Privacy-Constrained Remote Source Coding

Kittipong Kittichokechai and Giuseppe Caire
Technische Universität Berlin

Abstract—We consider the problem of revealing/sharing data in an efficient and secure way via a compact representation. The representation should ensure reliable reconstruction of the desired features/attributes while still preserve privacy of the secret parts of the data. The problem is formulated as a remote lossy source coding with a privacy constraint where the remote source consists of public and secret parts. Inner and outer bounds for the optimal tradeoff region of compression rate, distortion, and privacy leakage rate are given and shown to coincide for some special cases. When specializing the distortion measure to a logarithmic loss function, the resulting rate-distortion-leakage tradeoff for the case of identical side information forms an optimization problem which corresponds to the “secure” version of the so-called information bottleneck.

I. INTRODUCTION

With the prominence of the Internet and the rise of the Internet of Things (IoT), significant amount of data are being generated, stored, and exchanged over the networks. Proper data management has become one of the most important and challenging aspects in system design. Information contained in the data are usually valuable resources that can be harnessed. However, the extensive use of data incurs some privacy risks especially when sensitive information is involved. The ultimate goal is to utilize the data to its full extent while still preserving privacy of the sensitive information.

Sankar et al. [1] and du Pin Calmon and Fawaz [2] studied the utility-privacy tradeoff from an information theoretic perspective, by relating to the framework of secure lossy source coding [3]–[8]. Inspired by these works, we consider a problem of secure remote source coding where the remote source consists of public and secret parts. Inner and outer bounds for the optimal tradeoff region of compression rate, distortion, and privacy leakage rate are given and shown to coincide for some special cases. When specializing the distortion measure to a logarithmic loss function, the resulting rate-distortion-leakage tradeoff for the case of identical side information forms an optimization problem which corresponds to the “secure” version of the so-called information bottleneck.

Fig. 1. Privacy-constrained remote lossy source coding.

In the same spirit as how the rate-distortion theorem under logarithmic loss [9]–[11] is related to the information bottleneck problem [12], in this work, the rate-distortion-privacy leakage tradeoff under logarithmic loss distortion corresponds to the secure version of information bottleneck. It extends the information bottleneck by including the privacy constraint or alternatively it extends the dual privacy funnel problem [11] by including the compression rate constraint. The secure remote source coding problem under logarithmic loss distortion therefore gives an operational meaning to the secure information bottleneck problem.

Our problem is closely related to works on lossy source coding with a privacy constraint. For instance, Yamamoto considered secure source coding with remote sources to be extracted and protected [3], [4]. Villard and Piantanida [7] considered lossy secure source coding where the source sequence is reconstructed at the decoder satisfying a distortion criterion while limiting the leakage of the source to an eavesdropper below a certain level. Tandon et al. [13] considered privacy of side information at the first decoder against a secondary decoder in the Heegard-Berger setting [14], [15] and considered one-way and interactive data sharing settings where the source consists of public and private parts and characterized the set of all achievable distortion-leakage pairs. The main differences of our work from [11] are the presence of an eavesdropper with correlated side information and the fact that public and private information are considered as remote sources.

Notation: We denote the discrete random variables, their corresponding realizations or deterministic values, and their alphabets by the upper case, lower case, and calligraphic letters, respectively. The term $X_n^m$ denotes the sequence $\{X_m, \ldots, X_n\}$ when $m \leq n$, and the empty set otherwise. Also, we use the shorthand notation $X^n$ for $X_1^n$. The term
$X^n$ denotes the set $\{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$. Cardinality of the set $\mathcal{X}$ is denoted by $|\mathcal{X}|$. Notation $[1:K]$ denotes the set $\{1, 2, \ldots, K\}$. Finally, we use $X - Y - Z$ to denote that $(X, Y, Z)$ forms a Markov chain. Other notations follow the standard ones in [10].

II. SECURE REMOTE SOURCE CODING

A. Problem Formulation

Let us consider a secure remote source coding shown in Fig. 1. Source and side information alphabets, $\mathcal{X}$, $\mathcal{Y}_p$, $\mathcal{Y}_s$, $\mathcal{Y}$, $\mathcal{Z}$ are finite sets. Let $(X^n, Y^n_p, Y^n_s, Z^n)$ be $n$-length sequences which have i.i.d. components distributed according to some fixed distribution $P_{X,Y_p,Y_s,Y,Z}$. The sequence $X^n$ represents the data to be revealed or shared. Public and secret attributes/features associated with the data (but not accessible/allowed to be processed directly) are represented by $Y^n_p$ and $Y^n_s$, respectively. The rate-limited description $W$ is generated based on $X^n$. The decoder reconstructs the public attribute of the data based on $W$ and correlated side information $Y^n_s$. For generality, we consider an eavesdropper which has access to the (publicly) stored description and another correlated side information $Z^n$. The secure remote source coding should ensure the reconstruction quality of the public attribute within a prescribed distortion, and at the same time preserve privacy of the secret part by limiting the amount of information leakage rate at the eavesdropper $\mathcal{L}(I(Y^n_s; W, Z^n))$. We note that if $Z = Y$, the problem reduces to the case where we impose a privacy constraint against the legitimate receiver.

Let $d: \mathbb{R} \times \mathbb{R} \rightarrow [0, \infty)$ be a distortion measure. The distortion between $Y^n_p$ and its reconstruction $\hat{Y}^n_p$ is defined as

$$d^{(n)}(Y^n_p, \hat{Y}^n_p) = \frac{1}{n} \sum_{i=1}^{n} d(Y_{p,i}, \hat{Y}_{p,i}).$$

We are interested in characterizing the optimal tradeoff of the compression rate, incurred average distortion at the legitimate decoder, and information leakage rate at the eavesdropper.

**Definition 1:** A $([\mathcal{W}^{(n)}], n)$-code for secure remote source coding consists of

- an encoder $f^{(n)}: \mathcal{X}^n \rightarrow \mathcal{W}^{(n)}$,
- a decoder $g^{(n)}: \mathcal{W}^{(n)} \times Y^n \rightarrow \hat{Y}^{(n)}_{p,s}$. \diamond

**Definition 2:** A rate-distortion-leakage tuple $(R, D, L) \in \mathbb{R}^3_+$ is said to be achievable if, for any $\delta > 0$ there exists a sequence of $([\mathcal{W}^{(n)}], n)$-codes such that, for all sufficiently large $n$,

$$\frac{1}{n} \log |\mathcal{W}^{(n)}| \leq R + \delta, \quad (1)$$

$$E[d^{(n)}(Y^n_p, g^{(n)}(W, Y^n_s))] \leq D + \delta, \quad (2)$$

$$\frac{1}{n} I(Y^n_s; W, Z^n) \leq L + \delta, \quad (3)$$

where $W = f^{(n)}(X^n)$. The rate-distortion-leakage region $\mathcal{R}$ is defined as the closure of the set of all achievable tuples. \diamond

B. Results

**Theorem 1 (Inner bound):** An inner bound to the rate-distortion-leakage region $\mathcal{R}_{in}$ is given as a set of all tuples $(R, D, L) \in \mathbb{R}_+^3$ satisfying

$$R \geq I(X; V | Y) \quad (4)$$

$$D \geq E[d(Y_p, g(V, Y))] \quad (5)$$

$$L \geq I(Y_s; V, Y) + I(Z; X, Y_s | U) - I(Y; X, Y_s | U) - I(X; Z | V, Y) + I(X; Y | Y_s, Z), \quad (6)$$

for some $P_{X,Y_p,Y_s,Y,Z}P_{V | X}P_{U | V}$ and $g: V \times Y \rightarrow \hat{Y}_p$ with $|U| \leq |X| + 3$ and $|V| \leq (|X| + 3)(|X| + 2)$.

**Proof:** The proof is given in Appendix A and is based on a random coding argument where the achievable scheme follows the layered (superposition) coding with binning in [7]. The main difference lies in the analysis of achievable leakage rate.

**Remark 1:** The constraint in (6) can be rewritten as

$$L \geq I(X, Y_s; U, V, Y) + I(Z; X, Y_s | U) - I(Y; X, Y_s | U) - I(X; Z | Y_s) - I(X; V | Y_s, Z),$$

in which the terms on the right-hand side may be interpreted as follows. The term $I(X, Y_s; U, V, Y)$ corresponds to the leakage of $(X, Y_s)$ through the description which depends on the remaining uncertainty at the decoder (which in this case can decode $(U, V)$ and knows $Y$). The terms $I(Z; X, Y_s | U) - I(Y; X, Y_s | U)$ is the additional leakage of $(X, Y_s)$ through the difference of side information available at the eavesdropper and decoder given that the codeword $U$ can be decoded at the eavesdropper. Since we are only interested in the leakage of $Y_s$, the remaining terms correspond to the leakage reduction of $X$ that is “orthogonal” to that of $Y_s$. The layered coding here provides some degree of freedom to optimize achievable leakage rate for our general setting. \diamond

Next, we provide an outer bound to the rate-distortion-leakage region.

**Theorem 2 (Outer bound):** An outer bound to the rate-distortion-leakage region $\mathcal{R}_{out}$ is given as a set of all tuples $(R, D, L) \in \mathbb{R}_+^3$ satisfying (4), (5), and

$$L \geq I(Y_s; V, Y) + I(Z; X, Y_s | U) - I(Y; X, Y_s | U) - I(X; Z | V, Y_s, Y) + I(X; Y | T, Y_s, Z), \quad (7)$$

for some $P_{X,Y_p,Y_s,Y,Z}P_{T,V | X}P_{U | V}$ and $g: V \times Y \rightarrow \hat{Y}_p$.

**Proof:** The proof is based on standard properties of the entropy function and the Csiszar’s sum identity [16] and is given in Appendix B. \diamond

**Remark 2:** The results in Theorems 1 and 2 can be extended to a scenario where the sequences $(X^n, Y^n_p, Y^n_s)$ are available directly at the encoder. In this case, we can replace $X$ by $(X, Y_p, Y_s)$ in Theorems 1 and 2 and the joint distributions become those of the form $P_{X,Y_p,Y_s,Y,Z}P_{V | X}P_{U | V}$ and $P_{X,Y_p,Y_s,Y,Z}P_{T,V | X}P_{U | V}$, respectively.

Theorems 1 and 2 can also be generalized to the case where $Y_p$ and $Y_s$ are not “disjoint,” i.e., they share some common part. For instance, the decoder may wish to reconstruct some
attributes associated with the data that are considered as secret to the eavesdropper. We simply modify the setup by replacing $Y_p$ by $(Y_p, Y_c)$ and $Y_s$ by $(Y_s, Y_c)$, where $Y_c$ acts as a common part that is supposed to be reconstructed at the decoder and protected against the eavesdropper. Theorems $1$ and $2$ continue to hold with $Y_p$ replaced by $(Y_p, Y_c)$ and $Y_s$ replaced by $(Y_s, Y_c)$ and the joint distribution of relevant source and side information is given by $P_X, Y_p, Y_c, Y_s, Z$.

We see that inner and outer bounds in Theorems $1$ and $2$ do not match in general. In particular, there is a gap between the leakage rate bounds. The difficulty of proving the tight bound lies in the complex dependency of information available at the eavesdropper and the secret remote source $Y_s^n$. Nevertheless, there exist some special cases where the bounds are tight.

**Corollary 1:** For the sources and side information whose joint distributions satisfy $I(X; Y_s, Z | Y) = 0$, the rate-distortion-leakage region $\mathcal{R}$ is given as a set of all tuples $(R, D, L) \in \mathbb{R}_+^3$ satisfying (4), (5), and $L \geq I(Y_s; Z)$ for some $P_{X, Y_s, Z}P_{V | X}$ and $g : V \times Y \rightarrow \hat{Y}_p$ with $|V| \leq |X| + 1$.

Interestingly, in this case, the only leakage of $Y_s^n$ is from correlated side information $Z^n$. There is no additional leakage rate from the source description $W$ sent over the rate-limited link. This can be explained as follows. Our achievability scheme is based on the binning technique which renders $I(W; Y^n) \leq n\delta$. Since we have the Markov chain $W - X^n - Y^n - (Y_s^n, Z^n)$, from the data processing inequality, the additional leakage rate of $Y_s^n$ due to $W$ becomes negligible, i.e., $I(Y_s^n; W | Z^n) \leq I(Y_s^n, Z^n; W) \leq I(Y_s^n; W) \leq n\delta$.

**Proof:** The proof follows from specializing Theorems $1$ and $2$ to the case where $X - Y - (Y_s, Z)$ forms a Markov chain, where in achievability, we choose $U = \emptyset$.

**Corollary 2:** For the sources and side information whose joint distributions satisfy $I(X; Y_s, Z | Y) = 0$, the rate-distortion-leakage region $\mathcal{R}$ is given as a set of all tuples $(R, D, L) \in \mathbb{R}_+^3$ satisfying (4), (5), and

$$L \geq I(Y_s; V, Y) + I(Z; X, Y_s | U) - I(Y; X) - I(Y_s | U)$$

for some $P_{X, Y_s, Z}P_{V | X}$ and $g : V \times Y \rightarrow \hat{Y}_p$ with $|U| \leq |X| + 3$ and $|V| \leq (|X| + 3)(|X| + 2)$.

**Proof:** The achievability proof follows directly from Theorem $1$ with $I(X; Y | Y, Z) = 0$, while the converse follows from Theorem $2$ and the fact that $I(X; Y | Y, Z) \geq 0$.

**Remark 3:** Corollaries $1$ and $2$ hold also for the case of stochastic encoder where the description $W$ is randomly generated according to a conditional PMF $p(w | x^n)$. This follows from the fact that in the proof of Theorem $2$ we do not make any assumption regarding the deterministic encoder.

**Remark 4:** We see that the rate-distortion-leakage region is known for several classes of source and side information, e.g., those satisfying $I(X; Y_s, Z | X) = 0$ in Corollary $2$ which includes also the semi-deterministic mapping with $X$ being a deterministic function of $Y_s$. Moreover, the result in Corollary $2$ recovers several existing results in the secure lossy source coding literature, e.g.,

- when $Y = Z$, we may think of the privacy leakage constraint as one imposed at the legitimate decoder. If $X = (Y_p, Y_s)$, Corollary $2$ recovers the result of the utility-privacy tradeoff with side information in $[1]$. Furthermore, if $X = (Y_p, Y_s)$ and $Y = Z = \emptyset$, Corollary $2$ recovers the result in $[1]$.
- when $X = Y_p = Y_s$, Corollary $2$ recovers the result of secure lossy source coding problem studied in $[7]$.
- when $Y = Y_s$, the leakage term becomes $I(Y_s^n; V, Z^n)$ which is of the same type as the side information privacy considered in $[13]$. The main difference is that in $[13]$ side information privacy is considered at the secondary receiver who observes no additional side information. If the reconstruction constraint at the secondary receiver is neglected, then zero leakage rate is achievable by the Wyner-Ziv coding $[17]$. Interestingly, in our case where the eavesdropper has access to the additional side information, the layered random binning scheme turns out to be optimal. An achievable leakage rate in this case is $I(Y_s; U, Z) + I(V; Z | Y, U)$ which is larger than $I(Y_s; Z)$. This is due to the fact that conditioned on $Z^n$, the Wyner-Ziv bin indices are still correlated with side information $Y_s^n$, and thus revealing some information about $Y_s^n$ to the eavesdropper.

**Corollary 3:** When we set $Y_p = X$ and consider a lossless reconstruction of $X^n$ at the decoder, an inner bound to the rate-distortion region is given by the set of all $(R, L) \in \mathbb{R}_+^2$ satisfying

$$R \geq H(X | Y)$$

$$L \geq I(Y_s; X, Y) + I(Z; X, Y_s | U) - I(Y; X, Y_s | U)$$

$$+ I(X; Y_s | Z),$$

for some $P_{X, Y_s, Z}P_{U | X}$ with $|U| \leq |X|$.

The inner bound above can be proved similarly as in Theorem $1$. In fact, it can be obtained from Theorem $1$ by setting $Y_p = X = \emptyset$.

**Remark 5:** We note that the special case of lossless reconstruction above was considered recently in $[18]$ where an inner bound to the rate-equivocation region is provided. In general, the results in Corollary $3$ and $[18]$ Theorem $3$ do not match. As an example where $Y = Y_s$ and $Z = \emptyset$, it can be shown that Corollary $3$ implies that zero leakage rate is achievable (by choosing $U = \emptyset$). However, the achievable leakage rate according to $[18]$ Theorem $3$ can be strictly positive.

**C. Quadratic Gaussian Example**

We consider an example of the tradeoff in Corollary $2$ for Gaussian sources under quadratic distortion. Assuming that $X \sim N(0, N_x)$, $Y_s = X + \tilde{N}_s$, where $\tilde{N}_s \sim N(0, N_s)$ ⊥ $X$, and $Y_p = Y_s + \tilde{N}_p$, where $\tilde{N}_p \sim N(0, N_p - N_s) ⊥ Y_s$. Note that $X - Y_s - Y_p$ forms a Markov chain. Also, we assume that there is no side information, i.e., $Y = Z = \emptyset$. The tradeoff region in Corollary $2$ reduces to the set of all
(R, D, L) satisfying
\[
R \geq \frac{1}{2} \log \left( \frac{N_x}{D - N_p} \right),
\]
\[
L \geq \frac{1}{2} \log \left( \frac{N_x + N_s}{D - N_p + N_s} \right),
\]
for \( D > N_p \).

While our main results were proven for discrete memoryless sources, the extension to the quadratic Gaussian case is standard and it follows, e.g., [19] and [16]. For achievability, we set \( U = \emptyset \), choose \( V \) to be jointly Gaussian with \( X \), i.e., \( V = X + Q \), \( Q \sim \mathcal{N}(0, N_p) \), and choose the reconstruction function \( g(\cdot) \) to be an MMSE estimate of \( Y_p \) given \( V \). By letting \( D = E[d(Y_p, g(V))] \) and substituting it into the constraints on \( R \) and \( L \), we obtain the result above. The converse follows from utilizing the EPI [16] together with the fact that \( R \) and \( L \) are decreasing in \( h(X^n|W) \) and \( h(Y^n_p|W) \), respectively. For the more detailed proof, please see Appendix C.

From (9) and (10), we can also write the minimum distortion as a function of \( R \) and \( L \), i.e.,
\[
D_{\text{min}}(R, L) = \max\{N_p + N_s 2^{-2R}, N_p - N_s + (N_x + N_s) 2^{-2L}\}.
\]

For a fixed \( R \), the minimum distortion \( D_{\text{min}} \) decreases with \( L \), illustrating the utility-privacy tradeoff in terms of minimum achievable distortion and information leakage rate.

III. LOGARITHMIC LOSS DISTORTION

Logarithmic loss [9], [10] is a measure of quality of the “soft” estimate used in several applications [20], [21]. Under logarithmic loss, we employ a soft estimate of the source in terms of a probability distribution over the source alphabet. For a sequence \( X^n \in X^n \), we denote \( \hat{X}_i \), \( i = 1, \ldots, n \), the \( i \)th element of \( X^n \). Then \( X_i, i = 1, \ldots, n \) is a probability distribution on \( X \), i.e., \( \hat{X}_i : X \rightarrow [0, 1] \), and \( \hat{X}_i(x) \) is a probability distribution on \( X \) evaluated for the outcome \( x \in X \).

Definition 3 (Logarithmic loss [10]): The logarithmic loss distortion is defined as \( d(x, \hat{x}) = \log(\frac{1}{\hat{X}_i(x)}) = D_{KL}(\hat{X}_i || \hat{x}) \), where \( \hat{X}_i : X \rightarrow [0, 1] \) is an indicator function such that, for \( a \in X \), \( \hat{X}_i(a) = 1 \) if \( a = x \), and \( \hat{X}_i(a) = 0 \) otherwise. Using this definition for symbol-wise distortion, it is standard to define the distortion between sequences as \( d(n)(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i) \).

Under logarithmic loss, the average distortion and conditional entropy (equivocation at decoder) are closely related (see, e.g., Lemma 1 below [10]). This is reminiscent of what is known, e.g., for the Gaussian settings under quadratic distortion.

Lemma 1: Let \( C = (W, Y^n) \) be the argument of the reconstruction function \( g^{(n)}(\cdot) \), i.e., \( X^n = g^{(n)}(C) \), then under the logarithmic loss distortion measure, we get \( E[d^{(n)}(X^n, g^{(n)}(C))] \geq \frac{1}{n} \sum_{i=1}^{n} H(X_i|C) \).

The proof of this lemma follows from definition of logarithmic loss, i.e., \( d(x_i, \hat{g}_i^{(n)}(c)) \triangleq \log(\frac{1}{\hat{g}_i(x_i)}) \) where \( q \) is a probability measure on \( X \). Then the expected distortion conditioned on \( C = c \),
\[
E[d^{(n)}(X^n, g^{(n)}(c))|C = c] = E\left[\frac{1}{n} \sum_{i=1}^{n} d(x_i, g_i^{(n)}(c)) | C = c\right]
\]
\[
= \frac{1}{n} \sum_{x^n \in X^n} p(x^n|c) \frac{1}{n} \sum_{i=1}^{n} d(x_i, g_i^{(n)}(c))
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} \sum_{x_i \in X} p(x_i|c) \log \left( \frac{1}{q(x_i|c)} \right)
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} \sum_{x_i \in X} p(x_i|c) \log \left( \frac{p(x_i|c)}{q(x_i|c)} \right) \frac{1}{p(x_i|c)}
\]
\[
= \frac{1}{n} \sum_{i=1}^{n} D_{KL}(p(x_i|c) || q(x_i|c)) + \frac{1}{n} \sum_{i=1}^{n} H(X_i|C = c)
\]
\[
\geq \frac{1}{n} \sum_{i=1}^{n} H(X_i|C = c).
\]

By averaging both sides over all \( c \in C \), from the law of total expectation, the lemma is proved.

We note that Lemma 1 holds only for the case of symbol-by-symbol logloss distortion, i.e., \( d^{(n)}(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i) \), which is considered in the problem formulation.

Corollary 4: Consider the privacy-constrained remote source coding in Fig. 1. When the decoder and eavesdropper observe identical side information, i.e., \( Y = Z \) (can be seen as privacy leakage against the legitimate receiver), the rate-distortion-leakage region under logarithmic loss is given as a set of all tuples \( (R, D, L) \in \mathbb{R}_+^3 \) satisfying
\[
R \geq I(X; V) \quad \text{(11)}
\]
\[
D \geq H(Y_p|V, Y) \quad \text{(12)}
\]
\[
L \geq I(Y_p; V, Y) \quad \text{(13)}
\]

for some \( P_{X,Y_p,Y,V,X} \). By changing \( |\mathcal{Y}| = |X| + 2 \).

Proof: For achievability, we apply Corollary 1 under logarithmic loss distortion with \( Z = Y \). We choose \( g(\cdot) \) to be a conditional probability distribution on \( Y_p \), i.e., \( g(y,p|v,y) = p(y_p|v,y) \) which gives \( E[d(Y_p, g(V,Y))] = H(Y_p|V,Y) \).

The converse follows by setting \( V_i = (W, Y^n \setminus V) \) and applying Lemma 1 which gives \( E[d^{(n)}(Y^n_p, g^{(n)}(W, Y^n)))] = \frac{1}{n} \sum_{i=1}^{n} E[d(Y_p, g^{(n)}(W, Y^n))] \geq \frac{1}{n} \sum_{i=1}^{n} H(Y_p, |W, Y^n). \)

A. Secure Information Bottleneck

Based on Corollary 4, we can formulate an optimization problem where for a given \( P_{X,Y_p,Y,V} \), we wish to minimize the rate \( I(X; V|Y) \) over all \( P_{V|X} \) subject to the constraints.

1The proof of Theorem 1 holds for bounded distortion measures. However, it can still be extended to logarithmic loss distortion by perturbing the reconstruction probability distribution (see e.g., [22] Remark 3.4).
on the distortion and information leakage, i.e.,

\[
\begin{align*}
\min_{P_{V|X}} & \quad I(X; V | Y) \\
\text{s.t.} & \quad I(Y'_p; V, Y) \geq D' \\
& \quad I(Y'_s; V, Y) \leq L,
\end{align*}
\]

where in (15), the distortion constraint is rewritten as the “information” constraint with \(D' \triangleq H(Y'_p) - D\). The minimum above corresponds to the minimum achievable rate for fixed \(D'\) and \(L\) and may be termed as the rate-information-leakage function \(R_{\min}(D', L)\). We can see that the optimization problem is not convex, e.g., \(I(Y'_p; V, Y)\) is a convex function in \(P_{V|X}\) for a fixed \(P_{X,Y,Y'_p,Y'_s}\). Interestingly, the problem shares some similarity with the information bottleneck problem [12] where “information” constraint with \(\lambda\) may be termed as the rate-information-leakage function \(R_{\min}(D', L)\) is given by

\[
R_{\min}(D', L) = \log \left( \frac{\mathbb{E}[\exp(\frac{1}{\lambda} D)]}{\mathbb{E}[\exp(\frac{1}{\lambda} \mathbb{E}[D|V])]} \right)
\]

for which the maximum in (18) corresponds to the maximum achievable information for given \(R\) and \(L\) and may be termed as the information-rate-leakage function \(D'_{\max}(R, L)\).

In the following, we consider two simple examples under the assumptions that \(Y = \emptyset\) and \(X = Y'_p - Y'_s\) forms a Markov chain, in which we can express \(D'_{\max}(R, L)\) in closed form.

(i) Binary source: Let \(X \sim \text{Bernoulli}(1/2)\), \(Y_s\) be an output of a BSC\((p)\), \(p \in [0, 1/2]\), with input \(X\), and \(Y_p\) be an output of a BSC\((q)\), \(q \in [0, 1/2]\), with input \(Y_s\). The maximum achievable information for given \(R \in [0, H(X)]\) and \(L \in [0, H(Y_s)]\) is given by

\[
D'_{\max}(R, L) = H(Y_p) - \max \left\{ h(h^{-1}(H(Y_s) - L) \ast q), h(h^{-1}(H(X) - R) \ast p \ast q) \right\},
\]

where \(h(\cdot)\) is a binary entropy function with the inverse \(h^{-1}: [0, 1] \rightarrow [0, 1/2]\), and \(a \ast b \triangleq a(1 - b) + (1 - a)b\).

The achievability proof follows by letting \(V\) be an output of a BSC with input \(X\). Then letting \(D' = I(Y'_p; V) = H(Y_p) - h(h^{-1}(H(X|V)) \ast p \ast q) = H(Y_p) - h(h^{-1}(H(Y_s|V)) \ast q)\), we obtain the result above by substituting \(D'\) in the constraints on \(R\) and \(L\) and using the fact that \(h(h^{-1}(u) \ast q)\) is an increasing function in \(u\) for \(q \in [0, 1/2]\). The converse follows from Mrs. Gerber’s lemma [16]. For the more detailed proof, please see Appendix [12].

(ii) Gaussian source: Let \((X, Y_s, Y_p)\) be jointly Gaussian, i.e., \(X \sim \mathcal{N}(0, N_x), Y_s = X + N_s\), where \(N_s \sim \mathcal{N}(0, N_s) \perp X\), and \(Y_p = Y_s + N_p\), where \(N_p \sim \mathcal{N}(0, N_p - N_s) \perp Y_s\). The minimum achievable distortion for given \(R\) and \(L\) is given by

\[
D'_{\max}(R, L) = \min \left\{ \frac{1}{2} \log \left( \frac{N_x + N_p}{N_x 2^{-2R} + N_p} \right), \frac{1}{2} \log \left( \frac{N_x + N_p}{(N_x + N_s) 2^{-2R} + N_p - N_s} \right) \right\}.
\]

The proof follows similarly as in the binary case where in achievability we let \(V\) be jointly Gaussian with \(X\), i.e., \(V = X + Q\), \(Q \sim \mathcal{N}(0, N_q)\), and in the converse we use the conditional EPI [16]. Note that if the leakage constraint is neglected, e.g., letting \(L \rightarrow \infty\), (22) reduces to the optimal information-rate function in [25].
\( H(X^n | J, K, Y^n_s, Z^n) \)
\( \leq H(X^n, E|J, K, U^n(J), V^n(J, K), Y^n_s, Z^n) \)
\( \leq H(X^n|U^n, V^n, Y^n_s, Z^n, E) + H(E) \)

\[
= \Pr(E = 0)H(X^n|U^n, V^n, Y^n_s, Z^n, E = 0) + \Pr(E = 1)H(X^n|U^n, V^n, Y^n_s, Z^n, E = 1) + H(E)
\]
\( \leq H(X^n|U^n, V^n, Y^n_s, Z^n, E = 0) + \delta, H(X^n) + h(\delta_e) \)

\[
\leq H(X^n|U^n, V^n, Y^n_s, Z^n, E = 0) + n\delta_e \log |\mathcal{X}| + h(\delta_e)
\]
\[
= \sum_{(u^n, v^n, y^n_s, z^n) \in \mathcal{T}_n^{(n)}} p(u^n, v^n, y^n_s, z^n|E = 0)H(X^n|U^n = u^n, V^n = v^n, Y^n_s = y^n_s, Z^n = z^n, E = 0) + n\delta_e \log |\mathcal{X}| + h(\delta_e)
\]
\[
\leq \sum_{(u^n, v^n, y^n_s, z^n) \in \mathcal{T}_n^{(n)}} p(u^n, v^n, y^n_s, z^n|E = 0) \log |\mathcal{T}_n^{(n)}(X|u^n, v^n, y^n_s, z^n)| + n\delta_e \log |\mathcal{X}| + h(\delta_e)
\]
\[
\leq n(H(X|U, V, Y_s, Z) + \delta)'
\]

\textbf{APPENDIX A}

\textbf{PROOF OF THEOREM 11}

The proof is based on the achievable scheme used in [17], which is a layered (superposition) coding scheme with binning.

For random codebook generation, we fix \( P_U|X P_V|Y \) and the reconstruction function \( g(\cdot) \).

- Randomly and independently generate codewords \( u^n(j) \) for \( j \in [1:2^{n(I(X;U)+\delta_e})] \) according to the product distribution \( \prod_{i=1}^n P_U(u_i) \). We distribute the codewords uniformly at random into \( 2^{n(I(X;U)+\delta_e}) \) equal-sized bins \( b_U(w_1, w_2) \), \( w_1 \in [1:2^{n(I(X;U)+\delta_e})] \). Each bin contains \( 2^{n(I(U;Y)−\delta_e}) \) codewords, each indexed by \( w' \).

- For each \( j \), randomly and conditionally independently generate codewords \( v^n(j, k) \) where \( k \in [1:2^{n(I(X;U)+\delta_e})] \) according to the conditional product distribution \( \prod_{i=1}^n P_{V|U}(v_i|u_i(j)) \), and distribute these codewords uniformly at random into \( 2^{n(I(U;Y|V)+2\delta_e}) \) equal-sized bins \( b_V(j, w_2) \), \( w_2 \in [1:2^{n(I(U;Y|V)+2\delta_e})] \). Each bin \( b_V(j, w_2) \) contains \( 2^{n(I(U;Y−\delta_e}) \) codewords, each indexed by \( w'' \). There exists a one-to-one mapping between index \( k \) and the pair of bin/codeword indices \( (w_2, w'') \) such that, without loss of generality, we can identify \( k = (w_2, w'') \).

- The codebooks are revealed to all parties.

For encoding,

- Given \( X^n = x^n \), the encoder looks for codeword \( u^n(j) \) such that \((x^n, u^n(j))\) are jointly typical. From the covering lemma, with high probability, there exist such a codeword such that there are more than \( 2^{n/I(U;V)} \) codewords \( u^n(j) \). If there are more than one, the encoder selects one with the smallest index \( j \).

- Then based on \( x^n \) and the chosen \( u^n(j) \), the encoder looks for a codeword \( v^n(j, k) \) such that \((x^n, u^n(j)), v^n(j, k)\) are jointly typical. From the covering lemma, with high probability, there exists such a codeword since there are more than \( 2^{n(I(U;V))} \) codewords \( v^n(j, k) \).

- The encoder sends the bin indices \( w_1 \) and \( w_2 \) of the chosen codewords to the decoder. The total rate is thus equal to \( I(U;V) + I(U;Y) + 4\delta_e = I(U;V) + 4\delta_e \), where the equality follows from the Markov chain \( U - V - X - Y \).

For decoding,

- Based on \( Y^n = y^n \) and the bin indices \( (w_1, w_2) \), the decoder looks for a unique codeword \( u^n(j) \) in bin \( b_U(w_1) \) such that \((y^n, u^n(j))\) are jointly typical. From the packing lemma, there exists such a codeword \( u^n(j) \) with high probability since there are less than \( 2^{n/I(U;V)} \) codewords \( u^n(j) \) in each bin \( b_U(w_1) \).

- Then based on the decoded \( u^n(j) \), the decoder looks for a unique codeword \( v^n(j, k) \) in bin \( b_V(j, w_2) \) such that \((y^n, u^n(j), v^n(j, k))\) are jointly typical. From the packing lemma, there exists such a codeword \( v^n(j, k) \) with high probability since there are less than \( 2^{n/I(U;V)} \) codewords \( v^n(j, k) \) in each bin \( b_V(j, w_2) \).

- The decoder reconstructs \( \hat{y}_p \) such that \( \hat{y}_p(i, y_i) = g(v_i(j), y_i) \) for \( i = 1, \ldots, n \).

Let \( J = (W_1, W') \) and \( K = (W_2, W'') \) be the indices associated with the chosen codewords \( U^n(J) \) and \( V^n(J, K) \). From LLN, we have that the sequences \((X^n, Y^n, Y^n_s, Z^n, U^n(J), V^n(J, K))\) are jointly typical with high probability. Thus, using similar arguments as in [17] for a bounded distortion measure, the distortion constraint is satisfied if \( D \geq E[d(Y_p, g(V, Y))] \).

Before proceeding with the analysis of the leakage rate, we give a lemma which provides a bound on the \( n \)-letter
conditional entropy based on properties of jointly typical sequences.

**Lemma 2:** Let the index $J$ and the pair of indices $(J,K)$ be the indices specifying codewords $U^n$ and $V^n$, respectively. If $Pr((X^n, U^n(J), V^n(J, K), Y^n_s, Z^n) \in \mathcal{T}^{(n)}_s) \to 1$ as $n \to \infty$, we have that $H(X^n|J, K, Y^n_s, Z^n) \leq nH(X|U, V, Y_s, Z) + \delta_s$.

To prove the lemma, let $E$ be a binary random variable taking value 0 if $(X^n, U^n(J), V^n(J, K), Y^n_s, Z^n) \in \mathcal{T}^{(n)}_s$, and 1 otherwise. Since $(X^n, U^n(J), V^n(J, K), Y^n_s, Z^n) \in \mathcal{T}^{(n)}_s$ with high probability, we have $Pr(E = 1) \leq \delta_s$. The proof steps are given on top of the previous page where step $(a)$ follows from the fact that the codebook, $U^n$ and $V^n$ are functions of $J$ and $(J,K)$, $(b)$ follows from $Pr(E = 1) \leq \delta_s$ where $h(\cdot)$ is the binary entropy function, $(c)$ and $(d)$ follow from the property of jointly typical set [16] with $\delta_c', \delta_v' \to 0$ as $\epsilon \to 0$, and $\epsilon \to 0$ as $n \to \infty$.

The privacy leakage average over all randomly chosen codebooks can be bounded as follows:

$I(Y^n_s; W_1, W_2, Z^n) = H(Y^n_s) - H(Y^n_s|W_1, W_2, Z^n)$

$= H(Y^n_s) - H(Y^n_s, X^n|W_1, W_2, Z^n)$

$+ H(X^n|W_1, W_2, Y^n_s, Z^n)$

$\leq H(Y^n_s) - H(Y^n_s, X^n|J, Z^n) + H(W_2)$

$+ H(X^n|J, K, Y^n_s, Z^n)$

$+ I(X^n; W^n, W''|W_1, W_2, Y^n_s, Z^n)$

$\leq H(Y^n_s) - H(Y^n_s, X^n, Z^n) + H(J) + H(Z^n|J) + H(W_2)$

$+ H(X^n|J, K, Y^n_s, Z^n) + H(W^n, W''|W_1, W_2, Y^n_s, Z^n)$

$(a) \leq n[-H(X, Z|Y_s) + I(X; U) + H(Z|U) + I(X; V|U, Y)$

$+ H(X|U, V, Y_s, Z) + \delta_s']$

$+ H(W^n, W''|W_1, W_2, Y^n_s, Z^n)$

$(b) \leq n[P + \delta_s'] + I(W^n, W''|Y^n_s|W_1, W_2, Y^n_s, Z^n) + n\epsilon_n$

$\leq n[P + \delta_s'] + H(Y^n_s, Z^n) - H(Y^n_s|J, K, Y^n_s, Z^n) + n\epsilon_n$

$(c) \leq n[P + H(Y^n_s, Z) + \delta_s' - H(Y|U, V, Y_s, Z) + \delta_s] + n\epsilon_n$

$(d) = n[I(Y_s; V, Y) + I(Z; X, Y_s|U) - I(Y; X, Y_s|U)$

$- I(X; Z|V, Y_s, Y) + I(X; Y|Y_s, Z) + \delta_s'']$

$\leq n[L + \delta_s'']$

if $L \geq I(Y_s; V, Y) + I(Z; X, Y_s|U) - I(Y; X, Y_s|U) - I(X; Z|V, Y_s, Y) + I(X; Y|Y_s, Z)$, where $(a)$ follows from memoryless property of the sources, from the codebook generation with $J \in [1 : 2^{nH(1; X|U) + \delta_s}]$, $W_2 \in [1 : 2^{nH(X; U, V, Y_s) + \delta_s}]$, and from the bounds $H(Z^n|J) \leq nH(Z^{|U}) + \delta_s$ and $H(X^n|J, K, Y^n_s, Z^n) \leq nH(X^n|U, V, Y_s, Z) + \delta_s$ which can be shown similarly as in Lemma [2] $(b)$ follows from defining $P = -H(X, Z|Y_s) + I(X; U) + H(Z|U) + I(X; V|U, Y) + H(X|U, V, Y_s, Z)$ and Fano’s inequality $H(W^n, W''|W_1, W_2, Y^n_s, Y^n_s, Z^n) \leq n\epsilon_n$, which holds since given the codebook and $W_1, W_2, Y_s$, the decoder can decode $(W^n, W'')$ with high probability, $(c)$ follows from Lemma [3] below, and $(d)$ follows from the definition of $P$ and the Markov chain $U - V - X - (Y, Y_p, Y_s, Z)$.

**Lemma 3:** Given the codebook where $J$ is the codeword index of $U^n$, and $(J,K)$ is the codeword index of $V^n$, if $Pr((Y^n_s, U^n(J), V^n(J, K), Y^n_s, Z^n) \in \mathcal{T}^{(n)}_s) \to 1$ as $n \to \infty$, we have that $H(Y^n_s|J, K, Y^n_s, Z^n) \geq nH(Y^n_s|U, V, Y_s, Z) - \delta_c$.

**Proof of Lemma 2** We consider the following bound.

$H(Y^n_s|J, K, Y^n_s, Z^n)$

$= H(Y^n_s, J, K, Y^n_s, Z^n) - H(J, K) - H(Y^n_s, Z^n|J, K)$

$\geq nH(Y^n_s, Z) + I(X^n; J, K|Y^n_s, Z^n) - H(J, K)$

$- H(Y^n_s, Z^n|J, K)$

$= nH(X, Y, Z) - H(X^n|J, K, Y^n_s, Z^n) - H(J, K)$

$- H(Y^n_s, Z^n|J, K)$

$(a) \geq n[H(X, Y, Z) - H(X^n|U, V, Y, Z) - I(X; U, V)$

$- H(Y, Z|U, V, Y_s, Z) - \delta_c']$

$(b) = n[H(Y^n_s|U, V, Y_s, Z) - \delta_c']$,

where $(a)$ follows from the codebook generation where $J \in [1 : 2^{nH(I(X; U) + \delta_s)]}$ and $K \in [1 : 2^{nH(I(V; Z|U) + \delta_s)]$, and from bounding $H(X^n|J, K, Y^n_s, Z^n)$ and $H(Y^n_s, Z^n|J, K)$ similarly as in Lemma [2] and $(b)$ follows from the Markov chain $(U, V) \rightarrow X \rightarrow (Y, Y_s, Z)$.

The cardinality bounds can be proved using the support lemma [26].

**APPENDIX B**

**Proof of Theorem 1**

Let $(R, D, L)$ be an achievable tuple. Define $U_i \triangleq (W, Z^{i-1}, Y^n_s, V_i), V_i \triangleq (W, Z^{i-1}, Y^n_s)$ and $T_i = (W, Y^n_s, V^n_s, Z^{n-s})$ which satisfy $(V_i, T_i) \rightarrow X_i \rightarrow (Y_i, Y_{p,i}, Y_{s,i}, Z_i)$ and $U_i \rightarrow V_i \rightarrow (T_i, X_i, Y_i, Y_{p,i}, Y_{s,i}, Z_i)$ for all $i = 1, \ldots, n$. From properties of the entropy function we have that

$n(R + \delta_n) \geq H(W)$

$\geq I(X^n_s; Z^n_s; W^n)$

$= \sum_{i=1}^{n} H(X_i, Z_i|Y_i) - H(X_i, Z_i|W, X^{i-1}_s, Z^{i-1}_s, Y^n_s)$

$(a) \geq \sum_{i=1}^{n} H(X_i, Z_i|Y_i) - H(X_i, Z_i|V_i, Y_i)$

$\geq \sum_{i=1}^{n} I(X_i; V_i|Y_i)$,

where $(a)$ follows from the definition of $V_i$, and

$n(D + \delta_n) \geq nE[d_Y^n(Y_p^n, g^n(W, Y^n))]

\geq \sum_{i=1}^{n} E[d_Y^n(Y_{p,i}, g_i^n(W, Y^n))]

\geq \sum_{i=1}^{n} E[d_Y^n(Y_{p,i}, g_i(Y_i, V_i, Y_i))]$,
where \((a)\) follows from the definition of \(V_i\) which includes 
\((W, Y^n)\), implying that there exists a function \(g_i(\cdot)\) such that 
\(g_i(V_i, Y_i) = g_i^{(n)}(W, Y^n)\), and finally 
\[
\begin{align*}
\kappa (L + \delta) & \geq I(Y^n_s; X^n) \\
& \geq H(Y^n_s) - H(Y^n_s, X^n) \mid W, Z^n + H(X^n \mid Y^n_s, Z^n) \\
& \geq H(Y^n_s) - H(Y^n_s, X^n) \mid W + I(Y^n_s, X^n; Z^n) \\
& \quad + H(X^n \mid W, Y^n_s, Z^n) \\
& = H(Y^n_s) - H(Y^n_s, X^n) \mid W - I(Y^n_s, X^n; Y^n) \\
& \quad + I(Y^n_s, X^n; Z^n) + H(X^n \mid W, Y^n_s, Z^n), \\
\end{align*}
\]
which \((a)\) follows from the Markov chain \((Y^n_s - (X^n, Y^n) - W)
and \((b)\) follows from the Markov chain \((Y^n, Z^n) - (Y^n_s, X^n) - W).

Continuing the chain of inequalities, we get 
\[
n(L + \delta_n) \\
\geq \sum_{i=1}^{n} I(Y^n_s, X^n) - H(Y^n_s, X^n) \mid W + H(Z_i \mid X_i, Y_i, Y^n_s) \\
- I(Z_i \mid W \mid Z^i) + I(Y^n_s, X^n) \mid Y^n_s) + I(Y^n_s, X^n; Z^n) \\
+ I(Y^n_s, X^n; Z^n) + H(X^n \mid W, Y^n_s, Z^n), \\
\]
where \((c)\) follows from the Markov chain \((Y^n_s, Z^n) - (X^n, Y^n) - W, \) 
the definition of \(V_i,\) and the Markov chain 
\((Y^n_s - (W, X^n) \mid Y^n_s, Z^n) - X^n)\), \((d)\) follows from the 
Markov chain \((Z_i - (X_i, Y_i, Y^n_s) - V_i,\) the definition of \(T_i,\) and 
the facts that \(Z_i\) is independent of \(Z^i\) and \(Y_i\) is independent of \(Y^n_s), \((e)\) follows from the 
Csiszar’s sum identity \([16] \), and \((f)\) follows from the definition of \(U_i\) 
and the Markov chain \((Y_i, Z_i) - (X_i, Y^n_s) - U_i).\n
We end the proof by the standard time-sharing argument 
and letting \(n \to \infty\) and \(\delta_n \to 0.\n
**APPENDIX C**

**PROOF OF GAUSSIAN WITH QUADRATIC DISTORTION**

**Example**

Recalling that \(X \sim \mathcal{N}(0, N_s), Y_s = X + \tilde{N}_s,\) where \(\tilde{N}_s \sim \mathcal{N}(0, N_s) \perp X,\) and \(Y_p = Y_s + \tilde{N}_p,\) where \(\tilde{N}_p \sim \mathcal{N}(0, N_p - N_s) \perp Y_s.\)

**Proof of Achievability:** Let us choose \(V = X + Q, Q \sim \mathcal{N}(0, N_q)\) independent of \(X,\) and choose \(g(V)\) to be an MMSE estimate of \(Y_p\) given \(V.\) With these choices of \(V\) and \(g(\cdot),\) it can be shown that 
\[
I(X; V) = h(V) - h(V \mid X) \\
= \frac{1}{2} \log \left( \frac{N_x + N_q}{N_q} \right),
\]
and 
\[
I(Y_s; V) = h(Y_s) - h(Y_s \mid V) \\
= \frac{1}{2} \log \left( \frac{N_x + N_s}{N_s + N_p/N_q} \right),
\]
and finally 
\[
E[d(Y_p, g(V))] = E[(Y_p - g(V))^2] \\
= N_p + \frac{N_s N_q}{N_s + N_p/N_q}.
\]

Letting \(D = E[d(Y_p, g(V))] = N_p + \frac{N_s N_q}{N_s + N_p/N_q}\) and substituting it into the constraints on \(R\) and \(L\) complete the achievability part.

**Proof of Converse:** From the problem formulation, the joint 
PMF \(P_{X^n, Y^n, W, Y^n_p}^{X^n, Y^n_p} \) is given by 
\[
P_{X^n, Y^n_p, Y^n_p, W}^{X^n, Y^n_p} \mid Y^n_p = g(n) (W)\).

From the EPI, we have that 
\[
2^{h(Y_s \mid W)} \geq 2^{\frac{1}{2} h(X^n \mid W) + 2^{\frac{1}{2} h(\tilde{N}_s)}} \tag{23}
\]
and 
\[
2^{\frac{1}{2} h(Y_p \mid W)} \geq 2^{\frac{1}{2} h(Y_s \mid W) + 2^{\frac{1}{2} h(\tilde{N}_s)}} \tag{24}
\]
\[
\geq 2^{\frac{1}{2} h(X^n \mid W) + 2^{\frac{1}{2} h(\tilde{N}_s)} + 2^{\frac{1}{2} h(\tilde{N}_p)}} \tag{25}
\]
where the last inequality follows from \([23]\). Then it follows that 
\[
\kappa (R + \delta) \geq H(W) \\
\geq I(X^n; W) \\
= h(X^n) - h(X^n \mid W) \\
\geq h(X^n) \\
- \frac{n}{2} \log (2^{\frac{1}{2} h(Y_s \mid W)} - 2^{\frac{1}{2} h(\tilde{N}_s)} - 2^{\frac{1}{2} h(\tilde{N}_p)}), \tag{26}
\]
where the last inequality follows from \([25]\).
\[ n(L + \delta_n) \geq I(Y^n_s; W) \]
\[ = h(Y^n_s) - h(Y^n_s | W) \]
\[ \geq h(Y^n_s) - \frac{n}{2} \log(2^{2h(Y^n_s | W) - 2^{\frac{3}{2}h(S^n_s)})}, \] (27)
where the last inequality follows from (24).

\[ D + \delta_n \geq E[d^n(y^n_p, g^n(w)) | W] \]
\[ = \frac{1}{n} \sum_{i=1}^{n} E[(Y^n_p - g_i^n(W))^2] \]
\[ = \frac{1}{2\pi e} 2\log(2^{\frac{3}{2}} \sum_{i=1}^{n} E[(Y^n_p - g_i^n(W))^2]) \]
\[ (a) \geq \frac{1}{2\pi e} 2^{\frac{3}{2}} \sum_{i=1}^{n} \log(2\pi e E[(Y^n_p - g_i^n(W))^2]) \]
\[ (b) \geq \frac{1}{2\pi e} 2^{\frac{3}{2}} \sum_{i=1}^{n} h(Y^n_p | W) \]
\[ \geq \frac{1}{2\pi e} 2^{\frac{3}{2}} h(Y^n_p | W), \] (28)
where (a) follows from Jensen’s inequality (16) and the fact that \( \log(.) \) is a concave function, and (b) follows from the fact that \( E[\text{var}(Y^n_p | W)] \) is the MMSE over all possible estimator of \( Y^n_p \) for each \( i = 1, \ldots, n \).

Combining (28) with (26) and (27), and letting \( n \to \infty \) and \( \delta_n \to 0 \) complete the converse part.

**APPENDIX D**

**PROOF OF SECURE INFORMATION BOTTLENECK EXAMPLE: BINARY SOURCE**

Recalling that \( X \sim \text{Bernoulli}(1/2) \), \( Y_s \) is an output of a BSC\((p)\) with input \( X \), and \( Y_p \) is an output of a BSC\((q)\) with input \( Y_s \).

**Proof of Achievability:** Let \( V \) be an output of a BSC with input \( X \). We have that

\[ H(Y_p | V) = h(h^{-1}(H(X | V) \ast p \ast q)) \]
\[ = h(h^{-1}(H(Y_s | V) \ast q)). \] (29)

We let \( D' = H(Y_p) - H(Y_p | V) \). Combining (29) with the constraint \( R \geq H(X) - H(X | V) \) and noting that \( u \ast q \) is an increasing function in \( u \) for \( u \in [0, 1/2] \) for some fixed \( q \in [0, 1/2] \), we get \( D' \leq H(Y_p) - h(h^{-1}(H(X) - R) \ast p \ast q) \).

Similarly, combining (29) with \( L \geq H(Y_s) - H(Y_s | V) \) gives

\[ D' \leq H(Y_p) - h(h^{-1}(H(Y_s) - L) \ast q)). \] (30)

**Proof of Converse:** From Mrs. Gerber’s lemma, we have

\[ H(Y_s | V) \geq h(h^{-1}(H(X | V) \ast p)) \] (31)
\[ H(Y_p | V) \geq h(h^{-1}(H(Y_s | V) \ast q)). \] (32)

Since \( h(h^{-1}(u \ast q)) \) is an increasing function in \( u \) for \( u \in [0, 1/2] \), combining (31) and (32) gives

\[ H(Y_p | V) \geq h(h^{-1}(H(X | V) \ast p \ast q)). \] (33)

Then we have that

\[ D' \leq I(Y_p; V) \]
\[ \leq H(Y_p) - h(h^{-1}(H(Y_s | V) \ast q)) \]
\[ \leq H(Y_p) - h(h^{-1}(H(Y_s) - L) \ast q)), \]

where (a) follows from (32) and (b) follows from the facts that \( h(h^{-1}(H(Y_s | V) \ast q)) \) is an increasing function in \( H(Y_s | V) \) and that \( L \geq H(Y_s) - H(Y_s | V) \).

Similarly, we have

\[ D' \leq I(Y_p; V) \]
\[ \leq H(Y_p) - h(h^{-1}(H(X | V) \ast p \ast q)) \]
\[ \leq H(Y_p) - h(h^{-1}(H(X) - R) \ast p \ast q), \]

where (a) follows from (33) and (b) follows from the facts that \( h(h^{-1}(H(X | V) \ast p \ast q)) \) is an increasing function in \( H(X | V) \) and that \( R \geq H(X) - H(X | V) \).

**REFERENCES**

[1] L. Sankar, S. R. Rajagopalan, and H. V. Poor, “Utility-privacy tradeoffs in databases: An information-theoretic approach,” IEEE Trans. Inf. Forensic Secur., vol. 8, no. 6, pp. 838–852, June 2013.
[2] F. du Pin C名mon and N. Fawaz, “Privacy against statistical inference,” in Proc. Allerton Conf. Commun. Control Comput, Oct 2012.
[3] H. Yamamoto, “A source coding problem for sources with additional outputs to keep secret from the receiver or wiretappers (corresp.),” IEEE Trans. Inf. Theory, vol. 29, no. 6, pp. 918–923, Nov 1983.
[4] —, “Coding theorems for shannon’s cipher system with correlated source outputs, and common information,” IEEE Trans. Inf. Theory, vol. 40, no. 1, pp. 85–95, Jan 1994.
[5] V. Prabhakaran and K. Ramchandran, “On secure distributed source coding,” Proc. IEEE Inf. Theory Workshop, pp. 442–447, 2007.
[6] D. Gündüz, E. Erikp, and H. V. Poor, “Lossless compression with security constraints,” in Proc. IEEE ISIT, 2008, pp. 111–115.
[7] J. Villard and P. Piantanida, “Secure multiterminal source coding with side information at the eavesdropper,” IEEE Trans. Inf. Theory, vol. 59, no. 6, pp. 3668–3692, June 2013.
[8] C. Schieler and P. Cuff, “Rate-distortion theory for secrecy systems,” IEEE Trans. Inf. Theory, vol. 60, no. 12, pp. 7584–7605, Dec 2014.
[9] N. Merhav and M. Feder, “Universal prediction,” IEEE Trans. Inf. Theory, vol. 44, no. 6, pp. 2124–2147, Oct 1998.
[10] T. Courtade and T. Weissman, “Multiterminal source coding under logarithmic loss,” IEEE Trans. Inf. Theory, vol. 60, no. 1, pp. 740–761, Jan 2014.
[11] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Medard, “From the information bottleneck to the privacy funnel,” in IEEE Information Theory Workshop (ITW), 2014, Nov 2014, pp. 501–505.
[12] N. Tishby, F. C. Pereira, and W. Bialek, “The information bottleneck method,” in Proc. Allerton Conf. Commun. Control Comput, 1999.
[13] R. Tandon, L. Sankar, and H. V. Poor, “Discriminatory lossy source coding: Side information privacy,” IEEE Trans. Inf. Theory, vol. 59, no. 9, pp. 5665–5677, Sept 2013.
[14] C. Heegard and T. Berger, “Rate distortion when side information may be absent,” IEEE Trans. Inf. Theory, vol. 51, pp. 2727–2734, Nov 2005.
[15] B. Morfahf and L. Sankar, “Information-theoretic private interactive mechanism,” in Proc. Allerton Conf. Commun. Control Comput, 2015.
[16] A. El Gamal and Y. H. Kim, Network Information Theory. Cambridge University Press, 2011.
[17] A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” IEEE Trans. Inf. Theory, vol. 22, no. 1, pp. 1–10, Jan 1976.
[18] S. Asboodeh, F. Alajaji, and T. Linder, “Lossless secure source coding: Yamamoto’s setting,” in Proc. Allerton Conf. Commun. Control Comput, 2015.
[19] A. D. Wyner, “The rate-distortion function for source coding with side information at the decoder-part II: General sources,” Inf. Control, no. 38, pp. 60–80, 1978.

[20] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics). Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[21] T. Andre, M. Antonini, M. Barlaud, and R. Gray, “Entropy-based distortion measure and bit allocation for wavelet image compression,” IEEE Trans. Image Process., vol. 16, no. 12, pp. 3058–3064, Dec 2007.

[22] Y. Chia and K. Kittichokechai, “On secure source coding with side information at the encoder,” CoRR, 2013, abs/1307.0974.

[23] G. Chechik and N. Tishby, “Extracting relevant structures with side information,” Advances in Neural Information Processing Systems 15, 2002.

[24] N. Slonim and N. Tishby, “Agglomerative information bottleneck.” MIT Press, 1999, pp. 617–623.

[25] A. Winkelbauer and G. Matz, “Rate-information-optimal gaussian channel output compression,” in 48th Annual Conference on Information Sciences and Systems (CISS), March 2014.

[26] I. Csizér and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems. Cambridge University Press, 2011.