Lossless compression and secrecy by design

Yanina Y. Shkel
Princeton University
Rick S. Blum
Lehigh University
H. Vincent Poor
Princeton University

Abstract—The framework of secrecy by design is introduced and the fundamental limits of lossless data compression are characterized for this setting. The main idea behind secrecy by design is to begin with an operational secrecy constraint, which is modeled by a secrecy function \( f_s \), and then to derive fundamental limits for the performance of the resulting secrecy system. In the setting of lossless compression, it is shown that strong information-theoretic secrecy guarantees can be achieved using a reduced secret key size and a modular two-part coding strategy. Focusing on the non-asymptotic and asymptotic fundamental limits of lossless compression, variable-length lossless compression and fixed-length almost-lossless compression are studied. Although it is well known that the traditional fundamental limits of variable-length and almost lossless fixed-length compression are intimately related, this relationship collapses once the secrecy constraint is incorporated.

I. INTRODUCTION

We introduce a framework for secure communication which we call secrecy by design. The idea behind secrecy by design is to identify a function of data to be transmitted that we wish to secure from an eavesdropper, and to assure absolute information theoretic secrecy for this function. Thus, secrecy by design is a way to approach partial secrecy for settings in which total secrecy is not possible. Since complete information-theoretic secrecy can only be achieved with a secret key that is as long as the message [1], developing practical and theoretically sound approaches to partial secrecy is of paramount importance.

This work focuses on secure lossless data compression. However, secrecy by design could apply to other communication problems such as lossy data compression, data transmission, and multi-terminal (network) communication and compression. Our overarching motivation is the Internet of Things (IoT) setting in which the information processing components are resource constrained and have stringent delay requirements. Thus, the primary focus of this work is on the non-asymptotic secure compression where secrecy by design interfaces seamlessly with the information spectrum techniques for traditional compression [2].

A. Shannon cipher system

In his seminal work on “Communication theory of secrecy systems”, Shannon analyzed a secrecy system, see Figure 1, in which one of \( M \) messages is transmitted over a communication channel ‘in the clear’ [1]. The legitimate transmitter and receiver have an advantage over the eavesdropper in the from of a shared secret key, which is modeled by a random variable supported on \( \{1, \ldots, K\} \). 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. Moreover, if the shared secret key is equiprobable on \( \{1, \ldots, K\} \), perfect secrecy could be achieved with a one-time pad.

This result implies that for most communication systems ensuring perfect secrecy is impractical, and thus we need to develop a notion of partial secrecy. The original notion of partial secrecy proposed in [1] is equivocation, or entropy of the source given the message. Equivocation has been used as the default measure of partial secrecy in information-theoretic security [3]–[6]; however, it has also been criticized as a secrecy measure since operational implications of equivocations are not always clear [7], [8]. Other approaches to partial secrecy—motivated by rate-distortion theory, guessing, and maximum information leakage [8]–[14]–have also been proposed. The defining feature of our approach is that we start with an explicit operational requirement for partial secrecy; moreover, once we identify the function of the data that we wish to secure, we require absolute information theoretic secrecy for this function.

B. Secrecy by design

In the setting of this paper, the information source is modeled by a random variable \( X \) defined on a finite alphabet \( \mathcal{X} \). The operational secrecy constraint is modeled by a secrecy function

\[
\forall \quad f_s : \mathcal{X} \rightarrow \mathcal{Y}.
\]  

This work was supported in part by the U.S. National Science Foundation under Grants CNS-1702555, CNS-1702808 and ECCS-1647198.
We assume without loss of generality that $f_i$ is onto $\mathcal{Y}$, and use $Y = f_i(X)$ to denote the random variable associated with the output of the secrecy function. Throughout this paper we focus on a deterministic secrecy function; however, many of our results also extend to the case when $f_i$ is a random transformation.

We get to design a compressor-decompressor pair of mappings. The output of the compressor – lets call it $Z$ – could be fixed-length, taking values on $\{1, \ldots, M\}$, or variable-length, taking values on the space of all binary strings, $\{0, 1\}^*$. It is sent ‘in the clear’; that is, the eavesdropper has perfect access to it. The primary goal of this secure data compression system is to prevent the eavesdropper from learning anything about $f_i(X)$ after observing the message.

**Definition 1.** A random variable $Z$ is $f_i$-secure with respect to $X$ if

$$I(Z; f_i(X)) = 0.$$  \hspace{1cm} (2)

We will say that a compressor is $f_i$-secure if its output is $f_i$-secure and study lossless compressors that are required to be $f_i$-secure. Note that (2) holds if and only if $f_i(X)$ and $Z$ are independent. That is, if $Z$ is $f_i$-secure, then learning the value of $Z$ does not help with estimating $f_i(X)$ regardless of what the estimation loss is.

**C. Motivating examples**

The secrecy by design setting is very general, and can model a number of different operational requirements.

1) **Meta-data:** Let $A$ be some arbitrary finite alphabet, say $A = \{0, 1\}$, and let $X = A^n$ be an $n$-fold Cartesian product of $A$. For some $k < n$ define

$$f_i: A^n \rightarrow A^k$$  \hspace{1cm} (3)

to be the function that outputs the first $k$ elements in the vector $a^n$. This corresponds to the setting in which the header of the file contains the meta-data and this is what we are most interested in protecting.

2) **N-User Process:** Let $\mathcal{X} = \mathcal{Y} \times A$ where $\mathcal{Y} = \{1, \ldots, N\}$ is the index of one of $N$ users and $A$ is an arbitrary finite alphabet over which some process in the system is taking place. For example, $x = (y, a)$ could designate a job that user $y \in \mathcal{Y}$ requested from a list of possible jobs $A$. Suppose that

$$P_X(y, a) = P_y(a)$$  \hspace{1cm} (4)

and

$$f_i(y, a) = y.$$  \hspace{1cm} (5)

That is, every user has its own probability distribution over $A$ and we would like to protect the identity of the user. Note that this is a variation on the previous example since the index of the user could be viewed as the request meta-data.

3) **Covert Communication:** Let $X = \{0, 1, \ldots, N\}$ and suppose that

$$P_X(x) = \begin{cases} 1 - \delta, & x = 0 \\ \frac{\delta}{N}, & \text{otherwise} \end{cases}$$  \hspace{1cm} (6)

for some $0 < \delta < 1$. Define the secrecy function to be

$$f_i(x) = \begin{cases} 0, & x = 0 \\ 1, & \text{otherwise}. \end{cases}$$  \hspace{1cm} (7)

That is, with probability $1 - \delta$ the compressor has nothing to say to the decompressor and so it sends $x = 0$, and the rest of the times it sends one of $N$ equiprobable messages. The goal is to keep the eavesdropper from learning if communication is taking place at all.

4) **Empirical Mean:** Let $X = A^n = \{0, 1\}^n$ be as with the Meta-data example, and let

$$f_i(a^n) = \frac{1}{n} \sum_{i=1}^{n} a_i.$$  \hspace{1cm} (8)

That is, the eavesdropper is not allowed to learn anything new about the empirical mean of the compressed data. In the present setting we assume that legitimate users and the eavesdropper know the statistics of the data. Although the eavesdropper might already have a good prior belief about the empirical mean, the goal of the secrecy system is to keep the eavesdropper from refining this belief after seeing the message.

5) **Exponential Family:** Suppose $X \sim P_\beta$ where $P_\beta$ belongs to an exponential family which is parametrized by $\Theta$ and the secrecy function outputs the sufficient statistic with respect to this family. That is

$$P_\beta(x) = \exp \left( \beta^T \sigma(x) - \psi(\beta) \right) r(x)$$  \hspace{1cm} (9)

for some $\beta \in \Theta$ and the the secrecy function is given by

$$f_i(x) = \sigma(x).$$  \hspace{1cm} (10)

Note that this example actually subsumes some of the previous examples. For instance, under the Bernoulli i.i.d. distribution the Empirical Mean example is a special case of the Exponential Family example.

**D. Partial secrecy beyond equivocation**

Finally, we highlight a number of approaches to partial secrecy and remark on their connection to secrecy by design. Cryptographic approaches to partial secrecy have been the most technologically impactful. They primarily rely on computational hardness of certain problems: for example, the celebrated RSA algorithm is based on the practical difficulty of the factoring the product of two large prime numbers. Secrecy by design does not replace such cryptographic solutions, but could be used to supplement them.

The rate-distortion approach to partial secrecy has been proposed by Yamamoto [9], and later studied by Schieler and Cuff [10], [11]. In the non-asymptotic setting, it could be possible to frame secrecy by design as a rate-distortion problem by cleverly defining a suitable distortion criterion, and demanding that the eavesdropper suffer the maximal distortion
possible. However, it is not clear how the techniques developed in [9]–[11] would hold up in the non-asymptotic setting, and with the absolute secrecy requirement. Thus, the flavor of the two lines of work is distinctly different.

In their recent works Issa et. al. [14], [15] propose a leakage measure which is defined as the worst-case refinement in knowledge of some random transformation of data. In a sense, this is similar to our approach in that it identifies securing a function of the data as the goal of partial secrecy. However, the focus of [14] and [15] is on the worst case, and the role of the function is to define the measure, not an operational constraint. Again, the philosophy of the approach is fundamentally different from ours, as is the flavor of the results.

The rest of this paper is structured as follows. In Section II we define information theoretic quantities used throughout, overview the two-part coding construction which is at the heart of our upper ( achievability) bounds, and provide decomposition lemmas which are used in conjunction with the two-part coding. In Section III we provide lower and upper bounds on average length of $f_s$-secure variable-length codes, as well as on the probability of excess length for $f_s$-secure variable-length codes. In Section IV we provide lower and upper bounds on the probability of error for $f_s$-secure fixed-length codes. We particularize our bounds for the above examples in Section V and provide asymptotic analysis of our bounds in Section VI. We end with concluding remarks in Section VII.

II. PROBLEM OVERVIEW

In this section, we introduce the information theoretic measures needed to state our compression bounds. We overview the two-part coding strategy for $f_s$-secure compression and present basic decomposition lemmas that relate to the design of $f_s$-secure codes.

A. Entropy and information

Given a random variable $Y$ jointly distributed with $X$ we define information and conditional information as

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

and

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

respectively. The entropy and conditional entropy are given by

$$H(X) = E[i_X(X)],$$

and

$$H(X|Y) = E[i_{X|Y}(X|Y)],$$

respectively. Moreover,

$$H(X|Y = y) = E[i_{X|Y}(X|y)|Y = y].$$

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 [16].

We begin with a focus on 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. Let $Y = f_s(X)$ denote the value of the secrecy function. The fundamental limits for variable-length compression derived in this work are related to two basic quantities:

$$\log |Y|$$

and

$$\max_{y \in Y} H(X|Y = y).$$

That is, (18) and (19) will be shown to be lower bounds for the rate of lossless compression; these lower bounds are independent of the rate or the distribution of the secret key available to the legitimate users.

Throughout this work we emphasize the case when the secret key is equiprobable on $\{1, \ldots, K\}$. We also take the 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 $K = |Y|$ and in that case the fundamental limits of variable-length compression are related to

$$\log |Y| + \max_{y \in Y} H(X|Y = y).$$

To be precise, it will be shown in Section VI that (20) is exactly the asymptotic fundamental limit of variable-length compression when the secrecy function is assumed to have product form. Moreover, (20) will also make an appearance in the non-asymptotic bounds in Section III: the non-asymptotic fundamental limits will depend on the worst-case (over $y \in Y$) distribution of $i_{X|Y}(X|y)$, as well as $\log |Y|$.

When $Y$ is a deterministic function of $X$, the fundamental limit of traditional compression can be written as

$$H(X) = H(Y) + H(X|Y).$$

The penalty for partial secrecy becomes readily apparent when we compare (20) with the right hand side of (21). This penalty could be negligible or severe, depending on how the secrecy function $f_s: \mathcal{X} \to \mathcal{Y}$ interacts with the data.
Finally, 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. We will see in Section IV that allowing a small probability of compression error dramatically improves the performance of secure compressors. That is, the second term in (20) will become $H(X|Y)$: an average rather than worst case over $y \in \mathcal{Y}$.

**B. Two-part codes**

Recall that Shannon showed in [1] that perfect secrecy can be achieved with one-time-pad coding; see Figure 2. Since our setting is a generalization of [1], the one-time pad coding strategy is also an ingredient in our two-part codes for variable-length and fixed-length compression.

The first stage of our coding strategy uses the shared secret key to encode $Y = f_s(X)$ with a one-time pad. The second stage reconciles the remaining uncertainty about $X$ in a secure way. This is accomplished by constructing and then compressing a new random variable $Z$ which is independent of $Y$; See Figure 3. For variable-length lossless compression we require that $X$ can be reconstructed from $(Y, Z)$ without error, while for fixed-length compression a small probability of error is allowed. Observe that the penalty for partial secrecy that appears in the two terms in (20) corresponds to the two stages of this encoding strategy. 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 $Y$ we need local randomness at the compressor; and, this is the source of the penalty observed in the second term in (20).

**C. Secure decomposition lemmas**

At the heart of our results lies the following question. Given a random variable $X$ and a secrecy function $f_s$, can we decompose $X$ into $Y = f_s(X)$ and $Z$ in such a way that $(Y, Z)$ are independent while capturing all (or most) of the information about $X$? Consider the metadata example from Section I where $X = A^n = \{0, 1\}^n$ and the secrecy function outputs the first $k$ bits of $a^n$. If $A_i$’s are independent and identically distributed (i.i.d) the answer is trivially “yes” with $Z = A_{k+1}^n$.

Now, suppose $A_i$’s are first-order Markov and that $P_{A_{i|A_{i-1}}}(1|0) = P_{A_{i|A_{i-1}}}(0|1)$. Then, the same construction no longer works since $A^k$ and $A_{k+1}^n$ are not independent. However, letting

$$Z = (A_k \oplus A_{k+1}, A_{k+1} \oplus A_{k+2}, \ldots, A_{n-1} \oplus A_n) \quad (22)$$

produces the required $Z$.

The meta-data example for i.i.d. and Markov sources is very nice in that these sources interact with the secrecy function $f_s$ in a manageable way. Suppose we are given an arbitrary $X$ with an arbitrary secrecy function $f_s$. Is it always possible to produce such a decomposition? It turns out that the answer is again “yes”, as is shown in the next lemma.

**Lemma 1.** Given a random variable $X$ and a secrecy function $f_s: \mathcal{X} \rightarrow \mathcal{Y}$ there exists a random variable $Z$ on $\mathcal{Z}$ such that

1) $Z$ is $f_s$-secure; that is

$$I(f_s(X); Z) = 0, \quad (23)$$

2) $Z$ and $f_s(X)$ are sufficient to reconstruct $X$; that is

$$H(X|Z, f_s(X)) = 0, \quad (24)$$

3) The support of $Z$ is bounded by

$$|Z| \leq |\mathcal{X}| - |\mathcal{Y}| + 1. \quad (25)$$

**Proof.** Assume for simplicity that $\mathcal{X} = \{1, \ldots, |\mathcal{X}|\}$ and that $\mathcal{X}$ is sorted according to $P_X$. That is, $x \leq \tilde{x}$ if and only if $P_X(x) \geq P_X(\tilde{x})$. Let $n_y = |f_s^{-1}(y)|$ and for each $y \in \mathcal{Y}$ define

$$k_{i,y} = F_X^{-1}(y|x_i|), \quad 1 \leq i \leq n_y \quad (26)$$

where $x_i$ is the $i$th most probable element in $f_s^{-1}(y)$ (with ties broken arbitrarily) and

$$F_X^{-1}(a|y) = \Pr[X \leq a|Y = y] \quad (27)$$

is the cumulative distribution function of $X$ given $Y$. For simplicity we define $k_{0,y} = 0$.

Then

$$Z = \bigcup_{y \in \mathcal{Y}} \bigcup_{i=1}^{n_y} \{k_{i,y}\} \quad (28)$$

is a finite subset of the unit interval and, since $k_{n_y,y} = 1$ for all $y \in \mathcal{Y}$, $|Z| \leq |\mathcal{X}| - |\mathcal{Y}| + 1$.

Let $U_{(a,\tilde{a})}$ denote a uniform random variable on $(a, \tilde{a})$. If $x_i$ is the $i$th most likely element in $f_s^{-1}(y)$, then

$$Z = \min \{z: U_{(k_{i-1,y},k_{i,y})} \leq z\} \quad (29)$$

Note, that $Z$ may take on the value $k_{i,y}$ or it may take on some other value $z \in (k_{i-1,y},k_i,y)$ if such $z \in Z$ exists.

Equation (24) holds since this construction guarantees that each $x_i \in f_s^{-1}(y)$ is mapped to a disjoint segment of the unit interval; thus knowing $y$ together with $z$ is enough to reconstruct $x \in \mathcal{X}$.

Finally, (23) holds since for all $y \in \mathcal{Y}$ and all $z \in Z$

$$F_Z^{-1}(y|z) = F_Z(z) = z, \quad (30)$$

and therefore $Z$ and $Y$ are independent. \qed
Lemma 1 immediately gives a bound on entropy of $Z$:

$$H(Z) \leq \log (|\mathcal{X}| - |\mathcal{Y}| + 1). \quad (31)$$

This bound will be tight in some cases (e.g. Covert Communication introduced in Section I), but will not be tight in general.

The next two decomposition lemmas allow for a more refined estimate of the entropy of $Z$. We define the following quantity

$$\alpha_{XY}(L, y) = \mathbb{P} [x_{|Y|}(X|y) > \log L | Y = y] + \mathbb{E} \{x_{|Y|}(X|y) < \log L \} \exp \{x_{|Y|}(X|y) - \log L \} | Y = y \} \quad (32)$$

which appears in the statements of both lemmas.

**Lemma 2.** Let $(X, Y)$ be jointly distributed random variables and fix an integer $L$. There exists $Z$ such that

1) $$I(Y; Z) = 0,$$ \quad (33)

2) $$H(X|Z, Y) = 0,$$ \quad (34)

3) $$\mathbb{P}[i(z) > \log L] \leq \max_{y \in \mathcal{Y}} \alpha_{XY}(L, y). \quad (35)$$

Note that instead of bounding the support of $Z$, as in Lemma 1, Lemma 2 gives guarantees on the spectrum of the information of $Z$. The next lemma gives the distribution of $Z$ at the expense of relaxing the exact reconstruction condition.

**Lemma 3.** Let $(X, Y)$ be jointly distributed random variables and fix an integer $L$. There exists $Z$ such that

1) $$I(Y; Z) = 0,$$ \quad (36)

2) $$\mathbb{P}[i(z > |Z|, Y) > 0 | Y = y] \leq \alpha_{XY}(L, y) \quad (37)$$

for all $y \in \mathcal{Y}$,

3) and $Z$ is equiprobable on $Z = \{1, \ldots, L\}$. That is, $$H(Z) = L. \quad (38)$$

Lemmas 2 and 3 are proven in the Appendix, and will be the technical workhorses behind our achievability bounds in the next two sections. At the same time, we will derive corresponding lower bounds on the performance of secure compressors which attest to the goodness of the two-part coding approach.

III. VARIABLE-LENGTH COMPRESSION

We begin with defining the non-asymptotic fundamental limits of secure variable-length compression. Direct upper and lower bounds for the average length non-asymptotic fundamental limit are given in Theorems 1 and 2. These bounds are not tight in general, and tighter bounds on average length can be obtained by bounding the probability of excess code length. Thus, we use information spectrum techniques to obtain upper and lower bounds on the excess length non-asymptotic fundamental limit in Theorem 3 and 4.

A. Fundamental limits

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

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

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

where no codeword in the image of $f_e$ is a prefix of any another codeword.

Let $U$ be equiprobable on $\{1, \ldots, K\}$ and independent of $X$. The variable-length code $(f_e, f_d)$ is $f_e$-secure if

$$I(f_e(X, U); f_d(X)) = 0. \quad (39)$$

An $f_e$-secure code $(f_e, f_d)$ is lossless if

$$\mathbb{P}[f_d(f_e(X, U), U) = X] = 1. \quad (40)$$

We assume that, given a fixed key value $k$, $f_e(X, k)$ 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 sting $s \in \{0, 1\}^*$. We say that $(f_e, f_d)$ is an $(l, K)$-variable-length code if

$$\mathbb{E} [\ell(f_e(X, k))] \leq l, \quad \forall k \in \{1, \ldots, K\}. \quad (41)$$

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

$$\mathbb{P}[\ell(f_e(X, k)) > l] \leq \epsilon, \quad \forall k \in \{1, \ldots, K\}. \quad (42)$$
The non-asymptotic fundamental limits of lossless compression for the secrecy function \(f_s\) are given by
\[
\ell_s^*(K) = \inf \{ l : \exists f_s\text{-secure } (l, K)\text{-code} \}
\]
and
\[
\epsilon_s^*(l, K) = \inf \{ \epsilon : \exists f_s\text{-secure } (l, K, \epsilon)\text{-code} \},
\]
where \(Y = f_s(X)\) denotes the value of the secrecy function.

**Remark 1.** We could also define the non-asymptotic fundamental limits by changing (41) and (42) to
\[
E[\ell(f_c(X, U))] \leq l
\]
and
\[
P[\ell(f_c(X, U)) > l] \leq \epsilon.
\]
This is a less strict requirement since performance guarantees are averaged over all of the secret key values. It turns out that all of our lower bounds also hold for these less strict requirements.

**B. Bounds on average length**

We begin with a simple upper bound on the average length for secure variable-length codes.

**Theorem 1.**

\[
\ell_s^*(|\mathcal{Y}|) \leq [\log (|\mathcal{X}| - |\mathcal{Y}| + 1)] + [\log |\mathcal{Y}|] + 1
\]

**Proof.** The theorem follows from the two-part coding construction outlined in Section II and Lemma 1. \(\square\)

The next theorem identifies two fundamental lower bounds on the performance of secure variable-length codes.

**Theorem 2.** For any \(K \geq |\mathcal{Y}|\),
\[
\ell_s^*(K) \geq \max \left\{ \max_{y \in \mathcal{Y}} H(X|Y = y), \log |\mathcal{Y}| \right\}
\]
and \(f_s\)-secure codes do not exist for \(K < |\mathcal{Y}|\).

**Proof.** Fix an \(f_s\)-secure lossless code \((f_s, f_d)\) and let \(Z = f_s(X, U)\) denote the output of the compressor. The secrecy constraint (39) can be rewritten as
\[
I(Z; Y) = 0
\]
and therefore
\[
P_{Z|Y}(z|y) = P_Z(z)
\]
for all \(y, z\). As Shannon observed in [1], either \(P_Z(z) = 0\) or for every \(y \in \mathcal{Y}\) there needs to be a \(k \in \{1, \ldots, K\}\) such that \(P_{Z|U|Y}(z|k, y) = 0\). But, due to the lossless constraint, \(P_{Z|U|Y}(z|k, y) > 0\) for at most one \(y \in \mathcal{Y}\). It follows that \(K \geq |\mathcal{Y}|\).

Next, we show that
\[
H(Z) \geq \max \left\{ \max_{y \in \mathcal{Y}} H(X|Y = y), \log |\mathcal{Y}| \right\}
\]
which is sufficient for (47) to hold. First, observe that random variables \(X, Y, U\) and \(Z\) satisfy the conditions of Lemma 4 in the Appendix and so
\[
H(Z) \geq \max_{y \in \mathcal{Y}} H(X|Y = y).
\]

Secondly, on the one hand, (49) implies that
\[
\sum_{y \in \mathcal{Y}} P_{Z|Y}(z|y) = |\mathcal{Y}| P_Z(z).
\]

On the other hand,
\[
\sum_{y \in \mathcal{Y}} P_{Z|Y}(z|y) = \sum_{k=1}^{K} \sum_{y \in \mathcal{Y}} P_{Z|U|Y}(z|k, y) P_U(k)
\]
\[
= \sum_{k=1}^{K} P_U(k) \sum_{y \in \mathcal{Y}} P_{Z|U|Y}(z|k, y)
\]
\[
\leq \sum_{k=1}^{K} P_U(k) = 1
\]

since for a lossless code only one \(y\) can encode to \(z\) for any \(k\). Thus, we get
\[
P_Z(z) \leq \frac{1}{|\mathcal{Y}|}, \quad \forall z \in \mathcal{Z}
\]
and
\[
H(Z) \geq \log |\mathcal{Y}|.
\]

**Remark 2.** Although our primary focus is on the case \(K = |\mathcal{Y}|\), Theorem 2 holds for an arbitrary value of \(K\). Note also that the proof of Theorem 2 does not use the fact that \(U\) is equiprobable on \(\{1, \ldots, K\}\) or that \(Y\) is a deterministic function of \(X\). Thus, these lower bounds are fundamental to a more general secure lossless compression setting.

Theorems 1 adnd 2 are not tight in general, but they do give good bounds for some cases. We can get more insightful bounds on the average length fundamental limit (43) by leveraging bounds on the probability of excess length fundamental limit (44). That is, we can relate the two fundamental limits by
\[
\max_{l \in \{0, 1, \ldots\}} \ell_s^*(l, K) \leq \ell_s^*(K) \leq \min_{l \in \{0, 1, \ldots\}} \left( l + \kappa \epsilon_s^*(l, K) \right)
\]
where
\[
\kappa = \log \left( |\mathcal{X}| - K + 1 \right) + \lceil \log K \rceil + 1.
\]
Upper and lower bounds on \(\epsilon_s^*(l, K)\) are developed next.
C. Bounds on excess length

**Theorem 3.** Let \( K = |\mathcal{Y}| \), then for any \( L \in \{1, 2, \ldots \} \)
\[
\epsilon_{XY}(l, K) \leq \max \left\{ \frac{\alpha_{XY}(L, y)}{L} \right\} - \frac{2^\nu - 1}{L} \tag{60}
\]
where \( \nu = l - \lceil \log K \rceil \). Moreover,
\[
\epsilon_{XY}(l, K) \leq \max \left\{ \alpha_{XY}(2^\nu) \right\} \leq \max \min \left\{ \mathbb{P} \left[ \mathbb{P} \left[ |X| \left( |X| \left( \geq \nu \right) \right) \right] \right\} + 2^{-\tau} \tag{62}
\]
where \( \nu = l - \lceil \log K \rceil \).

**Proof.** We use a two-part code described in Section II by representing \( X \) as \( (Y, Z) \) where \( I(Y; Z) = 0 \), \( H(X|Z, Y) = 0 \) and
\[
\mathbb{P} \left[ |Z| > \log L \right] \leq \max \left\{ \alpha_{XY}(L, y) \right\} \tag{63}
\]
Recall that the existence of such \( Z \) is guaranteed by Lemma 2. We use \( \lceil \log K \rceil \) bits to encode \( Y \) with a one-time-pad. \( Z \) is encoded with a variable-length code that minimizes the probability or exceeding code length \( \nu = l - \lceil \log K \rceil \), as in [2]. That is, we encode the first \( 2^\nu \) most likely elements of \( Z \) with strings of length \( \nu \), and the rest with an arbitrary longer strings.

Let
\[
|z| \geq l \geq 0 \leq L \tag{64}
\]
and note that \( M \leq L \). If \( M \leq 2^\nu - 1 \) then
\[
\mathbb{P} \left[ \ell \left( f_\nu(X, k) \right) > l \right] \leq \mathbb{P} \left[ |Z| > \log L \right] \tag{65}
\]
Otherwise, \( P_Z(z) \geq \frac{1}{2} \) for the \( 2^\nu - 1 \) most probable \( z \in Z \) and we obtain
\[
\mathbb{P} \left[ \ell \left( f_\nu(X, k) \right) > l \right] \leq 1 - \frac{2^\nu - 1}{L} \tag{66}
\]
which completes the proof of (60).

Equation (61) is shown by taking \( L = 2^\nu \) and noting that \( M \geq 2^\nu - 1 \) only if \( P_Z(z) = \frac{1}{2} \) for all \( z \in Z \). In that case, \( M = 2^\nu \) and it is possible to represent all \( z \in Z \) with strings of length \( \nu \).

Finally, (62) is obtained by loosening (61). That is, fix any \( \tau > 0 \), then
\[
\mathbb{P} \left[ \alpha_{XY}(X|Y) > \nu \right] \geq \nu \|Y\| \geq \nu \mathbb{P} \left[ \mathbb{E} \left\{ \mathcal{I} \text{I}_{XY}(X|Y) < \nu \right\} \exp \mathcal{I}_{XY}(X|Y) - \nu \right] |Y = y] \tag{67}
\]
\[
\leq \mathbb{P} \left[ \alpha_{XY}(X|Y) > \nu \right] \geq \nu \|Y\| \geq \nu \mathbb{P} \left[ \mathbb{E} \left\{ \mathcal{I} \text{I}_{XY}(X|Y) < \nu \right\} \exp \mathcal{I}_{XY}(X|Y) - \nu \right] |Y = y] \tag{68}
\]
\[
\leq \mathbb{P} \left[ \alpha_{XY}(X|Y) > \nu \right] \geq \nu \|Y\| \geq \nu \mathbb{P} \left[ \mathbb{E} \left\{ \mathcal{I} \text{I}_{XY}(X|Y) < \nu \right\} \exp \mathcal{I}_{XY}(X|Y) - \nu \right] |Y = y] \tag{69}
\]
\[
\leq \mathbb{P} \left[ \alpha_{XY}(X|Y) > \nu \right] \geq \nu \|Y\| \geq \nu \mathbb{P} \left[ \mathbb{E} \left\{ \mathcal{I} \text{I}_{XY}(X|Y) < \nu \right\} \exp \mathcal{I}_{XY}(X|Y) - \nu \right] |Y = y] \tag{70}
\]
Note the correspondence between (62) and the lower bound in the next Theorem.

**Theorem 4.** Let \( K = |\mathcal{Y}| \), then
\[
\max_{\tau > 0, \nu \in \mathcal{Y}} \left\{ \mathbb{P} \left[ \alpha_{XY}(X|Y) \geq \nu \right] |Y| = y \} - 2^{-\tau} \right\} \leq \epsilon_{XY}(l, K) \tag{71}
\]
where \( \nu = \ell - \lceil \log K \rceil \).

**Proof.** Fix a lossless variable-length code \( (f_\nu, f_\nu) \) and let \( Z = f_\nu(X, U) \) 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 \( y \in \mathcal{Y} \) and \( \tau > 0 \). Define
\[
\mathcal{L} = \{x \in f_\nu^{-1}(y) : P_{X|Y}(x|y) \leq 2^{-\nu - \tau} \}, \tag{72}
\]
\[
\mathcal{C} = \{z \in \mathcal{Z} : \ell(z) \leq l\}, \tag{73}
\]
\[
\mathcal{C}_{x, k} = \{z \in \mathcal{C} : P_{X|Y}(x, z,k) > 0\}. \tag{74}
\]
Note that
\[
\sum_{k=1}^{K} \sum_{x \in \mathcal{L}} |\mathcal{C}_{x, k}| = \sum_{k=1}^{K} \sum_{x \in \mathcal{C}} \sum_{y \in \mathcal{Y}} 1 \{\mathcal{z} \in \mathcal{C}_{x, k}\} \tag{75}
\]
\[
= \sum_{k=1}^{K} \sum_{x \in \mathcal{C}} \sum_{y \in \mathcal{Y}} 1 \{\mathcal{z} \in \mathcal{C}_{x, k}\} \tag{76}
\]
\[
\leq \sum_{y \in \mathcal{Y}} 1 = |C|. \tag{77}
\]
Equation (77) holds because the secrecy assumption and the fact that \( K = |\mathcal{Y}| \) imply that \( 1 \{\mathcal{z} \in \mathcal{C}_{x, k}\} = 1 \) for a unique pair \((x, k)\). That is, there has to be an element from every set \( f_\nu^{-1}(y) \) mapping to each \( z \), thus an element from \( \mathcal{C} \) maps to \( z \) for only one value of \( k \). Because the compressor is lossless, at most one element of \( \mathcal{C} \) maps to \( z \). Then
\[
\mathbb{P} \left[ \alpha_{XY}(X|Y) \geq \nu \right] |Y| = y \} = \mathbb{P} \left[ X \in \mathcal{L} |Y = y \right] \tag{78}
\]
\[
= \mathbb{P} \left[ (X, Z) \in \mathcal{L} \times \mathcal{C} |Y = y \right] + \mathbb{P} \left[ (X, Z) \in \mathcal{L} \times \mathcal{C}^c |Y = y \right] \tag{79}
\]
\[
\leq \mathbb{P} \left[ (X, Z) \in \mathcal{L} \times \mathcal{C} |Y = y \right] + \mathbb{P} \left[ Z \in \mathcal{C}^c |Y = y \right] \tag{80}
\]
\[
\leq \mathbb{P} \left[ U = k \right] \mathbb{P} \left[ (X, Z) \in \mathcal{L} \times \mathcal{C} |Y = y, U = k \right] \tag{81}
\]
\[
+ \mathbb{P} \left[ Z \in \mathcal{C}^c |Y = y \right] \tag{82}
\]
\[
\leq \mathbb{P} \left[ (X, Z) \in \mathcal{L} \times \mathcal{C} |Y = y \right] \tag{83}
\]
\[
\leq \frac{1}{K} \mathbb{P} \left[ (X, Z) \in \mathcal{L} \times \mathcal{C} |Y = y \right] \tag{84}
\]
\[
= 2^{-\tau} + \mathbb{P} \left[ Z \in \mathcal{C}^c |Y = y \right] \tag{85}
\]
\[
= 2^{-\tau} + \mathbb{P} \left[ \ell \left( f_\nu(X, U) \right) \geq l \right] \tag{86}
\]
where (85) follows because \( Z \) and \( Y \) are independent.

\[ \square \]
Taking $|\mathcal{Y}| = 1$ recovers traditional compression. Particularizing Theorem 4 to this case we obtain
\[
\max_{\tau > 0} \left\{ \mathbb{P} \left[ \ell_{X}(X) \geq l + \tau - 2^{-\tau} \right] \leq \epsilon_{XY}(l, 1) \right\} \tag{87}
\]
which is exactly [2, Theorem 4].

IV. FIXED-LENGTH COMPRESSION

In this section we derive bounds on $f_{\epsilon}$-secure almost-lossless fixed-length compression. We will see that the worst case dependence on the value of the secrecy function, $y \in \mathcal{Y}$, that we observed in the variable-length setting will turn into average case dependence. Thus, allowing a small probability of error for the compression problem allows for potentially large gains in the rate of secure compression.

A. Fundamental limits

A fixed-length source code with $M$ codewords and a secret key of size $K$ is a pair of mappings:

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

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

An $f_{\epsilon}$-secure code $(f_{\epsilon}, f_{d})$ (c.f. (39)) is an $(M, K, \epsilon)$-almost-lossless code if
\[
\mathbb{P} \left[ f_{d} (f_{\epsilon}(X, k), k) \in f_{\epsilon}^{-1}(Y) \right] = 1 \quad \text{and} \quad \mathbb{P} \left[ f_{d} (f_{\epsilon}(X, k), k) \neq X \right] \leq \epsilon, \quad \forall k \in \{1, \ldots, K\}
\]
(88)
where $y = f_{\epsilon}(X)$ denotes the value of the secrecy function. In other words, we require that the value of a secrecy function is reconstructed exactly, but otherwise allow for a small probability of error. We assume that $f_{\epsilon}(x, k)$ 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
\[
\tilde{\epsilon}_{XY}(M, K) = \inf \{ \epsilon : \exists f_{\epsilon}$-secure $(M, K, \epsilon)$-code \} \tag{90}
\]
and
\[
\tilde{M}_{XY}(\epsilon, K) = \inf \{ M : \exists f_{\epsilon}$-secure $(M, K, \epsilon)$-code \} \tag{91}
\]
In this section we focus on bounding $\tilde{\epsilon}_{XY}(M, K)$, but the presented results could be used to bound $\tilde{M}_{XY}(\epsilon, K)$.

Remark 3. There are several meaningful ways to define secure fixed-length codes. For example, we could drop the requirement (88) and allow the decompressor to also make an error when reconstructing the value of the secrecy function. This would lead to slightly different fundamental limits; The log $|\mathcal{Y}|$ term will behave like $H(Y)$. In the present paper we choose to define secure fixed-length codes this way for a number of reasons. First, if a certain function of the data is important enough to be made secret, it should be important enough to be reconstructed losslessly. Secondly, this approach leads to the simplest exposition of our results, while still letting us compare almost lossless fixed-length coding with lossless variable-length coding under partial secrecy constraints.

B. Bounds on $\tilde{\epsilon}_{XY}(M, K)$

We can prove the following upper bound.

Theorem 5. Let $K = |\mathcal{Y}|$, then
\[
\tilde{\epsilon}_{XY}(M, K) \leq \min_{L \in \{1, 2, \ldots\}} \mathbb{E} \left[ \max \{ \alpha_{XY}(L, Y) - 1 + \frac{M'}{L} \} \right]
\]
(92)
where $M' = \left\lfloor \frac{M}{K} \right\rfloor$.

Proof. We use a two-part code described in Section II by representing $X$ as $(Y, Z)$ where $I(Y; Z) = 0$,
\[
\mathbb{P} \left[ f_{d} (f_{\epsilon}(X, k), k) \neq f_{\epsilon}(X) \right] \leq \epsilon, \quad \forall k \in \{1, \ldots, K\}
\]
(93)
and $Z$ is equiprobable on $Z = \{1, \ldots, L\}$. Recall that the existence of such $Z$ is guaranteed by Lemma 3.

We use $|\log K|$ bits to encode $Y$ with a one-time-pad. For a fixed $y \in \mathcal{Y}$, $Z$ is encoded with a fixed-length code that uses
\[
N = \min \{ M', L \}
\]
(94)
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
\[
S_{y} = \{ z : \mathbb{P} \left[ f_{d}(X, k) \neq f_{\epsilon}(X) \right] = 0 \}
\]
(97)
In other words, $S_{y}$ contains all $z \in \mathcal{X}$ that can be reconstructed without error for a given $z \in Z$ and $y \in \mathcal{Y}$.

If $|S_{y}| \leq N$ then all of $S_{y}$ is encoded losslessly (as well as some $z \in S_{y}$ which we ignore). Then
\[
\mathbb{P} \left[ f_{d} (f_{\epsilon}(X, k), k) \neq f_{\epsilon}(X) \right] \leq \mathbb{P} \left[ f_{d} (f_{\epsilon}(X, k), k) \neq f_{\epsilon}(X) \right] \leq \mathbb{P} \left[ f_{d} (f_{\epsilon}(X, k), k) \neq f_{\epsilon}(X) \right]
\]
(98)
which is exactly [2, Theorem 4].

Remark 4. There are several meaningful ways to define secure fixed-length codes. For example, we could drop the requirement (88) and allow the decompressor to also make an error when reconstructing the value of the secrecy function. This would lead to slightly different fundamental limits; The log $|\mathcal{Y}|$ term will behave like $H(Y)$. In the present paper we choose to define secure fixed-length codes this way for a number of reasons. First, if a certain function of the data is important enough to be made secret, it should be important enough to be reconstructed losslessly. Secondly, this approach leads to the simplest exposition of our results, while still letting us compare almost lossless fixed-length coding with lossless variable-length coding under partial secrecy constraints.
where $M' = \frac{M}{K}$.

**Proof.** Fix a fixed-length code $(f_x, f_y)$ and let $Z = \{1, \ldots, M, e\}$. Define a random variable

$$Z = \begin{cases} f_x(X, U), & f_y(f_x(X, U)) = X \\ e, & \text{otherwise} \end{cases}$$

(102)

Fix an arbitrary $y \in \mathcal{Y}$ and $\tau > 0$. Define

$$\mathcal{L} = \{ x \in f_x^{-1}(y) : P_{X|Y}(x|y) \leq 2^{-\log M' - \tau} \},$$

(103)

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

(104)

$$\mathcal{C}_{x,k} = \{ z \in \mathcal{C} : f_y(f_x(x, k)) = x \}.$$  

(105)

Note that

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

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

$$\leq \sum_{z \in \mathcal{C}} \sum_{x \in \mathcal{L}} 1 = |\mathcal{C}|,$$

(106)

(107)

(108)

Equation (108) holds because of the secrecy assumption and the fact that $K = |\mathcal{Y}|$. That is, there has to be an element from every $f_x^{-1}(y)$ 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

$$\mathbb{P}[f_{X|Y}(X|Y) \geq \log M' + \tau | Y = y] = \mathbb{P}[X \in \mathcal{L} | Y = y]$$

$$= \mathbb{P}[(X, Z) \in \mathcal{L} \times \mathcal{C} | Y = y] + \mathbb{P}[(X, Z) \in \mathcal{L} \times \mathcal{C}^c | Y = y]$$

$$\leq \mathbb{P}[(X, Z) \in \mathcal{L} \times \mathcal{C} | Y = y] + \mathbb{P}[Z \in \mathcal{C}^c | Y = y]$$

$$\leq \sum_{k=1}^{K} \mathbb{P}(U = k) \mathbb{P}[(X, Z) \in \mathcal{L} \times \mathcal{C} | Y = y, U = k]$$

$$+ \mathbb{P}[Z \in \mathcal{C}^c | Y = y]$$

$$= \frac{1}{K} \sum_{k=1}^{K} \mathbb{P}(X, Z) \in \mathcal{L} \times \mathcal{C} | Y = y, U = k]$$

$$+ \mathbb{P}[Z \in \mathcal{C}^c | Y = y]$$

$$\leq \frac{1}{K} \sum_{k=1}^{K} \sum_{x \in \mathcal{L}} 2^{-\log M' - \tau} |\mathcal{C}_{x,k}| + \mathbb{P}[Z \in \mathcal{C}^c | Y = y]$$

$$\leq 2\frac{2^{|Y|} \log M' - \tau}{K} + \mathbb{P}[Z \in \mathcal{C}^c | Y = y].$$

(109)

(110)

(111)

(112)

(113)

(114)

(115)

The second term in (115) is exactly the conditional probability of error. Averaging over all $y$ gives the desired result.

Taking $|\mathcal{Y}| = 1$ recovers traditional fixed-length compression. Particularizing Theorem 6 to this case yields

$$\max_{\tau > 0} \left( \mathbb{P}[X(X) \geq \log M + \tau] - 2^{-\tau} \right) \leq \tilde{e}_{XY}(M, 1).$$  

(116)

Since in the traditional compression setting the variable-length and fixed-length fundamental limits are intimately related (116) is exactly [2, Theorem 4].

C. Fixed-length and variable-length compression

It is observed in [2] that for traditional compression the probability of error for fixed-length coding and the probability of excess length in variable-length coding are related. However, once we introduce the partial secrecy constraint this relationship breaks down. Indeed, compare Theorems 3 and 4 with Theorems 5 and 6: 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.

V. EXAMPLES

In this section we evaluate the bounds from Sections III and IV for the Covert Communication, Exponential Family, and Empirical Mean examples.

A. Covert Communication

Recall that in this setting $\mathcal{X} = \{0, 1, \ldots, N\}$ and

$$P_X(x) = \begin{cases} 1 - \frac{\delta}{N}, & x = 0 \\ \frac{\delta}{N}, & \text{otherwise} \end{cases}$$

(117)

for some $0 < \delta < 1$. The secrecy function is given by

$$f_x(x) = \begin{cases} 0, & x = 0 \\ 1, & \text{otherwise}. \end{cases}$$

(118)

For secure variable-length coding the average length fundamental limit can be bounded directly:

$$\log N \leq \ell_{XY}^*(2) \leq \log N + 1$$

(119)

The lower bound in (119) is obtained from Theorem 2 and by noting that $H(X|Y = 1) = \log N$, while the upper bound is obtained from Theorem 1.

The excess length fundamental limit is

$$1 - \frac{2^{l-1}}{N} \leq \alpha_{XY}(l, 2) \leq 1 - \frac{2^{l-1}}{N} + 1 - \frac{1}{N}.$$  

(120)

The lower bound in (120) is obtained by applying Theorem 4 with $\tau = (1 - l) + \log N$. The upper bound follows from particularizing (60) in Theorem 3 to $L = N$ and noting that

$$\alpha_{XY}(N, 0) = \frac{1}{N}, \quad \text{and} \quad \alpha_{XY}(N, 1) = 0.$$  

(121)

For this example we can easily find the average length fundamental limit for various values of $K$. Letting $K \leq N + 1$ and assuming that $N$ is divisible by $K - 1$, we see by inspection that

$$\ell_{XY}^*(K) = \log N + \log \frac{K}{K-1}.$$  

(122)

By letting $K = N + 1$ it is possible to achieve an average length of $\log(n + 1) = \log |\mathcal{X}|$. Compare this with $\log n + 1$ when $K = 2$, which is the smallest secret key size at which a secure code exists.
Notice that all of the bounds for variable-length secure compression are independent of the parameter $\delta$. Indeed, the fundamental limits depend on the worst case distribution $P_X|Y(\cdot|y)$ which is independent of $\delta$. In the fixed-length setting the fundamental limits depend on the average case behavior of the distribution $P_X|Y(\cdot|y)$ and the parameter $\delta$. Namely, the fixed-length fundamental limit is bounded by

$$\delta \left(1 - \frac{M}{2N} \right) \leq \tilde{c}_{XY}(M, 2) \leq \delta \left(1 - \frac{M}{2N} \right) + \frac{1}{N} \quad (123)$$

where the upper and lower bounds are obtained from Theorems 5 and 6 in a similar fashion to (120).

### B. Exponential Family

We assume that $X \sim P_\beta$ and that $P_\beta$ belongs to an exponential family indexed by $\Theta$; that is, it can be written as

$$P_\beta(x) = \exp \left(\beta^T \sigma(x) - \psi(\beta)\right) r(x) \quad (124)$$

for some $\beta \in \Theta$. The secrecy function $f(x)$ is the sufficient statistic,

$$f(x) = \sigma(x), \quad (125)$$

and assume that for all $y \in Y$

$$r(x) = c_y, \forall x \in f^{-1}_s(y) \quad (126)$$

for some collection of constants $c_y$. Note that this is satisfied for a number of exponential families: for example, Bernoulli, Categorical, Poisson, as well as the exponential families resulting from taking $n$ i.i.d. draws from each of these distributions.

In this case all $x \in f^{-1}_s(y)$ are equiprobable. Let $M_y = |f^{-1}_s(y)|$ and observe that

$$\alpha_{XY}(L, y) = \left\{ \frac{M_y}{2}, M_y < L, 0, M_y = L, 1, M_y > L. \right\} \quad (127)$$

For secure variable-length coding the average length fundamental limit can be lower bounded using Theorem 2:

$$\log M_y \leq \ell_{XY}(K) \quad (128)$$

where

$$y^* = \arg \max_{y \in Y} M_y. \quad (129)$$

The excess length fundamental limit is bounded by

$$1 - \frac{2^l}{K M_y} \leq \tilde{c}_{XY}(l, K) \leq \frac{K' M_y}{2^l} \quad (130)$$

where $K = |\mathcal{Y}|$ and $K' = 2^{[\log K]}$. The lower bound in (130) is obtained by applying Theorem 4 with $\tau = \log M_y - l + \log K$. The upper bound follows from (61) in Theorem 3.

We can further tighten the upper bound in (130) by applying (60) in Theorem 3. Indeed, define

$$y_* = \arg \max_{y \in \mathcal{Y}, y \neq y^*} M_y. \quad (131)$$

Letting $L = M_{y^*}$ we obtain

$$\ell_{XY}(l, K) \leq \max \left\{ \frac{M_{y^*}}{M_y}, 1 - \frac{2^l}{K' M_{y^*}} + \frac{1}{M_y} \right\} \quad (132)$$

where $K = |\mathcal{Y}|$ and $K' = 2^{[\log K]}$. Note the nice correspondence between the second term in (132) and the lower bound in (130). In the cases when $\frac{M_{y^*}}{M_y}$ is the active term in (132) we can further optimize the upper bound (60) over $L > M_{y^*}$.

Finally, the fixed-length fundamental limit can be obtained in a similar fashion:

$$1 - \frac{M}{K} \mathbb{E} \left[ \frac{1}{M_y} \right] \leq \tilde{c}_{XY}(M, K) \quad (133)$$

and

$$\tilde{c}_{XY}(M, K) \leq \mathbb{E} \left[ \frac{M_y}{M_y} \mathbb{1}\{M_y < M'\} \right] + \mathbb{P} \{M_y > M'\} \quad (134)$$

where $K = |\mathcal{Y}|$ and $M' = \left[ \frac{M}{K} \right]$.

### C. Empirical Mean

Let $\mathcal{X} = \{0, 1\}^n$ and

$$f_s(y^n) = \frac{1}{n} \sum_{i=1}^n y_i. \quad (135)$$

If $X^n$ is i.i.d. Bernoulli($p$) then this is an exponential family and the bounds from the previous example immediately apply. Moreover, from Theorems 1 and 2:

$$\max \left\{ \frac{\log \binom{n}{k}}{k}, \log(n) \right\} \leq \ell_{XY}(n + 1) \quad (136)$$

$$\leq \log(2^n - n) + \log(n + 1). \quad (137)$$

It is easy to check that

$$\lim_{n \to \infty} \frac{\ell_{XY}(n + 1)}{n} = 1, \quad (138)$$

and asymptotically the compression of the 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.

### VI. ASYMPTOTIC FUNDAMENTAL LIMITS

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 separable; 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.
A. The product secrecy function
Let \( X \sim P_X \) and \( Y = f_s(X) \) where \( f_s: \mathcal{X} \to \mathcal{Y} \).

(139)

Given a single letter secrecy function (139), define an \( n \)-letter secrecy function

\[
\ell^n_s: \mathcal{X}^n \to \mathcal{Y}^n
\]

(140)

via

\[
\ell^n_s(x^n) = (f_s(x_1), \ldots, f_s(x_n)).
\]

(141)

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

\[
\begin{align*}
R^*_{XY} &= \lim_{n \to \infty} \frac{\ell^n_s(x^n)(\mathcal{Y}^n)}{n} \quad \text{for the separable setting.}
\end{align*}
\]

(142)

The asymptotic fixed-length fundamental limit is given by

\[
\tilde{R}_{XY} = \lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log \tilde{M}_{X^nY^n}(\epsilon, |\mathcal{Y}^n|). 
\]

(143)

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

**Theorem 7.**

\[
\begin{align*}
R^*_{XY} &= \log |\mathcal{Y}| + \max_{y \in \mathcal{Y}} H(X|Y = y) \\
\tilde{R}_{XY} &= \log |\mathcal{Y}| + H(X|Y).
\end{align*}
\]

(144) and (145)

Equation (144) is a direct consequence of Theorems 3 and 4, while (145) is a direct consequence of Theorems 5 and 6.

A detailed proof of Theorem 7 is given in the Appendix.

B. The separable secrecy function

Let \( X \sim P_X \) and assume

\[
f_s: \mathcal{X} \to \mathcal{Y} \subset \mathbb{R}.
\]

(146)

Given a single letter secrecy function (146), define an \( n \)-letter secrecy function

\[
\ell^n_s: \mathcal{X}^n \to \mathcal{Y}^n
\]

(147)

via

\[
\ell^n_s(x^n) = \frac{1}{n} \sum_{i=1}^{n} f_s(x_i).
\]

(148)

We assume that \( X^n \) is i.i.d. according to \( P_X \), and \( Y_n = \ell^n_s(X^n) \). The asymptotic variable-length fundamental limit is given by

\[
\mathcal{S}^*_{XY} = \lim_{n \to \infty} \frac{\ell^n_s(x^n)(\mathcal{Y}^n)}{n}.
\]

(149)

The asymptotic fixed-length fundamental limit is given by

\[
\tilde{S}_{XY} = \lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log \tilde{M}_{X^nY^n}(\epsilon, |\mathcal{Y}^n|).
\]

(150)

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

**Theorem 8.**

\[
\begin{align*}
S^*_{XY} &= \max_{y \in \mathcal{Y}} H(X|Y = y) \\
\tilde{S}_{XY} &= H(X).
\end{align*}
\]

(151) and (152)

A detailed proof of Theorem 8 is given in the Appendix. Comparing the separable setting to the product setting we see that the dependence on \( \log |\mathcal{Y}| \) terms disappears since \( |\mathcal{Y}| \) grows logarithmically in \( n \).

Note also that according to Theorem 2, \( S^*_{XY} \) can be lower bounded by \( \max_{y \in \mathcal{Y}} H(X|Y = y) \) 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 \( \tilde{S}_{XY} \) could not be improved even with a larger shared secret key.

VII. DISCUSSION

A. Contributions

In this paper 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 operational characterization of partial secrecy and to construct the compression system from the ground up to satisfy this requirement. We have applied the secrecy by design framework to the problem of lossless compression and have derived non-asymptotic and asymptotic fundamental limits on the rate of secure compression.

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. While in some cases it is possible to have partial secrecy with a small penalty on the rate of compression, in other cases (e.g. Covert Communication) the word “compression” is actually a misnomer: in order to achieve partial secrecy the source representation is actually inflated.

Our upper (achievability) bounds rely on the so-called two-part coding strategy. First, the source is decomposed into two parts: the first part is the secrecy function of the data that is secured with a one-time-pad, and the second part is designed to be independent of the secrecy function which means that it could be compressed with any traditional compression strategy. Our first significant technical contribution has been to show how to construct such decompositions. Our second significant technical contribution has been to derive matching lower bounds that attest to the goodness of the two-part coding strategy.
B. Future work

The secrecy by design framework could be applied to other problems in communication 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. When it comes to lossless compression, there are many open questions which merit a more detailed study.

1) Key slack and key distribution: In the present paper we focus on the shared secret key that is equiprobable on \{1, \ldots, |Y|\}, where \( Y \) is the range of the secrecy function. A notable exception to this assumption is Theorem 2 which holds for an arbitrary secret key size and distribution. Because of Theorem 2 we know that in some settings (c.f. Theorem 8) a larger secret key would not help very much. However, in other settings (c.f. Theorem 7) there is a significant gap between the best attainable compression rate with the shared secret key that is equiprobable on \{1, \ldots, |Y|\} and Theorem 2. In those cases it is also possible to study the tradeoffs between secret key size and compression rate. A related question is how to construct secure compressor with non-equiprobable shared secret key. This question could have interesting connections with other related fields, for example the study of Randomness Extraction in theoretical computer science.

2) Random secrecy function: In the present paper we focus on deterministic secrecy functions. However, we note that our main decomposition lemmas, Lemmas 2 and 3, do not assume that \( Y \) is a deterministic function of \( X \). In that sense, our results apply to the case where the value of the random secrecy function is known at the compressor and the decompressor, as was studied in [17]. However, the compression problem will have a distinctly different flavor, for example, if the value of secrecy function \( Y \) is known only at the compressor, the decompressor, or is not known at all.

3) Secure decomposition strategies: One of the contributions of this work is to show that good secure compressors could be constructed using the two-part coding strategy. The ingredients of the two-part construction, the one-time pad and traditional compression, are well studied and easy to implement. The main technical challenge that remains is the secure decomposition of the source into these two parts. In this work, we present Lemmas 1-3 which enable us to derive our theoretical bounds. However, there are a number of practical considerations with regard to such decompositions which may require different decomposition strategies. For example, some applications may require iterative decomposition strategies, or strategies which minimize the use of local randomness at the encoder.

C. Universality of secure compressors

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 III are already universal. That is, consider the Covert Communication and Exponential Family examples in Section V. In the case of Covert Communication the fundamental limits (as well as the code used to achieve them) are independent of the parameter \( \delta \). Likewise, in the Exponential Family example the fundamental limits (as well as the code used to achieve them) do not depend on the parameter \( \beta \). Thus, although secure lossless compressors incur a penalty for partial secrecy, this penalty does provide a useful bonus of making them universal with respect to a family of statistical models. This universality property of secure compressors could prove to be useful in the setting of secure inference.

APPENDIX

Lemma 4. Let \((X, Y)\) be jointly distributed random variables, and let \(U\) be an arbitrary random variable independent of \(X\) and \(Y\). Let \(Z\) be a random variable on \(Z\) such that

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

and

\[
H(X|Z, U, Y) = 0. \tag{154}
\]

Then

\[
H(Z) \geq \max_{y \in \mathcal{Y}} H(X|Y = y). \tag{155}
\]

Proof. Equation (153) implies that

\[
H(Z) = H(Z|Y) = H(Z|Y = y) \quad \forall y \in \mathcal{Y}. \tag{156}
\]

Equation (154) implies that

\[
H(X|Z, U, Y = y) = 0 \quad \forall y \in \mathcal{Y}. \tag{157}
\]

Finally, using the chain rule

\[
H(X, Z|U, Y = y) = H(Z|U, Y = y) + H(X|Z, U, Y = y)
\]

\[
\leq H(Z|Y = y) \tag{159}
\]

\[
= H(Z) \tag{160}
\]

and

\[
H(X|Y = y) = H(X|U, Y = y) \tag{161}
\]

\[
\leq H(X, Z|U, Y = y) \tag{162}
\]

\[
\leq H(Z). \tag{163}
\]

Since (163) holds for all \(y \in \mathcal{Y}\) it must hold for a maximum over \(\mathcal{Y}\).

Lemma 2. Assume without loss of generality that \(\mathcal{X} = \{1, \ldots, |X|\}\), fix a positive integer \(L\), and let \(Z = \{1, \ldots, L + |X|\}\). Define

\[
n_{x,y} = |P_{X|Y}(x|y)L| \tag{164}
\]

and note that

\[
N_y = \sum_{x \in \mathcal{X}} n_{x,y} \leq L, \quad \forall y \in \mathcal{Y}. \tag{165}
\]
For a fixed $y \in \mathcal{Y}$ partition $\{1, \ldots, N_y\}$ into subsets $\mathcal{M}_x$ such that
\[ |\mathcal{M}_x| = n_{x,y}, \quad \forall x \in \mathcal{X} \tag{166} \]
and construct and auxiliary $Z'$ on $Z$:
\[ P_{Z'|X}(z|x,y) = \begin{cases} \frac{1}{LP_{X|Y}(x|y)}, & z \in \mathcal{M}_x \\ \delta_x, & z = L + x \\ 0, & \text{otherwise} \end{cases} \tag{167} \]
where
\[ \delta_x = 1 - \frac{n_x}{LP_{X|Y}(x|y)} \leq \frac{1}{LP_{X|Y}(x|y)}. \tag{168} \]
This transformation satisfies
\[ P_{Z'|X}(z|x,y) = \frac{1}{L}, \quad z \leq N = \min_{y \in \mathcal{Y}} N_y. \tag{169} \]
Finally, $Z$ is obtained by applying transformation from Lemma 1 to $(Z', \mathcal{Y})$ and secrecy function
\[ f_x(z,y) = y. \tag{170} \]
By construction it still holds that
\[ P_Z(z) = P_{Z'|Y}(z|y) = \frac{1}{L}, \quad z \leq N = \min_{y \in \mathcal{Y}} N_y. \tag{171} \]
Moreover, it is straightforward to check that (33) and (34) hold.
To complete the proof we need to get a bound on $P[Z \geq N]$:
\[ P[Z \geq N] = P[Z \geq N|Y = y] = \max_{y \in \mathcal{Y}} P[Z \geq N, Y = y]. \tag{172} \]
Finally,
\[ P[Z \geq N, Y = y] = P[n_X = 0|Z \geq N, n_X = 0]Y = y \tag{173} \]
\[ \leq P[n_X = 0|Y = y] + P[Z \geq N, n_X \geq 1|Y = y] \tag{174} \]
\[ \leq P[x_{1:Y}(X|y) > \log L|Y = y] \tag{175} \]
\[ + \mathbb{E} \left[ \{ \exp (x_{1:Y}(X|y) - \log L) \} x_{1:Y}(X|y) < \log L \} | Y = y \right]. \]
The bound on the second term follow since
\[ P[Z \geq N, n_X \geq 1|Y = y] = \sum_{z \in \mathcal{Z} \times \mathcal{X}} P_{Z|X, Y}(z|x,y) P_{X|Y}(x|y) \{ z \geq N_y \} \{ n_z \geq 1 \} \tag{176} \]
\[ \leq \sum_{x \in \mathcal{X}} \delta_z P_{X|Y}(x|y) \{ n_z \geq 1 \} \tag{177} \]
\[ = \sum_{x \in \mathcal{X}} P_{X|Y}(x|y) \exp (i_{1:Y}(X|y) - \log L) \{ x \in F_y \} \tag{178} \]
\[ = \mathbb{E} \left[ \exp (i_{1:Y}(X|y) - \log L) \{ x \in F_y \} | Y = y \right]. \tag{179} \]
where $F_y = \{ x : i_{1:Y}(x|y) < \log L \}$. 

\textbf{Lemma 3.} We will ensure that (36) holds by constructing $Z$ in such a way that
\[ P[Z = z|Y = y] = P[Z = z] = \frac{1}{L} \tag{180} \]
for all $y \in \mathcal{Y}$ and all $z \in \mathcal{Z}$.
Let
\[ n_{x,y} = [P_{X|Y}(x|y) L] \tag{181} \]
and note that
\[ N_y = \sum_{x \in \mathcal{X}} n_{x,y} \leq L, \quad \forall y \in \mathcal{Y}. \tag{182} \]
For a fixed $y \in \mathcal{Y}$ partition $\{1, \ldots, N_y\}$ into subsets $\mathcal{M}_x$ such that
\[ |\mathcal{M}_x| = n_{x,y}, \quad \forall x \in \mathcal{X} \tag{183} \]
Then
\[ P_{Z|X, Y}(z|x,y) = \begin{cases} \frac{1}{LP_{X|Y}(x|y)} \delta_x, & z \in \{ N_y + 1, \ldots, L \} \\ 0, & \text{otherwise} \end{cases} \tag{184} \]
where
\[ \delta_x = 1 - \frac{n_x}{LP_{X|Y}(x|y) L} \leq \frac{1}{LP_{X|Y}(x|y) L}. \tag{185} \]
Note that this construction satisfies the required condition. Indeed, suppose $z \in \mathcal{M}_x$ for some $x \in \mathcal{X}$. Then
\[ P_{Z|Y}(z|y) = P_{X|Y}(x|y) P_{Z|X, Y}(z|x,y) \tag{186} \]
\[ = \frac{P_{X|Y}(x|y)}{LP_{X|Y}(x|y)} = \frac{1}{L}. \tag{187} \]
Suppose $z \in \{ N_y + 1, \ldots, L \}$. Then
\[ P_{Z|Y}(z|y) = \sum_{x \in \mathcal{X}} P_{X|Y}(x|y) \delta_x \tag{188} \]
\[ = \frac{1}{L - N_y} \sum_{x \in \mathcal{X}} \left( P_{X|Y}(x|y) - \frac{n_x}{L} \right) \tag{189} \]
\[ = \frac{1}{L - N_y} \left( 1 - \frac{N_y}{L} \right) = \frac{1}{L}. \tag{190} \]
Finally,
\[ P[i_{X|Y}(X|Z, y) > 0| Y = y] \leq P[Z \geq N_y|Y = y] \tag{191} \]
\[ \leq \alpha_{XY}(L, y). \tag{192} \]
Follows by exactly the same argument as in previous Lemma since $N_y$ is defined identically.

The proof of Theorems 7 and 8 makes use of McDiarmid’s inequality, see for example [18], applied to the conditional information.

\textbf{Theorem 9} (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 $U = f(X^n)$, where $f: \mathcal{X}^n \to \mathbb{R}$ is a measurable function satisfying the bounded difference assumption. That is,

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

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

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

When $(X^n, Y^n)$ are distributed i.i.d. according to $P_{XY}$ the function

$$f(x^n, y^n) = i_{X^n|Y^n}(x^n|y^n)$$

satisfies the bounded difference assumption with

$$d_i = d = \max_{x \in \mathcal{X}, y \in \mathcal{Y}} i_{X|Y}(x|y).$$

Applying McDiarmid’s inequality we thus obtain

$$P \left[ i_{X^n|Y^n}(X^n|y^n) - \sum_{i=1}^n H(X|Y = y_i) \geq n\gamma \right] \geq Y^n$$

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

$$P \left[ i_{X^n|Y^n}(X^n|y^n) - nH(X|Y) \right] \geq n\gamma \leq 2 \exp \left( -\frac{2\gamma n}{d^2} \right)$$

for any $\gamma > 0$.

**Theorem 7.** Fix $\gamma > 0$ and let

$$y^* = \arg \max_{y \in \mathcal{Y}} H(X|Y = y).$$

Setting

$$l_n = n \left( \log |\mathcal{Y}| + H(X|Y = y^n) + \gamma \right)$$

and applying Theorem 3 with $\tau = \frac{1}{2} n\gamma$ we have that

$$\epsilon_{X^n|Y^n}(l_n, |\mathcal{Y}^n|) \geq 2 \exp \left( -\frac{\gamma^2}{d^2} \right) + 2^{-\frac{1}{2} n\gamma^2}.$$

Indeed, for every $y^n \in \mathcal{Y}^n$

$$P \left[ i_{X^n|Y^n}(X^n|y^n) \geq nH(X|Y = y^n) + \frac{n\gamma}{2} |Y^n = y^n \right] \leq P \left[ i_{X^n|Y^n}(X^n|y^n) \geq H(X^n|Y^n = y^n) + \frac{n\gamma}{2} |Y^n = y^n \right]$$

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

On the other hand, letting

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

and applying Theorem 4 with $\tau = \frac{1}{2} n\gamma$ we have that

$$\epsilon_{X^n|Y^n}(l_n, |\mathcal{Y}^n|) \geq 1 - 2 \exp \left( -\frac{\gamma^2}{d^2} \right) - 2^{-\frac{1}{2} n\gamma^2}.$$
Finally, given the type \( u_n \) follows from the fact that \( H(U^n | U_n = u_n) \) is the lower bound in the proof of Theorem 7 and Lemma 5. Then

\[
\Pr \left[ \sum_{y^n} P_{XY}(y^n | y_n) H(X^n | Y = y_n) + \frac{1}{3} n \gamma Y^n = y^n \right] \leq 2 \exp \left( -\frac{2r}{d^2} n \right). \tag{218}
\]

Moreover,

\[
\Pr \left[ \sum_{y^n} P_{Y^n | U_n}(y^n | u_n) H(U_n | Y_n = y_n) + \frac{1}{3} n \gamma Y^n = y^n \right] = 0 \tag{219}
\]

since given the type \( u_n \) all sequences \( y^n \) are equiprobable. Finally, for \( n \) sufficiently large,

\[
\Pr \left[ \sum_{y^n} P_{Y^n | U_n}(y^n | u_n) H(U_n | Y_n = y_n) + \frac{1}{3} n \gamma Y^n = y^n \right] \leq \exp \left( -\frac{r}{6} n \right). \tag{220}
\]

Then

\[
\Pr \left[ \sum_{y^n} P_{Y^n | U_n}(y^n | y_n) H(X^n | Y = y_n) + \frac{1}{3} n \gamma Y^n = y^n \right] \leq 2 \exp \left( -A_n n \right). \tag{222}
\]

\[
\Pr \left[ \sum_{y^n} P_{Y^n | U_n}(y^n | y_n) H(X^n | Y = y_n) + \frac{1}{3} n \gamma Y^n = y^n \right] \leq 2 \exp \left( -A_n n \right). \tag{223}
\]

**Theorem 8.** Note that the lower bound on \( S_{XY}^* \) follows immediately from Theorem 2, while the lower bound on \( S_{XY} \) 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 7 and Lemma 5.

**REFERENCES**


