Source Coding for a Simple Network with Receiver Side Information

R. Timo[1,2], A. Grant[3], T. Chan[3] and G. Kramer[4]
[1]Department of Engineering, the Australian National University, Canberra, ACT, Australia.
[2]Networked Systems, NICTA, Canberra Research Laboratory, ACT, Australia.
[3]The Institute for Telecommunications Research, the University of South Australia, Adelaide, SA, Australia.
[4]Bell Laboratories, Alcatel-Lucent, Murray Hill, NJ, USA.
roy.timo@anu.edu.au, {alex.grant, terence.chan}@unisa.edu.au, gkr@research.bell-labs.com.

***Abstract*— We consider the problem of source coding with receiver side information for the simple network proposed by R. Gray and A. Wyner in 1974. In this network, a transmitter must reliably transport the output of two correlated information sources to two receivers using three noiseless channels: a public channel which connects the transmitter to both receivers, and two private channels which connect the transmitter directly to each receiver. We extend Gray and Wyner's original problem by permitting side information to be present at each receiver. We derive inner and outer bounds for the achievable rate region and, for three special cases, we show that the outer bound is tight.**

I. Introduction

The field of network source coding is centered on the following problem: given a noiseless communications network and a set of information sources, what is the best way to compress the output of each source for efficient and reliable transportation over the network? A solution to this type of problem needs to remove any temporal redundancy in each source, exploit any statistical correlations between different sources and optimize the use of limited channel capacities.

In network source coding, a *code* is a collection of rules that define how the output of each source is to be compressed, transported over the network and reconstructed. A code is said to be *reliable* if the output of each source can be reconstructed without error at each of its intended destinations. The *performance* of a reliable code is measured by the rates at which it sends data over each channel; an optimal code will send data at the smallest rates and thereby consume the least network capacity. An ordered collection of rates (one for each channel) is said to be *achievable* if there exists a reliable code which operates at these rates. The set of all achievable rates $\mathscr{R}$ is called the *achievable rate region* of the network, and its lower boundary $\overline{\mathscr{R}}$ provides a performance benchmark for the comparison of reliable codes.

NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

The work presented in this paper was undertaken by R. Timo while on visit at the Institute for Telecommunications Research, the University of South Australia, and Bell Laboratories, Alcatel-Lucent. R. Timo's travel was funded by student travel scholarships from NICTA, ARC Communications Research Network (ACoRN), and the University of South Australia.

The achievable rate region $\mathscr{R}$ is known for a small ad-hoc collection of networks; for most "real world" networks, $\mathscr{R}$ is unknown [1]. With the exception of [2], achievable rate regions have been studied on a network-by-network basis; researchers have designed and studied simple networks which isolate particular problems of interest. Two notable examples are: the separate coding of correlated sources [3], and the sharing of a finite capacity channel between multiple users [4]. It is hoped that solutions to these simple networks will yield practical and efficient codes for larger networks.

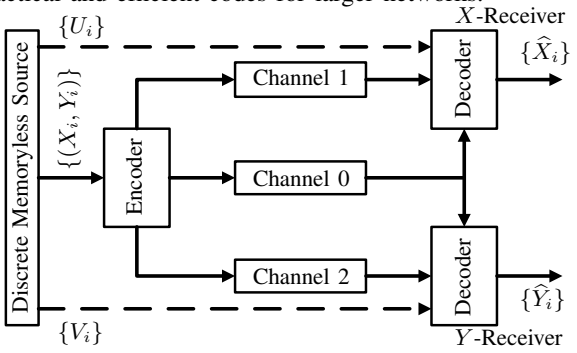


Fig. 1. Figure shows the network source coding problem proposed by R. Gray and A. Wyner [5]. The transmitter is connected to two receivers via three noiseless channels. The sequences $\{X_i\}$ and $\{Y_i\}$ are to be encoded at the transmitter, transported over the network and decoded at the x and y-receivers respectively. In this paper, we study an extension of this problem where "side information" $\{U_i\}$ and $\{V_i\}$ are present at each receiver. These additional information sources are marked with dashed lines in the figure.

We study the achievable rate region $\mathscr{R}$ of the network shown in Figure 1. A transmitter must transport the output of two correlated sources to two receivers using three noiseless channels: a public channel which connects the transmitter to both receivers, and two private channels which connect the transmitter directly to each receiver. The achievable rate region $\mathscr{R}$ of this network was found by R. Gray and A. Wyner [5] in 1974. They showed that an optimal code should endeavor to use the public channel to transport information common to both sources. As we will see, the intuition of this solution is lost when side information is introduced at each receiver; in particular, it is not clear how one should decompose the output of each source for transmission over the three channels.

An outline of the paper is as follows. To fix ideas, we briefly review [5] in Section II. In Section III, we formally define $\mathscr{R}$ for the network with side information. In Sections IV and V, we derive outer and inner bounds for $\mathscr{R}$ respectively. In Section VI, we ascertain $\mathscr{R}$ for one source, a degraded network and a complementary delivery network respectively. Finally, we conclude the paper in Section VII.

II. The Gray-Wyner Problem

Consider the network (without receiver side information) shown in Figure 1. We denote the capacities (in bits per second) of channels 0, 1 and 2 by C_0, C_1 and C_2 respectively. Finally, let $\mathscr{X}$ and $\mathscr{Y}$ be finite alphabets, and let $\mathscr{X}^n$ and $\mathscr{Y}^n$ denote their respective n-fold cartesian product spaces.

Suppose $\{(X_i, Y_i)\} \triangleq \{(X_i, Y_i);\ i = 1, 2, \ldots\}$ is a sequence of independent and identically distributed (i.i.d.) $\mathscr{X} \times \mathscr{Y}$ valued random variables emitted by a discrete memoryless source $\mathcal{Q}_{XY}(x, y) = \text{Prob}[X = x, Y = y]$. Suppose further that the random sequence $\{(X_i, Y_i)\}$ appears at the transmitter at the rate of one per second. It is desired that the transmitter delivers a reliable reproduction $\{\widehat{X}_i\} \triangleq \{\widehat{X}_i;\ i = 1, 2, \ldots\}$ of the sequence $\{X_i\}$ to the x-receiver, and a reliable reproduction $\{\widehat{Y}_i\} \triangleq \{\widehat{Y}_i;\ i = 1, 2, \ldots\}$ of the sequence $\{Y_i\}$ to the y-receiver. Assuming no delay constraints and unlimited computational power at the transmitter and receivers, the main problem is to ascertain which channel capacity triples (C_0, C_1, C_2) are both necessary and sufficient for each sequence to be reliably transported to its intended destination.

We assume the classic n-block source coding model where the sequence $\{(X_i, Y_i)\}$ is parsed and transported over the network in message blocks of length n (for some large integer n). Let $(X^n, Y^n) = (X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ denote the message at the transmitter, and let $\widehat{X}^n = \widehat{X}_1, \widehat{X}_2, \ldots, \widehat{X}_n$ and $\widehat{Y}^n = \widehat{Y}_1, \widehat{Y}_2, \ldots, \widehat{Y}_n$ denote the reconstructed messages at the x and y-receivers respectively.

For each $i = 0, 1, 2$, let $\mathscr{M}_i = \{1, 2, \ldots, |\mathscr{M}_i|\}$ be a finite index set for use on channel i. A network source code is a collection of mappings $(e^{(n)}, d_x^{(n)}, d_y^{(n)})$, where $e^{(n)} : \mathscr{X}^n \times \mathscr{Y}^n \to \mathscr{M}_0 \times \mathscr{M}_1 \times \mathscr{M}_2$ is the encoder at the transmitter; $d_x^{(n)} : \mathscr{M}_0 \times \mathscr{M}_1 \to \mathscr{X}^n$ is the decoder at the x-receiver; and $d_y^{(n)} : \mathscr{M}_0 \times \mathscr{M}_2 \to \mathscr{Y}^n$ is the decoder at the y-receiver. The transmitter encodes the pair (X^n, Y^n) with three indices $(M_0, M_1, M_2) = e^{(n)}(X^n, Y^n)$ which are sent over channels 0, 1 and 2 respectively. After receiving indices M_0 and M_1, the x-receiver reconstructs $\widehat{X}^n = d_x^{(n)}(M_0, M_1)$. Similarly, after receiving indices M_0 and M_2, the y-receiver reconstructs $\widehat{Y}^n = d_y^{(n)}(M_0, M_2)$. An error is said to occur if either $\widehat{X}^n \neq X^n$ or $\widehat{Y}^n \neq Y^n$, and the code is said to operate at a rate of $(1/n)\log_2 |\mathscr{M}_i|$ bits per source symbol on channel i (for $i = 0, 1, 2$).

A triple of rates (R_0, R_1, R_2) is said to be achievable if there exists a sequence of codes $\{(e^{(n)}, d_x^{(n)}, d_y^{(n)});\ n = 1, 2, \ldots\}$ such that the probability of error approaches zero and $(1/n)\log |\mathscr{M}_i|$ approaches R_i (for $i = 0, 1, 2$) as n goes to infinity.

Let $\mathscr{R}_{GW}$ denote the set of all achievable rate triples. It can be shown that $\mathscr{R}_{GW}$ is a closed convex subset of Euclidean three space, which is completely defined by its lower boundary $\overline{\mathscr{R}}_{GW}$ [5]:

$$\overline{\mathscr{R}}_{GW} \triangleq \big\{(R_0, R_1, R_2) \in \mathscr{R}_{GW} : (\widehat{R}_0, \widehat{R}_1, \widehat{R}_2) \in \mathscr{R}_{GW}, \\ \widehat{R}_i \leq R_i\ (i = 0, 1, 2) \to \widehat{R}_i = R_i\ (i = 0, 1, 2)\big\}\ .$$

Given $\mathcal{Q}_{XY}$ and a network with capacity triple (C_0, C_1, C_2), the sequences $\{X_i\}$ and $\{Y_i\}$ may be reliably reconstructed at the x and y-receivers respectively if and only if (C_0, C_1, C_2) lies above $\overline{\mathscr{R}}_{GW}$; thus, $\overline{\mathscr{R}}_{GW}$ defines exactly those capacity triples which are both necessary and sufficient for reliable communication.

Gray and Wyner [5] showed that to achieve rates (R_0, R_1, R_2) which lie on the lower boundary $\overline{\mathscr{R}}_{GW}$, the capacity of channel 0 should be prioritized for use by information common to both $\{X_i\}$ and $\{Y_i\}$. Specifically, they designed a coding scheme which used an auxiliary random variable W to represent the information transported over channel 0, and they showed any $(R_0, R_1, R_2) \in \overline{\mathscr{R}}_{GW}$ may be achieved by optimizing over the choice of W.

The formal description of $\mathscr{R}_{GW}$ in terms of W is as follow. Let $\mathscr{W}$ be a finite alphabet of cardinality $|\mathscr{W}| \leq |\mathscr{X}||\mathscr{Y}| + 2$, and let $\mathscr{P}_{GW}$ denote the family of probability functions on $\mathscr{W} \times \mathscr{X} \times \mathscr{Y}$ such that $\sum_w p(w, x, y) = \mathcal{Q}_{XY}(x, y)$. Now, for each $p \in \mathscr{P}_{GW}$, let

$$\mathscr{R}_{GW}^{(p)} \triangleq \left\{ (R_0, R_1, R_2)\ :\ \begin{array}{rcl} R_0 & \geq & I_p(X, Y; W) \\ R_1 & \geq & H_p(X|W) \\ R_2 & \geq & H_p(Y|W) \end{array} \right\}\ ,$$

where $I_p(\cdot;\cdot)$ denotes mutual information and $H_p(\cdot|\cdot)$ denotes conditional entropy (with respect to p).

Lemma 1: [5, Thm. 4] The achievable rate region $\mathscr{R}_{GW}$ of the Gray-Wyner Network is given by $\mathscr{R}_{GW} = \Big(\bigcup_{p \in \mathscr{P}_{GW}} \mathscr{R}_{GW}^{(p)}\Big)^c$, where $(\cdot)^c$ denotes the set closure.

It follows from Lemma 1 that $\mathscr{R}_{GW}$ is completely described by a single coding scheme which makes use of an auxiliary random variable W. As we will see, this coding scheme extends, in a natural way, to the network with side information. Unfortunately, however, this extension does not appear to completely describe the corresponding rate region.

III. Extension to the Side Information Case

Suppose $\mathscr{X}$, $\mathscr{Y}$, $\mathscr{U}$ and $\mathscr{V}$ are finite sets, and let $\mathscr{X}^n$, $\mathscr{Y}^n$, $\mathscr{U}^n$ and $\mathscr{V}^n$ denote their respective n-fold cartesian product spaces. Suppose further that $\{(X_i, Y_i, U_i, V_i)\}$ is a sequence of i.i.d. $\mathscr{X} \times \mathscr{Y} \times \mathscr{U} \times \mathscr{V}$ valued random variables emitted by a discrete memoryless source $\mathcal{Q}_{XYUV}(x, y, u, v) = \text{Prob}\big[X = x, Y = y, U = u, V = v\big]$. Finally, for each $i = 0, 1, 2$, let $\mathscr{M}_i = \{1, 2, \ldots, |\mathscr{M}_i|\}$ be a finite index set for channel i.

As before, a source code is a collection of mappings $(e^{(n)}, d_x^{(n)}, d_y^{(n)})$, where $e^{(n)} : \mathscr{X}^n \times \mathscr{Y}^n \to \mathscr{M}_0 \times \mathscr{M}_1 \times \mathscr{M}_2$ is the encoder at the transmitter; $d_x^{(n)} : \mathscr{M}_0 \times \mathscr{M}_1 \times \mathscr{U}^n \to \mathscr{X}^n$ is the decoder at the x-receiver; and $d_y^{(n)} : \mathscr{M}_0 \times \mathscr{M}_2 \times \mathscr{V}^n \to \mathscr{Y}^n$ is the decoder at the y-receiver. The transmitter encodes the

pair (X^n, Y^n) with indices $(M_0, M_1, M_2) = e^{(n)}(X^n, Y^n)$ which are sent over channels 0, 1 and 2 respectively. After receiving indices M_0 and M_1 as-well-as side information U^n, the x-receiver reconstructs $\widehat{X}^n = d_x^{(n)}(M_0, M_1, U^n)$. Similarly, after receiving M_0, M_2 and V^n, the y-receiver reconstructs $\widehat{Y}^n = d_y^{(n)}(M_0, M_2, V^n)$.

An error occurs if either $\widehat{X}^n \neq X^n$ or $\widehat{Y}^n \neq Y^n$. Let $P_{e,x} \triangleq \text{Prob}[\widehat{X}^n \neq X^n]$, $P_{e,y} \triangleq \text{Prob}[\widehat{Y}^n \neq Y^n]$ and $P_e \triangleq \max\{P_{e,x},\ P_{e,y}\}$.

Definition 1 (Achievable Rate): A rate triple (R_0, R_1, R_2) is said to be achievable if, for arbitrary $\epsilon > 0$ and sufficiently large n, there exists a code $(e^{(n)}, d_x^{(n)}, d_y^{(n)})$ with parameters $(n, |\mathscr{M}_0|, |\mathscr{M}_1|, |\mathscr{M}_2|, P_e)$ such that $P_e \leq \epsilon$ and $(1/n)\log|\mathscr{M}_i| \leq R_i + \epsilon$ for all $i = 0, 1, 2$. We let $\mathscr{R}$ denote the set of all achievable rate triples.

IV. An Outer Bound

Suppose $\mathscr{W}$ is a finite set of cardinality $|\mathscr{W}| \leq |\mathscr{X}||\mathscr{Y}| + 3$ and $\mathscr{P}$ is the family of probability functions on $\mathscr{W} \times \mathscr{X} \times \mathscr{Y} \times \mathscr{U} \times \mathscr{V}$ such that $p(w, x, y, u, v) = p(w|x, y)p(x, y, u, v)$ and

$$\mathcal{Q}_{XYUV}(x, y, u, v) = \sum_{w \in \mathscr{W}} p(w, x, y, u, v)$$

for all $p \in \mathscr{P}$. Now, for each $p \in \mathscr{P}$ let

$$\mathscr{R}_{out}^{(p)} = \Big\{(R_0, R_1, R_2) : \left.\begin{array}{rcl} R_0 & \geq & \max\big\{I_p(X,Y;W|U), I_p(X,Y;W|V)\big\} \\ R_0 + R_1 & \geq & \max\big\{I_p(X,Y;W|U), I_p(X,Y;W|V)\big\} \\ & & \quad + H_p(X|W,U), \\ R_0 + R_2 & \geq & \max\big\{I_p(X,Y;W|U), I_p(X,Y;W|V)\big\} \\ & & \quad + H_p(Y|W,V). \end{array}\right\}$$

Theorem 1 (Outer Bound): If (R_0, R_1, R_2) is an achievable rate triple, then there exists a $p \in \mathscr{P}$ such that $(R_0, R_1, R_2) \in \mathscr{R}_{out}^{(p)}$.

A. Proof Outline: Theorem 1

We show: if $\{(e^{(n)}, d_x^{(n)}, d_y^{(n)})\}$ is a sequence of codes where $P_e \to 0$ as $n \to \infty$, then there exists a $p \in \mathscr{P}$ such that $\big((1/n)\log|\mathscr{M}_0|,\ (1/n)\log|\mathscr{M}_1|,\ (1/n)\log|\mathscr{M}_2|\big) \in \mathscr{R}_{out}^{(p)}$.

Suppose $(e^{(n)}, d_x^{(n)}, d_y^{(n)})$ is a code with $(M_0, M_1, M_2) = e^{(n)}(X^n, Y^n)$, $\widehat{X}^n = d_x^{(n)}(M_0, M_1, U^n)$ and $\widehat{Y}^n = d_y^{(n)}(M_0, M_2, V^n)$, then

$$\begin{aligned} \log|\mathscr{M}_0| &\geq H(M_0|U^n) \geq I(X^n, Y^n; M_0|U^n) \\ &= \sum_{i=1}^{n} I(X_i, Y_i; M_0, X_1^{i-1}, X_1^{i-1}, U_1^{i-1}, U_{i+1}^n|U_i) \quad (1) \\ &\geq \sum_{i=1}^{n} I(X_i, Y_i; M_0|U_i) = \sum_{i=1}^{n} I(X_i, Y_i; W_i|U_i)\ , \quad (2) \end{aligned}$$

where (1) follows because $\{(X_i, Y_i, U_i, V_i)\}$ is drawn in an i.i.d. fashion and (2) follows by setting $W_i = M_0$. Similarly,

$$\log|\mathscr{M}_0| \geq \sum_{i=1}^{n} I(X_i, Y_i; W_i|V_i)\ . \quad (3)$$

On applying Fano's Inequality [6, Pg. 37] we get

$$H(X^n|M_0, M_1, U^n) \leq H(X^n|\widehat{X}^n) \leq n\delta(P_e, n)\ , \quad (4)$$

where $\delta(P_e, n) \triangleq (1/n) + P_e \log|\mathscr{X}||\mathscr{Y}|$. Similarly, we also have that $H(Y^n|M_0, M_1, V^n) \leq n\delta(P_e, n)$.

Now consider the series of Shannon (in)equalities (5) through (12). Note, (7) follows because $\{(X_i, Y_i, U_i, V)\}$ is drawn in an i.i.d. fashion and (11) follows since $M_0 \ominus (X_i, Y_i) \ominus (U_i, V_i)$ forms a Markov Chain and (4). From (10) and (12), it respectively follows that

$$\begin{aligned} &\frac{1}{n}\Big(\log|\mathscr{M}_0| + \log|\mathscr{M}_1|\Big) \\ &\geq \frac{1}{n}\sum_{i=1}^{n}\Big[I(X_i, Y_i; W_i|U_i) + H(X_i|W_i, U_i)\Big] - \delta(P_e, n)\ , \end{aligned}$$

and

$$\begin{aligned} &\frac{1}{n}\Big(\log|\mathscr{M}_0| + \log|\mathscr{M}_1|\Big) \\ &\geq \frac{1}{n}\sum_{i=1}^{n}\Big[I(X_i, Y_i; W_i|V_i) + H(X_i|W_i, U_i)\Big] - \delta(P_e, n)\ . \end{aligned}$$

Note, $(1/n)[\log|\mathscr{M}_0| + \log|\mathscr{M}_2|]$ may be bound in a similar manner. Following the time sharing principle given in [5, Pg. 1709], we may now construct a $p \in \mathscr{P}$ such that each inequality in the theorem holds as $n \to \infty$ and $P_e \to 0$. Finally, we may bound the cardinality of the auxiliary random variable W using the support lemma of Ahlswede and Körner [7, Lemma 3].

V. An Inner Bound

A natural extension of the code proposed by Gray and Wyner [5] yields the following inner bound for $\mathscr{R}$.

Let $\mathscr{W}$ and $\mathscr{P}$ be defined as in Section III. For $p \in \mathscr{P}$, let

$$\mathscr{R}_{in}^{(p)} = \Big\{(R_0, R_1, R_2) : \left.\begin{array}{rcl} R_0 & \geq & \max\big\{I_p(X,Y;W|U), I_p(X,Y;W|V)\big\} \\ R_1 & \geq & H_p(X|W,U), \\ R_2 & \geq & H_p(Y|W,V). \end{array}\right\},$$

and $\mathscr{R}_{in} = \big(\cup_{p \in \mathscr{P}} \mathscr{R}_{in}^{(p)}\big)^c$.

Theorem 2: $\mathscr{R} \supseteq \mathscr{R}_{in}$.

Remark 1: If $U = V$, then $\mathscr{R} = \mathscr{R}_{in}$.

Remark 2: Suppose $(X, Y) \ominus U \ominus V$ forms a Markov chain. It can be shown that a sum rate $R_0 + R_1 + R_2$ is achievable if and only if $R_0 + R_1 + R_2 \geq H(Y|V) + H(X|Y, U)$. (See [8] for the special case where $V =$ constant.) We may set $W = Y$ in Theorem 2 to achieve this sum rate.

Remark 3: Suppose $X = Y$. Sgarro [9] showed that the sum rate $R_0 + R_1 + R_2$ is achievable if and only if $R_0 + R_1 + R_2 \geq \max\{H(X|U),\ H(X|V)\}$. We may set $W = X = Y$ in Theorem 2 to achieve this sum rate.

Remark 4: Suppose $U = Y$ and $V = X$. Wyner *et. al.* [4] showed that the sum rate $R_0 + R_1 + R_2$ is achievable if and only if $R_0 + R_1 + R_2 \geq \max\{H(X|Y),\ H(Y|X)\}$. We may set $W = (X, Y)$ in Theorem 2 to achieve this sum rate.

$$\log|\mathscr{M}_0| + \log|\mathscr{M}_1| \geq H(M_0, M_1) = H(M_0, M_1|U^n) + I(M_0, M_1; U^n) \tag{5}$$
$$\geq I(X^n, Y^n; M_0, M_1|U^n) + I(M_0, M_1; U^n) \tag{6}$$
$$= \sum_{i=1}^{n} \big[I(X_i, Y_i; M_0, M_1, X_1^{i-1}, Y_1^{i-1}, U_1^{i-1}, U_{i+1}^n|U_i) + I(U_i; M_0, M_1, U_1^{i-1})\big] \tag{7}$$
$$\geq \sum_{i=1}^{n} \big[I(X_i, Y_i; M_0, M_1, U_1^{i-1}, U_{i+1}^n|U_i) + I(U_i; M_0)\big] \tag{8}$$
$$= \sum_{i=1}^{n} \big[I(X_i, Y_i; M_0|U_i) + I(X_i, Y_i; M_1, U_1^{i-1}, U_{i+1}^n|M_0, U_i) + I(U_i; M_0)\big] \tag{9}$$
$$\geq \sum_{i=1}^{n} \big[I(X_i, Y_i; M_0|U_i) + I(X_i; M_1, U_1^{i-1}, U_{i+1}^n|M_0, U_i) + I(U_i; M_0)\big] \tag{10}$$
$$= \sum_{i=1}^{n} \big[I(X_i, Y_i; M_0|V_i) + H(X_i|M_0, U_i) - n\delta(P_e, n)\big] \tag{11}$$
$$= \sum_{i=1}^{n} \big[I(X_i, Y_i; W_i|V_i) + H(X_i|W_i, U_i) - n\delta(P_e, n)\big] \tag{12}$$

Remark 5: The code, which yields the achievability of $\mathscr{R}_{in}$, is essentially a version of Heegard and Berger's "triple rate split code" given in [10, Thm. 2]. Indeed, we note that the problem of minimizing the sum rate $R_0+R_1+R_2$ is a special case of the two receiver generalized Kaspi-Heegard-Berger problem [10, Sec. VII].

A. Proof Outline: Theorem 2

1) Code Construction: Suppose $p \in \mathscr{P}$. Let R_0', R_1' and R_2' be non-negative integers whose values will be chosen later. Generate $2^{nR_0'}$ independent w-codewords of length n by choosing symbols i.i.d. from $\mathscr{W}$ according to p_W (the W-marginal of p). Label the resulting code book with the index m_0': $\mathscr{C}_{\mathscr{W}} \triangleq \{w^n(m_0') : 1 \leq m_0' \leq 2^{nR_0'}\}$. Similarly, generate $2^{nR_1'}$ and $2^{nR_2'}$ independent x and y-codewords using p_X and p_Y respectively: $\mathscr{C}_{\mathscr{X}} \triangleq \{x^n(m_1') : 1 \leq m_1' \leq 2^{nR_1'}\}$, and $\mathscr{C}_{\mathscr{Y}} \triangleq \{y^n(m_2') : 1 \leq m_2' \leq 2^{nR_2'}\}$.

Uniformly at random assign to each $w^n \in \mathscr{C}_{\mathscr{W}}$ a "bin label" from the set $\mathscr{M}_0 = \{1, 2, \ldots, 2^{\lfloor nR_0 \rfloor}\}$, and let $h_{\mathscr{W}} : \mathscr{C}_{\mathscr{W}} \to \mathscr{M}_0$ denote the induced mapping. Let $\mathscr{B}_{\mathscr{W}}(m_0)$ denote the set of w-codewords with bin label m_0: $\mathscr{B}_{\mathscr{W}}(m_0) \triangleq \{w^n \in \mathscr{C}_{\mathscr{W}} : h_{\mathscr{W}}(w^n) = m_0\}$, and let $\mathscr{B}_{\mathscr{W}}$ denote the collection of all w-bins. In the same way, assign one of $2^{\lfloor nR_1 \rfloor}$ and $2^{\lfloor nR_2 \rfloor}$ bin labels to each x and y-codeword, and define $h_{\mathscr{X}}$, $h_{\mathscr{Y}}$, $\mathscr{B}_{\mathscr{X}}$ and $\mathscr{B}_{\mathscr{Y}}$.

2) Encoding: The encoder assumes the messages x^n, y^n, u^n and v^n emitted by the source are ϵ-strong joint typical; that is, $(x^n, y^n, u^n, v^n) \in \mathcal{A}_\epsilon^{*(n)}(p_{XYUV})$. Let E_1 denote the event where this assumption is false. Then [6, Lem. 10.6.1]

$$\Pr\big[E_1\big] \leq \epsilon_1(n, \mathscr{X} \times \mathscr{Y} \times \mathscr{U} \times \mathscr{V}) \ , \tag{13}$$

where $\epsilon_1(n, \mathscr{X} \times \mathscr{Y} \times \mathscr{U} \times \mathscr{V}) \to 0$ in n for fixed $\epsilon > 0$.

The transmitter looks for a $w^n(m_0') \in \mathscr{C}_{\mathscr{W}}$ which is ϵ-strong joint typical with (x^n, y^n). If two-or-more such codewords exist, the transmitter selects the codeword with the smallest index. If no such codeword exists, an error is declared and the transmitter arbitrarily selects some $w_e^n(m_0') \in \mathscr{C}_{\mathscr{W}}$. Let E_2 denote this error event. Then [6, Lem. 10.6.2],

$$\Pr\big[E_2\big] \leq e^{-\left(2^{nR_0'}2^{-n(I(X,Y;W)+\epsilon_2)}\right)} \ , \tag{14}$$

where $\epsilon_2 \to 0$ as $\epsilon \to 0$ and $n \to \infty$. We assume $R_0' \geq I(X, Y; W) + \epsilon_2$, so that $\Pr[E_2] \to 0$ as $\epsilon \to 0$ and $n \to \infty$. After the transmitter selects $w^n(m_0') \in \mathscr{C}_{\mathscr{W}}$ it sends the index $m_0 = h_{\mathscr{W}}(w^n(m_0'))$ on channel 0.

The transmitter looks for a $x^n(m_1') \in \mathscr{C}_{\mathscr{X}}$ such that $x^n(m_1') = x^n$. If two-or-more such codewords exist, the transmitter selects the codeword with the smallest index. If no such codeword exists, an error is declared and the transmitter arbitrarily selects some $x_e^n(m_1') \in \mathscr{C}_{\mathscr{X}}$. Let $E_{3,x}$ denote this error event. Then,

$$\Pr\big[E_{3,x}\big] \leq e^{-\left(2^{nR_1'}2^{-n(H(X)+\epsilon_{3,x})}\right)} \ , \tag{15}$$

where $\epsilon_{3,x} \to 0$ as $\epsilon \to 0$ and $n \to \infty$. Choose $R_0' \geq H(X) + \epsilon_{3,x}$ arbitrarily, so that $\Pr[E_{3,x}] \to 0$ as $\epsilon \to 0$ and $n \to \infty$. The transmitter encode y^n is a similar fashion, and sends $m_1 = h_{\mathscr{X}}(x^n(m_1'))$ and $m_2 = h_{\mathscr{Y}}(y^n(m_2'))$ on channels 1 and 2 respectively.

3) Decoding: Given m_0 and u^n, the X-receiver looks for a unique $\widehat{w}^n \in \mathscr{B}_{\mathscr{W}}(m_0)$ which is jointly typical with u^n. If no such codeword can be found, an error is declared and the decoder arbitrarily selects some $\widehat{w}_e^n \in \mathscr{B}_{\mathscr{W}}(m_0)$. Let

- $E_{4,x}$: the codeword $w^n(m_0')$ chosen by the transmitter is not jointly typical with u^n, and
- $E_{5,x}$: there are two-or-more w-codewords in $\mathscr{B}_{\mathscr{W}}(m_0)$ which are jointly typical with u^n.

Consider $E_{4,x}$. Since $W \ominus (X,Y) \ominus U$ forms a Markov Chain under p, we have that [6, Lem. 15.8.1]

$$\Pr\big[E_{4,x}\big] \leq \epsilon_{4,x} \ , \tag{16}$$

where $\epsilon_{4,x} \to 0$ as $n \to \infty$.

Now consider $E_{5,x}$. We have that $u^n \in \mathcal{A}_\epsilon^{*(n)}(P_U)$. As before, the probability that a randomly generated w-codeword is jointly typical with u^n is upper bound by $2^{-n(I(W;U)+\epsilon_{5,x})}$, where $\epsilon_{5,x} \to 0$ as $n \to \infty$. Moreover, the number of codewords in each bin is at most $2^{n(R_0'-R_0)} + \epsilon_{5',x}$, where $\epsilon_{5',x} \to 0$ as $n \to \infty$ [11, Pg. 2766]. Hence,

$$\Pr\big[E_{5,x}\big] \leq 2^{-n(R_0-R_0'+I(W;U)-\epsilon_{5,x})} + \epsilon_{5',x} \ .$$

We need $R_0 - R_0' + I(W;U) - \epsilon_{5,x} \geq 0$, so that $\Pr\big[E_{5,x}\big] \to 0$ as $n \to \infty$. This requires

$$\begin{aligned} R_0 &\geq R_0' - I(W;U) + \epsilon_{5,x} \\ &\geq I(X,Y;W) - I(W;U) + \epsilon_2 + \epsilon_{5,x} && (17) \\ &= I(X,Y,U;W) - I(W;U) + \epsilon_2 + \epsilon_{5,x} && (18) \\ &= I(X,Y;W|U) + \epsilon_2 + \epsilon_{5,x} \ , && (19) \end{aligned}$$

where (17) follows because we selected $R_0' \geq I(X,Y;W) + \epsilon_2$, (18) follows because $W \ominus (X,Y) \ominus U$ forms a Markov Chain, and (19) follows from the chain rule for mutual information. Similarly, the y-receiver will correctly find a w-codeword with high probability if $R_0 \geq I(X,Y;W|V) + \epsilon_2 + \epsilon_{5,y}$, where $\epsilon_{5,y} \to 0$ as $\epsilon \to 0$ and $n \to 0$.

Given $\widehat{w}^n$, m_1 and u^n, the X-receiver looks for a unique $\widehat{x}^n \in \mathscr{B}_{\mathscr{X}}(m_1)$ which is jointly typical with $\widehat{w}^n$ and u^n. If there exists two-or-more such codewords, an error is declared and the decoder arbitrarily selects some $\widehat{x}_e^n \in \mathscr{B}_{\mathscr{X}}(m_1)$. Let $E_{6,x}$ denote this error event. It follows that

$$\Pr\big[E_{6,x}\big] \leq 2^{-n(R_1-R_1'+I(X;W,U)-\epsilon_{6,x})} + \epsilon_{6',x} \ , \tag{20}$$

where $\epsilon_{6,x} \to 0$ and $\epsilon_{6',x} \to 0$ as $\epsilon \to 0$ and $n \to \infty$. If $R_1 \geq H(X|W,U) + \epsilon_{6,x}$ it follows from (20) that $\Pr[E_{6,x}] \to 0$ as $n \to \infty$. Similarly, the y-receiver will correctly find $\widehat{y}^n$ with high probability if $R_2 \geq H(Y|W,V) + \epsilon_{6,y}$, where $\epsilon_{6,y} \to 0$ as $\epsilon \to 0$ and $n \to 0$.

VI. Three Simple Networks

A. Two Descriptions of $\mathscr{R}$ when $X = Y$

Let $\mathscr{P}$ and $\mathscr{R}_{out}^{(p)}$ be defined as in Section IV.

Theorem 3: If $X = Y$, then $\mathscr{R} = \Big(\cup_{p\in\mathscr{P}} \mathscr{R}_{out}^{(p)}\Big)^c$.

Now suppose $\mathscr{A}$ and $\mathscr{B}$ are finite sets of cardinalities $|\mathscr{A}| \leq |\mathscr{X}|+1$ and $|\mathscr{B}| \leq |\mathscr{X}|+1$. Let $\mathscr{P}^*$ denote the family of probability functions on $\mathscr{A}\times\mathscr{B}\times\mathscr{X}\times\mathscr{U}\times\mathscr{V}$ such that $p(a,b,x,u,v) = p(a,b|x)p(x,u,v)$ and $\mathcal{Q}_{XUV}(x,u,v) = \sum_{(a,b)\in\mathscr{A}\times\mathscr{B}} p(a,b,x,u,v)$. For each $p \in \mathscr{P}^*$, let

$$\mathscr{R}^{(p)*} = \left\{(R_0,R_1,R_2): \begin{array}{rcl} R_0 & \geq & \max\big\{H_p(X|A,U), H_p(X|B,V)\big\} \\ R_1 & \geq & I_p(X;A|U), \\ R_2 & \geq & I_p(X;B|V). \end{array}\right\} .$$

Theorem 4: If $X = Y$, then $\mathscr{R} = \big(\cup_{p\in\mathscr{P}^*} \mathscr{R}^{(p)*}\big)^c$.

B. $\mathscr{R}$ for a Type of Degraded Network

Let $\mathscr{P}$ and $\mathscr{R}_{out}^{(p)}$ be defined as in Section IV.

Theorem 5: If $Y = (X,Z)$ and $(X,Z) \ominus U \ominus V$ forms a Markov Chain, then $\mathscr{R} = \Big(\cup_{p\in\mathscr{P}} \mathscr{R}_{out}^{(p)}\Big)^c$.

C. $\mathscr{R}$ for a Complementary Delivery Network

Let $\mathscr{P}$ and $\mathscr{R}_{out}^{(p)}$ be defined as in Section IV.

Theorem 6: If $U = Y$ and $V = X$, then $\mathscr{R} = \Big(\cup_{p\in\mathscr{P}} \mathscr{R}_{out}^{(p)}\Big)^c$.

Now suppose $\mathscr{A}$ and $\mathscr{B}$ are finite sets of cardinalities $|\mathscr{A}| \leq |\mathscr{X}||\mathscr{Y}| + 1$ and $|\mathscr{B}| \leq |\mathscr{X}||\mathscr{Y}| + 1$. Let $\mathscr{P}^{**}$ denote the set of probability functions on $\mathscr{A}\times\mathscr{B}\times\mathscr{X}\times\mathscr{Y}$ such that $\mathcal{Q}_{XY}(x,y) = \sum_{(a,b)\in\mathscr{A}\times\mathscr{B}} p(a,b,x,y)$ is true for all (x,y) and $p \in \mathscr{P}^{**}$. For each $p \in \mathscr{P}^{**}$, let

$$\mathscr{R}^{(p)**} = \left\{(R_0,R_1,R_2): \begin{array}{rcl} R_0 & \geq & \max\big\{H_p(X|A,Y), H_p(Y|B,X)\big\} \\ R_1 & \geq & I_p(X;A|Y), \\ R_2 & \geq & I_p(Y;B|X). \end{array}\right\} .$$

Theorem 7: If $U = Y$ and $V = X$, then $\mathscr{R} = \big(\cup_{p\in\mathscr{P}^{**}} \mathscr{R}^{(p)**}\big)^c$.

VII. Conclusion

We investigated the achievable rate region $\mathscr{R}$ of a simple network with side information present at each receiver. Our first theorem gave an outer bound which, for three simple networks, was shown to be equal to $\mathscr{R}$. Our second result gave an inner bound which was obtained via an extension of the coding theorem given by Gray and Wyner [5].

References

[1] M. Effros, "Network Source Coding: A Perspective," *IEEE Inform. Theory Society Newsletter*, vol. 57, no. 4, pp. 15–23, December 2007.

[2] I. Csiszar and J. Körner, "Towards a General Theory of Source Networks," *IEEE Transactions on Information Theory*, vol. 26, no. 2, pp. 155–165, March 1980.

[3] D. Slepian and J. Wolf, "Noiseless Coding of Correlated Sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, July 1973.

[4] A. Wyner, J. Wolf, and F. Willems, "Communicating Via Processing Broadcast Satellite," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1243–1249, June 2002.

[5] R. Gray and A. Wyner, "Source Coding for a Simple Network," *Bell System Technical Journal*, vol. 53, no. 9, pp. 1681–1721, Nov. 1974.

[6] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. Wiley, 2006.

[7] R. F. Ahlswede and J. Körner, "Source Coding with Side Information and a Converse for Degraded Broadcast Channels," *IEEE Transactions on Information Theory*, vol. 21, no. 6, pp. 629–637, November 1975.

[8] R. Timo, A. Grant, and L. Hanlen, "Source Coding for a Noiseless Broadcast Channel with Partial Receiver Side Information," in *Proceedings IEEE Australian Communications Theory Workshop, AusCTW*, Adelaide, Australia, February 2007.

[9] A. Sgarro, "Source Coding with Side Information at Several Decoders," *IEEE Transactions on Information Theory*, vol. 23, no. 2, pp. 179–182, March 1977.

[10] C. Heegard and T. Berger, "Rate Distortion when Side Information May Be Absent," *IEEE Transactions on Information Theory*, vol. 31, no. 6, pp. 727–734, November 1985.

[11] M. Gastpar, "The Wyner Ziv Problem With Multiple Sources," *IEEE Transactions on Information Theory*, vol. 50, no. 11, pp. 2762–2768, November 2004.