Article

Utility–Privacy Trade-Offs with Limited Leakage for Encoder

Naruki Shinohara [†] and Hideki Yagi [*,†]

Department of Computer and Network Engineering, The University of Electro-Communications,
1-5-1 Chofugaoka, Chofu 182-8585, Tokyo, Japan; s1710291@mail.uec.jp
[*] Correspondence: h.yagi@uec.ac.jp; Tel.: +81-42-443-5366
[†] These authors contributed equally to this work.

Abstract: The utilization of databases such as IoT has progressed, and understanding how to protect the privacy of data is an important issue. As pioneering work, in 1983, Yamamoto assumed the source (database), which consists of public information and private information, and found theoretical limits (first-order rate analysis) among the coding rate, utility and privacy for the decoder in two special cases. In this paper, we consider a more general case based on the work by Shinohara and Yagi in 2022. Introducing a measure of privacy for the encoder, we investigate the following two problems: The first problem is the first-order rate analysis among the coding rate, utility, privacy for the decoder, and privacy for the encoder, in which utility is measured by the expected distortion or the excess-distortion probability. The second task is establishing the strong converse theorem for utility–privacy trade-offs, in which utility is measured by the excess-distortion probability. These results may lead to a more refined analysis such as the second-order rate analysis.

Keywords: utility–privacy trade-offs; source coding; Shannon theory; strong converse theorem

1. Introduction

1.1. Background

The utilization of database has progressed in our society and includes autonomous cars and the congestion data service over the Internet. At the same time, the risk of accidental or intentional leakage of private information has also increased rapidly. To protect private information, coding with a privacy constraint has been analyzed via an information-theoretic approach. In 1983, Yamamoto [1] introduced a framework to quantify the utility of databases and the privacy of personal information and analyzed the trade-offs between them. Decades later, in 2013, Sankar et al. [2] claimed the necessity of converting databases to protect privacy while maintaining the utility of data. Then, Yamamoto's framework [1] was re-recognized by Sankar et al. and other researchers. Using the **rate-distortion theory** in information theory, he revealed the optimal relationships (theoretical limits) among coding rate, utility, and privacy in two cases; (i) public information that can be open to the public and private information that should be protected from a third party are encoded, and (ii) only public information is encoded. However, since a more general case, i.e., where (iii) public information and a part of private information is encoded, had not been clarified, Shinohara and Yagi [3] derived the theoretical limits in such a case (see Figure 1). As a result, our characterization of the achievable region gives a "unified expression" because it includes the characteristics given in [1] in cases (i) and (ii) as special cases.

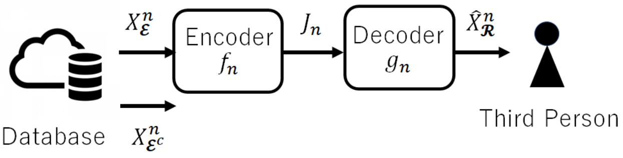

Figure 1. Privacy-constrained coding system.

check for updates

Citation: Shinohara, N.; Yagi, H. Utility–Privacy Trade-Offs with Limited Leakage for Encoder. *Entropy* **2023**, *25*, 921. https://doi.org/10.3390/e25060921

Academic Editors: Jun Chen and Sadaf Salehkalaibar

Received: 10 April 2023
Revised: 28 May 2023
Accepted: 6 June 2023
Published: 11 June 2023

1.2. Motivation and Contributions

By investigating case (iii), one can compare the theoretical limits corresponding to a variety of patterns of the encoded information. One can see that the achievable region in case (i) is the largest among all patterns. However, this may not be the case if **privacy leakage for the encoder** is constrained. Motivated by this observation, in this paper, we characterize the optimal trade-offs among coding rate, utility, privacy for the decoder, and **privacy for the encoder** in Section 3. The addressed problem corresponds to the case where there are some aggregators between the source and the encoder and the aggregator controls the data (source sequence) passing to the encoder. The obtained results indeed suggest that the best-encoded information can be in case (iii) if some restriction is imposed on the privacy leakage for the encoder.

One of the most important tasks in information-theoretic analysis for utility–privacy trade-offs is **second-order rate analysis** (e.g., [4–6]). In general, in second-order rate analysis, the **excess-distortion probability** is used as a measure of utility [4–6]. However, in the **first-order rate analysis** shown in [3], utility is measured by the **expected distortion**, so for second-order rate analysis, we need first to conduct first-order rate analysis, which replaces the expected distortion with the excess-distortion probability as the measure of utility. In Section 4, the theoretical limits coincide with the one in which utility is measured by expected distortion.

There is one more problem to solve before tackling second-order rate analysis: we need to clarify whether the boundary of the achievable region may vary or not, depending on the value of the excess-distortion probability. In Section 5, we establish the **strong converse** theorem, provided that utility is measured by the probability of excess distortion. For the sake of simplicity, we focus on the achievable region of utility and privacy for the decoder or a third party, which reveals an aspect of utility–privacy trade-offs. In the proof, we adopt a change in measure argument developed by Tyagi and Watanabe [7]. Contrary to the standard rate-distortion problem, the alphabets of the encoder's input and the decoder's output are different, so we extend the argument to incorporate this discrepancy. Although the strong converse theorem is shown for the rate region of utility and privacy, we can also derive the same result when the privacy of the encoder is involved.

For readers' convenience, Figure 2 shows the road map to the most important task: the second-order rate analysis. In summary, three contributions of this paper are as follows:

1. The rate analysis among the coding rate, utility, privacy for the decoder, and privacy for the encoder in which utility is measured using the expected distortion (Section 3).
2. The rate analysis among the coding rate, utility, privacy for the decoder, and privacy for the encoder in which utility is measured using the excess-distortion probability (Section 4).
3. The strong converse theorem for utility–privacy trade-offs in which utility is measured using the excess-distortion probability (Section 5).

1.3. Related Work

The analysis of the utility–privacy trade-offs using an information-theoretic approach was initiated by [2], which translates the rate-distortion problem with an equivocation constraint in [1] into the privacy and utility trade-off problem. In information-theoretic studies on coding with privacy and utility constraints, several measures for privacy and utility are adopted. One of the strong measures for privacy is differential privacy [8,9], and an extension and relaxation of differential privacy have been proposed in [10,11]. A weaker but useful privacy measure is the mutual information between the codeword and private information [1,2,12–14], which guarantees the average amount of leaked private information. Other examples of well-known privacy measures are maximal leakage [15], maximal α-leakage [16–18], and total variation [19]. Relationships among several measures for privacy have been revealed in [20]. On the other hand, well-known utility measures are average distortion [1–25], hard distortion [16,17], and log-loss distortion [26].

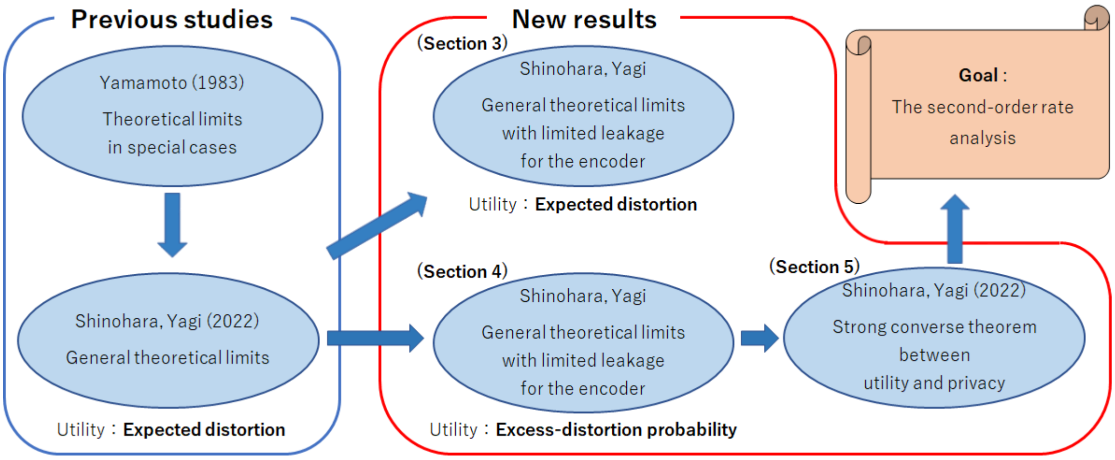
Previous studies
Yamamoto (1983)
Theoretical limits in special cases
Shinohara, Yagi (2022)
General theoretical limits
Utility：Expected distortion
New results
(Section 3)
Shinohara, Yagi
General theoretical limits with limited leakage for the encoder
Utility：Expected distortion
(Section 4)
Shinohara, Yagi
General theoretical limits with limited leakage for the encoder
Utility：Excess-distortion probability
(Section 5)
Shinohara, Yagi (2022)
Strong converse theorem between utility and privacy
Goal：
The second-order rate analysis

Figure 2. Road map for second-order rate analysis [1,3,8].

Coding systems in the utility–privacy problem are extended to the ones with the encoder's side information [2] and with the decoder's side information [25]. In [14], a related coding problem has been investigated, where both the encoder and the decoder can access a uniform secret key and the decoder can also access side information. Utility–privacy trade-off schemes are applied, for example, to the Internet of Energy [23] and to a system with informational self-determination [24].

A closely related study to this paper was given by Basciftci et al. [13], in which several release mechanisms of encoded information from the database were discussed. In particular, utility–privacy trade-offs (without the coding rate) were compared when the encoded information was (i) both private and public information, (ii) only public information, and (iv) only private information (see also the three cases described in Section 1.1). A sufficient condition under which the utility–privacy trade-offs coincide for cases (i) and (ii) was given.

1.4. Organization

This paper is organized as follows: In Section 2, we begin by introducing the notation and system model that are used in this paper. In Section 3, we give the first-order rate analysis among the coding rate, utility, privacy for the decoder, and privacy for the encoder in which utility is measured by the expected distortion. In Section 4, we tackle the first-order rate analysis among the coding rate, utility, privacy for the decoder, and privacy for the encoder in which utility is measured by the excess-distortion probability. Section 5 focuses on the strong converse theorem for utility–privacy trade-offs in which utility is measured by the excess-distortion probability. In Section 6, we discuss the significance of the encoded information with limited leakage for the encoder. Finally, in Section 7, the conclusion and future work are stated.

2. Notation and System Model

2.1. Information Source

Database d is described by a $K \times n$ matrix whose rows represent K attributes and columns represent n entries of data. Let $\mathcal{K} = \{1, 2, \ldots, K\}$ be the set of indexes of K attributes. The random variable for the lth attribute is denoted by X_l, which takes a value in a finite alphabet $\mathcal{X}_l$. For any subset $\mathcal{B} \subseteq \mathcal{K}$, the tuple of random variables $(X_l)_{l \in \mathcal{B}}$ is abbreviated as $X_{\mathcal{B}}$. Similarly, the Cartesian product of alphabets $\prod_{l \in \mathcal{B}} \mathcal{X}_l$ is abbreviated as $\mathcal{X}_{\mathcal{B}}$.

The K attributes can be divided into two groups; one may be open to the public and the other should be kept secret from a third party. Then, the set $\mathcal{K}$ is divided into disjoint sets $\mathcal{R}$ and $\mathcal{H}$. That is,

$$\mathcal{K} = \mathcal{R} \cup \mathcal{H}, \quad \mathcal{R} \cap \mathcal{H} = \emptyset, \quad \mathcal{X}_{\mathcal{K}} = \mathcal{X}_{\mathcal{R}} \times \mathcal{X}_{\mathcal{H}}, \tag{1}$$

where $\mathcal{X}_{\mathcal{R}}$ is the set of values that public (revealed) source symbols $X_{\mathcal{R}}$ take and $\mathcal{X}_{\mathcal{H}}$ is the set of values that private (hidden) source symbols $X_{\mathcal{H}}$ take.

We assume that the source sequence $X_{\mathcal{K}}^n = (X_{\mathcal{K},1}, X_{\mathcal{K},2}, \ldots, X_{\mathcal{K},n})$ is generated from a stationary and memoryless source $p_{X_{\mathcal{K}}}$. That is,

$$P_{X_{\mathcal{K}}^n}(x_{\mathcal{K}}^n) = \Pr\{X_{\mathcal{K}}^n = x_{\mathcal{K}}^n\} = \prod_{i=1}^{n} P_{X_{\mathcal{K}}}(x_{\mathcal{K},i}), \tag{2}$$

where $x_{\mathcal{K}}^n = (x_{\mathcal{K},1}, \ldots, x_{\mathcal{K},n}) \in \mathcal{X}_{\mathcal{K}}^n$. Taking the partition of attributes in (1) into account, the source sequence $X_{\mathcal{K}}^n$ is described as

$$X_{\mathcal{K}}^n = (X_{\mathcal{R}}^n, X_{\mathcal{H}}^n), \tag{3}$$

where

$$X_{\mathcal{R}}^n = (X_{\mathcal{R},1}, X_{\mathcal{R},2}, \ldots, X_{\mathcal{R},n}) \in \mathcal{X}_{\mathcal{R}}^n, \tag{4}$$

$$X_{\mathcal{H}}^n = (X_{\mathcal{H},1}, X_{\mathcal{H},2}, \ldots, X_{\mathcal{H},n}) \in \mathcal{X}_{\mathcal{H}}^n \tag{5}$$

are referred to as the revealed source sequence and the hidden source sequence, respectively. In the addressed coding system introduced in [22], the revealed symbols and a part of the hidden symbols are input to the encoder, and thus the encoded alphabet $\mathcal{E}$ satisfies $\mathcal{R} \subseteq \mathcal{E} \subseteq \mathcal{K}$. Similar to (3), $X_{\mathcal{K}}^n$ is sometimes described as

$$X_{\mathcal{K}}^n = (X_{\mathcal{E}}^n, X_{\mathcal{E}^{\mathrm{c}}}^n), \tag{6}$$

where $X_{\mathcal{E}}^n$ is the source sequence observed by the encoder and $\mathcal{E}^{\mathrm{c}} = \mathcal{K} \setminus \mathcal{E}$.

2.2. Encoder and Decoder

The coding system consists of encoder f_n and decoder g_n as in Figure 1. When the source sequence $X_{\mathcal{K}}^n = (X_{\mathcal{E}}^n, X_{\mathcal{E}^{\mathrm{c}}}^n)$ is generated from the stationary and memoryless source $p_{X_{\mathcal{K}}}$, the codeword $J_n = f_n(X_{\mathcal{E}}^n)$ is generated by the encoder

$$f_n\colon\ \mathcal{X}_{\mathcal{E}}^n \to \{1, 2, \ldots, M_n\} \tag{7}$$

and the reproduced sequence $\hat{X}_{\mathcal{R}}^n = g_n(J_n)$ is produced by decoder

$$g_n\colon\ \{1, 2, \ldots, M_n\} \to \hat{\mathcal{X}}_{\mathcal{R}}^n, \tag{8}$$

where M_n denotes the number of codewords.

3. First-Order Rate Analysis with Expected Distortion

3.1. Performance Measures

In this section, we mention the measure of the coding rate, utility, privacy for the decoder, and privacy for the encoder. Hereafter, let a pair of the encoder and decoder (f_n, g_n) be fixed.

For a given M_n, the coding rate is defined as

$$r_n := \frac{1}{n} \log M_n. \tag{9}$$

Let $d :\ \mathcal{X}_{\mathcal{R}} \times \hat{\mathcal{X}}_{\mathcal{R}} \to [0, \infty)$ be a distortion function between $x_{\mathcal{R}} \in \mathcal{X}_{\mathcal{R}}$ and $\hat{x}_{\mathcal{R}} \in \hat{\mathcal{X}}_{\mathcal{R}}$. The distortion between sequences $x_{\mathcal{R}}^n \in \mathcal{X}_{\mathcal{R}}^n$ and $\hat{x}_{\mathcal{R}}^n \in \hat{\mathcal{X}}_{\mathcal{R}}^n$ is defined as

$$d(x_{\mathcal{R}}^n, \hat{x}_{\mathcal{R}}^n) := \sum_{i=1}^{n} d(x_{\mathcal{R},i}, \hat{x}_{\mathcal{R},i}). \tag{10}$$

Then, the measure of utility is defined as

$$u_n := \mathbb{E}\left[\frac{1}{n}d(X^n_{\mathcal{R}}, \hat{X}^n_{\mathcal{R}})\right], \tag{11}$$

where $\mathbb{E}$ represents the expectation by the joint distribution of $(X^n_{\mathcal{R}}, \hat{X}^n_{\mathcal{R}})$.

In this system, the privacy of the hidden source sequence $X^n_{\mathcal{H}}$ should be protected when the codeword J_n is observed by decoder g_n. The measure of privacy for the decoder is defined as

$$l_n =: \frac{1}{n}I(X^n_{\mathcal{H}}; J_n), \tag{12}$$

where $I(X^n_{\mathcal{H}}; J_n)$ is the mutual information between $X^n_{\mathcal{H}}$ and J_n.

The privacy of the hidden source sequence $X^n_{\mathcal{H}}$ should be protected when the encoded information $X_{\mathcal{E}}$ is observed by encoder f_n. The measurement of privacy for the encoder is defined as

$$e_n := \frac{1}{n}I(X^n_{\mathcal{H}}; X^n_{\mathcal{E}}), \tag{13}$$

where $I(X^n_{\mathcal{H}}; X^n_{\mathcal{E}})$ is the mutual information between $X^n_{\mathcal{H}}$ and $X^n_{\mathcal{E}}$.

3.2. Achievable Region and Theorem

We define the achievable region for the first-order rate analysis with the expected distortion and state the obtained results.

Definition 1. *A tuple (R, D, L, E) is said to be ϵ-**achievable** (with respect to the expected distortion measure) if, for any given $\epsilon > 0$, there exists a sequence of codes (f_n, g_n) satisfying*

$$r_n \le R + \epsilon, \tag{14}$$
$$u_n \le D + \epsilon, \tag{15}$$
$$l_n \le L + \epsilon, \tag{16}$$
$$e_n \le E + \epsilon \tag{17}$$

for all sufficiently large n.

The technical meanings of each constraint in Definition 1 can be interpreted as follows: Equation (14) evaluates how much the source sequence is compressed, so this rate should be decreased. Equation (15) is the constraint corresponding to distortion being less than $D + \epsilon$. The smaller the distortion is, the better the utility is, so this condition should also be decreased. Equation (16) constrains the amount of leaked private information to the decoder. Since private information should be kept secret for the receiver, this quantity should be decreased as well. Equation (17) constrains the amount of private information leaked to the encoder. For the same reason as (16), this quantity should also be decreased.

Remark 1. *The minimum coding rate R for a fixed D corresponds to the rate-distortion function (Section 10 in [27]). Thus, in the proof of achievability, we evaluate the coding rate and the distortion with the argument in rate-distortion theory. This view is also important to correctly understand the numerical results in Section 6.1.*

Definition 2. *The closure of the set of ϵ-achievable tuples (R, D, L, E) is referred to as the ϵ-**achievable region** and is denoted by $\mathcal{C}_{\mathcal{E}}(\epsilon|P_{X_{\mathcal{K}}})$ and defines*

$$\mathcal{C}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) := \bigcap_{0<\epsilon<1} \mathcal{C}_{\mathcal{E}}(\epsilon|P_{X_{\mathcal{K}}}). \tag{18}$$

To characterize the achievable region, we define the following informational region.

Definition 3. *For any $\mathcal{E}$ such that $\mathcal{R} \subseteq \mathcal{E} \subseteq \mathcal{K}$, $\mathcal{S}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ is defined as*

$$\begin{aligned}\mathcal{S}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) = \{(R,D,L,E):\ & R \geq I(X_{\mathcal{E}};\hat{X}_{\mathcal{R}}),\\ & D \geq \mathbb{E}[d(X_{\mathcal{R}},\hat{X}_{\mathcal{R}})],\\ & L \geq I(X_{\mathcal{H}};\hat{X}_{\mathcal{R}}),\\ & E \geq I(X_{\mathcal{H}};X_{\mathcal{E}})\\ & \text{for some } P_{X_{\mathcal{E}},X_{\mathcal{E}^c}} \cdot P_{\hat{X}_{\mathcal{R}}|X_{\mathcal{E}}}\}.\end{aligned} \tag{19}$$

We establish the next theorem. For the proof of this theorem, please refer to Sections 3.3–3.5.

Theorem 1. *For any $\mathcal{E}$ such that $\mathcal{R} \subseteq \mathcal{E} \subseteq \mathcal{K}$, the achievable region of the coding system is given by*

$$\mathcal{C}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) = \mathcal{S}_{\mathcal{E}}(P_{X_{\mathcal{K}}}). \tag{20}$$

To clarify the relationship with the conventional result of Shinohara and Yagi [3], we mention the achievable region among the coding rate, utility, and privacy, which is derived by projecting the result of Theorem 1 onto the R-D-L hyperplane.

Definition 4. *For any $\mathcal{E}$ such that $\mathcal{R} \subseteq \mathcal{E} \subseteq \mathcal{K}$, we define*

$$\mathcal{C}_{\mathcal{E}}^{RDL}(\epsilon|P_{X_{\mathcal{K}}}) := \{(R,D,L):\ (R,D,L,E) \in \mathcal{C}_{\mathcal{E}}(\epsilon|P_{X_{\mathcal{K}}})\} \tag{21}$$

and

$$\mathcal{C}_{\mathcal{E}}^{RDL}(P_{X_{\mathcal{K}}}) := \bigcap_{0<\epsilon<1} \mathcal{C}_{\mathcal{E}}^{RDL}(\epsilon|P_{X_{\mathcal{K}}}). \tag{22}$$

Definition 5. *For any $\mathcal{E}$ such that $\mathcal{R} \subseteq \mathcal{E} \subseteq \mathcal{K}$, we define*

$$\begin{aligned}\mathcal{S}_{\mathcal{E}}^{RDL}(P_{X_{\mathcal{K}}}) = \{(R,D,L):\ & R \geq I(X_{\mathcal{E}};\hat{X}_{\mathcal{R}}),\\ & D \geq \mathbb{E}[d(X_{\mathcal{R}},\hat{X}_{\mathcal{R}})],\\ & L \geq I(X_{\mathcal{H}};\hat{X}_{\mathcal{R}})\\ & \text{for some } P_{X_{\mathcal{E}},X_{\mathcal{E}^c}} \cdot P_{\hat{X}_{\mathcal{R}}|X_{\mathcal{E}}}\}.\end{aligned} \tag{23}$$

Corollary 1. *For any $\mathcal{E}$ such that $\mathcal{R} \subseteq \mathcal{E} \subseteq \mathcal{K}$, the region $\mathcal{C}_{\mathcal{E}}^{RDL}(P_{X_{\mathcal{K}}})$ is given by*

$$\mathcal{C}_{\mathcal{E}}^{RDL}(P_{X_{\mathcal{K}}}) = \mathcal{S}_{\mathcal{E}}^{RDL}(P_{X_{\mathcal{K}}}). \tag{24}$$

Remark 2. *Corollary 1 suggests that the conventional result [3] can be obtained from $\mathcal{C}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$.*

Remark 3. *The derived characterization in (24) reduces to the characterization given in [1] when the encoded attribute $\mathcal{E}$ is either $\mathcal{K}$ or $\mathcal{R}$. Thus, (24) gives its generalization for $\mathcal{R} \subseteq \mathcal{E} \subseteq \mathcal{K}$.*

Examples to illustrate this result are shown in Section 6.1.

3.3. Proof Preliminaries for First-Order Rate Analysis

For preliminaries for coding theorems by the first-order rate analysis, we define strongly typical sequences that are necessary for the proof and show some properties. These proof preliminaries are also used in Section 4.

Definition 6 (Definition 2.1, [28])**.** *The type of a sequence $x^n \in \mathcal{X}^n$ of length n is the distribution P_{x^n} on $\mathcal{X}$ defined by*

$$P_{x^n}(a) := \frac{1}{n} N(a|x^n), \tag{25}$$

where $N(a|x^n)$ represents the number of occurrences of symbol $a \in \mathcal{X}$ in x^n. Likewise, the joint type of $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$ is the distribution $P_{x^n y^n}$ on $\mathcal{X} \times \mathcal{Y}$ defined by

$$P_{x^n y^n} := \frac{1}{n} N(a, b|x^n, y^n), \tag{26}$$

where $N(a, b|x^n, y^n)$ represents the number of the occurrences of $(a, b) \in \mathcal{X} \times \mathcal{Y}$ in the pair of sequences (x^n, y^n).

Definition 7 ((Conditional Type), [28], Definition 2.2)**.** *We define the conditional type of y^n given x^n as a stochastic matrix $V\colon\ \mathcal{X} \to \mathcal{Y}$ satisfying*

$$N(a, b|x^n, y^n) = N(a|x^n)V(b|a). \tag{27}$$

In particular, the conditional type of y^n given x^n is uniquely determined and given by

$$V(b|a) = \frac{N(a, b|x^n, y^n)}{N(a|x^n)} \tag{28}$$

if $N(a|x^n) > 0$ for any $a \in \mathcal{X}$.

Definition 8 ((Strongly Typical Sequences), [29], Definition 1.2.8)**.** *For any distribution P on $\mathcal{X}$, a sequence $x^n \in \mathcal{X}^n$ is said to be P-typical with constant $\delta > 0$ if*

$$\left| \frac{1}{n} N(a|x^n) - P(a) \right| \le \delta \quad \text{for every } a \in \mathcal{X} \tag{29}$$

and, in addition, no $a \in \mathcal{X}$ with $P(a) = 0$ occurs in x^n. The set of such sequences is denoted by $T^n_\delta(P)$. If X is a random variable with values in $\mathcal{X}$, we also refer to P-typical sequences as X-typical sequences and write $T^n_\delta(X)$.

Definition 9 ((Conditional Strongly Typical Sequences), [29], Definition 1.2.9)**.** *For a stochastic matrix $W\colon\ \mathcal{X} \to \mathcal{Y}$, a sequence $y^n \in \mathcal{Y}^n$ is said to be W-typical given $x^n \in \mathcal{X}^n$ with constant $\delta > 0$ if*

$$\left| \frac{1}{n} N(a, b|x^n, y^n) - \frac{1}{n} N(a|x^n) W(b|a) \right| \le \delta \tag{30}$$
$$\text{for every } a \in \mathcal{X}, b \in \mathcal{Y},$$

and, in addition, $N(a, b|x^n, y^n) = 0$ whenever $W(b|a) = 0$. The set of such sequences y^n is denoted by $T^n_\delta(W|x^n)$. Further, if X and Y are random variables with values in $\mathcal{X}$ and $\mathcal{Y}$, respectively, and $P_{Y|X} = W$, then they are also said to be $Y|X$-typical and written as $T^n_\delta(Y|X|x^n)$.

Hereafter, the set of conditional strongly typical sequences $T^n_\delta(Y|X|x^n)$ is abbreviated as $T^n_\delta(Y|x^n)$.

We state some lemmas that are used in this proof.

Lemma 1 ([29], Lemma 1.2.13)**.** *For any positive sequences $\{\delta_n\}_{n=1}^\infty$ and $\{\delta'_n\}_{n=1}^\infty$ such that $\delta_n \to 0$ and $\delta' \to 0$ as $n \to 0$, there exists a sequence $\epsilon_n = \epsilon_n(|\mathcal{X}, \mathcal{Y}|, \delta_n, \delta'_n) \to 0$ $(n \to \infty)$ such that for every distribution P on $\mathcal{X}$ and stochastic matrix $W\colon\ \mathcal{X} \to \mathcal{Y}$,*

$$\left|\frac{1}{n}\log|T^n_{\delta_n}(P)| - H(P)\right| \le \epsilon_n, \tag{31}$$

$$\left|\frac{1}{n}\log|T^n_{\delta'_n}(W|x^n)| - H(W|P)\right| \le \epsilon_n. \tag{32}$$

Lemma 2 ([29], Lemma 1.2.7)**.** *Let the variational distance between two distributions P and Q on $\mathcal{X}$ be defined as*

$$d_{\mathrm{v}}(P,Q) := \sum_{x\in\mathcal{X}} |P(x) - Q(x)|. \tag{33}$$

If $d_{\mathrm{v}}(P,Q) < \frac{1}{2}$, then

$$|H(P) - H(Q)| \le -d_{\mathrm{v}}(P,Q)\cdot\log\frac{d_{\mathrm{v}}(P,Q)}{|\mathcal{X}|}. \tag{34}$$

Lemma 3 ([29], Lemma 1.2.10)**.** *If $x^n \in T^n_\delta(X)$ and $y^n \in T^n_{\delta'}(Y|x^n)$, then $(x^n, y^n) \in T^n_{\delta+\delta'}(X,Y)$ and, consequently, $y^n \in T_{\delta''}(Y)$ for $\delta'' := (\delta+\delta')\cdot|\mathcal{X}|$.*

Lemma 4. *If $(x^n, y^n) \in T^n_\delta(X,Y)$, then $x^n \in T^n_{\delta_1}(X)$ and, consequently, $y^n \in T^n_{\delta_2}(Y|x^n)$ for $\delta_1 := |\mathcal{Y}|\cdot\delta$ and $\delta_2 := (|\mathcal{Y}|+1)\cdot\delta$.*

Lemma 5. *If $y^n \in T^n_\delta(Y)$ and $(x^n, y^n) \notin T^n_{2\delta}(X,Y)$, then $x^n \notin T^n_\delta(X|y^n)$.*

Lemma 6 ([29], Lemma 1.2.12 and Remark)**.** *For arbitrarily fixed $\delta > 0$ and every distribution P on $\mathcal{X}$ and stochastic matrix $W\colon\ \mathcal{X} \to \mathcal{Y}$*

$$\Pr\{X^n \in T^n_\delta(P)\} \ge 1 - 2|\mathcal{X}|\mathrm{e}^{-2\delta^2 n}, \tag{35}$$

$$\Pr\{Y^n \in T^n_\delta(W|x^n)|X^n = x^n\} \ge 1 - 2|\mathcal{X}|\cdot|\mathcal{Y}|\mathrm{e}^{-2\delta^2 n} \quad \text{for every } x^n \in \mathcal{X}^n. \tag{36}$$

3.4. Proof of Converse Part

In this part, we shall prove $\mathcal{C}_\mathcal{E}(P_{X_\mathcal{K}}) \subseteq \mathcal{S}_\mathcal{E}(P_{X_\mathcal{K}})$.

Let a tuple $(R, D, L, E) \in \mathcal{C}_\mathcal{E}(P_{X_\mathcal{K}})$ be arbitrarily fixed. Then, there exists an $(n, 2^{n(R+\epsilon)}, D+\epsilon, L+\epsilon, E+\epsilon)$ code that satisfies (14)–(17). Let Q be a uniform random variable over $\{1, 2, \ldots, n\}$ and let $p_i(x_{\mathcal{E},i}, x_{\mathcal{E}^c,i}, \hat{x}_{\mathcal{R},i})$ be the conditional distribution given $Q = i$. Evaluating the inequalities for R, we obtain

$$\begin{aligned}
R+\epsilon &\overset{\text{(a)}}{\geq} \frac{1}{n}\log M_n \\
&\overset{\text{(b)}}{\geq} \frac{1}{n}H(J_n) \\
&\geq \frac{1}{n}I(J_n;X_{\mathcal{E}}^n) \\
&\overset{\text{(c)}}{=} \frac{1}{n}\{H(X_{\mathcal{E}}^n)-H(X_{\mathcal{E}}^n|J_n,\hat{X}_{\mathcal{R}}^n)\} \\
&\overset{\text{(d)}}{=} \frac{1}{n}\sum_{i=1}^{n}H(X_{\mathcal{E},i})-\frac{1}{n}\sum_{i=1}^{n}H(X_{\mathcal{E},i}|X_{\mathcal{E}}^{i-1},J_n,\hat{X}_{\mathcal{R}}^n) \\
&\overset{\text{(e)}}{\geq} \frac{1}{n}\sum_{i=1}^{n}H(X_{\mathcal{E},i})-\frac{1}{n}\sum_{i=1}^{n}H(X_{\mathcal{E},i}|\hat{X}_{\mathcal{R},i}) \\
&\overset{\text{(f)}}{=} \sum_{i=1}^{n}\Pr\{Q=i\}H(X_{\mathcal{E},i}|Q=i) \\
&\quad -\sum_{i=1}^{n}\Pr\{Q=i\}H(X_{\mathcal{E},i}|\hat{X}_{\mathcal{R},i},Q=i) \\
&= H(X_{\mathcal{E},Q}|Q)-H(X_{\mathcal{E},Q}|\hat{X}_{\mathcal{R},Q},Q) \\
&\overset{\text{(g)}}{=} H(X_{\mathcal{E}})-H(X_{\mathcal{E},Q}|\hat{X}_{\mathcal{R},Q},Q) \\
&\overset{\text{(h)}}{\geq} H(X_{\mathcal{E}})-H(X_{\mathcal{E}}|\hat{X}_{\mathcal{R}}) \\
&= I(X_{\mathcal{E}};\hat{X}_{\mathcal{R}}),
\end{aligned} \tag{37}$$

(a) follows from (14),
(b) follows because $H(J_n)\leq\log|J_n|=\log M_n$,
(c) is due to the fact that $\hat{X}_{\mathcal{R}}^n=g(J_n)$,
(d) follows because each $X_{\mathcal{K},i}$ is independent and $\hat{X}_{\mathcal{R}}^n$ is a function of J_n,
(e) follows because conditioning reduces entropy,
(f) is due to the definition of Q,
(g) follows because $X_{\mathcal{E}}\perp Q$, and
(h) follows because conditioning reduces entropy, where $(X_{\mathcal{E}},\hat{X}_{\mathcal{R}})\sim\sum_{i=1}^{n}\Pr\{Q=i\}p_i(x_{\mathcal{E},i},\hat{x}_{\mathcal{R},i})=p(x_{\mathcal{E}},\hat{x}_{\mathcal{R}})$.

Similarly, evaluating D, L, and E, respectively, we obtain

$$\begin{aligned}
D+\epsilon &\overset{\text{(i)}}{\geq} \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} d(X_{\mathcal{R},i},\hat{X}_{\mathcal{R},i})\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[d(X_{\mathcal{R},i},\hat{X}_{\mathcal{R},i})] \\
&\overset{\text{(j)}}{=} \mathbb{E}_Q[\mathbb{E}[d(X_{\mathcal{R},i},\hat{X}_{\mathcal{R},i})|Q]] \\
&\overset{\text{(k)}}{=} \mathbb{E}[d(X_{\mathcal{R}},\hat{X}_{\mathcal{R}})], \qquad (38)\\
L+\epsilon &\overset{\text{(l)}}{\geq} \frac{1}{n}I(X^n_{\mathcal{H}};J_n) \\
&= \frac{1}{n}H(X^n_{\mathcal{H}}) - \frac{1}{n}H(X^n_{\mathcal{H}}|J_n) \\
&\overset{\text{(m)}}{=} H(X_{\mathcal{H}}) - \frac{1}{n}\sum_{i=1}^{n} H(X_{\mathcal{H},i}|X^{i-1}_{\mathcal{H}},J_n) \\
&\overset{\text{(n)}}{=} H(X_{\mathcal{H}}) - \frac{1}{n}\sum_{i=1}^{n} H(X_{\mathcal{H},i}|X^{i-1}_{\mathcal{H}},J_n,\hat{X}_{\mathcal{R},i}) \\
&\overset{\text{(o)}}{\geq} H(X_{\mathcal{H}}) - \frac{1}{n}\sum_{i=1}^{n} H(X_{\mathcal{H},i}|\hat{X}_{\mathcal{R},i}) \\
&\overset{\text{(p)}}{=} H(X_{\mathcal{H}}) - \sum_{i=1}^{n}\Pr\{Q=i\}H(X_{\mathcal{H},i}|\hat{X}_{\mathcal{R},i},Q=i) \\
&= H(X_{\mathcal{H}}) - H(X_{\mathcal{H},Q}|\hat{X}_{\mathcal{R},Q},Q) \\
&\overset{\text{(q)}}{\geq} H(X_{\mathcal{H}}) - H(X_{\mathcal{H}}|\hat{X}_{\mathcal{R}}) \\
&= I(X_{\mathcal{H}};\hat{X}_{\mathcal{R}}), \qquad (39)\\
E+\epsilon &\geq \frac{1}{n}I(X^n_{\mathcal{H}};X^n_{\mathcal{E}}) \\
&\overset{\text{(r)}}{=} \frac{1}{n}\sum_{i=1}^{n} I(X_{\mathcal{H},i};X^n_{\mathcal{E}}|X^{i-1}_{\mathcal{H}}) \\
&\overset{\text{(s)}}{=} \frac{1}{n}\sum_{i=1}^{n} I(X_{\mathcal{H},i};X_{\mathcal{E},i}) \\
&\overset{\text{(t)}}{=} I(X_{\mathcal{H}};X_{\mathcal{E}}), \qquad (40)
\end{aligned}$$

where

(i) is due to (15),
(j) is derived from the definition of Q,
(k) follows because $(X_{\mathcal{R}},\hat{X}_{\mathcal{R}}) \sim \sum_{i=1}^{n}\Pr\{Q=i\}p_i(x_{\mathcal{R},i},\hat{x}_{\mathcal{R},i}) = p(x_{\mathcal{R}},\hat{x}_{\mathcal{R}})$,
(l) is due to (16),
(m) follows because *i.i.d.* $P_{X^n_{\mathcal{K}}}$,
(n) follows because $\hat{X}^n_{\mathcal{R}} = g(J_n)$,
(o) follows from the fact that conditioning reduces entropy,
(p) is derived from the definition of Q, and
(q) follows because conditioning reduces entropy, where $(X_{\mathcal{H}},\hat{X}_{\mathcal{R}}) \sim \sum_{i=1}^{n}\Pr\{Q=i\}p_i(x_{\mathcal{H},i},\hat{x}_{\mathcal{R},i}) = p(x_{\mathcal{H}},\hat{x}_{\mathcal{R}})$,
(r) is due to chain rule for mutual information,
(s), (t) follow because *i.i.d.* $P_{X^n_{\mathcal{K}}}$.

It is readily shown that the Markov chain $X_{\mathcal{E}^c}–X_{\mathcal{E}}–\hat{X}_{\mathcal{R}}$ holds (cf. Appendix A). We complete the proof of the converse part.

3.5. Proof of Direct Part

In this part, we provide a sketch of the proof of $\mathcal{S}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) \subseteq \mathcal{C}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$.

Under an arbitrarily fixed distribution $P_{X_{\mathcal{E}},X_{\mathcal{E}^c}} \cdot P_{\hat{X}_{\mathcal{R}}|X_{\mathcal{E}}}$, any tuple $(R,D,L,E) \in \mathcal{S}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ is chosen such that

$$R > I(X_{\mathcal{E}};\hat{X}_{\mathcal{R}}), \tag{41}$$
$$D > \mathbb{E}[d(X_{\mathcal{R}},\hat{X}_{\mathcal{R}})], \tag{42}$$
$$L > I(X_{\mathcal{H}};\hat{X}_{\mathcal{R}}), \tag{43}$$
$$E > I(X_{\mathcal{H}};X_{\mathcal{E}}). \tag{44}$$

From (42) and (43), we can choose a sufficiently small $\epsilon > 0$ such that

$$D > \mathbb{E}[d(X_{\mathcal{R}},\hat{X}_{\mathcal{R}})] + \epsilon, \tag{45}$$
$$L > I(X_{\mathcal{H}};\hat{X}_{\mathcal{R}}) + \epsilon. \tag{46}$$

In addition, with this ϵ, some constant $0 < \tau < \frac{1}{2}$ is fixed such that

$$\tau(\log|\mathcal{X}_{\mathcal{H}}| + 5) + 4\tau\log\frac{|\mathcal{X}_{\mathcal{H}}| \cdot 2^R}{2\tau} < \epsilon. \tag{47}$$

We can also choose positive numbers $\delta(:= \delta(n))$ such that

$$(\delta(n) + \delta_1(n))|\mathcal{X}_{\mathcal{R}}|\cdot|\hat{\mathcal{X}}_{\mathcal{R}}|D_{\max} + \tau < \epsilon, \tag{48}$$
$$2\delta^2(n) \le R - I(X_{\mathcal{E}};\hat{X}_{\mathcal{R}}) - \frac{1}{n} - \tau, \tag{49}$$
$$\delta(n) \to 0, \tag{50}$$
$$\sqrt{n}\cdot\delta(n) \to \infty \tag{51}$$

as $n \to \infty$, where $\delta_1 := (|\mathcal{X}_{\mathcal{E}}| - |\mathcal{X}_{\mathcal{R}}|)\cdot\delta$ and $D_{\max} := \max_{a\in\mathcal{X}_{\mathcal{R}},b\in\hat{\mathcal{X}}_{\mathcal{R}}} d(a,b)$. Let $\delta(n) = \frac{c}{\sqrt{n}}\log n$ where c is a constant, and obviously (50) and (51) are satisfied.

Generation of codebook: Randomly generate $\hat{x}^n_{\mathcal{R}}(j)$ from the strongly typical sequences $T^n_\delta(\hat{X}_{\mathcal{R}})$ for $j = 1,2,\ldots,M_n := 2^{nR}$. Reveal the codebook $\mathcal{C} = \{\hat{x}^n_{\mathcal{R}}(1),\ldots,\hat{x}^n_{\mathcal{R}}(M_n)\}$ to the encoder and decoder.

Encoding: If a sequence $x^n_{\mathcal{E}} \in \mathcal{X}^n_{\mathcal{E}}$ satisfies $x^n_{\mathcal{K}} = (x^n_{\mathcal{E}}, x^n_{\mathcal{E}^c})$ with some $x^n_{\mathcal{E}^c} \in \mathcal{X}^n_{\mathcal{E}^c}$, we write $x^n_{\mathcal{E}} \prec x^n_{\mathcal{K}}$. Given $x^n_{\mathcal{K}}$, the encoder finds j such that $x^n_{\mathcal{E}} \in T^n_\delta(X_{\mathcal{E}}|\hat{x}_{\mathcal{R}}(j))$ and sets $f_n(x^n_{\mathcal{E}}) = j$ where $T^n_\delta(X_{\mathcal{E}}|\hat{x}_{\mathcal{R}}(j))$ is the conditional strongly typical sequences. If there exist multiple such j, $f_n(x^n_{\mathcal{E}})$ is set as the minimum one. If there are no such j, then $f_n(x^n_{\mathcal{E}}) = M_n$.

Decoding: When j is observed, the decoder sets the reproduced sequence as $\hat{X}^n_{\mathcal{R}} = \hat{x}^n_{\mathcal{R}}(j)$.

Evaluation: We define $\mathcal{A}(j)$, $\mathcal{B}(j)$, and $\tilde{\mathcal{A}}(j)$ as

$$\mathcal{A}(j) := \{x^n_{\mathcal{E}}\colon\ f_n(x^n_{\mathcal{E}}) = j\}, \tag{52}$$
$$\mathcal{B}(j) := \{x^n_{\mathcal{K}}\colon\ x^n_{\mathcal{E}} \prec x^n_{\mathcal{K}}, f_n(x^n_{\mathcal{E}}) = j\}, \tag{53}$$
$$\tilde{\mathcal{A}}(j) := \begin{cases} \{x^n_{\mathcal{K}}\colon\ x^n_{\mathcal{E}} \prec x^n_{\mathcal{K}}, f_n(x^n_{\mathcal{E}}) = j, x^n_{\mathcal{K}} \in T^n_{2\delta}(X_{\mathcal{K}}|\hat{x}^n_{\mathcal{R}}(j))\} \\ \qquad\qquad\qquad\qquad (j = 1,2,\ldots,M_n - 1) \\ \{x^n_{\mathcal{K}}\colon\ x^n_{\mathcal{K}} \in \mathcal{X}^n_{\mathcal{K}} \setminus \bigcup_{j=1}^{M_n-1}\tilde{\mathcal{A}}(j)\} \quad (j = M_n). \end{cases} \tag{54}$$

It is easily verified that $\mathcal{A}(j)$ for $j = 1,2,\ldots,M_n$ (also, $\mathcal{B}(j)$ and $\tilde{\mathcal{A}}(j)$) is disjoint. From the definitions of J_n, $\mathcal{A}(j)$, and $\mathcal{B}(j)$,

$$\Pr\{J_n = j\} = \Pr\{X^n_{\mathcal{E}} \in \mathcal{A}(j)\} = \Pr\{X^n_{\mathcal{K}} \in \mathcal{B}(j)\} \\ \text{for } j = 1,2,\ldots,M_n. \tag{55}$$

For sufficiently large n, we can prove (cf. Appendix B)

$$|\Pr\{X_{\mathcal{K}}^n \in \mathcal{B}(j)\} - \Pr\{X_{\mathcal{K}}^n \in \tilde{\mathcal{A}}(j)\}| \le 2|\mathcal{X}_{\mathcal{K}}| \cdot |\hat{\mathcal{X}}_{\mathcal{R}}| \mathrm{e}^{-2\delta^2 n} \quad \text{for } j = 1, 2, \ldots, M_n - 1. \tag{56}$$

For sufficiently large n, we can show that there exists a code (f_n, g_n) such that (cf. Appendix C)

$$\begin{aligned}
& r_n \le R, && (57)\\
& u_n \le \mathbb{E}[d(X_{\mathcal{R}}, \hat{X}_{\mathcal{R}})] + (\delta + \delta_1)|\mathcal{X}_{\mathcal{R}}| \cdot |\hat{\mathcal{X}}_{\mathcal{R}}| D_{\max} + \tau, && (58)\\
& e_n \le I(X_{\mathcal{H}}; X_{\mathcal{E}}), && (59)\\
& \Pr\left\{ X_{\mathcal{E}}^n \notin \bigcup_{j=1}^{M_n - 1} \mathcal{A}(j) \right\} \le (2|\mathcal{X}_{\mathcal{E}}| + 1)\mathrm{e}^{-2\delta^2 n}, && (60)\\
& \Pr\left\{ X_{\mathcal{K}}^n \notin \bigcup_{j=1}^{M_n - 1} \tilde{\mathcal{A}}(j) \right\} \le \tau, && (61)\\
& |\tilde{\mathcal{A}}(j)| \ge 2^{n\{H(X_{\mathcal{K}}|\hat{X}_{\mathcal{R}}) - \tau\}}. && (62)
\end{aligned}$$

For this code (f_n, g_n), we evaluate the privacy leakage against the decoder as

$$\begin{aligned}
l_n &:= \frac{1}{n} I(X_{\mathcal{H}}^n; J_n)\\
&= \frac{1}{n} H(X_{\mathcal{H}}^n) - \frac{1}{n} H(X_{\mathcal{H}}^n | J_n)\\
&\overset{\text{(a)}}{=} H(X_{\mathcal{H}}) - \frac{1}{n} \sum_{j=1}^{M_n} H(X_{\mathcal{H}}^n | X_{\mathcal{K}}^n \in \mathcal{B}(j)) \Pr\{X_{\mathcal{K}}^n \in \mathcal{B}(j)\}\\
&\overset{\text{(b)}}{\le} H(X_{\mathcal{H}}) - \frac{1}{n} \sum_{j=1}^{M_n} H(X_{\mathcal{H}}^n | X_{\mathcal{K}}^n \in \tilde{\mathcal{A}}(j)) \Pr\{X_{\mathcal{K}}^n \in \tilde{\mathcal{A}}(j)\}\\
&\quad + 4\tau \log \frac{|\mathcal{X}_{\mathcal{H}}| \cdot 2^R}{2\tau} && (63)\\
&\overset{\text{(c)}}{\le} H(X_{\mathcal{H}}) - \frac{1}{n} \sum_{j=1}^{M_n - 1} H(X_{\mathcal{H}}^n | X_{\mathcal{K}}^n \in \tilde{\mathcal{A}}(j)) \Pr\{X_{\mathcal{K}}^n \in \tilde{\mathcal{A}}(j)\}\\
&\quad + 4\tau \log \frac{|\mathcal{X}_{\mathcal{H}}| \cdot 2^R}{2\tau}\\
&= H(X_{\mathcal{H}}) - \frac{1}{n} \sum_{j=1}^{M_n - 1} \Bigg[-\sum_{x_{\mathcal{H}}^n} \Pr\{X_{\mathcal{H}}^n = x_{\mathcal{H}}^n | X_{\mathcal{K}}^n \in \tilde{\mathcal{A}}(j)\} \cdot\\
&\quad \log \Pr\{X_{\mathcal{H}}^n = x_{\mathcal{H}}^n | X_{\mathcal{K}}^n \in \tilde{\mathcal{A}}(j)\} \Bigg] \cdot \Pr\{X_{\mathcal{K}}^n \in \tilde{\mathcal{A}}(j)\} + 4\tau \log \frac{|\mathcal{X}_{\mathcal{H}}| \cdot 2^R}{2\tau}, && (64)
\end{aligned}$$

where

(a) follows because of *i.i.d.* $P_{X_{\mathcal{K}}^n}$,

(b) is due to the inequality proved in Appendix D,

(c) follows by removing the term for $j = M_n$.

Here, for any $x^n_{\mathcal{H}}$ satisfying $x^n_{\mathcal{K}} = (x^n_{\mathcal{R}}, x^n_{\mathcal{H}}) \in \tilde{\mathcal{A}}(j)$ with some $x^n_{\mathcal{R}}$, we can show that

$$\begin{aligned}
&\Pr\{X^n_{\mathcal{H}} = x^n_{\mathcal{H}} | X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)\} \\
&= \frac{\Pr\{X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j) | X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\} \Pr\{X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\}}{\Pr\{X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)\}} \\
&= \frac{\sum_{x^n_{\mathcal{R}}:\ (x^n_{\mathcal{R}}, x^n_{\mathcal{H}}) \in \tilde{\mathcal{A}}(j)} \Pr\{X^n_{\mathcal{R}} = x^n_{\mathcal{R}}, X^n_{\mathcal{H}} = x^n_{\mathcal{H}} | X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\}}{\sum_{(\tilde{x}^n_{\mathcal{R}}, \tilde{x}^n_{\mathcal{H}}) \in \tilde{\mathcal{A}}(j)} \Pr\{X^n_{\mathcal{R}} = \tilde{x}^n_{\mathcal{R}}, X^n_{\mathcal{H}} = \tilde{x}^n_{\mathcal{H}}\}} \cdot \Pr\{X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\} \\
&\overset{\text{(d)}}{=} \frac{\sum_{x^n_{\mathcal{R}}:\ (x^n_{\mathcal{R}}, x^n_{\mathcal{H}}) \in \tilde{\mathcal{A}}(j)} \Pr\{X^n_{\mathcal{R}} = x^n_{\mathcal{R}} | X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\}}{\sum_{(\tilde{x}^n_{\mathcal{R}}, \tilde{x}^n_{\mathcal{H}}) \in \tilde{\mathcal{A}}(j)} \Pr\{X^n_{\mathcal{R}} = \tilde{x}^n_{\mathcal{R}}, X^n_{\mathcal{H}} = \tilde{x}^n_{\mathcal{H}}\}} \cdot \Pr\{X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\} \\
&\overset{\text{(e)}}{\le} \frac{\sum_{x^n_{\mathcal{R}} \in T^n_{\delta_3}(X_{\mathcal{R}} | x^n_{\mathcal{H}}, \hat{x}^n_{\mathcal{R}}(j))} \Pr\{X^n_{\mathcal{R}} = x^n_{\mathcal{R}} | X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\}}{\sum_{(\tilde{x}^n_{\mathcal{R}}, \tilde{x}^n_{\mathcal{H}}) \in \tilde{\mathcal{A}}(j)} \Pr\{X^n_{\mathcal{R}} = \tilde{x}^n_{\mathcal{R}}, X^n_{\mathcal{H}} = \tilde{x}^n_{\mathcal{H}}\}} \cdot \Pr\{X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\} \qquad (65) \\
&\overset{\text{(f)}}{\le} \frac{2^{n\{H(X_{\mathcal{R}} | X_{\mathcal{H}}, \hat{X}_{\mathcal{R}}) + \tau\}} \cdot 2^{-n\{H(X_{\mathcal{R}} | X_{\mathcal{H}}) - \tau\}}}{2^{n\{H(X_{\mathcal{K}} | \hat{X}_{\mathcal{R}}) - \tau\}} \cdot 2^{-n\{H(X_{\mathcal{K}}) + \tau\}}} \cdot 2^{-n\{H(X_{\mathcal{H}}) - \tau\}} \\
&= 2^{-n\{H(X_{\mathcal{H}} | \hat{X}_{\mathcal{R}}) - 5\tau\}}, \qquad (66)
\end{aligned}$$

where

(d) follows from the fact that

$$\Pr\{X^n_{\mathcal{R}} = x^n_{\mathcal{R}}, X^n_{\mathcal{H}} = x^n_{\mathcal{H}} | X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\} = \Pr\{X^n_{\mathcal{R}} = x^n_{\mathcal{R}} | X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\},$$

(e) is due to the inequality proved in Appendix E, and

(f) follows because of the number of strongly typical sequences.

Therefore, from Equations (61), (64) and (66) we can obtain

$$\begin{aligned}
l_n &\le H(X_{\mathcal{H}}) - \frac{1}{n} \sum_{j=1}^{M_n - 1} \Bigg[n \sum_{x^n_{\mathcal{H}}} \Pr\{X^n_{\mathcal{H}} = x^n_{\mathcal{H}} | X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)\} \cdot \\
&\qquad \{H(X_{\mathcal{H}} | \hat{X}_{\mathcal{R}}) - 5\tau\} \Bigg] \cdot \Pr\{X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)\} + 4\tau \log \frac{|\mathcal{X}_{\mathcal{H}}| \cdot 2^R}{2\tau} \\
&= H(X_{\mathcal{H}}) - \Pr\left\{ X^n_{\mathcal{K}} \in \left(\bigcup_{j=1}^{M_n - 1} \tilde{\mathcal{A}}(j) \right) \right\} \cdot \{H(X_{\mathcal{H}} | \hat{X}_{\mathcal{R}}) - 5\tau\} + 4\tau \log \frac{|\mathcal{X}_{\mathcal{H}}| \cdot 2^R}{2\tau} \\
&\le H(X_{\mathcal{H}}) - (1 - \tau)\{H(X_{\mathcal{H}} | \hat{X}_{\mathcal{R}}) - 5\tau\} + 4\tau \log \frac{|\mathcal{X}_{\mathcal{H}}| \cdot 2^R}{2\tau} \\
&\le I(X_{\mathcal{H}}; \hat{X}_{\mathcal{R}}) + \tau(\log |\mathcal{X}_{\mathcal{H}}| + 5) + 4\tau \log \frac{|\mathcal{X}_{\mathcal{H}}| \cdot 2^R}{2\tau}. \qquad (67)
\end{aligned}$$

Since constants ϵ, τ, and δ are fixed to satisfy (45)–(48), from (44), (57)–(59) and (67), we obtain

$$\begin{aligned}
r_n &\le R, && (68) \\
u_n &\le \mathbb{E}[d(X_{\mathcal{R}}, \hat{X}_{\mathcal{R}})] + \epsilon < D, && (69) \\
l_n &< I(X_{\mathcal{H}}; \hat{X}_{\mathcal{R}}) + \epsilon < L, && (70) \\
e_n &\le I(X_{\mathcal{H}}; X_{\mathcal{E}}) < E. && (71)
\end{aligned}$$

Therefore, for the fixed distribution $P_{X_{\mathcal{E}}, X_{\mathcal{E}^c}} \cdot P_{\hat{X}_{\mathcal{R}} | X_{\mathcal{E}}}$ any tuple

$$(R,D,L,E) \in \{(R,D,L,E):\ \begin{aligned} & R > I(X_{\mathcal{E}};\hat{X}_{\mathcal{R}}), \\ & D > \mathbb{E}[d(X_{\mathcal{R}},\hat{X}_{\mathcal{R}})], \\ & L > I(X_{\mathcal{H}};\hat{X}_{\mathcal{R}}), \\ & E > I(X_{\mathcal{H}};X_{\mathcal{E}})\} =: \mathcal{S}^*_{\mathcal{E}}(P_{X_{\mathcal{K}}}) \end{aligned} \tag{72}$$

is achievable. Consequently, $\mathcal{S}^*_{\mathcal{E}}(P_{X_{\mathcal{K}}}) \subseteq \mathcal{C}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$. Taking the closure for the left-hand side (l.h.s.), we obtain $Cl(\mathcal{S}^*_{\mathcal{E}}(P_{X_{\mathcal{K}}})) \subseteq \mathcal{C}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ because $\mathcal{C}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ is a closed set. We conclude that $\mathcal{S}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) = \bigcup_p Cl(\mathcal{S}^*_{\mathcal{E}}(P_{X_{\mathcal{K}}})) \subseteq \mathcal{C}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ because the distribution $P_{X_{\mathcal{K}}} = P_{X_{\mathcal{E}},X_{\mathcal{E}^c}} \cdot P_{\hat{X}_{\mathcal{R}}|X_{\mathcal{E}}}$ is fixed arbitrarily. We complete the proof of the direct part.

4. First-Order Rate Analysis with Excess-Distortion Probability

4.1. Performance Measures

Hereafter, let the pair of the encoder and decoder (f_n, g_n) be fixed.

For a given M_n, the coding rate is defined as

$$r_n := \frac{1}{n}\log M_n. \tag{73}$$

Let $d:\ \mathcal{X}_{\mathcal{R}} \times \hat{\mathcal{X}}_{\mathcal{R}} \to [0,\infty)$ be a distortion function between $x_{\mathcal{R}} \in \mathcal{X}_{\mathcal{R}}$ and $\hat{x}_{\mathcal{R}} \in \hat{\mathcal{X}}_{\mathcal{R}}$. The distortion between sequences $x^n_{\mathcal{R}} \in \mathcal{X}^n_{\mathcal{R}}$ and $\hat{x}^n_{\mathcal{R}} \in \hat{\mathcal{X}}^n_{\mathcal{R}}$ is defined as

$$d(x^n_{\mathcal{R}},\hat{x}^n_{\mathcal{R}}) := \sum_{i=1}^{n} d(x_{\mathcal{R},i},\hat{x}_{\mathcal{R},i}). \tag{74}$$

Then, the measure of utility is defined as

$$u_n := \Pr\left\{\frac{1}{n}d(X^n_{\mathcal{R}},\hat{X}^n_{\mathcal{R}}) > D\right\}. \tag{75}$$

This measurement is called **excess-distortion probability** for $D \geq 0$.

In this system, the privacy of the hidden source sequence $X^n_{\mathcal{H}}$ should be protected when the codeword J_n is observed by decoder g_n. The measure of privacy for the decoder is defined as

$$l_n := \frac{1}{n}I(X^n_{\mathcal{H}};J_n), \tag{76}$$

where $I(X^n_{\mathcal{H}};J_n)$ is the mutual information between $X^n_{\mathcal{H}}$ and J_n.

The privacy of the hidden source sequence $X^n_{\mathcal{H}}$ should be protected when the encoded information $X_{\mathcal{E}}$ is observed by encoder f_n. The measurement of privacy for the encoder is defined as

$$e_n := \frac{1}{n}I(X^n_{\mathcal{H}};X^n_{\mathcal{E}}), \tag{77}$$

where $I(X^n_{\mathcal{H}};X^n_{\mathcal{E}})$ is the mutual information between $X^n_{\mathcal{H}}$ and $X^n_{\mathcal{E}}$.

4.2. Achievable Region and Theorem

We define the achievable region for the first-order rate analysis with the excess-distortion probability and state the obtained results.

Definition 10. *A tuple (R,D,L,E) is said to be **ϵ-achievable** (with respect to the excess-distortion probability) if, for any given $\epsilon > 0$, there exists a sequence of codes (f_n,g_n) satisfying*

$$r_n \le R + \epsilon, \tag{78}$$
$$u_n \le \epsilon, \tag{79}$$
$$l_n \le L + \epsilon, \tag{80}$$
$$e_n \le E + \epsilon \tag{81}$$

for all sufficiently large n.

The technical meanings of each constraint in Definition 10 can be interpreted as follows: Equation (78) evaluates how much the source sequence is compressed, so this rate should be decreased. Equation (79) is the constraint corresponding to the excess-distortion probability being less than ϵ, so this condition should also be decreased. Equation (80) constrains the amount of leaked private information to the decoder. Since private information should be kept secret for the receiver, this quantity should be decreased as well. Equation (81) constrains the amount of leaked private information to the encoder. For the same reason as (80), this quantity should also be decreased.

Definition 11. *The closure of the set of ϵ-achievable tuples (R, D, L, E) is referred to as the ϵ-**achievable region** and is denoted by $\mathcal{L}_{\mathcal{E}}(\epsilon|P_{X_{\mathcal{K}}})$ and define*

$$\mathcal{L}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) := \bigcap_{0<\epsilon<1} \mathcal{L}_{\mathcal{E}}(\epsilon|P_{X_{\mathcal{K}}}). \tag{82}$$

We establish the following theorem. For the proof of this theorem, please refer to Sections 4.3 and 4.4.

Theorem 2. *For any $\mathcal{E}$ such that $\mathcal{R} \subseteq \mathcal{E} \subseteq \mathcal{K}$, the achievable region of the coding system is given by*

$$\mathcal{L}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) = \mathcal{S}_{\mathcal{E}}(P_{X_{\mathcal{K}}}). \tag{83}$$

Remark 4. *From Theorems 1 and 2, we find that the achievable region in which utility is measured by the expected distortion is equal to the one in which utility is measured by the excess-distortion probability.*

Because in Section 6 we discuss the achievable region among coding rate, utility, and privacy, a characterization of the achievable region is derived by projecting the characterization in Theorem 2 onto the R-D-L hyperplane.

Definition 12. *For any $\mathcal{E}$ such that $\mathcal{R} \subseteq \mathcal{E} \subseteq \mathcal{K}$, we define*

$$\mathcal{L}^{RDL}_{\mathcal{E}}(\epsilon|P_{X_{\mathcal{K}}}) := \{(R, D, L) : (R, D, L, E) \in \mathcal{L}_{\mathcal{E}}(\epsilon|P_{X_{\mathcal{K}}})\} \tag{84}$$

and

$$\mathcal{L}^{RDL}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) := \bigcap_{0<\epsilon<1} \mathcal{L}^{RDL}_{\mathcal{E}}(\epsilon|P_{X_{\mathcal{K}}}). \tag{85}$$

Definition 13. *For any $\mathcal{E}$ such that $\mathcal{R} \subseteq \mathcal{E} \subseteq \mathcal{K}$, we define*

$$\begin{aligned}\mathcal{S}^{RDL}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) = \{(R, D, L) :\ & R \ge I(X_{\mathcal{E}}; \hat{X}_{\mathcal{R}}),\\ & D \ge \mathbb{E}[d(X_{\mathcal{R}}, \hat{X}_{\mathcal{R}})],\\ & L \ge I(X_{\mathcal{H}}; \hat{X}_{\mathcal{R}})\\ & \text{for some } P_{X_{\mathcal{E}}, X_{\mathcal{E}^c}} \cdot P_{\hat{X}_{\mathcal{R}}|X_{\mathcal{E}}}\}.\end{aligned} \tag{86}$$

Corollary 2. *For any $\mathcal{E}$ such that $\mathcal{R} \subseteq \mathcal{E} \subseteq \mathcal{K}$, the region $\mathcal{L}_{\mathcal{E}}^{RDL}(P_{X_{\mathcal{K}}})$ is given by*

$$\mathcal{L}_{\mathcal{E}}^{RDL}(P_{X_{\mathcal{K}}}) = \mathcal{S}_{\mathcal{E}}^{RDL}(P_{X_{\mathcal{K}}}). \tag{87}$$

Examples of numerical calculation of this result are shown in Section 6.1.

Since we focus on the achievable region between utility and privacy in the next section, a characterization of the achievable region is derived by further projecting the result of Theorem 2 onto the D-L plane.

Definition 14. *For any $\mathcal{E}$ such that $\mathcal{R} \subseteq \mathcal{E} \subseteq \mathcal{K}$, we define*

$$\mathcal{L}_{\mathcal{E}}^{DL}(\epsilon|P_{X_{\mathcal{K}}}) := \{(D, L) : \ (R, D, L, E) \in \mathcal{L}_{\mathcal{E}}(\epsilon|P_{X_{\mathcal{K}}})\} \tag{88}$$

and

$$\mathcal{L}_{\mathcal{E}}^{DL}(P_{X_{\mathcal{K}}}) := \bigcap_{0<\epsilon<1} \mathcal{L}_{\mathcal{E}}^{DL}(\epsilon|P_{X_{\mathcal{K}}}). \tag{89}$$

Definition 15. *For any $\mathcal{E}$ such that $\mathcal{R} \subseteq \mathcal{E} \subseteq \mathcal{K}$, we define*

$$\begin{aligned}\mathcal{S}_{\mathcal{E}}^{DL}(P_{X_{\mathcal{K}}}) = \{(D, L) : \ & D \geq \mathbb{E}[d(X_{\mathcal{R}}, \hat{X}_{\mathcal{R}})], \\ & L \geq I(X_{\mathcal{H}}; \hat{X}_{\mathcal{R}}) \\ & \text{for some } P_{X_{\mathcal{E}}, X_{\mathcal{E}^c}} \cdot P_{\hat{X}_{\mathcal{R}}|X_{\mathcal{E}}}\}. \end{aligned} \tag{90}$$

Corollary 3. *For any $\mathcal{E}$ such that $\mathcal{R} \subseteq \mathcal{E} \subseteq \mathcal{K}$, the region $\mathcal{L}_{\mathcal{E}}^{DL}(P_{X_{\mathcal{K}}})$ is given by*

$$\mathcal{L}_{\mathcal{E}}^{DL}(P_{X_{\mathcal{K}}}) = \mathcal{S}_{\mathcal{E}}^{DL}(P_{X_{\mathcal{K}}}). \tag{91}$$

4.3. Proof of Converse Part

From Section 3.4 (proof of the converse part), we have

$$\mathcal{C}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) \subseteq \mathcal{S}_{\mathcal{E}}(P_{X_{\mathcal{K}}}). \tag{92}$$

Let a tuple $(R, D, L, E) \in \mathcal{L}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ be arbitrarily fixed and $\epsilon > 0$ and $\epsilon' > 0$ be given. From the argument of the method of types, the sequences $x_{\mathcal{R}}^n$ are divided into two categories: distortion-typical or non-distortion-typical with some $\hat{x}_{\mathcal{R}}^n$. The sequences of the former categories satisfy $\frac{1}{n}d(x_{\mathcal{R}}^n, \hat{x}_{\mathcal{R}}^n) < D + \epsilon$ and the sequences of the latter one satisfy $\frac{1}{n}d(x_{\mathcal{R}}^n, \hat{x}_{\mathcal{R}}^n) < d_{\max}$ where $d_{\max} := \max_{x_{\mathcal{R}} \in \mathcal{X}_{\mathcal{R}},\, \hat{x}_{\mathcal{R}} \in \hat{\mathcal{X}}_{\mathcal{R}}} d(x_{\mathcal{R}}, \hat{x}_{\mathcal{R}})$. Then, the expected distortion is bounded from above as

$$\begin{aligned}\mathbb{E}\left[\frac{1}{n}d(X_{\mathcal{R}}^n, \hat{X}_{\mathcal{R}}^n)\right] &\leq D + \epsilon + \Pr\left\{\frac{1}{n}d(x_{\mathcal{R}}^n, \hat{x}_{\mathcal{R}}^n) > D + \epsilon\right\} \cdot d_{\max} \\ &\leq D + \epsilon + \Pr\left\{\frac{1}{n}d(x_{\mathcal{R}}^n, \hat{x}_{\mathcal{R}}^n) > D\right\} \cdot d_{\max} \\ &\overset{\text{(a)}}{\leq} D + \epsilon + \epsilon' d_{\max}, \end{aligned} \tag{93}$$

where (a) follows from (79) of ϵ-achievable in which utility is measured by the excess-distortion probability. Since $\epsilon + \epsilon' d_{\max}$ can be arbitrarily small with proper choices of ϵ and ϵ', (15) can be derived. This means

$$\mathcal{L}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) \subseteq \mathcal{C}_{\mathcal{E}}(P_{X_{\mathcal{K}}}). \tag{94}$$

From both inclusion relations,

$$\mathcal{L}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) \subseteq \mathcal{C}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) \subseteq \mathcal{S}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) \tag{95}$$

is evidently satisfied.

4.4. Proof of the Direct Part

In this part, we provide a sketch of the proof of $\mathcal{S}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) \subseteq \mathcal{L}_{\mathcal{E}}(\epsilon|P_{X_{\mathcal{K}}})$.

Under an arbitrarily fixed distribution $P_{X_{\mathcal{E}},X_{\mathcal{E}^c}} \cdot P_{\hat{X}_{\mathcal{R}}|X_{\mathcal{E}}}$, any tuple $(R, D, L, E) \in \mathcal{S}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ is chosen such that

$$R > I(X_{\mathcal{E}}; \hat{X}_{\mathcal{R}}), \tag{96}$$

$$D > \mathbb{E}[d(X_{\mathcal{R}}, \hat{X}_{\mathcal{R}})], \tag{97}$$

$$L > I(X_{\mathcal{H}}; \hat{X}_{\mathcal{R}}), \tag{98}$$

$$E > I(X_{\mathcal{H}}; X_{\mathcal{E}}). \tag{99}$$

From (97) and (98), we can choose a sufficiently small $\epsilon > 0$ such that

$$D > \mathbb{E}[d(X_{\mathcal{R}}, \hat{X}_{\mathcal{R}})] + \epsilon, \tag{100}$$

$$L > I(X_{\mathcal{H}}; \hat{X}_{\mathcal{R}}) + \epsilon. \tag{101}$$

In addition, with this ϵ, some constant $0 < \tau < \frac{1}{2}$ is fixed such that

$$\tau(\log |\mathcal{X}_{\mathcal{H}}| + 5) + 4\tau \log \frac{|\mathcal{X}_{\mathcal{H}}| \cdot 2^R}{2\tau} < \epsilon. \tag{102}$$

We can also choose positive numbers $\delta(:= \delta(n))$ such that

$$2\delta^2(n) \leq R - I(X_{\mathcal{E}}; \hat{X}_{\mathcal{R}}) - \frac{1}{n} - \tau, \tag{103}$$

$$\delta(n) \to 0, \tag{104}$$

$$\sqrt{n} \cdot \delta(n) \to \infty \tag{105}$$

as $n \to \infty$. Let $\delta(n) = \frac{c}{\sqrt{n}} \log n$ where c is a constant, and obviously (104) and (105) are satisfied.

Generation of codebook: Randomly generate $\hat{x}^n_{\mathcal{R}}(j)$ from the strongly typical sequences $T^n_\delta(\hat{X}_{\mathcal{R}})$ for $j = 1, 2, \ldots, M_n := 2^{nR}$. Reveal the codebook $\mathcal{C} = \{\hat{x}^n_{\mathcal{R}}(1), \ldots, \hat{x}^n_{\mathcal{R}}(M_n)\}$ to the encoder and decoder.

Encoding: If a sequence $x^n_{\mathcal{E}} \in \mathcal{X}^n_{\mathcal{E}}$ satisfies $x^n_{\mathcal{K}} = (x^n_{\mathcal{E}}, x^n_{\mathcal{E}^c})$ with some $x^n_{\mathcal{E}^c} \in \mathcal{X}^n_{\mathcal{E}^c}$, we write $x^n_{\mathcal{E}} \prec x^n_{\mathcal{K}}$. Given $x^n_{\mathcal{K}}$, the encoder finds j such that $x^n_{\mathcal{E}} \in T^n_\delta(X_{\mathcal{E}}|\hat{x}^n_{\mathcal{R}}(j))$ and sets $f_n(x^n_{\mathcal{E}}) = j$ where $T^n_\delta(X_{\mathcal{E}}|\hat{x}^n_{\mathcal{R}}(j))$ is the conditional strongly typical sequences. If there exist multiple such j, $f_n(x^n_{\mathcal{E}})$ is set as the minimum one. If there are no such j, then $f_n(x^n_{\mathcal{E}}) = M_n$.

Decoding: When j is observed, the decoder sets the reproduced sequence as $\hat{X}^n_{\mathcal{R}} = \hat{x}^n_{\mathcal{R}}(j)$.

Evaluation: We define $\mathcal{A}(j)$, $\mathcal{B}(j)$, and $\tilde{\mathcal{A}}(j)$ as

$$\mathcal{A}(j) := \{x^n_{\mathcal{E}} \colon\ f_n(x^n_{\mathcal{E}}) = j\}, \tag{106}$$

$$\mathcal{B}(j) := \{x^n_{\mathcal{K}} \colon\ x^n_{\mathcal{E}} \prec x^n_{\mathcal{K}}, f_n(x^n_{\mathcal{E}}) = j\}, \tag{107}$$

$$\tilde{\mathcal{A}}(j) := \begin{cases} \{x^n_{\mathcal{K}} \colon\ x^n_{\mathcal{E}} \prec x^n_{\mathcal{K}}, f_n(x^n_{\mathcal{E}}) = j, x^n_{\mathcal{K}} \in T^n_{2\delta}(X_{\mathcal{K}}|\hat{x}^n_{\mathcal{R}}(j))\} \\ \hfill (j = 1, 2, \ldots, M_n - 1) \\ \{x^n_{\mathcal{K}} \colon\ x^n_{\mathcal{K}} \in \mathcal{X}^n_{\mathcal{K}} \setminus \bigcup_{j=1}^{M_n - 1} \tilde{\mathcal{A}}(j)\} \quad (j = M_n). \end{cases} \tag{108}$$

It is easily verified that $\mathcal{A}(j)$ for $j = 1, 2, \ldots, M_n$ (and also $\mathcal{B}(j)$ and $\tilde{\mathcal{A}}(j)$) is disjoint. From the definitions of J_n, $\mathcal{A}(j)$, and $\mathcal{B}(j)$,

$$\Pr\{J_n = j\} = \Pr\{X^n_{\mathcal{E}} \in \mathcal{A}(j)\} = \Pr\{X^n_{\mathcal{K}} \in \mathcal{B}(j)\} \quad \text{for } j = 1, 2, \ldots, M_n. \tag{109}$$

For sufficiently large n, we can prove (cf. Appendix B)

$$|\Pr\{X^n_{\mathcal{K}} \in \mathcal{B}(j)\} - \Pr\{X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)\}| \leq 2|\mathcal{X}_{\mathcal{K}}| \cdot |\hat{\mathcal{X}}_{\mathcal{R}}| \mathrm{e}^{-2\delta^2 n} \quad \text{for } j = 1, 2, \ldots, M_n - 1. \tag{110}$$

For sufficiently large n, we can show that there exists a code (f_n, g_n) such that (cf. Appendix F)

$$r_n \leq R, \tag{111}$$

$$\Pr\left\{ X^n_{\mathcal{E}} \notin \bigcup_{j=1}^{M_n - 1} \mathcal{A}(j) \right\} \leq (2|\mathcal{X}_{\mathcal{E}}| + 1)\mathrm{e}^{-2\delta^2 n}, \tag{112}$$

$$u_n \leq (2|\mathcal{X}_{\mathcal{E}}| + 1)\mathrm{e}^{-2\delta^2 n}, \tag{113}$$

$$e_n \leq I(X_{\mathcal{H}}; X_{\mathcal{E}}), \tag{114}$$

$$\Pr\left\{ X^n_{\mathcal{K}} \notin \bigcup_{j=1}^{M_n - 1} \tilde{\mathcal{A}}(j) \right\} \leq \tau, \tag{115}$$

$$|\tilde{\mathcal{A}}(j)| \geq 2^{n\{H(X_{\mathcal{K}}|\hat{X}_{\mathcal{R}}) - \tau\}}. \tag{116}$$

For this code (f_n, g_n), we evaluate the privacy leakage against the decoder as

$$l_n := \frac{1}{n} I(X^n_{\mathcal{H}}; J_n) \tag{117}$$

$$\begin{aligned}
&= \frac{1}{n} H(X^n_{\mathcal{H}}) - \frac{1}{n} H(X^n_{\mathcal{H}}|J_n) \\
&\overset{\text{(a)}}{=} H(X_{\mathcal{H}}) - \frac{1}{n} H(X^n_{\mathcal{H}}|J_n) \\
&= H(X_{\mathcal{H}}) - \frac{1}{n} \sum_{j=1}^{M_n} H(X^n_{\mathcal{H}}|X^n_{\mathcal{K}} \in \mathcal{B}(j)) \Pr\{X^n_{\mathcal{K}} \in \mathcal{B}(j)\} \\
&\overset{\text{(b)}}{\leq} H(X_{\mathcal{H}}) - \frac{1}{n} \sum_{j=1}^{M_n} H(X^n_{\mathcal{H}}|X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)) \Pr\{X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)\} \\
&\quad + 4\tau \log \frac{|\mathcal{X}_{\mathcal{H}}| \cdot 2^R}{2\tau}
\end{aligned} \tag{118}$$

$$\begin{aligned}
&\overset{\text{(c)}}{\leq} H(X_{\mathcal{H}}) - \frac{1}{n} \sum_{j=1}^{M_n - 1} H(X^n_{\mathcal{H}}|X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)) \Pr\{X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)\} \\
&\quad + 4\tau \log \frac{|\mathcal{X}_{\mathcal{H}}| \cdot 2^R}{2\tau} \\
&= H(X_{\mathcal{H}}) - \frac{1}{n} \sum_{j=1}^{M_n - 1} \left[- \sum_{x^n_{\mathcal{H}}} \Pr\{X^n_{\mathcal{H}} = x^n_{\mathcal{H}}|X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)\} \cdot \log \Pr\{X^n_{\mathcal{H}} = x^n_{\mathcal{H}}|X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)\} \right] \cdot \\
&\quad \Pr\{X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)\} + 4\tau \log \frac{|\mathcal{X}_{\mathcal{H}}| \cdot 2^R}{2\tau},
\end{aligned} \tag{119}$$

where

(a) follows because of *i.i.d.* $P_{X^n_{\mathcal{K}}}$,

(b) is due to the inequality proved in Appendix D, and

(c) follows by removing the term for $j = M_n$.

Here, for any $x^n_{\mathcal{H}}$ satisfying $x^n_{\mathcal{K}} = (x^n_{\mathcal{R}}, x^n_{\mathcal{H}}) \in \tilde{\mathcal{A}}(j)$ with some $x^n_{\mathcal{R}}$, we can show that

$$\begin{aligned}
&\Pr\{X^n_{\mathcal{H}} = x^n_{\mathcal{H}} | X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)\} \\
&= \frac{\Pr\{X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j) | X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\} \Pr\{X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\}}{\Pr\{X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)\}} \\
&= \frac{\sum_{x^n_{\mathcal{R}}:\ (x^n_{\mathcal{R}}, x^n_{\mathcal{H}}) \in \tilde{\mathcal{A}}(j)} \Pr\{X^n_{\mathcal{R}} = x^n_{\mathcal{R}}, X^n_{\mathcal{H}} = x^n_{\mathcal{H}} | X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\}}{\sum_{(\tilde{x}^n_{\mathcal{R}}, \tilde{x}^n_{\mathcal{H}}) \in \tilde{\mathcal{A}}(j)} \Pr\{X^n_{\mathcal{R}} = \tilde{x}^n_{\mathcal{R}}, X^n_{\mathcal{H}} = \tilde{x}^n_{\mathcal{H}}\}} \cdot \Pr\{X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\} \\
&\overset{\text{(d)}}{=} \frac{\sum_{x^n_{\mathcal{R}}:\ (x^n_{\mathcal{R}}, x^n_{\mathcal{H}}) \in \tilde{\mathcal{A}}(j)} \Pr\{X^n_{\mathcal{R}} = x^n_{\mathcal{R}} | X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\}}{\sum_{(\tilde{x}^n_{\mathcal{R}}, \tilde{x}^n_{\mathcal{H}}) \in \tilde{\mathcal{A}}(j)} \Pr\{X^n_{\mathcal{R}} = \tilde{x}^n_{\mathcal{R}}, X^n_{\mathcal{H}} = \tilde{x}^n_{\mathcal{H}}\}} \cdot \Pr\{X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\} \\
&\overset{\text{(e)}}{\leq} \frac{\sum_{x^n_{\mathcal{R}} \in T^n_{\delta_3}(X_{\mathcal{R}} | x^n_{\mathcal{H}}, \hat{x}^n_{\mathcal{R}}(j))} \Pr\{X^n_{\mathcal{R}} = x^n_{\mathcal{R}} | X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\}}{\sum_{(\tilde{x}^n_{\mathcal{R}}, \tilde{x}^n_{\mathcal{H}}) \in \tilde{\mathcal{A}}(j)} \Pr\{X^n_{\mathcal{R}} = \tilde{x}^n_{\mathcal{R}}, X^n_{\mathcal{H}} = \tilde{x}^n_{\mathcal{H}}\}} \cdot \Pr\{X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\} \qquad (120) \\
&\overset{\text{(f)}}{\leq} \frac{2^{n\{H(X_{\mathcal{R}} | X_{\mathcal{H}}, \hat{X}_{\mathcal{R}}) + \tau\}} \cdot 2^{-n\{H(X_{\mathcal{R}} | X_{\mathcal{H}}) - \tau\}}}{2^{n\{H(X_{\mathcal{K}} | \hat{X}_{\mathcal{R}}) - \tau\}} \cdot 2^{-n\{H(X_{\mathcal{K}}) + \tau\}}} \cdot 2^{-n\{H(X_{\mathcal{H}}) - \tau\}} \\
&= 2^{-n\{H(X_{\mathcal{H}} | \hat{X}_{\mathcal{R}}) - 5\tau\}}, \qquad (121)
\end{aligned}$$

where

(d) follows from the fact that

$$\Pr\{X^n_{\mathcal{R}} = x^n_{\mathcal{R}}, X^n_{\mathcal{H}} = x^n_{\mathcal{H}} | X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\} = \Pr\{X^n_{\mathcal{R}} = x^n_{\mathcal{R}} | X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\},$$

(e) is due to the inequality proved in Appendix E, and

(f) follows because of the number of strongly typical sequences.

Therefore, from Equations (115), (119), and (121), we can obtain

$$\begin{aligned}
l_n &\leq H(X_{\mathcal{H}}) - \frac{1}{n} \sum_{j=1}^{M_n - 1} \left[n \sum_{x^n_{\mathcal{H}}} \Pr\{X^n_{\mathcal{H}} = x^n_{\mathcal{H}} | X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)\} \cdot \right. \\
&\qquad \left. \{H(X_{\mathcal{H}} | \hat{X}_{\mathcal{R}}) - 5\tau\} \right] \cdot \Pr\{X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)\} + 4\tau \log \frac{|\mathcal{X}_{\mathcal{H}}| \cdot 2^R}{2\tau} \\
&= H(X_{\mathcal{H}}) - \Pr\left\{ X^n_{\mathcal{K}} \in \left(\bigcup_{j=1}^{M_n - 1} \tilde{\mathcal{A}}(j) \right) \right\} \cdot \{H(X_{\mathcal{H}} | \hat{X}_{\mathcal{R}}) - 5\tau\} + 4\tau \log \frac{|\mathcal{X}_{\mathcal{H}}| \cdot 2^R}{2\tau} \\
&\leq H(X_{\mathcal{H}}) - (1 - \tau)\{H(X_{\mathcal{H}} | \hat{X}_{\mathcal{R}}) - 5\tau\} + 4\tau \log \frac{|\mathcal{X}_{\mathcal{H}}| \cdot 2^R}{2\tau} \\
&\leq I(X_{\mathcal{H}}; \hat{X}_{\mathcal{R}}) + \tau\{H(X_{\mathcal{H}} | \hat{X}_{\mathcal{R}}) + 5\} + 4\tau \log \frac{|\mathcal{X}_{\mathcal{H}}| \cdot 2^R}{2\tau}. \qquad (122)
\end{aligned}$$

Since constants ϵ, τ, and δ are fixed to satisfy (100)–(102), from (111), (113), and (122), we obtain

$$\begin{aligned}
r_n &\leq R, && (123) \\
u_n &\leq \epsilon, && (124) \\
l_n &< I(X_{\mathcal{H}}; \hat{X}_{\mathcal{R}}) + \epsilon < L, && (125) \\
e_n &\leq I(X_{\mathcal{H}}; X_{\mathcal{E}}) < E. && (126)
\end{aligned}$$

Therefore, for the fixed distribution $P_{X_{\mathcal{E}}, X_{\mathcal{E}^c}} \cdot P_{\hat{X}_{\mathcal{R}} | X_{\mathcal{E}}}$, any tuple

$$
\begin{aligned}
(R, D, L, E) \in \{(R, D, L, E) :\ & R > I(X_{\mathcal{E}}; \hat{X}_{\mathcal{R}}), \\
& D > \mathbb{E}[d(X_{\mathcal{R}}, \hat{X}_{\mathcal{R}})], \\
& L > I(X_{\mathcal{H}}; \hat{X}_{\mathcal{R}}), \\
& E > I(X_{\mathcal{H}}; X_{\mathcal{E}})\} =: \mathcal{S}^*_{\mathcal{E}}(P_{X_{\mathcal{K}}})
\end{aligned} \tag{127}
$$

is achievable. Consequently, $\mathcal{S}^*_{\mathcal{E}}(P_{X_{\mathcal{K}}}) \subseteq \mathcal{L}_{\mathcal{E}}(\epsilon|P_{X_{\mathcal{K}}})$. Taking the closure for the l.h.s., we obtain $Cl(\mathcal{S}^*_{\mathcal{E}}(P_{X_{\mathcal{K}}})) \subseteq \mathcal{L}_{\mathcal{E}}(\epsilon|P_{X_{\mathcal{K}}})$ because $\mathcal{L}_{\mathcal{E}}(\epsilon|P_{X_{\mathcal{K}}})$ is a closed set. We conclude that $\mathcal{S}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) = \bigcup_p Cl(\mathcal{S}^*_{\mathcal{E}}(P_{X_{\mathcal{K}}})) \subseteq \mathcal{L}_{\mathcal{E}}(\epsilon|P_{X_{\mathcal{K}}})$ because the distribution $P_{X_{\mathcal{K}}, \hat{X}_{\mathcal{R}}} = P_{X_{\mathcal{E}}, X_{\mathcal{E}^c}} \cdot P_{\hat{X}_{\mathcal{R}}|X_{\mathcal{E}}}$ is fixed arbitrarily. We complete the proof of the direct part.

5. Strong Converse Theorem for Utility–Privacy Trade-Offs

5.1. Another Expression of the Achievable Region

In Section 5.1, we clarify that the achievable region $\mathcal{L}^{DL}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ defined in (89) coincides with the region expressed with a tangent plane.

Definition 16. *For any $\mathcal{E}$ such that $\mathcal{R} \subseteq \mathcal{E} \subseteq \mathcal{K}$, the region $\mathcal{T}^{\mu}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ is defined as*

$$
T^{\mu}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) := \min\{I(X_{\mathcal{H}}; \hat{X}_{\mathcal{R}}) + \mu\mathbb{E}[d(X_{\mathcal{R}}, \hat{X}_{\mathcal{R}})] \text{ for some } P_{\hat{X}_{\mathcal{R}}|X_{\mathcal{E}}} \cdot P_{X_{\mathcal{E}^c} X_{\mathcal{E}}}\},
$$

where

$$
\mathcal{T}^{DL}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) := \bigcap_{\mu \geq 0} \{(L, D) :\ L + \mu D \geq T^{\mu}_{\mathcal{E}}(P_{X_{\mathcal{K}}})\}.
$$

Theorem 3. *For any $\mathcal{E}$ such that $\mathcal{R} \subseteq \mathcal{E} \subseteq \mathcal{K}$, the region $\mathcal{S}^{DL}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ defined in (90) is given by*

$$
\mathcal{S}^{DL}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) = \mathcal{T}^{DL}_{\mathcal{E}}(P_{X_{\mathcal{K}}}), \tag{128}
$$

and the achievable region $\mathcal{L}^{DL}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$, which is the projection region of the achievable region $\mathcal{L}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ onto the D-L plane, is given by

$$
\mathcal{L}^{DL}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) = \mathcal{T}^{DL}_{\mathcal{E}}(P_{X_{\mathcal{K}}}). \tag{129}
$$

Proof. Figure 3 illustrates the proof image using a graph. Let a constance $\mu \geq 0$ be fixed arbitrarily. Like in Figure 3, there exists a boundary point (D_μ, L_μ) of $\mathcal{S}^{DL}_{\mathcal{E}}$ tangent to the line with slope $-\mu$. The intercept of this tangent line is $L_\mu + \mu D_\mu$.

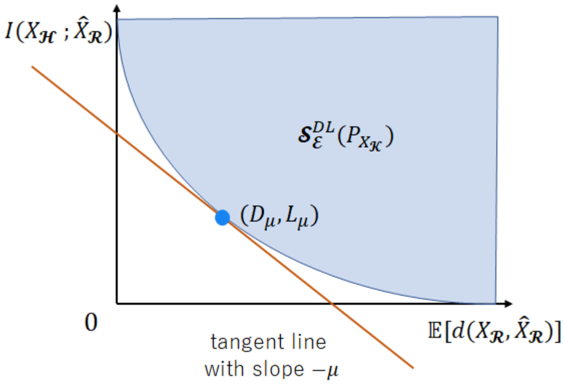
$I(X_{\mathcal{H}}; \hat{X}_{\mathcal{R}})$
$\mathcal{S}^{DL}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$
(D_μ, L_μ)
0
tangent line
with slope $-\mu$
$\mathbb{E}[d(X_{\mathcal{R}}, \hat{X}_{\mathcal{R}})]$

Figure 3. The region expressed with a tangent plane using the Legendre transformation.

The minimum $I(X_{\mathcal{H}};\hat{X}_{\mathcal{R}})+\mu\mathbb{E}[d(X_{\mathcal{R}},\hat{X}_{\mathcal{R}})]$ characterized by some distribution $P_{\hat{X}_{\mathcal{R}}|X_{\mathcal{E}}}$ coincides with $L_\mu+\mu D_\mu$. Therefore,

$$L_\mu+\mu D_\mu=\min\{I(X_{\mathcal{H}};\hat{X}_{\mathcal{R}})+\mu\mathbb{E}[d(X_{\mathcal{R}},\hat{X}_{\mathcal{R}})] \text{ for some } P_{\hat{X}_{\mathcal{R}}|X_{\mathcal{E}}}\cdot P_{X_{\mathcal{E}^c}X_{\mathcal{E}}}\}. \tag{130}$$

From (130), we obtain

$$\begin{aligned}\{(L,D):\ L+\mu D\geq L_\mu+\mu D_\mu\}=\{(L,D):\ L+\mu D\geq \min\{I(X_{\mathcal{H}};\hat{X}_{\mathcal{R}})&\\ +\mu\mathbb{E}[d(X_{\mathcal{R}},\hat{X}_{\mathcal{R}})] \text{ for some } P_{\hat{X}_{\mathcal{R}}|X_{\mathcal{E}}}\cdot P_{X_{\mathcal{E}^c}X_{\mathcal{E}}}\}\}.&\end{aligned} \tag{131}$$

Taking the intersection by $\mu\geq 0$ on the both sides of (131),

$$\bigcap_{\mu\geq 0}\{(L,D):\ L+\mu D\geq L_\mu+\mu D_\mu\}=\bigcap_{\mu\geq 0}\{(L,D):\ L+\mu D\geq T^{\mu}_{\mathcal{E}}(P_{X_{\mathcal{K}}})\}. \tag{132}$$

The l.h.s. of (131) shows the upper-right region in the first quadrant drawn by the tangent line with a slope $-\mu$ for $\mathcal{S}^{DL}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$. Since the l.h.s. of (132) is the intersection of the l.h.s. of (131), the l.h.s. of (132) represents $\mathcal{S}^{DL}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$. From Definition 16, the right-hand side (r.h.s.) of (132) is $\mathcal{T}^{DL}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$. As a result, (128) holds. Since $\mathcal{L}^{DL}_{\mathcal{E}}(P_{X_{\mathcal{K}}})=\mathcal{S}^{DL}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ from Corollary 3, likewise, (129) holds. □

5.2. Proof Preliminaries

In Section 5.2, we derive two fundamental properties of the minimization about two values and the inequalities about entropy and divergence to prove the strong converse theorem. In Proposition 1, we change the objective function $T^{\mu}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ of the region expressed with the tangent plane introduced in Section 5.1 onto the region expressed with divergence.

Proposition 1. *Let $\mu\leq 0$ be fixed arbitrarily. For any $\mathcal{E}$ such that $\mathcal{R}\subseteq\mathcal{E}\subseteq\mathcal{K}$,*

$$T^{\mu}_{\mathcal{E}}(P_{X_{\mathcal{K}}})=\sup_{\alpha>0}T^{\mu,\alpha}_{\mathcal{E}}(P_{X_{\mathcal{K}}}), \tag{133}$$

where

$$\begin{aligned}T^{\mu,\alpha}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) &:= \min_{P_{\tilde{X}_{\mathcal{E}^c}\tilde{X}_{\mathcal{E}}\tilde{\hat{X}}_{\mathcal{R}}}}[I(\tilde{X}_{\mathcal{H}};\tilde{\hat{X}}_{\mathcal{R}})+\mu\mathbb{E}[d(\tilde{X}_{\mathcal{R}},\tilde{\hat{X}}_{\mathcal{R}})]\\ &\qquad +\alpha D(P_{\tilde{X}_{\mathcal{E}^c}\tilde{X}_{\mathcal{E}}\tilde{\hat{X}}_{\mathcal{R}}}\parallel Q_{X_{\mathcal{E}^c}X_{\mathcal{E}}\tilde{\hat{X}}_{\mathcal{R}}})+D(P_{\tilde{X}_{\mathcal{E}^c}\tilde{X}_{\mathcal{E}}}\parallel P_{X_{\mathcal{E}^c}X_{\mathcal{E}}})]\\ &= \min_{P_{\tilde{X}_{\mathcal{E}^c}\tilde{X}_{\mathcal{E}}\tilde{\hat{X}}_{\mathcal{R}}}}[I(\tilde{X}_{\mathcal{H}};\tilde{\hat{X}}_{\mathcal{R}})+\mu\mathbb{E}[d(\tilde{X}_{\mathcal{R}},\tilde{\hat{X}}_{\mathcal{R}})]]\\ &\qquad +(\alpha+1)D(P_{\tilde{X}_{\mathcal{E}^c}\tilde{X}_{\mathcal{E}}}\parallel P_{X_{\mathcal{E}^c}X_{\mathcal{E}}})+\alpha I(\tilde{X}_{\mathcal{E}^c};\tilde{\hat{X}}_{\mathcal{R}}|\tilde{X}_{\mathcal{E}})],\end{aligned} \tag{134}$$

and $Q_{X_{\mathcal{E}^c}X_{\mathcal{E}}\tilde{\hat{X}}_{\mathcal{R}}}$ is the distribution induced from each $P_{\tilde{X}_{\mathcal{E}^c}\tilde{X}_{\mathcal{E}}\tilde{\hat{X}}_{\mathcal{R}}}$.

Proof. First, it is clear that $T^{\mu}_{\mathcal{E}}(P_{X_{\mathcal{K}}})\geq T^{\mu,\alpha}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ for all $\alpha>0$. To prove $T^{\mu}_{\mathcal{E}}(P_{X_{\mathcal{K}}})\leq T^{\mu,\alpha}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ for some $\alpha>0$, for $\alpha>0$, let $P^{\alpha}_{\tilde{X}_{\mathcal{E}^c}\tilde{X}_{\mathcal{E}}\tilde{\hat{X}}_{\mathcal{R}}}$ be the distribution that minimizes the r.h.s. of (134) and $Q^{\alpha}_{X_{\mathcal{E}^c}X_{\mathcal{E}}\tilde{\hat{X}}_{\mathcal{R}}}=P_{\tilde{\hat{X}}_{\mathcal{R}}|\tilde{X}_{\mathcal{E}}}P_{X_{\mathcal{E}^c}X_{\mathcal{E}}}$ be the estimated distribution. Since $G(P^{\alpha}_{\tilde{X}_{\mathcal{E}^c}\tilde{X}_{\mathcal{E}}\tilde{\hat{X}}_{\mathcal{R}}}):=I(\tilde{X}_{\mathcal{H}};\tilde{\hat{X}}_{\mathcal{R}})+\mathbb{E}[d(\tilde{X}_{\mathcal{R}},\tilde{\hat{X}}_{\mathcal{R}})]$ is non-negative and is bounded above, by setting $a=\log|\mathcal{X}_{\mathcal{H}}|+D_{\max}$, it must hold that

$$\alpha D(P^{\alpha}_{\tilde{X}_{\mathcal{E}^c}\tilde{X}_{\mathcal{E}}\tilde{\hat{X}}_{\mathcal{R}}}\parallel Q^{\alpha}_{X_{\mathcal{E}^c}X_{\mathcal{E}}\tilde{\hat{X}}_{\mathcal{R}}})\leq a$$

and thus

$$D(P^{\alpha}_{\tilde{X}_{\mathcal{E}^c}\tilde{X}_{\mathcal{E}}\tilde{\hat{X}}_{\mathcal{R}}}\parallel Q^{\alpha}_{X_{\mathcal{E}^c}X_{\mathcal{E}}\tilde{\hat{X}}_{\mathcal{R}}})\leq (a/\alpha).$$

Notice that any set of probability distributions on a finite alphabet forms a compact set. Because $G(P^{\alpha}_{\tilde{X}_{\mathcal{E}^c}\tilde{X}_{\mathcal{E}}\tilde{X}_{\mathcal{R}}})$ is a continuous function over a compact set, it is also uniformly continuous. Then, there exists a function $\Delta(t)$ satisfying $\Delta(t) \to 0$ as $t \to 0$ such that

$$\begin{aligned} T^{\mu,\alpha}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) &\geq G(P^{\alpha}_{\tilde{X}_{\mathcal{E}^c}\tilde{X}_{\mathcal{E}}\tilde{X}_{\mathcal{R}}}) \\ &\geq G(Q^{\alpha}_{X_{\mathcal{E}^c}X_{\mathcal{E}}\tilde{X}_{\mathcal{R}}}) - \Delta(a/\alpha) \\ &\geq T^{\mu}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) - \Delta(a/\alpha). \end{aligned}$$

Consequently, we obtain the desired inequality $T^{\mu}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) \leq \lim_{\alpha\to\infty} T^{\mu,\alpha}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ by taking $\alpha \to \infty$. □

In the following proposition, we show the inequalities satisfied between *i.i.d.* source $P_{X^n_{\mathcal{E}^c}X^n_{\mathcal{E}}}$ and arbitrary source $P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}}$.

Proposition 2. *For i.i.d. source $P_{X^n_{\mathcal{E}^c}X^n_{\mathcal{E}}}$, which has the common distribution $P_{X_{\mathcal{E}^c}X_{\mathcal{E}}}$ and arbitrary distribution $P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}}$, it holds that*

$$H(\tilde{X}^n_{\mathcal{E}^c}|\tilde{X}^n_{\mathcal{E}}) + D(P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}} \parallel P_{X^n_{\mathcal{E}^c}X^n_{\mathcal{E}}}) \geq n[H(\tilde{X}_{\mathcal{E}^c,J}|\tilde{X}_{\mathcal{E},J}) + D(P_{\tilde{X}_{\mathcal{E}^c,J}\tilde{X}_{\mathcal{E},J}} \parallel P_{X_{\mathcal{E}^c}X_{\mathcal{E}}})], \tag{135}$$

$$H(\tilde{X}^n_{\mathcal{H}}) + D(P_{\tilde{X}^n_{\mathcal{H}}\tilde{X}^n_{\mathcal{R}}} \parallel P_{X^n_{\mathcal{H}}X^n_{\mathcal{R}}}) \geq n[H(\tilde{X}_{\mathcal{H},J}) + D(P_{\tilde{X}_{\mathcal{H},J}\tilde{X}_{\mathcal{R},J}} \parallel P_{X_{\mathcal{H}}X_{\mathcal{R}}})], \tag{136}$$

where $J \sim \mathrm{unif}(1,\cdots,n)$ is the uniformly random variable over the set $\{1,2,\cdots,n\}$ for time-sharing and is assumed to be independent of all the other random variables involved.

Proof. The l.h.s. of (135) can be represented as

$$H(\tilde{X}^n_{\mathcal{E}^c}|\tilde{X}^n_{\mathcal{E}}) + D(P_{\tilde{X}^n_{\mathcal{E}^c}|\tilde{X}^n_{\mathcal{E}}} \parallel P_{X^n_{\mathcal{E}^c}|X^n_{\mathcal{E}}}|P_{\tilde{X}^n_{\mathcal{E}}}) + D(P_{\tilde{X}^n_{\mathcal{E}}} \parallel P_{X^n_{\mathcal{E}}}).$$

The sum of the first and second terms satisfies the following equation:

$$\begin{aligned} &H(\tilde{X}^n_{\mathcal{E}^c}|\tilde{X}^n_{\mathcal{E}}) + D(P_{\tilde{X}^n_{\mathcal{E}^c}|\tilde{X}^n_{\mathcal{E}}} \parallel P_{X^n_{\mathcal{E}^c}|X^n_{\mathcal{E}}}|P_{\tilde{X}^n_{\mathcal{E}}}) \\ &= \sum_{x^n_{\mathcal{E}^c},x^n_{\mathcal{E}}} P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}}(x^n_{\mathcal{E}^c},x^n_{\mathcal{E}}) \\ &\quad \cdot \left\{ \log \frac{1}{P_{\tilde{X}^n_{\mathcal{E}^c}|\tilde{X}^n_{\mathcal{E}}}(x^n_{\mathcal{E}^c}|x^n_{\mathcal{E}})} + \log \frac{P_{\tilde{X}^n_{\mathcal{E}^c}|\tilde{X}^n_{\mathcal{E}}}(x^n_{\mathcal{E}^c}|x^n_{\mathcal{E}})}{P_{X^n_{\mathcal{E}^c}|X^n_{\mathcal{E}}}(x^n_{\mathcal{E}^c}|x^n_{\mathcal{E}})} \right\} \\ &= \sum_{x^n_{\mathcal{E}^c},x^n_{\mathcal{E}}} P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}}(x^n_{\mathcal{E}^c},x^n_{\mathcal{E}}) \log \frac{1}{P_{X^n_{\mathcal{E}^c}|X^n_{\mathcal{E}}}(x^n_{\mathcal{E}^c}|x^n_{\mathcal{E}})} \\ &\overset{\text{(a)}}{=} \sum_{x^n_{\mathcal{E}^c},x^n_{\mathcal{E}}} P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}}(x^n_{\mathcal{E}^c},x^n_{\mathcal{E}}) \cdot \left\{ \sum_{j=1}^{n} \log \frac{1}{P_{X_{\mathcal{E}^c}|X_{\mathcal{E}}}(x_{\mathcal{E}^c,j}|x_{\mathcal{E},j})} \right\} \\ &\overset{\text{(b)}}{=} n \sum_{x_{\mathcal{E}^c},x_{\mathcal{E}}} P_{\tilde{X}_{\mathcal{E}^c,J}\tilde{X}_{\mathcal{E},J}}(x_{\mathcal{E}^c},x_{\mathcal{E}}) \log \frac{1}{P_{X_{\mathcal{E}^c}|X_{\mathcal{E}}}(x_{\mathcal{E}^c}|x_{\mathcal{E}})} \\ &= n \sum_{x_{\mathcal{E}^c},x_{\mathcal{E}}} P_{\tilde{X}_{\mathcal{E}^c,J}\tilde{X}_{\mathcal{E},J}}(x_{\mathcal{E}^c},x_{\mathcal{E}}) \\ &\quad \cdot \left\{ \log \frac{1}{P_{\tilde{X}_{\mathcal{E}^c,J}|\tilde{X}_{\mathcal{E},J}}(x_{\mathcal{E}^c}|x_{\mathcal{E}})} + \log \frac{P_{\tilde{X}_{\mathcal{E}^c,J}|\tilde{X}_{\mathcal{E},J}}(x_{\mathcal{E}^c}|x_{\mathcal{E}})}{P_{X_{\mathcal{E}^c}|X_{\mathcal{E}}}(x_{\mathcal{E}^c}|x_{\mathcal{E}})} \right\} \\ &= n\{H(\tilde{X}_{\mathcal{E}^c,J}|\tilde{X}_{\mathcal{E},J}) + D(P_{\tilde{X}_{\mathcal{E}^c,J}|\tilde{X}_{\mathcal{E},J}} \parallel P_{X_{\mathcal{E}^c}|X_{\mathcal{E}}}|P_{\tilde{X}_{\mathcal{E},J}})\}, \end{aligned} \tag{137}$$

where

(a) follows from the memoryless property of *i.i.d.* source $P_{X^n_{\mathcal{E}^c}X^n_{\mathcal{E}}}$;

(b) holds because $\frac{1}{n}\sum_{j=1}^{n} P_{\tilde{X}_{\mathcal{E}^c,j}\tilde{X}_{\mathcal{E},j}}(x_{\mathcal{E}^c},x_{\mathcal{E}}) = P_{\tilde{X}_{\mathcal{E}^c,J}\tilde{X}_{\mathcal{E},J}}(x_{\mathcal{E}^c},x_{\mathcal{E}})$.

The third term can be bounded from below as

$$\begin{aligned} D(P_{\tilde{X}^n_{\mathcal{E}}} \parallel P_{X^n_{\mathcal{E}}}) &= \sum_{j=1}^{n} D(P_{\tilde{X}_{\mathcal{E},j}|\tilde{X}^{j-1}_{\mathcal{E}}} \parallel P_{X_{\mathcal{E}}} | P_{\tilde{X}^{j-1}_{\mathcal{E}}}) \\ &\overset{\text{(c)}}{\geq} \sum_{j=1}^{n} D(P_{\tilde{X}_{\mathcal{E},j}} \parallel P_{X_{\mathcal{E}}}) \\ &\overset{\text{(d)}}{\geq} nD(P_{\tilde{X}_{\mathcal{E},J}} \parallel P_{X_{\mathcal{E}}}), \end{aligned} \tag{138}$$

where

(c) follows from the data processing inequality and
(d) holds because of Jensen's inequality.

From (137) and (138), (135) can be derived.

Likewise, the l.h.s. of (136) can be represented as

$$H(\tilde{X}^n_{\mathcal{H}}) + D(P_{\tilde{X}^n_{\mathcal{H}}} \parallel P_{X^n_{\mathcal{H}}}) + D(P_{\tilde{X}^n_{\mathcal{R}}|\tilde{X}^n_{\mathcal{H}}} \parallel P_{X^n_{\mathcal{R}}|X^n_{\mathcal{H}}} | P_{\tilde{X}^n_{\mathcal{H}}}),$$

The sum of the first and second terms satisfies

$$\begin{aligned} & H(\tilde{X}^n_{\mathcal{H}}) + D(P_{\tilde{X}^n_{\mathcal{H}}} \parallel P_{X^n_{\mathcal{H}}}) \\ &= \sum_{x^n_{\mathcal{H}}} P_{\tilde{X}^n_{\mathcal{H}}}(x^n_{\mathcal{H}}) \left\{ \log \frac{1}{P_{\tilde{X}^n_{\mathcal{H}}}(x^n_{\mathcal{H}})} + \log \frac{P_{\tilde{X}^n_{\mathcal{H}}}(x^n_{\mathcal{H}})}{P_{X^n_{\mathcal{H}}}(x^n_{\mathcal{H}})} \right\} \\ &= \sum_{x^n_{\mathcal{H}}} P_{\tilde{X}^n_{\mathcal{H}}}(x^n_{\mathcal{H}}) \log \frac{1}{P_{X^n_{\mathcal{H}}}(x^n_{\mathcal{H}})} \\ &= \sum_{x^n_{\mathcal{H}}} P_{\tilde{X}^n_{\mathcal{H}}}(x^n_{\mathcal{H}}) \cdot \left\{ \sum_{j=1}^{n} \log \frac{1}{P_{X_{\mathcal{H}}}(x_{\mathcal{H},j})} \right\} \\ &\overset{\text{(e)}}{=} n \sum_{x_{\mathcal{H}}} P_{\tilde{X}_{\mathcal{H},J}}(x_{\mathcal{H}}) \log \frac{1}{P_{X_{\mathcal{H}}}(x_{\mathcal{H}})} \\ &= n \sum_{x_{\mathcal{H}}} P_{\tilde{X}_{\mathcal{H},J}}(x_{\mathcal{H}}) \left\{ \log \frac{1}{P_{\tilde{X}_{\mathcal{H},J}}(x_{\mathcal{H}})} + \log \frac{P_{\tilde{X}_{\mathcal{H},J}}(x_{\mathcal{H}})}{P_{X_{\mathcal{H}}}(x_{\mathcal{H}})} \right\} \\ &= n\{H(\tilde{X}_{\mathcal{H},J}) + D(P_{\tilde{X}_{\mathcal{H},J}} \parallel P_{X_{\mathcal{H}}})\}, \end{aligned} \tag{139}$$

where

(e) holds because $\frac{1}{n}\sum_{j=1}^{n} P_{\tilde{X}_{\mathcal{H},j}}(x_{\mathcal{H}}) = P_{\tilde{X}_{\mathcal{H},J}}(x_{\mathcal{H}})$.

For the third term, it holds that

$$\begin{aligned} D(P_{\tilde{X}^n_{\mathcal{R}}|\tilde{X}^n_{\mathcal{H}}} \parallel P_{X^n_{\mathcal{R}}|X^n_{\mathcal{H}}} | P_{\tilde{X}^n_{\mathcal{H}}}) &= \sum_{j=1}^{n} D(P_{\tilde{X}_{\mathcal{R},j}|\tilde{X}^n_{\mathcal{H}}\tilde{X}^{j-1}_{\mathcal{R}}} \parallel P_{X_{\mathcal{R}}|X_{\mathcal{H}}} | P_{\tilde{X}^n_{\mathcal{H}}\tilde{X}^{j-1}_{\mathcal{R}}}) \\ &\overset{\text{(f)}}{\geq} \sum_{j=1}^{n} D(P_{\tilde{X}_{\mathcal{R},j}|\tilde{X}_{\mathcal{H},j}} \parallel P_{X_{\mathcal{R}}|X_{\mathcal{H}}} | P_{\tilde{X}_{\mathcal{H},j}}) \\ &\geq nD(P_{\tilde{X}_{\mathcal{R},J}|\tilde{X}_{\mathcal{H},J}} \parallel P_{X_{\mathcal{R}}|X_{\mathcal{H}}} | P_{\tilde{X}_{\mathcal{H},J}}), \end{aligned} \tag{140}$$

where

(f) follows from the log sum inequality.

From (139) and (140), we obtain (136). □

5.3. Strong Converse Theorem

We shall establish the strong converse theorem, which is the main result of this section. Before proving the theorem, we state the lemma of the key tool in the proof about a single-letterized $T^{\mu,\alpha}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ and a $T^{\mu,\alpha}_{\mathcal{E}}(P^n_{X_{\mathcal{K}}})$, which are introduced in Proposition 1.

Lemma 7. *For any $\mathcal{E}$ such that $\mathcal{R} \subseteq \mathcal{E} \subseteq \mathcal{K}$, all $n \in \mathbb{N}$, $\mu \geq 0$ and $\alpha > 0$, it holds that*

$$T^{\mu,\alpha}_{\mathcal{E}}(P^n_{X_{\mathcal{K}}}) \geq nT^{\mu,\alpha}_{\mathcal{E}}(P_{X_{\mathcal{K}}}).$$

As the main theorem of this section, we show the strong converse theorem for the utility–privacy trade-offs.

Theorem 4. *Strong converse theorem: For any $\mathcal{E}$ such that $\mathcal{R} \subseteq \mathcal{E} \subseteq \mathcal{K}$ and all $0 < \epsilon < 1$, it holds that*

$$\mathcal{L}^{DL}_{\mathcal{E}}(\epsilon|P_{X_{\mathcal{K}}}) = \mathcal{L}^{DL}_{\mathcal{E}}(P_{X_{\mathcal{K}}}).$$

Remark 5. *Theorem 4 suggests that regardless of the value of ϵ, the region $\mathcal{L}^{DL}_{\mathcal{E}}(\epsilon|P_{X_{\mathcal{K}}})$ is equal to $\mathcal{L}^{DL}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$.*

5.4. Proof of Lemma 7

Lemma 7 indicates that the function $T^{\mu,\alpha}_{\mathcal{E}}(P^n_{X_{\mathcal{K}}})$, whose argument $P^n_{X_{\mathcal{K}}}$ is a probability distribution over $\mathcal{X}^n_{\mathcal{K}}$, can be lower-bounded by the n-fold of a single-letterized function $T^{\mu,\alpha}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$. Before describing the detailed proof, we state the outline of the proof: (i) We first express the function $T^{\mu,\alpha}_{\mathcal{E}}(P^n_{X_{\mathcal{K}}})$ as the maximum of the difference of two functions denoted by G_1 and G_2 as in (142). (ii) Then, we show that the first function G_1 can be lower-bounded by the n-fold of its single-letterized function as in (143), while the second function G_2 can be upper-bounded by the n-fold of its single-letterized function as in (147). This outline of the proof is similar to the Proof of Theorem 4, 16 with a slight modification of the function G_2.

For a given distribution $P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}\hat{\tilde{X}}^n_{\mathcal{R}}}$, let functions $G_1(P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}})$ and $G_2(P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}\hat{\tilde{X}}^n_{\mathcal{R}}})$ be defined as

$$\begin{aligned} G_1(P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}}) &:= H(\tilde{X}^n_{\mathcal{H}}) + \alpha H(\tilde{X}^n_{\mathcal{E}^c}|\tilde{X}^n_{\mathcal{E}}) + (\alpha+1)D(P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}} \parallel P^n_{X_{\mathcal{E}^c}X_{\mathcal{E}}}), \\ G_2(P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}\hat{\tilde{X}}^n_{\mathcal{R}}}) &:= H(\tilde{X}^n_{\mathcal{H}}|\hat{\tilde{X}}^n_{\mathcal{R}}) - \mu\mathbb{E}[d(\tilde{X}^n_{\mathcal{R}}, \hat{\tilde{X}}^n_{\mathcal{R}})] + \alpha H(\tilde{X}^n_{\mathcal{E}}|\tilde{X}^n_{\mathcal{E}}, \hat{\tilde{X}}^n_{\mathcal{R}}). \end{aligned} \tag{141}$$

Using these functions, and in view of (134), $T^{\mu,\alpha}_{\mathcal{E}}(P^n_{X_{\mathcal{E}^c}X_{\mathcal{E}}})$ can be written as

$$T^{\mu,\alpha}_{\mathcal{E}}(P^n_{X_{\mathcal{E}^c}X_{\mathcal{E}}}) = \min_{P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}\hat{\tilde{X}}^n_{\mathcal{R}}}} \left[G_1(P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}}) - G_2(P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}\hat{\tilde{X}}^n_{\mathcal{R}}}) \right]. \tag{142}$$

For fixed $P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}\hat{\tilde{X}}^n_{\mathcal{R}}}$, from Proposition 2, it holds that

$$G_1(P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}}) \geq nG_1(P_{\tilde{X}_{\mathcal{E}^c,J}\tilde{X}_{\mathcal{E},J}}). \tag{143}$$

Next, we consider the function $G_2(P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}\hat{\tilde{X}}^n_{\mathcal{R}}})$. For the first term on the r.h.s. of (141), it holds that

$$\begin{aligned} H(\tilde{\mathrm{X}}^n_{\mathcal{H}}|\tilde{\hat{\mathrm{X}}}^n_{\mathcal{R}}) &= \sum_{j=1}^{n} H(\tilde{\mathrm{X}}_{\mathcal{H},j}|\tilde{\mathrm{X}}^{j-1}_{\mathcal{H}},\tilde{\hat{\mathrm{X}}}^n_{\mathcal{R}}) \\ &\leq \sum_{j=1}^{n} H(\tilde{\mathrm{X}}_{\mathcal{H},j}|\tilde{\hat{\mathrm{X}}}_{\mathcal{R},j}) \\ &= n\cdot\frac{1}{n}\sum_{j=1}^{n} H(\tilde{\mathrm{X}}_{\mathcal{H},j}|\tilde{\hat{\mathrm{X}}}_{\mathcal{R},j}) \\ &= nH(\tilde{\mathrm{X}}_{\mathcal{H},J}|\tilde{\hat{\mathrm{X}}}_{\mathcal{R},J},J) \\ &\leq nH(\tilde{\mathrm{X}}_{\mathcal{H},J}|\tilde{\hat{\mathrm{X}}}_{\mathcal{R},J}). \end{aligned} \tag{144}$$

The second term of (141) can be expressed as follows:

$$\begin{aligned} \mathbb{E}[d(\tilde{\mathrm{X}}^n_{\mathcal{R}},\tilde{\hat{\mathrm{X}}}^n_{\mathcal{R}})] &= \sum_{x^n_{\mathcal{R}},\hat{x}^n_{\mathcal{R}}} P_{\tilde{\mathrm{X}}^n_{\mathcal{R}}\tilde{\hat{\mathrm{X}}}^n_{\mathcal{R}}}(x^n_{\mathcal{R}},\hat{x}^n_{\mathcal{R}})\cdot\left\{\sum_{j=1}^{n} d(x_{\mathcal{R},j},\hat{x}_{\mathcal{R},j})\right\} \\ &= \sum_{j=1}^{n}\sum_{x_{\mathcal{R}},\hat{x}_{\mathcal{R}}} P_{\tilde{\mathrm{X}}_{\mathcal{R},j}\tilde{\hat{\mathrm{X}}}_{\mathcal{R},j}}(x_{\mathcal{R}},\hat{x}_{\mathcal{R}})d(x_{\mathcal{R}},\hat{x}_{\mathcal{R}}) \\ &\overset{\text{(a)}}{=} n\sum_{x_{\mathcal{R}},\hat{x}_{\mathcal{R}}} P_{\tilde{\mathrm{X}}_{\mathcal{R},J}\tilde{\hat{\mathrm{X}}}_{\mathcal{R},J}}(x_{\mathcal{R}},\hat{x}_{\mathcal{R}})d(x_{\mathcal{R}},\hat{x}_{\mathcal{R}}) \\ &= n\mathbb{E}[d(\tilde{\mathrm{X}}_{\mathcal{R},J},\tilde{\hat{\mathrm{X}}}_{\mathcal{R},J})], \end{aligned} \tag{145}$$

where

(a) follows from $\frac{1}{n}\sum_{j=1}^{n} P_{\tilde{\mathrm{X}}_{\mathcal{R},j}\tilde{\hat{\mathrm{X}}}_{\mathcal{R},j}}(x_{\mathcal{R}},\hat{x}_{\mathcal{R}}) = P_{\tilde{\mathrm{X}}_{\mathcal{R},J}\tilde{\hat{\mathrm{X}}}_{\mathcal{R},J}}(x_{\mathcal{R}},\hat{x}_{\mathcal{R}})$.

Moreover, for the third term of (141), it holds that

$$\begin{aligned} H(\tilde{\mathrm{X}}^n_{\mathcal{E}^c}|\tilde{\mathrm{X}}^n_{\mathcal{E}},\tilde{\hat{\mathrm{X}}}^n_{\mathcal{R}}) &= \sum_{j=1}^{n} H(\tilde{\mathrm{X}}_{\mathcal{E}^c,j}|\tilde{\mathrm{X}}^{j-1}_{\mathcal{E}^c},\tilde{\mathrm{X}}^n_{\mathcal{E}},\tilde{\hat{\mathrm{X}}}^n_{\mathcal{R}}) \\ &\leq \sum_{j=1}^{n} H(\tilde{\mathrm{X}}_{\mathcal{E}^c,j}|\tilde{\mathrm{X}}_{\mathcal{E},j},\tilde{\hat{\mathrm{X}}}_{\mathcal{R},j}) \\ &= n\cdot\frac{1}{n}\sum_{j=1}^{n} H(\tilde{\mathrm{X}}_{\mathcal{E}^c,j}|\tilde{\mathrm{X}}_{\mathcal{E},j},\tilde{\hat{\mathrm{X}}}_{\mathcal{R},j}) \\ &= nH(\tilde{\mathrm{X}}_{\mathcal{E}^c,J}|\tilde{\mathrm{X}}_{\mathcal{E},J},\tilde{\hat{\mathrm{X}}}_{\mathcal{R},J},J) \\ &\leq nH(\tilde{\mathrm{X}}_{\mathcal{E}^c,J}|\tilde{\mathrm{X}}_{\mathcal{E},J},\tilde{\hat{\mathrm{X}}}_{\mathcal{R},J}). \end{aligned} \tag{146}$$

From (144)–(146), we obtain

$$G_2(P_{\tilde{\mathrm{X}}^n_{\mathcal{E}^c}\tilde{\mathrm{X}}^n_{\mathcal{E}}\tilde{\hat{\mathrm{X}}}^n_{\mathcal{R}}}) \leq nG_2(P_{\tilde{\mathrm{X}}_{\mathcal{E}^c,J}\tilde{\mathrm{X}}_{\mathcal{E},J}\tilde{\hat{\mathrm{X}}}_{\mathcal{R},J}}). \tag{147}$$

Consequently, since (143) and (147) are satisfied for an arbitrary $P_{\tilde{\mathrm{X}}^n_{\mathcal{E}^c}\tilde{\mathrm{X}}^n_{\mathcal{E}}\tilde{\hat{\mathrm{X}}}^n_{\mathcal{R}}}$, the proof is completed.

5.5. Proof of Strong Converse Theorem

For any given $\epsilon > 0$, fix the rate pair $(D, L) \in \mathcal{L}^{DL}_{\mathcal{E}}(\epsilon|P_{\mathrm{X}_{\mathcal{K}}})$ arbitrarily. Then, by definition, there exists a code (f_n, g_n) satisfying (79) and (80). For this code (f_n, g_n), a set $\mathcal{D}$ is defined as

$$\mathcal{D} := \{(x^n_{\mathcal{E}^c}, x^n_{\mathcal{E}}):\ \ d(x^n_{\mathcal{R}}, g_n(f_n(x^n_{\mathcal{E}}))) \leq nD\}.$$

We derive a distribution $P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}}$ as

$$P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}}(x^n_{\mathcal{E}^c}, x^n_{\mathcal{E}}) := \frac{P^n_{X_{\mathcal{E}^c}X_{\mathcal{E}}}(x^n_{\mathcal{E}^c}, x^n_{\mathcal{E}})\mathbb{1}[(x^n_{\mathcal{E}^c}, x^n_{\mathcal{E}}) \in \mathcal{D}]}{P^n_{X_{\mathcal{E}^c}X_{\mathcal{E}}}(\mathcal{D})}.$$

It is obvious that the excess-distortion probability measured by $P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}}$ is 0; that is, $\tilde{X}^n_{\mathcal{R}}$ and $\tilde{\tilde{X}}^n_{\mathcal{R}} = g_n(f_n(\tilde{X}^n_{\mathcal{E}}))$ satisfy $\mathbb{E}[d(\tilde{X}^n_{\mathcal{R}}, \tilde{\tilde{X}}^n_{\mathcal{R}})] \le nD$. Thus, by imitating the proof approach of the standard weak converse theorem, it holds that

$$n(L + \mu D) \ge I(\tilde{X}^n_{\mathcal{H}}; \tilde{\tilde{X}}^n_{\mathcal{R}}) + \mu\mathbb{E}[d(\tilde{X}^n_{\mathcal{R}}, \tilde{\tilde{X}}^n_{\mathcal{R}})], \tag{148}$$

$$D(P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}} \parallel P^n_{X_{\mathcal{E}^c}X_{\mathcal{E}}}) = \log \frac{1}{P^n_{X_{\mathcal{E}^c}X_{\mathcal{E}}}(\mathcal{D})} \le \log \frac{1}{1-\epsilon}. \tag{149}$$

From (148), the following equation is obtained:

$$\begin{aligned} n(L + \mu D) &\overset{\text{(a)}}{\ge} I(\tilde{X}^n_{\mathcal{H}}; \tilde{\tilde{X}}^n_{\mathcal{R}}) + \mu\mathbb{E}[d(\tilde{X}^n_{\mathcal{R}}, \tilde{\tilde{X}}^n_{\mathcal{R}})] \\ &\quad + ((\alpha+1)D(P_{\tilde{X}^n_{\mathcal{E}^c}\tilde{X}^n_{\mathcal{E}}} \parallel P^n_{X_{\mathcal{E}^c}X_{\mathcal{E}}}) + \alpha I(\tilde{X}^n_{\mathcal{E}^c}; \tilde{\tilde{X}}^n_{\mathcal{R}} | \tilde{X}^n_{\mathcal{E}})) \\ &\quad - (\alpha+1)\log\frac{1}{1-\epsilon} \\ &\overset{\text{(b)}}{\ge} T^{\mu,\alpha}_{\mathcal{E}}(P^n_{X_{\mathcal{K}}}) - (\alpha+1)\log\frac{1}{1-\epsilon}, \end{aligned}$$

where

(a) follows from (149) and $I(\tilde{X}^n_{\mathcal{E}^c}; \tilde{\tilde{X}}^n_{\mathcal{R}} | \tilde{X}^n_{\mathcal{E}}) = 0$,
(b) is due to (134).

Since $T^{\mu,\alpha}_{\mathcal{E}}(P^n_{X_{\mathcal{K}}}) \ge nT^{\mu,\alpha}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ from Lemma 7, we have

$$L + \mu D \ge T^{\mu,\alpha}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) - \frac{(\alpha+1)}{n}\log\frac{1}{1-\epsilon},$$

and therefore

$$\sup_{\alpha>0}(L + \mu D) \ge \sup_{\alpha>0}\left[T^{\mu,\alpha}_{\epsilon}(P_{X_{\mathcal{K}}}) - \frac{(\alpha+1)}{n}\log\frac{1}{1-\epsilon}\right].$$

Because $T^{\mu}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) = \sup_{\alpha>0} T^{\mu,\alpha}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ from Proposition 1, it holds that for an arbitrary $\alpha > 0$,

$$L + \mu D \ge T^{\mu}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) - \frac{(\alpha+1)}{n}\log\frac{1}{1-\epsilon}.$$

Hence, it holds that

$$\begin{aligned} L + \mu D &\ge \lim_{n\to\infty}\left(T^{\mu}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) - \frac{(\alpha+1)}{n}\log\frac{1}{1-\epsilon}\right) \\ &= T^{\mu}_{\mathcal{E}}(P_{X_{\mathcal{K}}}) \qquad \text{for every } \mu \ge 0. \end{aligned} \tag{150}$$

For the set of (D, L) satisfying (150), varying $\mu \ge 0$ arbitrarily and taking the intersection, we have

$$(D, L) \in \bigcap_{\mu\ge 0}\{(D, L) :\ L + \mu D \ge T^{\mu}_{\mathcal{E}}(P_{X_{\mathcal{K}}})\}. \tag{151}$$

From Theorem 3, the r.h.s. of (151) is equal to $\mathcal{L}^{DL}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$. This proof is completed.

6. Discussion

6.1. Numerical Calculation of Coding Rate, Utility, and Privacy for Decoder

In this section, we show some numerical calculations of the achievable region $\mathcal{C}^{RDL}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ and $\mathcal{L}^{RDL}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ in Corollaries 1 and 2, respectively. In general, it is difficult to compute the achievable region $\mathcal{C}^{RDL}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ and $\mathcal{L}^{RDL}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$. Nevertheless, to obtain some insight, let us consider the three tractable but essential cases. In these calculations, the number of public attributes is one ($|\mathcal{R}| = 1$) and the number of private attributes is two ($|\mathcal{H}| = 2$). We assume that each of the attributes is binary. Here, note again that the coding rate R acts like the rate-distortion function in rate-distortion theory (cf. (Section 10 in [27])). For fixed D and L, a smaller coding rate is better.

In the first example, we calculated the L-D graph of theoretical limits in case (i) $\mathcal{E} = \mathcal{K}$, case (ii) $\mathcal{E} = \mathcal{R}$, and case (iii) $\mathcal{R} \subset \mathcal{E} \subset \mathcal{K}$ (Figure 4). As a result, the achievable privacy leakage L becomes small as D becomes large if we do not impose any restrictions on the value of R. For a given D, the privacy leakage for the decoder in case (i) $\mathcal{E} = \mathcal{K}$ is the smallest, and the one in case (ii) $\mathcal{E} = \mathcal{R}$ is the largest in all cases. The second example calculated the R-D graph of theoretical limits in cases (i), (ii), and (iii) (Figure 5). We can see that the minimum coding rates for a given D coincide in all cases if we do not impose any restrictions on the value of L. In the third example, we calculated the optimal privacy leakage L for fixed D and the corresponding coding rates R in cases (i), (ii), and (iii) (Tables 1–3). As a result, the optimal privacy leakage in cases (i) and (iii) is smaller than the one in case (ii), whereas for the optimal privacy leakage, the achievable coding rates in cases (i) and (iii) is larger than the one in case (ii).

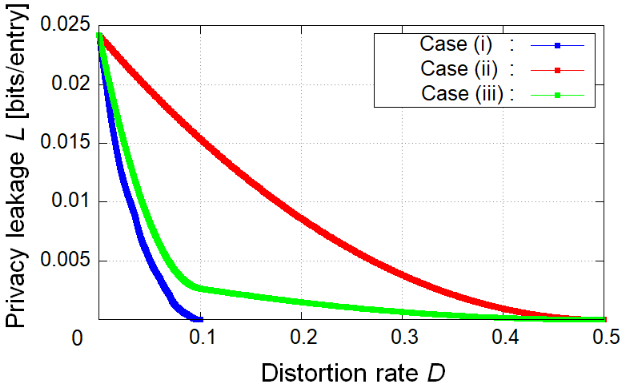

Figure 4. Utility–privacy trade-off region in cases (i), (ii), and (iii).

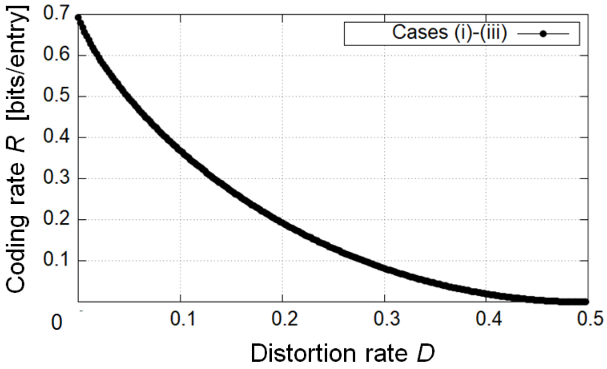

Figure 5. Utility–coding-rate trade-off region in cases (i), (ii), and (iii). The curves coincide in all cases.

Table 1. Minimum L and its corresponding R for $D = 0.0500$.

Cases	Leakage L	Coding Rate R
case (ii)	0.019512	0.494629
case (iii)	0.008298	0.527700
case (i)	0.005107	0.539478

Table 2. Minimum L and its corresponding R for $D = 0.100$.

Cases	Leakage L	Coding Rate R
case (ii)	0.015378	0.368062
case (iii)	0.002656	0.418826
case (i)	0.000000	0.429490

Table 3. Minimum L and its corresponding R for $D = 0.1500$.

Cases	Leakage L	Coding Rate R
case (ii)	0.011748	0.270436
case (iii)	0.002032	0.294424
case (i)	0.000000	0.382211

Next, we discuss these results. In Figure 4, in comparison with each case, we can verify that for a given D, the more private information is encoded, the smaller the achievable minimum privacy leakage is. Figure 5 suggests that if the coding rate should be minimized, it suffices to encode only the public attributes. This result is evident from Corollaries 1 and 2 because the condition on the choice of test channel $P_{\hat{X}_{\mathcal{R}}|X_{\mathcal{E}}}$ in case (i) is weaker than the one in case (ii), and if an appropriate test channel is taken in case (i), it is also appropriate in case (ii). It is indicated that the achievable region in case (ii) is also the one in cases (i) and (iii). The opposite is not the case. From Tables 1–3, we can confirm the trade-off between the optimal privacy leakage L for a fixed D and the corresponding coding rate R in comparison with each case.

Summarizing the foregoing arguments, we have discussed the relationship between utility and privacy in Figure 4, the one between utility and coding rate in Figure 5, and the one between privacy and coding rate in Tables 1–3. From the discussion about Figure 5, some readers may suspect that case (i) is the best-encoded information because the achievable region in cases (ii) and (iii) is the one in case (i). This is true if we do not consider the leakage for the encoder. However, this is not true if we consider the leakage for the encoder, that is, the measurement of privacy for the encoder (see (12) or (76)). In the next section, we discuss this point in detail.

6.2. Significance of Limited Leakage for Encoder

In this section, we discuss the significance of evaluating the leakage for the encoder. Our goal of this discussion is to show that the best-encoded information may be case (iii) $\mathcal{R} \subset \mathcal{E} \subset \mathcal{K}$ if we take the limited leakage for the encoder into consideration.

The first issue is the amount of encoded information. Some readers may think that it is better that more encoded information is inputted into the encoder. However, there are pros and cons.

Pros: The achievable regions $\mathcal{C}_{\mathcal{E}}^{RDL}(P_{X_{\mathcal{K}}})$ and $\mathcal{L}_{\mathcal{E}}^{RDL}(P_{X_{\mathcal{K}}})$ become larger.
Cons: The leakage for the encoder increases.

From this point of view, we can come up with the idea that there exists the best-encoded information in case (iii) $\mathcal{R} \subset \mathcal{E} \subset \mathcal{K}$ if we impose some constraint on the leakage for the encoder. This idea is the key point of this paper.

The second issue is the significance of the limited leakage for the encoder. Figure 6 shows the Hasse diagram, which represents the inclusion relation about the index sets of attributes. The Hasse diagram is often used to represent inclusion relations, for example, $\mathcal{R} \subset \mathcal{E}_2 \subset \mathcal{E}_1 \subset \mathcal{K}$.

We can also regard Figure 6 as the Hasse diagram that represents the inclusion relation for the achievable regions $\mathcal{C}_{\mathcal{E}}^{RDL}(P_{X_{\mathcal{K}}})$ and $\mathcal{L}_{\mathcal{E}}^{RDL}(P_{X_{\mathcal{K}}})$ because the index sets of attributes ($\mathcal{R} \subseteq \mathcal{E} \subseteq \mathcal{K}$) corresponds to the encoded information ($X_{\mathcal{E}}$) and the encoded information corresponds to the achievable region ($\mathcal{C}_{\mathcal{E}}^{RDL}(P_{X_{\mathcal{K}}})$ and $\mathcal{L}_{\mathcal{E}}^{RDL}(P_{X_{\mathcal{K}}})$). In addition, the diagram in Figure 6 has another property, which is that the superordinate sets have a larger amount of privacy leakage for the encoder than the subordinate sets since the index sets of attributes correspond to the privacy leakage for the encoder.

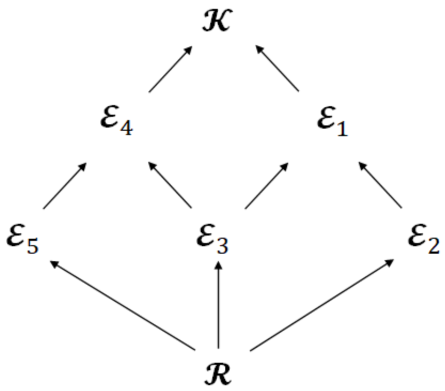

Figure 6. Hasse diagram that represents the inclusion relation for the index sets of attributes.

Let us consider a practical application. We assume that the data aggregator, that is, the encoder, tries to gather encoded information from some application user and hopes to develop the utility of the application while limiting the amount of leakage for $X_{\mathcal{H}}^n$ by $E \geq 0$, that is, $e_n \leq E$. More precisely, for a given E, we want to find which subsets of $\mathcal{K}$ are sufficient to characterize

$$\mathcal{C}^{RDL}(P_{X_{\mathcal{K}}}|E) := \bigcup_{\mathcal{R}\subseteq\mathcal{E}\subseteq\mathcal{K}} \{(R, D, L) : (R, D, L, E) \in \mathcal{C}_{\mathcal{E}}(P_{X_{\mathcal{K}}})\},$$
$$\mathcal{L}^{RDL}(P_{X_{\mathcal{K}}}|E) := \bigcup_{\mathcal{R}\subseteq\mathcal{E}\subseteq\mathcal{K}} \{(R, D, L) : (R, D, L, E) \in \mathcal{L}_{\mathcal{E}}(P_{X_{\mathcal{K}}})\},$$

where $\mathcal{C}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ and $\mathcal{L}_{\mathcal{E}}(P_{X_{\mathcal{K}}})$ are defined in Definitions 2 and 11, respectively. The process is as follows.

Step 1: Check the user's requirements and impose the restriction on the privacy leakage for the encoder.

Figure 7 shows the Hasse diagram for Step 1. The blue dotted line means the border line satisfies the restriction of the privacy leakage for the encoder. Therefore, the index sets $\mathcal{E}_1$ and $\mathcal{K}$ are not suitable as the index sets of encoded information.

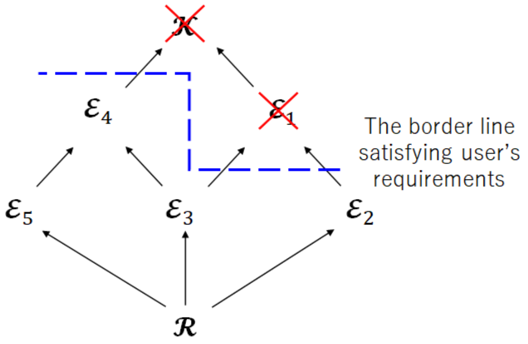

Figure 7. Hasse diagram for Step 1.

Step 2: Check the inclusion relation between index sets.

Figure 8 shows the Hasse diagram for Step 2. From Figure 6, we can find that

$$\mathcal{R} \subset \mathcal{E}_2, \;\; \mathcal{R} \subset \mathcal{E}_3, \;\; \mathcal{R} \subset \mathcal{E}_5, \;\; \mathcal{E}_3 \subset \mathcal{E}_4, \;\; \mathcal{E}_5 \subset \mathcal{E}_4.$$

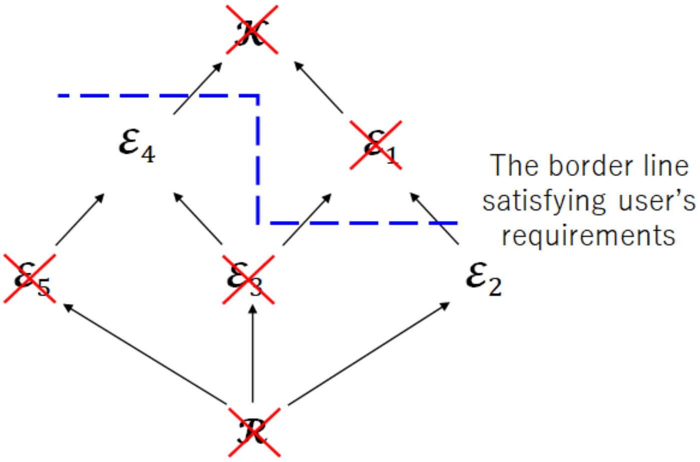

Figure 8. Hasse diagram for Step 2.

Therefore, the index sets $\mathcal{R}$, $\mathcal{E}_3$, and $\mathcal{E}_5$ are not suitable as the index sets of encoded information.

Figure 9 shows the Hasse diagram obtained after Step 2. From Figure 9, the remaining index sets are $\mathcal{E}_2$ and $\mathcal{E}_4$. Therefore, if we impose restriction on privacy leakage for the encoder, the index sets $\mathcal{E}_2$ or $\mathcal{E}_4$ form the Pareto area in this multi-objective optimization problem. In other words, there exists a system that satisfies the user's requirements E of the maximum amount of leakage to the encoder, and the achievable regions are given by $\mathcal{C}^{RDL}(P_{X_\mathcal{K}}|E) = \mathcal{C}^{RDL}_{\mathcal{E}_2}(P_{X_\mathcal{K}}) \cup \mathcal{C}^{RDL}_{\mathcal{E}_4}(P_{X_\mathcal{K}})$ and $\mathcal{L}^{RDL}(P_{X_\mathcal{K}}|E) = \mathcal{L}^{RDL}_{\mathcal{E}_2}(P_{X_\mathcal{K}}) \cup \mathcal{L}^{RDL}_{\mathcal{E}_4}(P_{X_\mathcal{K}})$.

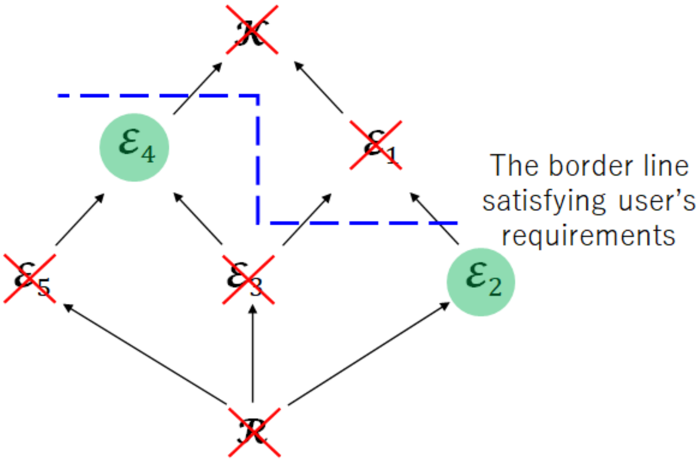

Figure 9. Hasse diagram obtained after Step 2.

From the discussion above, we mention that the best-encoded information is case (iii) $\mathcal{R} \subset \mathcal{E} \subset \mathcal{K}$ if we take the limited leakage for the encoder into account. This concept is one of the most important novelties in this paper.

If E satisfies some condition, then $\mathcal{C}^{RDL}(P_{X_\mathcal{K}}|E)$ can be characterized by the expressions given by Yamamoto [1] (cf. Remark 3). More specifically, the region $\mathcal{C}^{RDL}(P_{X_\mathcal{K}}|E)$ can be given by

$$\mathcal{C}^{RDL}(P_{X_\mathcal{K}}|E) = \mathcal{S}^{RDL}_{\mathcal{K}}(P_{X_\mathcal{K}})$$

if $E \geq H(X_{\mathcal{K}})$ and

$$\mathcal{C}^{RDL}(P_{X_{\mathcal{K}}}|E) = \mathcal{S}^{RDL}_{\mathcal{R}}(P_{X_{\mathcal{K}}})$$

if $H(X_{\mathcal{R}}) \leq E < H(X_{\mathcal{E}})$ for any $\mathcal{R} \subset \mathcal{E}$ with $\mathcal{R} \neq \mathcal{E}$, where the regions $\mathcal{S}^{RDL}_{\mathcal{K}}(P_{X_{\mathcal{K}}})$ and $\mathcal{S}^{RDL}_{\mathcal{R}}(P_{X_{\mathcal{K}}})$ are given in [1] (cf. Remark 3).

6.3. Discussion on Measures for Privacy Leakage

This paper adopts the mutual information as the measure of privacy leakage as in (12), (13), (76), and (77). However, some less likely data can be leaked even though the database satisfies the theoretical limit of privacy leakage. For example, let (X, Y) be a pair of correlated random variables whose $I(X;Y)$ is very small. However, there may exist a pair of (x_1, y_1) such that $Y = y_1$ can imply $X = x_1$ with high probability. To put it differently, the receiver can tell the value of X if it observes $Y = y_1$. The theoretical limit evaluated with mutual information cannot prevent such a scenario. To circumvent this scenario, we suggest the other measurement adopted in related studies. A promising candidate to avoid this problem is to employ Rényi information of higher orders [30], maximal leakage [15], and maximal α-leakage [16–18,21].

7. Conclusions

In this paper, we strengthened the results in [3] mainly by establishing three coding theorems in a privacy-constrained source coding problem. In Sections 3 and 4, two theorems are made about the first-order rate analysis in which utility is measured by the expected distortion or the excess-distortion probability for case (iii), $\mathcal{R} \subset \mathcal{E} \subset \mathcal{K}$. The novelty is the introduction of the measure of privacy for the encoder along with the use of the excess-distortion probability. The obtained characterization reduces to the one given in [3] derived based on the expected distortion when the leakage for the encoder is not limited, and the result shows that employing an excess-distortion probability does not change the achievable region from the one with an expected distortion. In Section 5, we establish the strong converse theorem for utility–privacy trade-offs. Although the described result is for the projected plane of utility and privacy for the decoder for simplicity, we can also incorporate the measure of privacy for the encoder. Finally, we discuss the significance of the encoded information considering limited leakage for the encoder. The argument suggests that the best-encoded information can be case (iii) $\mathcal{R} \subset \mathcal{E} \subset \mathcal{K}$ if some constraint is imposed on the privacy leakage for the encoder.

As future work, the second-order rate analysis for utility–privacy trade-offs is an interesting research topic [4–6]. Moreover, the strong converse theorem and the second-order rate analysis for the four-dimensional region of coding rate, utility, privacy for the decoder, and privacy for the encoder are more challenging tasks. It is also worth analyzing the achievable region with the other privacy measures such as Rényi information [30], maximal leakage [15], and maximal α-leakage [16–18,21]. This paper analyzed the theoretical limits of coding, but understanding how to achieve the theoretical limits remains open. The construction of good codes is also an important subject. Extensions of this paper's scenario to coding with side information [2,25] are also of interest.

Author Contributions: N.S. contributed to the conceptualization of the research goals and aims, the visualization, the formal analysis of the results, and the review and editing. H.Y. contributed to the conceptualization of the ideas, the validation of the results, and the supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by JSPS KAKENHI Grant Numbers JP20K04462 and JP18H01438.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of the Markov Chain $X_{\mathcal{E}^c}$–$X_{\mathcal{E}}$–$\hat{X}_{\mathcal{R}}$ in Converse Part of Theorem 1

Let $p_i(x_{\mathcal{E},i}, x_{\mathcal{E}^c,i}, \hat{x}_{\mathcal{R},i})$ be the conditional distribution given $Q = i$,

$$\begin{aligned}
p_i(x_{\mathcal{E},i}, x_{\mathcal{E}^c,i}, \hat{x}_{\mathcal{R},i}) &= \sum_{\substack{x_{\mathcal{E},k}:\\ k\neq i}} \sum_{\substack{x_{\mathcal{E}^c,k}:\\ k\neq i}} \sum_{\substack{\hat{x}_{\mathcal{R},k}:\\ k\neq i}} p(x^n_{\mathcal{E}}, x^n_{\mathcal{E}^c}, \hat{x}^n_{\mathcal{R}}) \\
&= \sum_{\substack{x_{\mathcal{E},k}:\\ k\neq i}} \sum_{\substack{x_{\mathcal{E}^c,k}:\\ k\neq i}} p(x^n_{\mathcal{E}}, x^n_{\mathcal{E}^c}, \hat{x}_{\mathcal{R},i}) \\
&\overset{(a)}{=} \sum_{\substack{x_{\mathcal{E},k}:\\ k\neq i}} \sum_{\substack{x_{\mathcal{E}^c,k}:\\ k\neq i}} p_i(x^n_{\mathcal{E}}, \hat{x}_{\mathcal{R},i}) p(x^n_{\mathcal{E}^c}|x^n_{\mathcal{E}}) \\
&= \sum_{\substack{x_{\mathcal{E},k}:\\ k\neq i}} p_i(x^n_{\mathcal{E}}, \hat{x}_{\mathcal{R},i}) \sum_{\substack{x_{\mathcal{E}^c,k}:\\ k\neq i}} p(x^n_{\mathcal{E}^c}|x^n_{\mathcal{E}}) \\
&\overset{(b)}{=} \sum_{\substack{x_{\mathcal{E},k}:\\ k\neq i}} p_i(x^n_{\mathcal{E}}, \hat{x}_{\mathcal{R},i}) \cdot \\
&\quad \sum_{\substack{x_{\mathcal{E}^c,k}:\\ k\neq i}} \left(\prod_{l=1}^{n} p(x_{\mathcal{E}^c,l}|x_{\mathcal{E},l}) \right) \\
&= p_i(x_{\mathcal{E},i}, \hat{x}_{\mathcal{R},i}) p(x_{\mathcal{E}^c,i}|x_{\mathcal{E},i}) \\
&= p(x_{\mathcal{E},i}) p(x_{\mathcal{E}^c,i}|x_{\mathcal{E},i}) p_i(\hat{x}_{\mathcal{R},i}|x_{\mathcal{E},i}) \\
&= p(x_{\mathcal{E},i}, x_{\mathcal{E}^c,i}) p_i(\hat{x}_{\mathcal{R},i}|x_{\mathcal{E},i}), \qquad \text{(A1)}
\end{aligned}$$

where

(a) is due to the Markov chain $X^n_{\mathcal{E}^c}$–$X^n_{\mathcal{E}}$–$\hat{X}_{\mathcal{R},i}$ and
(b) follows from the stationary memoryless source.

Therefore, we can obtain the Markov chain $X_{\mathcal{E}^c,i}$–$X_{\mathcal{E},i}$–$\hat{X}_{\mathcal{R},i}$. For the marginal distribution, we can show that

$$\begin{aligned}
p(x_{\mathcal{E}}, x_{\mathcal{E}^c}, \hat{x}_{\mathcal{R}}) &\overset{(c)}{=} \frac{1}{n} \sum_{i=1}^{n} p_i(x_{\mathcal{E}}, x_{\mathcal{E}^c}, \hat{x}_{\mathcal{R}}) \\
&\overset{(d)}{=} \frac{1}{n} \sum_{i=1}^{n} p_i(x_{\mathcal{E}}, x_{\mathcal{E}^c}) p_i(\hat{x}_{\mathcal{R}}|x_{\mathcal{E}}) \\
&\overset{(e)}{=} p(x_{\mathcal{E}}, x_{\mathcal{E}^c}) \cdot \frac{1}{n} \sum_{i=1}^{n} p_i(\hat{x}_{\mathcal{R}}|x_{\mathcal{E}}) \\
&\overset{(f)}{=} p(x_{\mathcal{E}}, x_{\mathcal{E}^c}) p(\hat{x}_{\mathcal{R}}|x_{\mathcal{E}}), \qquad \text{(A2)}
\end{aligned}$$

where

(c) follows because

$$p(x_{\mathcal{E}}, x_{\mathcal{E}^c}, \hat{x}_{\mathcal{R}}) = \sum_{i=1}^{n} \Pr\{Q = i\} p_i(x_{\mathcal{E}}, x_{\mathcal{E}^c}, \hat{x}_{\mathcal{R}}), \qquad \text{(A3)}$$

(d) is due to the Markov chain $X_{\mathcal{E}^c,i}$–$X_{\mathcal{E},i}$–$\hat{X}_{\mathcal{R},i}$,
(e) follows from the stationary memoryless source, and

(f) follows because

$$p(\hat{x}_{\mathcal{R}}|x_{\mathcal{E}}) = \sum_{i=1}^{n} \Pr\{Q = i\} p_i(\hat{x}_{\mathcal{R}}|x_{\mathcal{E}}). \tag{A4}$$

Therefore, we can obtain the Markov chain $X_{\mathcal{E}^c}$–$X_{\mathcal{E}}$–$\hat{X}_{\mathcal{R}}$. We complete the proof.

Appendix B. Proof of Equation (56)

From $\tilde{\mathcal{A}}(j) \subseteq \mathcal{B}(j)$ for $j = 1, 2, \ldots, M_n - 1$,

$$\Pr\{X^n_{\mathcal{K}} \in \mathcal{B}(j)\} = \Pr\{X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)\} + \Pr\{X^n_{\mathcal{K}} \in \mathcal{B}(j) \setminus \tilde{\mathcal{A}}(j)\}. \tag{A5}$$

If $x^n_{\mathcal{K}} \in \mathcal{B}(j) \setminus \tilde{\mathcal{A}}(j)$, then $x^n_{\mathcal{E}} \in T^n_\delta(X_{\mathcal{E}}|\hat{x}^n_{\mathcal{R}}(j))$ and $(x^n_{\mathcal{E}}, x^n_{\mathcal{E}^c}) \notin T^n_{2\delta}(X_{\mathcal{K}}|\hat{x}^n_{\mathcal{R}}(j))$, and thus we have $x^n_{\mathcal{E}^c} \notin T^n_\delta(X_{\mathcal{E}^c}|x^n_{\mathcal{E}}, \hat{x}^n_{\mathcal{R}}(j))$ from Lemma 5. Then,

$$\begin{aligned} x^n_{\mathcal{K}} \in \mathcal{B}(j) \setminus \tilde{\mathcal{A}}(j) \Longrightarrow & x^n_{\mathcal{E}} \in T^n_\delta(X_{\mathcal{E}}|\hat{x}^n_{\mathcal{R}}(j)), \\ & x^n_{\mathcal{E}^c} \notin T^n_\delta(X_{\mathcal{E}^c}|x^n_{\mathcal{E}}, \hat{x}^n_{\mathcal{R}}(j)) \end{aligned} \tag{A6}$$

We can prove that

$$\begin{aligned} &\Pr\{X^n_{\mathcal{K}} \in \mathcal{B}(j) \setminus \tilde{\mathcal{A}}(j)\} \\ &\le \Pr\{X^n_{\mathcal{E}} \in T^n_\delta(X_{\mathcal{E}}|\hat{x}^n_{\mathcal{R}}(j)), X^n_{\mathcal{E}^c} \notin T^n_\delta(X_{\mathcal{E}^c}|X^n_{\mathcal{E}}, \hat{x}^n_{\mathcal{R}}(j))\} \\ &= \sum_{x^n_{\mathcal{E}} \in T^n_\delta(X_{\mathcal{E}}|\hat{x}^n_{\mathcal{R}}(j))} \Pr\{X^n_{\mathcal{E}} = x^n_{\mathcal{E}}\} \cdot \\ &\quad \Pr\{X^n_{\mathcal{E}^c} \notin T^n_\delta(X_{\mathcal{E}^c}|x^n_{\mathcal{E}}, \hat{x}^n_{\mathcal{R}}(j)) | X^n_{\mathcal{E}} = x^n_{\mathcal{E}}\} \\ &\overset{\text{(a)}}{=} \sum_{x^n_{\mathcal{E}} \in T^n_\delta(X_{\mathcal{E}}|\hat{x}^n_{\mathcal{R}}(j))} \Pr\{X^n_{\mathcal{E}} = x^n_{\mathcal{E}}\} \cdot \\ &\quad \Pr\{X^n_{\mathcal{E}^c} \notin T^n_\delta(X_{\mathcal{E}^c}|x^n_{\mathcal{E}}, \hat{x}^n_{\mathcal{R}}(j)) | X^n_{\mathcal{E}} = x^n_{\mathcal{E}}, \hat{X}^n_{\mathcal{R}} = \hat{x}^n_{\mathcal{R}}(j)\} \\ &\overset{\text{(b)}}{\le} \sum_{x^n_{\mathcal{E}} \in T^n_\delta(X_{\mathcal{E}}|\hat{x}^n_{\mathcal{R}}(j))} \Pr\{X^n_{\mathcal{E}} = x^n_{\mathcal{E}}\} \cdot 2|\mathcal{X}_{\mathcal{E}^c}| \cdot |\mathcal{X}_{\mathcal{E}}| \cdot |\hat{\mathcal{X}}_{\mathcal{R}}| \mathrm{e}^{-2\delta^2 n} \\ &\le 2|\mathcal{X}_{\mathcal{K}}| \cdot |\hat{\mathcal{X}}_{\mathcal{R}}| \mathrm{e}^{-2\delta^2 n}, \end{aligned} \tag{A7}$$

where

(a) is due to the Markov chain $X^n_{\mathcal{E}^c}$–$X^n_{\mathcal{E}}$–$\hat{X}^n_{\mathcal{R}}$ and
(b) follows from Lemma 6.

From Equations (A5) and (A7), we can obtain

$$|\Pr\{X^n_{\mathcal{K}} \in \mathcal{B}(j)\} - \Pr\{X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)\}| \le 2|\mathcal{X}_{\mathcal{K}}| \cdot |\hat{\mathcal{X}}_{\mathcal{R}}| \mathrm{e}^{-2\delta^2 n}. \tag{A8}$$

We complete the proof of (56).

Appendix C. Proof of Existence of Code Satisfying Equations (57)–(62)

We first set $M_n := 2^{nR}$ and $r_n := \frac{1}{n} \log M_n$. Then, we obviously have (57). From the union upper bound,

$$\begin{aligned}&\Pr\left\{X_{\mathcal{E}}^n \notin \bigcup_{j=1}^{M_n-1}\mathcal{A}(j)\right\}\\ &\le \Pr\{X_{\mathcal{E}}^n \notin T_\delta^n(X_{\mathcal{E}})\}\\ &\quad+\Pr\{X_{\mathcal{E}}^n \in T_\delta^n(X_{\mathcal{E}}), X_{\mathcal{E}}^n \notin T_\delta^n(X_{\mathcal{E}}|\hat{x}_{\mathcal{R}}^n(j))\\ &\qquad\text{for all } j=1,2,\dots,M_n-1\}. \end{aligned}\tag{A9}$$

From Lemma 6, the first term in (A9) is bounded as

$$\Pr\{X_{\mathcal{E}}^n \notin T_\delta^n(X_{\mathcal{E}})\} \le 2|\mathcal{X}_{\mathcal{E}}|\mathrm{e}^{-2\delta^2 n}.\tag{A10}$$

We consider the expectation of the second term in (A9) by random coding. Hereafter, we denote the random variable corresponding to the reproduced sequence $\hat{x}_{\mathcal{R}}^n(j)$ as $\hat{X}_{\mathcal{R}}^n(j)$. For notational simplicity, we use the abbreviation

$$\begin{aligned}&\Pr\{X_{\mathcal{E}}^n \notin T_\delta^n(X_{\mathcal{E}}|\hat{X}_{\mathcal{R}}^n(j))\\ &\qquad\text{for all } j=1,2,\dots,M_n-1|X_{\mathcal{E}}^n=x_{\mathcal{E}}^n\}\\ &=\Pr\{x_{\mathcal{E}}^n \notin T_\delta^n(X_{\mathcal{E}}|\hat{X}_{\mathcal{R}}^n(j))\\ &\qquad\text{for all } j=1,2,\dots,M_n-1\},\end{aligned}\tag{A11}$$

and then

$$\begin{aligned}&\mathbb{E}[\Pr\{X_{\mathcal{E}}^n \in T_\delta^n(X_{\mathcal{E}}), X_{\mathcal{E}}^n \notin T_\delta^n(X_{\mathcal{E}}|\hat{X}_{\mathcal{R}}^n(j))\\ &\qquad\text{for all } j=1,2,\dots,M_n-1\}]\\ &=\sum_{x_{\mathcal{E}}^n\in T_\delta^n(X_{\mathcal{E}})} p(x_{\mathcal{E}}^n)\mathbb{E}\Big[\Pr\{X_{\mathcal{E}}^n \notin T_\delta^n(X_{\mathcal{E}}|\hat{X}_{\mathcal{R}}^n(j))\\ &\qquad\text{for all } j=1,2,\dots,M_n-1|X_{\mathcal{E}}^n=x_{\mathcal{E}}^n\}\Big]\\ &\overset{\text{(a)}}{=}\sum_{x_{\mathcal{E}}^n\in T_\delta^n(X_{\mathcal{E}})} p(x_{\mathcal{E}}^n)\mathbb{E}\Big[\Pr\{x_{\mathcal{E}}^n \notin T_\delta^n(X_{\mathcal{E}}|\hat{X}_{\mathcal{R}}^n(j))\\ &\qquad\text{for all } j=1,2,\dots,M_n-1\}\Big]\\ &=\sum_{x_{\mathcal{E}}^n\in T_\delta^n(X_{\mathcal{E}})} p(x_{\mathcal{E}}^n)\left(\prod_{j=1}^{M_n-1}\mathbb{E}[\Pr\{x_{\mathcal{E}}^n \notin T_\delta^n(X_{\mathcal{E}}|\hat{X}_{\mathcal{R}}^n(j))\}]\right)\\ &\overset{\text{(b)}}{=}\sum_{x_{\mathcal{E}}^n\in T_\delta^n(X_{\mathcal{E}})} p(x_{\mathcal{E}}^n)\big(\mathbb{E}[\Pr\{x_{\mathcal{E}}^n \notin T_\delta^n(X_{\mathcal{E}}|\hat{X}_{\mathcal{R}}^n(1))\}]\big)^{M_n-1}\\ &\overset{\text{(c)}}{\le}\exp\left\{-2^{n(R-I(X_{\mathcal{E}};\hat{X}_{\mathcal{R}})-\frac{1}{n}-\tau)}\right\}\\ &\overset{\text{(d)}}{\le}\exp\left\{-2^{2\delta^2 n}\right\},\end{aligned}\tag{A12}$$

where

(a) is owing to (A11),
(b) is due to the symmetry about indexes of random coding,
(c) follows from the same way as in (Section 3.6.3 in [31]), and
(d) because δ is fixed to satisfy (49).

From (A10) and (A12), we obtain

$$\mathbb{E}\left[\Pr\left\{X_{\mathcal{E}}^n \notin \bigcup_{j=1}^{M_n-1} \mathcal{A}(j)\right\}\right] \leq (2|\mathcal{X}_{\mathcal{E}}|+1)\mathrm{e}^{-2\delta^2 n}. \tag{A13}$$

Therefore, there exists at least one codebook satisfying (60) in the ensembles obtained by random coding.

Hereafter, codebook $\mathcal{C}$ is fixed to satisfy (60). That is, codebook $\mathcal{C}$ satisfies

$$\Pr\left\{X_{\mathcal{E}}^n \notin \bigcup_{j=1}^{M_n-1} \mathcal{A}(j)\right\} \leq (2|\mathcal{X}_{\mathcal{E}}|+1)\mathrm{e}^{-2\delta^2 n}. \tag{A14}$$

We evaluate the distortion function for each j.

(i) $j = 1, 2, \ldots, M_n - 1$:

$$\begin{aligned} & d(x_{\mathcal{R}}^n, \hat{x}_{\mathcal{R}}^n(j)) \\ & = \frac{1}{n} \sum_{a \in \mathcal{X}_{\mathcal{R}}} \sum_{b \in \hat{\mathcal{X}}_{\mathcal{R}}} N(a, b | x_{\mathcal{R}}^n, \hat{x}_{\mathcal{R}}^n(j)) d(a, b) \\ & \overset{\text{(e)}}{\leq} \sum_{a \in \mathcal{X}_{\mathcal{R}}} \sum_{b \in \hat{\mathcal{X}}_{\mathcal{R}}} P_{X_{\mathcal{R}}, \hat{X}_{\mathcal{R}}}(a, b) d(a, b) \\ & \quad + (\delta + \delta_1)|\mathcal{X}_{\mathcal{R}}| \cdot |\hat{\mathcal{X}}_{\mathcal{R}}| D_{\max} \\ & = \mathbb{E}[d(X_{\mathcal{R}}, \hat{X}_{\mathcal{R}})] + (\delta + \delta_1)|\mathcal{X}_{\mathcal{R}}| \cdot |\hat{\mathcal{X}}_{\mathcal{R}}| D_{\max}, \end{aligned} \tag{A15}$$

where

(e) because from Lemma 4, if $x_{\mathcal{E}}^n \in T_\delta^n(X_{\mathcal{E}} | \hat{x}_{\mathcal{R}}^n(j))$, then $x_{\mathcal{R}}^n \in T_{\delta_1}^n(X_{\mathcal{R}} | \hat{x}_{\mathcal{R}}^n(j))$ and from Lemma 3, if $\hat{x}_{\mathcal{R}}^n(j) \in T_\delta^n(\hat{X}_{\mathcal{R}})$ and $x_{\mathcal{R}}^n \in T_{\delta_1}^n(X_{\mathcal{R}} | \hat{x}_{\mathcal{R}}^n(j))$, then $(x_{\mathcal{R}}^n, \hat{x}_{\mathcal{R}}^n(j)) \in T_{\delta+\delta_1}^n(X_{\mathcal{R}}, \hat{X}_{\mathcal{R}})$.

(ii) $j = M_n$:

$$\begin{aligned} d(x_{\mathcal{R}}^n, \hat{x}_{\mathcal{R}}^n(M_n)) & = \frac{1}{n} \sum_{i=1}^n d(x_{\mathcal{R},i}, \hat{x}_{\mathcal{R},i}) \\ & \overset{\text{(f)}}{\leq} D_{\max}, \end{aligned} \tag{A16}$$

where

(f) is due to the definition of $D_{\max} := \max_{a \in \mathcal{X}_{\mathcal{R}}, b \in \hat{\mathcal{X}}_{\mathcal{R}}} d(a, b)$.

We consider $\Pr\{J_n = M_n\}$. From (A14),

$$\begin{aligned} \Pr\{J_n = M_n\} & = \Pr\{X_{\mathcal{E}}^n \in \mathcal{A}(M_n)\} \\ & = \Pr\left\{X_{\mathcal{E}}^n \notin \bigcup_{j=1}^{M_n-1} \mathcal{A}(j)\right\} \\ & \leq (2|\mathcal{X}_{\mathcal{E}}|+1)\mathrm{e}^{-2\delta^2 n}. \end{aligned} \tag{A17}$$

Therefore, we can confirm

$$\lim_{n \to \infty} \Pr\{J_n = M_n\} = 0. \tag{A18}$$

From (i) and (ii), we can evaluate utility u_n as below.

$$\begin{aligned} u_n &:= \mathbb{E}\big[d(X^n_{\mathcal{R}}, \hat{X}^n_{\mathcal{R}})\big] \\ &\le \sum_{j=1}^{M_n-1} \Pr\{J_n = j\} \cdot \Big(\mathbb{E}[d(X_{\mathcal{R}}, \hat{X}_{\mathcal{R}})] \\ &\quad + (\delta+\delta_1)|\mathcal{X}_{\mathcal{R}}| \cdot |\hat{\mathcal{X}}_{\mathcal{R}}| D_{\max}\Big) + \Pr\{J_n = M_n\} \cdot D_{\max} \\ &\overset{\text{(g)}}{\le} \mathbb{E}[d(X_{\mathcal{R}}, \hat{X}_{\mathcal{R}})] + (\delta+\delta_1)|\mathcal{X}_{\mathcal{R}}| \cdot |\hat{\mathcal{X}}_{\mathcal{R}}| D_{\max} + \tau \end{aligned} \tag{A19}$$

for all sufficiently large n, where

(g) follows from (A18).

Thus, we obtain (58).

We can evaluate the privacy leakage against the encoder as shown below.

$$\begin{aligned} e_n &:= \frac{1}{n} I(X^n_{\mathcal{H}}; X^n_{\mathcal{E}}) \\ &\overset{\text{(h)}}{=} \frac{1}{n} \sum_{i=1}^{n} I(X_{\mathcal{H},j}; X^n_{\mathcal{E}} | X^{j-1}_{\mathcal{H}}) \\ &\overset{\text{(i)}}{=} \frac{1}{n} \sum_{i=1}^{n} I(X_{\mathcal{H},j}; X_{\mathcal{E},j}) \\ &\overset{\text{(j)}}{=} I(X_{\mathcal{H}}; X_{\mathcal{E}}), \end{aligned} \tag{A20}$$

where

(h) is due to chain rule for mutual information and

(i), (j) follows because *i.i.d.* $P_{X^n_{\mathcal{K}}}$.

Thus, we have (59).

Next, we show that the probability that random vector $X^n_{\mathcal{K}}$ is not included in the set $\bigcup_{j=1}^{M_n-1} \tilde{\mathcal{A}}(j)$ is sufficiently small. First, notice that

$$\begin{aligned} x^n_{\mathcal{K}} \notin \bigcup_{j=1}^{M_n-1} \tilde{\mathcal{A}}(j) \Longrightarrow x^n_{\mathcal{E}} &\notin \bigcup_{j=1}^{M_n-1} \mathcal{A}(j) \\ &\text{or} \\ & x^n_{\mathcal{E}} \in \mathcal{A}(j_0), \\ & (x^n_{\mathcal{E}}, x^n_{\mathcal{E}^c}) \notin T^n_{2\delta}(X_{\mathcal{K}} | \hat{x}^n_{\mathcal{R}}(j_0)) \\ & \text{for } j_0 = f_n(x^n_{\mathcal{E}}), \end{aligned} \tag{A21}$$

where j_0 is the index such that $f_n(x^n_{\mathcal{E}}) = j_0$ for $1 \le j_0 \le M_n - 1$. Therefore, by the union upper bound,

$$\begin{aligned} &\Pr\left\{ X^n_{\mathcal{K}} \notin \bigcup_{j=1}^{M_n-1} \tilde{\mathcal{A}}(j) \right\} \\ &\le \Pr\left\{ X^n_{\mathcal{E}} \notin \bigcup_{j=1}^{M_n-1} \mathcal{A}(j) \right\} \\ &\quad + \Pr\{X^n_{\mathcal{E}} \in \mathcal{A}(j_0), (X^n_{\mathcal{E}}, X^n_{\mathcal{E}^c}) \notin T^n_{2\delta}(X_{\mathcal{K}} | \hat{x}^n_{\mathcal{R}}(j_0)) \\ &\qquad \text{for } j_0 = f_n(X^n_{\mathcal{E}})\}. \end{aligned} \tag{A22}$$

We evaluate each term in (A22).

(i) The first term:

$$\Pr\left\{ X_{\mathcal{E}}^n \notin \bigcup_{j=1}^{M_n-1} \mathcal{A}(j) \right\} \overset{\text{(k)}}{\leq} (2|\mathcal{X}_{\mathcal{E}}|+1)\mathrm{e}^{-2\delta^2 n}, \tag{A23}$$

where

(k) is because of (A14).

(ii) The second term:
If the event in the second term occurs, $x_{\mathcal{E}}^n \in T_\delta^n(X_{\mathcal{E}}|\hat{x}_{\mathcal{R}}^n(j_0))$ and $(x_{\mathcal{E}}^n, x_{\mathcal{E}^c}^n) \notin T_{2\delta}^n(X_{\mathcal{K}}|\hat{x}_{\mathcal{R}}^n(j_0))$. Therefore, from Lemma 5, $x_{\mathcal{E}^c}^n \notin T_\delta^n(X_{\mathcal{E}^c}|x_{\mathcal{E}}^n, \hat{x}_{\mathcal{R}}^n(j_0))$ holds. Hence,

$$\begin{aligned}
&\Pr\{X_{\mathcal{E}}^n \in \mathcal{A}(j_0), (X_{\mathcal{E}}^n, X_{\mathcal{E}^c}^n) \notin T_{2\delta}^n(X_{\mathcal{K}}|\hat{x}_{\mathcal{R}}^n(j_0)) \\
&\quad \text{for } j_0 = f_n(X_{\mathcal{E}}^n)\} \\
&\leq \Pr\{X_{\mathcal{E}}^n \in \mathcal{A}(j_0), X_{\mathcal{E}^c}^n \notin T_\delta^n(X_{\mathcal{E}^c}|X_{\mathcal{E}}^n, \hat{x}_{\mathcal{R}}^n(j_0))\} \\
&\leq \sum_{j=1}^{M_n-1} \sum_{x_{\mathcal{E}}^n \in \mathcal{A}(j)} \Pr\{X_{\mathcal{E}}^n = x_{\mathcal{E}}^n\} \cdot \\
&\quad \Pr\{X_{\mathcal{E}^c}^n \notin T_\delta^n(X_{\mathcal{E}^c}|x_{\mathcal{E}}^n, \hat{x}_{\mathcal{R}}^n(j)) | X_{\mathcal{E}}^n = x_{\mathcal{E}}^n\} \\
&\overset{\text{(l)}}{=} \sum_{j=1}^{M_n-1} \sum_{x_{\mathcal{E}}^n \in \mathcal{A}(j)} \Pr\{X_{\mathcal{E}}^n = x_{\mathcal{E}}^n\} \cdot \\
&\quad \Pr\{X_{\mathcal{E}^c}^n \notin T_\delta^n(X_{\mathcal{E}^c}|x_{\mathcal{E}}^n, \hat{x}_{\mathcal{R}}^n(j)) | X_{\mathcal{E}}^n = x_{\mathcal{E}}^n, \hat{X}_{\mathcal{R}}^n = \hat{x}_{\mathcal{R}}^n(j)\} \\
&\overset{\text{(m)}}{\leq} \sum_{j=1}^{M_n-1} \sum_{x_{\mathcal{E}}^n \in \mathcal{A}(j)} \Pr\{X_{\mathcal{E}}^n = x_{\mathcal{E}}^n\} \cdot 2|\mathcal{X}_{\mathcal{E}^c}| \cdot |\mathcal{X}_{\mathcal{E}}| \cdot |\hat{\mathcal{X}}_{\mathcal{R}}| \mathrm{e}^{-2\delta^2 n} \\
&\overset{\text{(n)}}{\leq} 2|\mathcal{X}_{\mathcal{K}}| \cdot |\hat{\mathcal{X}}_{\mathcal{R}}| \mathrm{e}^{-2\delta^2 n},
\end{aligned} \tag{A24}$$

where

(l) is due to the Markov chain $X_{\mathcal{E}^c}^n - X_{\mathcal{E}}^n - \hat{X}_{\mathcal{R}}^n$,
(m) follows since $x_{\mathcal{E}}^n \in T_\delta^n(X_{\mathcal{E}}|\hat{x}_{\mathcal{R}}^n(j_0))$ and Lemma 6, and
(n) follows because $\mathcal{A}(j)$ are disjoint for each j.

From (A22)–(A24),

$$\Pr\left\{ X_{\mathcal{K}}^n \notin \bigcup_{j=1}^{M_n-1} \tilde{\mathcal{A}}(j) \right\} \leq 4|\mathcal{X}_{\mathcal{K}}| \cdot |\hat{\mathcal{X}}_{\mathcal{R}}| \mathrm{e}^{-2\delta^2 n}. \tag{A25}$$

Therefore, for sufficiently large n,

$$\Pr\left\{ X_{\mathcal{K}}^n \notin \bigcup_{j=1}^{M_n-1} \tilde{\mathcal{A}}(j) \right\} \leq \tau, \tag{A26}$$

and we obtain (61).

From Lemma 1, for sufficiently large n to stochastic matrix $W: \hat{\mathcal{X}}_{\mathcal{R}} \to \mathcal{X}_{\mathcal{K}}$ and $\hat{x}_{\mathcal{R}}^n(j) \in T_\delta^n(\hat{X}_{\mathcal{R}})$ we can show that

$$\left| \frac{1}{n} \log |T_{\delta_2}^n(X_{\mathcal{K}}|\hat{x}_{\mathcal{R}}^n(j))| - H(X_{\mathcal{K}}|\hat{X}_{\mathcal{R}}) \right| \leq \tau, \tag{A27}$$

$$\delta_2 := \frac{\delta}{|\mathcal{X}_{\mathcal{E}^c}|}.$$

We can also show from (A27) that

$$2^{n\{H(X_{\mathcal{K}}|\hat{X}_{\mathcal{R}})-\tau\}} \le |T^n_{\delta_2}(X_{\mathcal{K}}|\hat{x}^n_{\mathcal{R}}(j))| \le 2^{n\{H(X_{\mathcal{K}}|\hat{X}_{\mathcal{R}})+\tau\}}. \tag{A28}$$

From the definition of $\tilde{\mathcal{A}}(j)$ and $T^n_{\delta_2}(X_{\mathcal{K}}|\hat{x}^n_{\mathcal{R}}(j))$ and Lemma 3, for $j = 1, 2, \ldots, M_n - 1$, we have

$$x^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j) \Longleftrightarrow \begin{cases} x^n_{\mathcal{E}} \in T^n_{\delta}(X_{\mathcal{E}}|\hat{x}^n_{\mathcal{R}}(j)) \\ x^n_{\mathcal{K}} \in T^n_{2\delta}(X_{\mathcal{K}}|\hat{x}^n_{\mathcal{R}}(j)) \end{cases} \tag{A29}$$

$$x^n_{\mathcal{K}} \in T^n_{\delta_2}(X_{\mathcal{K}}|\hat{x}^n_{\mathcal{R}}(j)) \Longrightarrow \begin{cases} x^n_{\mathcal{E}} \in T^n_{\delta}(X_{\mathcal{E}}|\hat{x}^n_{\mathcal{R}}(j)) \\ x^n_{\mathcal{K}} \in T^n_{2\delta}(X_{\mathcal{K}}|\hat{x}^n_{\mathcal{R}}(j)) \end{cases} \tag{A30}$$

This means

$$\begin{aligned} &T^n_{\delta_2}(X_{\mathcal{K}}|\hat{x}^n_{\mathcal{R}}(j)) \subseteq \tilde{\mathcal{A}}(j) \\ \Longrightarrow\ &|T^n_{\delta_2}(X_{\mathcal{K}}|\hat{x}^n_{\mathcal{R}}(j))| \le |\tilde{\mathcal{A}}(j)|. \end{aligned} \tag{A31}$$

Therefore, from (A28) and (A31),

$$|\tilde{\mathcal{A}}(j)| \ge 2^{n\{H(X_{\mathcal{K}}|\hat{X}_{\mathcal{R}})-\tau\}}, \tag{A32}$$

and we obtain (62).

Appendix D. Derivation of Inequality in Equation (63)

We derive the inequality in (63). To write notation concisely, for every $x^n_{\mathcal{H}} \in \mathcal{X}^n_{\mathcal{H}}$ and each $j = 1, 2, \ldots, M_n$, we define $P_n(j)$, $Q_n(j)$, $\tilde{P}_n(x^n_{\mathcal{H}}, j)$, and $\tilde{Q}_n(x^n_{\mathcal{H}}, j)$ as follows:

$$P_n(j) := \Pr\{X^n_{\mathcal{K}} \in \mathcal{B}(j)\}, \tag{A33}$$

$$Q_n(j) := \Pr\{X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)\}, \tag{A34}$$

$$\tilde{P}_n(x^n_{\mathcal{H}}, j) := \Pr\{X^n_{\mathcal{H}} = x^n_{\mathcal{H}}, X^n_{\mathcal{K}} \in \mathcal{B}(j)\}, \tag{A35}$$

$$\tilde{Q}_n(x^n_{\mathcal{H}}, j) := \Pr\{X^n_{\mathcal{H}} = x^n_{\mathcal{H}}, X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j)\}. \tag{A36}$$

Then, using the notation in [5], we can write each entropy as

$$H(X^n_{\mathcal{K}} \in \mathcal{B}(J_n)) = H(P_n), \tag{A37}$$

$$H(X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(J_n)) = H(Q_n), \tag{A38}$$

$$H(X^n_{\mathcal{H}}, X^n_{\mathcal{K}} \in \mathcal{B}(J_n)) = H(\tilde{P}_n), \tag{A39}$$

$$H(X^n_{\mathcal{H}}, X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(J_n)) = H(\tilde{Q}_n). \tag{A40}$$

The variational distance between distributions P_n and Q_n is

$$\begin{aligned} d_{\mathrm{v}}(P_n, Q_n) &= \sum_{j=1}^{M_n} |P_n(j) - Q_n(j)| \\ &= \sum_{j=1}^{M_n-1} |P_n(j) - Q_n(j)| \\ &\quad + |P_n(M_n) - Q_n(M_n)|. \end{aligned} \tag{A41}$$

We evaluate each term in (A41).

(i) The first term:

$$\begin{aligned}
&\sum_{j=1}^{M_n-1} |P_n(j) - Q_n(j)| \\
&= \sum_{j=1}^{M_n-1} \Pr\{X_{\mathcal{K}}^n \in \mathcal{B}(j) \setminus \tilde{\mathcal{A}}(j)\} \\
&\overset{(a)}{=} \Pr\left\{ X_{\mathcal{K}}^n \in \bigcup_{j=1}^{M_n-1} \mathcal{B}(j) \setminus \tilde{\mathcal{A}}(j) \right\} \\
&= \Pr\left\{ X_{\mathcal{K}}^n \in \bigcup_{j=1}^{M_n-1} \mathcal{B}(j) \right\} - \Pr\left\{ X_{\mathcal{K}}^n \in \bigcup_{j=1}^{M_n-1} \tilde{\mathcal{A}}(j) \right\} \\
&= \left(1 - \Pr\left\{ X_{\mathcal{K}}^n \in \bigcup_{j=1}^{M_n-1} \tilde{\mathcal{A}}(j) \right\} \right) \\
&\quad - \left(1 - \Pr\left\{ X_{\mathcal{K}}^n \in \bigcup_{j=1}^{M_n-1} \mathcal{B}(j) \right\} \right) \\
&= \Pr\left\{ X_{\mathcal{K}}^n \notin \bigcup_{j=1}^{M_n-1} \tilde{\mathcal{A}}(j) \right\} - \Pr\left\{ X_{\mathcal{K}}^n \notin \bigcup_{j=1}^{M_n-1} \mathcal{B}(j) \right\} \\
&\leq \Pr\left\{ X_{\mathcal{K}}^n \notin \bigcup_{j=1}^{M_n-1} \tilde{\mathcal{A}}(j) \right\} \\
&\overset{(b)}{\leq} \tau,
\end{aligned} \tag{A42}$$

where

(a) follows because $\mathcal{B}(j) \setminus \tilde{\mathcal{A}}(j)$ is disjoint for each $j = 1, 2, \ldots, M_n - 1$,

(b) is owing to (61).

(ii) The second term:

$$\begin{aligned}
|P_n(M_n) - Q_n(M_n)| &\overset{(c)}{=} Q_n(M_n) - P_n(M_n) \\
&\leq Q_n(M_n) \\
&= \Pr\left\{ X_{\mathcal{K}}^n \notin \bigcup_{j=1}^{M_n-1} \tilde{\mathcal{A}}(j) \right\} \\
&\overset{(d)}{\leq} \tau,
\end{aligned} \tag{A43}$$

where

(c) follows because $\mathcal{B}(M_n) \subseteq \tilde{\mathcal{A}}(M_n)$ and

(d) follows from (61).

From (A42) and (A43), the variational distance between P_n and Q_n is bounded from above as

$$\begin{aligned}
d_{\mathrm{v}}(P_n, Q_n) &\leq \tau + \tau \\
&= 2\tau.
\end{aligned} \tag{A44}$$

Next, the variational distance between distributions $\tilde{P}_n$ and $\tilde{Q}_n$ is

$$\begin{aligned} d_{\mathrm{v}}(\tilde{P}_n, \tilde{Q}_n) &= \sum_{j=1}^{M_n} \sum_{x_{\mathcal{H}}^n \in \mathcal{X}_{\mathcal{H}}^n} \left|\tilde{P}_n(x_{\mathcal{H}}^n, j) - \tilde{Q}_n(x_{\mathcal{H}}^n, j)\right| \\ &= \sum_{j=1}^{M_n-1} \sum_{x_{\mathcal{H}}^n \in \mathcal{X}_{\mathcal{H}}^n} \left|\tilde{P}_n(x_{\mathcal{H}}^n, j) - \tilde{Q}_n(x_{\mathcal{H}}^n, j)\right| \\ &\quad + \sum_{x_{\mathcal{H}}^n \in \mathcal{X}_{\mathcal{H}}^n} \left|\tilde{P}_n(x_{\mathcal{H}}^n, M_n) - \tilde{Q}_n(x_{\mathcal{H}}^n, M_n)\right|. \end{aligned} \tag{A45}$$

We evaluate each term in (A45).

(i) The first term:

$$\begin{aligned} &\sum_{j=1}^{M_n-1} \sum_{x_{\mathcal{H}}^n \in \mathcal{X}_{\mathcal{H}}^n} \left|\tilde{P}_n(x_{\mathcal{H}}^n, j) - \tilde{Q}_n(x_{\mathcal{H}}^n, j)\right| \\ &= \sum_{j=1}^{M_n-1} \sum_{x_{\mathcal{H}}^n \in \mathcal{X}_{\mathcal{H}}^n} \Pr\{X_{\mathcal{H}}^n = x_{\mathcal{H}}^n, X_{\mathcal{K}}^n \in \mathcal{B}(j) \setminus \tilde{\mathcal{A}}(j)\} \\ &\overset{\text{(e)}}{=} \sum_{x_{\mathcal{H}}^n \in \mathcal{X}_{\mathcal{H}}^n} \Pr\left\{ X_{\mathcal{H}}^n = x_{\mathcal{H}}^n, X_{\mathcal{K}}^n \in \bigcup_{j=1}^{M_n-1} \mathcal{B}(j) \setminus \tilde{\mathcal{A}}(j) \right\} \\ &= \Pr\left\{ X_{\mathcal{K}}^n \in \bigcup_{j=1}^{M_n-1} \mathcal{B}(j) \setminus \tilde{\mathcal{A}}(j) \right\} \\ &= \Pr\left\{ X_{\mathcal{K}}^n \in \bigcup_{j=1}^{M_n-1} \mathcal{B}(j) \right\} - \Pr\left\{ X_{\mathcal{K}}^n \in \bigcup_{j=1}^{M_n-1} \tilde{\mathcal{A}}(j) \right\} \\ &= \left(1 - \Pr\left\{ X_{\mathcal{K}}^n \in \bigcup_{j=1}^{M_n-1} \tilde{\mathcal{A}}(j) \right\} \right) \\ &\quad - \left(1 - \Pr\left\{ X_{\mathcal{K}}^n \in \bigcup_{j=1}^{M_n-1} \mathcal{B}(j) \right\} \right) \\ &= \Pr\left\{ X_{\mathcal{K}}^n \notin \bigcup_{j=1}^{M_n-1} \tilde{\mathcal{A}}(j) \right\} - \Pr\left\{ X_{\mathcal{K}}^n \notin \bigcup_{j=1}^{M_n-1} \mathcal{B}(j) \right\} \\ &\le \Pr\left\{ X_{\mathcal{K}}^n \notin \bigcup_{j=1}^{M_n-1} \tilde{\mathcal{A}}(j) \right\} \\ &\overset{\text{(f)}}{\le} \tau, \end{aligned} \tag{A46}$$

where

(e) follows since $\mathcal{B}(j) \setminus \tilde{\mathcal{A}}(j)$ is disjoint for each $j = 1, 2, \ldots, M_n - 1$,

(f) is due to (61).

(ii) The second term:

$$
\begin{aligned}
&\sum_{x_{\mathcal{H}}^n \in \mathcal{X}_{\mathcal{H}}^n} |\tilde{P}_n(x_{\mathcal{H}}^n, M_n) - \tilde{Q}_n(x_{\mathcal{H}}^n, M_n)| \\
&\overset{\text{(g)}}{=} \sum_{x_{\mathcal{H}}^n \in \mathcal{X}_{\mathcal{H}}^n} \left(\tilde{Q}_n(x_{\mathcal{H}}^n, M_n) - \tilde{P}_n(x_{\mathcal{H}}^n, M_n)\right) \\
&\leq \sum_{x_{\mathcal{H}}^n \in \mathcal{X}_{\mathcal{H}}^n} \tilde{Q}_n(x_{\mathcal{H}}^n, M_n) \\
&= Q_n(M_n) \\
&= \Pr\left\{ X_{\mathcal{K}}^n \notin \bigcup_{j=1}^{M_n - 1} \tilde{\mathcal{A}}(j) \right\} \\
&\overset{\text{(h)}}{\leq} \tau,
\end{aligned}
\tag{A47}
$$

where

(g) follows because $\mathcal{B}(M_n) \subseteq \tilde{\mathcal{A}}(M_n)$ and

(h) is due to (61).

From (A46) and (A47), the variational distance between $\tilde{P}_n$ and $\tilde{Q}_n$ is bounded from above as

$$
\begin{aligned}
d_{\text{v}}(\tilde{P}_n, \tilde{Q}_n) &\leq \tau + \tau \\
&= 2\tau.
\end{aligned}
\tag{A48}
$$

As a result, from Lemma 2 and the relation of each entropy,

$$
|H(X_{\mathcal{K}}^n \in \mathcal{B}(J_n)) - H(X_{\mathcal{K}}^n \in \tilde{\mathcal{A}}(J_n))| \leq -2\tau \log \frac{2\tau}{M_n}, \tag{A49}
$$

$$
\begin{aligned}
&|H(X_{\mathcal{H}}^n, X_{\mathcal{K}}^n \in \mathcal{B}(J_n)) - H(X_{\mathcal{H}}^n, X_{\mathcal{K}}^n \in \tilde{\mathcal{A}}(J_n))| \\
&\leq -2\tau \log \frac{2\tau}{|\mathcal{X}_{\mathcal{H}}|^n \cdot M_n}.
\end{aligned}
\tag{A50}
$$

From (A49), (A50), and the chain rule of entropy,

$$
\begin{aligned}
&|H(X_{\mathcal{H}}^n | X_{\mathcal{K}}^n \in \mathcal{B}(J_n)) - H(X_{\mathcal{H}}^n | X_{\mathcal{K}}^n \in \tilde{\mathcal{A}}(J_n))| \\
&= |\{H(X_{\mathcal{H}}^n, X_{\mathcal{K}}^n \in \mathcal{B}(J_n)) - H(X_{\mathcal{K}}^n \in \mathcal{B}(J_n))\} \\
&\quad - \{H(X_{\mathcal{H}}^n, X_{\mathcal{K}}^n \in \tilde{\mathcal{A}}(J_n)) - H(X_{\mathcal{K}}^n \in \tilde{\mathcal{A}}(J_n))\}| \\
&= |\{H(X_{\mathcal{H}}^n, X_{\mathcal{K}}^n \in \mathcal{B}(J_n)) - H(X_{\mathcal{H}}^n, X_{\mathcal{K}}^n \in \tilde{\mathcal{A}}(J_n))\} \\
&\quad + \{H(X_{\mathcal{K}}^n \in \tilde{\mathcal{A}}(J_n)) - H(X_{\mathcal{K}}^n \in \mathcal{B}(J_n))\}| \\
&\overset{\text{(i)}}{\leq} |H(X_{\mathcal{H}}^n, X_{\mathcal{K}}^n \in \mathcal{B}(J_n)) - H(X_{\mathcal{H}}^n, X_{\mathcal{K}}^n \in \tilde{\mathcal{A}}(J_n))| \\
&\quad + |H(X_{\mathcal{K}}^n \in \tilde{\mathcal{A}}(J_n)) - H(X_{\mathcal{K}}^n \in \mathcal{B}(J_n))| \\
&\leq -2\tau \log \frac{2\tau}{M_n} - 2\tau \log \frac{2\tau}{|\mathcal{X}_{\mathcal{H}}|^n \cdot M_n} \\
&= -4\tau \log \frac{2\tau}{|\mathcal{X}_{\mathcal{H}}|^n \cdot M_n} \\
&= 4\tau \log \frac{|\mathcal{X}_{\mathcal{H}}|^n \cdot M_n}{2\tau},
\end{aligned}
\tag{A51}
$$

where

(i) is because of the triangle inequality.

Therefore, we obtain

$$\begin{aligned}\frac{1}{n}H(X^n_{\mathcal{H}}|J_n) &= \frac{1}{n}H(X^n_{\mathcal{H}}|X^n_{\mathcal{K}} \in \mathcal{B}(J_n)) \\ &\geq \frac{1}{n}H(X^n_{\mathcal{H}}|X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(J_n)) - \frac{4\tau}{n}\log\frac{|\mathcal{X}_{\mathcal{H}}|^n \cdot M_n}{2\tau} \\ &\overset{\text{(j)}}{>} \frac{1}{n}H(X^n_{\mathcal{H}}|X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(J_n)) - \frac{4\tau}{n}\log\frac{|\mathcal{X}_{\mathcal{H}}|^n \cdot 2^{nR}}{(2\tau)^n} \\ &= \frac{1}{n}H(X^n_{\mathcal{H}}|X^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(J_n)) - 4\tau\log\frac{|\mathcal{X}_{\mathcal{H}}| \cdot 2^R}{2\tau},\end{aligned} \tag{A52}$$

where

(j) follows from the definition that $M_n = 2^{nR}$ and $2\tau < 1$.

We complete the derivation of (63).

Appendix E. Proof of Equation (65)

First of all, we shall show

$$\begin{aligned} x^n_{\mathcal{K}} \in \tilde{\mathcal{A}}(j) \Longrightarrow x^n_{\mathcal{R}} \in T^n_{\delta_3}(X_{\mathcal{R}}|x^n_{\mathcal{H}}, \hat{x}^n_{\mathcal{R}}(j)), \\ \delta_3 := (|\mathcal{X}_{\mathcal{H}}| + 1) \cdot 2\delta. \end{aligned} \tag{A53}$$

By the definition of $\tilde{\mathcal{A}}(j)$,

$$\tilde{\mathcal{A}}(j) \subseteq T^n_{2\delta}(X_{\mathcal{K}}|\hat{x}^n_{\mathcal{R}}(j)) \quad \text{for } j = 1, 2, \ldots, M_n - 1. \tag{A54}$$

Thus, from Lemma 4, any $x^n_{\mathcal{R}}$ such that $(x^n_{\mathcal{R}}, x^n_{\mathcal{H}}) \in \tilde{\mathcal{A}}(j)$ satisfies

$$x^n_{\mathcal{R}} \in T^n_{\delta_3}(X_{\mathcal{R}}|x^n_{\mathcal{H}}, \hat{x}^n_{\mathcal{R}}(j)). \tag{A55}$$

That is, given $x^n_{\mathcal{H}} \in \mathcal{X}^n_{\mathcal{H}}$ and $\hat{x}^n_{\mathcal{R}}(j) \in \hat{\mathcal{X}}^n_{\mathcal{R}}$, $x^n_{\mathcal{R}} \in \mathcal{X}^n_{\mathcal{R}}$ and $x^n_{\mathcal{K}} = (x^n_{\mathcal{R}}, x^n_{\mathcal{H}}) \in \tilde{\mathcal{A}}(j)$ are conditional strongly typical sequences. Then, we obtain (A53), and

$$\begin{aligned} &\sum_{x^n_{\mathcal{R}}:\, (x^n_{\mathcal{R}}, x^n_{\mathcal{H}}) \in \tilde{\mathcal{A}}(j)} \Pr\{X^n_{\mathcal{R}} = x^n_{\mathcal{R}}|X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\}\Pr\{X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\} \\ &\leq \sum_{x^n_{\mathcal{R}} \in T^n_{\delta_3}(X_{\mathcal{R}}|x^n_{\mathcal{H}}, \hat{x}^n_{\mathcal{R}}(j))} \Pr\{X^n_{\mathcal{R}} = x^n_{\mathcal{R}}|X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\}\Pr\{X^n_{\mathcal{H}} = x^n_{\mathcal{H}}\}. \end{aligned} \tag{A56}$$

Therefore, we obtain (65).

Appendix F. Proof of the Existence of Code Satisfying Equations (111)–(116)

We first set $M_n := 2^{nR}$ and $r_n := \frac{1}{n}\log M_n$. Then, we obviously have (111). From the union upper bound,

$$\begin{aligned} &\Pr\left\{ X^n_{\mathcal{E}} \notin \bigcup_{j=1}^{M_n - 1} \mathcal{A}(j) \right\} \\ &\leq \Pr\{X^n_{\mathcal{E}} \notin T^n_{\delta}(X_{\mathcal{E}})\} \\ &\quad + \Pr\{X^n_{\mathcal{E}} \in T^n_{\delta}(X_{\mathcal{E}}), X^n_{\mathcal{E}} \notin T^n_{\delta}(X_{\mathcal{E}}|\hat{x}^n_{\mathcal{R}}(j)) \\ &\qquad \text{for all } j = 1, 2, \ldots, M_n - 1\}. \end{aligned} \tag{A57}$$

From Lemma 6, the first term in (A57) is bounded as

$$\Pr\{X^n_{\mathcal{E}} \notin T^n_{\delta}(X_{\mathcal{E}})\} \leq 2|\mathcal{X}_{\mathcal{E}}|\mathrm{e}^{-2\delta^2 n}. \tag{A58}$$

We consider the expectation of the second term in (A57) by random coding. Hereafter, we denote the random variable corresponding to the reproduced sequence $\hat{x}^n_{\mathcal{R}}(j)$ as $\hat{X}^n_{\mathcal{R}}(j)$. For notational simplicity, we use the abbreviation

$$\begin{aligned}&\Pr\{X^n_{\mathcal{E}} \notin T^n_\delta(X_{\mathcal{E}}|\hat{X}^n_{\mathcal{R}}(j))\\ &\quad \text{for all } j = 1,2,\dots,M_n-1|X^n_{\mathcal{E}} = x^n_{\mathcal{E}}\}\\ &= \Pr\{x^n_{\mathcal{E}} \notin T^n_\delta(X_{\mathcal{E}}|\hat{X}^n_{\mathcal{R}}(j))\\ &\qquad \text{for all } j = 1,2,\dots,M_n-1\},\end{aligned} \tag{A59}$$

and then

$$\begin{aligned}&\mathbb{E}[\Pr\{X^n_{\mathcal{E}} \in T^n_\delta(X_{\mathcal{E}}), X^n_{\mathcal{E}} \notin T^n_\delta(X_{\mathcal{E}}|\hat{X}^n_{\mathcal{R}}(j))\\ &\quad \text{for all } j = 1,2,\dots,M_n-1\}]\\ &= \sum_{x^n_{\mathcal{E}} \in T^n_\delta(X_{\mathcal{E}})} p(x^n_{\mathcal{E}})\mathbb{E}\Big[\Pr\{X^n_{\mathcal{E}} \notin T^n_\delta(X_{\mathcal{E}}|\hat{X}^n_{\mathcal{R}}(j))\\ &\quad \text{for all } j = 1,2,\dots,M_n-1|X^n_{\mathcal{E}} = x^n_{\mathcal{E}}\}\Big]\\ &\overset{\text{(a)}}{=} \sum_{x^n_{\mathcal{E}} \in T^n_\delta(X_{\mathcal{E}})} p(x^n_{\mathcal{E}})\mathbb{E}\Big[\Pr\{x^n_{\mathcal{E}} \notin T^n_\delta(X_{\mathcal{E}}|\hat{X}^n_{\mathcal{R}}(j))\\ &\quad \text{for all } j = 1,2,\dots,M_n-1\}\Big]\\ &= \sum_{x^n_{\mathcal{E}} \in T^n_\delta(X_{\mathcal{E}})} p(x^n_{\mathcal{E}})\left(\prod_{j=1}^{M_n-1}\mathbb{E}[\Pr\{x^n_{\mathcal{E}} \notin T^n_\delta(X_{\mathcal{E}}|\hat{X}^n_{\mathcal{R}}(j))\}]\right)\\ &\overset{\text{(b)}}{=} \sum_{x^n_{\mathcal{E}} \in T^n_\delta(X_{\mathcal{E}})} p(x^n_{\mathcal{E}})\left(\mathbb{E}[\Pr\{x^n_{\mathcal{E}} \notin T^n_\delta(X_{\mathcal{E}}|\hat{X}^n_{\mathcal{R}}(1))\}]\right)^{M_n-1}\\ &\overset{\text{(c)}}{\le} \exp\Big\{-2^{n(R-I(X_{\mathcal{E}};\hat{X}_{\mathcal{R}})-\frac{1}{n}-\tau)}\Big\}\\ &\overset{\text{(d)}}{\le} \exp\Big\{-2^{2\delta^2 n}\Big\},\end{aligned} \tag{A60}$$

where

(a) is owing to (A59),
(b) is due to the symmetry about indexes of random coding,
(c) follows from the same way as in ([31], Section 3.6.3), and
(d) because δ is fixed to satisfy (103).

From (A58) and (A60), we obtain

$$\mathbb{E}\left[\Pr\left\{X^n_{\mathcal{E}} \notin \bigcup_{j=1}^{M_n-1}\mathcal{A}(j)\right\}\right] \le (2|\mathcal{X}_{\mathcal{E}}|+1)\mathrm{e}^{-2\delta^2 n}. \tag{A61}$$

Therefore, there exists at least one codebook satisfying (112) in the ensembles obtained using random coding.

Hereafter, codebook $\mathcal{C}$ is fixed to satisfy (112). That is, codebook $\mathcal{C}$ satisfies

$$\Pr\left\{X^n_{\mathcal{E}} \notin \bigcup_{j=1}^{M_n-1}\mathcal{A}(j)\right\} \le (2|\mathcal{X}_{\mathcal{E}}|+1)\mathrm{e}^{-2\delta^2 n}. \tag{A62}$$

For a fixed codebook $\mathcal{C}$, we divide the sequences $x^n_{\mathcal{E}} \in \mathcal{X}^n_{\mathcal{E}}$ into three categories.

- Strongly typical sequences $x^n_{\mathcal{E}} \in T^n_\delta(X_{\mathcal{E}})$ such that there exists a codeword $\hat{X}^n_{\mathcal{R}}(j_o)$ for some $j_o = 1,2,\dots,M_n-1$ that is conditionally strongly typical with $x^n_{\mathcal{E}}$. In this case,

from Lemma 3, $(x_{\mathcal{E}}, \hat{x}^n_{\mathcal{R}}(j_o)) \in T^n_{2\delta}(X_{\mathcal{E}}, \hat{X}_{\mathcal{R}}(j_o))$. Since the codeword is jointly strongly typical with $x^n_{\mathcal{E}}$, the continuity of the distortion as a function of the joint distribution ensures that they are also typical distortions (see [2], Chapters 10.5 and 10.6). Hence, the distortion between these $x^n_{\mathcal{E}}$ and their codewords is bounded by $D + \delta'$ where δ' goes to 0 as $n \to \infty$. In the first-order analysis, that is, $n \to \infty$, we can regard $D + \delta'$ as D.
- Strongly typical sequences $x^n_{\mathcal{E}} \in T^n_{\delta}(X_{\mathcal{E}})$ such that $f_n(x^n_{\mathcal{E}}) = M_n$.
- Non-strongly typical sequences $x^n_{\mathcal{E}} \notin T^n_{\delta}(X_{\mathcal{E}})$.

The sequences in the second and third categories are encoded as $f_n(x^n_{\mathcal{E}}) = M_n$. The sequences of third categories are the sequences that can be bounded by such the distortion $d_{\max}$ as in excess of D. Then, the excess-distortion probability is evaluated as

$$\Pr\left\{\frac{1}{n}d(X^n_{\mathcal{R}}, \hat{X}^n_{\mathcal{R}}) > D\right\} < \Pr\{X^n_{\mathcal{E}} \in \mathcal{A}(M_n)\} \tag{A63}$$

$$= \Pr\left\{X^n_{\mathcal{E}} \notin \bigcup_{j=1}^{M_n-1} \mathcal{A}(j)\right\} \tag{A64}$$

$$\leq (2|\mathcal{X}_{\mathcal{E}}| + 1)\mathrm{e}^{-2\delta^2 n}. \tag{A65}$$

Hence, for an appropriate choice of ϵ and n, we can ensure the excess-distortion probability of all badly represented sequences are as small as we want. We obtain (113).

We can evaluate privacy leakage against the encoder as below.

$$\begin{aligned} e_n &:= \frac{1}{n} I(X^n_{\mathcal{H}}; X^n_{\mathcal{E}}) \\ &\overset{\text{(e)}}{=} \frac{1}{n}\sum_{i=1}^{n} I(X_{\mathcal{H},j}; X^n_{\mathcal{E}} | X^{j-1}_{\mathcal{H}}) \\ &\overset{\text{(f)}}{=} \frac{1}{n}\sum_{i=1}^{n} I(X_{\mathcal{H},j}; X_{\mathcal{E},j}) \\ &\overset{\text{(g)}}{=} I(X_{\mathcal{H}}; X_{\mathcal{E}}), \end{aligned} \tag{A66}$$

where

(e) is due to chain rule for mutual information and
(f), (g) follows because *i.i.d.* $P_{X^n_{\mathcal{K}}}$.

Thus, we have (114).

Next, we show that the probability that random vector $X^n_{\mathcal{K}}$ is not included in the set $\bigcup_{j=1}^{M_n-1} \tilde{\mathcal{A}}(j)$ and is sufficiently small. First, notice that

$$\begin{aligned} x^n_{\mathcal{K}} \notin \bigcup_{j=1}^{M_n-1} \tilde{\mathcal{A}}(j) \Longrightarrow & x^n_{\mathcal{E}} \notin \bigcup_{j=1}^{M_n-1} \mathcal{A}(j) \\ & \text{or} \\ & x^n_{\mathcal{E}} \in \mathcal{A}(j_0), \\ & (x^n_{\mathcal{E}}, x^n_{\mathcal{E}^c}) \notin T^n_{2\delta}(X_{\mathcal{K}} | \hat{x}^n_{\mathcal{R}}(j_0)) \\ & \text{for } j_0 = f_n(x^n_{\mathcal{E}}), \end{aligned} \tag{A67}$$

where j_0 is the index such that $f_n(x^n_{\mathcal{E}}) = j_0$ for $1 \leq j_0 \leq M_n - 1$. Therefore, by the union's upper bound,

$$\begin{aligned}
&\Pr\left\{X^n_{\mathcal{K}} \notin \bigcup_{j=1}^{M_n-1} \tilde{\mathcal{A}}(j)\right\} \\
&\le \Pr\left\{X^n_{\mathcal{E}} \notin \bigcup_{j=1}^{M_n-1} \mathcal{A}(j)\right\} \\
&\quad + \Pr\{X^n_{\mathcal{E}} \in \mathcal{A}(j_0), (X^n_{\mathcal{E}}, X^n_{\mathcal{E}^c}) \notin T^n_{2\delta}(X_{\mathcal{K}}|\hat{x}^n_{\mathcal{R}}(j_0)) \\
&\qquad \text{for } j_0 = f_n(X^n_{\mathcal{E}})\}. \qquad \text{(A68)}
\end{aligned}$$

We evaluate each term in (A68).

(i) The first term:

$$\Pr\left\{X^n_{\mathcal{E}} \notin \bigcup_{j=1}^{M_n-1} \mathcal{A}(j)\right\} \overset{\text{(h)}}{\le} (2|\mathcal{X}_{\mathcal{E}}|+1)\mathrm{e}^{-2\delta^2 n}, \qquad \text{(A69)}$$

where

(h) is because of (A62).

(ii) The second term:
If the event in the second term occurs, $x^n_{\mathcal{E}} \in T^n_{\delta}(X_{\mathcal{E}}|\hat{x}^n_{\mathcal{R}}(j_0))$ and $(x^n_{\mathcal{E}}, x^n_{\mathcal{E}^c}) \notin T^n_{2\delta}(X_{\mathcal{K}}|\hat{x}^n_{\mathcal{R}}(j_0))$. Therefore, from Lemma 5, $x^n_{\mathcal{E}^c} \notin T^n_{\delta}(X_{\mathcal{E}^c}|x^n_{\mathcal{E}}, \hat{x}^n_{\mathcal{R}}(j_0))$ holds. Hence,

$$\begin{aligned}
&\Pr\{X^n_{\mathcal{E}} \in \mathcal{A}(j_0), (X^n_{\mathcal{E}}, X^n_{\mathcal{E}^c}) \notin T^n_{2\delta}(X_{\mathcal{K}}|\hat{x}^n_{\mathcal{R}}(j_0)) \\
&\quad \text{for } j_0 = f_n(X^n_{\mathcal{E}})\} \\
&\le \Pr\{X^n_{\mathcal{E}} \in \mathcal{A}(j_0), X^n_{\mathcal{E}^c} \notin T^n_{\delta}(X_{\mathcal{E}^c}|X^n_{\mathcal{E}}, \hat{x}^n_{\mathcal{R}}(j_0))\} \\
&\le \sum_{j=1}^{M_n-1} \sum_{x^n_{\mathcal{E}} \in \mathcal{A}(j)} \Pr\{X^n_{\mathcal{E}} = x^n_{\mathcal{E}}\} \cdot \\
&\quad \Pr\{X^n_{\mathcal{E}^c} \notin T^n_{\delta}(X_{\mathcal{E}^c}|x^n_{\mathcal{E}}, \hat{x}^n_{\mathcal{R}}(j))|X^n_{\mathcal{E}} = x^n_{\mathcal{E}}\} \\
&\overset{\text{(i)}}{=} \sum_{j=1}^{M_n-1} \sum_{x^n_{\mathcal{E}} \in \mathcal{A}(j)} \Pr\{X^n_{\mathcal{E}} = x^n_{\mathcal{E}}\} \cdot \\
&\quad \Pr\{X^n_{\mathcal{E}^c} \notin T^n_{\delta}(X_{\mathcal{E}^c}|x^n_{\mathcal{E}}, \hat{x}^n_{\mathcal{R}}(j))|X^n_{\mathcal{E}} = x^n_{\mathcal{E}}, \hat{X}^n_{\mathcal{R}} = \hat{x}^n_{\mathcal{R}}(j)\} \\
&\overset{\text{(j)}}{\le} \sum_{j=1}^{M_n-1} \sum_{x^n_{\mathcal{E}} \in \mathcal{A}(j)} \Pr\{X^n_{\mathcal{E}} = x^n_{\mathcal{E}}\} \cdot 2|\mathcal{X}_{\mathcal{E}^c}| \cdot |\mathcal{X}_{\mathcal{E}}| \cdot |\hat{\mathcal{X}}_{\mathcal{R}}|\mathrm{e}^{-2\delta^2 n} \\
&\overset{\text{(k)}}{\le} 2|\mathcal{X}_{\mathcal{K}}| \cdot |\hat{\mathcal{X}}_{\mathcal{R}}|\mathrm{e}^{-2\delta^2 n}, \qquad \text{(A70)}
\end{aligned}$$

where

(i) is due to the Markov chain $X^n_{\mathcal{E}^c} - X^n_{\mathcal{E}} - \hat{X}^n_{\mathcal{R}}$,
(j) follows since $x^n_{\mathcal{E}} \in T^n_{\delta}(X_{\mathcal{E}}|\hat{x}^n_{\mathcal{R}}(j_0))$ and Lemma 6,
(k) follows because $\mathcal{A}(j)$ is disjoint for each j.

From (A68)–(A70),

$$\Pr\left\{X^n_{\mathcal{K}} \notin \bigcup_{j=1}^{M_n-1} \tilde{\mathcal{A}}(j)\right\} \le 4|\mathcal{X}_{\mathcal{K}}| \cdot |\hat{\mathcal{X}}_{\mathcal{R}}|\mathrm{e}^{-2\delta^2 n}. \qquad \text{(A71)}$$

Therefore, for sufficiently large n,

$$\Pr\left\{ X_{\mathcal{K}}^n \notin \bigcup_{j=1}^{M_n-1} \tilde{\mathcal{A}}(j) \right\} \le \tau, \tag{A72}$$

and we obtain (115).

From Lemma 1, for sufficiently large n to stochastic matrix $W : \hat{\mathcal{X}}_{\mathcal{R}} \to \mathcal{X}_{\mathcal{K}}$ and $\hat{x}_{\mathcal{R}}^n(j) \in T_{\delta}^n(\hat{X}_{\mathcal{R}})$ we can show that

$$\left| \frac{1}{n} \log |T_{\delta_2}^n(X_{\mathcal{K}}|\hat{x}_{\mathcal{R}}^n(j))| - H(X_{\mathcal{K}}|\hat{X}_{\mathcal{R}}) \right| \le \tau, \tag{A73}$$
$$\delta_2 := \frac{\delta}{|\mathcal{X}_{\mathcal{E}^c}|}.$$

We can also show from (A73) that

$$2^{n\{H(X_{\mathcal{K}}|\hat{X}_{\mathcal{R}})-\tau\}} \le |T_{\delta_2}^n(X_{\mathcal{K}}|\hat{x}_{\mathcal{R}}^n(j))| \le 2^{n\{H(X_{\mathcal{K}}|\hat{X}_{\mathcal{R}})+\tau\}}. \tag{A74}$$

From the definition of $\tilde{\mathcal{A}}(j)$ and $T_{\delta_2}^n(X_{\mathcal{K}}|\hat{x}_{\mathcal{R}}^n(j))$ and Lemma 4, for $j = 1, 2, \ldots, M_n - 1$, we have

$$x_{\mathcal{K}}^n \in \tilde{A}(j) \Longleftrightarrow \begin{cases} x_{\mathcal{E}}^n \in T_{\delta}^n(X_{\mathcal{E}}|\hat{x}_{\mathcal{R}}^n(j)) \\ x_{\mathcal{K}}^n \in T_{2\delta}^n(X_{\mathcal{K}}|\hat{x}_{\mathcal{R}}^n(j)) \end{cases} \tag{A75}$$

$$x_{\mathcal{K}}^n \in T_{\delta_2}^n(X_{\mathcal{K}}|\hat{x}_{\mathcal{R}}^n(j)) \Longrightarrow \begin{cases} x_{\mathcal{E}}^n \in T_{\delta}^n(X_{\mathcal{E}}|\hat{x}_{\mathcal{R}}^n(j)) \\ x_{\mathcal{K}}^n \in T_{2\delta}^n(X_{\mathcal{K}}|\hat{x}_{\mathcal{R}}^n(j)) \end{cases} \tag{A76}$$

This means

$$\begin{aligned} & T_{\delta_2}^n(X_{\mathcal{K}}|\hat{x}_{\mathcal{R}}^n(j)) \subseteq \tilde{\mathcal{A}}(j) \\ \Longrightarrow\ & |T_{\delta_2}^n(X_{\mathcal{K}}|\hat{x}_{\mathcal{R}}^n(j))| \le |\tilde{\mathcal{A}}(j)|. \end{aligned} \tag{A77}$$

Therefore, from (A74) and (A77),

$$|\tilde{\mathcal{A}}(j)| \ge 2^{n\{H(X_{\mathcal{K}}|\hat{X}_{\mathcal{R}})-\tau\}}, \tag{A78}$$

and we obtain (116).

References

1. Yamamoto, H. A source coding problem for sources with additional outputs to keep secret from the receiver or wiretappers. *IEEE Trans. Inf. Theory* **1983**, *29*, 918–923. [CrossRef]
2. Sankar, L.; Rajagopalan, S.R.; Poor, H.V. Utility–Privacy tradeoff in databases: An information-theoretic approach. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 838–852. [CrossRef]
3. Shinohara, N.; Yagi, H. Unified expression of utility–privacy trade-off in privacy-constrained source coding. In Proceedings of the 2022 International Symposium on Information Theory and Its Applications (ISITA2022), Tsukuba, Japan, 17–19 October 2022; pp. 198–202.
4. Ingber, A.; Kochman, Y. The dispersion of lossy source coding. In Proceedings of the 2011 Data Compression Conference, Snowbird, UT, USA, 29–31 March 2011; pp. 53–62.
5. Kostina, V.; Verdú, S. Fixed length lossy compression in the finite blocklength regime: Discrete memoryless sources. *IEEE Trans. Inf. Theory* **2012**, *58*, 3309–3338. [CrossRef]
6. Watanabe, S. Second-order region for Gray-Wyner network. *IEEE Trans. Inf. Theory* **2017**, *63*, 1006–1018. [CrossRef]
7. Tyagi, H.; Watanabe, S. Strong converse using change of measure arguments. *IEEE Trans. Inf. Theory* **2020**, *66*, 689–703. [CrossRef]
8. Dwork, C.; McSherry, F.; Nissim, K.; Smith, A. Calibrating noise to sensitivity in private data analysis. In Proceedings of the 3rd Conference Theory Cryptograph (TCC), New York, NY, USA, 4–7 March 2006; pp. 265–284. [CrossRef]
9. Dwork, C. Differential privacy. In Proceedings of the 33rd International Conference Automata, Languages and Programming (ICALP), Venice, Italy, 10–14 July 2006; pp. 1–12. [CrossRef]

10. Soria-Comas, J.; Domingo-Ferrer, J.; Sánchez, D.; Megías, D. Individual differential privacy: A utility-preserving formulation of differential privacy guarantees. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 1418–1429.
11. Kalantari, K.; Sankar, L.; Sarwate, A.D. Robust privacy-utility tradeoffs under differential privacy and hamming distortion. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2816–2830.
12. Makhdoumi, A.; Fawaz, N. Privacy-utility tradeoff under statistical uncertainty. In Proceedings of the 2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 2–4 October 2013; pp. 1627–1634. [CrossRef]
13. Basciftci, Y.O.; Wang, Y.; Ishwar, P. On privacy-utility tradeoffs for constrained data release mechanisms. In Proceedings of the 2016 Information Theory and Applications Workshop (ITA), La Jolla, CA, USA, 31 January–5 February 2016; pp. 1–6. [CrossRef]
14. Günlü, O.; Schaefer, R.F.; Boche, H.; Poor, H.V. Secure and private source coding with private key and decoder side information. In Proceedings of the 2022 IEEE Information Theory Workshop (ITW), Mumbai, India, 6–9 November 2022; pp. 226–231.
15. Issa, I.; Wagner, A.B.; Kamath, S. An operational approach to information leakage. *IEEE Trans. Inf. Theory* **2020**, *66*, 1625–1657.
16. Liao, J.; Kosut, O.; Sankar, L.; Calmon, F.P. Privacy under hard distortion constraints. In Proceedings of the 2018 IEEE Information Theory Workshop (ITW2018), Guangzhou, China, 25–29 November 2018; pp. 1–5.
17. Liao, J.; Kosut, O.; Sankar, L.; Calmon, F.P. Tunable measures for information leakage and applications to privacy-utility tradeoffs. *IEEE Trans. Inf. Theory* **2019**, *65*, 8043–8066.
18. Saeidian, S.; Cervia, G.; Oechtering, T.J.; Skoglund, M. Quantifying membership privacy via information leakage. *IEEE Trans. Inf. Forensics Secur.* **2020**, *16*, 3096–3108.
19. Rassouli, B.; Gündüz, D. Optimal utility–privacy trade-off with total variation distance as a privacy measure. *IEEE Trans. Inf. Forensics Secur.* **2019**, *15*, 594–603.
20. Wang, W.; Ying, L.; Zhang, J. On the relation between identifiability, differential privacy, and mutual-information privacy. *IEEE Trans. Inf. Theory* **2016**, *62*, 5018–5029.
21. Liao, J.; Sankar, L.; Kosut, O.; Calmon, F.P. Maximal α-leakage and its properties. In Proceedings of the 2020 IEEE Conference on Communications and Network Security (CNS), Virtual, 29 June–1 July 2020; pp. 1–6.
22. Shinohara, N.; Yagi, H. Strong converse theorem for utility–privacy trade-offs. In Proceedings of the 45th Symposium on Information Theory and Its Applications (SITA2022), Noboribetsu, Japan, 29 November–2 December 2022; pp. 338–343. [CrossRef]
23. Guan, Z.; Si, G.; Wu, J.; Zhu, L.; Zhang, Z.; Ma, Y. Utility–privacy tradeoff based on random data obfuscation in internet of energy. *IEEE Access* **2017**, *5*, 3250–3262.
24. Asikis, T.; Pournaras, E. Optimization of privacy-utility trade-offs under informational self-determination. *Future Gener. Comput. Syst.* **2020**, *109*, 488–499. [CrossRef]
25. Lu, J.; Xu, Y.; Zhu, Z. On scalable source coding problem with side information privacy. In Proceedings of the 2022 14th International Conference on Wireless Communications and Signal Processing (WCSP), Nanjing, China, 1–3 November 2022; pp. 415–420. [CrossRef]
26. Makhdoumi, A.; Salamatian, S.; Fawaz, N.; Médard, M. From the information bottleneck to the privacy funnel. In Proceedings of the 2014 IEEE Information Theory Workshop (ITW), Hobart, Australia, 2–5 November 2014; pp. 501–505. [CrossRef]
27. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John & Wiley Sons, Inc.: Hoboken, NJ, USA, 2006. [CrossRef]
28. Uyematsu, T. *Gendai Shannon Riron*, 1st ed.; Baihukan: Tokyo, Japan, 1998. (In Japanese) [CrossRef]
29. Csizar, L.; Korner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2011.
30. Sason, I.; Verdú, S. Improved bounds on lossless source coding and guessing moments via Rényi measures. *IEEE Trans. Inf. Theory* **2018**, *64*, 4323–4346.
31. El Gamal, A.; Kim, Y.H. *Network Information Theory*, 1st ed.; Cambridge University Press: Cambridge, UK, 2011. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.