Some Results on Source Coding Problem with Privacy Condition

Bahman Moraffah
Department of Electrical Engineering
Arizona State University
Tempe, AZ, USA

July 21, 2014

Abstract

This article presents one-way source coding problem with additional condition on source output[1], Wyner-Ziv problem with additional output to be kept private from receivers[2] and finally find region for two-way communication systems with additional conditions on sources output at both locations. The region for above cases will be introduced such that they satisfy distortion prescription and equivocation constraints.

keywords: Rate-Distortion theory, Wyner-Ziv Problem, Kaspi Problem, Equivocation.

1 Introduction

Let us consider a source coding problem. The systems that will be considered are one-way and two-way communication systems with correlated sources. We go over one-way communication systems and Wyner-Ziv which have been introduced in [1],[2], then generalize the results to interactive source coding.In this article entropy or equivalently mutual information is considered as a measure of privacy.It is shown [4] interaction might help in this sense that by increasing the number of messages we can achieve less sum-rate. We show interaction might help to achieve more equivocation through an example. Let $\{X_k, X'_k\}_{k=1}^{\infty}$ be a sequence of i.i.d.

random variables where $X_k = (X_{k_r}, X_{k_h})$ and $X'_k = (X'_{k_r}, X'_{k_h})$ taking values in finite sets. The communication system in figure[1] will be analyzed.

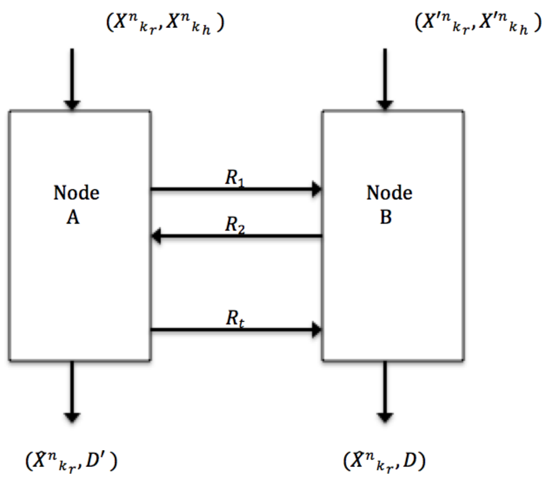

Figure 1: Two-way source coding scheme

Where other source output at each node must be kept private.

2 Formal Statement of The Problem and The Results

Let $\{\left(X_{k_{r_k}}, X_{k_{h_k}}\right), \left(X'_{k_{r_k}}, X'_{k_{h_k}}\right)\}_{k=1}^{\infty}$ be a sequence of i.i.d. random variables taking values in finite set $\mathcal{X}_{k_r}, \mathcal{X}_{k_h}, \mathcal{X}'_{k_r}, \mathcal{X}'_{k_r}.A\left(n, t, \{e_i\}_{i=1}^{t}, g_A, g_B, D, D', e, e'\right)$ equivocation - distortion code consists of

$$e_j : \mathcal{X}_{k_r} * \mathcal{X}_{k_h} \bigotimes_{i=1}^{j-1} \mathcal{M}_i \rightarrow \mathcal{M}_j \qquad j : odd \tag{1}$$

$$e_j : \mathcal{X}'_{k_r} * \mathcal{X}'_{k_h} \bigotimes_{i=1}^{j-1} \mathcal{M}_i \rightarrow \mathcal{M}_j \qquad j : even \tag{2}$$

$$g_A : \mathcal{X}_{k_r} * \mathcal{X}_{k_h} \bigotimes_{i=1}^{t} \mathcal{M}_i \rightarrow \mathcal{X}'_{k_r} \tag{3}$$

$$g_B : \mathcal{X'}_{k_r} * \mathcal{X'}_{k_h} \bigotimes_{i=1}^{t} \mathcal{M}_i \to \mathcal{X}_{k_r} \tag{4}$$

where $\mathcal{M}_i = \{1, 2, ..., M_i\}$ The average distortion of the code is given by

$$\Delta_x = E\left(\frac{1}{n}\sum_{k=1}^{n} d_x\left(X_{k_{r_k}}, \hat{X}_{k_{r_k}}\right)\right)$$

$$\Delta_{x'} = E\left(\frac{1}{n}\sum_{k=1}^{n} d_{x'}\left(X'_{k_{r_k}}, \hat{X}'_{k_{r_k}}\right)\right)$$

where $d_x, d_{x'}$ are per-letter distortion measure and measure of privacy is equivocation rate

$$\Delta E_1 = \frac{1}{n}H\left(X^n_{k_h} \mid M^t, X'^n_{k_r}X'^n_{k_h}\right)$$

$$\Delta E_2 = \frac{1}{n}H\left(X'^n_{k_h} \mid M^t, X^n_{k_r}, X^n_{k_h}\right)$$

(R_1, R_2, D, D', e, e') is achievable if there exists a $(n, t, \{e_i\}_{i=1}^t, g_A, g_B, D, D', e, e')$ code such that for any $\epsilon > 0$ and sufficiently large n, $\frac{1}{n}\log(\mid \mathcal{M}_j \mid) \leq R_j + \epsilon$ j = 1,...,t

$$R_{1\to2} = \sum_{j:odd} R_j$$

$$R_{2\to1} = \sum_{j:even} R_j$$

$$\Delta_x \leq D + \epsilon$$

$$\Delta_{x'} \leq D' + \epsilon$$

$$\Delta E_1 \leq e_1 - \epsilon$$

$$\Delta E_2 \leq e_2 - \epsilon$$

or equivalent conditions for last two conditions are

$$\frac{1}{n}I\left(X^n_{k_h}; M^t, X'^n_{k_r}, X'^n_{k_h}\right) \leq L_1 - \epsilon$$

$$\frac{1}{n}I\left(X'^n_{k_h}; M^t, X^n_{k_r}, X^n_{k_h}\right) \leq L_2 - \epsilon$$

Let us define$\mathcal{R}^*$ as set of all achievable $(R_{1\to2}, R_{2\to1}, D, D', e, e')$.
In addition of that $\mathcal{R}^*_{D-e}$ is distortion - equivocation achievable region. and equivocation function $E^*_{1\to2}(D, D')$, $E^*_{2\to1}(D, D')$and Rate-distortion-equivocation function $R^*_{1\to2}(D, D', e, e')$,$R^*_{2\to1}(D, D', e, e')$ as following:

$$R^*_{1\to2}(D, D', e, e') = \min_{(R_{1\to2}, R_{2\to1}, D, D', e, e') \in \mathcal{R}^*} R_{1\to2}$$

$$R^*_{2\to 1}(D, D', e, e') = \min_{(R_{2\to 1}, R_{1\to 2}, D, D', e, e')\in \mathcal{R}^*} R_{2\to 1}$$

$$E^*_{1\to 2}(D, D') = \max_{(D, D', e, e')\in \mathcal{R}^*_{D-e}} e$$

$$E^*_{2\to 1}(D, D') = \max_{(D, D', e, e')\in \mathcal{R}^*_{D-e}} e'$$

where $\mathcal{R}^*_{D-e} = \{(D, D', e, e') : (R_{1\to 2}, R_{2\to 1}, D, D', e, e')$ is achievable for some $R_{1\to 2} \geq 0 R_{2\to 1} \geq 0\}$.
Before considering general two-way communication systems we narrow down the problem to two special cases.
Consider one-way communication system Fig[2].

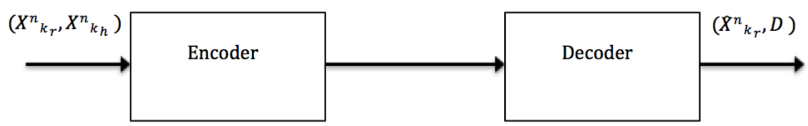

Figure 2: One-Way communication systems

Proposition 1: Consider one-way communication in Fig[2], we have

$$R^*(D, e) = \min_{P\left(\hat{X}_{k_r}|X_{k_h}, X_{k_r}\right): Ed\left(X_{k_r}, \hat{X}_{k_r}\right)\leq D, H\left(X_{k_h}|\hat{X}_{k_r}\right)\geq e} I\left(X_{k_r}, X_{k_h}; \hat{X}_{k_r}\right) \quad (5)$$

$E^*(D) = \max_{P\left(X_{k_r}, X_{k_h}, \hat{X}_{k_r}\right)\in \mathcal{P}(D)} H\left(X_{k_h} \mid \hat{X}_{k_r}\right)$

where $\mathcal{P}(D) := \bigcup_{H\left(X_{k_h}|\hat{X}_{k_r}\right)\leq e\leq H\left(X_{k_h}\right)} \mathcal{P}(D, e)$ where $\mathcal{P}(D, e)$is the family of probability distribution$P\left(\hat{X}_{k_r} \mid X_{k_h}, X_{k_r}\right)$such that

$$Ed\left(X_{k_r}, \hat{X}_{k_r}\right) \leq D$$

$$H\left(X_{k_h}|\hat{X}_{k_r}\right) \geq e$$

Proof : It has been proven in [1].

Now, consider Wyner-Ziv problem showed in Fig[3].

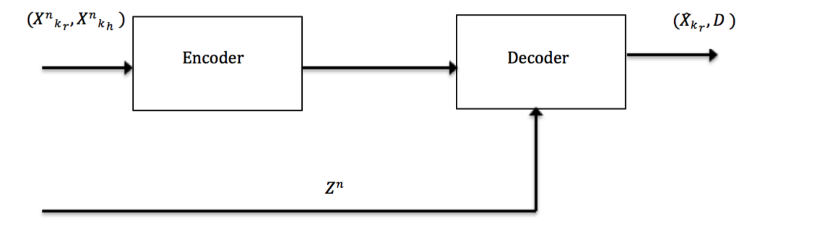

Figure 3: Source coding problem with side information at decoder

Proposition 2: Consider source Coding Problem with side information at decoder Fig[3] we have,

$$R^*(D, e) \geq I(X_{k_r}, X_{k_h}; U \mid Z)$$
$$E^*(D) \leq H(X_{k_h} \mid U, Z)$$

For some distribution $P(u \mid x_{k_r}, x_{k_h})$ such that there is a function $\hat{x}_{k_r} = f(u, z)$ for which $E\left(d\left(X_{k_r}, \hat{X}_{k_r}\right)\right) \leq D$ and $\mid \mathcal{U} \mid \leq \mid \mathcal{X}_k \mid +1$ where $\mid \mathcal{X}_k \mid$ is the cardinality of $\mathcal{X}_{k_r} \bigcup \mathcal{X}_{k_h}$

Proof: It has been proven in [2, Theorem 2]

Now, consider the general interaction source coding Fig[1]. We have the following Theorem:

Theorem: For distortion (D_1, D_2) the set of achievable $(R_{1\to 2}, R_{2\to 2}, E_{1\to 2}, E_{2\to 1})$ is given by:

$$R_{1\to 2} \geq R'_{1\to 2}(D_1, D_2, e_1, e_2) = I(X_1, Y_1; U^t | X_2, Y_2) \quad (6)$$

$$R_{2\to 1} \geq R'_{2\to 1}(D_1, D_2, e_1, e_2) = I(X_2, Y_2; U^t | X_1, X_2) \quad (7)$$

$$E_{1\to 2} \leq E'_{1\to 2}(D_1, D_2) = H(Y_1 | U^t, X_2, Y_2) \quad (8)$$

$$E_{2\to 1} \leq E'_{1\to 2}(D_1, D_2) = H(Y_2 | U^t, X_1, Y_1) \quad (9)$$

for some conditional pmf $\prod_{k=1}^{t} P(u_k | x_{j_k}, u^{k-1})$ and two functions $\hat{X}_1 = G(U^t, X_2, Y_2)$, $\hat{X}_2 = F(U^t, X_1, Y_1)$ such that $E(d_{x_1}(X_1, \hat{X}_1)) \leq D_1$, $E(d_{x_2}(X_2, \hat{X}_2)) \leq D_2$ and $\mid \mathcal{U}_k \mid \leq \mid \mathcal{X}_{j_k} \mid .(\prod_{j=1}^{k-1} \mid \mathcal{U}_j \mid) + 1$ where $j_k = 1$ if k is odd and $\mid \mathcal{X}_{j_k} \mid = \mid \mathcal{X}_1 \cup \mathcal{Y}_1 \mid$ and $j_k = 2$ if k is even and $\mid \mathcal{X}_{j_k} \mid = \mid \mathcal{X}_2 \cup \mathcal{Y}_2 \mid$

Proof: *Converse:* we now develop lower and upper bound on the rate and equivocation respectively. We show that given a $(n, t, \{e_k\}_{k=1}^{t}, g_A, g_B, D_1, D_2, e_1, e_2)$ code there exists a $P(X_1, X_2, Y_1, Y_2, U^t)$ such that the rate and equivocation of the system are bounded as below:

$$n(R_{1\to 2}+\epsilon) \geq \sum_{j:odd} H(M_j) \geq H(M_1, M_3, ..., M_{t-1} \mid X_2^n, Y_2^n)$$
$$\geq I(X_1^n, Y_1^n; M_1, M_3, ..., M_{t-1} \mid X_2^n, Y_2^n)$$
$$= H(X_1^n, Y_1^n) - H(X_1^n, Y_1^n \mid M_1, M_3, ..., M_{t-1}, X_2^n, Y_2^n)$$
$$= H(X_1^n, Y_1^n) - H(X_1^n, Y_1^n \mid M_1, M_2, ..., M_t, X_2^n, Y_2^n) \tag{10}$$
$$= \sum_{i=1}^{n} H(X_{1,i}, Y_{1,i}) - H(X_{1,i}, Y_{1,i} \mid M_1, ..., M_t, X^{i-1}{}_{1,i}, Y^{i-1}{}_{1,i}, X_2^n, Y_2^n)$$
$$\geq \sum_{i=1}^{n} H(X_{1,i}, Y_{1,i}) - H(X_{1,i}, Y_{1,i} \mid M^t, X^{i-1}{}_{1,i}, Y^{i-1}{}_{1,i}, X_{2,i}, Y_{2,i}, X_{2,i+1}^n, Y_{2,i+1}^n) \tag{11}$$

Now consider $U_{1i} = (X_{2,i+1}^n, Y_{2,i+1}^n, X_1^{i-1}, Y^{i-1}{}_1, M_1)$,$U_{ki} = M_k \forall k = 2, ..t$

So,

$$R_{1\to 2} + \epsilon \geq \frac{1}{n}\sum_{i=1}^{n} I(X_{1,i}, X_{2,i}; U_i^t \mid Y_{1,i}, Y_{2,i}) \tag{12}$$

We also have

$$E_{1\to 2} - \epsilon \leq \frac{1}{n} H(Y_1^n \mid M^t, X_2^n, Y_2^n) = \frac{1}{n}\sum_{i=1}^{n} H(X_{1,i} \mid M^t, X_1^n, Y_2^n, Y_{2,i+1}^n)$$
$$\leq \frac{1}{n}\sum_{i=1}^{n} H(Y_{2,i} \mid U^t, X_{2,i}, Y_{2,i})$$

$I(X_1, Y_1; U^t \mid X_2, Y_2)$ is non-increasing convex function of (D_1, D_2) and non-decreasing convex function of (e_1, e_2) [1].

<u>Define:</u>

$$Ed_{x_1}(X_{1,i}, \hat{X}_{1,i}) = d_i$$
$$Ed_{x_2}(X_{2,i}, \hat{X}'_{2,i}) = d'_i$$
$$H(Y_{1,i} \mid U^t, X_{2,i}, Y_{2,i}) = e_i$$
$$H(Y_{2,i} \mid U^t, X_{1,i}, Y_{1,i}) = e'_i$$

, then

$$D_1 + \epsilon \geq \frac{1}{n}\sum_{i=1}^{n} Ed_x(X_{1,i}, \hat{X}_{1,i}) = \frac{1}{n}\sum_{i=1}^{n} d_i$$

similarly, we have:

$$D_2 + \epsilon \geq \frac{1}{n}\sum_{i=1}^{n} Ed_{x'}(X_{2,i}, \hat{X}'_{2,i}) = \frac{1}{n}\sum_{i=1}^{n} d'_i$$

$$E_{1\to 2} - \epsilon \leq \frac{1}{n}\sum_{i=1}^{n} H(Y_{1,i} \mid U^t, X_{2,i}, Y_{2,i}) = \frac{1}{n}\sum_{i=1}^{n} e_i$$

$$R_{1\to 2} + \epsilon \geq \frac{1}{n}\sum_{i=1}^{n} I(X_{1,i}, Y_{1,i}; U_i^t \mid X_{2,i}, Y_{2,i})$$

$$\geq \sum_{i=1}^{n} \frac{1}{n} R'_{1\to 2}(d_i, d'_i, e_i, e'_i) \tag{13}$$

$$\geq \; R'_{1\to 2}(\frac{1}{n}\sum_{i=1}^{n} d_i, \frac{1}{n}\sum_{i=1}^{n} d'_i, \frac{1}{n}\sum_{i=1}^{n} e_i, \frac{1}{n}\sum_{i=1}^{n} e'_i) \tag{14}$$

$$\geq R'_{1\to 2}(D_1 + \epsilon, D_2 + \epsilon, e_1 - \epsilon, e_2 - \epsilon) \tag{15}$$

(10): $M_2, M_4, ...M_t$are functions of $M_1, M_3, ...M_{t-1}$andX_2^n, Y_2^n
(13): definition of the problem
(14): Jensen's inequality
(15)$R_{1\to 2}$ is non-increasing function of D_1, D_2 and non-decreasing function of e_1, e_2.

Achievability: For this proof we use binning coding scheme introduced in [4]. By using this method we know

$$R_{1\to 2} \geq I(X_1, Y_1; U^t | X_2, Y_2)$$

$$R_{2\to 1} \geq I(X_2, Y_2; U^t | X_1, Y_1)$$

are achievable. For this code, let us evaluate the equivocation rate. we have to prove:

$$\lim_{n\to +\infty} \frac{1}{n} H(Y_1^n \mid M^t, X_2^n, Y_2^n) \geq H(Y_1 | U^t, X_2 Y_2) - \epsilon$$

or equivalently

$$\lim_{n\to +\infty} \frac{1}{n} I(Y_1^n; M^t, X_2^n, Y_2^n) \leq I(Y_1 | U^t, X_2, Y_2) + \epsilon$$

We consider $I(Y_1^n; M^t, U_1^n, ..., U_t^n, X_2^n, Y_2^n)$

$$I(Y_1^n; M^t, U_1^n, ..., U_t^n, X_2^n, Y_2^n) = I(Y_1^n; M^t, X_2^n, Y_2^n) + I(Y_1^n; U_1^n, ..., U_t^n \mid M^t, X_2^n, Y_2^n)$$

$$= I(Y_1^n; M^t, X_1^n, Y_2^n) \tag{16}$$

$$I(Y_1^n; U_1^n, ..., U_t^n, X_2^n, Y_2^n) + I(Y_1^n; M^t \mid U_1^n, ..., U_t^n, X_2^n, Y_2^n)$$

we know that:

$$I(Y_1^n; U^t, X_2^n, Y_2^n) = nI(Y_1 | U^t, X_2, Y_2) \tag{17}$$

then because $I(Y_1^n; M^t \mid U_1^n, ..., U_t^n, X_2^n, Y_2^n) \geq 0$, we have:

$$I(Y_1^n; M^t, X_1^n, Y_2^n) \leq nI(Y_1 | U^t, X_2, Y_2)$$

(16): encoding scheme implies the decodability of $U_1^n, ..., U_t^n$ as follows: upon receiving the bin index. Decoder finds the unique u_1^n in the received bin such that (u_1^n, y^n) are jointly typical, then find u_2^n such that (u_1^n, u_2^n, y^n) are jointly typical. We keep doing this till we have a path with length t. So, we can decode unique $U_1^n, ..., U_t^n$ correctly with probability of error goes to zero. According to fano's inequality :

$$H(U_1^n, ..., U_t^n \mid M_1, M_2, ..., M_t, X^n) \leq n\delta(n)$$

$$H(U_1^n, ..., U_t^n \mid M_1, M_2, ..., M_t, Y^n) \leq n\delta(n)$$

$$H(U_1^n, ..., U_t^n \mid M_1, M_2, ..., M_t, X^n, Y^n) \leq n\delta(n)$$

where $n\delta(n) \to 0$ when $n \to \infty$. So, (16) $\to 0$when $n \to \infty$.

(17): because

$$I(Y_1^n; U^t, X_2^n, Y_2^n) = H(Y_1^n) - H(Y_1^n \mid U^t, X_2^n, Y_2^n)$$

so we just need to prove:

$$H(Y_1^n \mid U^t, X_2^n, Y_2^n) = nH(Y_1 \mid U^t, X_2, Y_2)$$

For this we have:

$$H(Y_1^n \mid U^t, X_2^n, Y_2^n) = \textstyle\sum_{u,x_2,y_2} P(U^t = u, X_2^n = x_2, Y_2^n = y_2) H(Y_1^n \mid u, x_2, y_2)$$
$$= \textstyle\sum_{u,x_2,y_2 \in \mathcal{T}_{U^t,X_2,Y_2}} P(u, x_2, y_2) H(Y_1^n \mid u, x_2, y_2)$$
$$+ \textstyle\sum_{u,x_2,y_2 \notin \mathcal{T}_{U^t,X_2,Y_2}} P(u, x_2, y_2) H(Y_2^n \mid u, x_2, y_2)$$

$$\leq \sum_{u,x_2,y_2 \in \mathcal{T}_{U^t,X_2,Y_2}} P(u, x_2, y_2) H(Y_1^n \mid u, x_2, y_2) + nH(Y_1)\delta(n) \tag{18}$$

$$= nH(Y_1)\delta(n)$$

$$+ \sum_{u,x_2,y_2 \in \mathcal{T}_{U^t,X_2,Y_2}} P(u,x_2,y_2)[-\sum_{y_1 \in \mathcal{T}_{Y_1|u,x_2,y_2}} P(y_1 \mid u,x_2,y_2)\log(P(y_1 \mid u,x_2,y_2)) \tag{19}$$

$$+ \sum_{u,x_2,y_2 \notin \mathcal{T}_{Y_1|u,x_2,y_2}} P(y_1 \mid u,x_2,y_2)\log(P(y_1 \mid u,x_2,y_2))] \tag{20}$$

$$\leq nH(Y_1 \mid U^t, X_2, Y_2) + n\epsilon(n) \tag{21}$$

where $n\epsilon(n) \to 0$ when $n \to \infty$.

(18):$P(u, x_2, y_2 \notin \mathcal{T}) \leq \delta(n)$ where $n\delta(n) \to 0$ when $n \to \infty$.
the first term in the last inequality holds because of our coding scheme and second term can be less than$\delta'(n)$.

3 Conclusions

The characterization of the achievable rate, distortion and equivocation region for the two-way source coding problem depicted in Fig. [1] does not involve the block length n. for the finite number of messages t we established the rate-distortion-equivocation single letter characterization. The rate-distortion-equivocation region for infinite number of messages is still unknown.

4 References

[1] H. Yamamoto, A Source Coding Problem for Source with Additional outputs to keep secret from receivers and wiretappers, IEEE Trans. Inform Theory, Vol. 29, No.6, pp. 918-923, Nov. 1983

[2] L. Sankar, S.R. Rajagopalan and H.V. Poor, Utility-Privacy Tradeoff in Database: An Information Theoretic Approach, IEEE Trans.Inform Theory, 2012

[3] N. Ma, Interaction Source Coding for Function Computation in Networks, Ph.D. Dissertation, Boston University, 2011

[4] A.H. Kaspi, Two-way Source Coding with a fidelity Criterion, IEEE Trans. Inform. Theory, 735-740, 1985

[5] A. Elgamal, Y.H. Kim, Network Information Theory,Cambridge University Press,2011