ISIT'2013 1569718143

Secure Source Coding with a Public Helper

Kittipong Kittichokechai[†], Yeow-Khiang Chia[‡], Tobias J. Oechtering[†], Mikael Skoglund[†], and Tsachy Weissman[§]

[†] School of Electrical Engineering and the ACCESS Linnaeus Center, KTH Royal Institute of Technology, Sweden

[‡] Modulation and Coding Dept., Institute for Infocomm Research, Singapore

[§] Electrical Engineering Department, Stanford University, USA

***Abstract*—We consider secure multi-terminal source coding problems in the presence of a public helper. Two main scenarios are studied: 1) source coding with a helper where the coded side information from the helper is eavesdropped by an external eavesdropper, 2) triangular source coding with a helper where the helper is considered as a public terminal. We are interested in how the helper can support the source transmission subject to a constraint on the amount of information leaked due to its public nature. We characterize the tradeoff between transmission rate, incurred distortion, and information leakage rate at the helper/eavesdropper in the form of a rate-distortion-leakage region for various classes of problems.**

I. Introduction

Nowadays the Internet is an essential part of our daily life. We rely on many online services which inevitably create huge amounts of information flow in the network. With this huge amount of information, the main tasks for network designers are to ensure that the data can be transmitted reliably and also securely across the network. The latter requirement is becoming increasingly acute, especially when sensitive information is involved. Let us imagine a network in which information flows from one node to another through a number of intermediate nodes. The system design generally makes use of these intermediate nodes to help the transmission. However, these nodes might be public devices or terminals which we cannot fully trust with access to significant amounts of our information. This scenario leads to a natural tradeoff between cooperation and secrecy in the system and motivates the study of secure communication and compression in the presence of a public helper. More specifically, we consider a secure lossy source coding problem involving a public helper under an information leakage rate constraint. An overview of the problem settings and contributions of this work is as follows.

A. Contribution

1) Secure Source Coding with a Public Helper: First we consider lossy source coding with a public helper problems, as depicted in Fig. 1, which are essentially extensions of the one-helper problem [1, Ch.10], [2] in the presence of an eavesdropper. The setting is motivated by a scenario where the helper can only provide side information through a rate-limited communication link which is not secure due to its *public* nature, i.e., it can be eavesdropped by an external eavesdropper. In the "one-sided helper" setting, the helper communicates through a public link only to the decoder, while in the "two-sided helper" case, the helper *broadcasts* the same coded side information to both encoder and decoder. We characterize the rate-distortion-leakage tradeoff under the logarithmic loss function [3] for the one-sided helper case and under general distortion for the two-sided helper case. Interestingly, the achievable schemes for the original one-helper problems are also optimal in the presence of an eavesdropper. However, the set of optimizing input distributions may change due to the leakage constraint.

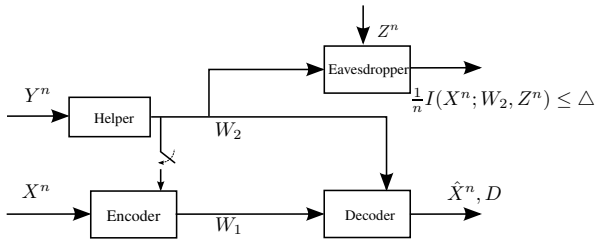

Fig. 1. Secure source coding with one-sided/two-sided public helper.

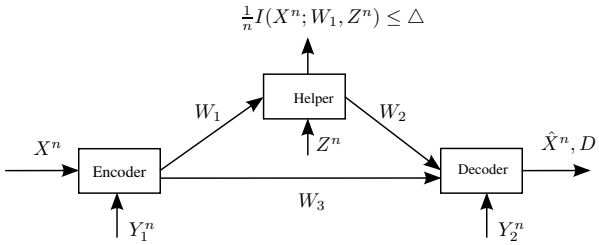

Fig. 2. Secure triangular source coding with public helper.

2) Secure Triangular/Cascade Source Coding with a Public Helper: Next, we consider problems of triangular source coding with a public helper, as shown in Fig. 2. In contrast to the previous settings, where the focus is on leakage at an external eavesdropper, we address the problem of information leakage at a legitimate user. The setting is motivated by a scenario where the helper is a public terminal that forwards the information as the protocol requests from the encoder to the decoder. However, the helper might be curious and not ignore the data which may not be intended for him. We characterize the rate-distortion-leakage regions for various settings summarized below based on different side information patterns available to the encoder, the helper, and the decoder.

Setting (A): We solve the problem under the logarithmic loss

This work was partially supported by the Swedish Research Council and the NSF Center for Science of Information (CSoI).

distortion by assuming that Y_1 is constant and $Y_2 = Y$. We also assume that $X - Y - Z$ forms a Markov chain and show that the forwarding scheme at the helper (setting $W_2 = W_1$) is optimal. Note that the Markov assumption $X - Y - Z$ in this setting can be relevant in scenarios where the decoder is a fusion center collecting all correlated side information.

Setting (B): We assume that the side information $Y_1^n = Y_2^n = Y^n$, and solve the problem under the logarithmic loss distortion. Again we assume that $X - Y - Z$ forms a Markov chain and show that the forwarding scheme at the helper is optimal. Interestingly, we note that although the availability of the side information at the encoder does not improve the rate-distortion tradeoff, this side information can be used for a secret key generation at the encoder and the decoder. In our coding scheme, the secret key is used to scramble part of the message sent to the helper, and thus decrease the leakage.

Setting (C): We assume that the side information at the helper is also available at the encoder, i.e., $Y_1 = Z$ and we let $Y_2 = Y$. In this case we assume that $X - Z - Y$ forms a Markov chain and solve the problem under general distortion. Due to $X - Z - Y$, we show that the decode-and-re encode type scheme at the helper is optimal. That is, it is meaningful to take into account Z^n at the helper in relaying information.

We note that our settings (A)-(C) are different from the conventional triangular/cascade source coding problem in that the decoding constraint at the helper is replaced by the secrecy constraint. Also, the cascade settings can be seen as special cases of our triangular settings when the private link from the encoder to the decoder is removed (setting W_3 constant).

B. Related Work

Multi-terminal source coding problems have been studied extensively in various settings. We refer to [1, Ch.10] for the lossless distributed source coding, and to [1, Ch.11-12] for the lossy case. The problem of lossy distributed source coding is still open in general. There exist only a few special cases which can be solved completely (cf., e.g., [1, Ch.12]. Recently, [3] introduced a logarithmic loss distortion[1] as a new and interesting distortion measure for lossy distributed source coding and solved the problem completely under this distortion measure. As many of our settings are solved under the logarithmic loss distortion, its definition and important properties, taken from [3], are briefly given at the end of this section. As for other related multi-terminal source coding problems, [4] studied and established the rate-distortion regions for the cascade and triangular source coding problems without side information. Variations of the cascade and triangular settings have been studied in recent years (see, e.g., [5, 6]).

Recently, security has become an important issue in system design, i.e., when the goal is to design a communication system that is both reliable and secure. Physical layer security was introduced based on the fact that the signals available at the legitimate receiver and the eavesdropper usually possess different characteristics. Information theoretic study of physical layer security was pioneered by Wyner in [7], which introduces and solves the problem of coding for the Wiretap channel. Several extensions to other multiuser channels are considered in [8]. More recently, due to potential applications in areas such as privacy in sensor networks and databases (see, e.g., [9]), an idea of physical layer security from the source coding perspective was also studied, i.e., source coding with side information subject to an additional secrecy constraint. Secure lossless distributed source coding was studied in [10–12], and the lossy case was recently considered in [13, 14]. Note that the one-sided/two-sided helper settings in Fig. 1 where the eavesdropper observes instead the link from an encoder to a decoder are studied in [12, 13].

[1]Some motivation for the use of logarithmic loss measure for source coding can also be found in [3].

Definition 1 (logarithmic loss [3]): For logarithmic loss distortion measure, we let the reconstruction alphabet $\hat{\mathcal{X}}$ be the set of probability distribution over the source alphabet $\mathcal{X}$. That is $\hat{X}$ is a probability distribution on $\mathcal{X}$, i.e., $\hat{X} : \mathcal{X} \to [0, 1]$, and $\hat{X}(x)$ is a probability distribution on $\mathcal{X}$ evaluated for the outcome $x \in \mathcal{X}$. The logarithmic loss distortion measure is defined as $d(x, \hat{x}) = \log(\frac{1}{\hat{x}(x)})$. Using this definition for symbol-wise distortion, we define the distortion between sequences as $d^{(n)}(x^n, \hat{x}^n) = \frac{1}{n}\sum_{i=1}^{n} d(x_i, \hat{x}_i)$.

In the following, we present some lemmas in [3] that are essential in proving our results under the logarithmic loss distortion. Lemma 1 is used in the achievability proof, while Lemma 2 is used for upper bounding the conditional entropy in the converse proof.

Lemma 1: Let U be the argument of the reconstruction function $g(\cdot)$, then under the logarithmic loss distortion measure, we get $E[d(X, g(U))] = H(X|U)$.

Lemma 2: Let $Z = (W_1, W_2)$ be the argument of the reconstruction function $g^{(n)}(\cdot)$, then under the logarithmic loss distortion measure, we get $E[d^{(n)}(X^n, g^{(n)}(Z))] \geq \frac{1}{n}H(X^n|Z)$.

II. Secure Source Coding with One-sided/Two-sided Public Helper

In this section, we consider source coding with a helper whose link to a decoder (and an encoder) is public and can therefore be observed by an external eavesdropper.

A. Secure Source Coding with One-sided Public Helper

Let us consider the setting in Fig. 1 when the switch is open. Source, side information, and reconstruction alphabets, $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \hat{\mathcal{X}}$ are assumed to be finite, and source and side information sequences (X^n, Y^n, Z^n) are assumed to be i.i.d. according to $P_{X,Y,Z}$.

Definition 2: A $(|\mathcal{W}_1^{(n)}|, |\mathcal{W}_2^{(n)}|, n)$-code for secure source coding with one-sided helper link consists of:

- A stochastic encoder $F_1^{(n)}$ which takes X^n as an input and generates $W_1 \in \mathcal{W}_1^{(n)}$ according to a conditional pmf $p(w_1|x^n)$,
- A stochastic helper $F_2^{(n)}$ which takes Y^n as an input and generates $W_2 \in \mathcal{W}_2^{(n)}$ according to $p(w_2|y^n)$, and
- A decoder $g^{(n)} : \mathcal{W}_1^{(n)} \times \mathcal{W}_2^{(n)} \to \hat{\mathcal{X}}^n$,

where $\mathcal{W}_1^{(n)}$ and $\mathcal{W}_2^{(n)}$ are finite sets.

Let $d : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty)$ be the single-letter distortion measure. The average *distortion* between the source sequence and its reconstruction at the decoder is defined as

$$E\left[d^{(n)}\left(X^n, \hat{X}^n\right)\right] \triangleq \frac{1}{n} E\left[\sum_{i=1}^{n} d(X_i, \hat{X}_i)\right],$$

where $d^{(n)}(\cdot)$ is the distortion function.

The *information leakage rate* at the eavesdropper who has access to W_2 and Z^n is measured by the normalized mutual information $\frac{1}{n} I(X^n; W_2, Z^n)$.

Definition 3: The rate-distortion-leakage tuple $(R_1, R_2, D, \triangle) \in \mathbb{R}_+^4$ is said to be *achievable* if for any $\delta > 0$ and all sufficiently large n there exists a $(|\mathcal{W}_1^{(n)}|, |\mathcal{W}_2^{(n)}|, n)$ code such that

$$\frac{1}{n} \log \left|\mathcal{W}_i^{(n)}\right| \leq R_i + \delta, i = 1, 2,$$
$$E[d^{(n)}(X^n, g^{(n)}(W_1, W_2))] \leq D + \delta,$$
$$\text{and} \quad \frac{1}{n} I(X^n; W_2, Z^n) \leq \triangle + \delta.$$

The *rate-distortion-leakage region* $\mathcal{R}_{\text{one-sided}}$ is the set of all achievable tuples.

Theorem 1 (logarithmic loss): The rate-distortion-leakage region under logarithmic loss $\mathcal{R}_{\text{one-sided, logloss}}$ is the set of all tuples $(R_1, R_2, D, \triangle) \in \mathbb{R}_+^4$ for which there exists a random variable $U \in \mathcal{U}$ such that $U - Y - (X, Z)$ forms a Markov chain and

$$R_2 \geq I(Y; U), \tag{1a}$$
$$R_1 \geq [H(X|U) - D]^+, \tag{1b}$$
$$\triangle \geq I(X; U, Z), \tag{1c}$$

where $[a]^+ := \max\{0, a\}$. The cardinality of the set $\mathcal{U}$ can be upperbounded by $|\mathcal{U}| \leq |\mathcal{Y}| + 2$.

Proof sketch: The achievable scheme consists of the rate-distortion code for lossy transmission of y^n via the codeword u^n at rate $I(Y; U) + \epsilon$, and the Wyner-Ziv code at rate $I(X; V|U) + 2\epsilon$ for lossy transmission of x^n with u^n as side information at the decoder. We can show that the distortion D and the leakage rate $\triangle$, satisfying $D \geq E[d(X, g(U, V))], \triangle \geq I(X; U, Z)$, are achievable. The analysis of leakage rate is given below where the mutual information averaged over all codebooks $\mathcal{C}_n$ is given by

$$\begin{aligned}
&I(X^n; W_2, Z^n|\mathcal{C}_n) \\
&= H(X^n|\mathcal{C}_n) - H(X^n, W_2, Z^n|\mathcal{C}_n) + H(W_2, Z^n|\mathcal{C}_n) \\
&\leq H(X^n|\mathcal{C}_n) - H(X^n, Z^n|\mathcal{C}_n) - I(W_2; Y^n|X^n, Z^n, \mathcal{C}_n) \\
&\quad + H(W_2|\mathcal{C}_n) + H(Z^n|W_2, \mathcal{C}_n) \\
&\overset{(a)}{=} H(X^n) - H(X^n, Y^n, Z^n) + H(Y^n|W_2, X^n, Z^n, \mathcal{C}_n) \\
&\quad + H(W_2|\mathcal{C}_n) + H(Z^n|W_2, \mathcal{C}_n) \\
&\overset{(b)}{\leq} n[H(X) - H(X, Y, Z) + H(Y|U, X, Z) + \delta_\epsilon \\
&\quad + I(Y; U) + \epsilon + H(Z|U) + \delta_\epsilon] \\
&\overset{(c)}{=} n[I(X; U, Z) + \delta'_\epsilon],
\end{aligned}$$

where (a) follows from the facts that (X^n, Y^n, Z^n) are independent of the codebook, (b) follows from the i.i.d. property of (X^n, Y^n, Z^n), from the codebook generation that we have $W_2 \in [1 : 2^{n(I(Y;U)+\epsilon)}]$, and from bounding the terms $H(Y^n|W_2, X^n, Z^n, \mathcal{C}_n)$ and $H(Z^n|W_2, \mathcal{C}_n)$ (proofs are given in [15]), (c) follows from the Markov chain $U - Y - (X, Z)$, and that $\delta'_\epsilon := \epsilon + 2\delta_\epsilon \to 0$ as $\epsilon \to 0$.

Due to the property of logarithmic loss distortion function (Lemma 1), we have $E[d(X, g(U, V))] = H(X|U, V)$. If $H(X|U) < D$, the encoder does not need to send anything, i.e., setting V constant. If $H(X|U) > D$, we define $V = X$ with probability $p = 1 - \frac{D}{H(X|U)}$ and constant otherwise. Then we get $H(X|U, V) = D$ and $I(X; V|U) = H(X|U) - D$. Therefore, we obtain the desired achievable rate-distortion-leakage expressions as in (1).

The converse proof follows standard steps by utilizing the fact that, for logarithmic loss distortion function, $E[d(X^n, g(W_1, W_2))] \geq \frac{1}{n} H(X^n|W_1, W_2)$ (Lemma 2), and by defining $U_i \triangleq (W_2, X^{i-1})$. We refer readers to the complete proof in [15]. For the bound on the cardinality of the set $\mathcal{U}$, it can be shown by using the support lemma [1, Appendix] that it suffices that $\mathcal{U}$ should have $|\mathcal{Y}| - 1$ elements to preserve P_Y, plus three more for $H(Y|U), H(X|U)$, and $H(X|U, Z)$.

Other Results: For the case of one-sided helper in Fig. 1, the Gaussian setting (jointly Gaussian sources and quadratic distortion function) can also be solved under an additional assumption of $Y - X - Z$. The complete characterization of rate-distortion-leakage region is given in [15]. Also, by letting $D \to 0$, we obtain the result for the lossless setting. In this case, the scheme for lossless one-helper problem [1, Ch. 10] is in fact optimal.

B. Secure Source Coding with Two-sided Public Helper

Let us consider the setting in Fig. 1 when the switch is closed. Since the problem setting is similar to that of the one-sided helper case, details are omitted. The main difference is that the coded side information $W_2 \in \mathcal{W}_2^{(n)}$ is given to both the encoder and the decoder. The encoding function becomes $F_1^{(n)}$ which takes (X^n, W_2) as input and generates W_1 according to $p(w_1|x^n, w_2)$. We characterize the rate-distortion-leakage region for a general distortion function.

Theorem 2: The rate-distortion-leakage region $\mathcal{R}_{\text{two-sided}}$ is the set of all tuples $(R_1, R_2, D, \triangle) \in \mathbb{R}_+^4$ for which there exist random variables $U \in \mathcal{U}$ and $\hat{X} \in \hat{\mathcal{X}}$ such that $U - Y - (X, Z)$ and $\hat{X} - (U, X) - (Y, Z)$ form Markov chains and

$$R_2 \geq I(Y; U), \tag{2a}$$
$$R_1 \geq I(X; \hat{X}|U), \tag{2b}$$
$$D \geq E[d(X, \hat{X})], \tag{2c}$$
$$\triangle \geq I(X; U, Z). \tag{2d}$$

The cardinality of the alphabet of the auxiliary random variable can be upperbounded by $|\mathcal{U}| \leq |\mathcal{Y}| + 3$.

Proof sketch: The achievable scheme consists of the rate-distortion code for lossy transmission of y^n via u^n at rate

$I(Y;U)+\epsilon$. Since w_2 is given to both encoder and decoder, the source coding with side information known at both encoder and decoder at rate $I(X;\hat{X}|U)+2\epsilon$ is employed for lossy transmission of x^n with u^n as side information. The achievable leakage proof and the converse proof follow similarly as that of one-sided helper case, and are omitted. Similarly as in [2], we note that by following the converse proof of one-sided helper case, we in fact proved the outer bound which has the same expressions as in (2), but with the joint distribution satisfying $U-Y-(X,Z)$ and $\hat{X}-(U,X,Y)-Z$. Clearly, this outer bound includes the achievable region due to the larger set of distributions. To show that the outer bound is also included in the achievable region, we show that there exists a distribution, induced by the one in the outer bound, of the form satisfying the Markov conditions in the achievable region such that the constraints on $(R_1,R_2,D,\triangle)$ in (2) hold. See [2] and [15] for more details. For the bound on $|\mathcal{U}|$, by the support lemma, it suffices that $\mathcal{U}$ should have $|\mathcal{Y}|-1$ elements to preserve P_Y, plus four more for $H(Y|U), I(X;\hat{X}|U), H(X|U,Z)$, and the distortion constraint.

III. Secure Triangular/Cascade Source Coding with a Public Helper

In this section, we consider triangular source coding where the helper plays a role in relaying information from an encoder to a decoder subject to the leakage rate constraint, as depicted in Fig. 2. Clearly, there exists a tradeoff between the amount of information leakage at the helper and the helper's ability to support the transmission. We characterize the tradeoff between rate, distortion, and information leakage rate in the form of rate-distortion-leakage region for different special cases (settings (A)-(C)).

A. General Problem Formulation

Let us consider a setting in Fig. 2. Source and side information sequences (X^n,Y_1^n,Y_2^n,Z^n) are assumed to be i.i.d. according to $P_{X,Y_1,Y_2,Z}$.

Definition 4: A $(|\mathcal{W}_1^{(n)}|,|\mathcal{W}_2^{(n)}|,|\mathcal{W}_3^{(n)}|,n)$-code for secure triangular source coding with a public helper consists of:

- A stochastic encoder $F_1^{(n)}$ which takes (X^n,Y_1^n) as input and generates $W_1\in\mathcal{W}_1^{(n)}$ according to a conditional pmf $p(w_1|x^n,y_1^n)$,
- A stochastic helper $F_2^{(n)}$ which takes (W_1,Z_1^n) as input and generates $W_2\in\mathcal{W}_2^{(n)}$ according to $p(w_2|w_1,z^n)$,
- A stochastic encoder $F_3^{(n)}$ which takes (X^n,Y_1^n) as input and generates $W_3\in\mathcal{W}_3^{(n)}$ according to $p(w_3|x^n,y_1^n)$, and
- A decoder $g^{(n)}:\mathcal{W}_2^{(n)}\times\mathcal{W}_3^{(n)}\times\mathcal{Y}_2^{(n)}\to\hat{\mathcal{X}}^n$,

where $\mathcal{W}_1^{(n)},\mathcal{W}_2^{(n)}$, and $\mathcal{W}_3^{(n)}$ are finite sets.

The information leakage rate at the helper who has access to W_1 and Z^n is measured by $\frac{1}{n}I(X^n;W_1,Z^n)$.

Definition 5: The rate-distortion-leakage tuple $(R_1,R_2,R_3,D,\triangle)\in\mathbb{R}_+^5$ is said to be *achievable* if for any $\delta>0$ and all sufficiently large n there exists a $(|\mathcal{W}_1^{(n)}|,|\mathcal{W}_2^{(n)}|,|\mathcal{W}_3^{(n)}|,n)$ code such that

$$\frac{1}{n}\log\left|\mathcal{W}_i^{(n)}\right|\le R_i+\delta, i=1,2,3,$$
$$E[(X^n,g^{(n)}(W_2,W_3,Y_2^n))]\le D+\delta,$$
$$\text{and}\quad \frac{1}{n}I(X^n;W_1,Z^n)\le\triangle+\delta.$$

The *rate-distortion-leakage* region is defined as a set of all achievable tuples.

B. Main Result

We characterize the rate-distortion-leakage regions for some special cases of the triangular source coding settings described earlier.

Theorem 3 (setting (A), logarithmic loss): The rate-distortion-leakage region $\mathcal{R}_{\text{tri(A), X-Y-Z, logloss}}$ under logarithmic loss distortion and the Markov relation $X-Y-Z$ is the set of all tuples $(R_1,R_2,R_3,D,\triangle)\in\mathbb{R}_+^5$ that satisfy

$$R_1\ge[H(X|Y)-D-R_3]^+, \tag{3a}$$
$$R_2\ge[H(X|Y)-D-R_3]^+, \tag{3b}$$
$$\triangle\ge I(X;Z)+[H(X|Y)-D-R_3]^+. \tag{3c}$$

Proof: Sketch of Achievability: The Wyner-Ziv coding at rate of $I(X;U|Y)+2\epsilon=H(X|Y)-D+2\epsilon$ is performed to satisfy the distortion constraint, where the equality is due to the choice of U and the property of logarithmic loss distortion. If $H(X|Y)-D>R_3$, we perform rate-splitting on the Wyner-Ziv index. That is, we split the index into two parts, namely $w_1\in[1:2^{n(H(X|Y)-D-R_3+\epsilon)}]$, and $w_3\in[1:2^{n(R_3+\epsilon)}]$. The indices w_1 and w_3 are sent over the cascade link and the private (triangular) link, respectively. Then the helper forwards the index w_1 to the decoder. It can be seen that the rate and distortion constraints are satisfied. As for the analysis of leakage rate averaged over all codebooks $\mathcal{C}_n$, we get

$$\begin{aligned}I(X^n;W_1,Z^n|\mathcal{C}_n)&\le I(X^n;Z^n|\mathcal{C}_n)+H(W_1|Z^n,\mathcal{C}_n)\\&\overset{(a)}{\le} n[I(X;Z)+H(X|Y)-D-R_3+\epsilon],\end{aligned}$$

where (a) follows from the facts that (X^n,Z^n) are i.i.d. and independent of the codebook, and from the codebook generation that we have $W_1\in[1:2^{n(H(X|Y)-D-R_3+\epsilon)}]$.

On the other hand, if $H(X|Y)-D<R_3$, we send the Wyner-Ziv bin index over the private link, and send nothing over the cascade links, i.e., $R_1\ge0,R_2\ge0$ are achievable. The corresponding leakage rate is $\frac{1}{n}I(X^n;W_1,Z^n|\mathcal{C}_n)=\frac{1}{n}I(X^n;Z^n)=I(X;Z)$. The converse proof is given in [15]. ■

Theorem 4 (setting (B), logarithmic loss): The rate-distortion-leakage region $\mathcal{R}_{\text{tri(B), X-Y-Z, logloss}}$ under logarithmic loss distortion and $X-Y-Z$ assumption is the set of all tuples $(R_1,R_2,R_3,D,\triangle)\in\mathbb{R}_+^5$ that satisfy

$$R_1\ge[H(X|Y)-D-R_3]^+, \tag{4a}$$
$$R_2\ge[H(X|Y)-D-R_3]^+, \tag{4b}$$
$$\triangle\ge I(X;Z)+[H(X|Y)-D-R_3-H(Y|X,Z)]^+. \tag{4c}$$

Remark 1: It is interesting to note that although the availability of common side information Y^n at the encoder does not help in terms of rate-distortion tradeoff for a logarithmic loss distortion (like in the Gaussian case for Wyner-Ziv setting [16], [17]), it helps to reduce the leakage at the helper by allowing the encoder and the decoder to generate a secret key. We can see this from the leakage constraint (4c) above where the leakage rate consists of contributions from the eavesdropper's side information $I(X;Z)$ and from the eavesdropped source description which is partially protected by the secret key of rate $\min\{H(Y|X,Z), H(X|Y) - D - R_3\}$. This role of side information at the encoder in another secure source coding setting is also studied in [18].

Proof of Theorem 4: *Sketch of Achievability*: The proof follows similarly as in previous triangular case with the additional steps of secret key generation using y^n. We note also that the achievable scheme below is similar to that found in [18]. That is, the Wyner-Ziv coding at rate of $I(X;U|Y) + 2\epsilon = H(X|Y) - D + 2\epsilon$ is performed to satisfy the distortion constraint. Then we perform rate-splitting on the Wyner-Ziv index by splitting it into two parts, namely $w_1 \in [1 : 2^{n(H(X|Y)-D-R_3+\epsilon)}]$, and $w_3 \in [1 : 2^{n(R_3+\epsilon)}]$. If $H(X|Y) - D - R_3 > H(Y|X,Z)$, we further split w_1 into $w_{11} \in [1 : 2^{n(H(X|Y)-D-R_3-H(Y|X,Z)+\epsilon)}]$ and $w_{12} \in [1 : 2^{nH(Y|X,Z)}]$. Then the secret key k is generated by randomly and independently partitioning sequences in $\mathcal{Y}^n$ into $2^{nH(Y|X,Z)}$ bins and choosing k as the corresponding bin index of the given y^n. The encoder sends w_{11} and $w_{12} \oplus k$ over the cascade link, and w_3 over the private link, where $\oplus$ denotes addition over the $2^{nH(Y|X,Z)}$ field. The helper forwards the index w_{11} and $w_{12} \oplus k$ to the decoder. The decoder can recover w_{12} from its key generated by y^n. We can show that the tuples satisfying (4) where $[a]^+ = a$ in (4c) are achievable. If $H(X|Y) - D - R_3 < H(Y|X,Z)$, the secret key is generated by randomly and independently partitioning sequences in $\mathcal{Y}^n$ into $2^{n(H(X|Y)-D-R_3+\epsilon)}$ bins and choosing the corresponding bin index of given y^n as a key. The encoder sends $w_1 \oplus k$ over the cascade link, and w_3 over the private link. The helper forwards $w_1 \oplus k$ to the decoder. We can show that the tuples satisfying (4) where $[a]^+ = 0$ in (4c) are achievable. Detailed achievability and converse proofs are given in [15]. ■

Theorem 5 (setting (C)): The rate-distortion-leakage region $\mathcal{R}_{\text{tri(C),X-Z-Y}}$ is the set of all tuples $(R_1, R_2, R_3, D, \triangle) \in \mathbb{R}_+^5$ for which there exist random variables $U \in \mathcal{U}$ and $V \in \mathcal{V}$ such that $(U,V,X) - Z - Y$ forms a Markov chain, and a function $g : \mathcal{U} \times \mathcal{V} \times \mathcal{Y} \to \hat{\mathcal{X}}$ and

$$R_1 \geq I(X;U|Z), \tag{5a}$$
$$R_2 \geq I(X,Z;U|Y), \tag{5b}$$
$$R_3 \geq I(X,Z;V|U,Y), \tag{5c}$$
$$D \geq E[d(X, g(U,V,Y))], \tag{5d}$$
$$\triangle \geq I(X;U,Z). \tag{5e}$$

The cardinalities of $\mathcal{U}$ and $\mathcal{V}$ can be upperbounded by $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{Z}| + 3$ and $|\mathcal{V}| \leq (|\mathcal{X}||\mathcal{Z}| + 3)(|\mathcal{X}||\mathcal{Z}| + 1)$.

Proof: Since the achievable scheme essentially follows the *decode and re-bin* scheme of [6], it is omitted. As for the analysis of leakage rate, we get

$$\begin{aligned} I(X^n; W_1, Z^n|\mathcal{C}_n) &\leq I(X^n; Z^n|\mathcal{C}_n) + H(W_1|Z^n, \mathcal{C}_n) \\ &\overset{(a)}{\leq} n[I(X;Z) + I(X;U|Z) + \epsilon] = n[I(X;U,Z) + \epsilon], \end{aligned}$$

where (a) follows from the facts that (X^n, Z^n) are i.i.d. and independent of the codebook, and that we have $W_1 \in [1 : 2^{n(I(X;U|Z)+\epsilon)}]$. The converse proof is given in [15]. ■

Other Results: Apart from results for the logarithmic loss distortion, the Gaussian result for settings (A) can also be obtained. Furthermore, based on Fig. 2, we may consider a slightly different scenario where the encoder *broadcasts* the same source description to the helper and the decoder. Depending on the side information pattern, the helper may become useless in terms of providing extra information to the decoder. The complete characterization of the rate-distortion-leakage regions for above cases are given in [15].

REFERENCES

[1] A. E. Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[2] H. Permuter, Y. Steinberg, and T. Weissman, "Two-way source coding with a helper," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2905 –2919, June 2010.

[3] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *IEEE Trans. Inf. Theory*, to appear (arXiv:1110.3069 [cs.IT]).

[4] H. Yamamoto, "Source coding theory for cascade and branching communication systems," *IEEE Trans. Inf. Theory*, vol. 27, pp. 299–308, 1981.

[5] D. Vasudevan, C. Tian, and S. N. Diggavi, "Lossy source coding for a cascade communication system with side informations," in *Proc. Allerton Conf. Commun. Control Comput*, 2006.

[6] Y.-K. Chia, H. H. Permuter, and T. Weissman, "Cascade, triangular, and two-way source coding with degraded side information at the second user," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 189–206, 2012.

[7] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[8] Y. Liang, H. V. Poor, and S. Shamai (Shitz), "Information theoretic security," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 4–5, pp. 355–580, 2008.

[9] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "Utility-privacy tradeoff in databases: An information-theoretic approach," *CoRR, abs/1102.3751*.

[10] V. Prabhakaran and K. Ramchandran, "On secure distributed source coding," *Proc. IEEE Inf. Theory Workshop*, pp. 442–447, 2007.

[11] D. Gündüz, E. Erkip, and H. V. Poor, "Lossless compression with security constraints," *Proc. IEEE ISIT*, pp. 111–115, 2008.

[12] R. Tandon, S. Ulukus, and K. Ramchandran, "Secure source coding with a helper," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2178–2187, 2013.

[13] J. Villard and P. Piantanida, "Secure multiterminal source coding with side information at the eavesdropper," *submitted to IEEE Trans. Inf. Theory*, Mar 2011.

[14] E. Ekrem and S. Ulukus, "Secure lossy source coding with side information," in *Proc. Allerton Conf. Commun. Control Comput.*, 2011.

[15] K. Kittichokechai, Y. K. Chia, T. J. Oechtering, M. Skoglund, and T. Weissman, "Secure source coding with a public helper," 2013, in preparation. To be posted online at ArXiv.

[16] 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.

[17] 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.

[18] Y. K. Chia and K. Kittichokechai, "On secure lossy source coding with side information at the encoder," 2013, in preparation. To be posted online at ArXiv.