Refinements and Extensions of Ziv's Model of Perfect Secrecy for Individual Sequences

Neri Merhav

The Andrew & Erna Viterbi Faculty of Electrical and Computer Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 3200003, ISRAEL
E–mail: merhav@technion.ac.il

Abstract

We refine and extend Ziv's model and results regarding perfectly secure encryption of individual sequences. According to this model, the encrypter and the legitimate decrypter share in common a secret key, not shared with the unauthorized eavesdropper, who is aware of the encryption scheme and has some prior knowledge concerning the individual plaintext source sequence. This prior knowledge, combined with the cryptogram, is harnessed by eavesdropper which implements a finite-state machine as a mechanism for accepting or rejecting attempted guesses of the source plaintext. The encryption is considered perfectly secure if the cryptogram does not provide any new information to the eavesdropper that may enhance its knowledge concerning the plaintext beyond his prior knowledge. Ziv has shown that the key rate needed for perfect secrecy is essentially lower bounded by the finite-state compressibility of the plaintext sequence, a bound which is clearly asymptotically attained by Lempel-Ziv compression followed by one-time pad encryption. In this work, we consider some more general classes of finite-state eavesdroppers and derive the respective lower bounds on the key rates needed for perfect secrecy. These bounds are tighter and more refined than Ziv's bound and they are attained by encryption schemes that are based on different universal lossless compression schemes. We also extend our findings to the case where side information is available to the eavesdropper and the legitimate decrypter, but may or may not be available to the encrypter as well.

1 Introduction

Theoretical frameworks focusing on individual sequences and finite-state encoders and decoders have undergone extensive exploration, diverging from the conventional probabilistic paradigm used in modeling sources and channels. This divergence has been particularly evident across various information-theoretic domains, such as 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]. These references merely scratch the surface of a vast body of literature. Conversely, the realm of information-theoretic security, from Shannon's seminal contribution [26] to more contemporary research [27], [28], [29], [30], [31], remains almost totally entrenched in the probabilistic framework. While these works represent a mere fraction of the extensive literature, they exemplify the nearly exclusive reliance on probabilistic models within this field.

To the best of the author's knowledge, there are only two exceptions to this prevailing paradigm, documented in an unpublished memorandum by Ziv [32] and a subsequent work [33]. Ziv's memorandum presents a unique approach wherein the plaintext source, to be encrypted using a secret key, is treated as an individual sequence. The encrypter is modeled as a general block encoder, while the eavesdropper employs a finite-state machine (FSM) as a message discriminator. That memorandum postulates that the eavesdropper possesses certain prior knowledge about the plaintext, expressed as a set of "acceptable messages", hereafter referred to as the *acceptance set*. In other words, before observing the cryptogram, the eavesdropper uncertainty about the plaintext sequence is that it could be any member is in this set of acceptable messages.

This assumption about prior knowledge available to the eavesdropper is fairly realistic in real life. Consider, for example, the case where the plaintext alphabet is the latin alphabet, but the eavesdropper furthermore knows that the plaintext must be a piece of text in the Italian language. In this case, her prior knowledge, first and foremost, allows her to reject every candidate string of symbols that includes the letters 'j', 'k', 'w', 'x' and 'y', which are not used in Italian. Another example, which is common to English and some other languages, is that the letter 'q' must be followed by 'u'. In the same spirit, some additional rules of grammar can be invoked, like limitations on the number of successive consonant (or vowel) letters in a word, a limitation on the length of a word, and so on.

Now, according to Ziv's approach, perfectly secure encryption amounts to a situation where the presence of the cryptogram does not reduce the uncertainty associated with the acceptance set. In other words, having intercepted the cryptogram, the eavesdropper learns nothing about the plaintext that she did not not know before. The size of the acceptance set can be thought of as a quantifier of the level of uncertainty: a larger set implies greater uncertainty. The aforementioned

FSM is used to discriminate between acceptable and unacceptable strings of plaintext symbols that can be obtained by examining various key bit sequences. Accordingly, perfect security amounts to maintaining the size of the acceptance set unchanged, and consequently, the uncertainty level, in the presence of the cryptogram. The principal finding in Ziv's work is that the asymptotic key rate required 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]. Clearly, this lower bound is asymptotically achieved by employing one-time pad encryption (that is, bit-by-bit XOR with key bits) of the bit-stream obtained from LZ data compression of the plaintext source, mirroring Shannon's classical probabilistic result which asserts that the minimum required key rate equals the entropy rate of the source.

In the subsequent work [33], the concept of perfect secrecy for individual sequences was approached differently. Instead of a finite-state eavesdropper with predefined knowledge, it is assumed that the encrypter can be realized by an FSM which is sequentially fed by the plaintext source and random key bits. A notion of "finite-state encryptability" is introduced (in the spirit of the analogous finite-state compressibility of [10]), which designates the minimum key rate which must be consumed by any finite-state encrypter, such that probability law of the cryptogram would be independent of the plaintext input, and hence be perfectly secure. Among the main results of [33], it is asserted and proved that the finite-state encryptability of an individual sequence is essentially bounded from below by its finite-state compressibility, a bound which is once again attained asymptotically by LZ compression followed by one-time pad encryption.

In this work, we revisit Ziv's approach to perfect secrecy for individual sequences [32]. After presenting his paradigm in detail, we proceed to refine and generalize his findings in certain aspects. First, we consider several more general classes of finite-state discriminators that can be employed by the eavesdropper. These will lead to tight lower bounds on the minimum key rate to be consumed by the encrypter, which will be matched by encryption schemes that are based some other universal data compression schemes. The resulting gaps between the lower bounds and the corresponding upper bounds (i.e., the redundancy rates) would converge faster. Among these more general classes of finite-state machines, we will consider finite-state machines that are equipped with counters, as well as periodically time-varying finite-state machines with counters. Another direction of generalizing Ziv's findings is the incorporation of side information (SI) that is available

both at the eavesdropper and the legitimate decrypter, but may or may not be available at the encrypter as well.

The outline of this article is as follows. In Section 2, we formulate the model setting, establish the notation, and provide a more detailed background on Ziv's model and results in [32]. In Section 3, which is the main section of this article, we present the refinements and extensions to other types of FSMs, including FSMs with counters (Subsection 3.1), shift-register FSMs with counters (Subsection 3.2), and periodically time-varying FSMs with counters (Subsection 3.3). Finally, in Section 4, we further extend some of our findings to the case where SI is available at both the legitimate decrypter and the eavesdropper, but not necessarily at the encrypter.

2 Formulation, Notation and Background

Consider the following version of Shannon's cipher system model, adapted to the entryption of individual sequences, as proposed by Ziv [32]. An individual (deterministic) plaintext sequence, $\boldsymbol{x} = (x_0, \ldots, x_{n-1})$ (n - positive integer), is encrypted using a random key K, whose entropy is $H(K)$, to generate a cryptogram, $W = T(\boldsymbol{x}, K)$, where the mapping $T(\cdot, K)$ is invertible given K, namely, $\boldsymbol{x}$ can be reconstructed by the legitimate decoder, who has access to K, by applying the inverse function, $\boldsymbol{x} = T^{-1}(W, K)$. The plaintext symbols, x_i, $i = 0, 1, 2, \ldots, n-1$, take on values in a finite alphabet, $\mathcal{X}$, of size α. Thus, $\boldsymbol{x}$ is a member of $\mathcal{X}^n$, the n-th Cartesian power of $\mathcal{X}$, whose cardinality is α^n. Without essential loss of generality, we assume that K is a uniformly distributed random variable taking on values in a set $\mathcal{K}$ whose cardinality is $2^{H(K)}$. Sometimes, it may be convenient to consider $\mathcal{K}$ to be the set $\{0, 1, \ldots, 2^{H(K)} - 1\}$. A specific realization of the key, K, will be denoted by k.

An eavesdropper, who knows the mapping T, but not the realization of the key, K, is in the quest of learning as much as possible about $\boldsymbol{x}$ upon observing W. It is assumed that the eavesdropper also has some prior knowledge about $\boldsymbol{x}$, even before observing W. In particular, the eavesdropper knows that the plaintext source string $\boldsymbol{x}$ must be a member of a certain subset of $\mathcal{X}^n$, denoted $\mathcal{A}_n$, which is referred to as the *acceptance set*.

Ziv models the eavesdropper by a cascade of a guessing decrypter and a finite-state message discriminator, which work together as follows. At each step, the eavesdropper examines a certain

key, $k \in \mathcal{K}$, by generating an estimated plaintext, $\hat{\boldsymbol{x}} = T^{-1}(W, k)$, and then feeding $\hat{\boldsymbol{x}}$ into the message discriminator to examine whether or not $\hat{\boldsymbol{x}} \in \mathcal{A}_n$. If the answer is affirmative, $\hat{\boldsymbol{x}}$ is accepted as a candidate, otherwise, it is rejected. Upon completing this step, the eavesdropper moves on to the next key, $k+1$, and repeats the same process, etc. The message discriminator is modeled as a finite-state machine, which implements the following recursive equations for $i = 0, 1, 2, \ldots, n-1$:

$$u_i = f(z_i, \hat{x}_i) \tag{1}$$

$$z_{i+1} = g(z_i, \hat{x}_i), \tag{2}$$

where $z_0, z_1, z_2, \ldots, z_{n-1}$ is a sequence of states, $z_i \in \mathcal{S}$, $i = 0, 1, 2, \ldots, n-1$, $\mathcal{S}$ being a set of s states (and with the initial state, z_0, as a fixed member of $\mathcal{S}$), $u_0, u_1, \ldots, u_{n-1}$ is a binary output sequence, $f : \mathcal{S} \times \mathcal{X} \to \{0, 1\}$ is the output function, and $g : \mathcal{S} \times \mathcal{X} \to \mathcal{S}$ is the next-state function. If $u_0, u_1, u_2, \ldots, u_{n-1}$ is the all-zero sequence, $\hat{\boldsymbol{x}}$ is accepted, namely, $\hat{\boldsymbol{x}} \in \mathcal{A}_n$, otherwise, as soon as $u_i = 1$ for some $0 \le i \le n-1$, $\hat{\boldsymbol{x}}$ is rejected. In other words, $\mathcal{A}_n$ is defined to be the set of all $\{\hat{\boldsymbol{x}}\}$ for which the response of the finite-state discriminator is the all-zero sequence, $\boldsymbol{u} = (0, 0, \ldots, 0)$.

Example 1. Let $\mathcal{X} = \{0, 1\}$ and suppose that membership of $\boldsymbol{x}$ in $\mathcal{A}_n$ forbids the appearance of more than two successive zeroes. Then, a simple discriminator can detect the violence of this rule using the finite-state machine defined by $\mathcal{S} = \{0, 1, 2\}$ and

$$g(z, \hat{x}) = \begin{cases} (z+1) \bmod 3 & \hat{x} = 0 \\ 0 & \hat{x} = 1 \end{cases} \tag{3}$$

$$f(z, \hat{x}) = \begin{cases} 1 & z = 2 \text{ and } \hat{x} = 0 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

More generally, consider the set of binary sequences that comply with the so called (d, k)-constraints, well known from the literature of magnetic recording (see, e.g., [34] and references therein), namely, binary sequences, where the runs of successive zeroes must be of length at least d and at most k, where d and k (not to be confused with the notation of the encryption key) are positive integers with $d \le k$. We shall return to this example later. □

Ziv defines perfect secrecy for individual sequences as a situation where even upon observing W, the eavesdropper's uncertainty about $\boldsymbol{x}$ is not reduced. In the mathematical language, let us

denote

$$T^{-1}(W) \triangleq \{T^{-1}(W,k),\ k \in \mathcal{K}\}, \tag{5}$$

and

$$\mathcal{A}_n(W) \triangleq \mathcal{A}_n \bigcap T^{-1}(W), \tag{6}$$

then, perfect secrecy is defined as a situation where

$$A_n(W) = A_n, \tag{7}$$

or equivalently,

$$\mathcal{A}_n \subseteq T^{-1}(W). \tag{8}$$

To demonstrate these concepts, consider the following example.

Example 2. Let $n = 4$, $\mathcal{X} = \{0, 1\}$, $\boldsymbol{x} = (1111)$, $2^{H(K)} = 8$, $k = (1111)$, and then for a one-time pad encrypter,

$$W = T(\boldsymbol{x}, k) = \boldsymbol{x} \oplus k = (1111) \oplus (1111) = (0000), \tag{9}$$

where $\oplus$ denotes bit-wise XOR (modulo 2 addition). Let the set $\mathcal{K}$ of all 8 possible key strings be given by

$$\mathcal{K} = \left\{ \begin{array}{c} 1111 \\ 1000 \\ 1100 \\ 1001 \\ 0000 \\ 0111 \\ 0011 \\ 0110 \end{array} \right\}. \tag{10}$$

Obviously, the decryption is given by $T^{-1}(W, k) = W \oplus k$. Since $W = (0000)$, then $T^{-1}(W) = \mathcal{K}$. Following Example 1, suppose that $\mathcal{A}_4$ is the set of all binary vectors of length $n = 4$, which do not contain runs of more than two zeroes. There are only 3 binary vectors of length 4 that contain a succession of more than 2 (i.e., 3 or 4) zeroes, namely, (0000), (1000), and (0001). Thus, $|\mathcal{A}_4| = 2^4 - 3 = 13$. On the other hand,

$$|\mathcal{A}_4(W)| = |\mathcal{A}_4 \bigcap T^{-1}(W)| \leq |T^{-1}(W)| = |\mathcal{K}| = 8 < 13, \tag{11}$$

which means that this encryption system is not perfectly secure. The reason is that the key space, $\mathcal{K}$, is not large enough. $\square$

Clearly, the best one can do in the quest of minimizing $H(K)$, without compromising perfect secrecy, is to design the encrypter in such a way that

$$T^{-1}(W) = \mathcal{A}_n, \tag{12}$$

for every cryptogram W that can possibly be obtained from some combination of $\boldsymbol{x}$ and k. Conceptually, this can be obtained by mapping $\mathcal{A}_n$ to the set of all binary sequences of length $H(K)$ by means of a fixed-rate data compression scheme and applying one-time pad encryption to the compressed sequence. Here, and throughout the sequel, we neglect integer length constraints associated with large numbers, and so, $H(K)$ is assumed integer without essential loss of generality and optimality.

Remark 1. For readers familiar with the concepts and the terminology of coding for systems with (d, k) constraints (and other readers may skip this remark without loss of continuity), it is insightful to revisit the second part of Example 1: If $\mathcal{A}_n$ is the set of binary n-sequences that satisfy a certain (d, k) constraint, then optimal encryption for $\mathcal{A}_n$ pertains to compression using the inverse mapping of a channel encoder for the same (d, k) constraint, namely, the corresponding channel decoder, which is followed by one-time pad encryption. The minimum key rate needed is then equal to the capacity of the constrained system, which can be calculated either algebraically, as the logarithm of the Perron-Frobenius eigenvalue of the state adjacency matrix of the state transition diagram, or probabilistcally, as the maximum entropy among all stationary Markov chains that are supported by the corresponding state transition graph [34]. $\square$

Ziv's main result in [32] is that for a finite-state discriminator, if $\boldsymbol{x} \in \mathcal{A}_n$, the cardinality of $\mathcal{A}_n$ cannot be exponentially smaller than $2^{LZ(\boldsymbol{x})}$ (see Appendix for the proof), where $LZ(\boldsymbol{x})$ is the length (in bits) of the compressed version of $\boldsymbol{x}$ using the 1978 version of the Lempel-Ziv algorithm (the LZ78 algorithm) [10], and so, the key rate needed in order to completely encrypt all members

of $\mathcal{A}_n$ is lower bounded by

$$\begin{aligned} R &\triangleq \frac{H(K)}{n} \\ &= \frac{\log|T^{-1}(W)|}{n} \\ &\geq \frac{\log|\mathcal{A}_n|}{n} \\ &\geq \frac{LZ(\boldsymbol{x})}{n} - \epsilon_n \end{aligned} \tag{13}$$

where ϵ_n is a positive sequence tending to zero as $n \to \infty$ at the rate of $\frac{\log(\log n)}{\log n}$, and where here and throughout the sequel, the notation $|\mathcal{E}|$ for a finite set $\mathcal{E}$, designates the cardinality of $\mathcal{E}$. The first inequality of (13) follows from (8). Obviously, this bound is essentially attained by LZ78 compression of $\boldsymbol{x}$, followed by one-time pad encryption using $LZ(\boldsymbol{x})$ key bits. As can be seen, the gap, ϵ_n, between the upper bound and the lower bound on R is $O\left(\frac{\log(\log n)}{\log n}\right)$, which tends to zero rather slowly.

Remark 2. For an infinite sequence $\boldsymbol{x} = x_0, x_1, x_2, \ldots$, asymptotic results are obtained in [32] by a two-stage limit: First, consider a finite sequence of total length $m \cdot n$, which is divided into m non-overlapping n-blocks, where the above described mechanism is applied to each n-block separately. The asymptotic minimum key rate is obtained by a double limit superior, which is taken first, for $m \to \infty$ for a given n, and then for $n \to \infty$. In this work, we will have in mind a similar double limit, but we shall not mention it explicitly at every relevant occasion. Instead, we will focus on the behavior of a single n-block.

3 More General Finite-State Discriminators

This section, which is the main section in this article, is devoted to describe several more general classes of finite-state discriminators along with derivations of their respective more refined bounds.

3.1 FSMs with Counters

While Ziv's model for a finite-state discriminator is adequate for rejecting sequences with certain forbidden patterns (like a succession of more than two zeroes in the above examples), it is not sufficient to handle situations like the following one. Suppose that encrypter applies a universal lossless

compression algorithm for memoryless sources, followed by one-time pad encryption. Suppose also that the universal compression scheme is a two-part code, where the first part encodes the index of the type class of $\boldsymbol{x}$, using a number of bits that is proportional to $\log n$, and the second part represents the index of $\boldsymbol{x}$ within the type class, assuming that the encoder and the decoder have agreed on some ordering ahead of time. In this case, the length of the cryptogram (in bits), which is equal to the length of the compressed data, is about

$$L = H(K) \approx n\hat{H}(\boldsymbol{x}) + \frac{\alpha - 1}{2} \cdot \log n, \tag{14}$$

where $\hat{H}(\boldsymbol{x})$ is the empirical entropy of $\boldsymbol{x}$ (see, e.g., [5] and references therein). The eavesdropper, being aware of the encryption scheme, observes the length L of the cryptogram, and immediately concludes that $\boldsymbol{x}$ must be a sequence whose empirical entropy is

$$H_0 = \frac{1}{n} \cdot \left(L - \frac{\alpha - 1}{2} \cdot \log n \right). \tag{15}$$

In other words, in this case,

$$\mathcal{A}_n = \left\{ \hat{\boldsymbol{x}} : \ \hat{H}(\hat{\boldsymbol{x}}) = H_0 \right\}. \tag{16}$$

Therefore, every sequence whose empirical distribution pertains to empirical entropy different from H_0 should be rejected. To this end, our discriminator should be able to gather empirical statistics, namely, to count occurrences of symbols (or more generally, count combinations of symbols and states) and not just to detect a forbidden pattern that might have occurred just once in $\boldsymbol{x}$.

This motivates us to broaden the class of finite-state discriminators to be considered, in the following fashion. We consider discriminators that consist of a next-state function,

$$z_{i+1} = g(z_i, \hat{x}_i), \tag{17}$$

as before, but instead of the binary output function, f, of [32], these discriminators are equipped with a set of $\alpha \cdot s$ counters that count the number of joint occurrences of all $(x, z) \in \mathcal{X} \times \mathcal{S}$ for $i = 0, 1, 2, \ldots, n-1$, i.e.,

$$n(x, z) = \sum_{i=1}^{n} \mathcal{I}\{\hat{x}_i = x, \ z_i = z\}, \qquad x \in \mathcal{X}, \ z \in \mathcal{S}, \tag{18}$$

where $\mathcal{I}\{A\}$, for a generic event A, denotes its indicator function, namely, $\mathcal{I}\{A\} = 1$ if A is true and $I\{A\} = 0$ if not. A sequence $\hat{\boldsymbol{x}}$ is accepted (resp. rejected) if the matrix of counts,

$\{n(x,z),\ x\in\mathcal{X},\ z\in\mathcal{S}\}$, satisfies (resp. violates) a certain condition. In the example of the previous paragraph, a sequence is accepted if

$$\hat{H}(\hat{\boldsymbol{x}})=\sum_{x\in\mathcal{X}}\frac{n(x)}{n}\log\left(\frac{n}{n(x)}\right)=H_0, \tag{19}$$

where $n(x)=\sum_{z\in\mathcal{S}}n(x,z)$. Ziv's discriminator model is clearly a special case of this model: let $(x_\star,z_\star)$ be any combination of input and state such that $f(z_\star,x_\star)=1$ in Ziv's model (that is, a "forbidden" combination). Then, in terms of the proposed extended model, a sequence is rejected whenever $n(x_\star,z_\star)\geq 1$.

Clearly, since $\boldsymbol{x}\in\mathcal{A}_n$ and since membership in $\mathcal{A}_n$ depends solely on the matrix of counts, $\{n(x,z),\ x\in\mathcal{X},\ z\in\mathcal{S}\}$, it follows that all $\{\hat{\boldsymbol{x}}\}$ that share the same counts as these of $\boldsymbol{x}$ must also be members of $\mathcal{A}_n$. The set of all $\{\hat{\boldsymbol{x}}\}$ of length n with the same counts, $\{n(x,z),\ x\in\mathcal{X},\ z\in\mathcal{S}\}$, as these of $\boldsymbol{x}$, is called the *finite-state type class* w.r.t. the FSM g (see also [5]), and it is denoted by $\mathcal{T}_g(\boldsymbol{x})$. Since $\mathcal{A}_n\supseteq\mathcal{T}_g(\boldsymbol{x})$,

$$|\mathcal{A}_n|\geq|\mathcal{T}_g(\boldsymbol{x})|. \tag{20}$$

It is proved in [5] (Lemma 3 therein), that if $n(x,z)\geq n(z)\delta(n)$ for every $(x,z)\in\mathcal{X}\times\mathcal{S}$, where $\delta(n)>0$ may even be a vanishing sequence, then

$$|\mathcal{T}_g(\boldsymbol{x})|\geq\exp_2\left\{n\hat{H}(X|Z)-\frac{s(\alpha-1)}{2}\cdot\log(2\pi n)\right\}, \tag{21}$$

where

$$\hat{H}(X|Z)=-\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{S}}\frac{n(x,z)}{n}\log\frac{n(x,z)}{n(z)}, \tag{22}$$

with $n(z)=\sum_{x\in\mathcal{X}}n(x,z)$ and with the conventions that $0\log 0\stackrel{\Delta}{=}0$ and $0/0\stackrel{\Delta}{=}0$. It therefore follows that the key rate needed for perfect secrecy is lower bounded by

$$\begin{aligned}R &= \frac{H(K)}{n}\\ &\geq \frac{\log|\mathcal{A}_n|}{n}\\ &\geq \frac{\log|\mathcal{T}_g(\boldsymbol{x})|}{n}\\ &\geq \hat{H}(X|Z)-\frac{s(\alpha-1)}{2}\cdot\frac{\log(2\pi n)}{n}.\end{aligned} \tag{23}$$

This lower bound can be asymptotically attained by an encrypter that applies universal loss compression for finite-state sources with the next state function g, followed by one-time pad encryption.

This universal lossless compression scheme is based on a conceptually simple extension of the above-mentioned universal scheme for the class of memoryless sources: one applies a two-part code where the first part includes a code for the index of the type class w.r.t. g (using a number of bits that is proportional to $\log n$), and the second part encodes the index of the location of $\hat{\boldsymbol{x}}$ within $\mathcal{T}_g(\boldsymbol{x})$ according to a predefined order agreed between the encoder and the decoder (see [5] for more details). More precisely, the compression ratio that can be achieved, which is also the key rate consumed, is upper bounded by (see Section III of [5]):

$$R \leq \hat{H}(X|Z) + \frac{s(\alpha-1)}{2} \cdot \frac{\log n}{n} + O\left(\frac{1}{n}\right), \tag{24}$$

which tells us that the gap between the lower bound and the achievability is proportional to $\frac{\log n}{n}$, which decays much faster than the aforementioned $O\left(\frac{\log(\log n)}{\log n}\right)$ convergence rate of the gap in Ziv's approach. Moreover, for most sequences in most type classes, $\{\mathcal{T}_g(\boldsymbol{x})\}$, the coding rate (which is also the key rate in one-time-pad encryption) of $\hat{H}(X|Z) + \frac{s(\alpha-1)}{2} \cdot \frac{\log n}{n}$ is smaller than $LZ(\boldsymbol{x})/n$ since the former quantity is essentially also a lower bound to the compression ratio of any lossless compression scheme for most individual sequences in $\mathcal{T}_g(\boldsymbol{x})$, for almost all such type classes [5, Theorem 1]. The converse inequality between $\hat{H}(X|Z)$ and $LZ(\boldsymbol{x})/n$, which follows from Ziv's inequality (see [35, Lemma 13.5.5] and [36]) holds up to an $O\left(\frac{\log(\log n)}{\log n}\right)$ term again.

Remark 3. To put Ziv's result in perspective, one may wish to envision a class of discriminators defined by a given dictionary of c distinct 'words', which are the c phrases in Ziv's derivation [32] (see also Appendix). A possible definition of such a discriminator is that it accepts only n-sequences of plaintext that are formed by concatenating words from the given dictionary, allowing repetitions. These are no longer finite-state discriminators. In this case, Ziv's derivation is applicable under the minor modification that the various phrases should be classified only in terms of their length, but without further partition according to initial and final states (z and z' in the derivation of the appendix). This is equivalent to assuming that $s = 1$, and accordingly, the term $\frac{2c \log s}{n}$ in the last equation of the appendix should be omitted. Of course, the resulting bound can still be matched by LZ78 compression followed by one time-pad encryption.

3.2 Shift-Register Machines with Counters

Since the eavesdropper naturally does not cooperate with the encrypter, the latter might not know the particular FSM, g, used by the former, and therefore, it is instructive to derive key-rate bounds that are independent of g. To this end, consider the following. Given $\boldsymbol{x}$ and a fixed positive integer ℓ ($\ell \ll n$), let us observe the more refined joint empirical distribution,

$$\hat{P}_{X^{\ell+1}Z^{\ell+1}}(a_0,\ldots,a_\ell,s_0,\ldots,s_\ell) = \frac{1}{n}\sum_{i=0}^{n-1}\mathcal{I}\{x_i = a_0,\ldots,x_{i\oplus\ell}=a_\ell, z_i = s_0,\ldots,z_{i\oplus\ell}=s_\ell\}, \quad (25)$$

as well as all partial marginalizations derived from this distribution, where here, $\oplus$ denotes addition modulo n so to create a periodic extension of $\boldsymbol{x}$ (hence redefining $z_0 = g(z_{n-1}, x_{n-1})$). Accordingly, the previously defined empirical conditional entropy, $\hat{H}(X|Z)$ is now denoted $\hat{H}(X_1|Z_1)$, which is also equal to $\hat{H}(X_2|Z_2)$, etc., due to the inherent shift-invariance property of the empirical joint distribution extracted under the periodic extension of $\boldsymbol{x}$. Consider now the following chain of inequalities:

$$\begin{aligned}
& \hat{H}(X_\ell|X_0,X_1,\ldots,X_{\ell-1}) - \hat{H}(X_\ell|Z_\ell) \\
\leq\ & \frac{1}{\ell+1}\sum_{j=0}^{\ell}[\hat{H}(X_j|X_0,\ldots,X_{j-1}) - \hat{H}(X_j|X_0,\ldots,X_{j-1},Z_0)] && (26)\\
=\ & \frac{1}{\ell+1}[\hat{H}(X_0,\ldots,X_\ell) - \hat{H}(X_0,\ldots,X_\ell|Z_0)] && (27)\\
=\ & \frac{\hat{I}(Z_0;X_0,\ldots,X_\ell)}{\ell+1} && (28)\\
\leq\ & \frac{\hat{H}(Z_0)}{\ell+1} && (29)\\
\leq\ & \frac{\log s}{\ell+1}, && (30)
\end{aligned}$$

where $\hat{I}(\cdot;\cdot)$ denotes empirical mutual information and where in the second line, the term corresponding to $j=0$ should be understood to be $[\hat{H}(X_0) - \hat{H}(X_0|Z_0)]$. The first inequality follows because given $(X_0,\ldots,X_{j-1},Z_0)$, one can reconstruct $Z_1,Z_2,\ldots,Z_j$ by j recursive applications of the next-state function, g. Therefore,

$$\begin{aligned}
\hat{H}(X_j|X_0,\ldots,X_{j-1},Z_0) &= \hat{H}(X_j|X_0,\ldots,X_{j-1},Z_0,Z_1,\ldots,Z_j)\\
&\leq \hat{H}(X_j|Z_j) && (31)\\
&= \hat{H}(X_\ell|Z_\ell). && (32)
\end{aligned}$$

Equivalently,

$$\hat{H}(X_\ell|Z_\ell) \geq \hat{H}(X_\ell|X_0,\ldots,X_{\ell-1}) - \frac{\log s}{\ell+1}, \tag{33}$$

and so, combining this with eqs. (20) and (21), we get

$$\log|\mathcal{A}_n| \geq n\left[\hat{H}(X_\ell|X_0,\ldots,X_{\ell-1}) - \frac{\log s}{\ell+1}\right] - \frac{s(\alpha-1)}{2}\cdot\log(2\pi n). \tag{34}$$

The advantage of this inequality is in its independence upon the arbitrary next-state function g. In fact, we actually replaced the arbitrary FSM, g, by a particular FSM – the shift-register FSM, whose state is $z_i = (x_{i-\ell}, x_{i-\ell+1}, \ldots, x_{i-1})$ at the cost of a gap of $\frac{\log s}{\ell+1}$, which can be kept arbitrarily small if the size of the shift-register, ℓ, is sufficiently large compared to the memory size, $\log s$, of g.

Remark 4. The fact that $\hat{H}(X_\ell|X_0,\ldots,X_{\ell-1})$ cannot be much larger than $\hat{H}(X_\ell|Z_\ell)$ for large enough ℓ actually suggests that whatever the state of any FSM g can possibly "remember" from the past of $\boldsymbol{x}$ is essentially captured by the recent past, and not by the remote past. While this is not surprising in the context of the probabilistic setting, especially if the underlying random process is ergodic and hence has a vanishing memory of the remote past, this finding is not quite trivial when it comes to arbitrary individual sequences. □

Returning to our derivations, in view of the first three lines of (13), the key rate needed for perfect secrecy is lower bounded by

$$R \geq \frac{\log|\mathcal{A}_n|}{n} \geq \hat{H}(X_\ell|X_0,\ldots,X_{\ell-1}) - \frac{\log s}{\ell+1} - \frac{s(\alpha-1)}{2n}\cdot\log(2\pi n), \tag{35}$$

and since this holds for any ℓ in some fixed range $1 \leq \ell \leq l$ (i.e., where l is independent of n),

$$R \geq \max_{1\leq\ell\leq l}\left[\hat{H}(X_\ell|X_0,\ldots,X_{\ell-1}) - \frac{\log s}{\ell+1}\right] - \frac{s(\alpha-1)}{2n}\cdot\log(2\pi n), \tag{36}$$

Note that if $\boldsymbol{x}$ is an "ℓ_0-th order Markovian sequence" in the sense that $\hat{H}(X_\ell|X_0,\ldots,X_{\ell-1})$ is almost fixed for all $\ell_0 \leq \ell \leq l$ with $l \gg \log s$, then ℓ_0 is the preferred choice for ℓ as it essentially captures the best attainable key rate.

This lower bound can be asymptotically attained by universal lossless compression for ℓ-th order Markov types [37], [38], [39, Section VII.A], followed by one-time pad encryption, where the

achieved rate is

$$R \leq \hat{H}(X_\ell | X_0, \ldots, X_{\ell-1}) + \frac{\alpha^{\ell-1}(\alpha-1)}{2} \cdot \frac{\log n}{n} + O\left(\frac{1}{n}\right). \tag{37}$$

In this case, $\mathcal{A}_n$ is the ℓ-th order Markov type of $\boldsymbol{x}$ and the finite-state discriminator of the eavesdropper is a shift-register machine, that checks whether the ℓ-th order Markov type of each $\hat{\boldsymbol{x}}$ has the matching conditional empirical conditional entropy of order ℓ.

In this result, there is a compatibility between the converse bound and the achievability bound in the sense that both are given in terms of FSMs with a fixed number of states that does not grow with n. Among all possible finite-state machines, we have actually single out the shift-register machine universally, at the cost of a controllable asymptotic gap of $\frac{\log s}{\ell+1}$, but otherwise, the bound is explicit and it is clear how to approach it. If we wish to keep this gap below a given $\epsilon > 0$, then we select $\ell = \lceil (\log s)/\epsilon \rceil - 1$. In this sense, it is another refinement of Ziv's result.

3.3 Periodically Time-Varying FSMs with Counters

So far, we have considered time-invariance FSM, where the function g remains fixed over time. We now expand the scope to consider the class of discriminators implementable by periodically time-varying FSMs, defined as follows:

$$z_{i+1} \quad = \quad g(z_i, \hat{x}_i, i \bmod l), \tag{38}$$

where $i = 0, 1, 2, \ldots$ and l is a positive integer that designates the period of the time-varying finite-state machine. Conceptually, this is not really more general than the ordinary, time-invariant FSM defined earlier, because the state of the modulo-ℓ clock can be considered part of the entire state, in other words, this is a time-invariant FSM with $s \cdot l$ states, indexed by the ordered pair $(z_i, i \bmod l)$. The reason it makes sense to distinguish between the state z_t and the state of the clock is because the clock does not store any information regarding past input data. This is to say that in the context of time-varying finite-state machines, we distinguish between the amount of memory of past input ($\log s$ bits) and the period l. Both parameters manifest the richness of the class of machines, but in different manners. Indeed, the parameters s and l will play very different and separate roles in the converse bound to be derived below.

Remark 5. Clearly, the earlier considered time-invariant finite-state machine is obtained as the special case pertaining to $l = 1$, or to the case where l is arbitrary, but the next-state functions, $g(\cdot,\cdot,0), g(\cdot,\cdot,1), \ldots, g(\cdot,\cdot,l-1)$, are all identical. $\square$

First, observe that a periodic FSM with period l can be viewed as time-invariant FSM in the level of l-blocks, $\{x_{il}, x_{il+1}, \ldots, x_{il+l-1}\}$, $i = 0, 1, \ldots$, and hence also in the level of ℓ-blocks where ℓ is an integer multiple of l. Accordingly, let ℓ be an arbitrary integer multiple of l, but at the same time, assume that ℓ divides n. Denote $m = n/\ell$, and define the counts,

$$m(z, z', x^\ell) = \sum_{i=0}^{n/\ell-1} \mathcal{I}\{z_{i\ell} = z,\ z_{i\ell+\ell} = z',\ \hat{x}_{i\ell}^{i\ell+\ell-1} = x^\ell\}, \qquad z, z' \in \mathcal{S},\ x^\ell \in \mathcal{X}^\ell. \tag{39}$$

In fact, there is a certain redundancy in this definition because z' is a deterministic function of (z, x^ℓ) obtained by ℓ recursive applications of the (time-varying) next-state function g. Concretely, $m(z, z', x^\ell) = m(z, x^\ell)$ iff z' matches (z, x^ℓ) and $m(z, z', x^\ell) = 0$ otherwise. Nonetheless, we adopt this definition for the sake of clarity of combinatorial derivation to be carried out shortly. In particular, our derivation will be based on grouping together all $\{x^\ell\}$, which for a given z, yield the same z'. Accordingly, we also denote $m(z, z') = \sum_{x^\ell \in \mathcal{X}^\ell} m(z, z', x^\ell)$.

Suppose that the acceptance/rejection criterion that defines $\mathcal{A}_n$ is based on the counts, $\{m(z, z', x^\ell),\ z, z' \in \mathcal{S},\ x^\ell \in \mathcal{X}^\ell\}$, and then the smallest $\mathcal{A}_n$ that contains $\boldsymbol{x}$ is the type class of $\boldsymbol{x}$ pertaining to $\{m(z, z', x^\ell),\ z, z' \in \mathcal{S},\ x^\ell \in \mathcal{X}^\ell\}$. The various sequences in this type class are obtained by permuting distinct ℓ-tuples $\{\hat{x}_{i\ell}^{i\ell+\ell-1},\ i = 0, 1, \ldots, m-1\}$ that begin at the same state, z, and end at the same state, z'. Let us define the empirical distribution,

$$\hat{P}(z, z', x^\ell) = \frac{m(z, z', x^\ell)}{m}, \qquad z \in \mathcal{S},\ a^\ell \in \mathcal{X}^\ell, \tag{40}$$

and the joint entropy,

$$\hat{H}(Z, Z', X^\ell) = -\sum_{z, z', x^\ell} \hat{P}(z, z', x^\ell) \log \hat{P}(z, z', x^\ell), \tag{41}$$

and let

$$\hat{H}(X^\ell | Z, Z') = \hat{H}(Z, Z', X^\ell) - \hat{H}(Z, Z'), \tag{42}$$

where $\hat{H}(Z, Z')$ is the marginal empirical entropy of (Z, Z'). Then, using the method of types [39],

we have:

$$
\begin{aligned}
|\mathcal{A}_n| &\geq \prod_{z,z'\in\mathcal{S}} \frac{m(z,z')!}{\prod_{x^\ell\in\mathcal{X}^\ell} m(z,z',x^\ell)!} && (43)\\
&\geq \prod_{z,z'\in\mathcal{S}} \left[(m+1)^{-\alpha^\ell}\cdot \exp_2\left\{\sum_{x^\ell} m(z,z',x^\ell)\log\frac{m(z,z')}{m(z,z',x^\ell)}\right\}\right] && (44)\\
&\geq (m+1)^{-s^2\alpha^\ell}\cdot 2^{m\hat{H}(X^\ell|Z,Z')} && (45)\\
&= (m+1)^{-s^2\alpha^\ell}\cdot 2^{m[\hat{H}(X^\ell)-I(Z,Z';X^\ell)]} && (46)\\
&\geq (m+1)^{-s^2\alpha^\ell}\cdot 2^{m[\hat{H}(X^\ell)-H(Z,Z')]} && (47)\\
&\geq (m+1)^{-s^2\alpha^\ell}\cdot 2^{m[\hat{H}(X^\ell)-2\log s]} && (48)\\
&= \exp_2\left\{n\left[\frac{\hat{H}(X^\ell)}{\ell}-\frac{2\log s}{\ell}-\frac{\ell s^2\alpha^\ell}{n}\log\left(\frac{n}{\ell}+1\right)\right]\right\}, && (49)
\end{aligned}
$$

and so,

$$
R\geq\frac{\log|\mathcal{A}_n|}{n}\geq\frac{\hat{H}(X^\ell)}{\ell}-\frac{2\log s}{\ell}-\frac{\ell s^2\alpha^\ell}{n}\log\left(\frac{n}{\ell}+1\right), \tag{50}
$$

which, for $\ell\gg 2\log s$, can be essentially matched by universal compression for block-memoryless sources and one-time pad encryption using exactly the same ideas as before. Once again, we have derived a lower bound that is free of dependence on the particular FSM, g. Recall that ℓ divides n and that it is also a multiple of l, but otherwise, ℓ is arbitrary. Hence we may maximize this lower bound w.r.t. ℓ subject to these constraints. Alternatively, we may rewrite the lower bound as

$$
R\geq\max_{\{q \text{ divides } n/l\}}\left\{\frac{\hat{H}(X^{ql})}{ql}-\frac{2\log s}{ql}-\frac{qls^2\alpha^{ql}}{n}\log\left(\frac{n}{ql}+1\right)\right\}. \tag{51}
$$

Clearly, in view of Remark 5, the above lower bound applies also to the case of a time-invariant FSM, g, but then there would be some mismatch between the upper and lower bound because to achieve the lower bound, one must gather more detailed empirical statistics, namely, empirical statistics of blocks together with states, rather than just single symbols with states.

4 Side Information

Some of the results presented in the previous sections extend to the case where side information (SI) is available at the legitimate decrypter and at the eavesdropper. In principle, it may or may not be available to the encrypter, and we consider first the case where it is available. We assume the

SI sequence to be of length n, and denote it by $\boldsymbol{y} = (y_0, y_1, \ldots, y_{n-1})$, $y_i \in \mathcal{Y}$, $i = 0, 1, \ldots, n-1$. It is related to the plaintext, $\boldsymbol{x}$, to be encrypted, but like $\boldsymbol{x}$, it is a deterministic, individual sequence. The SI alphabet $\mathcal{Y}$ is finite and its cardinality, $|\mathcal{Y}|$, is denoted by β. Here, the acceptance set depends on $\boldsymbol{y}$ and denoted accordingly by $\mathcal{A}_n(\boldsymbol{y})$. The encrypter is therefore a mapping $W = T(\boldsymbol{x}, \boldsymbol{y}, K)$, which is invertible given $(\boldsymbol{y}, K)$, allowing the decrypter to reconstruct $\boldsymbol{x} = T^{-1}(W, \boldsymbol{y}, K)$.

Consider first a natural extension of Ziv's model for a finite-state discriminator, which for $i = 0, 1, \ldots, n-1$, implements the recursion,

$$u_i = f(z_i, \hat{x}_i, y_i) \tag{52}$$

$$z_{i+1} = g(z_i, \hat{x}_i, y_i). \tag{53}$$

Similarly as before, the acceptance set, $\mathcal{A}_n(\boldsymbol{y})$ is the set of all $\{\boldsymbol{x}\}$, which in the presence of the given $\boldsymbol{y}$, yield the all-zero output sequence, $\boldsymbol{u} = (u_0, u_1, \ldots, u_{n-1}) = (0, 0, \ldots, 0)$.

Let $c(\boldsymbol{x}, \boldsymbol{y})$ denote the number of phrases of joint parsing of

$$(\boldsymbol{x}, \boldsymbol{y}) = ((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n))$$

into distinct phrases, and let $c_l(\boldsymbol{x}|\boldsymbol{y})$ denote the number of occurrences of $\boldsymbol{y}(l)$ - the l-th distinct phrase of $\boldsymbol{y}$, which is also the number of different phrases of $\boldsymbol{x}$ that appear jointly with $\boldsymbol{y}(l)$. Let $c(\boldsymbol{y})$ denote the number of different $\{\boldsymbol{y}(l)\}$, that is, $\sum_{l=1}^{c(\boldsymbol{y})} c_l(\boldsymbol{x}|\boldsymbol{y}) = c(\boldsymbol{x}, \boldsymbol{y})$. Finally, let $c_{lzz'}(\boldsymbol{x}|\boldsymbol{y})$ denote the number of $\boldsymbol{x}$-phrases that are aligned with $\boldsymbol{y}(l)$, beginning at state z and ending at state z'. Then, similarly as in the derivation in [32] (see also Appendix),

$$|\mathcal{A}_n(\boldsymbol{y})| \geq \prod_{l=1}^{c(\boldsymbol{y})} \prod_{z,z'} c_{lzz'}(\boldsymbol{x}|\boldsymbol{y})^{c_{lzz'}(\boldsymbol{x}|\boldsymbol{y})}, \tag{54}$$

and so, defining the auxiliary RV's (Z_l, Z_l'), $l = 1, \ldots, c(\boldsymbol{y})$, as being jointly distributed according to

$$Q_l(z, z') = \frac{c_{lzz'}(\boldsymbol{x}|\boldsymbol{y})}{c_l(\boldsymbol{x}|\boldsymbol{y})}, \quad z, z \in \mathcal{S}, \tag{55}$$

we have

$$\begin{aligned} \log|\mathcal{A}_n(\boldsymbol{y})| &\geq \sum_{l=1}^{c(\boldsymbol{y})} \sum_{z,z'} c_{lzz'}(\boldsymbol{x}|\boldsymbol{y}) \log c_{lzz'}(\boldsymbol{x}|\boldsymbol{y}) \\ &= \sum_{l=1}^{c(\boldsymbol{y})} c_l(\boldsymbol{x}|\boldsymbol{y}) \sum_{z,z'} \frac{c_{lzz'}(\boldsymbol{x}|\boldsymbol{y})}{c_l(\boldsymbol{x}|\boldsymbol{y})} \left[\log \frac{c_{lzz'}(\boldsymbol{x}|\boldsymbol{y})}{c_l(\boldsymbol{x}|\boldsymbol{y})} + \log c_l(\boldsymbol{x}|\boldsymbol{y})\right] \end{aligned}$$

$$
\begin{aligned}
&= \sum_{l=1}^{c(\boldsymbol{y})} c_l(\boldsymbol{x}|\boldsymbol{y}) \log c_l(\boldsymbol{x}|\boldsymbol{y}) - \sum_{l=1}^{c(y)} c_l(\boldsymbol{x},\boldsymbol{y}) H(Z_l, Z_l') \\
&\geq \sum_{l=1}^{c(\boldsymbol{y})} c_l(\boldsymbol{x}|\boldsymbol{y}) \log c_l(\boldsymbol{x}|\boldsymbol{y}) - 2 \cdot \sum_{l=1}^{c(y)} c_l(\boldsymbol{x},\boldsymbol{y}) \log s \\
&= \sum_{l=1}^{c(\boldsymbol{y})} c_l(\boldsymbol{x}|\boldsymbol{y}) \log c_l(\boldsymbol{x}|\boldsymbol{y}) - 2c(\boldsymbol{x},\boldsymbol{y}) \log s \\
&\geq \sum_{l=1}^{c(\boldsymbol{y})} c_l(\boldsymbol{x}|\boldsymbol{y}) \log c_l(\boldsymbol{x}|\boldsymbol{y}) - \frac{2n \log s}{(1-\epsilon_n) \log n}, \qquad (56)
\end{aligned}
$$

where the last inequality follows from [10] (see also [35, Lemma 13.5.3]) with $\epsilon_n = O\left(\frac{\log(\log n)}{\log n}\right)$. This lower bound can be asymptotically attained by the conditional version of the LZ algorithm (see [40] and [41]), followed by one-time pad encryption. If $\boldsymbol{y}$ is unavailable at the encrypter, $\boldsymbol{x}$ can still be compressed into about

$$
u(\boldsymbol{x}|\boldsymbol{y}) = \sum_{l=1}^{c(\boldsymbol{y})} c_l(\boldsymbol{x}|\boldsymbol{y}) \log c_l(\boldsymbol{x}|\boldsymbol{y}) \qquad (57)
$$

bits before the one-time pad encryption, using Slepian-Wolf encoding [35, Section 15.4] and reconstructing (with high probability) after decryption using a universal decoder that uses $u(\boldsymbol{x}|\boldsymbol{y})$ as a decoding metric, see [42].

Unfortunately, and somewhat surprisingly, a direct extension of the results of Subsections 3.1 and 3.2 to the case with SI turns out to be rather elusive. The reason is the lack of a single-letter formula for the exponential growth rate of the cardinality of a finite-state conditional type class of $\boldsymbol{x}$-vectors that is defined by joint counts of the form $n(x,y,z) = \sum_{i=1}^{n} \mathcal{I}\{x_i = x, y_i = y, z_i = z\}$, even in the simplest special case where $z_i = x_{i-1}$. In a nutshell, this quantity depends on $\boldsymbol{y}$ in a complicated manner, which cannot be represented in terms of an empirical distribution of fixed dimension that does not grow with n. However, we can obtain at least a lower bound if we treat these cases as special cases of the periodically time-varying FSMs with counters in view of Remark 5.

Finally, consider the class of discriminators implementable by periodically time-varying FSMs with SI, defined as follows:

$$
z_{i+1} = g(z_i, \hat{x}_i, y_i, i \bmod l), \qquad (58)
$$

where $i = 0, 1, 2, \ldots$ and l is as in Subsection 3.3. The extension of the derivation in Subsection 3.3 is quite straightforward - one has to count permutations of ℓ-vectors of $\boldsymbol{x}$ that, not only share the same initial and final states, but are also aligned to the same ℓ-blocks of $\boldsymbol{y}$. The resulting bound is given by:

$$R \geq \frac{\log |\mathcal{A}_n(\boldsymbol{y})|}{n} \geq \frac{\hat{H}(X^\ell|Y^\ell)}{\ell} - \frac{2\log s}{\ell} - \frac{\ell s^2 \alpha^\ell \beta^\ell}{n} \log\left(\frac{n}{\ell} + 1\right), \tag{59}$$

which, for $\ell \gg 2\log s$, can be essentially matched by universal compression for block-memoryless sources in the presence of SI and one-time pad encryption. In the absence of SI at the encrypter, one may use universal Slepian-Wolf coding for block-memoryless sources, which is a direct extension of universal Slepian-Wolf coding for memoryless sources, see, e.g., [43].

Appendix - Proof of the Last Step of Eq. (13)

Since reference [32] is unpublished, then for the sake of completeness, we provide here the essential steps of Ziv's proof of the inequality,

$$\log |\mathcal{A}_n| \geq LZ(\boldsymbol{x}) - n\epsilon_n, \tag{A.1}$$

which must hold for every finite-state discriminator with no more than s states, where $\epsilon_n = O\left(\frac{\log(\log n)}{\log n}\right)$ for fixed s. The final steps are different from those of [32] by adopting a somewhat simpler approach, that is presented in Section 13.5 of [35].

Let $\boldsymbol{x}$ be parsed into c distinct phrases,

$$(x_0, x_1, \ldots, x_{n_1-1}), (x_{n_1}, x_{n_1+1}, \ldots, x_{n_2-1}), \ldots, (x_{n_c}, x_{n_c+1}, \ldots, x_{n-1})$$

with the possible exception of the last phrase which might be incomplete, and let $c_{lzz'}$ denote the number of phrases of length l wherein the initial state of the FSM g is z and the final state is z' (of course, $\sum_{l,z,z'} c_{lzz'} = c$). If the discriminator accepts $\boldsymbol{x}$, namely, if $\boldsymbol{u} = (0, 0, \ldots, 0)$, then for every (l, z, z'), each of the $c_{lzz'}$ phrases can be replaced by any of the other such phrases to obtain new sequences of length n for which the output is also $\boldsymbol{u} = (0, 0, \ldots, 0)$, and hence must be accepted too. Now, let (L, Z, Z') denote a triple of auxiliary random variables jointly distributed according to the probability distribution

$$Q(l, z, z') = \frac{c_{lzz'}}{c}, \qquad z, z' \in \mathcal{S},\ l = 1, 2, \ldots \tag{A.2}$$

and let $H(L, Z, Z')$ denote the joint entropy of (L, Z, Z'). Then,

$$|\mathcal{A}_n| \geq \prod_{l,z,z'} (c_{lzz'})^{c_{lzz'}}, \tag{A.3}$$

where $(c_{lzz'})^{c_{lzz'}}$ on the right-hand side is the number of ways each one of the $c_{lzz'}$ phrases of length l, starting at state z and ending at state z' can be replaced by any other phrase with the same qualifiers. Consequently,

$$\begin{aligned}
\log|\mathcal{A}_n| &\geq \sum_{l,z,z'} c_{lzz'} \log c_{lzz'} && \text{(A.4)}\\
&= c \cdot \sum_{l,z,z'} \frac{c_{lzz'}}{c} \left[\log \frac{c_{lzz'}}{c} + \log c\right] && \text{(A.5)}\\
&= c \log c - c \cdot H(L, Z, Z') && \text{(A.6)}\\
&\geq c \log c - c[H(L) + H(Z) + H(Z')] && \text{(A.7)}\\
&\geq c \log c - c \cdot H(L) - 2c \log s. && \text{(A.8)}
\end{aligned}$$

Now, the entropy of L, given that $\boldsymbol{E}\{L\} = n/c$, cannot be larger than $(n/c+1)\log(n/c+1) - (n/c)\log(n/c) \leq 1 + \log(n/c+1)$ (see, Lemma 13.5.4 and eqs. (13.120)-(13.122) in [35]). Thus,

$$\log|\mathcal{A}_n| \geq c \log c - 2c \log s - c - c \log\left(\frac{n}{c} + 1\right), \tag{A.9}$$

and then

$$\begin{aligned}
R &\geq \frac{\log|\mathcal{A}_n|}{n}\\
&\geq \frac{c \log c}{n} - \frac{2c}{n} \log s - \frac{c}{n} \log\left(\frac{n}{c} + 1\right)\\
&= \frac{c \log c}{n} - \frac{2c}{n} \log s - O\left(\frac{\log\log n}{\log n}\right), && \text{(A.10)}
\end{aligned}$$

where the last line is obtained similarly as in eqs. (13.123)-(13.124) in [35]. The final step of eq. (13) is obtained from the fact that $LZ(\boldsymbol{x})$ is upper bounded by $c \log c$ plus terms that, after normalizing by n, are are negligible compared to $\frac{\log(\log n)}{\log n}$ – see Theorem 2 in [10].

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] 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.

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

[4] 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.

[5] 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.

[6] 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.

[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] 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.

[14] 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.

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

[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* **2005**, pp. 1516–1520, Adelaide, Australia.

[26] 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).

[27] Hellman, M. E. "An extension of the Shannon theory approach to cryptography," *IEEE Trans. Inform. Theory* **1977**, vol. IT–23, no. 3, pp. 289–294.

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

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

[30] Massey, J. L. "An introduction to contemporary cryptology," *Proc. IEEE* **1988**, vol. 76, vol. 5, pp. 533–549.

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

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

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

[34] Marcus, B.; Roth, R.; Siegel, P. H. *Constrained Systems and Coding for Recording Channels*, Technion-I.I.T., Department of Computer Science, 1998.

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

[36] 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.

[37] Davisson, L. D.; Longo, G.; Sgarro, A. "The error exponent for noiseless encoding of finite ergodic Markov sources," *IEEE Trans. Inform. Theory* **1981**, vol. IT–27, no. 4, pp. 431–438.

[38] Natarajan, S. "Large deviations, hypotheses testing, and source coding for finite Markov chains," *IEEE Trans. Inform. Theory* **1985**, vol. IT–31, no. 3, pp. 360–365.

[39] I. Csiszár, "The method of types," *IEEE Trans. Inform. Theory* **1998**, vol. 44, no. 6, pp. 2505–2523.

[40] Ziv, J. "Universal decoding for finite-state channels," *IEEE Trans. Inform. Theory* **1985**, vol. IT–31, no. 4, pp. 453–460.

[41] Uyematsu, T.; Kuzuoka, S. "Conditional Lempel-Ziv complexity and its application to source coding theorem with side information," *IEICE Trans. Fundamentals* **2003**, vol. E86-A, no. 10, pp. 2615–2617.

[42] Merhav, N. "Universal Slepian-Wolf coding for individual sequences," submitted to *IEEE Trans. Inform. Theory*, **2024**. Available on-line at: `https://arxiv.org/pdf/2403.07409`

[43] Draper, S. C. "Universal incremental Slepian–Wolf coding," *Proc. Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, October 2004.