An Information-Theoretical Analysis of the Minimum Cost to Erase Information*

Tetsunao Matsuta† Tomohiko Uyematsu‡

SUMMARY We normally hold a lot of confidential information in hard disk drives and solid-state drives. When we want to erase such information to prevent the leakage, we have to overwrite the sequence of information with a sequence of symbols independent of the information. The overwriting is needed only at places where overwritten symbols are different from original symbols. Then, the cost of overwrites such as the number of overwritten symbols to erase information is important. In this paper, we clarify the minimum cost such as the minimum number of overwrites to erase information under weak and strong independence criteria. The former (resp. the latter) criterion represents that the mutual information between the original sequence and the overwritten sequence normalized (resp. not normalized) by the length of the sequences is less than a given desired value.

Key words: data erasure, distortion-rate function, information erasure, information spectrum, random number generation

1 Introduction

Since services and activities using various types of information have increased, we normally hold a lot of confidential information. For example, storage devices such as hard disk drives (HDDs), solid-state drives (SSDs) and USB flash drives of individuals and companies hold personal addresses, names, phone numbers, e-mail addresses, credit card numbers, etc. When we want to discard, refurbish or just increase the security of these devices, we will usually erase information to prevent the leakage.

In order to erase information, we have to overwrite the sequence of information with a sequence of symbols independent of the information. Commonly used methods of erasure are to overwrite information with uniform random numbers or repeated specific patterns such as all zeros and all ones. There are several standards [3, 4, 5, 6, 7] to erase information. Although most of these standards propose to repeat overwriting many times, overwriting data once is adequate to erase information for modern storage devices (see, e.g., [7, Section 2.3]).

The overwriting is needed only at places where overwritten symbols are different from original symbols, e.g., 0 to 1 or 1 to 0 for binary sequences. If there are so many overwritten symbols, the overwriting damages devices, shortens the storage life and may also take write time. This is crucial for devices with a limited number of writes such as SSDs and USB flash drives. Thus, we want to reduce the number of overwritten symbols when we erase information. Here comes a natural question: "What is the minimum number of overwritten symbols?".

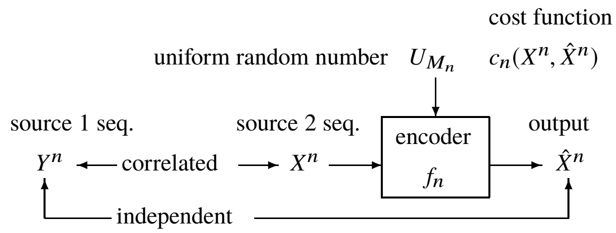

Figure 1: Information Erasure Model

In this paper, we clarify the minimum cost such as the minimum number or time of overwrites to erase information. As we stated in the above, for a binary sequence, the overwriting occurs at places where overwritten symbols are different from original symbols. In this case, a proper measure of the cost is the Hamming distance between the original sequence and the overwritten sequence. From this point of view, the information erasure can be modeled by correlated sources as Fig. 1 which actually is a somewhat general model. In this model, sequences emitted from source 1 and source 2 represent confidential information and information to be erased, respectively. For example, source 1 and source 2 are regarded as a fingerprint and its quantized image, respectively. When two correlated sources are identical, the model corresponds to the above mentioned situation. As shown in this figure, the encoder can observe one of the sequences. The encoder outputs a sequence that represents the overwritten sequence. Here, we allow the encoder to observe a uniform random number of limited size to generate an independent sequence. Then, the cost can be measured by a function of the input source sequence and the output sequence of the encoder.

For this information erasure model, we consider a *weak* and a *strong* independence criteria. The former (resp. the latter) criterion represents that the mutual information between the source sequence and the output sequence of the encoder normalized (resp. not normalized) by the length (blocklength) of sequences is less than a given desired value. For the weak independence criterion, we consider the *average* cost and the *worst-case* cost. The former cost represents the expectation of the cost with respect to the sequences. The later cost represents the limit superior in probability [8] of the cost. Then, by using information-spectrum quantities [8], we characterize the minimum average and the minimum worst-case costs for general sources, where the block length is unlimited. For the strong independence criterion, by

*Portions of this paper were presented at the 39th Symposium on Information Theory and Its Applications [1], and at the 2017 IEEE Information Theory Workshop [2].

†tetsu@ict.e.titech.ac.jp
‡uematsu@ict.e.titech.ac.jp
The authors are with Dept. of Information and Communications Engineering, Tokyo Institute of Technology, Tokyo, 152-8552 Japan.

employing a stochastic encoder, we give a single-letter characterization of the minimum average cost for stationary memoryless sources, where the blocklength is unlimited. On the other hand, for the strong (same as the weak in this case) independence criterion, we also consider the non-asymptotic minimum average cost for a given finite blocklength. Then, we give a single-letter characterization of it for stationary memoryless sources. We show that the minimum average and the minimum worst-case costs can be characterized by the *distortion-rate function* for the lossy source coding problem (see. e.g., [8]) when the two correlated sources are identical. This means that our problem setting gives a new point of view of the lossy source coding problem. We also show that for stationary memoryless sources, there exists a sufficient condition such that the optimal method of erasure from the point of view of the cost is to overwrite the source sequence with repeated identical symbols.

There are some studies [9, 10] to investigate a relationship between coding and information leakage, but their problem settings are fundamentally different from this study which considers the minimum cost of information erasure.

The rest of this paper is organized as follows. In Section 2, we give some notations and formal definitions of the minimum average and the minimum worst-case costs under the weak independence criterion. Then, we characterize these costs for general sources. In Section 3, we give the formal definition of the minimum average cost under the strong independence criterion. We also give the formal definition of the non-asymptotic minimum average cost. Then, we give a single-letter characterization of these costs and some results obtained from this characterization. In Section 4, we show proofs for characterizations of minimum costs under the weak independence criterion. In Section 5, we conclude the paper.

2 Minimum Costs to Erase Information under the Weak Independence Criterion

In this section, we consider the minimum average and the minimum worst-case costs under the weak independence criterion, and characterize these costs for general sources. We show some special cases of these costs in this section.

2.1 Problem Formulation

In this section, we provide the formal setting of the information erasure and define the minimum average and the minimum worst-case costs under the weak independence criterion.

Unless otherwise stated, we use the following notations throughout this paper (not just this section). The probability distribution of a random variable (RV) X is denoted by the subscript notation P_X, and the conditional probability distribution for X given an RV Y is denoted by $P_{X|Y}$. The n-fold Cartesian product of a set $\mathcal{X}$ is denoted by $\mathcal{X}^n$ while an n-length sequence of symbols $(a_1, a_2, \cdots, a_n)$ is denoted by a^n. The sequence of RVs $\{X^n\}_{n=1}^{\infty}$ is denoted by the bold-face letter $\mathbf{X}$. Hereafter, log means the natural logarithm.

Let $\mathcal{X}$, $\mathcal{Y}$ and $\hat{\mathcal{X}}$ be finite sets, and $\mathcal{U}_{M_n} = \{1, 2, \cdots, M_n\}$. Let U_{M_n} be an RV uniformly distributed on $\mathcal{U}_{M_n}$, and (X^n, Y^n) be a pair of RVs on $\mathcal{X}^n \times \mathcal{Y}^n$ such that (X^n, Y^n) is independent of U_{M_n}. The pair $(\mathbf{X}, \mathbf{Y}) = \{(X^n, Y^n)\}_{n=1}^{\infty}$ of a sequence of RVs represents a pair of general sources [8] that is not required to satisfy the consistency condition.

Before we show several definitions, we introduce the limit superior and the limit inferior in probability [8].

Definition 1 (Limit superior/inferior in probability)**.** For an arbitrary sequence $\mathbf{Z} = \{Z^n\}_{n=1}^{\infty}$ of real-valued RVs, we respectively define the limit superior and the limit inferior in probability by

$$\operatorname*{p\text{-}lim\,sup}_{n\to\infty} Z_n \triangleq \inf\left\{\alpha : \lim_{n\to\infty} \Pr\{Z_n > \alpha\} = 0\right\},$$
$$\operatorname*{p\text{-}lim\,inf}_{n\to\infty} Z_n \triangleq \sup\left\{\beta : \lim_{n\to\infty} \Pr\{Z_n < \beta\} = 0\right\}.$$

For the information erasure model (Fig. 1), let $f_n : \mathcal{X}^n \times \mathcal{U}_{M_n} \to \hat{\mathcal{X}}^n$ be an encoder, and $c_n : \mathcal{X}^n \times \hat{\mathcal{X}}^n \to [0, \infty)$ be a cost function satisfying

$$\sup_{n\geq 1} \sup_{(x^n, \hat{x}^n)\in\mathcal{X}^n\times\hat{\mathcal{X}}^n} c_n(x^n, \hat{x}^n) \triangleq c_{\max} < \infty.$$

We define the worst-case cost by the limit superior in probability of the cost, i.e.,

$$\operatorname*{p\text{-}lim\,sup}_{n\to\infty} c_n(X^n, f_n(X^n, U_{M_n})).$$

Then, we introduce two types of achievability.

Definition 2. For real numbers $R, \Gamma, \epsilon \geq 0$, we say (R, Γ) is ϵ-weakly achievable in the sense of the average cost if and only if there exist a sequence of integers $\{M_n\}_{n=1}^{\infty}$ and a sequence of encoders $\{f_n\}_{n=1}^{\infty}$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \leq R, \tag{1}$$
$$\limsup_{n\to\infty} \frac{1}{n} I(Y^n, f_n(X^n, U_{M_n})) \leq \epsilon, \tag{2}$$
$$\limsup_{n\to\infty} \mathrm{E}[c_n(X^n, f_n(X^n, U_{M_n}))] \leq \Gamma,$$

where $I(X;Y)$ denotes the mutual information between RVs X and Y, and $\mathrm{E}[\cdot]$ denotes the expectation.

Definition 3. For real numbers $R, \Gamma, \epsilon \geq 0$, we say (R, Γ) is ϵ-weakly achievable in the sense of the worst-case cost if and only if there exist a sequence of integers $\{M_n\}_{n=1}^{\infty}$ and a sequence of encoders $\{f_n\}_{n=1}^{\infty}$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \leq R,$$
$$\limsup_{n\to\infty} \frac{1}{n} I(Y^n; f_n(X^n, U_{M_n})) \leq \epsilon, \tag{3}$$
$$\operatorname*{p\text{-}lim\,sup}_{n\to\infty} c_n(X^n, f_n(X^n, U_{M_n})) \leq \Gamma.$$

We adopt the mutual information normalized by the blocklength n in these definitions (i.e., (2) and (3)). This is a somewhat weak criterion of independence compared with the mutual information itself (not normalized by the blocklength). The stronger version of this criterion will be considered in the later section.

Now, we define the minimum average and the minimum worst-case costs under the weak independence criterion.

Definition 4. We define the minimum average cost as

$$C_{\mathrm{a}}(\epsilon, R) \triangleq \inf\{\Gamma : (R, \Gamma) \text{ is } \epsilon\text{-weakly achievable in the sense of the average cost}\}.$$

Definition 5. We define the minimum worst-case cost as

$$C_{\mathrm{w}}(\epsilon, R) \triangleq \inf\{\Gamma : (R, \Gamma) \text{ is } \epsilon\text{-weakly achievable in the sense of the worst-case cost}\}.$$

2.2 Minimum Average and Minimum Worst-Case Costs

In this section, we characterize the minimum average and the minimum worst-case costs. To this end, for given sequences $(\mathbf{Y}, \mathbf{X}, \hat{\mathbf{X}})$ of RVs, we define

$$I(\mathbf{Y}, \hat{\mathbf{X}}) \triangleq \limsup_{n\to\infty} \frac{1}{n} I(Y^n, \hat{X}^n),$$

$$\overline{H}(\hat{\mathbf{X}}|\mathbf{X}) \triangleq \text{p-}\limsup_{n\to\infty} \frac{1}{n} \log \frac{1}{P_{\hat{X}^n|X^n}(\hat{X}^n|X^n)},$$

$$c(\mathbf{X}, \hat{\mathbf{X}}) \triangleq \limsup_{n\to\infty} \mathrm{E}[c_n(X^n, \hat{X}^n)],$$

$$\overline{c}(\mathbf{X}, \hat{\mathbf{X}}) \triangleq \text{p-}\limsup_{n\to\infty} c_n(X^n, \hat{X}^n),$$

and denote by $\mathbf{Y} - \mathbf{X} - \hat{\mathbf{X}}$ that the Markov chain $Y^n - X^n - \hat{X}^n$ holds for all $n \geq 1$.

For the minimum costs under the weak independence criterion, we have the following two theorems.

Theorem 1. For a pair of general sources $(\mathbf{X}, \mathbf{Y})$ and any real numbers $\epsilon, R \geq 0$, we have

$$C_{\mathrm{a}}(\epsilon, R) = \inf_{\substack{\hat{\mathbf{X}}: \mathbf{Y}-\mathbf{X}-\hat{\mathbf{X}}, \\ I(\mathbf{Y}, \hat{\mathbf{X}}) \leq \epsilon, \overline{H}(\hat{\mathbf{X}}|\mathbf{X}) \leq R}} c(\mathbf{X}, \hat{\mathbf{X}}).$$

Theorem 2. For a pair of general sources $(\mathbf{X}, \mathbf{Y})$ and any real numbers $\epsilon, R \geq 0$, we have

$$C_{\mathrm{w}}(\epsilon, R) = \inf_{\substack{\hat{\mathbf{X}}: \mathbf{Y}-\mathbf{X}-\hat{\mathbf{X}}, \\ I(\mathbf{Y}; \hat{\mathbf{X}}) \leq \epsilon, \overline{H}(\hat{\mathbf{X}}|\mathbf{X}) \leq R}} \overline{c}(\mathbf{X}, \hat{\mathbf{X}}).$$

Since proofs of theorems are rather long, we postpone these to Section 4. The only difference of two theorems is using a function $c(\mathbf{X}, \hat{\mathbf{X}})$ or $\overline{c}(\mathbf{X}, \hat{\mathbf{X}})$.

According to [11, Theorem 8 c), d), and e)], it holds that $\overline{H}(\hat{\mathbf{X}}|\mathbf{X}) \leq \log |\hat{\mathcal{X}}|$. Hence, the following two corollaries follow immediately.

Corollary 1. When $\mathbf{X} = \mathbf{Y}$ and $R \geq \log |\hat{\mathcal{X}}|$, we have

$$C_{\mathrm{a}}(\epsilon, R) = \inf_{\hat{\mathbf{X}}: I(\mathbf{X}, \hat{\mathbf{X}}) \leq \epsilon} c(\mathbf{X}, \hat{\mathbf{X}}).$$

Corollary 2. When $\mathbf{X} = \mathbf{Y}$ and $R \geq \log |\hat{\mathcal{X}}|$, we have

$$C_{\mathrm{w}}(\epsilon, R) = \inf_{\hat{\mathbf{X}}: I(\mathbf{X}; \hat{\mathbf{X}}) \leq \epsilon} \overline{c}(\mathbf{X}, \hat{\mathbf{X}}).$$

Right-hand sides of Corollaries 1 and 2 can be regarded as the *distortion-rate* function for the variable-length coding under the average distortion criterion (see, e.g., [8, Remark 5.7.2]) and the maximum distortion criterion (see, e.g., the proof of [8, Theorem 5.6.1]), respectively. This fact allows us to apply many results of the distortion-rate function to our study. For example, according to the proof of [8, Theorem 5.8.1], the minimum costs for stationary memoryless sources are given by the next corollary.

Corollary 3. Let $\mathbf{X} = \mathbf{Y}$ and $R \geq \log |\hat{\mathcal{X}}|$. Further, let $\mathbf{X}$ be a stationary memoryless source induced by an RV X on $\mathcal{X}$, and $c_n : \mathcal{X}^n \times \hat{\mathcal{X}}^n \to [0, \infty)$ be an additive cost function defined by

$$c_n(x^n, \hat{x}^n) \triangleq \frac{1}{n} \sum_{i=1}^{n} c(x_i, \hat{x}_i),$$

where $c : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty)$. Then, we have

$$C_{\mathrm{a}}(\epsilon, R) = C_{\mathrm{w}}(\epsilon, R) = \min_{\hat{X}: I(X; \hat{X}) \leq \epsilon} \mathrm{E}[c(X, \hat{X})].$$

We also consider a mixed source $\mathbf{X}$ of two sources $\mathbf{X}_1$ and $\mathbf{X}_2$ defined by

$$P_{X^n}(x^n) = \alpha P_{X_1^n}(x^n) + (1 - \alpha) P_{X_2^n}(x^n),$$

where $\alpha \in [0, 1]$. According to [8, Remark 5.10.2], we have the next corollary.

Corollary 4. Let $\mathbf{X} = \mathbf{Y}$ and $R \geq \log |\hat{\mathcal{X}}|$. Further, let $c_n : \mathcal{X}^n \times \hat{\mathcal{X}}^n \to [0, \infty)$ be a subadditive cost function that satisfies

$$c_{n+m}((x_1^n, x_2^m), (\hat{x}_1^n, \hat{x}_2^m)) \leq c_n(x_1^n, \hat{x}_1^n) + c_m(x_2^m, \hat{x}_2^m),$$

and $C_{\mathrm{a}}(\epsilon, R|\mathbf{X})$ be the minimum average cost when $\mathbf{X} = \mathbf{Y}$. Then, for a mixed source $\mathbf{X}$ of two stationary sources $\mathbf{X}_1$ and $\mathbf{X}_2$, we have

$$C_{\mathrm{a}}(\epsilon, R|\mathbf{X}) = \inf_{\substack{(\epsilon_1, \epsilon_2) \in [0,\infty)^2: \\ \alpha\epsilon_1 + (1-\alpha)\epsilon_2 \leq \epsilon}} \left(\alpha C_{\mathrm{a}}(\epsilon_1, R|\mathbf{X}_1) + (1 - \alpha) C_{\mathrm{a}}(\epsilon_2, R|\mathbf{X}_2)\right).$$

3 Minimum Costs to Erase Information under the Strong Independence Criterion

In this section, we consider the minimum average cost under the strong independence criterion. In order to clarify the fundamental limit of average costs, we assume that an encoder is a stochastic encoder in this section. In other words, we consider the case where the size of the uniform random number is sufficiently large. We also assume that a source is a stationary memoryless source. Then, we give a single-letter characterization of the minimum average cost and some results obtained from this characterization.

3.1 Problem Formulation

In this section, we define minimum average cost under the strong independence criterion.

Let $(\mathbf{X}, \mathbf{Y})$ be the pair of stationary memoryless sources, i.e., $\{(X_i, Y_i)\}_{n=1}^{\infty}$ be independent copies of a pair of RVs (X, Y) on $\mathcal{X} \times \mathcal{Y}$. For the sake of brevity, we simply express the sources as (X, Y). Let $f_n : \mathcal{X}^n \to \hat{\mathcal{X}}^n$ be a *stochastic* encoder, and $c_n : \mathcal{X}^n \times \hat{\mathcal{X}}^n \to [0, \infty)$ be an additive cost function as defined in Corollary 3, i.e., $c_n(x^n, \hat{x}^n) \triangleq \frac{1}{n} \sum_{i=1}^{n} c(x_i, \hat{x}_i)$, where $c : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty)$ is an arbitrary function.

The achievablility under the strong independence criterion is defined as follows.

Definition 6. For real numbers $\Gamma, \epsilon \geq 0$, we say Γ is ϵ-strongly achievable in the sense of the average cost if and only if there exists a sequence of stochastic encoders $\{f_n\}_{n=1}^{\infty}$ such that

$$\limsup_{n\to\infty} I(Y^n, f_n(X^n)) \leq \epsilon, \quad (4)$$
$$\limsup_{n\to\infty} \mathrm{E}[c_n(X^n, f_n(X^n))] \leq \Gamma,$$

where the expectation is with respect to the sequence X^n and the output of the stochastic encoder f_n.

The difference from the previous section is to use the strong independence criterion in (4).

The minimum average cost under the strong independence criterion is defined as follows.

Definition 7. We define the minimum average cost as

$$C_{\mathrm{a}}^*(\epsilon) \triangleq \inf\{\Gamma : \Gamma \text{ is } \epsilon\text{-strongly achievable in the sense of the average cost}\}.$$

Remark 1. We only consider the average cost in this section. This is because the minimum worst-case cost coincides with the minimum average cost after all for stationary memoryless sources. This is similar to Corollary 3.

We also consider the non-asymptotic version of the achievablity defined as follows.

Definition 8. For an integer $n \geq 1$, and real numbers $\Gamma, \epsilon \geq 0$, we say Γ is (n, ϵ)-strongly achievable in the sense of the average cost if and only if there exists a stochastic encoder f_n such that

$$I(Y^n; f_n(X^n)) \leq \epsilon, \quad (5)$$
$$\mathrm{E}[c_n(X^n, f_n(X^n))] \leq \Gamma.$$

Remark 2. Definition 8 adopts the strong independence criterion in (5). However, this is not important in the non-asymptotic setting because this criterion is regarded as the weak criterion if we set ϵ as $n\epsilon$.

The non-asymptotic minimum average cost is defined as follows.

Definition 9. We define the non-asymptotic minimum average cost for a given finite blocklength $n \geq 1$ as

$$C_{\mathrm{a}}^*(n, \epsilon) \triangleq \inf\{\Gamma : \Gamma \text{ is } (n, \epsilon)\text{-strongly achievable in the sense of the average cost}\}.$$

3.2 Minimum Average Costs

In this section, we give a single-letter characterization of minimum average costs $C_{\mathrm{a}}^*(\epsilon)$ and $C_{\mathrm{a}}^*(n, \epsilon)$. Since this characterization is given by employing usual information-theoretical techniques, this might not be of the main interest. However, results obtained from it are interesting and insightful.

First of all, we show a single-letter characterization of the non-asymptotic minimum average cost $C_{\mathrm{a}}^*(n, \epsilon)$.

Theorem 3. For a pair of stationary memoryless sources (X, Y), any integer $n \geq 1$, and any real number $\epsilon \geq 0$, we have

$$C_{\mathrm{a}}^*(n, \epsilon) = \min_{\substack{\hat{X}: Y - X - \hat{X}, \\ I(Y;\hat{X}) \leq \frac{\epsilon}{n}}} \mathrm{E}[c(X, \hat{X})].$$

Proof. First, we show the converse part. If Γ is (n, ϵ)-strongly achievable in the sense of the average cost, there exists f_n such that

$$I(Y^n; \hat{X}^n) \leq \epsilon,$$
$$\mathrm{E}[c_n(X^n, \hat{X}^n)] \leq \Gamma, \quad (6)$$

where $\hat{X}^n = f_n(X^n)$. We note that

$$\begin{aligned} I(Y^n; \hat{X}^n) &= \sum_{i=1}^{n} I(Y_i; \hat{X}^n | Y^{i-1}) \\ &= \sum_{i=1}^{n} I(Y_i; \hat{X}^n, Y^{i-1}) \\ &\geq \sum_{i=1}^{n} I(Y_i; \hat{X}_i), \end{aligned} \quad (7)$$

where the second equality comes from the fact that Y_i is independent of Y^{i-1}, i.e., $I(Y_i; Y^{i-1}) = 0$. On the other hand, let Q be an RV on $\{1, 2, \cdots, n\}$ and $(Q, Y, X, \hat{X})$ be RVs on $\{1, \cdots, n\} \times \mathcal{Y} \times \mathcal{X} \times \hat{\mathcal{X}}$ such that $P_{QYX\hat{X}}(i, y, x, \hat{x}) = \frac{1}{n} P_{Y_i X_i \hat{X}_i}(y, x, \hat{x})$. Then, we have

$$\epsilon \geq \sum_{i=1}^{n} I(Y_i; \hat{X}_i) = nI(Y; \hat{X}|Q) \geq nI(Y; \hat{X}), \quad (8)$$

where the first inequality comes from (7) and the last inequality comes from the fact that Q is independent of Y. Thus, from (6), we have

$$\Gamma \geq \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[c(X_i, \hat{X}_i)] \geq \min_{\substack{\hat{X}: Y - X - \hat{X}, \\ I(Y;\hat{X}) \leq \frac{\epsilon}{n}}} \mathrm{E}[c(X, \hat{X})],$$

where the last inequality comes from (8) and the fact that $Y - X - \hat{X}$. Since this inequality holds for any (n, ϵ)-strongly achievable Γ, we have

$$C_a^*(n, \epsilon) \geq \min_{\substack{\hat{X}: Y - X - \hat{X}, \\ I(Y;\hat{X}) \leq \frac{\epsilon}{n}}} \mathrm{E}[c(X, \hat{X})].$$

Next, we show the direct part. Let $\hat{X}$ be an RV on $\hat{\mathcal{X}}$ such that $Y - X - \hat{X}$ and

$$I(Y; \hat{X}) \leq \frac{\epsilon}{n}.$$

Then, the direct part is obvious, if we define the encoder as

$$f_n(x^n) = \hat{x}^n \text{ with probability } \prod_{i=1}^{n} P_{\hat{X}|X}(\hat{x}_i|x_i).$$

For this encoder, we have

$$\begin{aligned} I(Y^n; f_n(X^n)) &= nI(Y; \hat{X}) \leq \epsilon, \\ \mathrm{E}[c_n(X^n, f_n(X^n))] &= \mathrm{E}[c(X, \hat{X})]. \end{aligned}$$

Thus, $\mathrm{E}[c(X, \hat{X})]$ is (n, ϵ)-strongly achievable for any $\hat{X}$ such that $Y - X - \hat{X}$ and $I(Y; \hat{X}) \leq \frac{\epsilon}{n}$. This implies that

$$C_a^*(n, \epsilon) \leq \min_{\substack{\hat{X}: Y - X - \hat{X}, \\ I(Y;\hat{X}) \leq \frac{\epsilon}{n}}} \mathrm{E}[c(X, \hat{X})]. \qquad \square$$

Next, we give a single-letter characterization of the minimum average cost $C_a^*(\epsilon)$ which shows that it is impossible to reduce the minimum cost by allowing information leakage.

Theorem 4. For a pair of stationary memoryless sources (X, Y) and any $\epsilon \geq 0$, we have

$$C_a^*(\epsilon) = \min_{\substack{\hat{X}: Y - X - \hat{X}, \\ I(Y;\hat{X}) = 0}} \mathrm{E}[c(X, \hat{X})].$$

Proof. If Γ is ϵ-strongly achievable in the sense of the average cost, there exists f_n such that for any $\delta > 0$ and all sufficiently large $n > 0$,

$$\begin{aligned} I(Y^n; \hat{X}^n) &\leq \epsilon + \delta, \\ \mathrm{E}[c_n(X^n, \hat{X}^n)] &\leq \Gamma + \delta, \end{aligned}$$

where $\hat{X}^n = f_n(X^n)$. By noting that $\delta > 0$ is arbitrary and $\min_{\hat{X}: Y - X - \hat{X}, I(Y;\hat{X}) \leq \epsilon} \mathrm{E}[c(X, \hat{X})]$ is continuous at $\epsilon = 0$ (see A), the rest of the proof can be done in the same way as the proof of Theorem 3. Hence, we omit the details. $\square$

According to Theorem 3 and Theorem 4, it holds that for any $n \geq 1$ and $\epsilon \geq 0$,

$$C_a^*(\epsilon) = C_a^*(n, 0).$$

Hence, we only consider $C_a^*(n, \epsilon)$ because $C_a^*(\epsilon)$ is a special case of it.

As in the previous section, the next corollary follows immediately.

Corollary 5. When $X = Y$, we have

$$C_a^*(n, \epsilon) = \min_{\hat{X}: I(X;\hat{X}) \leq \frac{\epsilon}{n}} \mathrm{E}[c(X, \hat{X})]. \tag{9}$$

According to this corollary and Corollary 3, when $\mathbf{X} = \mathbf{Y}$ and $\mathbf{X}$ is a stationary memoryless source, it holds that for any $\epsilon \geq 0$,

$$C_a(\epsilon, R) = C_w(\epsilon, R) = C_a^*(1, \epsilon).$$

Since the right-hand side of (9) is the distortion-rate function, we have some closed-from expressions of the minimum cost (see. e.g., [8] and [12]). For example, let $\mathcal{X} = \hat{\mathcal{X}} = \{0, 1\}$, $P_X(0) = p$, and $c(x, \hat{x}) = \mathbf{1}\{x \neq \hat{x}\}$, where $p \in [0, 1/2]$ and $\mathbf{1}\{\cdot\}$ denotes the indicator function. Then, we have

$$C_a^*(n, \epsilon) = h^{-1}(|h(p) - \epsilon/n|^+), \tag{10}$$

where $|x|^+ = \max\{0, x\}$, $h(p) = -p \log p - (1 - p) \log(1 - p)$, and $h^{-1} : [0, \log 2] \to [0, 1/2]$ is the inverse function of h.

Furthermore, according to Corollary 5, when $X = Y$, it holds that

$$C_a^*(n, 0) = \min_{\hat{x} \in \hat{\mathcal{X}}} \mathrm{E}[c(X, \hat{x})] \triangleq \Gamma_{\min}, \quad \forall n \geq 1,$$

where the first equality comes from the fact that X and $\hat{X}$ are independent. Interestingly, this can be achieved by a certain *deterministic* encoder as follows: Let $\tilde{x} = \mathrm{argmin}_{\hat{x} \in \hat{\mathcal{X}}} \mathrm{E}[c(X, \hat{x})]$ and define an encoder $f_n^{(\mathrm{r})}$ as

$$f_n^{(\mathrm{r})}(x^n) \triangleq (\tilde{x}, \cdots, \tilde{x}), \quad \forall x^n \in \mathcal{X}^n.$$

Then, this encoder achieves $C_a^*(n, 0)$ $(= \Gamma_{\min})$, i.e., we have

$$I(Y^n; f_n^{(\mathrm{r})}(X^n)) = 0, \tag{11}$$

$$\begin{aligned} \mathrm{E}[c_n(X^n, f_n^{(\mathrm{r})}(X^n))] &= \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[c(X_i, \tilde{x})] \\ &= \mathrm{E}[c(X, \tilde{x})] = \Gamma_{\min}. \end{aligned} \tag{12}$$

This means that when $X = Y$, the optimal method of erasure is to overwrite the source sequence with repeated identical symbols using $f_n^{(\mathrm{r})}$. We note that $f_n^{(\mathrm{r})}$ gives the minimum average cost among encoders using repeated identical symbols.

Next, we give a sufficient condition such that $C_a^*(n, 0)$ can be achieved by the encoder $f_n^{(\mathrm{r})}$. Then, we show that the case where $X = Y$ is a special case of the sufficient condition. To this end, we define the *weak independence* introduced by Berger and Yeung [13].

Definition 10 (Weak independence)**.** For a pair (X, Y) of RVs, let $P_{Y|X}(\cdot|x) = (P_{Y|X}(y|x) : y \in \mathcal{Y})$ be the xth row of the stochastic matrix $P_{Y|X}$. Then, we say Y is *weakly independent* of X if the rows $P_{Y|X}(\cdot|x)$ $(x \in \mathcal{X})$ are linearly dependent.

Remark 3. If X is binary, then Y is weakly independent of X if and only if Y and X are independent [13, Remark 3].

The weak independence has a useful property for independence of a triple of RVs satisfying a Markov chain. This property is shown in the next lemma.

Lemma 1 ([13, Theorem 4])**.** Let $|\hat{\mathcal{X}}| \geq 2$. Then, for a pair (X, Y) of RVs, there exists an RV $\hat{X}$ satisfying

1. $Y - X - \hat{X}$
2. Y and $\hat{X}$ are independent
3. X and $\hat{X}$ are not independent

if and only if Y is weakly independent of X.

Now, we give a sufficient condition.

Theorem 5. If Y is *not* weakly independent of X, the optimal method of erasure is to overwrite the source sequence with repeated identical symbols using $f_n^{(\mathrm{r})}$, i.e., it holds that

$$I(Y^n; f_n^{(\mathrm{r})}(X^n)) = 0,$$
$$\mathrm{E}[c_n(X^n, f_n^{(\mathrm{r})}(X^n))] = C_{\mathrm{a}}^*(n, 0).$$

Proof. Since we immediately obtain that $I(Y^n; f_n^{(\mathrm{r})}(X^n)) = 0$ and $\mathrm{E}[c_n(X^n, f_n^{(\mathrm{r})}(X^n))] = \Gamma_{\min}$ (see (11) and (12)), we only have to show that $C_{\mathrm{a}}^*(n, 0) = \Gamma_{\min}$.

Since Y is not weakly independent of X, there not exists an RV $\hat{X}$ simultaneously satisfying three conditions in Lemma 1. This implies that for any $\hat{X}$ such that $Y - X - \hat{X}$ and $I(Y; \hat{X}) = 0$, it must satisfy that $I(X; \hat{X}) = 0$. This is because if $I(X; \hat{X}) > 0$, $\hat{X}$ simultaneously satisfies three conditions in Lemma 1.

Thus, we have

$$\begin{aligned} C_{\mathrm{a}}^*(n, 0) &= \min_{\substack{\hat{X}: Y - X - \hat{X}, \\ I(Y; \hat{X}) = 0}} \mathrm{E}[c(X, \hat{X})] \\ &\overset{(a)}{=} \min_{\substack{\hat{X}: Y - X - \hat{X}, \\ I(Y; \hat{X}) = 0, I(X; \hat{X}) = 0}} \mathrm{E}[c(X, \hat{X})] \\ &\overset{(b)}{=} \min_{\substack{\hat{X}: Y - X - \hat{X}, \\ I(Y; \hat{X}) = 0, I(X; \hat{X}) = 0}} \sum_{\hat{x} \in \hat{\mathcal{X}}} P_{\hat{X}}(\hat{x}) \mathrm{E}[c(X, \hat{x})] \\ &\geq \Gamma_{\min}, \end{aligned}$$

where (a) comes from the above argument and (b) follows since X and $\hat{X}$ are independent.

Since the opposite direction is obvious by setting $\hat{X} = \tilde{x}$ with probability 1, this completes the proof. □

If $X = Y$, Y is not weakly independent of X. Thus, this is a special case of this sufficient condition. According to Remark 3, we can also show that if X is binary, the encoder $f_n^{(\mathrm{r})}$ is optimal as long as Y and X are not independent.

On the other hand, if Y is weakly independent of X, $C_{\mathrm{a}}^*(n, 0)$ cannot be achieved by the repeated symbols using the encoder $f_n^{(\mathrm{r})}$ in general. To show this fact, we give an example such that $C_{\mathrm{a}}^*(n, 0) < \Gamma_{\min}$. Let $\mathcal{Y} = \{0, 1\}$, $\mathcal{X} = \hat{\mathcal{X}} = \{0, 1, 2\}$, $c(x, \hat{x}) = \mathbf{1}\{x \neq \hat{x}\}$, $P_X(x) = 1/3$ for all $x \in \{0, 1, 2\}$, and

$$P_{Y|X} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix},$$

Table 1: This table shows that $f_n^{(\mathrm{r})}$ is optimal or not in the sense that it whenever can achieve the minimum average cost $C_{\mathrm{a}}^*(n, \epsilon)$ or not for each corresponding condition. WI is an abbreviation for "weakly independent".

	Y is not WI of X	Y is WI of X
$\epsilon = 0$	optimal	not optimal
$\epsilon > 0$	not optimal	not optimal

where the xth row and the yth column denotes the conditional probability $P_{Y|X}(y|x)$. Then, we have $\Gamma_{\min} = 2/3$. We note that Y is weakly independent of X. On the other hand, we consider an RV $\hat{X}$ such that $Y - X - \hat{X}$, and

$$P_{\hat{X}|X} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/6 & 2/3 & 1/6 \\ 1/2 & 0 & 1/2 \end{bmatrix},$$

where the xth row and the $\hat{x}$th column denotes the conditional probability $P_{\hat{X}|X}(\hat{x}|x)$. Then, one can easily check that Y is independent of $\hat{X}$, and

$$C_{\mathrm{a}}^*(n, 0) \leq \mathrm{E}[c(X, \hat{X})] = 1/2 < \Gamma_{\min}. \tag{13}$$

Hence, the encoder $f_n^{(\mathrm{r})}$ is no longer optimal.

Further, if we allow a little bit of leakage of information, i.e., $\epsilon > 0$, the encoder $f_n^{(\mathrm{r})}$ is no longer optimal even if Y is not weakly independent of X. This is because in general, it holds that $C_{\mathrm{a}}^*(n, \epsilon) < \Gamma_{\min}$ for $\epsilon > 0$ (see (10) and also (13)).

The optimality of the encoder $f_n^{(\mathrm{r})}$ is summarized in Table 1.

4 Proofs of Theorems

In this section, we prove Theorems 1 and 2.

4.1 Fundamental Lemmas for the Random Number Generation

In this section, we introduce some lemmas to prove Theorems 1 and 2. Since proofs of these lemmas are similar to the proofs in [8, Section 2], we will omit the details.

For two probability distributions P and Q on the same set $\mathcal{X}$, we define the variational distance $d(P, Q)$ as

$$d(P, Q) \triangleq \sum_{x \in \mathcal{X}} |P(x) - Q(x)|.$$

For all lemmas in this section, let $(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \{(X^n, Y^n, Z^n)\}_{n=1}^{\infty}$ be a triple of sequences of RVs, where (X^n, Y^n, Z^n) is a triple of RVs on $\mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n$. For this triple, we define

$$\mathcal{S}_n(\alpha) \triangleq \left\{(x^n, z^n) \in \mathcal{X}^n \times \mathcal{Z}^n : \frac{1}{n} \log \frac{1}{P_{X^n|Z^n}(x^n|z^n)} \geq \alpha \right\},$$

$$\mathcal{T}_n(\beta) \triangleq \Big\{(y^n, z^n) \in \mathcal{Y}^n \times \mathcal{Z}^n : \frac{1}{n}\log\frac{1}{P_{Y^n|Z^n}(y^n|z^n)} \le \beta\Big\}.$$

The next lemma is an extended version of [8, Lemma 2.1.1].

Lemma 2. For any integer $n \ge 1$ and any real numbers $\gamma > 0$ and $a \in \mathbb{R}$, there exists a mapping $\varphi_n : \mathcal{X}^n \times \mathcal{Z}^n \to \mathcal{Y}^n$ satisfying

$$\begin{aligned} d(P_{Y^nZ^n}, P_{\tilde{Y}^nZ^n}) \le{} & 2\Pr\{(X^n, Z^n) \notin \mathcal{S}_n(a+\gamma)\} \\ & + 2\Pr\{(Y^n, Z^n) \notin \mathcal{T}_n(a)\} + 2e^{-n\gamma}, \end{aligned}$$

where $\tilde{Y}^n = \varphi_n(X^n, Z^n)$.

Proof. Since this lemma can be easily proved in the same manner as the proof of [8, Lemma 2.1.1], we omit the details. □

The next lemma gives a sufficient condition to simulate the correlation of a pair of RVs from another RV.

Lemma 3. If $\underline{H}(\mathbf{X}|\mathbf{Z}) > \overline{H}(\mathbf{Y}|\mathbf{Z})$, there exists a mapping $\varphi_n : \mathcal{X}^n \times \mathcal{Z}^n \to \mathcal{Y}^n$ satisfying

$$\lim_{n\to\infty} d(P_{Y^nZ^n}, P_{\tilde{Y}^nZ^n}) = 0,$$

where $\tilde{Y}^n = \varphi_n(X^n, Z^n)$ and

$$\underline{H}(\mathbf{X}|\mathbf{Z}) = \operatorname{p-lim\,inf}_{n\to\infty} \frac{1}{n}\log\frac{1}{P_{X^n|Z^n}(X^n|Z^n)}.$$

Proof. Since this lemma can be easily proved by using Lemma 2 and the same manner as the proof of [8, Theorem 2.1.1], we omit the details. □

The next lemma is an extended version of [8, Lemma 2.1.2].

Lemma 4. For any integer $n \ge 1$, any real numbers $\gamma > 0$ and $a \in \mathbb{R}$, and any mapping $\varphi_n : \mathcal{X}^n \times \mathcal{Z}^n \to \mathcal{Y}^n$, it holds that

$$\begin{aligned} d(P_{Y^nZ^n}, P_{\tilde{Y}^nZ^n}) \ge{} & 2\Pr\{(Y^n, Z^n) \notin \mathcal{T}_n(a+\gamma)\} \\ & - 2\Pr\{(X^n, Z^n) \in \mathcal{S}_n(a)\} - 2e^{-n\gamma}, \end{aligned}$$

where $\tilde{Y}^n = \varphi_n(X^n, Z^n)$.

Proof. Since this lemma can be easily proved in the same manner as the proof of [8, Lemma 2.1.2], we omit the details. □

According to this lemma, we have the next lemma which is an information spectrum version of the fact that

$$H(X|Z) \ge H(\varphi(X, Z)|Z)$$

for any function φ, where $H(X|Z)$ is the conditional entropy of X given Z.

Lemma 5. Let $\varphi_n : \mathcal{X}^n \times \mathcal{Z}^n \to \mathcal{Y}^n$ be an arbitrary mapping and set $\tilde{Y}^n = \varphi_n(X^n, Z^n)$ and $\tilde{\mathbf{Y}} = \{\tilde{Y}^n\}_{n=1}^\infty$. Then, it holds that

$$\overline{H}(\mathbf{X}|\mathbf{Z}) \ge \overline{H}(\tilde{\mathbf{Y}}|\mathbf{Z}).$$

Proof. Since this lemma can be easily proved by using Lemma 4 and the same manner as the proof of [8, Corollary 2.1.2], we omit the details. □

4.2 Direct Part

In this section, we first show that

$$C_a(\epsilon, R) \le \inf_{\substack{\hat{\mathbf{X}}:\mathbf{Y}-\mathbf{X}-\hat{\mathbf{X}},\\ I(\mathbf{Y};\hat{\mathbf{X}})\le\epsilon, \overline{H}(\hat{\mathbf{X}}|\mathbf{X})\le R}} c(\mathbf{X}, \hat{\mathbf{X}}). \tag{14}$$

In other words, we show the direct part of the proof of Theorem 1.

For given R and ϵ, let $\hat{\mathbf{X}}$ be a sequence of RVs such that

$$\mathbf{Y} - \mathbf{X} - \hat{\mathbf{X}}, \tag{15}$$

$$\overline{H}(\hat{\mathbf{X}}|\mathbf{X}) \le R, \tag{16}$$

$$I(\mathbf{Y}; \hat{\mathbf{X}}) \le \epsilon. \tag{17}$$

For an arbitrarily fixed $\delta > 0$, let $\{M_n\}_{n=1}^\infty$ be a sequence of integers such that

$$M_n = \left\lceil e^{n(R+\delta)} \right\rceil. \tag{18}$$

Then, we have

$$\underline{H}(\mathbf{U}|\mathbf{X}) = R + \delta > \overline{H}(\hat{\mathbf{X}}|\mathbf{X}),$$

where $\mathbf{U} = \{U_{M_n}\}_{n=1}^\infty$ and the inequality comes from (16). Thus, according to Lemma 3, there exists a sequence of functions $f_n : \mathcal{X}^n \times \mathcal{U}_{M_n} \to \hat{\mathcal{X}}^n$ such that

$$\lim_{n\to\infty} d(P_{\hat{X}^nX^n}, P_{\tilde{X}^nX^n}) = 0,$$

where $\tilde{X}^n = f_n(X^n, U_{M_n})$. Since $Y^n - X^n - \hat{X}^n$ and $Y^n - X^n - \tilde{X}^n$, we also have

$$\lim_{n\to\infty} d(P_{\hat{X}^nX^nY^n}, P_{\tilde{X}^nX^nY^n}) = \lim_{n\to\infty} d(P_{\hat{X}^nX^n}, P_{\tilde{X}^nX^n}) = 0.$$

Hence from the continuity of the mutual information (see, e.g., [14, Lemma 2.7]), we have

$$I(\mathbf{Y}; \tilde{\mathbf{X}}) = I(\mathbf{Y}; \hat{\mathbf{X}}) \le \epsilon, \tag{19}$$

where $\tilde{\mathbf{X}} = \{\tilde{X}^n\}_{n=1}^\infty$ and the last inequality comes from (17). We also have

$$\begin{aligned} & c(\mathbf{X}, \tilde{\mathbf{X}}) - c(\mathbf{X}, \hat{\mathbf{X}}) \\ & = \limsup_{n\to\infty} \mathrm{E}[c_n(X^n, \tilde{X}^n)] - \limsup_{n\to\infty} \mathrm{E}[c_n(X^n, \hat{X}^n)] \\ & \le \limsup_{n\to\infty} (\mathrm{E}[c_n(X^n, \tilde{X}^n)] - \mathrm{E}[c(X^n, \hat{X}^n)]) \\ & \le \limsup_{n\to\infty} d(P_{\hat{X}^nX^n}, P_{\tilde{X}^nX^n}) c_{\max} = 0. \end{aligned} \tag{20}$$

According to (18), (19), and (20), there exist $\{M_n\}_{n=1}^\infty$ and $\{f_n\}_{n=1}^\infty$ such that

$$\limsup_{n\to\infty} \frac{1}{n}\log M_n \le R + \delta,$$

$$I(\mathbf{Y}; \tilde{\mathbf{X}}) \le \epsilon,$$

$$c(\mathbf{X}, \tilde{\mathbf{X}}) \le c(\mathbf{X}, \hat{\mathbf{X}})$$

for any sequence $\hat{\mathbf{X}}$ of RVs satisfying (15), (16), and (17). This means that $(R+\delta, c(\mathbf{X}, \hat{\mathbf{X}}))$ is ϵ-weakly achievable for any $\delta > 0$. Then, by using the usual diagonal line argument [8], we can show that $(R, c(\mathbf{X}, \hat{\mathbf{X}}))$ is also ϵ-weakly achievable. This implies (14).

For the same RV $\tilde{X}^n = f_n(X^n, U_{M_n})$ as above, we also have

$$\begin{aligned}\overline{c}(\mathbf{X}, \tilde{\mathbf{X}}) &= \inf\{\alpha : \lim_{n\to\infty} \Pr\{c_n(X^n, \tilde{X}^n) > \alpha\} = 0\} \\ &= \inf\{\alpha : \lim_{n\to\infty} \Pr\{c_n(X^n, \hat{X}^n) > \alpha\} = 0\} \\ &= \overline{c}(\mathbf{X}, \hat{\mathbf{X}}), \end{aligned} \tag{21}$$

where the second equality comes from the fact that

$$\limsup_{n\to\infty} \Pr\{c_n(X^n, \tilde{X}^n) > \alpha\} = \limsup_{n\to\infty} \Pr\{c_n(X^n, \hat{X}^n) > \alpha\}.$$

Thus, by replacing (20), $c(\mathbf{X}, \tilde{\mathbf{X}})$, and $c(\mathbf{X}, \hat{\mathbf{X}})$ with (21), $\overline{c}(\mathbf{X}, \tilde{\mathbf{X}})$, and $\overline{c}(\mathbf{X}, \hat{\mathbf{X}})$, respectively, and repeating the same argument as above, we also have

$$C_{\mathrm{w}}(\epsilon, R) \le \inf_{\substack{\hat{\mathbf{X}}:\mathbf{Y}-\mathbf{X}-\hat{\mathbf{X}},\\ I(\mathbf{Y};\hat{\mathbf{X}})\le\epsilon, \overline{H}(\hat{\mathbf{X}}|\mathbf{X})\le R}} \overline{c}(\mathbf{X}, \hat{\mathbf{X}}).$$

This is the direct part of the proof of Theorem 2.

4.3 Converse Part

In this section, we first show that

$$C_{\mathrm{a}}(\epsilon, R) \ge \inf_{\substack{\hat{\mathbf{X}}:\mathbf{Y}-\mathbf{X}-\hat{\mathbf{X}},\\ I(\mathbf{Y};\hat{\mathbf{X}})\le\epsilon, \overline{H}(\hat{\mathbf{X}}|\mathbf{X})\le R}} c(\mathbf{X}, \hat{\mathbf{X}}). \tag{22}$$

In other words, we show the converse part of the proof of Theorem 1.

If (R, Γ) is ϵ-weakly achievable, there exist sequences of integers $\{M_n\}_{n=1}^{\infty}$ and encoders $\{f_n\}_{n=1}^{\infty}$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n \le R, \tag{23}$$

$$I(\mathbf{Y}; \hat{\mathbf{X}}) \le \epsilon, \tag{24}$$

$$c(\mathbf{X}, \hat{\mathbf{X}}) \le \Gamma, \tag{25}$$

where $\hat{\mathbf{X}} = \{\hat{X}^n\}_{n=1}^{\infty}$ and $\hat{X}^n = f_n(X^n, U_{M_n})$.

According to Lemma 5, we have

$$\overline{H}(\mathbf{U}|\mathbf{X}) \ge \overline{H}(\hat{\mathbf{X}}|\mathbf{X}).$$

On the other hand, due to (23), we have

$$\overline{H}(\mathbf{U}|\mathbf{X}) \le R.$$

Thus, we have

$$\overline{H}(\hat{\mathbf{X}}|\mathbf{X}) \le R. \tag{26}$$

Now, by combining (24), (25), (26), and the fact that $\hat{\mathbf{X}}$ satisfies $\mathbf{Y} - \mathbf{X} - \hat{\mathbf{X}}$, we have

$$\Gamma \ge c(\mathbf{X}, \hat{\mathbf{X}}) \ge \inf_{\substack{\hat{\mathbf{X}}:\mathbf{Y}-\mathbf{X}-\hat{\mathbf{X}},\\ I(\mathbf{Y};\hat{\mathbf{X}})\le\epsilon, \overline{H}(\hat{\mathbf{X}}|\mathbf{X})\le R}} c(\mathbf{X}, \hat{\mathbf{X}}).$$

Since this inequality holds for any Γ such that (R, Γ) is ϵ-weakly achievable, we have (22).

By replacing $c(\mathbf{X}, \hat{\mathbf{X}})$ with $\overline{c}(\mathbf{X}, \hat{\mathbf{X}})$ and repeating the same argument as above, we also have

$$C_{\mathrm{w}}(\epsilon, R) \ge \inf_{\substack{\hat{\mathbf{X}}:\mathbf{Y}-\mathbf{X}-\hat{\mathbf{X}},\\ I(\mathbf{Y};\hat{\mathbf{X}})\le\epsilon, \overline{H}(\hat{\mathbf{X}}|\mathbf{X})\le R}} \overline{c}(\mathbf{X}, \hat{\mathbf{X}}). \tag{27}$$

This is the converse part of the proof of Theorem 2.

5 Conclusion

In this paper, we introduced the information erasure model and considered minimum costs under the weak and the strong independence criteria. For the weak independence criterion, we characterized the minimum average and the minimum worst-case costs for general sources by using information-spectrum quantities. On the other hand, for the strong independence criterion, we gave a single-letter characterization of the minimum average cost for stationary memoryless sources. By using this characterization, we gave a sufficient condition such that the optimal method of erasure is to overwrite the source sequence with repeated identical symbols.

Acknowledgment

The authors would like to thank Prof. H. Yamamoto for teaching us his result [10], Prof. Y. Oohama for teaching us the paper [13], and Prof. H. Yagi for teaching us the paper [9].

References

[1] T. Matsuta and T. Uyematsu, “On the minimum cost to erase information: An information theoretic approach,” Proc. 39th Symp. on Inf. Theory and its Apps. (SITA2016), pp.176–181, Dec. 2016.

[2] T. Matsuta and T. Uyematsu, “On the minimum worst-case cost and the minimum average cost to erase information,” Proc. 2017 IEEE Inf. Theory Workshop, pp.254–258, Nov. 2017.

[3] U. S. Department of Defense, 5220.22-M National Industrial Security Program Operating Manual, Jan. 1995.

[4] P. Gutmann, “Secure deletion of data from magnetic and solid-state memory,” Proc. Sixth USENIX Security Symp., San Jose, CA, pp.77–90, July 1996.

[5] B. Schneier, Applied Cryptography: Protocols, Algorithms, and Source Code in C, John Wiley & Sons, Inc., New York, NY, USA, 1996.

[6] U. S. Air Force, Air Force System Security Instruction 5020, 1998.

[7] U. S. National Institute of Standards and Technology, Special Publication 800-88: Guidelines for Media Sanitization, Sep. 2006.

[8] T. S. Han, Information-Spectrum Methods in Information Theory, Springer, 2003.

[9] K. Kalantari, L. Sankar, and O. Kosut, “On information-theoretic privacy with general distortion cost functions,” Proc. IEEE Int. Symp. on Inf. Theory, pp.2865–2869, June 2017.

[10] H. Yamamoto, “A source coding problem for sources with additional outputs to keep secret from the receiver or wiretappers (corresp.),” IEEE Trans. Inf. Theory, vol.29, no.6, pp.918–923, Nov. 1983.

[11] S. Verdú and T. S. Han, “A general formula for channel capacity,” IEEE Trans. Inf. Theory, vol.40, no.4, pp.1147–1157, Jul. 1994.

[12] T. M. Cover and J. A. Thomas, Elements of Information Theory, 2nd ed., Wiley, New York, 2006.

[13] T. Berger and R.W. Yeung, “Multiterminal source encoding with encoder breakdown,” IEEE Trans. Inf. Theory, vol.35, no.2, pp.237–244, Mar. 1989.

[14] I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, 2nd ed., Cambridge University Press, 2011.

A Continuity at $\epsilon = 0$

In this appendix, we show that

$$\lim_{\epsilon \downarrow 0} \min_{\substack{\hat{X}: Y - X - \hat{X},\\ I(Y;\hat{X}) \le \epsilon}} \mathrm{E}[c(X,\hat{X})] = \min_{\substack{\hat{X}: Y - X - \hat{X},\\ I(Y;\hat{X}) = 0}} \mathrm{E}[c(X,\hat{X})]. \tag{28}$$

Let $\{\epsilon_n\}_{n=1}^{\infty}$ be a sequence such that $\epsilon_n > 0$ and $\epsilon_n \to 0$. Then, we have

$$\lim_{\epsilon \downarrow 0} \min_{\substack{\hat{X}: Y - X - \hat{X},\\ I(Y;\hat{X}) \le \epsilon}} \mathrm{E}[c(X,\hat{X})] = \lim_{n\to\infty} \min_{\substack{\hat{X}: Y - X - \hat{X},\\ I(Y;\hat{X}) \le \epsilon_n}} \mathrm{E}[c(X,\hat{X})]. \tag{29}$$

Let $P_{\hat{X}^{(n)}|X} : \mathcal{X} \to \hat{\mathcal{X}}$ be a conditional probability distribution such that

$$\mathrm{E}[c(X,\hat{X}^{(n)})] = \min_{\substack{\hat{X}: Y - X - \hat{X},\\ I(Y;\hat{X}) \le \epsilon_n}} \mathrm{E}[c(X,\hat{X})], \tag{30}$$

$$I(Y;\hat{X}^{(n)}) \le \epsilon_n. \tag{31}$$

Then, for the sequence $\{P_{\hat{X}^{(n)}|X}\}_{n=1}^{\infty}$, there exists a convergent subsequence $\{P_{\hat{X}^{(n_k)}|X}\}_{k=1}^{\infty}$ such that $P_{\hat{X}^{(n_k)}|X} \to P_{\tilde{X}|X}$ $(k \to \infty)$, where $P_{\tilde{X}|X} : \mathcal{X} \to \hat{\mathcal{X}}$ is also a conditional probability distribution. Then, by the continuity, we have

$$\begin{aligned}\mathrm{E}[c(X,\tilde{X})] &= \lim_{k\to\infty} \mathrm{E}[c(X,\hat{X}^{(n_k)})] \\ &\overset{\text{(a)}}{=} \lim_{\epsilon \downarrow 0} \min_{\substack{\hat{X}: Y - X - \hat{X},\\ I(Y;\hat{X}) \le \epsilon}} \mathrm{E}[c(X,\hat{X})],\end{aligned}$$

and

$$\begin{aligned}I(Y;\tilde{X}) &= \lim_{k\to\infty} I(Y;\hat{X}^{(n_k)}) \\ &\overset{\text{(b)}}{\le} \lim_{k\to\infty} \epsilon_{n_k} = 0,\end{aligned}$$

where (a) comes from (29) and (30), and (b) comes from (31). Thus, we have

$$\begin{aligned}\min_{\substack{\hat{X}: Y - X - \hat{X},\\ I(Y;\hat{X}) = 0}} \mathrm{E}[c(X,\hat{X})] &\le \mathrm{E}[c(X,\tilde{X})] \\ &= \lim_{\epsilon \downarrow 0} \min_{\substack{\hat{X}: Y - X - \hat{X},\\ I(Y;\hat{X}) \le \epsilon}} \mathrm{E}[c(X,\hat{X})]. \end{aligned} \tag{32}$$

Since the opposite direction is obvious, we have (28) from (32).