The Shannon Cipher System with a Guessing Wiretapper: General Sources

Manjesh Kumar Hanawal
Department of ECE
Indian Institute of Science
Bangalore, Karnataka 560012, India
Email: manjesh@ece.iisc.ernet.in

Rajesh Sundaresan
Department of ECE
Indian Institute of Science
Bangalore, Karnataka 560012, India
Email: rajeshs@ece.iisc.ernet.in

***Abstract*—The Shannon cipher system is studied in the context of general sources using a notion of computational secrecy introduced by Merhav & Arikan. Bounds are derived on limiting exponents of guessing moments for general sources. The bounds are shown to be tight for iid, Markov, and unifilar sources, thus recovering some known results. A close relationship between error exponents and correct decoding exponents for fixed rate source compression on the one hand and exponents for guessing moments on the other hand is established.**

I. INTRODUCTION

We consider the classical cipher system of Shannon [1]. Let $X^n = (X_1, \cdots, X_n)$ 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 encryption function f_n is invertible for any fixed 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. The key rate for the system is $k/n = R$. A wiretapping attacker 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 provides the natural test mechanism. We assume that the attacker is allowed an unlimited number of guesses.

Merhav & Arikan [2] studied discrete memoryless sources (DMS) in the above setting and characterized the best attainable moments of the number of guesses required by an attacker. In particular, they showed that for a DMS with the governing single letter PMF P on $\mathbb{X}$, the value of the optimal exponent for the ρth moment ($\rho > 0$) is given by

$$E(R, \rho) = \max_{Q} \left\{ \rho \min\{H(Q), R\} - D(Q \parallel P) \right\}. \qquad (1)$$

The maximization is over all PMFs Q on $\mathbb{X}$, $H(Q)$ is the Shannon entropy of Q, and $D(Q \parallel P)$ is the Kullback-Leibler divergence between Q and P. They also showed that $E(R, \rho)$ increases linearly in R for $R \le H(P)$, continues to increase in a concave fashion for $R \in [H(P), H^{'}]$, where $H^{'}$ is a threshold, and is constant for $R > H^{'}$. Unlike the classical equivocation rate analysis, atypical sequences do affect the behavior of $E(R, \rho)$ for $R \in [H(P), H^{'}]$ and perfect secrecy is obtained, i.e., cryptogram is uncorrelated with the message, only for $R > H^{'} > H(P)$. Merhav & Arikan also determined the best achievable performance based on the probability of a large deviation in the number of guesses, and showed that it equals the Legendre-Fenchel transform of $E(R, \rho)$ as a function of ρ. Sundaresan [3] extended the above results to unifilar sources. Hayashi & Yamamoto [4] proved coding theorems for the Shannon cipher system with correlated outputs (X^n, Z^n) where the wiretapper is interested in X^n while the receiver in Z^n.

In this paper, we extend Merhav & Arikan's notion of computational secrecy to general sources. One motivation is that secret messages typically come from the natural languages which can be well-modelled as sources with memory, for e.g., a Markov source of appropriate order. Another motivation is that the study of general sources clearly brings out the connection between guessing and compression, as discussed next.

As with other studies of general sources, *information spectrum* plays crucial role in this paper. We show that $E(R, \rho)$ is closely related to (a) the error exponent of a rate-R source code, and (b) the correct decoding exponent of a rate-R source code, when exponentiated probabilities are considered (see Sec. III-A2). In particular, the exponents in (a) and (b) appear in the first and second terms when we rewrite $E(R, \rho)$ for a DMS as

$$E(R, \rho) = \max \left\{ \rho R - \min_{Q: H(Q) > R} D(Q \parallel P), \right.$$
$$\left. \min_{Q: H(Q) \le R} \{\rho H(Q) - D(Q \parallel P)\} \right\}.$$

This brings out the fundamental connection between source coding exponents and key-rate constrained guessing exponents. Further, unlike the case for the probability of a large deviation in the number of guesses [2, Sec. V], both the error exponent and the correct decoding exponent determine

978-1-4244-4313-0/09/$25.00 ©2009 IEEE

$E(R,\rho)$. We extend the above result to general sources by getting upper and lower bounds on $E(R,\rho)$. We then show that these are tight for DMS, Markov and unifilar sources. The bounds may be of interest even if they are not tight because the upper bound specifies the amount of effort need by an attacker and the lower bound specifies the secrecy strength of the cryptosystem to a designer.

The limiting case as $\rho \downarrow 0$ in (b) yields classical framework for probability of correct decoding. This special case is related to the work of Han [5] and Iriyama [6] who studied the dual problem of rates required to meet a specified error exponent or a specified correct decoding exponent.

The paper is organized as follows. Section II relates our problem to a modification of Campbell's compression problem [7]. Section III gives bounds on the limits of exponential rate of guessing moments, in terms of information spectrum quantities. Section IV evaluates the bounds for some specific examples. All proofs can be found in the technical report [8].

II. Guessing with key-rate constraints and source compression

In this section, we state the problem precisely, and establish a connection between guessing and source compression subject to a new cost criterion.

Let $\mathbb{X}^n$ denote the set of messages and $\mathcal{M}(\mathbb{X}^n)$ the set of PMFs on $\mathbb{X}^n$. By a source, we mean a sequence of PMFs $(P_n : n \in \mathbb{N})$, where $P_n \in \mathcal{M}(\mathbb{X}^n)$ [1]. Let X^n denote a message put out by the source and U^k the secure key of k 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. The key rate for the system is $k/n = R$.

For a given cryptogram $Y = y$, define a *guessing strategy*

$$G_n(\cdot \mid y) : \ \mathbb{X}^n \to \{1, 2, \cdots |\mathbb{X}|^n\}$$

as a bijection that denotes the order in which elements of $\mathbb{X}^n$ are guessed. $G_n(x^n \mid y) = g$ indicates that x^n is the gth guess, when the cryptogram is y. With knowledge of P_n, the encryption function f_n, and the cryptogram Y, the attacker can completely calculate all the posterior probabilities of plaintexts $P_{X^n|Y}(\cdot \mid y)$ given the cryptogram. The attacker's optimal guessing strategy is then to guess in the decreasing order of these posterior probabilities $P_{X^n|Y}(\cdot \mid y)$. Let us denote this optimal attack strategy as G_{f_n}. Let $(f_n : n \in \mathbb{N})$ denote the sequence of encryption functions known to the attacker, where $\mathbb{N}$ denotes the set of natural numbers. We assume that attacker employs the optimal guessing strategy.

For a given $\rho > 0$, key rate $R > 0$, define the normalized guessing exponent

$$E_n^g(R,\rho) := \sup_{f_n} \frac{1}{n} \log \mathbb{E}\left[G_{f_n}(X^n \mid Y)^\rho\right].$$

The supremum is taken over all encryption functions. Further define performance limits of guessing moments as in [2]:

$$E_u^g(R,\rho) := \limsup_{n\to\infty} E_n^g(R,\rho) \tag{2}$$

$$E_l^g(R,\rho) := \liminf_{n\to\infty} E_n^g(R,\rho). \tag{3}$$

We next define the related compression quantities. A length function $L_n : \mathbb{X}^n \to \mathbb{N}$ is a mapping that satisfies Kraft's inequality:

$$\sum_{x^n \in \mathbb{X}^n} \exp\{-L_n(x)\} \le 1.$$

For a given $\rho > 0$, code rate $R > 0$, define

$$E_n^s(R,\rho) := \min_{L_n} \frac{1}{n} \log \mathbb{E}\left[\exp\left\{\rho \min\left\{L_n(X^n), nR\right\}\right\}\right]. \tag{4}$$

The minimum is taken over all length functions. We may interpret the cost of using length $L_n(x^n)$ as $\min\{\exp\{L_n(x^n), nR\}\}$, i.e., the cost is exponential in L_n, but saturates at $\exp\{nR\}$ and so all lengths larger than $\exp\{nR\}$ enjoy a saturated cost. Then $E_n^s(R,\rho)$ is the minimum normalized exponent of the ρth moment of this new compression cost. In analogy with (2) and (3) we define

$$E_u^s(R,\rho) = \limsup_{n\to\infty} E_n^s(R,\rho)$$

$$E_l^s(R,\rho) = \liminf_{n\to\infty} E_n^s(R,\rho)$$

The following equivalence between compression and guessing is immediate from proof of [3, Cor. 9].

Theorem 1 (Guessing-Compression Equivalence): For any $\rho > 0$ and $R > 0$, we have $E_u^s(R,\rho) = E_u^g(R,\rho)$ and $E_l^s(R,\rho) = E_l^g(R,\rho)$.

Thus, the problem of finding the optimal guessing exponent is the same as that of finding the optimal exponent for the coding problem in (4). When $R \ge \log|\mathbb{X}|$, the coding problem in (4) reduces to the one considered by Campbell in [7].

In the rest of the paper we focus on the equivalent compression problem and find bounds on E_u^s and E_l^s.

III. Growth Exponent for the Modified Compression Problem

We begin with some words on notation. Recall that $\mathcal{M}(\mathbb{X}^n)$ denotes the set of PMFs on $\mathbb{X}^n$. The Shannon entropy for a $P_n \in \mathcal{M}(\mathbb{X}^n)$ is

$$H(P_n) = -\sum_{x^n \in \mathbb{X}^n} P_n(x^n) \log P_n(x^n)$$

and the Rényi entropy of order $\alpha \ne 1$ is

$$H_\alpha(P_n) = \frac{1}{1-\alpha} \log\left(\sum_{x^n \in \mathbb{X}^n} P_n(x^n)^\alpha\right). \tag{5}$$

The Kullback-Leibler divergence or relative entropy between two PMFs Q_n and P_n is

$$D(Q_n \parallel P_n) = \begin{cases} \displaystyle\sum_{x^n \in \mathbb{X}^n} Q_n(x^n) \log \frac{Q_n(x^n)}{P_n(x^n)}, & \text{if } Q_n \ll P_n, \\ \infty, & \text{otherwise,} \end{cases}$$

[1] Sometimes we use P_{X^n} in place of P_n when we refer to the distribution of random vector X^n.

where $Q_n \ll P_n$ means Q_n is absolutely continuous with respect to P_n. $(X^n : n \in \mathbb{N})$ denotes a sequence of random variables on $\mathbb{X}^n$ with corresponding sequence of probability measures denoted by $\mathbf{X} := (P_{X^n} : n \in \mathbb{N})$. Thus $\mathbf{X}$ is a source and X^n its n-letter message output. Abusing notation, we let $\mathcal{M}(\mathbb{X}^{\mathbb{N}})$ denote the set of all sequences $\mathbf{Y} = (P_{Y^n} : n \in \mathbb{N})$ of probability measures, and for each $\mathbf{B} := (B_n \subseteq \mathbb{X}^n : n \in \mathbb{N})$, we define

$$\mathcal{M}(\mathbf{B}) := \left\{ \mathbf{Y} \in \mathcal{M}(\mathbb{X}^{\mathbb{N}}) : \lim_{n\to\infty} P_{Y^n}(B_n) = 1 \right\}.$$

In the rest of this section $\mathbf{X}$ is a fixed source. For any $\mathbf{Y} \in \mathcal{M}(\mathbf{B})$ and $\rho > 0$, define

$$E_u(\mathbf{Y}, \mathbf{X}, \rho) := \limsup_{n\to\infty} \frac{1}{n}\{\rho H(P_{Y^n}) - D(P_{Y^n} \parallel P_{X^n})\}$$

$$E_l(\mathbf{Y}, \mathbf{X}, \rho) := \liminf_{n\to\infty} \frac{1}{n}\{\rho H(P_{Y^n}) - D(P_{Y^n} \parallel P_{X^n})\}.$$

We first state an upper bound on E_u^s.

Proposition 2 (Upper Bound): Let $R > 0$ and $\rho > 0$. Then

$$E_u^s(R, \rho) \leq \min_{0\leq\theta\leq\rho} \left[(\rho - \theta)R + \max_{\mathbf{Y}\in\mathcal{M}(\mathbb{X}^{\mathbb{N}})} E_u(\mathbf{Y}, \mathbf{X}, \theta) \right]$$

A. Lower Bound on E_l^s

We now state a lower bound on E_l^s. For a given distribution P_{Y^n}, let $T_R(Y^n)$ denote the first $M := \lfloor |\mathbb{X}|^{nR} \rfloor$ elements, when they are arranged in the decreasing order of probabilities. We denote the probability of this set by F_{Y^n}, i.e.,

$$F_{Y^n} = \sum_{x^n \in T_R(Y^n)} P_{Y^n}(x^n),$$

and the probability of complement of this set $T_R^c(Y^n)$ by $F_{Y^n}^c$.

Proposition 3 (Lower Bound): For a given $\rho > 0$ and rate $R > 0$, we have

$$E_l^s(R, \rho) \geq \max \left\{ \rho R + \liminf_{n\to\infty} \frac{1}{n} \log F_{X^n}^c, \right.$$
$$\left. (1 + \rho) \liminf_{n\to\infty} \frac{1}{n} \log \sum_{x^n \in T_R(X^n)} P_{X^n}^{\frac{1}{1+\rho}}(x^n) \right\}. \quad (6)$$

Remark 1: The first term contains limit infimum of the error exponent for a rate-R source code. The second exponent is the correct decoding exponent for a rate-R code when $\rho \downarrow 0$.

In the subsequent subsections we further lower bound each of the two terms under max on the right side of (6). For an arbitrary source we first recall the source coding error exponent. We also identify the growth rate of sum of exponentiated probabilities of the correct decoding set. We then relate them to the terms in the lower bound obtained in (6). We largely follow the approach and notation of Iriyama [6], which we now describe.

For the given $\mathbf{X}$ and a $\mathbf{Y} = (P_{Y^n} : n \in \mathbb{N})$, we define the upper divergence $D_u(\cdot \parallel \cdot)$ and lower divergence $D_l(\cdot \parallel \cdot)$ by

$$D_u(\mathbf{Y} \parallel \mathbf{X}) := \limsup_{n\to\infty} \frac{1}{n} D(P_{Y^n} \parallel P_{X^n})$$

$$D_l(\mathbf{Y} \parallel \mathbf{X}) := \liminf_{n\to\infty} \frac{1}{n} D(P_{Y^n} \parallel P_{X^n}).$$

For a $\mathbf{Y} = (P_{Y^n} : n \in \mathbb{N})$, denote *spectral sup-entropy-rate* [5, Sec. II], [9] as

$$\overline{H}(\mathbf{Y}) := \inf \left\{ \theta : \lim_{n\to\infty} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{Y^n}(Y^n)} > \theta \right\} = 0 \right\}.$$

Also define, as in [6, Sec. II], the following quantity which determines the performance under mismatched compression:

$$\underline{R}(\mathbf{Y}, \mathbf{X}) := \sup \left\{ \theta : \lim_{n\to\infty} \Pr \left\{ \frac{1}{n} \log \frac{1}{P_{X^n}(Y^n)} < \theta \right\} = 0 \right\}.$$

1) Decoding Error Exponent: In this subsection we recall the decoding error exponent for fixed-rate encoding of an arbitrary source. We identify the first term in (6) as composed of the exponent of minimum probability of decoding error, and lower bound it, or alternatively upper bound the error exponent. This is made precise in the following definitions. The key rate R plays the role of source coding rate.

By an (n, M_n, ϵ_n)-code we mean an encoding mapping

$$\phi_n : \mathbb{X}^n \to \{1, 2, \cdots, M_n\}$$

and a decoding mapping

$$\psi_n : \{1, 2, \cdots M_n\} \to \mathbb{X}^n$$

with probability of error $\epsilon_n := \Pr\{\psi_n(\phi_n(X^n)) \neq X^n\}$. R is r-achievable if for all $\eta > 0$ there exists a sequence of (n, M_n, ϵ_n)-codes such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \frac{1}{\epsilon_n} \geq r \quad \textit{and} \quad \limsup_{n\to\infty} \frac{1}{n} \log M_n \leq R + \eta. \quad (7)$$

The *infimum fixed-length coding rate* for exponent r is

$$\hat{R}(r|\mathbf{X}) = \inf\{R : R \text{ is } r\text{-achievable}\}.$$

On the other hand, the *supremum fixed-length coding exponent* for rate R is

$$\hat{E}(R|\mathbf{X}) = \sup\{r : R \text{ is } r\text{-achievable}\}.$$

Han [9, Sec. 1.9] and Iriyama [6] use a pessimistic definition for fixed rate source coding, i.e., the limit infimum in (7), and obtain expressions for the infimum coding rate. For our bounds we need optimistic definitions. Iriyama [6, Eqn. (13)] obtained a lower bound on the infimum coding rate $\hat{R}(r|\mathbf{X})$ under the optimistic definitions. We however work with the error exponent, and obtain an upper bound on supremum coding exponent. This suffices to lower bound the first term in (6).

Obviously, the best exponent is obtained by encoding only the highest M realizations and hence

$$\hat{E}(R|\mathbf{X}) = \limsup_{n\to\infty} \frac{1}{n} \log \frac{1}{F_{X^n}^c}$$

so that

$$-\hat{E}(R|\mathbf{X}) = \liminf_{n\to\infty} \frac{1}{n} \log F_{X^n}^c.$$

The following Proposition upper bounds the supremum coding exponent.

Proposition 4: For any rate $R > 0$,

$$\hat{E}(R|\mathbf{X}) \leq \inf_{\mathbf{Y}: \underline{R}(\mathbf{Y}, \mathbf{X}) - D_u(\mathbf{Y}\parallel\mathbf{X}) > R} D_u(\mathbf{Y} \parallel \mathbf{X}).$$

Remark 2: Notice that when $R \geq \log |\mathbb{X}|$, we have an infimum over an empty set and hence $\hat{E}(R|\mathbf{X}) = \infty$.

2) Correct Decoding Exponent: We now study a generalization of the exponential rate for probability of correct decoding.

For a given (n, M_n, ϵ_n)-code, let

$$A_n := \{x^n \in \mathbb{X}^n : \psi_n(\phi_n(x^n)) = x^n\}$$

denote the set of correctly decoded sequences. For a given $\rho > 0$, R is (r, ρ)-admissible if for every $\eta > 0$ there exists a sequence of (n, M_n, ϵ_n)-codes such that

$$(1+\rho) \liminf_{n\to\infty} \frac{1}{n} \log \sum_{x^n \in A_n} P_{X^n}^{\frac{1}{1+\rho}}(x^n) \geq r \tag{8}$$

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \leq R + \eta. \tag{9}$$

Define *infimum fixed-length admissible rate* for a given r and $\rho > 0$ as

$$R^*(r, \rho|\mathbf{X}) = \inf\{R : R \text{ is } (r,\rho)\text{-admissible}\}.$$

Clearly, the set $\{R : R \text{ is } (r,\rho)\text{-admissible}\}$ is closed and so $R^*(r, \rho|\mathbf{X})$ is (r, ρ)-admissible.
Define *supremum fixed-length coding exponent* for a given R and ρ as

$$E^*(R, \rho|\mathbf{X}) = \sup\{r : R \; is \; (r,\rho)\text{-admissible}\}.$$

Remark 3: The choice of limit infimum in (8) makes the definition of admissibility pessimistic. For $\rho \downarrow 0$ the above definitions reduce to the special case of exponential rate for probability of correct decoding (see [9, Sec. 1.10]).

Clearly, A_n should be $T_R(X^n)$ to maximise the left side of (8) and hence

$$E^*(R, \rho|\mathbf{X}) = (1+\rho) \liminf_{n\to\infty} \frac{1}{n} \log \sum_{x^n \in T_R(X^n)} P_{X^n}^{\frac{1}{1+\rho}}(x^n).$$

The following Proposition gives an expression for $E^*(R, \rho|\mathbf{X})$ and generalizes [6, Thm. 4] to any arbitrary $\rho > 0$. En route to its derivation we find the expression for $R^*(r, \rho|\mathbf{X})$.

Proposition 5: For any $\rho > 0$ and $r > 0$,

$$R^*(r, \rho|\mathbf{X}) = \inf_{\mathbf{Y}: E_l(\mathbf{Y}, \mathbf{X}, \rho) \geq r} \overline{H}(\mathbf{Y}) \tag{10}$$

$$E^*(R, \rho|\mathbf{X}) = \sup_{\overline{H}(\mathbf{Y}) \leq R} E_l(\mathbf{Y}, \mathbf{X}, \rho). \tag{11}$$

B. Summary of Bounds on E_u^s and E_l^s

We now combine the Propositions 2-5 of the previous subsections to obtain the main result of the paper.

Theorem 6: For a given $\rho > 0$ and $R > 0$,

$$\begin{aligned} \max\Bigg\{\rho R - &\inf_{\mathbf{Y}: \underline{R}(\mathbf{Y},\mathbf{X}) - D_u(\mathbf{Y}\|\mathbf{X}) > R} D_u(\mathbf{Y} \parallel \mathbf{X}), \\ &\qquad\qquad \sup_{\overline{H}(\mathbf{Y}) \leq R} E_l(\mathbf{Y}, \mathbf{X}, \rho)\Bigg\} \\ \leq\;& E_l^s(R, \rho) \leq E_u^s(R, \rho) \\ \leq\;& \min_{0 \leq \theta \leq \rho} \left\{(\rho - \theta)R + \max_{\mathbf{Y}} E_u(\mathbf{Y}, \mathbf{X}, \theta)\right\}. \end{aligned} \tag{12}$$

IV. EXAMPLES

In this section we evaluate the bounds for some examples where they are tight, and recover some known results.

Example 1 (Perfect Secrecy): First consider the perfect secrecy case, for example $R \geq \log|\mathbb{X}|$. Because of Remark 2 and because we may take $\theta = \rho$ in the upper bound in (12), the limiting exponential rate of guessing moments simplifies to

$$\begin{aligned} \sup_{\mathbf{Y}} E_l(\mathbf{Y}, \mathbf{X}, \rho) &\leq E_l^s(R, \rho) \\ &\leq E_u^s(R, \rho) \leq \max_{\mathbf{Y}} E_u(\mathbf{Y}, \mathbf{X}, \rho). \end{aligned}$$

In a related work we proved in [10, Prop. 7] that whenever the *information spectrum* of the source satisfies the large deviation property with rate function I, the lower and upper bounds coincide, and limiting guessing exponent equals the Legendre-Fenchel dual of the scaled rate function $I_1(t) := (1+\rho)I(t)$, i.e.,

$$E_u^s(\rho) = E_l^s(\rho) = \sup_{t \in \mathbb{R}}\{\rho t - I_1(t)\}.$$

In the next examples, we consider the case $R < \log|\mathbb{X}|$.

Example 2 (An iid source): This example was first studied by Merhav & Arikan [2]. Recall that an iid source is one for which $P_n(x^n) = \prod_{i=1}^n P_1(x_i)$, where P_1 denotes the marginal of X_1. We will now evaluate each term in (12).

We first argue that

$$\begin{aligned} &\inf_{\mathbf{Y}: \underline{R}(\mathbf{Y},\mathbf{X}) - D_u(\mathbf{Y}\|\mathbf{X}) > R} D_u(\mathbf{Y} \parallel \mathbf{X}) \\ &\qquad = \inf_{P_Y : H(P_Y) > R} D(P_Y \parallel P_1). \end{aligned} \tag{13}$$

To prove "$\geq$" in (13) we use the following result:

$$\inf_{\mathbf{Y}: \underline{R}(\mathbf{Y},\mathbf{X}) - D_u(\mathbf{Y}\|\mathbf{X}) > R} D_u(\mathbf{Y} \parallel \mathbf{X}) \geq \inf_{\mathbf{Y}: H_l(\mathbf{Y}) > R} D_u(\mathbf{Y} \parallel \mathbf{X}), \tag{14}$$

where $H_l(\mathbf{Y}) = \liminf_{n\to\infty} \frac{1}{n} H(P_{Y^n})$. Proof of above inequality follows from a straightforward manipulation of [6, Cor. 1], and is therefore omitted. From (14) it is sufficient to prove

$$\inf_{\mathbf{Y}: H_l(\mathbf{Y}) > R} D_u(\mathbf{Y} \parallel \mathbf{X} \geq \inf_{P_Y : H(P_Y) > R} D(P_Y \parallel P_1). \tag{15}$$

Let $\mathbf{Y}$ be such that $H_l(\mathbf{Y}) > R$. Construct a source $\hat{\mathbf{Y}}$ such that, $P_{\hat{Y}_i} = P_{Y_i}$ for $1 \leq i \leq n$ and $\hat{Y}_1, \hat{Y}_2, \cdots, \hat{Y}_n$ are independent. Let $\mathbf{Z}$ be another source such that, $Z_1, Z_2, \cdots, Z_n$ is an iid sequence with distribution

$$P_{Z_j} = \frac{1}{n} \sum_{i=1}^n P_{Y_i}, \quad j = 1, 2, \cdots, n.$$

Now, by convexity of divergence, we have

$$\begin{aligned} D(P_{Y^n} \parallel P_{X^n}) &\geq D(P_{\hat{Y}^n} \parallel P_{X^n}) \geq D(P_{Z^n} \parallel P_{X^n}) \\ &= nD(P_{Z_1} \parallel P_1) \end{aligned} \tag{16}$$

and by concavity of Shannon entropy

$$H(P_{Y^n}) \leq \sum_{i=1}^n H(P_{Y_i}) = H(P_{\hat{Y}^n}) \leq H(P_{Z^n}) = nH(P_{Z_1}). \tag{17}$$

Normalize by n take limsup in (16) and liminf in (17) to get $D_u(\mathbf{Y} \parallel \mathbf{X}) \geq D(P_{Z_1} \parallel P_1)$ and $H(P_{Z_1}) > R$. From these we conclude (15). Following a similar procedure as in [6, Example 1], we can show the other direction. Also the remaining terms in (12) can be shown to satisfy

$$\begin{aligned}\sup_{\mathbf{Y}:\overline{H}(\mathbf{Y})\leq R} & E_l(\mathbf{Y},\mathbf{X},\rho) \\ & \geq \sup_{P_Y:H(P_Y)\leq R} \{\rho H(P_Y) - D(P_Y \parallel P_1)\} \qquad (18)\end{aligned}$$

$$\sup_{\mathbf{Y}} E_u(\mathbf{Y},\mathbf{X},\theta) = \sup_{P_Y}\{\theta H(P_Y) - D(P_Y \parallel P_1)\}. \qquad (19)$$

Substitution of (13) and (18) in the lower bound of (12) yields

$$\begin{aligned}E_l^s(R,\rho) \geq \max\Big\{&\rho R - \inf_{P_Y:H(P_Y)>R} D(P_Y \parallel P_1), \\ & \sup_{P_Y:H(P_Y)\leq R} \{\rho H(P_Y) - D(P_Y \parallel P_1)\Big\} \\ = \sup_{P_Y} & \{\rho \min\{H(P_Y), R\} - D(P_Y \parallel P_1)\} \qquad (20)\end{aligned}$$

Similarly substitution of (19) in the upper bound of (12) yields

$$\begin{aligned}E_u^s&(R,\rho) \\ &\leq \min_{0\leq\theta\leq\rho} \Big\{(\rho-\theta)R + \sup_{P_Y}\{\theta H(P_Y) - D(P_Y \parallel P_1)\}\Big\} \\ &= \sup_{P_Y}\Big\{\rho \min_{0\leq\theta\leq\rho}\{(\rho-\theta)R + \theta H(P_Y)\} \\ &\qquad\qquad -D(P_Y \parallel P_1)\Big\} \qquad (21) \\ &= \sup_{P_Y}\{\rho \min\{H(P_Y,R)\} - D(P_Y \parallel P_1)\}, \qquad (22)\end{aligned}$$

where the interchange of sup and min in (21) holds because the function within braces is linear in θ and concave in P_Y. From (20) and (22), we recover Merhav & Arikan's result (1) for an iid source [2, Eqn. (3)].

Example 3 (Markov source): In this we focus on an irreducible stationary Markov source, taking values on $\mathbb{X}$, with transition probability matrix π.

Let $\mathcal{M}_s(\mathbb{X}^2)$ denote the set of *stationary* PMFs defined by

$$\begin{aligned}\mathcal{M}_s\left(\mathbb{X}^2\right) = \Big\{&Q \in \mathcal{M}\left(\mathbb{X}^2\right) : \\ & \sum_{x_1\in\mathbb{X}} Q(x_1,x) = \sum_{x_2\in\mathbb{X}} Q(x,x_2), \forall x \in \mathbb{X}\Big\}.\end{aligned}$$

Denote the common marginal by q and let

$$\eta(\cdot \mid x_1) := \begin{cases} Q(x_1,\cdot)/q(x_1), & \text{if } q(x_1) \neq 0, \\ 1/|\mathbb{X}|, & \text{otherwise.}\end{cases}$$

We may then denote $Q = q \times \eta$, where q is the distribution of X_1 and η the conditional distribution of X_2 given X_1. Following steps similar to the iid case, we have

$$E(R,\rho) = \sup_{Q\in\mathcal{M}_s(\mathbb{X}^2)} \Big\{\rho \min\{H(\eta \mid q), R\} - D(\eta \parallel \pi \mid q)\Big\},$$

where

$$H(\eta \mid q) := \sum_{x\in\mathbb{X}} q(x) H(\eta(\cdot \mid x))$$

is the conditional one-step entropy, and

$$D(\eta \parallel \pi \mid q) = \sum_{x_1\in\mathbb{X}} q(x_1) D(\eta(\cdot \mid x_1) \parallel \pi(\cdot \mid x_1)).$$

For a unifilar source the underlying state space forms a Markov chain and the entropy and divergence of the source equals those of the underlying Markov state space source [11, Thm. 6.4.2]. The arguments for the Markov source are now directly applicable to a unifilar source.

Acknowledgments

This work was supported by the Defence Research and Development Organisation, 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).

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 guessingwiretapper," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1860–1866, Sep. 1999.

[3] R. Sundaresan, "Guessing based on length functions," in *Proceedings of the Conference on Managing Complexity in a Distributed World, MCDES*, Bangalore, India, May 2008; *also available as* DRDO-IISc Programme in Mathematical Engineering Technical Report No. TR-PME-2007-02, Feb. 2007.
http://pal.ece.iisc.ernet.in/PAM/tech_rep07/TR-PME-2007-02.pdf.

[4] Y. Hayashi and H. Yamamoto, "Coding theorems for the Shannon cipher system with a guessing wiretapper and correlated source outputs," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2808–2817, Jun. 2008.

[5] T. S. Han, "The reliability functions of the general source with fixed-length coding," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 2117–2132, Sep 2000.

[6] K. Iriyama, "Probability of error for the fixed-length source coding of general sources," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1537–1543, May 2001.

[7] L. L. Campbell, "A coding theorem and Rényi's entropy," *Information and Control*, vol. 8, pp. 423–429, 1965.

[8] M. K. Hanawal and R. Sundaresan, "The shannon cipher system with a guessing wiretapper: General sources," *Submitted to IEEE* Transactions on information theory, 2009, *also available as* DRDO-IISc Programme in Mathematical Engineering Technical Report No. TR-PME-2009-04, Jan. 2009.
http://pal.ece.iisc.ernet.in/PAM/tech_rep09/TR-PME-2009-04.pdf.

[9] T. S. Han, *Information-Spectrum Methods in InformationTheory*. Springer-Verlog, 2003.

[10] M. K. Hanawal and R. Sundaresan, "Guessing revisited: A large deviations approach," *Submitted to IEEE* Transactions on information theory, 2008, *also available as* DRDO-IISc Programme in Mathematical Engineering Technical Report No. TR-PME-2008-08, Dec. 2008.
http://pal.ece.iisc.ernet.in/PAM/tech_rep08/TR-PME-2008-08.pdf.

[11] R. Ash, *Information Theory*. Interscience Publishers, 1965.