

DRDO–IISc Programme on
Advanced Research in Mathematical
Engineering

Guessing Based On Length Functions

(TR-PME-2007-02)

by

Rajesh Sundaresan

Department of Electrical Communication Engineering, Indian Institute of Science,
Bangalore

26 February 2007



Indian Institute of Science
Bangalore 560 012

Guessing Based On Length Functions

Rajesh Sundaresan

26 February 2007

ABSTRACT

A guessing wiretapper's performance on a Shannon cipher system is analyzed for a source with memory. Close relationships between guessing functions and length functions are first established. Subsequently, asymptotically optimal encryption and attack strategies are identified and their performances analyzed for sources with memory. The performance metrics are exponents of guessing moments and probability of large deviations. The metrics are then characterized for unifilar sources. Universal asymptotically optimal encryption and attack strategies are also identified for unifilar sources. Guessing in the increasing order of Lempel-Ziv coding lengths is proposed for finite-state sources, and shown to be asymptotically optimal. Finally, competitive optimality properties of guessing in the increasing order of description lengths and Lempel-Ziv coding lengths are demonstrated.

Keywords: cipher systems, compression, cryptography, guessing, Lempel-Ziv code, length function, minimum description length, sources with memory, source coding, unifilar, universal source coding

1 INTRODUCTION

We consider the classical Shannon cipher system [1]. Let $X^n = (X_1, \dots, X_n)$ be a message where each letter takes values on a finite set \mathbb{X} . This message should be communicated securely from a transmitter to a receiver, both of which have access to a common secure key U^k of k purely random bits independent of X^n . The transmitter computes the cryptogram $Y = f_n(X^n, U^k)$ and sends it to the receiver over a public channel. The cryptogram may be of variable length. The function f_n is invertible given U^k . The receiver, knowing Y and U^k , computes $X^n = f_n^{-1}(Y, U^k)$. The functions f_n and f_n^{-1} are published. An attacker (wiretapper) has access to the cryptogram Y , knows f_n and f_n^{-1} , and attempts to identify X^n without knowledge of U^k . The attacker can use knowledge of the statistics of X^n . We assume that the attacker has a test mechanism that tells him whether a guess \hat{X}^n is correct

or not. For example, the attacker may wish to attack an encrypted password or personal information to gain access to, say, a computer account, or a bank account via internet, or a classified database [2]. In these situations, successful entry into the system or a failure provides the natural test mechanism. We assume that the attacker is allowed an unlimited number of guesses. Given the probability mass function (PMF) of X^n , the function f_n , and the cryptogram Y , the attacker can determine the posterior probabilities of the message $P_{X^n|Y}(\cdot | y)$. His best guessing strategy having observed $Y = y$ is then to guess in the decreasing order of these posterior probabilities $P_{X^n|Y}(\cdot | y)$. The key rate for the system is $k/n = R$ which represents the number of bits of key used to communicate one message letter.

Merhav and Arikan [2] study discrete memoryless sources (DMS) in the above setting and characterize the best attainable moments of the number of guesses that the attacker has to submit before success. In particular, they show that for a DMS with the governing single letter PMF P on \mathbb{X} , the value of the optimal guessing exponent is given by

$$E(R, \rho) = \max_Q [\rho \min\{H(Q), R\} - D(Q \| P)],$$

where the maximization is over all PMFs Q on \mathbb{X} , $H(Q)$ is the Shannon entropy of the PMF Q , and $D(Q \| P)$ is the Kullback-Leibler divergence between Q and P . They also show that $E(R, \rho)$ equals ρR for $R < H(P)$, and equals the constant $\rho H_{1/(1+\rho)}(P)$ for $R > H(P)$. When $R < H(P)$, the key rate is not sufficiently large, and an exhaustive key-search attack is asymptotically optimal. When $R > H(P)$, the randomness introduced by the key is near perfect, and the cryptogram is useless to the attacker. The attacker submits guesses based directly on the message statistics, and $\rho H_{1/(1+\rho)}(P)$ is known to be the optimal guessing exponent in this scenario [3], where $H_{1/(1+\rho)}(P)$ is the Rényi entropy of the DMS P . For $H(P) < R < H(P_\rho)$, the optimal strategy makes use of both the key and the message statistics. P_ρ is the PMF of an auxiliary DMS given by (47). Merhav and Arikan [2] also determine the best achievable performance based on the large deviations of the number of guesses for success, and show that it equals the Fenchel-Legendre transform of $E(R, \rho)$ as a function of ρ .

Secret messages typically come from the natural languages which can be well-modeled as sources with memory, for e.g., a Markov source of an appropriate order. In this report, we extend the results of Merhav and Arikan [2] to sources with memory. As a first step towards this, we first consider the perfect secrecy scenario (for e.g., those analogous to $R \geq H(P_\rho)$ in the DMS case), and identify a tight relationship between the number of guesses for success and a lossless source coding length function. Specifically, we sandwich the number of guesses on either side by a suitable length function. Arikan's result [3] that the best value of the guessing exponent for memoryless sources is the Rényi entropy of an appropriate order immediately

follows by recognizing that it is the least value of an average exponential coding length problem proposed and solved by Campbell [4]. Our approach based on length functions has the benefit of showing that guessing in the increasing order of lengths of compressed strings can yield a good attack strategy for sources with memory. In particular, guessing in the increasing order of Lempel-Ziv code lengths [5] for finite-state sources and increasing description lengths for unifilar sources [6] are asymptotically optimal in a sense made precise in the sequel. In the perfect secrecy context, our approach is closely related to that of Weinberger et al [7] who use Lempel-Ziv code lengths for universal ordering of sequences. However, the quantities studied here and in [7] differ. We study the asymptotics of the ratio of the expected number of guesses (resp. moments) under a universal guessing scheme over the optimal expected number of guesses (resp. moments). Weinberger et al [7] consider the asymptotics of the expected value of the point-wise ratio of the same quantities. Consequently, our asymptotics involves Rényi entropies while theirs involves Shannon entropy.

Next, we establish similar connections between guessing and source compression for the key-constrained scenarios (i.e., those analogous to $R < H(P_\rho)$ in the memoryless case). We then study guessing exponents for the cipher system on sources with memory, and then specialize our results to show that all conclusions of Merhav and Arikan in [2] for memoryless sources extend to unifilar sources. We also consider the large deviations performance of the number of guesses and show that attacks based on the Lempel-Ziv coding lengths and minimum description lengths are asymptotically optimal for finite-state and unifilar sources, respectively. We then establish competitive optimality results for guessing based on these two length functions.

The report is organized as follows. In Section 2 we study guessing under perfect secrecy and establish the relationship between guessing and source compression. In Section 3, we study the key-rate constrained system, establish optimal strategies for both parties for sources with memory, and study the relationship between guessing and a new source coding problem. In Section 4, we characterize the performance for unifilar sources. In Section 5, we study the large deviations performance and establish the optimality properties of guessing based on Lempel-Ziv and minimum description lengths. Section 6 summarizes the report and presents some open problems.

2 Guessing under perfect secrecy and source compression

Let us first consider the following ideal setting where $k = nR \geq n \log |\mathbb{X}|$. Enumerate all the sequences in \mathbb{X}^n from 0 to $|\mathbb{X}|^n - 1$ and let the function f_n be the bit-wise XOR of the key bits and the bits representing the index of the message. The cryptogram is the message whose index is the output of f_n .

The decryption function is also clear - simply XOR the bits representing the cryptogram with the key bits. Such an encryption renders the cryptogram completely useless to an attacker who does not have knowledge of the key. The attacker's optimal strategy is to guess the message in the decreasing order of message probabilities. In case the attacker does not have access to the message probabilities, a robust strategy is needed. We first relate the problem of guessing to one of source compression. As we will see soon, robust source compression strategies lead to robust guessing strategies.

For ease of exposition, and because we have perfect encryption, let us assume that the message space is simply \mathbb{X} . The extension to strings of length n is straightforward.

A guessing function

$$G : \mathbb{X} \rightarrow \{1, 2, \dots, |\mathbb{X}|\}$$

is a bijection that denotes the order in which the elements of \mathbb{X} are guessed. If $G(x) = i$, then the i th guess is x . A length function

$$L : \mathbb{X} \rightarrow \mathbb{N}$$

is one that satisfies Kraft's inequality

$$\sum_{x \in \mathbb{X}} 2^{-L(x)} \leq 1. \quad (1)$$

To each guessing function G , we associate a PMF Q_G on \mathbb{X} and a length function L_G as follows.

Definition 1 *Given a guessing function G , we say Q_G defined by*

$$Q_G(x) = c^{-1} \cdot G(x)^{-1}, \quad \forall x \in \mathbb{X}, \quad (2)$$

is the PMF on \mathbb{X} associated with G . The quantity c in (2) is the normalization constant. We say L_G defined by

$$L_G(x) = \lceil -\log Q_G(x) \rceil, \quad \forall x \in \mathbb{X}, \quad (3)$$

is the length function associated with G . □

Observe that

$$c = \sum_{a \in \mathbb{X}} G(a)^{-1} = \sum_{i=1}^{|\mathbb{X}|} \frac{1}{i} \leq 1 + \ln |\mathbb{X}|, \quad (4)$$

and therefore the PMF in (2) is well-defined. We record the intimate relationship between these associated quantities in the following result. (This is also available in the proof of [7, Th. 1, p.382]).

Proposition 2 *Given a guessing function G , the associated quantities satisfy*

$$c^{-1} \cdot Q_G(x)^{-1} = G(x) \leq Q_G(x)^{-1}, \quad (5)$$

$$L_G(x) - 1 - \log c \leq \log G(x) \leq L_G(x). \quad (6)$$

□

Proof: The first equality in (5) follows from the definition in (2), and the second inequality from the fact that $c \geq 1$.

The upper bound in (6) follows from the upper bound in (5) and from (3). The lower bound in (6) follows from

$$\begin{aligned} \log G(x) &= \log (c^{-1} \cdot Q_G(x)^{-1}) \\ &= -\log Q_G(x) - \log c \\ &\geq (\lceil -\log Q_G(x) \rceil - 1) - \log c \\ &= L_G(x) - 1 - \log c. \end{aligned}$$

■

We now associate a guessing function G_L to each length function L .

Definition 3 *Given a length function L , we define the associated guessing function G_L to be the one that guesses in the increasing order of L -lengths. Messages with the same L -length are ordered using an arbitrary fixed rule, say the lexicographic order on \mathbb{X} . We also define the associated PMF Q_L on \mathbb{X} to be*

$$Q_L(x) = \frac{2^{-L(x)}}{\sum_{a \in \mathbb{X}} 2^{-L(a)}}. \quad (7)$$

□

Proposition 4 *For a length function L , the associated PMF and the guessing function satisfy the following:*

1. G_L guesses messages in the decreasing order of Q_L -probabilities;
- 2.

$$\log G_L(x) \leq \log Q_L(x)^{-1} \leq L(x). \quad (8)$$

□

Proof: The first statement is clear from the definition of G_L and from (7).

Letting $1\{E\}$ denote the indicator function of an event E , we have as a consequence of statement 1) that

$$\begin{aligned} G_L(x) &\leq \sum_{a \in \mathbb{X}} 1\{Q_L(a) \geq Q_L(x)\} \\ &\leq \sum_{a \in \mathbb{X}} \frac{Q_L(a)}{Q_L(x)} \\ &= Q_L(x)^{-1}, \end{aligned} \tag{9}$$

which proves the left inequality in (8). This inequality was known to Wyner [8].

The last inequality in (8) follows from (7) and Kraft's inequality (1) as follows:

$$Q_L(x)^{-1} = 2^{L(x)} \cdot \sum_{a \in \mathbb{X}} 2^{-L(a)} \leq 2^{L(x)}.$$

■

Let $\{L(x) \geq B\}$ denote the set $\{x \in \mathbb{X} \mid L(x) \geq B\}$. We then have the following easy to verify corollary to Propositions 2 and 4.

Corollary 5 *For a given G , its associated length function L_G , and any $B \geq 1$, we have*

$$\begin{aligned} \{L_G(x) \geq B + 1 + \log c\} \\ \subseteq \{G(x) \geq 2^B\} \\ \subseteq \{L_G(x) \geq B\}. \end{aligned} \tag{10}$$

Analogously, for a given L , its associated guessing function G_L , and any $B \geq 1$, we have

$$\{G_L(x) \geq 2^B\} \subseteq \{L(x) \geq B\}. \tag{11}$$

□

The inequalities between the associates in (6) and (8) indicate the direct relationship between guessing moments and Campbell's coding problem [4], and that the Rényi entropies are the optimal growth exponents for guessing moments. See (14) below. They also establish a simple and new result: the minimum expected value of the logarithm of the number of guesses is close to the Shannon entropy.

We now demonstrate other relationships between guessing moments and average exponential coding lengths which will be useful in establishing universality properties.

Proposition 6 *Let L be any length function on \mathbb{X} , G_L the guessing function associated with L , P a PMF on \mathbb{X} , $\rho \in (0, \infty)$, L^* the length function that*

minimizes $\mathbb{E}[2^{\rho L^*(X)}]$, where the expectation is with respect to P , G^* the guessing function that proceeds in the decreasing order of P -probabilities and therefore the one that minimizes $\mathbb{E}[G^*(X)^\rho]$, and c as in (4). Then

$$\frac{\mathbb{E}[G_L(X)^\rho]}{\mathbb{E}[G^*(X)^\rho]} \leq \frac{\mathbb{E}[2^{\rho L(X)}]}{\mathbb{E}[2^{\rho L^*(X)}]} \cdot 2^{\rho(1+\log c)}. \quad (12)$$

Analogously, let G be any guessing function, and L_G its associated length function. Then

$$\frac{\mathbb{E}[G(X)^\rho]}{\mathbb{E}[G^*(X)^\rho]} \geq \frac{\mathbb{E}[2^{\rho L_G(X)}]}{\mathbb{E}[2^{\rho L^*(X)}]} \cdot 2^{-\rho(1+\log c)}. \quad (13)$$

Also,

$$\left| \frac{1}{\rho} \log \mathbb{E}[G^*(X)^\rho] - \frac{1}{\rho} \log \mathbb{E}[2^{\rho L^*(X)}] \right| \leq 1 + \log c. \quad (14)$$

□

Proof: Observe that

$$\begin{aligned} \mathbb{E}[2^{\rho L(X)}] &\geq \mathbb{E}[G_L(X)^\rho] \\ &\geq \mathbb{E}[G^*(X)^\rho] \end{aligned} \quad (15)$$

$$\geq \mathbb{E}[2^{\rho L_{G^*}(X)}] 2^{-\rho(1+\log c)} \quad (16)$$

$$\geq \mathbb{E}[2^{\rho L^*(X)}] 2^{-\rho(1+\log c)}, \quad (17)$$

where (15) follows from (8), and (16) from the left inequality in (6). The result in (12) immediately follows. A similar argument shows (13). Finally, (14) follows from the inequalities leading to (17) by setting $L = L^*$. ■

Thus if we have a length function whose performance is close to optimal, then its associated guessing function is close to guessing optimal. The converse is true as well. Moreover, the optimal guessing exponent is within $1 + \log c$ of the optimal coding exponent for the length function.

Let us now consider strings of length n . Let \mathbb{X}^n denote the set of messages and consider $n \rightarrow \infty$. Let \mathbb{T} denote a class of sources. It is now easy to see that universality (within the class) in the average exponential coding rate sense implies existence of a universal guessing strategy that achieves the optimal exponent for guessing. For each source in the class, let P_n be its restriction to strings of length n and let L_n^* denote an optimal length function that attains the minimum value $\mathbb{E}[2^{\rho L_n^*(X^n)}]$ among all length functions, the expectation being with respect to P_n . On the other hand, let L_n be a sequence of length functions for the class of sources that does not depend on

the actual source within the class. Suppose further that the length sequence L_n is asymptotically optimal, i.e.,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n\rho} \log \mathbb{E} \left[2^{\rho L_n(X^n)} \right] \\ = \lim_{n \rightarrow \infty} \frac{1}{n\rho} \log \mathbb{E} \left[2^{\rho L_n^*(X^n)} \right], \end{aligned}$$

for every source belonging to the class. L_n is thus “universal” for (i.e., asymptotically optimal for all sources in) the class. An application of (12) by denoting c in (12) as c_n followed by the observation $(1 + \log c_n)/n \rightarrow 0$ shows that the sequence of guessing strategies G_{L_n} is asymptotically optimal for the class, i.e.,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n\rho} \log \mathbb{E} [G_{L_n}(X^n)^\rho] \\ = \lim_{n \rightarrow \infty} \frac{1}{n\rho} \log \mathbb{E} [G^*(X^n)^\rho]. \end{aligned}$$

Arikan and Merhav [9] provide a universal guessing strategy for the class of discrete memoryless sources (DMS). For the class of unifilar sources with a known number of states, the minimum description length encoding is asymptotically optimal for Campbell’s coding length problem (see Merhav [6]). It follows as a consequence of the above argument that guessing in the increasing order of description lengths is asymptotically optimal. (See also the development in Section 4). The left side of (12) is the extra factor in the expected number of guesses (relative to the optimal value) due to lack of knowledge of the specific source in class. Our prior work [10] characterizes this loss as a function of the uncertainty class.

3 Guessing with key-rate constraints and source compression

We continue to consider strings of length n . Let X^n be a message and U^k the secure key of purely random bits independent of X^n . Recall that the transmitter computes the cryptogram $Y = f_n(X^n, U^k)$ and sends it to the receiver over a public channel. Given a PMF of X^n , the function f_n , and the cryptogram Y , the attacker’s optimal strategy is to guess in the decreasing order of posterior probabilities $P_{X^n|Y}(\cdot | y)$. Let us denote this optimal attack strategy as G_{f_n} . The key rate for the system is $k/n = R < \log |\mathbb{X}|$. If the attacker does not know the source statistics, a robust guessing strategy is needed. The following is a first step in this direction.

Proposition 7 *Let L_n be an arbitrary length function on \mathbb{X}^n . There is a guessing list G such that for any encryption function f_n , we have*

$$G(x^n | y) \leq 2 \min \left\{ 2^{nR}, 2^{L_n(x^n)} \right\}.$$

□

Proof: We use a technique of Merhav and Arikan [2]. Let G_{L_n} denote the associated guessing function that proceeds in the increasing order of the lengths and completely ignores the cryptogram. Let G_{L_n} proceed in the order x_1^n, x_2^n, \dots . By Proposition 4, we need at most $2^{L_n(x^n)}$ guesses to identify x^n .

Consider the alternative exhaustive key-search attack defined by the following guessing list:

$$f_n^{-1}(y, u_1^k), f_n^{-1}(y, u_2^k), \dots,$$

where u_1^k, u_2^k, \dots is an arbitrary ordering of the keys. This strategy identifies x^n in at most 2^{nR} guesses.

Finally, let $G(\cdot | y)$ be the list that alternates between the two lists, skipping those already guessed, i.e., the one that proceeds in the order

$$\left\{ x_1^n, f_n^{-1}(y, u_1^k), x_2^n, f_n^{-1}(y, u_2^k), \dots \right\}. \quad (18)$$

Clearly, for every x^n , we need at most twice the minimum of the two original lists. ■

We now look at a weak converse to the above in the expected sense. Our proof also suggests an asymptotically optimal encryption strategy for sources with memory.

Proposition 8 *Fix $n \in \mathbb{N}$, $\rho > 0$, and let c_n denote the constant in (4) as a function of n with \mathbb{X}^n replacing \mathbb{X} . There is an encryption function f_n and a length function L_n such that every guessing strategy $G(\cdot | y)$ (and in particular G_{f_n}) satisfies*

$$\begin{aligned} & \mathbb{E}[G(X^n | Y)^\rho] \\ & \geq \frac{1}{(2c_n)^\rho(2+\rho)} \mathbb{E} \left[\left(\min \left\{ 2^{L_n(X^n)}, 2^{nR} \right\} \right)^\rho \right]. \end{aligned}$$

□

Proof: The proof is an extension of Merhav and Arikan's proof of [2, Th. 1] to sources with memory. The idea is to identify an encryption mechanism that maps messages of roughly equal probability to each other.

Let P_n be any PMF on \mathbb{X}^n . Enumerate the elements of \mathbb{X}^n in the decreasing order of their probabilities. For convenience, let $M = 2^{nR}$. If M

does not divide $|\mathbb{X}|^n$, append a few dummy messages of zero probability to make the number of messages N a multiple of M . Index the messages from 0 to $N - 1$. Henceforth, we identify a message by its index.

Divide the messages into groups of M so that message m belongs to group T_j , where $j = \lfloor m/M \rfloor$, and $\lfloor \cdot \rfloor$ is the floor function. Enumerate the key streams from 0 to $M - 1$, so that $0 \leq u \leq M - 1$. The function f_n is now defined as follows. For $m = jM + i$ set

$$f_n(jM + i, u) \triangleq jM + (i \oplus u),$$

where $i \oplus u$ is the bit-wise XOR operation. Thus messages in group T_j are encrypted to messages in the same group. The index i identifying the specific message in group T_j , i.e., the last nR bits of m , are encrypted via bit-wise XOR with the key stream. Given u and the cryptogram, decryption is clear – perform bit-wise XOR with u on the last nR bits of y .

Given a cryptogram y , the only information that the attacker gleans is that the message belongs to the group determined by y . Indeed, if $y \in T_j$

$$P_n \{Y = y\} = \frac{1}{M} P_n \{X^n \in T_j\}$$

and therefore

$$P_n \{X^n = m \mid Y = y\} = \begin{cases} \frac{P_n \{X^n = m\}}{P_n \{X^n \in T_j\}}, & \lfloor m/M \rfloor = j, \\ 0, & \text{otherwise,} \end{cases}$$

decreases with m for $m \in T_j$, and is 0 for $m \notin T_j$. The attacker's best strategy $G_{f_n}(\cdot \mid y)$ is therefore to restrict his guesses to T_j and guess in the order $jM, jM + 1, \dots, jM + M - 1$. Thus, when $x^n = jM + i$, the optimal attack strategy requires $i + 1$ guesses.

We now analyze the performance of this attack strategy as follows.

$$\begin{aligned}
& \mathbb{E}[G_{f_n}(X^n|Y)^\rho] \\
&= \sum_{j=0}^{N/M-1} \sum_{i=0}^{M-1} P_n\{X^n = jM + i\}(i+1)^\rho \\
&\geq \sum_{j=0}^{N/M-1} \sum_{i=0}^{M-1} P_n\{X^n = (j+1)M - 1\}(i+1)^\rho \tag{19}
\end{aligned}$$

$$\geq \sum_{j=0}^{N/M-1} P_n\{X^n = (j+1)M - 1\} \frac{M^{1+\rho}}{1+\rho} \tag{20}$$

$$\geq \frac{1}{1+\rho} \sum_{j=0}^{N/M-1} \sum_{i=0}^{M-1} P_n\{X^n = (j+1)M + i\} M^\rho \tag{21}$$

$$= \frac{1}{1+\rho} \sum_{m=M}^{N-1} P_n\{X^n = m\} M^\rho \tag{22}$$

where (19) follows because the arrangement in the decreasing order of probabilities implies that

$$P_n\{X^n = jM + i\} \geq P_n\{X^n = (j+1)M - 1\}$$

for $i = 0, \dots, M-1$. Inequality (20) follows because

$$\sum_{i=0}^{M-1} (i+1)^\rho = \sum_{i=1}^M i^\rho \geq \int_0^M z^\rho dz = \frac{M^{1+\rho}}{1+\rho},$$

(21) follows because by the decreasing probability arrangement

$$P_n\{X^n = (j+1)M - 1\} \geq \frac{1}{M} \sum_{i=0}^{M-1} P_n\{X^n = (j+1)M + i\}.$$

Thus (22) implies that

$$\begin{aligned}
& \sum_{m=0}^{N-1} P_n\{X^n = m\} (\min\{m+1, M\})^\rho \\
&= \sum_{m=0}^{M-1} P_n\{X^n = m\} (m+1)^\rho + \sum_{m=M}^{N-1} P_n\{X^n = m\} M^\rho \\
&\leq \mathbb{E}[G_{f_n}(X^n|Y)^\rho] + (1+\rho) \mathbb{E}[G_{f_n}(X^n|Y)^\rho] \\
&= (2+\rho) \mathbb{E}[G_{f_n}(X^n|Y)^\rho], \tag{23}
\end{aligned}$$

Set G_P to be the guessing function that guesses in the decreasing order of P -probabilities without regard to Y , i.e., $G_P(m) = m + 1$. Let L_{G_P} be the associated length function. Now use (23) and (6) to get

$$\begin{aligned}
& \mathbb{E}[G_{f_n}(X^n|Y)^\rho] \\
& \geq \frac{1}{2+\rho} \mathbb{E}[(\min\{G_P(X^n), M\})^\rho] \\
& \geq \frac{1}{2+\rho} \mathbb{E}\left[\left(\min\left\{\frac{2^{L_{G_P}(X^n)}}{2c_n}, M\right\}\right)^\rho\right] \\
& \geq \frac{1}{(2c_n)^\rho(2+\rho)} \mathbb{E}\left[\left(\min\left\{2^{L_{G_P}(X^n)}, M\right\}\right)^\rho\right].
\end{aligned}$$

Since G_{f_n} is the strategy that minimizes $\mathbb{E}[G(X^n | Y)^\rho]$, the proof is complete. \blacksquare

For a given $\rho > 0$, key rate $R > 0$, encryption function f_n , define

$$E_n(R, \rho) \triangleq \sup_{f_n} \frac{1}{n} \log \mathbb{E}[G_{f_n}(X^n | Y)^\rho].$$

Propositions 7 and 8 naturally suggest the following coding problem: identify

$$E_{n,l}(R, \rho) \triangleq \min_{L_n} \frac{1}{n} \log \mathbb{E}\left[\left(\min\left\{2^{L_n(X^n)}, 2^{nR}\right\}\right)^\rho\right]. \quad (24)$$

Analogous to (14), we can relate $E_n(R, \rho)$ and $E_{n,l}(R, \rho)$ for a specified key rate R . The following is a corollary to Propositions 7 and 8.

Corollary 9 *For a given $R, \rho > 0$, we have*

$$|E_{n,l}(R, \rho) - E_n(R, \rho)| \leq \frac{\log(2^{2\rho} c_n^\rho (2+\rho))}{n}.$$

\square

Proof: Let L_n^* be the length function that achieves $E_{n,l}(R, \rho)$. By Proposition 7, and after taking expectations, we have the guessing strategy $G(\cdot | y)$ that satisfies

$$\begin{aligned}
& \mathbb{E}\left[\left(\min\left\{2^{L_n^*(X^n)}, 2^{nR}\right\}\right)^\rho\right] \\
& \geq \sup_{f_n} \frac{1}{2^\rho} \mathbb{E}[G(X^n | Y)^\rho] \\
& \geq \sup_{f_n} \frac{1}{2^\rho} \mathbb{E}[G_{f_n}(X^n | Y)^\rho] \\
& \geq \frac{1}{2^{2\rho} c_n^\rho (2+\rho)} \mathbb{E}\left[\left(\min\left\{2^{L_n(X^n)}, 2^{nR}\right\}\right)^\rho\right]
\end{aligned}$$

for a particular f_n and L_n guaranteed by Proposition 8

$$\geq \frac{1}{2^{2\rho} c_n^\rho (2 + \rho)} \mathbb{E} \left[\left(\min \left\{ 2^{L_n^*(X^n)}, 2^{nR} \right\} \right)^\rho \right].$$

Take logarithms and normalize by n to get the bound. \blacksquare

The magnitude of the difference between $E_n(R, \rho)$ and $E_{n,l}(R, \rho)$ vanishes as $n \rightarrow \infty$. Thus, the problem of finding the optimal guessing exponent is the same as that of finding the optimal exponent for a coding problem. When $R \geq \log |\mathbb{X}|$, the coding problem in (24) reduces to the one considered by Campbell in [4]. Proposition 7 shows that the optimal length function attaining the minimum in (24) yields an asymptotically optimal attack strategy on the cipher system. Moreover, the encryption strategy in Proposition 8 is asymptotically optimal.

The following Proposition upper bounds the guessing effort needed to identify the correct message for sources with memory. A sharper result analogous to the DMS case is shown later for unifilar sources.

Proposition 10 *For a given $R, \rho > 0$, we have*

$$\limsup_{n \rightarrow \infty} E_n(R, \rho) \leq \min \left\{ \rho R, \limsup_{n \rightarrow \infty} E_n(\rho) \right\}, \quad (25)$$

where

$$E_n(\rho) \triangleq \min_{L_n} \frac{1}{n} \log \mathbb{E} \left[2^{\rho L_n(X^n)} \right].$$

\square

Proof: By Corollary 9, it is sufficient to show that the sequence $E_{n,l}(R, \rho)$ is upperbounded by the sequence on the right side of (25). Let L_n^* be the length function that minimizes $\mathbb{E} [2^{\rho L_n(X^n)}]$. Observe that $\min \{ 2^{\rho n R}, x \}$ is a concave function of x for a fixed ρ and R . Jensen's inequality then yields

$$\mathbb{E} \left[\min \left\{ 2^{\rho n R}, 2^{\rho L_n^*(X^n)} \right\} \right] \leq \min \left\{ 2^{\rho n R}, \mathbb{E} \left[2^{\rho L_n^*(X^n)} \right] \right\}.$$

Take logarithms, normalize by n , and use the definition of $E_{n,l}(\rho, R)$ to get

$$\begin{aligned} E_{n,l}(R, \rho) &\leq \frac{1}{n} \log \left(\min \left\{ 2^{\rho n R}, \mathbb{E} \left[2^{\rho L_n^*(X^n)} \right] \right\} \right) \\ &= \min \left\{ \rho R, \frac{1}{n} \log \mathbb{E} \left[2^{\rho L_n^*(X^n)} \right] \right\}. \end{aligned}$$

Now take the limsup as $n \rightarrow \infty$ to complete the proof. \blacksquare

Our results thus far are applicable to a rather general class of sources with memory. In the next section, we specialize our results to the important class of unifilar sources. If the source is a DMS with defining PMF P , then the second term within the min in (25) is known to be $\rho H_{1/(1+\rho)}(P)$, where $H_{1/(1+\rho)}(P)$ is Rényi's entropy of order $1/(1+\rho)$ for the source. For unifilar sources, we soon show that the limsup can be replaced by a limit which equals ρ times a generalization of the Rényi entropy for such a source.

4 Unifilar Sources

In this section, we generalize the DMS results of Merhav and Arikan [2] to unifilar sources. We first make some definitions largely following Merhav's notation in [6].

Let $x^n = (x_1, \dots, x_n)$ be a string taking values in \mathbb{X}^n . The string x^n needs to be guessed. Let $s^n = (s_1, \dots, s_n)$ be another sequence taking values in \mathbb{S}^n where $|\mathbb{S}| < \infty$. Let $s_0 \in \mathbb{S}$ be a fixed initial state. A probabilistic source P_n is *finite-state* with $|\mathbb{S}|$ states [6] if the probability of observing the sequence pair (x^n, s^n) is given by

$$P_n(x^n, s^n) = \prod_{i=1}^n P(x_i, s_i \mid s_{i-1}),$$

where $P(x_i, s_i \mid s_{i-1})$ is the joint probability of letter x_i and state s_i given the previous state s_{i-1} . The dependence of P_n on the initial state s_0 is implicit. Typically, the letter sequence x^n is observable and the state sequence s^n is not. Let H denote the entropy-rate of a finite-state source, i.e.,

$$H \triangleq - \lim_{n \rightarrow \infty} n^{-1} \sum_{x^n \in \mathbb{X}^n} P_n(x^n) \log P_n(x^n).$$

A finite-state source is *unifilar* [11, p.187] if the state s_i is given by a deterministic mapping $\phi : \mathbb{X} \times \mathbb{S} \rightarrow \mathbb{S}$ as

$$s_i = \phi(x_i, s_{i-1}),$$

and the mapping $x \mapsto \phi(x, s)$ is one-to-one¹ for each $s \in \mathbb{S}$. Given s_0 and the sequence x^n , the state sequence is uniquely determined. Moreover, given s_0 and the state sequence s^n , x^n is uniquely determined. An important example of a unifilar source is a k th order Markov source where $s_i = (x_i, x_{i-1}, \dots, x_{i-k+1})$.

Fix $x^n \in \mathbb{X}^n$. For $s \in \mathbb{S}, x \in \mathbb{X}$, let

$$Q_{x^n}(x, s) = \frac{1}{n} \sum_{i=1}^n 1\{x_i = x, s_{i-1} = s\},$$

where $1\{A\}$ is the indicator function of the event A . Q_{x^n} is thus an empirical PMF on $\mathbb{S} \times \mathbb{X}$. Let

$$Q_{x^n}(s) = \sum_{x \in \mathbb{X}} Q_{x^n}(x, s).$$

The use of Q_{x^n} for both the joint and the marginal PMFs is an abuse of notation. The context should make the meaning clear. Let

$$q_{x^n}(x \mid s) = \begin{cases} Q_{x^n}(x, s)/Q_{x^n}(s), & Q_{x^n}(s) > 0, \\ 0, & Q_{x^n}(s) = 0 \end{cases}$$

¹The definition in [6] does not restrict ϕ to be one-to-one.

denote the empirical letter probability given the state. (Given that ϕ is one-to-one, this actually defines a transition probability matrix on the state space). Denote the empirical conditional entropy as

$$H(Q_{x^n}) = - \sum_{s \in \mathbb{S}} \sum_{x \in \mathbb{X}} Q_{x^n}(x, s) \log q_{x^n}(x|s),$$

and the conditional Kullback-Leibler divergence between the empirical conditional PMF and the one-step transition matrix $P(x|s)$ as

$$D(Q_{x^n} \parallel P) = \sum_{s \in \mathbb{S}} \sum_{x \in \mathbb{X}} Q_{x^n}(x, s) \log \frac{q_{x^n}(x|s)}{P(x|s)}.$$

Given that we are dealing with multiple random variables, $H(Q)$ and $D(Q \parallel P)$ usually stand for joint entropy and Kullback-Leibler divergence of a pair of joint distributions. We however alert the reader that they stand for conditional values in our notation.

Let us further define the type T_{x^n} of a sequence x^n as follows:

$$T_{x^n} = \{a^n \in \mathbb{X}^n \mid Q_{a^n} = Q_{x^n}\}.$$

For the unifilar source under consideration, it is easy to see that

$$P_n(x^n) = 2^{-n(H(Q_{x^n}) + D(Q_{x^n} \parallel P))}, \quad (26)$$

i.e., all elements of the same type have the same probability. Moreover, for a fixed type Q_{x^n} , if we set $P(x|s) = q_{x^n}(x|s)$ and observe that for the resulting unifilar source matched to x^n , we have $1 \geq P_n\{T_{x^n}\} = |T_{x^n}|P_n(x^n)$, we easily deduce from (26) that

$$|T_{x^n}| \leq 2^{nH(Q_{x^n})}. \quad (27)$$

Consequently, for any unifilar P_n ,

$$P_n\{T_{x^n}\} \leq 2^{-nD(Q_{x^n} \parallel P)}. \quad (28)$$

Using the fact that the mapping $x \mapsto \phi(x, s)$ is one-to-one for each s , it is possible to get the following useful lower bounds on the size and probability of a type for unifilar sources.

Lemma 11 (Merhav [6, Lemma 1], Gutman [12, Lemma 1]) *For a unifilar source, there exists a sequence $\varepsilon(n) = \Theta(n^{-1} \log n)$ such that*

$$\left| \frac{1}{n} \log P_n\{T_{x^n}\} + D(Q_{x^n} \parallel P) \right| \leq \varepsilon(n) \quad (29)$$

for every $x^n \in \mathbb{X}^n$. □

Consequently, we also have ([6, eqn. (17)]):

$$\left| \frac{1}{n} \log |T_{x^n}| - H(Q_{x^n}) \right| \leq \varepsilon(n). \quad (30)$$

Let us now define in a fashion analogous to the DMS case

$$E(R, \rho) \triangleq \max_Q [\rho h(Q, R) - D(Q \| P)] \quad (31)$$

where $h(Q, R) = \min\{H(Q), R\}$, Q is a joint PMF on $\mathbb{S} \times \mathbb{X}$ with letter probabilities given the state identified by $q(x | s)$, and $H(Q)$ is the conditional entropy

$$H(Q) = - \sum_{s \in \mathbb{S}} \sum_{x \in \mathbb{X}} Q(x, s) \log q(x | s).$$

$P(x|s)$ is the conditional PMF that defines the unifilar source. The string s_0 is irrelevant in the definition of $E(R, \rho)$.

We now state and prove a generalization of the Merhav and Arikan result [2, Th. 1].

Theorem 12 *For any unifilar source, any $\rho > 0$,*

$$\lim_{n \rightarrow \infty} E_n(R, \rho) = E(R, \rho).$$

□

Proof: We show that the limiting value of $E_{n,l}(R, \rho)$ exists for the corresponding coding problem and equals $E(R, \rho)$. Corollary 9 then implies that $E_n(R, \rho)$ for the guessing problem has the same limiting value.

Let L_n be a minimal length function that attains $E_{n,l}(R, \rho)$. Arrange the elements of \mathbb{X}^n in the decreasing order of their probabilities. Furthermore, ensure that all sequences belonging to the same type occur together. Enumerate the sequences from 0 to $|\mathbb{X}|^n - 1$. Henceforth we refer to a message by its index.

We claim that we may assume L_n is a nondecreasing function of the message index. Suppose this is not the case. Let j be the first index where the nondecreasing property is violated, i.e. $L_n(i) \leq L_n(i+1)$ for $i = 1, \dots, j-1$, and $L_n(j) > L_n(j+1)$. Identify the smallest index j^* that satisfies $L_n(j^*) > L_n(j+1)$. Modify the lengths as follows: set $L'_n(j^*) = L_n(j+1)$, then $L'_n(i+1) = L_n(i)$ for $i = j^*, \dots, j$, and leave the rest unchanged. Call the new set of lengths L_n . In effect, we have “bubbled” $L_n(j+1)$ towards the smaller indices to the nearest location that does not violate the nondecreasing condition. The new set of lengths will have the same or lower $\mathbb{E} \left[(\min\{2^{L_n(X^n)}, 2^{nR}\})^\rho \right]$. By the optimality of the original set of lengths, the new lengths are also optimal. Furthermore, as a consequence

of the modification, the location of the first index where $L_n(i) \not\leq L_n(i+1)$ has strictly increased. Continue the process until it terminates; it will after a finite number of steps. The resulting L_n is nondecreasing and optimal.

Next, observe that

$$2^{L_n(i)} \geq i + 1 \quad (32)$$

because the length functions are such that the sequences are uniquely decipherable. Another way to see (32) is to observe that index i is the $i+1$ st guess when guessing in the increasing order of L_n as prescribed by the indices, and therefore (8) implies (32).

We then have the following sequence of inequalities

$$\begin{aligned} & \sum_{a^n \in \mathbb{X}^n} P_n(a^n) \left(\min \{ 2^{L_n(a^n)}, 2^{nR} \} \right)^\rho \\ & \geq P_n(x^n) \sum_{a^n \in T_{x^n}} \left(\min \{ 2^{L_n(a^n)}, 2^{nR} \} \right)^\rho \end{aligned} \quad (33)$$

$$\geq P_n(x^n) \sum_{i=i_0(T_{x^n})}^{i_0(T_{x^n})+|T_{x^n}|-1} \left(\min \{ i+1, 2^{nR} \} \right)^\rho \quad (34)$$

$$\geq P_n(x^n) \sum_{i=1}^{|T_{x^n}|} \left(\min \{ i, 2^{nR} \} \right)^\rho \quad (35)$$

$$\begin{aligned} & \geq P_n(x^n) \int_0^{|T_{x^n}|} \left(\min \{ y, 2^{nR} \} \right)^\rho dy \\ & \geq P_n(x^n) |T_{x^n}| \frac{1}{1+\rho} \left(\min \{ |T_{x^n}|, 2^{nR} \} \right)^\rho \end{aligned} \quad (36)$$

$$\geq P\{T_{x^n}\} \frac{1}{1+\rho} \left(\min \{ 2^{nH(Q_{x^n})-n\varepsilon(n)}, 2^{nR} \} \right)^\rho \quad (37)$$

$$\geq \frac{2^{-2n\varepsilon(n)}}{1+\rho} 2^{n(\rho \min\{H(Q_{x^n}), R\} - D(Q_{x^n} \| P))}, \quad (38)$$

where (33) follows by restricting the sum to sequences in type T_{x^n} , (34) follows because of (32) and by setting $i_0(T_{x^n})$ as the starting index of type T_{x^n} . We can do this because our ordering clustered all sequences of the same type. Inequality (35) holds because every term under the summation is lower bounded by the corresponding term on the right side. Inequality (36) follows because of the following. For simplicity, let $|T_{x^n}| = N$ and $2^{nR} = M$. When $N \leq M$,

$$\frac{1}{N} \int_0^N y^\rho dy = \frac{N^\rho}{1+\rho},$$

and when $N > M$,

$$\begin{aligned}
& \frac{1}{N} \int_0^N (\min\{y, M\})^\rho dy \\
&= \frac{1}{N} \int_0^M y^\rho dy + \frac{1}{N} \int_M^N M^\rho dy \\
&= \frac{M}{N} \frac{M^\rho}{1+\rho} + \left(1 - \frac{M}{N}\right) M^\rho \\
&\geq \frac{M^\rho}{1+\rho}.
\end{aligned}$$

Inequality (37) follows from (30) and (38) follows from (29).

The type T_{x^n} in (38) is arbitrary. Moreover, $D(Q \parallel P)$ and $H(Q)$ are continuous functions of Q , and the set of rational empirical functions $\{Q_{x^n}\}$ become dense in the class of unifilar sources with $|\mathbb{S}|$ states and $|\mathbb{X}|$ alphabets, as $n \rightarrow \infty$. From (38) and the above facts, we get $\liminf_{n \rightarrow \infty} E_{n,l}(R, \rho) \geq E(R, \rho)$.

To show the other direction, we define a universal encoding for the class of unifilar sources on state space \mathbb{S} with alphabet \mathbb{X} . Given a sequence x^n , encode each one of the $|\mathbb{S}|(|\mathbb{X}| - 1)$ source parameters $\{q_{x^n}(x | s)\}$ estimated from x^n . Each parameter requires $\log(n+1)$ bits. Then use $nH(Q_{x^n})$ bits to encode the index of x^n within the type T_{x^n} . The resulting description length can be set to

$$L_n^*(x^n) = nH(Q_{x^n}) + |\mathbb{S}|(|\mathbb{X}| - 1) \log(n+1),$$

where we have ignored constants arising from integral length constraints. We call this strategy the minimum description length coding and L_n^* the minimum description lengths.

L_n^* depends on x^n only through its type T_{x^n} . Moreover, there are at most $(n+1)^{|\mathbb{S}|(|\mathbb{X}|-1)}$ types. Using these facts, (27), and (28), we get

$$\mathbb{E} \left[\left(\min \left\{ 2^{L_n^*(X^n)}, 2^{nR} \right\} \right)^\rho \right] \quad (39)$$

$$\leq (n+1)^{(1+\rho)|\mathbb{S}|(|\mathbb{X}|-1)} \quad (40)$$

$$\cdot \max_{T_{x^n} \subseteq \mathbb{X}^n} P\{T_{x^n}\} \min \left\{ 2^{n\rho H(Q_{x^n})}, 2^{n\rho R} \right\} \quad (41)$$

$$\leq (n+1)^{(1+\rho)|\mathbb{S}|(|\mathbb{X}|-1)} 2^{nE(R, \rho)}. \quad (42)$$

Take logarithms and normalize by n to get

$$\limsup_{n \rightarrow \infty} E_{n,l}(R, \rho) \leq E(R, \rho).$$

This completes the proof. ■

The minimum description length coding works without knowledge of the true source parameters. Knowledge of the transition function ϕ is sufficient. In the context of guessing, the optimal attack strategy does not depend on knowledge of the source parameters. Interlacing the exhaustive key-search attack with the attack based on increasing description lengths is asymptotically optimal. Incidentally, the encryption strategy of Merhav and Arikan [2, Th. 1] uses only type information for encoding, and is applicable to unifilar sources. The same arguments in the proof of [2, Th. 1] go to show that their encryption strategy is asymptotically optimal for unifilar sources.

Let us define the quantity

$$E(\rho) \triangleq \max_Q [\rho H(Q) - D(Q \parallel P)]. \quad (43)$$

Observe that $E(\rho) = E(R, \rho)$ for $R \geq \log |\mathbb{X}|$, i.e., $E(\rho)$ determines the guessing exponent under perfect encryption. The following result identifies useful properties of these functions.

Proposition 13 *$E(\rho)$ is a convex function of ρ . $E(\rho, R)$ is a convex function of ρ and a concave function of R .* \square

Proof: Equation (43) is a maximum of affine functions of ρ and is therefore convex in ρ . The same is the case for $E(R, \rho)$. To see the concavity of $E(R, \rho)$ in R , write (31) as done in [2, Sec. IV] as

$$\begin{aligned} E(R, \rho) &= \max_Q \left[\rho \min_{0 \leq \theta \leq \rho} \left[\frac{\theta}{\rho} H(Q) + \left(1 - \frac{\theta}{\rho}\right) R \right] - D(Q \parallel P) \right] \\ &= \max_Q \min_{0 \leq \theta \leq \rho} [\theta H(Q) + (\rho - \theta) R - D(Q \parallel P)] \\ &= \min_{0 \leq \theta \leq \rho} \max_Q [\theta H(Q) + (\rho - \theta) R - D(Q \parallel P)] \quad (44) \\ &= \min_{0 \leq \theta \leq \rho} [E(\theta) + (\rho - \theta) R]. \quad (45) \end{aligned}$$

The maximization and minimization interchange in (44) is justified because the term within square brackets, sum of a scaled conditional entropy and the negative of a conditional divergence, is indeed concave in Q and affine in θ . Since (45) is a minimum of affine functions in R , it is concave in R . \blacksquare

It is easy to see the following fact for a unifilar source:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\sum_{x^n \in \mathbb{X}^n} P_n(x^n)^{1/(1+\rho)} \right)^{1+\rho} = E(\rho). \quad (46)$$

That the left side in (46) is at least as large as the right side follows from the proof in [6, Appendix B] and the observation that $\rho H(Q) - D(Q \parallel P)$ is continuous in Q and that the set of rational empirical PMFs Q_{x^n} is dense

in the set of unifilar sources with state space \mathbb{S} and alphabet \mathbb{X} , as $n \rightarrow \infty$. The other direction is an easy application of the method of types. The initial state which is implicit in P_n does not affect the value of the limit (as one naturally expects in this Markov case). In the memoryless case, i.e., when $s_i = x_i$, and $P(x|s)$ is independent of s , this quantity converges to $E(\rho) = \rho H_{1/(1+\rho)}(P)$ where $H_{1/(1+\rho)}(P)$ is the Rényi entropy of the DMS P on \mathbb{X} .

Analogous to a DMS case, we can characterize the behavior of $E(R, \rho)$ as a function of R for a particular source P .

Proposition 14 *For a given $\rho > 0$ and a unifilar source, let $E'(\rho)$ exist. Then*

$$E(R, \rho) = \begin{cases} \rho R, & R < H, \\ (\rho - \theta_0)R + E(\theta_0), & H \leq R \leq E'(\rho), \\ E(\rho), & R > E'(\rho) \end{cases}$$

where $\theta_0 \in [0, \rho]$ in the second case. \square

Proof: Indeed, from (45) it is clear by the continuity of the term within square brackets that for all values of R , $E(R, \rho) = (\rho - \theta_0)R + E(\theta_0)$ for some $\theta_0 \in [0, \rho]$, and the second case is directly proved.

Suppose $R < H$. Then we may choose $Q = P$ in (31) to get $E(R, \rho) \geq \rho R$. However, (25) indicates that $E(R, \rho) \leq \rho R$, which leads us to conclude that $E(R, \rho) = \rho R$ when $R < H$.

Next observe that $E(R, \rho) \leq E(\rho)$ is direct for all values of R , and in particular for $R > E'(\rho)$. To show the reverse direction, (45) yields

$$\begin{aligned} E(R, \rho) &= \min_{0 \leq \theta \leq \rho} [E(\theta) + (\rho - \theta)R] \\ &= E(\rho) + \min_{0 \leq \theta \leq \rho} (\rho - \theta) \left(R - \frac{E(\rho) - E(\theta)}{\rho - \theta} \right). \end{aligned}$$

The proof will be complete if we can show that the term within parentheses is nonnegative for $0 \leq \theta \leq \rho$. This holds because of the following. By the convexity of $E(\theta)$, the largest value of $(E(\rho) - E(\theta))/(\rho - \theta)$ for the given range of θ is $E'(\rho)$ (see for example, Royden [13, Lemma 5.5.16]), and this is upper bounded by R . \blacksquare

For a DMS, Merhav and Arikan [2] show that $E'(\rho) = H(P_\rho)$, where P_ρ is the PMF given by

$$P_\rho(x) = \frac{P(x)^{1/(1+\rho)}}{\sum_{a \in \mathbb{X}} P(a)^{1/(1+\rho)}}. \quad (47)$$

They also show that θ_0 is the unique solution to $R = H(P_{\theta_0})$.

5 Large Deviations Performance

5.1 General Sources With Memory

We now study the problem of large deviations in guessing and its relation to source compression. Our goal is to extend the large deviations results of Merhav and Arikan [2] to sources with memory using the tight relationship between guessing functions and length functions. We begin with the following general result.

Proposition 15 *1. When $B > R > 0$, there is an attack strategy that satisfies*

$$\sup_{f_n} P_n \{G(X^n | Y) \geq 2^{nB}\} = 0$$

for all sufficiently large n .

2. When $B \leq R$, there is an attack strategy that satisfies

$$\begin{aligned} \sup_{f_n} P_n \{G(X^n | Y) \geq 2^{nB}\} \\ \leq \min_{L_n} P_n \{L_n(X^n) \geq nB - 1\}. \end{aligned}$$

3. When $B < R$, there is an encryption function f_n such that

$$\begin{aligned} P_n \{G_{f_n}(X^n | Y) \geq 2^{nB}\} \\ \geq \frac{1}{3} \cdot \min_{L_n} P_n \{L_n(X^n) \geq nB + 1 + \log c_n\}. \end{aligned}$$

□

Remarks: When $B = R$, the large deviations behavior of guessing and coding may differ. If we define

$$F_n(R, B) \triangleq \inf_{f_n} \left[-\frac{1}{n} \log P_n \{G_{f_n}(X^n | Y) \geq 2^{nB}\} \right] \quad (48)$$

and

$$F_{n,l}(B) \triangleq \max_{L_n} \left[-\frac{1}{n} \log P_n \{L_n(X^n) \geq 2^{nB}\} \right], \quad (49)$$

then $F_n(R, B) = \infty$ for all sufficiently large n if $R < B$. When $R > B$, $F_n(R, B)$ is bounded between $F_{n,l}(B - 1/n)$ and $F_{n,l}(B + (1 + \log c_n)/n)$ ignoring vanishing terms.

Proof: Observe first that for any encryption function, the strategy (18) requires at most 2^{nR+1} guesses. If $B > R$, $2^{nB} > 2^{nR+1}$ for all sufficiently large n , and therefore

$$\sup_{f_n} P_n \{G(X^n | Y) \geq 2^{nB}\} = 0.$$

When $B \leq R$, the same strategy with an optimal L_n that minimizes $P_n\{L_n(X^n) \geq nB - 1\}$ requires $G(x^n | y) \leq 2 \min\{2^{L(x^n)}, 2^{nR}\}$ guesses. Hence

$$\{G(x^n | y) \geq 2^{nB}\} \subseteq \{L_n(x^n) \geq nB - 1\}$$

and therefore

$$P_n\{G(X^n | Y) \geq 2^{nB}\} \leq P_n\{L_n(X^n) \geq nB - 1\}.$$

Since this is true for any encryption function f_n , the second statement follows. The attack $G(\cdot | y)$ given by (18) interlaces guesses in the increasing order of the L_n that attains the minimum in $\min_{L_n} P_n\{L_n(X^n) \geq nB - 1\}$ with the exhaustive key-search strategy.

Next, let $B < R$ and consider the encryption strategy given in the proof of Proposition 8 with $N = M\lceil |\mathbb{X}|^n/M \rceil$ (with dummy messages possibly appended) and $M = 2^{nR}$. Let G_{P_n} denote guessing in the decreasing order of P_n -probabilities. Once again we refer to messages by their indices. For the optimal guessing strategy G_{f_n} , we have

$$\begin{aligned} & P_n\{G_{f_n}(X^n | Y) \geq 2^{nB}\} \\ &= \sum_{j=0}^{N/M-1} \sum_{i=2^{nB}-1}^{M-1} P_n\{X^n = jM + i\} \\ &\geq \sum_{j=0}^{N/M-1} P_n\{X^n = (j+1)M - 1\} (M - 2^{nB}) \\ &\geq \sum_{j=0}^{N/M-1} \sum_{i=0}^{M-1} P_n\{X^n = (j+1)M + i\} \frac{M - 2^{nB}}{M} \\ &= \left(1 - \frac{2^{nB}}{M}\right) \sum_{m=M}^{N-1} P_n\{X^n = m\} \\ &\geq \frac{1}{2} \sum_{m=M}^{N-1} P_n\{X^n = m\}, \end{aligned}$$

where the last inequality follows because $B < R$. (When $B = R$, the lower bound is 0 and this technique does not work). Also, rather trivially,

$$P_n\{G_{f_n}(X^n | Y) \geq 2^{nB}\} \geq \sum_{m=2^{nB}-1}^{M-1} P_n\{X^n = m\}.$$

Putting these together, we get

$$\begin{aligned} \sum_{m=2^{nB}-1}^{N-1} P_n\{X^n = m\} &= P_n\{G_{P_n}(X^n) \geq 2^{nB}\} \\ &\leq 3P_n\{G_{f_n}(X^n | Y) \geq 2^{nB}\}. \end{aligned}$$

Since $\{L_{G_{P_n}}(x^n) \geq nB + 1 + \log c_n\} \subseteq \{G_{P_n}(x^n) \geq 2^{nB}\}$, we get

$$\begin{aligned} & P_n\{G_{f_n}(X^n | Y) \geq 2^{nB}\} \\ & \geq \frac{1}{3} \cdot P_n\{L_{G_{P_n}}(X^n) \geq nB + 1 + \log c_n\} \\ & \geq \frac{1}{3} \cdot \min_{L_n} P_n\{L_n(X^n) \geq nB + 1 + \log c_n\}, \end{aligned}$$

and this concludes the proof. \blacksquare

5.2 Unifilar Sources

In this subsection, we specialize the result of Proposition 15 to unifilar sources.

Corollary 16 *For a unifilar source,*

$$F(R, B) \triangleq \lim_{n \rightarrow \infty} F_n(R, B) = \begin{cases} \infty, & B > R, \\ F(B), & B < R, \end{cases}$$

where

$$F(B) \triangleq \min_{Q: H(Q) \geq B} D(Q \| P)$$

is the source coding error exponent for the unifilar source. \square

Proof: This follows straightforwardly from the remarks immediately following Proposition 15 if we can show that $\lim_{n \rightarrow \infty} F_{n,l}(B) = F(B)$ and that $F(B)$ is continuous in $(0, \log |\mathbb{X}|)$. This was proved by Merhav in [6, Sec. III]. \blacksquare

We remark that the optimal attack strategy does not depend on the source parameters. Guessing in the increasing order description lengths, interlaced with the exhaustive key-search attack is an asymptotically optimal attack. Furthermore, as is the case for guessing moments, the encryption strategy of Merhav and Arikan [2, Th. 2] is easily verified to be an asymptotically optimal encryption strategy for unifilar sources when $B < R$.

$E(R, \rho)$ and $F(R, B)$ for unifilar sources are related via the Fenchel-Legendre transform, i.e.,

$$E(R, \rho) = \sup_{B > 0} [\rho B - F(R, B)]$$

and

$$F(R, B) = \sup_{\rho > 0} [\rho B - E(R, \rho)].$$

The proof is identical to that of [2, Th. 3] where this result is proved for DMSs.

5.3 Finite-State Sources

We now consider the larger class of finite state sources. The Lempel-Ziv coding strategy [5] asymptotically achieves the entropy rate of a finite-state source without knowledge of the source parameters. It is therefore natural to consider its use in attacking a cipher system that attempts to securely transmit a message put out by a finite-state source. Our next goal is to show that guessing in the increasing order of Lempel-Ziv coding lengths has an interesting universality property.

Let $U_{LZ} : \mathbb{X}^n \rightarrow \mathbb{N}$ be the length function for the Lempel-Ziv code [5]. The following theorem due to Merhav [6] indicates that the Lempel-Ziv algorithm is asymptotically optimal in achieving the minimum probability of buffer overflow.

Theorem 17 (Merhav [6]) *For any length function L_n , every finite-state source P_n , every $B_n \in (nH, n \log |\mathbb{X}|)$ where H is the entropy-rate of the source P_n , and all sufficiently large n ,*

$$\begin{aligned} P_n \{U_{LZ}(X^n) \geq B_n + n\varepsilon(n)\} \\ \leq (1 + \delta(n)) \cdot P_n \{L_n(X^n) \geq B_n\} \end{aligned} \quad (50)$$

where $\varepsilon(n) = \Theta(1/\sqrt{\log n})$ is a positive sequence that depends on $|\mathbb{X}|$ and $|\mathbb{S}|$, and $\delta(n) = n^2 2^{-n\varepsilon(n)}$. \square

Remark: Merhav's result [6, Th. 1] assumes that $B_n = nB$ for a constant $B \in (H, \log |\mathbb{X}|)$, but the proof is valid for any sequence $B_n \in (nH, n \log |\mathbb{X}|)$.

Let G_{LZ} be the short-hand notation for the more cumbersome $G_{U_{LZ}}$, the guessing function associated with U_{LZ} . Let c_n be as given in (4) with \mathbb{X}^n replacing \mathbb{X} . Furthermore, for the key-constrained cipher system, let $G_{LZ}(\cdot | y)$ denote the attack of guessing in the order prescribed by G_{LZ} interlaced with the exhaustive key-search attack. Observe that $G_{LZ}(\cdot | y)$ needs knowledge of f_n .

Theorem 18 *For any guessing function G_n , every finite-state source P_n , every $B \in (H, \log |\mathbb{X}|)$ where H is the entropy-rate of the source P_n , and all sufficiently large n ,*

$$\begin{aligned} P_n \{n^{-1} \log G_{LZ}(X^n) \geq B + \varepsilon(n) + \gamma(n)\} \\ \leq (1 + \delta(n)) \cdot P_n \{n^{-1} \log G_n(X^n) \geq B\} \end{aligned} \quad (51)$$

where $\varepsilon(n)$ and $\delta(n)$ are the sequences in (50), and $\gamma(n) = (1 + \log c_n)/n = \Theta(n^{-1} \log n)$.

For the key-rate constrained cipher system, let $B < R$. Then for any encryption function, we have

$$\begin{aligned} & P_n \{n^{-1} \log G_{LZ}(X^n | Y) \geq B + 1/n + \varepsilon(n) + \gamma(n)\} \\ & \leq 3(1 + \delta(n)) \cdot \sup_{f_n} P_n \{n^{-1} \log G_{f_n}(X^n | Y) \geq B\} \end{aligned} \quad (52)$$

for all sufficiently large n . \square

Remark: Thus the Lempel-Ziv coding strategy provides an asymptotically optimal universal attack strategy for the class of finite-state sources, in the sense of attaining the limiting value of (48), if the limit exists.

Proof: Observe that

$$\begin{aligned} & (1 + \delta(n)) P_n \{G_n(X^n) \geq 2^{nB}\} \\ & \geq (1 + \delta(n)) P_n \{L_{G_n}(X^n) \geq nB + 1 + \log c_n\} \end{aligned} \quad (53)$$

$$\geq P_n \{U_{LZ}(X^n) \geq nB + 1 + \log c_n + n\varepsilon(n)\} \quad (54)$$

$$\geq P_n \{G_{LZ}(X^n) \geq 2^{nB+n\varepsilon(n)+n\gamma(n)}\}, \quad (55)$$

where (53) follows from the first inclusion in (10), and (54) from (50). The last inequality (55) follows from (11). This proves the first part.

To show the second part, we use Proposition 15.3 and Theorem 17 as follows: for all sufficiently large n ,

$$\begin{aligned} & 3(1 + \delta(n)) \sup_{f_n} P_n \{G_{f_n}(X^n | Y) \geq 2^{nB}\} \\ & \geq (1 + \delta(n)) P_n \{L_n(X^n) \geq nB + n\gamma(n)\} \\ & \geq P_n \{U_{LZ}(X^n) \geq nB + n\gamma(n) + n\varepsilon(n)\} \\ & \geq P_n \{G_{LZ}(X^n | Y) \geq 2^{nB+1+n\gamma(n)+n\varepsilon(n)}\} \end{aligned}$$

where the last inequality holds for any arbitrary encryption function with $G_{LZ}(\cdot | y)$ being the interlaced attack strategy. \blacksquare

Observe that $\varepsilon(n) + \gamma(n) = \Theta(1/\sqrt{\log n})$. For unifilar sources, a result analogous to Theorem 18 can be shown with $\varepsilon(n) + \gamma(n) = \Theta(n^{-1} \log n)$. Guessing for this class of sources proceeds in the order of increasing description lengths. This conclusion follows from a result analogous to Theorem 17 on the asymptotic optimality of minimum description coding (see Merhav [6, Sec. III]).

5.4 Competitive Optimality

We now demonstrate a competitive optimality property for G_{LZ} . From [6, eqn. (28)] extended to finite-state sources, we have for any competing code

L_n

$$\begin{aligned} & P_n\{U_{LZ}(X^n) > L_n(X^n) + n\varepsilon(n)\} \\ & \leq P_n\{U_{LZ}(X^n) < L_n(X^n) + n\varepsilon(n)\} \end{aligned} \quad (56)$$

where $\varepsilon(n) = \Theta((\log \log n)/(\log n))$. From (8) and (6), we get

$$U_{LZ}(x^n) \geq \log G_{LZ}(x^n)$$

and

$$\log G(x^n) \geq L_G(x^n) - 1 - \log c_n,$$

respectively. We therefore conclude that

$$\begin{aligned} & \{\log G_{LZ}(x^n) > \log G(x^n) + n(\varepsilon(n) + \gamma(n))\} \\ & \subseteq \{U_{LZ}(x^n) > L_G(x^n) + n\varepsilon(n)\} \end{aligned}$$

and that

$$\begin{aligned} & \{U_{LZ}(x^n) < L_G(x^n) + n\varepsilon(n)\} \\ & \subseteq \{\log G_{LZ}(x^n) < \log G(x^n) + n(\varepsilon(n) + \gamma(n))\}. \end{aligned}$$

From these two inclusions and (56), we easily deduce the following result.

Theorem 19 *For any finite-state source and any competing guessing function G , we have*

$$\begin{aligned} & P_n\{\log G_{LZ}(X^n) > \log G(X^n) + n\varepsilon'(n)\} \\ & \leq P_n\{\log G_{LZ}(X^n) < \log G(X^n) + n\varepsilon'(n)\} \end{aligned}$$

where $\varepsilon'(n) = \varepsilon(n) + \gamma(n)$. □

For unifilar sources, the above sequence of arguments for minimum description length coding and [6, eqn. (28)] imply that we may take $\varepsilon'(n) = \Theta(n^{-1} \log n)$.

6 Concluding Remarks

In this report, we studied two measures of cryptographic security based on guessing, for sources with memory. The first one was based on guessing moments and the second on large deviations performance of the number of guesses. We identified an asymptotically optimal encryption strategy that orders the messages in the decreasing order of their probabilities, enumerates them, and then encrypts as many least-significant bits as there are key bits. We also identified an optimal attack strategy based on a length function that

attains the optimal value for a source coding problem. Both these strategies need knowledge of the message probabilities.

We then specialized our results to the case of unifilar sources, gave formulas for computing the two measures of performance, and argued that the optimal encryption strategy as well as the optimal attack strategy depended on the source parameters only through the number of states and letters, i.e., the optimal encryption and attack strategies are universal for this class.

We also showed that an attack strategy based on the Lempel-Ziv coding lengths is asymptotically optimal for the class of finite state sources. Finally, we provided competitive optimality results for guessing in the order of increasing description lengths and Lempel-Ziv lengths.

We end this report with a short list of related open problems.

- Consider a modification to the encryption technique of Proposition 8 where the messages are enumerated in the increasing order of their Lempel-Ziv lengths instead of message probabilities. Does this ordering lead to an asymptotically optimal encryption strategy? Such a strategy would not depend on the specific knowledge of source parameters.
- It would be of interest to see if the results on guessing moments for unifilar sources can be extended to finite-state sources.
- The large deviations behavior of guessing when $B = R$ is not well-understood and might be worth investigating.
- As mentioned in [2], one might wish to consider a scenario where only a noisy version of the cryptogram is available to the attacker. The degradation in the attacker's performance could be quantified.

Acknowledgements This work was supported by the Defence Research & Development Organisation (DRDO), Ministry of Defence, Government of India, under the DRDO-IISc Programme on Advanced Research in Mathematical Engineering, and by the University Grants Commission under Grant Part (2B) UGC-CAS-(Ph.IV).

The author thanks an anonymous referee for pointing out reference [8] and for suggestions that simplified the proof of Proposition 2.

References

- [1] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, no. 3, pp. 565–715, Oct. 1949.
- [2] N. Merhav and E. Arikan, "The Shannon cipher system with a guessing wiretapper," *IEEE Trans. Inform. Theory*, vol. 45, no. 6, pp. 1860–1866, Sep. 1999.
- [3] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 99–105, Jan. 1996.

- [4] L. L. Campbell, "A coding theorem and Rényi's entropy," *Information and Control*, vol. 8, pp. 423–429, 1965.
- [5] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inform. Theory*, vol. 24, no. 5, pp. 530–536, Sept. 1978.
- [6] N. Merhav, "Universal coding with minimum probability of codeword length overflow," *IEEE Trans. Inform. Theory*, vol. 37, no. 3, pp. 556 – 563, May 1991.
- [7] M. J. Weinberger, J. Ziv, and A. Lempel, "On the optimal asymptotic performance of universal ordering and of discrimination of individual sequences," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 380–385, Mar. 1992.
- [8] A.D.Wyner, "An upper bound on the entropy series," *Information and Control*, vol. 20(2), pp. 176–181, Mar. 1972.
- [9] E. Arikan and N. Merhav, "Guessing subject to distortion," *IEEE Trans. Inform. Theory*, vol. IT-44, pp. 1041–1056, May 1998.
- [10] R. Sundaresan, "Guessing under source uncertainty," *IEEE Trans. Inform. Theory*, vol. 53, no. 1, pp. 269–287, Jan. 2007.
- [11] R. Ash, *Information Theory*. Interscience Publishers, 1965.
- [12] M. Gutman, "Asymptotically optimal classification for multiple tests with empirically observed statistics," *IEEE Trans. Inform. Theory*, vol. 35, no. 2, pp. 401–408, Mar. 1989.
- [13] H. L. Royden, *Real Analysis*. New York: Macmillan, 1988.