Universal Encryption of Individual Sequences Under Maximal Leakage

Neri Merhav

The Andrew & Erna Viterbi Faculty of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, ISRAEL
E–mail: merhav@ee.technion.ac.il

Abstract

We consider the Shannon cipher system in the framework of individual sequences and finite-state encrypters under the metric of maximal leakage of information. A lower bound and an asymptotically matching upper bound on the leakage are derived, which lead to the conclusion that asymptotically minimum leakage can be attained by Lempel-Ziv compression followed by one-time pad encryption of the compressed bit-stream.

1 Introduction

Theoretical frameworks centered on the combination of individual sequences and finite-state encoders and decoders, have been thoroughly explored, marking a significant departure from the traditional probabilistic models typically employed in source and channel modeling. This shift has been particularly noticeable in a variety of information-theoretic fields, including data compression [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], source/channel simulation [11], [12], classification [13], [14], [15], prediction [16], [17], [18], [19], [20], [21], denoising [22], and even channel coding [23], [24], [25]. For a concise recent overview, see [26]. These references only skim the surface of a much larger body of work. In sharp contrast, the field of information-theoretic security, from Shannon's pioneering work [27] to more recent studies [28], [29], [30], [31], [32], has remained almost entirely rooted in the probabilistic framework. Although these examples represent just a small fraction of the extensive literature, they highlight the near-exclusive reliance on probabilistic models within this domain.

Only two notable exceptions to this prevailing paradigm are known to the author: one is an unpublished memorandum by Ziv [33] and the other is a subsequent work [34]. Ziv's memorandum introduces a distinctive approach where the plaintext source, intended for encryption with a secret key, is treated as an individual sequence. In this model, the encrypter is conceptualized as a general block encoder, while the eavesdropper employs a finite-state machine (FSM) to distinguish between messages. Ziv hypothesizes that the eavesdropper has some prior knowledge of the plaintext, which is expressed as a set of "acceptable messages," referred to as the acceptance set. In other words, prior to observing the ciphertext, the eavesdropper's uncertainty about the plaintext is that it could be any member of this set of acceptable messages. According to Ziv's framework, perfectly secure encryption occurs when the presence of the ciphertext does not reduce the uncertainty about the acceptance set. In essence, even after intercepting the ciphertext, the eavesdropper learns nothing new about the plaintext that she did not already know. The size of the acceptance set serves as a measure of uncertainty: a larger set corresponds to greater uncertainty. The FSM is then used to distinguish between acceptable and unacceptable plaintext sequences based on various key bit sequences. Consequently, perfect security is defined as maintaining the size of the acceptance set, and thus the uncertainty, unchanged in the presence of the ciphertext. Ziv's primary finding is that the asymptotic key rate necessary for perfectly secure encryption, according to this definition, cannot be lower (up to asymptotically vanishing terms) than the Lempel-Ziv (LZ) complexity of the plaintext source [10]. Notably, this lower bound can be asymptotically achieved using one-time pad encryption (i.e., bit-by-bit XOR with key bits) on the bit-stream generated by LZ data compression of the plaintext, echoing Shannon's classical probabilistic result that the minimum key rate required is equal to the source's entropy rate. More recently, Ziv's methodology has been refined and expanded in several directions in [35].

In the follow-up work [34], the concept of perfect secrecy for individual sequences was approached from a different perspective. Rather than assuming a finite-state eavesdropper with predefined knowledge, this framework posits that the encrypter itself can be modeled as a FSM, which is sequentially fed both the plaintext source and random key bits. A new concept, "finite-state encryptability", is introduced, inspired by the analogous idea of finite-state compressibility in [10]. This concept defines the minimum key rate that must be used by any finite-state encrypter to ensure that a certain form of normalized empirical mutual information between the plaintext and

ciphertext tends to zero as the block length grows. Among the key results in [34], it is established and proven that the finite-state encryptability of an individual sequence is fundamentally bounded from below by its finite-state compressibility. This lower bound is again asymptotically achieved by applying LZ compression to the plaintext and then one-time pad encryption of the compressed bits.

In this paper, we adopt the same model setting as in [34], but with a different security metric: the maximum leakage of information, which was first introduced by Issa, Wagner, and Kamath in [36] and then further explored in several more recent works, including [37], [38], [39], [40], and [41], among others. This metric is closely related to, and similarly motivated by, the earlier security measure proposed in [42], which defines security as a scenario where the correct decoding exponent of the plaintext is not improved by the availability of the ciphertext, compared to that of blind guessing. For more details, see the last paragraph of Subsection 2.2.1. The maximum leakage metric is defined in a more general form and has a relatively straightforward expression, as demonstrated in [36] and further clarified in the following sections. As will be discussed in the sequel, the maximum leakage metric is particularly well-suited for the individual-sequence setting considered here, as it is weakly dependent on the probability distribution of the plaintext, depending only on its support.

We derive both a lower bound and an asymptotically matching upper bound on the leakage, leading yet again to the conclusion that asymptotically optimal performance can be achieved by applying LZ compression followed by one-time pad encryption of the compressed bit-stream, and so, considering also the above mentioned earlier works, [33], [34], and [35], one of the messages of this work is that one-time pad encryption on top of LZ compression forms an asymptotically optimal cipher system from many aspects. That said, we believe that the deeper and more interesting contribution of this work is the converse theorem (Theorem 1 in the sequel) and its proof, asserting that the key rate that must be consumed to encrypt an individual sequence cannot be much smaller than the LZ complexity of the sequence minus the allowed normalized maximal information leakage.

The outline of the remaining part of this paper is as follows. In Section 2, we establish notation conventions, provide some necessary background, and formulate the problem studied in this work. In Section 3, we assert the main results and discuss them. Finally, in Section 4 we prove Theorem 1, which is the converse theorem.

2 Notation Conventions, Background and, Problem Formulation

2.1 Notation Conventions

Throughout this paper, scalar random variables (RV's) will be denoted by capital letters, their sample values will be denoted by the respective lower case letters, and their alphabets will be denoted by the respective calligraphic letters. A similar convention will apply to random vectors and their sample values, which will be denoted with same symbols superscripted by the dimension. Thus, for example, A^m (m – positive integer) will denote a random m-vector $(A_1, ..., A_m)$, and $a^m = (a_1, ..., a_m)$ is a specific vector value in $\mathcal{A}^m$, the m–th Cartesian power of $\mathcal{A}$. The notations a_i^j and A_i^j, where i and j are integers and $i \leq j$, will designate segments $(a_i, \ldots, a_j)$ and $(A_i, \ldots, A_j)$, respectively, where for $i = 1$, the subscript will be omitted (as above). For $i > j$, a_i^j (or A_i^j) will be understood as the null string. The notation $[u]_+$ for a real u will stand for $\max\{0, u\}$. Logarithms and exponents, throughout this paper, will be understood to be taken to the base 2 unless specified otherwise.

Sources and channels will be denoted generically by the letter P or Q, subscripted by the name of the RV and its conditioning, if applicable, exactly like in ordinary textbook notation standards, e.g., $P_{X^m}(x^m)$ is the probability function of X^m at the point $X^m = x^m$, $P_{X|W^m}(x|w^m)$ is the conditional probability of $X = x$ given $W^m = w^m$, and so on. Whenever clear from the context, these subscripts will be omitted. Information theoretic quantities, like entropies and mutual informations, will be denoted following the usual conventions of the information theory literature, e.g., $H(K^m)$, $I(V; X^m|W^m)$, and so on.

In the sequel $x^n = (x_1, \ldots, x_n)$ will designate an individual sequence to be encrypted. The components, $\{x_i\}$ of x^n all take values in a finite alphabet, $\mathcal{X}$, whose cardinality will be denoted by α.

2.2 Background

Before the exposition of the main results and their proofs, we revisit key terms and details related to the notion of maximal leakage of information and the 1978 version of the LZ algorithm, also known as the LZ78 algorithm [10], which is the central building block in this work.

2.2.1 Maximal Leakage of Information

As mentioned in the Introduction, in this paper, we adopt the maximal leakage [36] as our secrecy metric. For a probabilistic plaintext source, the maximal leakage from a secret random variable X, distributed according to $\{P_X(x),\ x \in \mathcal{X}\}$, to another random variable Y, available to an adversary, and which is conditionally distributed given $X = x$ according to $\{P_{Y|X}(y|x),\ x \in \mathcal{X},\ y \in \mathcal{Y}\}$, is defined as

$$\mathcal{L}(X \to Y) \stackrel{\Delta}{=} \sup_{U-X-Y-\hat{U}} \log \frac{\Pr\{\hat{U} = U\}}{\max_{u \in \mathcal{U}} P_U(u)}, \tag{1}$$

where the supremum is over all finite-alphabet random variables U and $\hat{U}$, with the Markov structure $U - X - Y - \hat{U}$. In other words, it is the maximum possible difference between the logarithm of the probability of correctly guessing some (possibly randomized) function of X based on Y and correctly guessing it blindly.

In Theorem 1 of [36], it was asserted and proved that the leakage can be calculated relatively easily using the formula:

$$\mathcal{L}(X \to Y) = \log \left[\sum_{y \in \mathcal{Y}} \max_{\{x:\ P_X(x) > 0\}} P_{Y|X}(y|x) \right]. \tag{2}$$

Clearly, if $P_{Y|X}(y|x)$ is independent of x for all $y \in \mathcal{Y}$ then $\mathcal{L}(X \to Y) = 0$, which is the case of perfect secrecy. In general, the smaller is $\mathcal{L}(X \to Y)$, the more secure the system is. In [36], it is shown that the maximal leakage has many interesting properties, one of them is that it satisfies a data processing inequality (see Lemma 1 of [36]). It is also shown in Section III of [36] that the maximal leakage has several additional operative meanings in addition to the original one explained above.

Note that the dependence on the distribution of the secret random variable, P_X, is rather weak, as it depends only on its support. When passing from single variables to vectors of length n, $\mathcal{L}(X^n \to Y^n)$ is defined in the same manner except that x, y, $\mathcal{X}$, $\mathcal{Y}$, $P_X(\cdot)$, and $P_{Y|X}(\cdot|\cdot)$ are replaced by x^n, y^n, $\mathcal{X}^n$, $\mathcal{Y}^n$, $P_{X^n}(\cdot)$, and $P_{Y^n|X^n}(\cdot|\cdot)$, respectively. In this case, the weak dependence of $\mathcal{L}(X^n \to Y^n)$ on P_{X^n} makes it natural to use when P_{X^n} is uncertain, or completely unknown, or even non-existent, such as in the individual sequence setting considered here. In this case, we

adopt the simple definition

$$\mathcal{L}(x^n \to Y^n) \stackrel{\Delta}{=} \log \left[\sum_{y^n \in \mathcal{Y}^n} \max_{x^n \in \mathcal{X}^n} P_{Y^n|X^n}(y^n|x^n) \right], \tag{3}$$

corresponding to the full support $\mathcal{X}^n$ for x^n, which accounts for a worst-case approach. The operational significance of maximal information leakage in the our setting can then be understood in two ways: (i) Considering the definition (1), it allows arbitrary probability distributions (without any assumed structure) on x^n, including those that put almost all their mass on a single (unknown) arbitrary sequence, in the spirit of the individual-sequence setting considered here. (ii) Referring to the formula (3), it is evident that the leakage vanishes whenever $P_{Y^n|X^n}(y^n|x^n)$ is independent of x^n, which is an indisputable characterization for perfect secrecy in the individual-sequence setting too, where no distribution at all is assumed on x^n.

As mentioned in the Introduction, in [42] a somewhat different security metric was proposed, but it is intimately related to the maximal information leakage considered here. In [42], the idea was to define a system as secure if the probability of guessing X correctly is essentially the same if Y is present or absent. (More precisely, if X and Y are random vectors of dimension n, then a system is considered secure if the correct decoding exponent of X in the presence of Y is the same as if Y is absent.) Specifically, the correct decoding probability of X based on Y is

$$P_{\mathrm{c}} = \sum_y \max_x P_{XY}(x,y), \tag{4}$$

which is closely related to

$$\begin{aligned} 2^{\mathcal{L}(X\to Y)} &= \sum_y \max_x P_{Y|X}(y|x) \\ &= |\mathcal{X}| \cdot \sum_y \max_x \frac{P_{Y|X}(y|x)}{|\mathcal{X}|} \\ &\stackrel{\Delta}{=} |\mathcal{X}| \cdot \sum_y \max_x P_X(x) P_{Y|X}(y|x) \\ &= \frac{\sum_y \max_x P_X(x) P_{Y|X}(y|x)}{1/|\mathcal{X}|} \\ &= \frac{P_{\mathrm{c}}^{\mathrm{i}}}{P_{\mathrm{c}}^{\mathrm{u}}}, \end{aligned} \tag{5}$$

where $P_X(\cdot)$ is understood to designate the uniform distribution across $\mathcal{X}$, and accordingly, $P_{\mathrm{c}}^{\mathrm{i}}$ stands for the probability of correct decoding of a uniformly distributed X by an informed observer,

namely, one that has access to Y, whereas $P_{\rm c}^{\rm u} = 1/|\mathcal{X}|$ denotes the probability of correct blind guessing the value of X (in the absence of Y).

2.2.2 Lempel-Ziv Parsing

The incremental parsing procedure in the LZ78 algorithm is a sequential method applied to an input vector x^n over a finite alphabet. In this process, each new phrase is defined as the shortest substring that has not appeared previously as a complete parsed phrase, except possibly for the final (incomplete) phrase. For example, applying incremental parsing to the sequence $x^{15} = \text{abbabaabbaaabaa}$ yields a,b,ba,baa,bb,aa,ab,aa. Let $c(x^n)$ designate the total number of phrases formed from x^n using the incremental parsing procedure (in the example above, $c(x^{15}) = 8$). Also, let $LZ(x^n)$ stand for the length of the LZ78 binary compressed representation for x^n. By Theorem 2 of [10], the following inequality holds:

$$\begin{aligned}
LZ(x^n) &\leq [c(x^n)+1]\log\{2\alpha[c(x^n)+1]\} \\
&= c(x^n)\log[c(x^n)+1] + c(x^n)\log(2\alpha) + \\
&\quad \log\{2\alpha[c(x^n)+1]\} \\
&= c(x^n)\log c(x^n) + c(x^n)\log\left[1+\frac{1}{c(x^n)}\right] + c(x^n)\log(2\alpha) + \log\{2\alpha[c(x^n)+1]\} \\
&\leq c(x^n)\log c(x^n) + \log e + \frac{n(\log\alpha)\log(2\alpha)}{(1-\varepsilon_n)\log n} + \log[2\alpha(n+1)] \\
&\stackrel{\Delta}{=} c(x^n)\log c(x^n) + n\cdot\epsilon(n), \qquad (6)
\end{aligned}$$

where we remind that α is the cardinality of $\mathcal{X}$, and where ε_n and $\epsilon(n)$ both tend to zero as $n \to \infty$. Stated differently, the LZ code-length for x^n is upper bounded by an expression whose main term is $c(x^n)\log c(x^n)$. On the other hand, $c(x^n)\log c(x^n)$ is also the dominant term of a lower bound (see Theorem 1 of [10]) to the shortest code-length attainable by any information lossless finite-state encoder with no more than s states, provided that $\log(s^2)$ is very small compared to $\log c(x^n)$. Accordingly, we henceforth refer to $c(x^n)\log c(x^n)$ as the unnormalized *LZ complexity* of x^n whereas the normalized LZ complexity is defined as

$$\rho_{\rm LZ}(x^n) \stackrel{\Delta}{=} \frac{c(x^n)\log c(x^n)}{n}. \qquad (7)$$

2.3 Problem Formulation

Similarly as in [34], we adopt the following model of finite-state encryption. A finite–state encrypter is defined by a sextuplet

$$E = (\mathcal{X}, \mathcal{Y}, \mathcal{Z}, f, g, \Delta),$$

where $\mathcal{X}$ is a finite input alphabet of size $\alpha = |\mathcal{X}|$, $\mathcal{Y}$ is a finite set of variable-length binary strings, including possibly the empty string λ (of length zero), $\mathcal{Z}$ is a finite set of states, $f : \mathcal{Z} \times \mathcal{X} \times \{0,1\}^* \to \mathcal{Y}$ is the output function, $g : \mathcal{Z} \times \mathcal{X} \to \mathcal{Z}$ is the next-state function, and $\Delta : \mathcal{Z} \times \mathcal{X} \to 0, 1, 2, \ldots$ specifies the number of key bits consumed per step. When two infinite sequences, $\boldsymbol{x} = x_1, x_2, \ldots, x_i \in \mathcal{X}$, henceforth the *plain-text sequence* (or, the source sequence), and $\boldsymbol{u} = u_1, u_2, \ldots, u_i \in \{0,1\}$, $i = 1, 2, \ldots$, henceforth the *key sequence*, are fed into an encrypter E, it produces an infinite output sequence $\boldsymbol{y} = y_1, y_2, \ldots, y_i \in \mathcal{Y}$, henceforth the *ciphertext*, while passing through an infinite sequence of states $\boldsymbol{z} = z_1, z_2, \ldots, z_i \in \mathcal{Z}$, according to the following recursive equations, implemented for $i = 1, 2, \ldots$

$$t_i = t_{i-1} + \Delta(z_i, x_i), \qquad t_0 \stackrel{\Delta}{=} 0 \tag{8}$$

$$k_i = (u_{t_{i-1}+1}, u_{t_{i-1}+2}, \ldots, u_{t_i}) \tag{9}$$

$$y_i = f(z_i, x_i, k_i) \tag{10}$$

$$z_{i+1} = g(z_i, x_i) \tag{11}$$

where the initial state, z_1, is assumed fixed, and will be labeled $z_\star$ hereafter, and where it is understood that if $\Delta(z_i, x_i) = 0$, then $k_i = \lambda$, the null word of length zero, namely, no key bits are used in the i–th step. By the same token, if $y_i = \lambda$, no output is produced at this step, i.e., the system is idling and only the state evolves in response to the input. In other words, at each time instant t, when the state is z_i, the encrypter is fed by the current plain-text symbol x_i and it consumes the next $\Delta(z_i, x_i)$ previously unused key bits. It then updates the next state z_{i+1} and produces an output y_i.

In summary, at each time step i: the current state is z_i, the encrypter receives input x_i, consumes the next $\Delta(z_i, x_i)$ unused key bits from $\boldsymbol{u}$ to form k_i, produces output y_i, and transitions to the next state z_{i+1}.

Remark 1. Note that the evolution of the state variable z_i depends solely on the source inputs $\{x_i\}$ and is independent of the key bits. This design choice reflects the intended role of z_i, which is to retain memory of the source sequence x^n, allowing the encrypter to exploit empirical correlations and repetitive patterns within the plaintext. In contrast, maintaining memory of past key bits—which are assumed to be independent and identically distributed (i.i.d.)—offers no practical benefit and is therefore omitted. That said, the model can be naturally extended to include two state variables: one that evolves based only on the source sequence $\{x_i\}$ (as in the current setup), and another that evolves based on both $\{x_i\}$ and the consumed key bits $\{k_i\}$. In such a framework, the first state variable would continue to govern the update of the index t_i, while the second could influence the output function, allowing for more expressive or adaptive encryption mechanisms.

An encrypter with s states, or an s–state encrypter, E, is one with $|\mathcal{Z}| = s$. It is assumed that the plain-text sequence $\boldsymbol{x}$ is deterministic (i.e., an individual sequence), whereas the key sequence $\boldsymbol{u}$ is purely random, i.e., for every positive integer n, $P_{U^n}(u^n) = 2^{-n}$.

A few additional notation conventions will be convenient: By $f(z_i, x_i^j, k_i^j)$, $(i \leq j)$ we refer to the vector y_i^j produced by E in response to the inputs x_i^j and k_i^j when the initial state is z_i. Similarly, the notation $g(z_i, x_i^j)$ will mean the state z_{j+1} and $\Delta(z_i, x_i^j)$ will designate $\sum_{\ell=i}^{j} \Delta(z_\ell, x_\ell)$ under the same circumstances.

As explained in Subsection 2.2, we adopt the maximal leakage of information as our security metric given by

$$\mathcal{L}(x^n \to Y^n) \stackrel{\Delta}{=} \ln \left[\sum_{y^n} \max_{x^n \in \mathcal{X}^n} P_{Y^n|X^n}(y^n|x^n) \right]. \tag{12}$$

An encryption system E is said to be *perfectly secure* if for every positive integer n, $\mathcal{L}(x^n \to Y^n) = 0$. If $\mathcal{L}(x^n \to Y^n) \to 0$ as $n \to \infty$, we say that the encryption system is asymptotically secure.

An encrypter is referred to as *information lossless* (IL) if for every $z_i \in \mathcal{Z}$, every sufficiently large n and all pairs (x_i^{i+n}, k_i^{i+n}), the quadruple $(z_i, k_i^{i+n}, f(z_i, x_i^{i+n}, k_i^{i+n}), g(z_i, x_i^{i+n}))$ uniquely determines x_i^{i+n}. Given an encrypter E and an input string x^n, the encryption key rate of x^n w.r.t. E is defined as

$$\sigma_E(x^n) \stackrel{\Delta}{=} \frac{\ell(k^n)}{n} = \frac{1}{n} \sum_{i=1}^{n} \ell(k_i), \tag{13}$$

where $\ell(k_i) = \Delta(z_i, x_i)$ is the length of the binary string k_i and $\ell(k^n) = \sum_{i=1}^{n} \ell(k_i)$ is the total length of k^n.

Remark 2. It is worth noting that the definition of information losslessness used here is more relaxed, and thus more general, than the one given in [10]. In [10], the requirement must hold for every positive integer n whereas in the present context, it is only required to hold for all sufficiently large n. The absence of information losslessness in the stricter sense of [10] does not contradict the ability of the legitimate decoder to reconstruct the source. Rather, it implies that reconstructing x^n may require more than just the tuple $(z_i, y_i^{i+n}, k_i^{i+n}, z_{i+n+1})$, for example, some additional data from times later than $i + n + 1$ may be needed.

The set of all perfectly secure, IL encrypters $\{E\}$ with no more than s states will be denoted by $\mathcal{E}(s)$. The minimum of $\sigma_E(x^n)$ over all encrypters in $\mathcal{E}(s)$ will be denoted by $\sigma_s(x^n)$, i.e.,

$$\sigma_s(x^n) = \min_{E \in \mathcal{E}(s)} \sigma_E(x^n). \tag{14}$$

Finally, let

$$\sigma_s(\boldsymbol{x}) = \limsup_{n \to \infty} \sigma_s(x^n), \tag{15}$$

and define the *finite–state encryptability* of $\boldsymbol{x}$ as

$$\sigma(\boldsymbol{x}) = \lim_{s \to \infty} \sigma_s(\boldsymbol{x}). \tag{16}$$

Our purpose is to characterize these quantities and to point out how they can be achieved in principle.

3 Main Results

Our converse theorem, whose proof appears in Section 4, is the following.

Theorem 1. For every information lossless encrypter E with no more than s states,

$$\begin{aligned} \frac{\mathcal{L}(x^n \to Y^n)}{n} \quad \geq \quad & \Bigg[\max_{x^n \in \mathcal{X}^n} \{\rho_{\mathrm{LZ}}(x^n) - \sigma_E(x^n)\} - \\ & \delta_s(n) - \frac{(\alpha s - 1)\log(n+1)}{n} - \frac{\log s}{n} \Bigg]_+, \end{aligned} \tag{17}$$

where $\delta_s(n) \leq O\left(\frac{\log(\log n)}{\log n}\right)$ for every fixed s. Equivalently, if $0 \leq \mathcal{L}(x^n \to Y^n) \leq n\lambda$ for some given constant $\lambda \geq 0$, then for every $x^n \in \mathcal{X}^n$ and every information lossless encrypter $E \in \mathcal{E}(s)$,

$$\sigma_E(x^n) \geq \rho_{\mathrm{LZ}}(x^n) - \lambda - \delta_s(n) - \frac{(\alpha s - 1)\log(n+1)}{n} - \frac{\log s}{n}. \tag{18}$$

As for achievability, consider first an arbitrary lossless compression scheme that compresses x^n at a compression ratio of $\rho(x^n) = L(x^n)/n$, and then applies one-time pad encryption to $[L(x^n) - n\lambda]_+$ compressed bits. Let y^n denote the resulting (partially) encrypted compressed representation of x^n. Then, obviously, the length of y^n, denoted $L(y^n)$, is equal to $L(x^n)$ and so, denoting $L_{\max} = \max_{x^n \in \mathcal{X}^n} L(x^n)$, we have:

$$\begin{aligned}
\exp_2\{\mathcal{L}(x^n \to y^n)\} &= \sum_{y^n \in \mathcal{Y}^n} \max_{x^n} P_{Y^n|X^n}(y^n|x^n) \\
&= \sum_{\ell=1}^{L_{\max}} \sum_{\{y^n:\ L(y^n)=\ell\}} \max_{x^n} P_{Y^n|X^n}(y^n|x^n) \\
&= \sum_{\ell=1}^{L_{\max}} \sum_{\{y^n:\ L(y^n)=\ell\}} 2^{-[\ell - n\lambda]_+} \\
&\leq \sum_{\ell=1}^{L_{\max}} 2^{\ell} 2^{-[\ell - n\lambda]_+} \\
&\leq \sum_{\ell=1}^{L_{\max}} 2^{\ell} 2^{-(\ell - n\lambda)} \\
&= L_{\max} \cdot 2^{n\lambda},
\end{aligned} \tag{19}$$

and so,

$$\mathcal{L}(x^n \to y^n) \leq n\lambda + \log L_{\max}. \tag{20}$$

If $L_{\max} = O(n)$, then the dominant term is clearly $n\lambda$.

Remark 3. The condition that $L_{\max} = O(n)$ is easy to satisfy always by a minor modification of any given compression scheme (if it does not satisfy the condition in the first place). First, test whether $L(x^n) < \lceil n \log \alpha \rceil$ or $L(x^n) \geq \lceil n \log \alpha \rceil$. If $L(x^n) < \lceil n \log \alpha \rceil$ add a header bit '0' before the compressed representation of x^n; otherwise, add a header bit '1' and then the uncompressed binary representation of x^n using $\lceil n \log \alpha \rceil$ bits. The resulting code-length would then be

$L'(x^n) = \min\{L(x^n), \lceil n \log \alpha \rceil\} + 1$ bits.

If the compression scheme is chosen to be the LZ78 algorithm then,

$$\sigma_E(x^n) \leq \rho_{\mathrm{LZ}}(x^n) - \lambda + O\left(\frac{\log\log n}{\log n}\right), \tag{21}$$

which essentially meets the converse bound (18). We have therefore proved the following direct theorem.

Theorem 2. Given $\lambda \geq 0$, there exists a universal encrypter that satisfies

$$\mathcal{L}(x^n \to Y^n) \leq n\lambda + \log n + O(1), \tag{22}$$

and for every $x^n \in \mathcal{X}^n$,

$$\sigma_E(x^n) \leq \rho_{\mathrm{LZ}}(x^n) - \lambda + O\left(\frac{\log\log n}{\log n}\right). \tag{23}$$

Discussion. A few comments are now in order.

1. We established both a lower bound and an asymptotically matching upper bound on the information leakage, leading once again to the conclusion that asymptotically optimal performance can be achieved by applying Lempel-Ziv (LZ) compression followed by one-time pad encryption of the compressed bitstream. Together with earlier works such as [33], [34], and [35], this reinforces the message that one-time pad encryption applied after LZ compression yields an asymptotically optimal cipher system in several important respects. That said, we believe the deeper and more significant contribution of this work lies in the converse theorem (Theorem 1), which shows that the key rate required to securely encrypt an individual sequence cannot be substantially smaller than its LZ complexity minus the permitted normalized maximal information leakage.

2. Similarly as in [10], formally there is a certain gap between the converse theorem and the achievability scheme in its basic form, when examined from the viewpoint of the number of states, s, relative to n. While s should be small relative to n for the lower bound to be essentially $\rho_{\mathrm{LZ}}(x^n)$ (see Subsection 2.2 above), the number of states actually needed to implement LZ78 compression

for a sequence of length n is basically exponential in n. In [10], the gap is closed in the limit of $s \to \infty$ (after taking the limit $n \to \infty$) by subdividing the sequence into blocks and restarting the LZ algorithm at the beginning of every block. A similar comment applies here too in the double limit of achieving $\sigma(\boldsymbol{x})$.

3. As discussed in [35] in a somewhat different context, for an alternative to the use of the LZ78 algorithm, it can be shown that asymptotically optimum performance can also be attained by a universal compression scheme for the class of k-th order Markov sources, where k is chosen sufficiently large. In this case, $\rho_{\mathrm{LZ}}(x^n)$ in Theorems 1 and 2 should be replaced by the k-th order empirical entropy of order k and some redundancy terms should be modified. But one of these redundancy terms is $\frac{\log s}{k+1}$, which means that in order to compete with the best encrypter with s states, k must be chosen significantly larger than $\log s$, so as to make this term reasonably small.

4. It is speculated that it may not be difficult extend our findings in several directions, including: lossy reconstruction, the presence of side information at either parties, the combination of both, and successive refinement systems in the spirit of [41]. Other potentially interesting extensions are in broadening the scope of the FSM model to larger classes of machines, including: FSMs with counters, shift-register machines with counters, and periodically time-varying FSMs with counters, as was done in Section III of [35]. Research work in some of these directions is deferred to future studies.

4 Proof of Theorem 1

First, observe that

$$\sigma_E(x^n) = \frac{1}{n}\sum_{i=1}^{n}\Delta(z_i, x_i) = \sum_{x,z}\hat{P}(x,z)\Delta(z,x), \tag{24}$$

where $\hat{P} = \{\hat{P}(x,z),\ x \in \mathcal{X},\ z \in \mathcal{Z}\}$ is the joint empirical distribution of (x,z) derived from (x^n, z^n). It is therefore seen that $\sigma_E(x^n)$ depends on x^n only via $\hat{P}$. Accordingly, in the sequel, we will also use the alternative notation $\sigma_E(\hat{P})$ when we wish to emphasize the dependence on $\hat{P}$. Let $\mathcal{T}(x^n)$ denote the set of $\tilde{x}^n \in \mathcal{X}^n$, that together with their associated state sequences, share the same empirical PMF $\hat{P}$ as that of x^n along with its state sequence. Similarly as with $\sigma_E(\cdot)$, we

also denote it by $\mathcal{T}(\hat{P})$. In the sequel, we will make use of the inequality

$$\frac{\log|\mathcal{T}(x^n)|}{n} \geq \rho_{\mathrm{LZ}}(x^n) - \delta_s(n), \tag{25}$$

where $\delta_s(n) \to 0$ as $n \to \infty$ for fixed s at the rate of $\frac{\log(\log n)}{\log n}$. The proof of eq. (25), which appears in various forms and variations in earlier papers (see, e.g., [43]), is provided in the appendix for the sake of completeness (see also the related Ziv's inequality in Lemma 13.5.5 of [44]).

For later use, we also define the following sets.

$$\mathcal{P}(y^n) = \{\hat{P}:\ \mathcal{T}(\hat{P}) \cap f^{-1}(y^n) \neq \emptyset\}, \tag{26}$$

$$\mathcal{Y}(\hat{P}) = \{y^n:\ \mathcal{T}(\hat{P}) \cap f^{-1}(y^n) \neq \emptyset\}, \tag{27}$$

where

$$\begin{aligned} f^{-1}(y^n) \quad &= \quad \{x^n:\ f(z_\star, x^n, k^n) = y^n \\ &\qquad \text{for some } k^n \in \{0,1\}^{n\sigma_E(x^n)}\}. \end{aligned} \tag{28}$$

Now, observe that

$$\begin{aligned} |\mathcal{Y}(\hat{P})| \quad &\geq \quad \left|\left\{y^n:\ y^n = f(z_\star, x^n, 0^{n\sigma_E(x^n)})\right.\right. \\ &\qquad \left.\left.\text{for some } x^n \in \mathcal{T}(\hat{P})\right\}\right| \\ &\geq \quad \max_{z\in\mathcal{Z}} \left|\left\{y^n:\ y^n = f(z_\star, x^n, 0^{n\sigma_E(x^n)})\right.\right. \\ &\qquad \left.\left.\text{and } g(z_\star, x^n) = z \text{ for some } x^n \in \mathcal{T}(\hat{P})\right\}\right| \\ &\stackrel{\Delta}{=} \quad \max_{z\in\mathcal{Z}} |\mathcal{Y}_z(\hat{P})| \\ &\geq \quad \frac{|\mathcal{T}(\hat{P})|}{s}, \end{aligned} \tag{29}$$

where the last inequality follows from the following consideration: Let x^n exhaust all members of $\mathcal{T}(\hat{P})$. For each such x^n, let $y^n = f(z_\star, x^n, 0^{n\sigma_E(x^n)})$. Now for every $z \in \mathcal{Z}$, let $\mathcal{T}_z(\hat{P})$ denote the subset of $\mathcal{T}(\hat{P})$ for which $z_{n+1} = g(z_\star, x^n) = z$, and we have already defined $\mathcal{Y}_z(P)$ to denote the set of corresponding output sequences, $\{y^n\}$. Obviously, since $\{\mathcal{T}_z(\hat{P})\}_{z\in\mathcal{Z}}$ form a partition of $\mathcal{T}(\hat{P})$, then for some $z = z^*$, $|\mathcal{T}_{z^*}(\hat{P})| \geq |\mathcal{T}(\hat{P})|/s$, and so,

$$\max_z |\mathcal{Y}_z(\hat{P})| \quad \geq \quad |\mathcal{Y}_{z^*}(\hat{P})|$$

$$
\begin{aligned}
&= |\mathcal{T}_{z^*}(\hat{P})| \\
&\geq \frac{|\mathcal{T}(\hat{P})|}{s}, \qquad (30)
\end{aligned}
$$

where the equality is since the mapping between x^n and y^n is one-to-one given that $k^n = 0^{n\sigma_E(x^n)}$, $z_1 = z_\star$, and $z_{n+1} = z^*$ by the information losslessness postulated, provided that n is sufficiently large as required. Now, let $\mathcal{Y}_n^+$ denote that set of all $y^n \in \mathcal{Y}^n$ for which $P_{Y^n|X^n}(y^n|x^n) > 0$ for some $x^n \in \mathcal{X}^n$. Then,

$$
\begin{aligned}
\exp_2\{\mathcal{L}(x^n \to y^n\} &= \sum_{y^n \in \mathcal{Y}_n^+} \max_{x^n} P_{Y^n|X^n}(y^n|x^n) \\
&= \sum_{y^n \in \mathcal{Y}_n^+} \max_{x^n \in \phi^{-1}(y^n)} P_{Y^n|X^n}(y^n|x^n) \\
&= \sum_{y^n \in \mathcal{Y}_n^+} \max_{\hat{P} \in \mathcal{P}(y^n)} \max_{x^n \in \phi^{-1}(y^n) \cap \mathcal{T}(\hat{P})} P_{Y^n|X^n}(y^n|x^n) \\
&\overset{\text{(a)}}{\geq} \sum_{y^n \in \mathcal{Y}_n^+} \max_{\hat{P} \in \mathcal{P}(y^n)} \max_{x^n \in \phi^{-1}(y^n) \cap \mathcal{T}(\hat{P})} 2^{-n\sigma_E(\hat{P})} \\
&= \sum_{y^n \in \mathcal{Y}_n^+} \max_{\hat{P} \in \mathcal{P}(y^n)} 2^{-n\sigma_E(\hat{P})} \\
&\geq \frac{1}{M_n} \sum_{y^n \in \mathcal{Y}_n^+} \sum_{\hat{P} \in \mathcal{P}(y^n)} 2^{-n\sigma_E(\hat{P})} \\
&= \frac{1}{M_n} \sum_{\hat{P}} \sum_{y^n \in \mathcal{Y}(\hat{P})} 2^{-n\sigma_E(\hat{P})} \\
&= \frac{1}{M_n} \sum_{\hat{P}} |\mathcal{Y}(\hat{P})| \cdot 2^{-n\sigma_E(\hat{P})} \\
&\overset{\text{(b)}}{\geq} \frac{1}{M_n s} \sum_{\hat{P}} |\mathcal{T}(\hat{P})| \cdot 2^{-n\sigma_E(\hat{P})} \\
&\geq \frac{1}{M_n s} \cdot \max_{x^n \in \mathcal{X}^n} |\mathcal{T}(x^n)| \cdot 2^{-n\sigma_E(x^n)} \\
&\overset{\text{(c)}}{\geq} \frac{1}{M_n s} \cdot \max_{x^n \in \mathcal{X}^n} 2^{n[\rho_{\mathrm{LZ}}(x^n) - \delta_s(n)]} \cdot 2^{-n\sigma_E(x^n)} \\
&= \exp_2 \left\{ n \cdot \max_{x^n \in \mathcal{X}^n} \left[\rho_{\mathrm{LZ}}(x^n) - \sigma_E(x^n) - \delta_s(n) - \right.\right. \\
&\qquad \left.\left. \frac{\log M_n}{n} - \frac{\log s}{n} \right] \right\} \\
&\geq \exp_2 \left\{ n \cdot \max_{x^n \in \mathcal{X}^n} \left[\rho_{\mathrm{LZ}}(x^n) - \sigma_E(x^n) - \delta_s(n) - \right.\right.
\end{aligned}
$$

$$\frac{(\alpha s-1)\log(n+1)}{n}-\frac{\log s}{n}\Bigg]\Bigg\}, \tag{31}$$

where in (a) we used the fact that $P_{Y^n|X^n}(y^n|x^n)>0$ implies $P_{Y^n|X^n}(y^n|x^n)\geq 2^{-n\sigma_E(x^n)}$ (because $P_{Y^n|X^n}(y^n|x^n)>0$ implies that there is at least one $k^n\in\{0,1\}^{n\sigma_E(x^n)}$ such that $f(z_\star,x^n,k^n)=y^n$ and the probability of each such k^n is $2^{-n\sigma_E(x^n)}$), and where M_n is the number of different type classes, $\{\hat{P}\}$, which is upper bounded by $(n+1)^{\alpha s-1}$. In (b) we used eq. (29) and in (c) we used eq. (25). Finally, the operator $[\cdot]_+$ that appears in the assertion of Theorem 1 is due to the additional trivial lower bound $\mathcal{L}(x^n\to Y^n)\geq 0$. This completes the proof of Theorem 1.

Appendix – Proof of Eq. (25)

Consider the LZ78 incremental parsing procedure applied to x^n and let $c_{\ell zz'}$, $\ell\in\mathcal{N}$, $z,z'\in\mathcal{Z}$, denote the number of phrases of length ℓ, which start at state z and end at state z'. Clearly, $\sum_{\ell,z,z'}c_{\ell zz'}=c(x^n)$, for which we will use the shorthand notation c in this appendix.

Given that $x^n\in\mathcal{T}(\hat{P})$, one can generate other members of $\mathcal{T}(\hat{P})$ by permuting phrases of the same length which start at the same state and end at the same state. Thus, $|\mathcal{T}(\hat{P})|\geq\prod_{\ell,z,z'}(c_{\ell zz'}!)$, and so,

$$\begin{aligned}
\log|\mathcal{T}(\hat{P})| &\geq \sum_{\ell,z,z'}\log(c_{\ell zz'}!)\\
&\geq \sum_{\ell,z,z'}c_{\ell zz'}\log\frac{c_{\ell zz'}}{e}\\
&= \sum_{\ell,z,z'}c_{\ell zz'}\log c_{\ell zz'}-c\log e\\
&= c\sum_{\ell,z,z'}\frac{c_{\ell zz'}}{c}\left[\log\frac{c_{\ell zz'}}{c}+\log c\right]-c\log e\\
&= c\log c-cH(L,Z,Z')-c\log e,
\end{aligned} \tag{A.1}$$

where $H(L,Z,Z')$ is the joint entropy of the auxiliary random variables L, Z, and Z', jointly distributed according to the distribution $\pi(\ell,z,z')=c_{\ell zz'}/c$, $\ell\in\mathcal{N}$, $z,z'\in\mathcal{Z}$. To further bound $\log|\mathcal{T}(\hat{P})|$ from below, we now derive an upper bound to $H(L,Z,Z')$:

$$\begin{aligned}
H(L,Z,Z') &\leq H(L)+H(Z)+H(Z')\\
&\leq H(L)+2\log s
\end{aligned}$$

$$
\begin{aligned}
&\leq (1+EL)\log(1+EL) - (EL)\log(EL) + 2\log s \\
&= \left(1+\frac{n}{c}\right)\log\left(1+\frac{n}{c}\right) - \frac{n}{c}\log\frac{n}{c} + 2\log s \\
&= \frac{n}{c}\log\left(1+\frac{c}{n}\right) + \log\left(\frac{n}{c}+1\right) + 2\log s \\
&\leq \log\left(\frac{n}{c}+1\right) + \log(s^2 e), \qquad \text{(A.2)}
\end{aligned}
$$

where the third inequality is due to Lemma 13.5.4 of [44], the following equality is due to the relation $EL = \sum_{\ell,z,z'} \ell c_{\ell z z'}/c = n/c$, and the last inequality is due to an application of the inequality $\log(1+u) \leq u \log e$ for all $u > -1$. It follows that

$$
\begin{aligned}
\frac{\log|\mathcal{T}(\hat{P})|}{n} &\geq \rho_{\text{LZ}}(x^n) - \frac{c}{n}\log\left(\frac{n}{c}+1\right) - \frac{c}{n}\log(s^2 e) \\
&\geq \rho_{\text{LZ}}(x^n) - \frac{c}{n}\log\frac{n}{c} - \frac{c}{n}\log\left(1+\frac{c}{n}\right) - \frac{c}{n}\log(s^2 e) \\
&\geq \rho_{\text{LZ}}(x^n) - \frac{c}{n}\log\frac{n}{c} - \left(\frac{c}{n}\right)^2 \log e - \frac{c}{n}\log(s^2 e) \\
&= \rho_{\text{LZ}}(x^n) - \delta_s(n), \qquad \text{(A.3)}
\end{aligned}
$$

where

$$
\delta_s(n) \triangleq \frac{c}{n}\log\frac{n}{c} + \left(\frac{c}{n}\right)^2 \log e + \frac{c}{n}\log(s^2 e). \qquad \text{(A.4)}
$$

Since $\frac{c}{n} \leq \frac{\log\alpha}{\log n}(1+o(1))$ (see eq. (6) of [10] and Lemma 13.5.3 of [44] and reference therein), the second and the third terms of $\delta_s(n)$ are bounded by $O(1/\log^2 n)$ and $O(1/\log n)$, respectively. The first term of $\delta_s(n)$ is upper bounded by $O\left(\frac{\log(\log n)}{\log n}\right)$ (see eq. (13.124) of [44]).

References

[1] Kieffer, J. C.; Yang, E.-h. "Sequential codes, lossless compression of individual sequences, and Kolmogorov complexity," Technical Report 1993–3, Information Theory Research Group, University of Minnesota.

[2] Merhav, N.; Ziv, J. "On the Wyner–Ziv problem for individual sequences," *IEEE Trans. Inform. Theory* **2006**, vol. 52, no. 3, pp. 867–873.

[3] Reani, A.; Merhav, N. "Efficient on–line schemes for encoding individual sequences with side information at the decoder," *IEEE Trans. Inform. Theory* **2011**, vol. 57, no. 10, pp. 6860–6876.

[4] Weinberger, M. J.; Merhav, N.; Feder, M. "Optimal sequential probability assignment for individual sequences," *IEEE Trans. Inform. Theory* **1994**, vol. 40, no. 2, pp. 384-396.

[5] Weissman, T.; Merhav, N. "On limited–delay lossy coding and filtering of individual sequences," *IEEE Trans. Inform. Theory* **2002**, vol. 48, no. 3, pp. 721–733.

[6] Yang, E.-h.; Kieffer, J. C. "Simple universal lossy data compression schemes derived from the Lempel–Ziv algorithm," *IEEE Trans. Inform. theory* **1996**, vol. 42, pp. 239–245.

[7] Ziv, J. "Coding theorems for individual sequences," *IEEE Trans. Inform. Theory* **1978**, vol. IT–24, no. 4, pp. 405–412.

[8] Ziv, J. "Distortion–rate theory for individual sequences," *IEEE Trans. Inform. Theory* **1980**, vol. IT–26, no. 2, pp. 137–143.

[9] Ziv J. "Fixed-rate encoding of individual sequences with side information," *IEEE Transactions on Information Theory* **1984**, vol. IT–30, no. 2, pp. 348–452.

[10] Ziv, J.; Lempel, A. "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inform. Theory* **1978**, vol. IT–24, no. 5, pp. 530–536.

[11] Martín, A.; Merhav, N.; Seroussi, G.; Weinberger, M. J. "Twice–universal simulation of Markov sources and individual sequences," *IEEE Trans. Inform. Theory* **2010**, vol. 56, no. 9, pp. 4245–4255.

[12] Seroussi, G. "On universal types," *IEEE Trans. Inform. Theory* **2006**, vol. 52, no. 1, pp. 171–189.

[13] Merhav, N. "Universal detection of messages via finite–state channels," *IEEE Trans. Inform. Theory* **2000**, vol. 46, no. 6, pp. 2242–2246.

[14] Ziv, J. "Compression, tests for randomness, and estimating the statistical model of an individual sequence," *Proc. Sequences* **1990**, R. M. Capocelli Ed., New York: Springer Verlag, pp. 366–373.

[15] Ziv J.; Merhav, N. "A measure of relative entropy between individual sequences with application to universal classification," *IEEE Trans. Inform. Theory* **1993**, vol. 39, no. 4, pp. 1270–1279.

[16] Feder, M.; Merhav, N.; Gutman, M. "Universal prediction of individual sequences," *IEEE Trans. Inform. Theory* **1992**, vol. 38, no. 4, pp. 1258–1270.

[17] Haussler, D.; Kivinen, J.; Warmuth, M. K. "Sequential prediction of individual sequences under general loss functions," *IEEE Trans. Inform. Theory* **1998**, vol. 44, no. 5, pp. 1906–1925.

[18] Merhav, N.; Feder, M. "Universal schemes for sequential decision from individual data sequences," *IEEE Trans. Inform. Theory* **1993**, vol. 39, no. 4, pp. 1280–1291.

[19] Weissman, T.; Merhav, N. "Universal prediction of binary individual sequences in the presence of noise," *IEEE Trans. Inform. Theory* **2001**, vol. 47, no. 6, pp. 2151–2173.

[20] Weissman, T.; Merhav, N.; Somekh-Baruch, A. "Twofold universal prediction schemes for achieving the finite–state predictability of a noisy individual binary sequence," *IEEE Trans. Inform. Theory* **2001**, vol. 47, no. 5, pp. 1849-1866.

[21] Ziv, J.; Merhav, N. "On context–tree prediction of individual sequences," *IEEE Trans. Inform. Theory* **2007**, vol. 53, no. 5, pp. 1860–1866.

[22] Weissman, T.; Ordentlich, E.; Seroussi, G.; Verdú, S.; Weinberger, M. J. "Universal denoising: known channel," *IEEE Trans. Inform. Theory* **2005**, vol. 51, no. 1, pp. 5–28.

[23] Lomnitz, Y.; Feder, M. "Universal communication over individual channels," *IEEE Trans. Inform. Theory* **2011**, vol. 57, no. 11, pp. 7333-7358.

[24] Lomnitz,Y.; Feder, M. "Universal communication – part I: modulo additive channels," *IEEE Trans. Inform. Theory* **2013**, vol. 59, no. 9, pp. 5488-5510.

[25] Shayevitz, O.; Feder, M. "Communicating using feedback over a binary channel with arbitrary noise sequence," *Proc. ISIT 2005*, pp. 1516–1520, Adelaide, Australia, 2005.

[26] Merhav, N. "On Jacob Ziv's individual-sequence approach to information theory," to appear in *IEEE BITS the Information Theory Magazine*, 2025.

[27] Shannon, C. E. "Communication theory of secrecy systems," *Bell Systems Technical Journal* **1948**, vol. 27, pp. 479–523, (Part I); pp. 623–656, (Part II).

[28] Hellman, M. E. "An extension of the Shannon theory approach to cryptography," *IEEE Trans. Inform. Theory* **1997**, vol. IT–23, no. 3, pp. 289–294.

[29] Lempel, A. "Cryptology in transition," *Computing Surveys* **1979**, vol. 11, no. 4, pp. 285–303.

[30] Liang, Y.; Poor, H. V.; Shamai (Shitz), S. "Information theoretic security," *Foundations and Trends in Communications and Information Theory* **2005**, vol. 5, no. 4–5, pp. 355–580, 2009.

[31] Massey, J. L. "An introduction to contemporary cryptology," *Proc. IEEE* **1988**, vol. 76, no. 5, pp. 533–549.

[32] Yamamoto, H. "Information theory in cryptology," *IEICE Trans.* **1991**, vol. E74, no. 9, pp. 2456–2464.

[33] Ziv, J. "Perfect secrecy for individual sequences," unpublished manuscript, **1978**.

[34] Merhav, N. "Perfectly secure encryption of individual sequences," *IEEE Trans. Inform. Theory* **2013**, vol. 59, no. 3, pp. 1302–1310.

[35] Merhav, N. "Refinements and extensions of Ziv's model of perfect secrecy for individual sequences," *Entropy*, **2024**, 26(6), 503. `https://doi.org/10.3390/e26060503`
June 9, 2024.

[36] Issa, I.; Wagner, A. B.; Kamath, S. "An operational approach to information leakage," *IEEE Trans. Inform. Theory* **2020**, vol. 66, no. 3, pp. 1625-1657.

[37] Bloch, M.; Günlü, O.; Yener, A.; Oggier, F.; Poor, H. V.; Sankar, L.; Schaefer, R. F. "An overview of information-theoretic security and privacy: metrics, limits and applications," *IEEE J. Selected Areas in Inform. Theory* **2021**, vol. 2, no. 1, pp. 5–22.

[38] Esposito, A. R.; Gastpar, M.; Issa, I. "Generalization error bounds via Rényi-, f-divergences and maximal leakage," *IEEE Trans. Inform. Theory* **2021**, vol. 67, no. 8, pp. 4986–5004.

[39] Kurri, G. R.; Sankar, L.; Kosut, O. "An operational approach to information leakage via generalized gain functions," *IEEE Trans. Inform. Theory* **2024**, vol. 70, no. 2, pp. 1349–1375.

[40] Saeidian, S.; Cervia, G.; Oechtering, T. J.; Skoglund, M. "Pointwise maximal leakage," *IEEE Trans. Inform. Theory* **2023**, vol. 69, no. 12, pp. 8054–8080.

[41] Wu, Z.; Bai, L.; Zhou, L. "Successive refinement of Shannon cipher system under under maximal leakage," *IEEE Trans. Inform. Theory* **2025**, vol. 71, no. 3, pp. 1487–1503.

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

[43] Plotnik, E.; Weinberger, M. J.; Ziv, J. "Upper bounds on the probability of sequences emitted by finite-state sources and on the redundancy of the Lempel-Ziv algorithm," *IEEE Trans. Inform. Theory* **1992**, vol. 38, no. 1, pp. 66–72.

[44] Cover, T. M.; Thomas, J. A. *Elements of Information Theory*, John Wiley & Sons, Hoboken, New Jersey, 2006.