Maximal Leakage Minimization for The Shannon Cipher System

Ibrahim Issa[1], Sudeep Kamath[2], and Aaron B. Wagner[1]
[1]School of Electrical and Computer Engineering, Cornell University, Ithaca, New York
[2] Department of Electrical Engineering, Princeton University, Princeton, New Jersey
Emails: ii47@cornell.edu, sukamath@princeton.edu, wagner@cornell.edu

***Abstract*—A variation of the Shannon cipher system, in which lossy communication is allowed and performance of an encryption scheme is measured in terms of *maximal leakage* (recently proposed by the authors [1]), is investigated. The asymptotic behavior of normalized maximal leakage is studied, and a single-letter characterization of the optimal limit is derived. Moreover, asymptotically-optimal encryption schemes are demonstrated.**

I. Introduction

Shannon [2] introduced the secrecy system shown in Figure 1, which has since become known as the Shannon cipher system. The setup consists of a transmitter and a legitimate receiver that are linked by a public noiseless channel and share a common key, and an eavesdropper who has access to the public channel and is aware of the source statistics and the used encryption schemes. The encryption schemes must allow the legitimate receiver to perfectly reconstruct the source sequence.

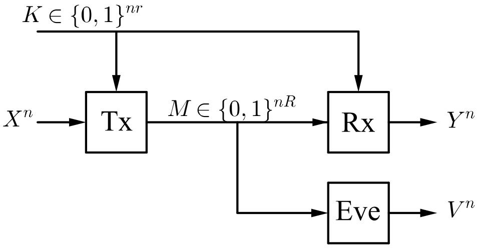

Fig. 1. The Shannon cipher system

Shannon showed that *perfect secrecy* (i.e., making the public message M and the source sequence X^n statistically independent) requires a key rate that is at least as large as the message rate, which is typically not possible in practice. It is necessary then to quantify *imperfect* or *partial* secrecy. To that end, Shannon used equivocation, $H(X^n|M)$, as a secrecy metric. However, its use is not well motivated [3], and several works have argued for the use of other metrics. More recently, we [1] have introduced *maximal leakage* as a proper metric to measure the leakage of information from a discrete random variable X, to another, Y. Unlike equivocation, maximal leakage, denoted by $\mathcal{L}(X \to Y)$, is given by an operational definition (cf. Definition 1): it is the multiplicative increase in the probability of guessing correctly a (possibly randomized) function of X after observing Y, maximized over all such functions. Furthermore, it admits a simple form (cf. Theorem 1), and it has properties consistent with an axiomatic view of a leakage metric [1]: it satisfies the data processing inequality as well as additivity over independent pairs $\{(X_i, Y_i)\}$, and it is equal to zero if and only if X and Y are independent. Maximal leakage is further strengthened by the robustness of its definition [1]: it is unchanged if the adversary can make *several* guesses, or if s/he wants to only *approximate* the true function value. Further discussion of its motivation and its properties can be found in our previous paper [1]. Moreover, the maximal leakage formula arises from a different definition in the computer security literature [4,5], and is given there a close but different interpretation.

Therefore, in this paper, we use maximal leakage to assess the best partial secrecy that can be obtained in the Shannon cipher system, for any encryption scheme. Moreover, we allow for lossy communication by introducing a distortion function d at the legitimate receiver. For a given distortion level D, we require that the probability of violating the distortion constraint decays as $2^{-n\alpha}$, for a given $\alpha > 0$. Then, for a given D and α, we study the asymptotic behavior of the normalized maximal leakage.

For a discrete memoryless source (DMS), we derive the optimal (i.e., minimal) limit of the normalized maximal leakage. The scheme we propose for the primary user (i.e., the transmitter–legitimate receiver pair) operates on a type-by-type basis. With each type, we associate a good rate-distortion code. The codebooks are then divided into bins, and the key is used to randomize, within a bin, the choice of codeword associated with a particular source sequence. However, types with low enough probability are discarded, i.e., a dummy message is associated with all the source sequences belonging to such types. We also derive the optimal limit when the requirement of a decaying probability of violating the distortion constraint is replaced with an expected distortion constraint. Other works have considered the Shannon cipher system allowing for lossy compression [3,6]–[8], although they differ from the present work in how they measure partial secrecy.

II. Preliminaries

We briefly review our results on maximal leakage [1].

Definition 1 (Maximal Leakage): Given a joint distribution P_{XY} on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, the *maximal leakage* from X to Y is defined as

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

where U and $\hat{U}$ are in the same finite, but arbitrary, alphabet.

The maximization is intended to represent a worst-case analysis, and models cases in which the conditional probability distribution $P_{U|X}$ of the variable of interest to the eavesdropper is not known. Maximal leakage is characterized by the following theorem.

Theorem 1 ([1]): For any joint distribution P_{XY} on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, the maximal leakage from X to Y is equal to the Sibson mutual information of order infinity, $I_\infty(X;Y)$. That is,

$$\mathcal{L}(X\to Y) = \log \sum_{y\in\mathcal{Y}} \max_{\substack{x\in\mathcal{X}:\\ P_X(x)>0}} P_{Y|X}(y|x).$$

The theorem reveals two of the more useful aspects of maximal leakage from an engineering perspective: minimizing $\mathcal{L}(X\to Y)$ over $P_{Y|X}$, for a fixed P_X, amounts to minimizing a convex function, and $\mathcal{L}(X\to Y)$ depends on P_X only through its support. The latter is notable in that P_X is often not known exactly in practice.

Finally, we set some notation for the remainder of the paper. In the following, $\mathcal{Z}$ is an arbitrary discrete set, and Z is a random variable over $\mathcal{Z}$.

- For a sequence $z^n \in \mathcal{Z}^n$, Q_{z^n} is the empirical PMF of z^n, also referred to as its type.
- $\mathcal{Q}^n_{\mathcal{Z}}$ is the set of types in $\mathcal{Z}^n$, i.e., the set of rational PMF's with denominator n.
- For $Q_Z \in \mathcal{Q}^n_{\mathcal{Z}}$, the type class of Q_Z is $T_{Q_Z} \triangleq \{z^n \in \mathcal{Z}^n : Q_{z^n} = Q_Z\}$.
- $\mathbf{E}_Q[\cdot]$, $H_Q(\cdot)$, and $I_Q(\cdot;\cdot)$ denote respectively expectation, entropy, and mutual information taken with respect to distribution Q.

III. Shannon Cipher System

Let $\mathcal{X}$ and $\mathcal{Y}$ be the alphabets associated with the transmitter and the legitimate receiver, respectively. The transmitter and the legitimate receiver are connected through a noiseless channel of rate R, and share common randomness $K_n \in \mathcal{K}_n = \{0,1\}^{nr}$, where K_n is uniformly distributed over $\mathcal{K}_n$, and $r > 0$ is the rate of the key. The transmitter observes an n-length message $X^n = (X_1, X_2, \cdots, X_n)$, independent of K_n, and wants to transmit a quantized version of it. Let f and h be, respectively, the transmitter's encoding and the receiver's decoding functions. The transmitter then sends a message $M_n = f(X^n, K_n)$, $M_n \in \mathcal{M}_n = \{0,1\}^{nR}$, and the receiver generates a reconstruction $Y^n = h(M_n, K_n)$. Note that we allow the functions f and h to be randomized (beyond the randomness in K_n). For a given distortion function $d : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}_+$, distortion level D, and distortion excess probability α, we require that $\mathbf{Pr}(d(X^n,Y^n) > D) \leq \exp\{-n\alpha\}$, where $d(X^n,Y^n) = \frac{1}{n}\sum_{i=1}^n d(X_i,Y_i)$.

An eavesdropper intercepts the message M. We assume s/he knows the source statistics as well as the encoding and decoding functions, but does not have access to the key K_n.

The primary user aims to minimize the maximal leakage to the eavesdropper $\mathcal{L}(X^n \to M_n)$. We characterize the asymptotically-optimal normalized maximal leakage under the following assumptions[1]:

(A1) The source is memoryless.
(A2) The alphabets $\mathcal{X}$ and $\mathcal{Y}$ are finite.
(A3) The distortion function d is bounded, i.e., there exists $D_{\max}$ such that, for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, $d(x,y) \leq D_{\max}$. Moreover, $D \geq D_{\min}$, where $D_{\min} = \max_{x\in\mathcal{X}} \min_{y\in\mathcal{Y}} d(x,y)$.
(A4) $R > \max_{Q:D(Q||P)\leq\alpha} R(Q,D)$, where $R(Q,D)$ is the rate distortion function for source distribution Q.

We denote the optimal limit by $L(P,D,\overrightarrow{R},\alpha)$, where P is the source distribution, and $\overrightarrow{R} = (R,r)$:

$$L(P,D,\overrightarrow{R},\alpha) = \lim_{n\to\infty} \min_{\{f_n\}} \frac{1}{n}\mathcal{L}\left(X^n \to f(X^n,K_n)\right),$$

where $\{f_n\}$ is restricted to the class of functions ensuring the feasibility of the primary user's problem.

Without loss of generality, we assume that $P_X(x) > 0$ for all $x \in \mathcal{X}$. Our main result is then the characterization of the optimal limit as follows:

Theorem 2: Under assumptions (A1)-(A4), for any DMS P, and distortion function d with associated distortion level $D \geq D_{\min}$ and distortion excess probability $\alpha > 0$:

$$L(P,D,\overrightarrow{R},\alpha) = \max_{Q:D(Q||P)\leq\alpha} [R(Q,D) - r]^+, \quad (1)$$

where $[a]^+ = \max\{0,a\}$.

Note that the case $\alpha = \infty$ (i.e., when the distortion constraint is imposed almost surely) is included in the theorem. Moreover, in that case, the theorem holds even if the source is not memoryless, as long as the support of X^n is $\mathcal{X}^n$. This follows from the fact

[1] Note that it is necessary to have $R \geq \max_{Q:D(Q||P)\leq\alpha} R(Q,D)$ for the primary user's problem to be feasible.

that $\mathcal{L}(X^n \to M_n)$ and the constraint, when imposed almost surely, depend on the distribution of X^n only through its support. Therefore, solving for any specific distribution on that support is equivalent to solving for all distributions on the same support.

A. Achievability Proof

We will slightly abuse notation and shorten $L(P, D, \overrightarrow{R}, \alpha)$ to L in the following. We now show that the right-hand side of (1) upper-bounds L.

The scheme we consider is similar to the one proposed in [3, Section IV.B]. To that end, consider any $\epsilon > 0$ and let n be large enough such that we can construct a rate-distortion code $\mathcal{C}^n_{Q_X}$, for each type $Q_X \in \mathcal{Q}^n_{\mathcal{X}}$, satisfying the following: each sequence $x^n \in T_{Q_X}$ is covered and $|\mathcal{C}^n_{Q_X}| \leq 2^{n(R(Q_X,D)+\epsilon)}$. Such construction is guaranteed by the type covering lemma (Lemma 2.4.1 in [9]). We divide the codebook $\mathcal{C}^n_{Q_X}$ into $\lceil |\mathcal{C}^n_{Q_X}| / 2^{nr} \rceil$ bins, each of size 2^{nr}, except for possibly the last one. We denote by $\mathcal{C}^n_{Q_X}(i, .)$ the ith partition of the codebook, and by $\mathcal{C}^n_{Q_X}(i, j)$ the jth codeword in the ith partition. For each $x^n \in T_{Q_X}$, let i_{x^n} and j_{x^n} denote, respectively, the index of the partition containing the codeword associated with x^n and the index of the codeword within the partition (Note that if more than one codeword can be associated with x^n, we fix any one of them arbitrarily). Finally, let $m(Q_X, i, j)$ be a message consisting of the following:

- $\lceil \log |\mathcal{Q}^n_{\mathcal{X}}| \rceil$ bits to describe the type Q_X.
- $\lceil \log \lceil |\mathcal{C}^n_{Q_X}| / 2^{nr} \rceil \rceil$ bits to describe the index i, where $1 \leq i \leq \lceil |\mathcal{C}^n_{Q_X}| / 2^{nr} \rceil$.
- $\lceil \log |\mathcal{C}^n_{Q_X}(i, .)| \rceil$ bits to describe the index j, where $0 \leq j \leq \exp \lceil \log |\mathcal{C}^n_{Q_X}(i, .)| \rceil - 1$.

Now, for any $\delta \in \mathbb{R}$, let $\mathcal{Q}(\alpha, \delta) = \{Q_X : D(Q_X || P) \leq \alpha + \delta\}$, $\mathcal{Q}_n(\alpha, \delta) = \{Q_X \in \mathcal{Q}^n_{\mathcal{X}} : D(Q_X || P) \leq \alpha + \delta\}$, and consider the following lemma.

Lemma 1:

$$\lim_{\delta \to 0} \max_{Q_X \in \mathcal{Q}(\alpha,\delta)} R(Q_X, D) = \max_{Q_X \in \mathcal{Q}(\alpha,0)} R(Q_X, D).$$

Proof: This follows directly from the continuity of $R(Q_X, D)$ in Q_X under (A3) [9, Lemma 2.2.2]. ■

Now let $\delta > 0$ be such that $\max_{Q_X \in \mathcal{Q}(\alpha,\delta)} R(Q_X, D) < R$ (Such δ exists by Lemma 1 and (A4)). Finally, for each sequence x^n, let $s(x^n) = \lceil \log |\mathcal{C}^n_{Q_X}(i_{x^n}, .)| \rceil$, and let $K_{s(x^n)}$ be the first $s(x^n)$ bits of K_n. The transmitter encodes as follows. Given x^n, if $Q_{x^n} \in \mathcal{Q}_n(\alpha, \delta)$, then

$$f(x^n, K_n) = m\left(Q_{x^n}, i_{x^n}, j_{x^n} \oplus K_{s(x^n)}\right), \tag{2}$$

where the XOR-operation is performed bitwise. Note that, in this case, the legitimate receiver can retrieve the type of the transmitted sequence and the index of the bin from the first two parts of the message, and the index of the sequence within the bin using the last part of the message and the key K_n, so that $h(M_n, K_n) = \mathcal{C}^n_{Q_{x^n}}(i_{x^n}, j_{x^n})$. Now, consider an $m_0 \in \mathcal{M}_n$ that has not been used by the previous encoding (Assumption (A4) and the choice of δ ensures the existence of such m_0). Then, for all x^n such that $Q_{x^n} \notin \mathcal{Q}_n(\alpha, \delta)$,

$$f(x^n, K_n) = m_0. \tag{3}$$

Remark 1: To verify that the suggested scheme satisfies the excess distortion probability constraint, consider the following:

$$\begin{aligned}\mathbf{Pr}(d(X^n, Y^n) > D) &\leq \sum_{Q_X \notin \mathcal{Q}_n(\alpha,\delta)} P(Q)\\ &\leq \sum_{Q_X \notin \mathcal{Q}_n(\alpha,\delta)} 2^{-nD(Q_X||P)}\\ &\leq (n+1)^{|\mathcal{X}|} 2^{-n(\alpha+\delta)}\\ &< 2^{-n\alpha},\end{aligned}$$

where the last inequality holds for large enough n.

Effectively, we are leaking the first two parts of the message Q_{X^n} and i_{X^n}, and *hiding* completely the last part j_{X^n}. Since there are only polynomially many types, the first part does not affect the normalized leakage. The second part, however, consists roughly of $R(Q, D) - r$ bits, whenever $R(Q, D) > r$; otherwise, i.e., when $R(Q, D) \leq r$, there is only one bin and there is no information to be leaked.

For a more rigorous analysis, let P_f be the induced joint probability distribution of (X^n, M_n). Then, for x^n satisfying $Q_{x^n} \in \mathcal{Q}_n(\alpha, \delta)$, we get from (2):

$$P_f\left(m(Q_{x^n}, i_{x^n}, j) \middle| x^n\right) = 2^{-s(x^n)}, \;\; 0 \leq j \leq 2^{s(x^n)} - 1.$$

Let $S(x^n) = 2^{s(x^n)}$. Note that we can equivalently denote $S(x^n)$ by $S(Q_{x^n}, i_{x^n})$, since the dependence on the sequence is only through the type and the index of the bin. Therefore, we get

$$\begin{aligned}&\exp\{\mathcal{L}(X^n \to M_n)\}\\ &= \sum_{m \in \mathcal{M}_n} \max_{x^n \in \mathcal{X}^n} P_f(m|x^n)\\ &= \max_{x^n \in \mathcal{X}^n} P_f(m_0|x^n) +\\ &\quad \sum_{\substack{Q_X \in\\ \mathcal{Q}_n(\alpha,\delta)}} \sum_{i=1}^{\lceil |\mathcal{C}^n_{Q_X}|/2^{nr} \rceil} \sum_{j=0}^{S(Q_X,i)-1} \max_{x^n \in \mathcal{X}^n} P_f(m(Q_X, i, j)|x^n)\\ &= 1 + \sum_{\substack{Q_X \in\\ \mathcal{Q}_n(\alpha,\delta)}} \sum_{i=1}^{\lceil |\mathcal{C}^n_{Q_X}|/2^{nr} \rceil} \sum_{j=0}^{S(Q_X,i)-1} S(Q_X, i)^{-1}\\ &\leq 1 + \sum_{Q_X \in \mathcal{Q}_n(\alpha,\delta)} (2^{n(R(Q_X,D)+\epsilon-r)} + 1)\\ &\leq 1 + 2 \sum_{Q_X \in \mathcal{Q}_n(\alpha,\delta)} 2^{n \max\{R(Q_X,D)+\epsilon-r, 0\}}\end{aligned}$$

$$\leq 4(n+1)^{|\mathcal{X}|} \exp\{n \max_{Q_X \in \mathcal{Q}_n(\alpha,\delta)} [R(Q_X, D) + \epsilon - r]^+\}. \tag{4}$$

Taking the limit as n tends to infinity, and noting that ϵ and δ were arbitrary, we get that

$$L \leq \max_{Q: D(Q||P) \leq \alpha} [R(Q, D) - r]^+,$$

where the inequality follows from Lemma 1 and the following lemma (the proof of which is omitted).

Lemma 2:

$$\lim_{n\to\infty} \max_{Q \in \mathcal{Q}^n_{\mathcal{X}}: D(Q||P) \leq \alpha} R(Q, D) = \max_{Q: D(Q||P) \leq \alpha} R(Q, D).$$

B. Converse Proof

We now show that L is lower-bounded by the right-hand side of (1). To that end, consider any valid encoding function f. To lower-bound $\mathcal{L}(X^n \to M_n)$, we consider a specific $P_{U|X^n}$ which is constructed as follows. Let $p^\star = \operatorname{argmin}_{x^n} P(x^n)$, and for each $x^n \in \mathcal{X}^n$, let $k(x^n) = P(x^n)/p^\star$, and let $\mathcal{U} = \bigcup_{x^n \in \mathcal{X}^n} \{(x^n, 1), (x^n, 2), \ldots, (x^n, \lceil k(x^n) \rceil)\}$. For each $u = (i_u, j_u) \in \mathcal{U}$ and $x^n \in \mathcal{X}$, let $P_{U|X^n}(u|x^n)$ be:

$$P_{U|X^n}((i_u, j_u)|x^n)$$
$$= \begin{cases} \frac{p^\star}{P(x^n)}, & i_u = x^n,\ 1 \leq j_u \leq \lfloor k(x^n) \rfloor, \\ 1 - \frac{(\lceil k(x^n) \rceil - 1)p^\star}{P(x^n)}, & i_u = x^n,\ j_u = \lceil k(x^n) \rceil, \\ 0, & i_u \neq x^n,\ 1 \leq j_u \leq \lceil k(i_u) \rceil. \end{cases}$$

Remark 2: Note that if $\lfloor k(x^n) \rfloor = \lceil k(x^n) \rceil$, the corresponding expressions coincide. Moreover, the given $P_{U|X^n}$ achieves the supremum in the definition of $\mathcal{L}(X^n \to M_n)$ [1], although this is not needed here.

Therefore, $\max_{u \in \mathcal{U}} P_U(u) = p^\star$. We will also consider a sub-optimal guessing function for U. The scheme is as follows: the eavesdropper first tries to guess the key K_n by choosing an element uniformly at random from $\{0,1\}^{nr}$. We denote this guess by $\tilde{K}_n$. Then, proceeding by assuming that the key guess was correct, s/he tries to guess the sequence x^n using a guessing function given by Lemma 3 below. We denote this stage by g_1. Finally, again proceeding by assuming that the source sequence guess was correct, the eavesdropper attempts to guess U by using the MAP rule. We denote this stage by g_2, and we get for each $x^n \in \mathcal{X}^n$,

$$g_2(x^n) = (x^n, 1),$$
$$\text{and } \mathbf{Pr}(g_2(x^n) = U^n | x^n) = p^\star / P(x^n). \tag{5}$$

Lemma 3: There exists a function $g_1 : \mathcal{Y}^n \to \mathcal{X}^n$ such that, for all (x^n, y^n) satisfying $d(x^n, y^n) \leq D$, $\mathbf{Pr}\left(x^n = g(y^n)\right) \geq c_n 2^{-n(H_{Q_{x^n}}(X) - R(Q_{x^n}, D))}$, where $c_n = (n+1)^{-|\mathcal{X}||\mathcal{Y}|(|\mathcal{X}|+1)}$.

Proof: This is an application of Lemma 5 in [3]. In particular, we set in Lemma 5 $\mathcal{V}$ to be $\mathcal{X}$, d_e to be the Hamming distortion function, and D_e to be zero. Then, $I_{P^\star_n(Q_{x^n y^n})}(X; V|Y)$ (as defined in [3]) satisfies:

$$\begin{aligned} & I_{P^\star_n(Q_{x^n y^n})}(X; V|Y) \\ & = H_{Q_{x^n y^n}}(X|Y) \\ & = H_{Q_{x^n}}(X) - H_{Q_{x^n}}(X) + H_{Q_{x^n y^n}}(X|Y) \\ & \leq H_{Q_{x^n}}(X) - R(Q_{x^n}, D). \qquad \blacksquare \end{aligned}$$

To analyze the above scheme, fix $\epsilon > 0$, and let P_f denote the induced joint probability on (X^n, K_n, M_n). Furthermore, without loss of generality, we can assume that the decoding function h is a deterministic function of M_n and K_n. Finally, define

$$\mathcal{M}_D(x^n, k) = \{m \in \mathcal{M}_n : d(x^n, h(m,k)) \leq D\},$$
$$x^n \in \mathcal{X}^n, k \in \mathcal{K}_n, \tag{6}$$
$$\text{and } \mathcal{A} = \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : d(x^n, y^n) > D\}. \tag{7}$$

Letting g be the concatenation of the two stages, we get

$$\begin{aligned} & \mathbf{Pr}(U = g(M)) \\ & = \sum_{x^n \in \mathcal{X}^n} \sum_{u \in \mathcal{U}} \sum_{k \in \mathcal{K}_n} \sum_{m \in \mathcal{M}_n} P(x^n) P_{U|X^n}(u|x^n) P_{K_n}(k). \\ & \qquad P_f(m|x^n, k) P(u = g(m)|x^n, m, k) \\ & \geq \sum_{x^n \in \mathcal{X}^n} \sum_{u \in \mathcal{U}} \sum_{k \in \mathcal{K}_n} \sum_{m \in \mathcal{M}_D(x^n,k)} P(x^n) P_{U|X^n}(u|x^n). \\ & \qquad P_{K_n}(k) P_f(m|x^n, k) P(u = g(m)|x^n, m, k) \\ & \geq \sum_{x^n \in \mathcal{X}^n} \sum_{u \in \mathcal{U}} \sum_{k \in \mathcal{K}_n} \sum_{m \in \mathcal{M}_D(x^n,k)} P(x^n) P_{U|X^n}(u|x^n). \\ & \qquad P_{K_n}(k) P_f(m|x^n, k) P(\tilde{K}_n = k). \\ & \qquad P(g_1(h(m,k)) = x^n) P(g_2(x^n) = u|x^n) \\ & \overset{(a)}{\geq} c_n \sum_{x^n \in \mathcal{X}^n} \sum_{k \in \mathcal{K}_n} \sum_{m \in \mathcal{M}_D(x^n,k)} P(x^n) P_{K_n}(k) P_f(m|x^n, k). \\ & \qquad 2^{-nr} 2^{-n(H_{Q_{x^n}}(X) - R(Q_{x^n}, D))} p^\star / P(x^n) \\ & = c_n p^\star 2^{-nr} \sum_{Q_X \in \mathcal{Q}^n_{\mathcal{X}}} \sum_{x^n \in T_{Q_X}} \sum_{k \in \mathcal{K}_n} \sum_{m \in \mathcal{M}_D(x^n,k)} P(x^n). \\ & \qquad P_{K_n}(k) P_f(m|x^n, k) 2^{-n(H_{Q_X}(X) - R(Q_X, D))} / P(x^n) \\ & = c_n p^\star 2^{-nr} \sum_{Q_X \in \mathcal{Q}^n_{\mathcal{X}}} \sum_{x^n \in T_{Q_X}} \sum_{k \in \mathcal{K}_n} \sum_{m \in \mathcal{M}_D(x^n,k)} P(x^n). \\ & \qquad P_{K_n}(k) P_f(m|x^n, k) 2^{n(R(Q_X, D) + D(Q_X||P))} \\ & = c_n p^\star 2^{-nr} \sum_{Q_X \in \mathcal{Q}^n_{\mathcal{X}}} 2^{n(R(Q_X, D) + D(Q_X||P))}. \\ & \qquad P_f(\mathcal{A}^c \cap T_{Q_X}), \end{aligned} \tag{8}$$

where (a) follows from Lemma 3, (5), and (6). Now, note that for any Q,

$$\begin{aligned} & P_f(\mathcal{A}^c | T_Q) \\ & = 1 - P_f(\mathcal{A} | T_Q) \\ & \geq 1 - \min\{1, P_f(\mathcal{A}) / P(T_Q)\} \end{aligned}$$

$$\geq 1-\min\{1, 2^{-n(\alpha-D(Q||P)-\frac{|\mathcal{X}|}{n}\log(n+1))}\}$$
$$=\max\{0, 1-2^{-n(\alpha-D(Q||P)-\frac{|\mathcal{X}|}{n}\log(n+1))}\}.$$

Then, continuing (8), we get

$$\begin{aligned}
&\mathbf{Pr}(U=g(M)) \\
&\geq c_n p^\star 2^{-nr} \sum_{Q_X\in\mathcal{Q}_{\mathcal{X}}^n} 2^{n(R(Q_X,D)+D(Q_X||P))} P(T_{Q_X}) \cdot \\
&\qquad \max\{0, 1-2^{-n(\alpha-D(Q_X||P)-\frac{|\mathcal{X}|}{n}\log(n+1))}\} \\
&\overset{(a)}{\geq} c_n' p^\star 2^{-nr} \sum_{Q_X\in\mathcal{Q}_n(\alpha,-\epsilon)} 2^{nR(Q_X,D)} \cdot \\
&\qquad (1-2^{-n(\alpha-D(Q_X||P)-\frac{|\mathcal{X}|}{n}\log(n+1))}) \\
&\overset{(b)}{\geq} c_n' p^\star 2^{-nr} \sum_{Q_X\in\mathcal{Q}_n(\alpha,-\epsilon)} 2^{nR(Q_X,D)} (1/2) \\
&\geq (c_n' p^\star/2) \max_{Q_X\in\mathcal{Q}_n(\alpha,-\epsilon)} \exp\{n(R(Q_X,D)-r)\},
\end{aligned} \tag{9}$$

where (a) and (b) hold for large enough n, and $c_n' = (n+1)^{-|\mathcal{X}|}c_n$. Finally taking the ratio of $\mathbf{Pr}(U = g(M))$ and $\max_u P_U(u)$, and taking the limit as n tends to infinity, and noting that ϵ is arbitrary, we get

$$L \geq \max_{Q:D(Q||P)\leq\alpha} R(Q,D) - r,$$

where the inequality follows from Lemmas 1 and 2. Since L is positive by definition,

$$\begin{aligned}
L &\geq [\max_{Q:D(Q||P)\leq\alpha} R(Q,D) - r]^+ \\
&= \max_{Q:D(Q||P)\leq\alpha} [R(Q,D) - r]^+.
\end{aligned}$$

C. Varying The Distortion Constraint

We briefly discuss a variation of the above problem. Instead of requiring a decaying probability of violating the distortion constraint, we could require that the distortion constraint holds only in expectation—as is common in many works in the literature. In that case, we modify assumption (A4) to be:

(A4') $R > R(P, D)$.

Then, the following theorem holds.

Theorem 3: Under assumptions (A1)-(A3) and (A4'), for any DMS P, and distortion function d with associated distortion level $D \geq D_{\min}$:

$$L(P, D, \overrightarrow{R}) = [R(P, D) - r]^+. \tag{10}$$

The achievability argument follows by a similar manner as the one given in III.A. However, instead of encoding on a type-by-type basis, we simply use a good rate-distortion code that satisfies the expected distortion requirement. As above, we divide the codebook into bins of size 2^{nr}, except for possibly the last one. Then, an analysis similar to (4) yields $L(P, D, \overrightarrow{R}) \leq [R(P, D) - r]^+$. As for the lower bound, we use the fact that $I_\infty(X; Y) \geq I(X; Y)$ [10]. This problem, with mutual information replacing maximal leakage, has already been solved by Schieler and Cuff [11]. More specifically, Corollary 5 of [11] yields that the optimal normalized mutual information is indeed given by $[R(P, D) - r]^+$.

Acknowledgment

This research was supported by the US National Science Foundation under grant CCF-1065352.

References

[1] I. Issa, S. Kamath, and A. B. Wagner, "An operational measure of information leakage," in *Proc. of 50th Ann. Conf. on Information Sciences and Systems (CISS) (to appear)*, Mar. 2016.

[2] C. E. Shannon, "Communication theory of secrecy systems," *Bell system technical journal*, vol. 28, no. 4, pp. 656–715, 1949.

[3] I. Issa and A. B. Wagner, "Measuring secrecy by the probability of a successful guess," Aug. 2015. [Online]. Available: http://arxiv.org/abs/1507.02342

[4] C. Braun, K. Chatzikokolakis, and C. Palamidessi, "Quantitative notions of leakage for one-try attacks," *Electronic Notes in Theoretical Computer Science*, vol. 249, pp. 75–91, 2009.

[5] M. S. Alvim, K. Chatzikokolakis, A. Mciver, C. Morgan, C. Palamidessi, and G. Smith, "Additive and multiplicative notions of leakage, and their capacities," in *IEEE 27th Computer Security Foundations Symposium*, July 2014, pp. 308–322.

[6] H. Yamamoto, "Rate-distortion theory for the Shannon cipher system," *Information Theory, IEEE Transactions on*, vol. 43, no. 3, pp. 827–835, 1997.

[7] C. Schieler and P. Cuff, "The henchman problem: measuring secrecy by the minimum distortion in a list," Nov. 2014. [Online]. Available: http://arxiv.org/abs/1410.2881

[8] N. Weinberger and N. Merhav, "A large deviations approach to secure lossy compression," May 2015. [Online]. Available: http://arxiv.org/abs/1504.05756

[9] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Budapest: Akadémiai Kiadó, 1997.

[10] S. Verdú, "α-mutual information," in *Information Theory and Applications Workshop (ITA), 2015*, Feb 2015, pp. 1–6.

[11] C. Schieler and P. Cuff, "Rate-distortion theory for secrecy systems," *Information Theory, IEEE Transactions on*, vol. 60, no. 12, pp. 7584–7605, Dec 2014.