathcal{S}$ 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 $\mathcal{S}$ which satisfies (29), (30), and

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

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

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

The proofs of Lemmas 2 and 3 are given in Appendix B.

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

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

According to Lemmas 1 and 2,

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

Then,

$$P_Z(z) = \begin{cases} 1 \quad & z = z_1, \\ \frac{1}{4} \quad & z \in \{z_2, z_3\}. \end{cases}$$

and

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

**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 a more refined 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 $Z$.

To state these theorems we define

$$\alpha(P_Y, L) = \mathbb{E} \left[ 1 \left\{ r_Y(Y) < \log L \right\} \exp \left( r_Y(Y) - \log L \right) \right] + \mathbb{P} \left[ r_Y(Y) > \log L \right]$$

where $P_Y$ is an arbitrary distribution and any real $L \geq 1$.

**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 $L \geq 1$ there exists a random variable $Z$ satisfying (29), (30), and

$$\mathbb{P}[r_Z(Z) > \log L] \leq \max_{s \in \mathcal{S}} \alpha(P_X|S = s, L) \leq \max_{s \in \mathcal{S}} \min_{\tau > 0} \left\{ \mathbb{P} \left[ r_X(X|S) > \log L - \tau | S = s \right] + \exp(-\tau) \right\}.$$
Theorem 4 relates the distribution of $\mathbb{I}_Z(Z)$ to the distribution to the worst-case (over $s \in S$) distribution of $\mathbb{I}_{X|S}(X|s)$. Observe a nice correspondence between (53) 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 $X$ and a finite alphabet $S$, respectively. If a random variable $Z$ satisfies (29) and (30) then

$$
P_{[Z](Z) > \log L] \geq \max_{s \in S \cup \{0\}, \tau > 0} \left\{ \mathbb{P}[X|S(Z|S) > \log L + \tau | S = s] - \exp(-\tau) \right\}
$$

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 achieved when $S = \{0, e\}$.

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 $X$ and a finite alphabet $S$, respectively. Fix an integer $L \geq 1$. Then, there exists a random variable $Z$ on $Z$ satisfying (29) and

$$
P_{[X|Z}(X|Z, S) > 0 | S = s) \leq \alpha(P_{X|S}(s, L))
$$

for all $s \in S$. Moreover, $Z$ is equiprobable on $Z = \{1, \ldots, L\}$. 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 [37]. We focus on the perfect privacy functions $g_0(P_{X|S})$ and $h_0(P_{X|S})$ defined in (21) and (23), respectively. We begin by reviewing the necessary and sufficient conditions for $g_0(P_{X|S})$ and $h_0(P_{X|S})$ to be nonzero. We then show that $g_0(P_{X|S})$ and $h_0(P_{X|S})$ 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 [42] (see also [25, Theorem 10]).

**Theorem 4** ([25], [42]). Let $X$ and $S$ be two jointly distributed random variables. Then

$$
g_0(P_{X|S}) > 0
$$

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 $|S| < |X|$ since in that case there will always be linearly dependent rows in the matrix $[P_{S|X}(\cdot|x)]$. On the other hand, if $|S| \geq |X|$ perfect privacy is only achievable if the joint distribution $P_{X|S}$ has a very particular structure.

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

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

for some $0 \leq \epsilon \leq 1$. The construction in Lemma 1 gives the lower bound

$$
h_0(P_{X|S}) \geq 1 - h\left(\frac{1}{2}(1 - \epsilon)\right).
$$

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

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

and according to Theorem 4

$$
g_0(P_{X|S}) = 0.
$$

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 $h_0(P_{X|S})$.

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 alphabet $X$ and finite alphabet $S$, respectively. Then

$$
h_0(P_{X|S}) > 0
$$

if and only if

$$
H(X|S) > 0.
$$

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 [27] by explicitly constructing a $Z$ on $Z = \{z_1, z_2\}$ in a manner similar to the proof of Theorem 4 given in [25].
The functions $g_0(P_{X|S})$ and $h_0(P_{X|S})$ 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 [27]. Moreover, [27] shows that $g_0(P_{X|S})$ and $h_0(P_{X|S})$ 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 $g_0(P_{X|S})$ and $h_0(P_{X|S})$ collapses into one.

**B. New lower and upper bounds on perfect privacy**

In order to further explore the relationship between $h_0(P_{X|S})$ and $g_0(P_{X|S})$, we begin with the following basic, but important, observation. Let $X$, $S$, and $Z$ be arbitrary random variables. Then, by the chain rule

$$I(X;S,Z) = I(X;Z) + I(S;Z|X)$$

(64)

$$= I(S;Z) + I(X;Z|S)$$

(65)

$$= I(S;Z) + H(X|S) - H(X|S,Z).$$

(66)

Therefore, for any $S$-secure $Z$ it holds that

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

(67)

Since (67) holds for any $S$-secure $Z$, $H(X|S)$ is an upper bound on $h_0(P_{X|S})$ and on $g_0(P_{X|S})$. This is formalized for $h_0(P_{X|S})$ in Theorem 7 and was shown for $g_0(P_{X|S})$ in [25]. Equation (67) could also be used to lower bound $h_0(P_{X|S})$, as is shown in the following Theorem.

**Theorem 6.** Let $X$ and $S$ be two jointly distributed random variables supported on finite $X$ and $S$, respectively, then

$$h_0(P_{X|S}) \geq H(X|S) - H(X|S,Z) - I(S;Z|X) = H(X) - H(S).$$

(68)

The lower bound (68) 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 $X$

$$g_0(P_{X|S}) = h_0(P_{X|S}) = H(X|S)$$

(69)

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 1 where $X^n$ is i.i.d. Bernoulli($p$) and

$$S_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$  

(70)

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

$$g_0(P_{X|S_n}) = h_0(P_{X|S_n}) = \sum_{k=0}^{n} p^k (1-p)^{n-k} \log \left( \frac{n}{k} \right).$$

(71)

Using basic techniques from the method of types and Stirling’s approximation [31] it can be shown that

$$\lim_{n \to \infty} \frac{g_0(P_{X^n|S_n})}{n} = \lim_{n \to \infty} \frac{h_0(P_{X^n|S_n})}{n} = h(p).$$

(72)

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 $X$ and $S$, respectively. Then

$$g_0(P_{X|S}) \leq h_0(P_{X|S}) \leq H(X|S).$$

(73)

Moreover, the following are equivalent:

1) $g_0(P_{X|S}) = H(X|S)$,
2) $h_0(P_{X|S}) = g_0(P_{X|S})$,
3) $h_0(P_{X|S}) = H(X|S)$.

The proof of Theorem 7 is given in Appendix C. It is based on (67) 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 $g_0(P_{X|S})$. 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 $h_0(P_{X|S})$ must also satisfy $H(X|S,Z) = 0$.

The necessary and sufficient conditions for equalities in (73) 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 $X = \{0,1\}$ and $S = \{0,1\}$. Suppose that $(X,S)$ are jointly distributed according to

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

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

$$g_0(P_{X|S}) = h_0(P_{X|S}) = h(\epsilon).$$

(75)

The random variable $Z$ on $Z = \{z_1,z_2\}$ that achieves (75) is produced by the joint distribution

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

(76)

In general, there does not appear to be a better way to characterize necessary and sufficient conditions for $h_0(P_{X|S})$ and $g_0(P_{X|S})$ to achieve (73) 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

$$S_x = \{ s \in S : P_{X|S}(x|s) > 0 \} .$$

$g_0(P_{X|S}) = H(X|S)$ only if for all $x \in \mathcal{X}$,

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

However, suppose $|S| = 2$. Then $g_0(P_{X|S}) = H(X|S)$ if and only if (78) 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 logarithm loss [38]–[41]. 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 $\hat{P} \in \mathcal{P}(X)$ is given by

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

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

$$\hat{P} : Z \rightarrow \mathcal{P}(\mathcal{X})$$

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

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

$$\mathbb{E} [\ell(X, \hat{P})] = \mathbb{E} \left[ \log \frac{1}{P(X|Z)} \right]$$

$$= \mathbb{E} \left[ \log \frac{P_{X|Z}(X|Z)}{P(X|Z)} \right] + H(X|Z)$$

$$= D(P_{X|Z} \| \hat{P}) + H(X|Z)$$

$$\geq H(X|Z).$$

Equation (83) 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

$$I(X; Z) = H(X) - H(X|Z).$$

In light of this discussion, Theorem 7 shows that

$$\inf_{Z: \ell(X, \hat{P})} \inf_{\hat{P} : Z \rightarrow \mathcal{P}(X)} \mathbb{E} [\ell(X, \hat{P})]$$

$$\geq \inf_{Z: \ell(X, \hat{P})} \inf_{\hat{P} : Z \rightarrow \mathcal{P}(X)} \mathbb{E} [\ell(X, \hat{P})] \geq I(X; S).$$

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. APPLICATION TO COMPRESSION**

We apply the 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 $K = |S|$ 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:

Encoder: $c : \mathcal{X} \times \{1, \ldots, K\} \rightarrow \{0,1\}^*$

Decoder: $d : \{0,1\}^* \times \{1, \ldots, K\} \rightarrow \mathcal{X}$

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

$$I(c(X,U);S) = 0.$$ (88)

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

$$\mathbb{P}[d(c(X,U),U) = X] = 1.$$ (89)

We assume that, given a fixed key value $k$, $c(x, k)$ can 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

$$\mathbb{E} [\ell(c(X,k))] \leq l, \quad \forall k \in \{1, \ldots, K\}.$$ (90)

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

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

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

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

and

$$c^s_{XS}(l,K) = \inf \{ \epsilon : \exists S \text{-secure } (l,K,\epsilon) \text{-code} \}.$$ (93)

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
\[
\ell_{XS}(n) \leq \sum_{s \in \mathcal{S}} H(S|S = s) + 1 + \lfloor \log |\mathcal{S}| \rfloor. \tag{94}
\]
If $X$ is finite then
\[
\ell_{XS}^*(n) \leq \lfloor \log \left(|\mathcal{X}| - 1 \right) \rfloor + \lfloor \log |\mathcal{S}| \rfloor, \tag{95}
\]
and if $S$ is also a deterministic function of $X$ then
\[
\ell_{XS}^*(n) \leq \lfloor \log \left(|\mathcal{X}| - |\mathcal{S}| + 1 \right) \rfloor + \lfloor \log |\mathcal{S}| \rfloor. \tag{96}
\]

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

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
\[
\ell_{XS}(K) \geq \max_{s \in \mathcal{S}} H(X|S = s) \tag{97}
\]
holds for any $K \in \{1, 2, \ldots \}$.

If $S$ is a deterministic function of $X$ then for any $K \geq |\mathcal{S}|$,
\[
\ell_{XS}^*(K) \geq \log |\mathcal{S}|. \tag{98}
\]
and $f_{\tau}$-secure codes do not exist for $K < |\mathcal{S}|$.

Lower bound (97) is a corollary of Lemma 3 with a caveat of accounting for the shared secret key and (98) is a corollary of (1). Theorem 9 is proved in Appendix D. We highlight that, although our primary focus is on the case $K = |\mathcal{S}|$, (97) holds for an arbitrary value of $K$ and its proof does not use the fact that $U$ is equiprobable on $\{1, \ldots, K\}$. 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 10. Applying Theorems 8 and 9 with $X \leftarrow X^n$ and $S \leftarrow S$,
\[
\max_{\tau \in \{0, 1\}} \max_{k \in \{0, 1\}} \log n + \log (n + 1) \leq \ell_{XY}^*(n + 1) \tag{99}
\]
\[
\leq \log (2^n - n) + \log (n + 1). \tag{100}
\]
It is easy to check that
\[
\lim_{n \to \infty} \frac{\ell_{XS}^*(n + 1)}{n} = 1. \tag{101}
\]
Asymptotically the compression of this source is impossible even under a partial secrecy constraint. However, the partial secrecy is achieved with a reduced secret key size of $K = n + 1$ as opposed to $K = 2^n$ which would be required for full secrecy.

Theorems 8 and 9 are not tight in general. However, as is the case in traditional compression [36], it is possible to get better asymptotic bounds on (92) by leveraging bounds on the probability of excess length fundamental limit (93). Such upper and lower bounds on (93) 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 $\nu \in \{1, 2, \ldots \}$
\[
c_{XS}(l, |S|) \leq \max_{s \in \mathcal{S}} \left\{ \max_{\nu \geq 1} \left\{ \alpha(P_{X|S=s}, \nu) \right\} + 2^{\nu - 1} \right\} \tag{102}
\]
where $\nu = l - \lfloor \log |\mathcal{S}| \rfloor$. Moreover,
\[
c_{XS}(l, |S|) \leq \max_{s \in \mathcal{S}} \left\{ \alpha(P_{X|S=s}, 2^\nu) \right\} \tag{103}
\]
\[
\leq \max_{s \in \mathcal{S}} \min_{\tau > 0} \left\{ \mathbb{P} \left[ i_{X|S}(X|S) \geq \nu - \tau |S = s\right] \geq 2^{-\tau} \right\} \tag{104}
\]
where $\nu = l - \lfloor \log |\mathcal{S}| \rfloor$.

Note the correspondence between (104) 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
\[
\max_{\nu \geq 1} \left\{ \mathbb{P} \left[ i_{X|S}(X|S) \geq \nu + \tau |S = s\right] \geq 2^{-\tau} \right\} \leq c_{XS}(l, |S|) \tag{105}
\]
where $\nu = l - \log |\mathcal{S}|$.

Theorems 10 and 11 are proved in Appendix D. Taking $|\mathcal{S}| = 1$ recovers traditional compression. Particularizing Theorem 11 to this case we obtain
\[
\max_{\tau > 0} \left\{ \mathbb{P} \left[ i_{X}(X) \geq l + \tau \right] \geq 2^{-\tau} \right\} \leq c_{XS}(l, 1) \tag{106}
\]
which is exactly [36, Theorem 4].

We show in Appendix A that the asymptotic fundamental limit of lossless variable-length compression is
\[
\log |\mathcal{S}| + \max_{s \in \mathcal{S}} H(X|S = s) \tag{107}
\]
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:

Encoder: $c: \mathcal{X} \times \{1, \ldots, K\} \to \{1, \ldots, M\}$

Decoder: $d: \{1, \ldots, M\} \times \{1, \ldots, K\} \to \mathcal{X}$

Suppose $S \equiv f_{\tau}(X)$. An $f_{\tau}$-secure code $(c, d)$ (see (88)) is an $(M, K, \epsilon)$-almost-lossless code if
\[
\mathbb{P} \left[ d(c(X, k)) \in f_{\tau}^{-1}(S) \right] = 1 \tag{108}
\]
\[
\mathbb{P} \left[ d(c(X, k)) \neq X \right] \leq \epsilon, \quad \forall k \in \{1, \ldots, K\}. \tag{109}
\]
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 $c(x, k)$ can 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 $f_\epsilon$-secure almost-lossless compression are given by

$$
\hat{\epsilon}_{XS}(M, K) = \inf \{ \epsilon : \exists f_\epsilon \text{-secure } (M, K, \epsilon) \text{-code} \} \quad (110)
$$

and

$$
\hat{M}_{XS}(\epsilon, K) = \inf \{ M : \exists f_\epsilon \text{-secure } (M, K, \epsilon) \text{-code} \}. \quad (111)
$$

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

$$
\hat{M}_{XS}(\epsilon, K) = \inf \{ M : \hat{\epsilon}_{XS}(M, K) \leq \epsilon \}. \quad (112)
$$

**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. Let $S$ be a deterministic function of $X$. Then

$$
\hat{\epsilon}_{XS}(M, |S|) \leq \min_{L \in \{1, \ldots \}} \mathbb{E} \left[ \max \left\{ \alpha(P_{X|S=y, L}), 1 - \frac{M'}{L} \right\} \right] \quad (113)
$$

$$
\leq \mathbb{E} \left[ \{ \alpha(P_{X|S=y, M'}) \} \right] \quad (114)
$$

$$
\leq \mathbb{E} \left[ \min_{\tau > 0} \{ P_{X|S=y}(X|S) \geq \log M' - \tau + 2^{-\tau} \} \right] \quad (115)
$$

$$
\leq \min_{\tau > 0} \{ P_{X|S=y}(X|S) \geq \log M' - \tau + 2^{-\tau} \} \quad (116)
$$

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

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

$$
\max_{\tau > 0} \{ P_{X|S=y}(X|S) \geq \log M' + \tau - 2^{-\tau} \} \leq \hat{\epsilon}_{XS}(M, |S|) \quad (117)
$$

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

It is observed in [36] 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 $|S| = 1$ recovers traditional fixed-length compression. Particularizing Theorem 13 to this case yields

$$
\max_{\tau > 0} \{ P_{X}(X) \geq \log M + \tau - 2^{-\tau} \} \leq \hat{\epsilon}_{XY}(M, 1) \quad (118)
$$

which is exactly the application of [36, 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 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

$$
\log |S| + H(X|S) \quad (119)
$$

whenever $K = |S|$. The tradition 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 recovers the variable-length setting.

**C. Secrecy by design versus leakage**

In [30, 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 [30, 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 [30] it is possible to leverage fixed-length compression to achieve variable-length compression with an asymptotically negligible increase in leakage. Comparing [30, Case 7] to (107) 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 (97) 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 [30] 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, [30] 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 (107) and [30] shows that, in general, a small leakage regime is not a good approximation for 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 which let us represent and 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 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 is 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 which 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.

ACKNOWLEDGEMENTS

This work was supported in part by the U.S. National Science Foundation under Grants CNS-1702555, CCF 0939370 and CCF 1513915. 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 Shkel would like to thank Robert Gallager for helpful discussions and Young-Han Kim for pointing out the prior examples of Functional Representation Lemma in information-theoretic literature.

APPENDIX

A. Asymptotic fundamental limits for compression

In this section we derive asymptotic fundamental limits for variable-length and fixed-length secure compression. We consider two asymptotic settings. In the first setting the secrecy function is assumed to be an $n$-fold product of the single-letter secrecy function. In the second setting the secrecy function is assumed to be additive; this assumption holds, for example, in the case when the source distribution is a member of an exponential family and the secrecy function is its sufficient statistic.

1) The product secrecy function: Let $X \sim P_X$ and $S = f_s(X)$ where

$$f_s : X \rightarrow S.$$ (120)

Given a single letter secrecy function (120), define an $n$-letter secrecy function

$$f_s^n : X^n \rightarrow S^n$$ (121)

via

$$f_s^n(x^n) = (f_s(x_1), \ldots, f_s(x_n)).$$ (122)

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

$$\mathcal{R}_{XS}^* = \lim_{n \rightarrow \infty} \frac{\ell_{X^nS^n}([S^n])}{n}.$$ (123)

The asymptotic fixed-length fundamental limit is given by

$$\bar{\mathcal{R}}_{XS} = \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \log \tilde{M}_{X^nS^n}(\epsilon, |S^n|).$$ (124)

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

Theorem 14.

$$\mathcal{R}_{XS}^* = \log |S| + \max_{s \in S} H(X|S = s)$$ (125)

and

$$\bar{\mathcal{R}}_{XS} = \log |S| + H(X|S).$$ (126)

Equation (125) is a direct consequence of Theorems 10 and 11, while (126) is a a direct consequence of Theorems 12 and 13. A detailed proof of Theorem 14 is given in Appendix E.
2) The additive secrecy function: Let \( X \sim P_X \) and assume
\[
f_s: \mathcal{X} \to \mathcal{S} \subset \mathbb{R}.
\]
(127)
Given a single letter secrecy function (127), define an \( n \)-letter secrecy function
\[
f_s^n: \mathcal{X}^n \to \mathcal{S}_n
\]
(128)
via
\[
f_s^n(x^n) = \frac{1}{n} \sum_{i=1}^n f_s(x_i).
\]
(129)
We assume that \( X^n \) is i.i.d. according to \( P_X \), and \( S_n = f_s^n(X^n) \). The asymptotic variable-length fundamental limit is given by
\[
S_{\mathcal{X}S}^* = \lim_{n \to \infty} \frac{f_s^n(X^n,S_n)(|S_n|)}{n}.
\]
(130)
The asymptotic fixed-length fundamental limit is given by
\[
\tilde{S}_{\mathcal{X}S} = \lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log |M_{\mathcal{X}S}(\epsilon,|S_n|)|.
\]
(131)
The next theorem characterizes the two fundamental limits for the additive setting.

Theorem 15.
\[
S_{\mathcal{X}S}^* = \max_{s \in \mathcal{S}} H(X|S = s)
\]
(132)
and
\[
\tilde{S}_{\mathcal{X}S} = H(X).
\]
(133)
A detailed proof of Theorem 15 is given in the Appendix E. Comparing the additive setting to the product setting we see that the dependence on \( \log |\mathcal{S}| \) terms disappears since \( |\mathcal{S}_n| \) grows logarithmically in \( n \).

Note also that according to Theorem 9, \( S_{\mathcal{X}S}^* \) can be lower bounded by \( \max_{s \in \mathcal{S}} H(X|S = s) \) independent of the size or the distribution of the shared secret key. Thus, asymptotically, there is no advantage to using extra secret key in this setting. Likewise, entropy is the fundamental limit of traditional secure compression. Thus, partial secrecy does not impact the compression rate in this setting, and \( \tilde{S}_{\mathcal{X}S} \) could not be improved even with a larger shared secret key.

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 \( x \in \mathcal{X} \) is mapped to a disjoint segment of the unit interval and thus \( X \) is a deterministic function of \((S,Z)\). More explicitly, we define a deterministic function \( g: \mathcal{S} \times \mathcal{Z} \to \mathcal{X} \) via
\[
g(s,z) = \arg\min_{x \in \mathcal{X}_s} \{ k_{x,s}: z \leq k_{x,s} \}.
\]
(134)
For each \( s \in \mathcal{S} \) define
\[
W_s = g(s,Z)
\]
(135)
and observe that \( W_s \) is distributed according to \( P_{X|S=s} \). Moreover, by construction, \( Z \) is a deterministic function of the set \( \{W_s\}_{s \in \mathcal{S}} \).

\[
H(Z) \leq H(\{W_s\}_{s \in \mathcal{S}})
\]
(136)
\[
\leq \sum_{s \in \mathcal{S}} H(W_s)
\]
(137)
\[
= \sum_{s \in \mathcal{S}} H(X|S = s)
\]
(138)
and this shows (45).

Lemma 3. It follows from (29) that
\[
H(Z) = H(Z|S = s) = H(Z|S = s) \quad \forall s \in \mathcal{S}
\]
(139)
and from (30) that
\[
H(X|Z, S = s) = 0 \quad \forall s \in \mathcal{S}.
\]
(140)
Using these two observations and the chain rule
\[
H(X|S = s) \leq H(X, Z|S = s)
\]
(141)
\[
= H(Z|S = s) + H(X|Z, S = s)
\]
(142)
\[
= H(Z|S = s)
\]
(143)
\[
= H(Z).
\]
(144)

Since the lower bound on \( H(Z) \) holds for all \( s \in \mathcal{S} \) it must hold for a maximum over \( \mathcal{S} \).

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 \( L \geq 1 \), and let \( \mathcal{Z} = \{1, \ldots, [L] + |\mathcal{X}|\} \). Define
\[
n_{x,s} = |P_{X|Y}(x|s)L|
\]
(145)
and note that
\[
N_s = \sum_{x \in \mathcal{X}} n_{x,s} \leq L, \quad \forall s \in \mathcal{S}.
\]
(146)
For a fixed \( s \in \mathcal{S} \) partition \( \{1, \ldots, N_s\} \) into subsets \( \mathcal{M}_x \) such that
\[
|\mathcal{M}_x| = n_{x,s}, \quad \forall x \in \mathcal{X}
\]
(147)
and construct and auxiliary \( Z' \) on \( \mathcal{Z} \):
\[
P_{Z'|Xs}(z|x,s) = \begin{cases} 
\frac{1}{L P_{X|s}(x|s)}, & z \in \mathcal{M}_x \\
\frac{1}{L} & z = [L] + x \\
0, & \text{otherwise}
\end{cases}
\]
(148)
where
\[
\delta_{x,s} = 1 - \frac{n_{x,s}}{L P_{X|s}(x|s)} \leq \frac{1}{L P_{X|s}(x|s)}.
\]
(149)
This transformation satisfies
\[
P_{Z'|s}(z|s) = \frac{1}{L}, \quad z = N = \min_{s \in \mathcal{S}} N_s
\]
(150)
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
\[
P_Z(z) = P_{Z|S}(z|s) = \frac{1}{L}, \quad z \leq N = \min_{s \in S} N_s,
\]
and (29) and (30) hold by Lemma 1.
To complete the proof we need to get a bound on $P[Z \geq N]$:
\[
P[Z \geq N] = P[Z \geq N|S = s'] = \max_{s \in S} P[Z \geq N_S|S = s].
\]
(152)

Then
\[
P[Z \geq N_S|S = s] = P[n_{X,S} = 0|P[Z \geq N_S, n_{X,S} = 0|S = s]
+ P[n_{X,S} \geq 1|P[Z \geq N_S, n_{X,S} \geq 1|S = s]
\leq P[n_{X,S} = 0|S = s] + P[Z \geq N_S, n_{X,S} \geq 1|S = s]
\leq P [I_{X|S}(X|S) > \log L|S = s]
+ E \left\{ \left[ \exp \left( I_{X|S}(X|S) - \log L \right) \right] \right\} \left| X \in F_S \right| |S = s
\]
(154)

The bound on the second term in (154) follow since
\[
P[Z \geq N_S|S = s] = P[n_{X,S} = 0|P[Z \geq N_S, n_{X,S} = 0|S = s]
\leq P[n_{X,S} = 0|S = s] + P[Z \geq N_S, n_{X,S} \geq 1|S = s]
\leq P [I_{X|S}(X|S) > \log L|S = s]
+ E \left\{ \left[ \exp \left( I_{X|S}(X|S) - \log L \right) \right] \right\} \left| X \in F_S \right| |S = s
\]
(155)

\[\text{Theorem 3. We will ensure that (29) holds by constructing } Z \text{ in such a way that}\]
\[
P[Z = z|S = s] = P[Z = z] = \frac{1}{L}
\]
(175)

Finally, (53) is obtained by loosening (52). That is, fix any
\[\tau > 0,\]
then
\[
P[I_{X|S}(X|S) > \log L|S = s]
= \left[ \exp \left( I_{X|S}(X|S) - \log L \right) \right] \left| X \in F_S \right| |S = s
\]
(160)

\[\text{Theorem 2. Fix an arbitrary } s \in \mathcal{S}, \tau > 0 \text{ and define}\]
\[
\mathcal{L} = \{x \in \mathcal{X} : \log L + \tau \leq I_{X|S}(X|s) < \infty\}
\]
(164)
\[
\mathcal{C} = \{z \in \mathcal{Z} : I_Z(z) \leq \log L\}.
\]
(165)
Suppose \( z \in \{N_s + 1, \ldots, L\} \). Then
\[
P_{Z|S}(z|s) = \sum_{x \in X} P_{X|S}(x|s) \frac{\delta_{x,s}}{L - N_s} \quad (183)
\]
\[
= \frac{1}{L - N_s} \sum_{x \in X} \left( P_{X|S}(x|s) - \frac{n_{x,s}}{L} \right) \quad (184)
\]
\[
= \frac{1}{L - N_s} \left( 1 - \frac{N_s}{L} \right) = \frac{1}{L}. \quad (185)
\]
Finally,
\[
P_{X|Z}(x|s) > 0 | S = s \leq P[Z \geq N_s | S = s] \leq \alpha(P_{X|S|S=s}, L). \quad (186)
\]
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
\[
I(X; Z) \leq I(X, S; Z) = I(S; Z) = 0. \quad (188)
\]
Suppose \( X \) is not a deterministic function of \( S \). Then, there must exist \( \tilde{x} \in X \) and \( \tilde{s} \in S \) such that \( 0 < P_{X|S}(\tilde{x}|\tilde{s}) < 1 \). Let us assume without loss of generality that \( \tilde{x} = 1 \), and let \( \tilde{s} = 1 \). The particular construction used in the proof of Lemma 1 has the property that \( P_{X|S}(\tilde{x}, \tilde{z}) > 0 \) whenever \( P_{X|S}(\tilde{x}|\tilde{s}) > 0 \). Thus, \( P_{X|Z}(\tilde{x}|\tilde{z}, \tilde{s}) > P_{X|S}(\tilde{x}|\tilde{s}) > 0 \). Moreover, \( P_{X|Z}(\tilde{x}|\tilde{z}, \tilde{s}) = 0 \) whenever \( P_{X|S}(\tilde{x}|\tilde{s}) = 0 \).

Putting this all together we obtain
\[
P_{X|Z}(\tilde{x}|\tilde{z}) = \sum_{S \in S} P_{X|S}(\tilde{x}|\tilde{z}, s) P_S(s) \quad (189)
\]
\[
\leq \sum_{S \in S} P_{X|S}(\tilde{x}|\tilde{s}) P_S(s) = P_X(\tilde{x}) \quad (190)
\]
and \( Z \) is not independent from \( X \).

Theorem 6. Let \( Z \) be the random variable guaranteed by Lemma 1. Then
\[
I(X; Z) = H(X|S) - H(X|S, Z) - I(Z; S|X) \quad (191)
\]
\[
= H(X|S) - I(Z; S|X) \quad (192)
\]
\[
= H(X|S) - H(S|X) + H(S|X, Z) \quad (193)
\]
\[
\geq H(X|S) - H(S|X) \quad (194)
\]
where (192) follows since Lemma 1 guarantees that \( X \) is a deterministic function of \( (S, Z) \). This proves the lower bound in (68).

To prove the if and only if part observe that (67) holds for any \( S \)-secure \( Z \) and so
\[
I(X; Z) \leq H(X|S) - I(Z; S|X) \quad (195)
\]
\[
= H(X|S) - H(S|X) + H(S|X, Z) \quad (196)
\]
\[
= H(X|S) - H(S|X) \quad (197)
\]
whenever \( S \) is a deterministic function of \( X \).

Finally, suppose \( S \) is not a deterministic function of \( X \). Then, there exists \( \tilde{x} \in X \) and \( \tilde{s}_1, \tilde{s}_2 \in S \) such that \( P_{X|S}(\tilde{x}|\tilde{s}_1) > 0, P_{X|S}(\tilde{x}|\tilde{s}_2) > 0, \) and \( \tilde{s}_1 \neq \tilde{s}_2 \). Assume without loss of generality that \( \tilde{x} = 1 \), and let \( \tilde{z} = 1 \). The particular construction used in the proof of Lemma 1 has the property that \( P_{S|XZ}(\tilde{s}_1, \tilde{x}) > 0 \) and \( P_{S|XZ}(\tilde{s}_2, \tilde{x}) > 0 \). This implies that \( H(S|X, Z) > 0 \) since \( S \) is not a deterministic function of \( (X, Z) \) and the inequality in (68) is strict.

\begin{lemma}
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 \( \tilde{Z} \) is \( (Z^*, S) \)-secure. Then \( Z = (Z^*, \tilde{Z}) \) is \( S \)-secure.
\end{lemma}

\begin{proof}
\[
I(S; Z) = I(S; Z^*, \tilde{Z}) \quad (198)
\]
\[
= I(S; Z^*) + I(S; \tilde{Z}|Z^*) \quad (199)
\]
\[
= H(\tilde{Z}|Z^*) - H(\tilde{Z}|Z^*, S) \quad (200)
\]
\[
= H(\tilde{Z}) - H(\tilde{Z}) = 0. \quad (201)
\]
\end{proof}

\begin{lemma}
Let \( X \) and \( S \) be two jointly distributed random variables supported on finite or countably infinite \( X \) and \( S \), respectively. Suppose \( Z \) is an optimizer for \( h_0(P_{X|S}) \), then
\[
H(X|S) = 0. \quad (202)
\]
\begin{proof}
Suppose \( Z^* \) is an optimizer for \( h_0(P_{X|S}) \) 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 \( X \leftrightarrow X \) and \( S \leftrightarrow (S, Z^*) \). According to Theorem 5
\[
I(X; \tilde{Z}) > 0. \quad (203)
\]
Let \( Z = (Z^*, \tilde{Z}) \) and observe that \( Z \) is \( S \)-secure by Lemma 4. Finally,
\[
I(S; Z) = I(X; Z^*, \tilde{Z}) \quad (204)
\]
\[
= I(X; Z^*) + I(X; \tilde{Z}|Z^*) \quad (205)
\]
\[
= I(X; Z^*) + I(X, Z^*; \tilde{Z}) \quad (206)
\]
\[
\geq I(X; Z^*) + I(X; \tilde{Z}) \quad (207)
\]
\[
> I(X; Z^*). \quad (208)
\]
This contradicts the optimality of \( Z^* \).
\end{proof}

\begin{theorem}
The upper bound follows directly from (67) and (1) \( \Rightarrow \) (2) is a trivial consequence of (73). Moreover, it is shown in [27] that computing \( g_0(P_{X|S}) \) and \( h_0(P_{X|S}) \) could be reduced to a standard linear program and therefore \( g_0(P_{X|S}) \) and \( h_0(P_{X|S}) \) are always achieved by some \( Z \).

2) \( \Rightarrow \) 3): Suppose \( h_0(P_{X|S}) = g_0(P_{X|S}) \) and let \( Z \) be an \( S \)-secure random variable that achieves \( g_0(P_{X|S}) \). 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 \( h_0(P_{X|S}) \) and by Lemma 5 it satisfies \( H(X|Z, S) = 0 \). It follows from (67) that (73) holds with equality.

3) \( \Rightarrow \) 1): Suppose \( h_0(P_{X|S}) = H(X|S) \) and let \( Z \) be an \( S \)-secure random variable that achieves \( h_0(P_{X|S}) \). Note that in
Finally, using the chain rule and Equation (209) implies that Equation (209) implies that

\[ H(Z) \geq \max_{s \in S} H(X|S = s). \]  

Proof. Equation (209) implies that

\[ H(Z) = H(Z|S) = H(Z|S = s) \quad \forall s \in S. \]  

Equation (210) implies that

\[ H(X|Z, U, S = s) = 0 \quad \forall s \in S. \]  

Finally, using the chain rule

\[ H(X, Z|U, S = s) = H(Z|U, S = s) + H(X|Z, U, S = s) \]  

\[ \leq H(Z|S = s) \]  

\[ = H(Z) \]  

and

\[ H(X|S = s) = H(X|U, S = s) \]  

\[ \leq H(X, Z|U, S = s) \]  

\[ \leq H(Z). \]  

Since (219) holds for all \( s \in S \) it must hold for a maximum over \( S \).

Theorem 9. Fix an \( f_\alpha \)-secure lossless code \((c, d)\) and let \( Z = c(X, U)\) denote the output of the compressor. The secrecy constraint (88) can be rewritten as

\[ I(Z; S) = 0 \]  

and therefore

\[ P_{Z|S}(z|s) = P_Z(z) \]  

for all \( s, z \). As Shannon observed in [1], either \( P_Z(z) = 0 \) or for every \( s \in S \) 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 \( s \in S \). It follows that \( K \geq |S| \).

Next, we show that

\[ H(Z) \geq \max \left\{ \max_{s \in S} H(X|S = s), \log |S| \right\} \]  

which is sufficient for (97) and (98) to hold. First, observe that random variables \( X, S, U \) and \( Z \) satisfy the conditions of Lemma 6 in the Appendix and so

\[ H(Z) \geq \max_{s \in S} H(X|S = s). \]  

Secondly, on the one hand, (221) implies that

\[ \sum_{s \in S} P_{Z|S}(z|s) = |S| P_Z(z). \]  

On the other hand,

\[ \sum_{s \in S} P_{Z|S}(z|s) = \sum_{s \in S} \sum_{k=1}^{K} P_{Z|US}(z|k, s) P_U(k) \]  

\[ = \sum_{k=1}^{K} P_U(k) \sum_{s \in S} P_{Z|US}(z|k, s) \leq \sum_{k=1}^{K} P_U(k) = 1 \]  

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

\[ P_Z(z) \leq \frac{1}{|S|}, \forall z \in Z \]  

and

\[ H(Z) \geq \log |S|. \]  

\( \square \)

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

\[ \Pr[I(Z) > \log L] \leq \max_{\alpha}(P_{X|S = s}, L). \]  

Recall that the existence of such \( Z \) is guaranteed by Theorem 1. We use \( \lceil \log |S| \rceil \) 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 \( \nu = L - \lceil \log |S| \rceil \), as in [36]. That is, we encode the first \( 2^{\nu} - 1 \) most likely elements of \( Z \) with strings of length \( \nu \), and the rest with an arbitrary longer strings.

Let

\[ |z: I_Z(z) \leq \log L| = M \]  

and note that \( M \leq L \). If \( M \leq 2^\nu - 1 \) then

\[ \Pr[\ell(c(X, k)) > l] \leq \Pr[I(Z) > \log L]. \]  

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

\[ \Pr[\ell(c(X, k)) > l] \leq 1 - \frac{2^\nu - 1}{L} \]  

which completes the proof of (102).
$M = 2^n$ and it is possible to represent all $z \in Z$ with strings of length $n$. Finally, (104) is obtained by loosening (103). \hfill \Box

**Theorem 11.** Fix a lossless variable-length code $(c, d)$ and let $Z = c(X, U)$ 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 $s \in S$ and $\tau > 0$. Define

$$L = \{x \in f_s^{-1}(s) : P_{X|S}(x|s) \leq 2^{-n-\tau}\},$$  
$$C = \{z \in Z : \ell(z) \leq \ell\},$$

$$C_{x,k} = \{z \in C : P_{X,Z|S}(x,z|s,k) > 0\}.$$

Note that

$$\sum_{k=1}^{K} \sum_{x \in L} |C_{x,k}| = \sum_{k=1}^{K} \sum_{x \in L} \sum_{z \in C} 1 \{z \in C_{x,k}\} = \sum_{z \in C} \sum_{x \in L} \sum_{k=1}^{K} 1 \{z \in C_{x,k}\} \leq \sum_{z \in C} 1 = |C|.$$

Equation (239) holds because the secrecy assumption and the fact that $K = |S|$ imply that $1 \{z \in C_{x,k}\} = 1$ for a unique pair $(x, k)$. That is, there is to be an element from every $f_s^{-1}(s)$ 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

$$P\left[X|S\right](X|S) \geq \nu + \tau |S = s| = P\left[X \in \mathcal{L} | S = s\right] = \sum_{k=1}^{K} \sum_{x \in L} \sum_{z \in C} 1 \{z \in C_{x,k}\} = \sum_{z \in C} \sum_{x \in L} \sum_{k=1}^{K} 1 \{z \in C_{x,k}\} \leq \sum_{z \in C} 1 = |C|.$$

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

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

messages in the following way: $N$ of the elements of $Z$ are encoded losslessly, while the remaining $Z - N$ 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 $Z$ which will be encoded losslessly, let

$$X_s = \{z : P_{X|Z}(X|Z,S)(X|Z,S) > 0 \land Y = y = 0\}.$$  

In other words, $X_s$ contains all $x \in X$ that can be reconstructed without error for a given $z \in Z$ and $s \in S$. If $|X_s| \leq N$ then all of $X_s$ is encoded losslessly (as well as some $z \in X'_s$ which we ignore). Then

$$P \left[d(c(X,k), k) \neq X\right] = P \left[\ell(c(X,k)) \neq \ell(c(X,k))\right] = P \left[|X|Z,S(X|Z,S) > 0 \land \ell(S) = s\right].$$

If $|X_s| > N$ than arbitrary $N$ elements of $X_s$ are encoded losslessly. Note that in this case $N = M'$ and we obtain

$$P \left[d(c(X,k), k) \neq X\right] \leq 1 - \frac{M'}{L}.$$

Taking the expectation over $S$ completes the proof of (113). Equation (114) follows by particularizing (113) with $M' = L$. Equation (115) follows by the same relaxation as in Theorem 10. \hfill \Box

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

$$Z = \begin{cases} c(X,U), & \text{if } \ell(c(X,U)) = \ell \in C, \\ e, & \text{otherwise} \end{cases}$$

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

$$\mathcal{L} = \{x \in f_s^{-1}(s) : P_{X|S}(x|s) \leq 2^{-n-\tau}\},$$

$$C = \{1, \ldots, M\},$$

$$C_{x,k} = \{z \in C : d(c(x,k)) = x\}.$$  

Note that

$$\sum_{k=1}^{K} \sum_{x \in L} |C_{x,k}| = \sum_{k=1}^{K} \sum_{x \in L} \sum_{z \in C} 1 \{z \in C_{x,k}\} = \sum_{z \in C} \sum_{x \in L} \sum_{k=1}^{K} 1 \{z \in C_{x,k}\} \leq \sum_{z \in C} 1 = |C|.$$  

Equation (260) holds because of the secrecy assumption and the fact that $K = |S|$. That is, there has to be an element from every $f_s^{-1}(y)$ 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

$$P \left[|X|S(X|S) \geq \log M' + \tau |S = s\right] = P \left[X \in \mathcal{L} | S = s\right].$$
for every $s^n \in S^n$ and $\gamma > 0$. Moreover,

$$\Pr \left[ |I_{X^nS^n}(X^nS^n) - nH(X|S^n)| \geq n\gamma \right] \leq 2 \exp \left( \frac{-2\gamma}{d^2n} \right)$$

for any $\gamma > 0$.

**Theorem 14.** Fix $\gamma > 0$ and let

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

Setting

$$l_n = n (\log |S| + H(X|S = s^*) + \gamma)$$

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

$$\epsilon^*_X \in S^n(l_n, |S^n|) \leq 2 \exp \left( \frac{-\gamma}{d^2n} \right) + 2^{-\frac{1}{2}n\gamma}.$$  

Indeed, for every $s^n \in S^n$

$$\Pr \left[ I_{X^nS^n}(X^nS^n) \geq nH(X|S^n = s^n) + \frac{n\gamma}{2} \right]$$

$$\leq \Pr \left[ I_{X^nS^n}(X^nS^n) \geq H(X|S^n = s^n) + \frac{n\gamma}{2} \right]$$

$$\leq 2 \exp \left( \frac{-\gamma}{d^2n} \right).$$

On the other hand, letting

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

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

$$\epsilon^*_X \in S^n(l_n, |S^n|) \geq 1 - 2 \exp \left( \frac{-\gamma}{d^2n} \right) - 2^{-\frac{1}{2}n\gamma}.$$  

Note that

$$\Pr \left[ I_{X^nS^n}(X^nS^n) < nH(X|S^n = s^*) - \frac{n\gamma}{2} \right]$$

$$\leq 2 \exp \left( \frac{-2\gamma}{d^2n} \right)$$

follows from (274) and therefore

$$\Pr \left[ I_{X^nS^n}(X^nS^n) \geq nH(X|S^n = s^n) - \frac{n\gamma}{2} \right]$$

$$\leq 1 - 2 \exp \left( \frac{-\gamma}{d^2n} \right).$$  

Putting this together we obtain

$$\lim_{n \to \infty} \frac{\epsilon^*_X \in S^n(|S^n|)}{n} \leq \log |S| + \max_{s \in S} H(X|S = s) + \gamma$$

$$+ \lim_{n \to \infty} \left( 2 \exp \left( \frac{-\gamma}{d^2n} \right) n \left( \log |X| + \log |S| \right) \right)$$

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

and

$$\lim_{n \to \infty} \frac{\epsilon^*_X \in S^n(|S^n|)}{n} \geq \log |S| + \max_{s \in S} H(X|S = s) - \gamma.$$  

for any $s^n \in S^n$. Moreover,
Noting that (286) and (287) hold for all $\gamma > 0$ gives (125). Equation (126) follows similarly by applying (275) to Theorems 12 and 13.

Next, we show a McDiarmid-like inequality for the additive case.

**Lemma 7.** Let $s_n = t_n^2(x^n)$ be given by (129) for some $x^n \in X^n$. Then for sufficiently large $n$

$$
\mathbb{P} \left[ I_{X^n|S_n} (X^n|S_n) \geq H(X^n|S = s_n) + nr|S_n = s_n \right] 
\leq 2 \exp (-A_r n)
$$

(288)

where $A_r$ is a positive constant that depends on $r$.

**Proof.** Fix $x^n \in X^n$, $s_n = t_n^2(x^n)$ and let

$$
s^n = (f_x(x_1), \ldots, f_x(x_n)).
$$

(289)

Let $u_n$ denote the type of the sequence $y^n \in S^n$ and $U_n$ be the alphabet of all such types. Note that

$$
t_{X^n|S_n} (x^n|s^n) = t_{X^n|S_n} (x^n|s^n) + t_{Y^n|U_n} (y^n|u_n)
$$

and averaging both sides with respect to $P_{X^n|S_n = s_n}$ gives

$$
H(X^n|S = s_n) \geq H(X^n|S = s^n) + H(S^n|U_n = u_n) + H(U_n|S = s_n).
$$

(291)

Therefore,

$$
\mathbb{P} \left[ I_{X^n|S_n} (X^n|S_n) \geq H(X^n|S = s_n) + nr|S_n = s^n \right] 
\leq \mathbb{P} \left[ I_{X^n|S_n} (X^n|S^n) \geq H(X^n|S = s^n) + \frac{1}{3} nr|S^n = s^n \right]
$$

$$
+ \mathbb{P} \left[ I_{S^n|U_n} (S^n|u_n) \geq H(S^n|U_n = u_n) + \frac{1}{3} nr|S^n = s^n \right]
$$

$$
+ \mathbb{P} \left[ I_{U_n|S_n} (U_n|s_n) \geq H(U_n|S = s_n) + \frac{1}{3} nr|S^n = s^n \right].
$$

(292)

By McDiarmid’s inequality

$$
\mathbb{P} \left[ I_{X^n|S^n} (X^n|S^n) \geq H(X^n|S = s^n) + \frac{1}{3} nr|S^n = s^n \right] 
\leq 2 \exp \left( -\frac{2r}{d^2} n \right).
$$

(293)

Moreover,

$$
\mathbb{P} \left[ I_{S^n|U_n} (S^n|u_n) \geq H(S^n|U_n = u_n) + \frac{1}{3} nr|S^n = s^n \right] = 0
$$

(294)

since given the type $u_n$ all sequences $s^n$ are equiprobable. Finally for $n$ sufficiently large,

$$
\mathbb{P} \left[ I_{U_n|S_n} (U_n|s_n) \geq H(U_n|S = s_n) + \frac{1}{3} nr|S^n = s^n \right] 
\leq \mathbb{P} \left[ I_{U_n|S_n} (U_n|s_n) \geq \log |U_n| + \frac{1}{6} nr|S^n = s^n \right]
$$

(295)

$$
\leq \exp \left( -\frac{r}{6} n \right).
$$

(296)

Then

$$
\mathbb{P} \left[ I_{X^n|S_n} (X^n|S_n) \geq H(X^n|S = s_n) + \frac{1}{2} nr|S_n = s_n \right] 
= \sum_{s_n} \mathbb{P}_S (s_n) \mathbb{P} \left[ I_{X^n|S_n} (X^n|S_n) \geq H(X^n|S = s_n) + \frac{1}{2} nr|S^n = s^n \right]
$$

(297)

$$
\leq 2 \exp (-A_r n).
$$

(298)

**Theorem 15.** Note that the lower bound on $S^n_{XY}$ follows immediately from Theorem 9, while the lower bound on $S^n_{XS}$ follows from the fact that $H(X)$ is the lower bound in traditional compression without any secrecy constraints. The upper bounds follow by the same sequence of steps as the proof of Theorem 14 and Lemma 7.

**References**


