

# On Achievable Rates for Multicast in the Presence of Side Information

Mayank Bakshi, Michelle Effros  
 Department of Electrical Engineering  
 California Institute of Technology  
 Pasadena, California 91125, USA  
 Email: {mayank, effros}@caltech.edu

**Abstract**—We investigate the network source coding rate region for networks with multiple sources and multicast demands in the presence of side information, generalizing earlier results on multicast rate regions without side information. When side information is present only at the terminal nodes, we show that the rate region is precisely characterized by the cut-set bounds and that random linear coding suffices to achieve the optimal performance. When side information is present at a non-terminal node, we present an achievable region. Finally, we apply these results to obtain an inner bound on the rate region for networks with general source-demand structures.

## I. INTRODUCTION

The rate region for a multicast network was characterized for independent sources by Ahlswede et al. in [1], wherein the achievability of the cut-set bounds was shown. For dependent sources, Ho et al. [2] proved that cut-set bounds are again tight for multicast demands.

In this paper, we generalize these results to incorporate jointly distributed source and side information random variables in multicast networks. In Theorem 1, we show that the cut-set region is achievable with linear coding when each sink is a terminal node with access to a distinct side information random variable. (See Fig. 1.) Theorem 2 gives an alternative proof of this achievability result that does not rely on linear codes; this approach allows us to use unstructured random binning in the derivations that follow. Theorem 3 generalizes Theorem 2 to allow side information at one non-sink node. Theorem 4 applies the above results to find an achievable region for networks with general source-demand structures.

## II. PRELIMINARIES

A network  $\mathcal{N} = (V, E)$  is a directed, acyclic graph with vertex set  $V$  and noiseless edge set  $E \subseteq (V \times V)$ . For each  $v \in V$ , we use  $\Gamma_i(v) \subseteq E$  and  $\Gamma_o(v) \subseteq E$  to denote the incoming and outgoing edges, respectively, for node  $v$ . We likewise use  $\Gamma_i(A) = \cup_{v \in A} \Gamma_i(v) \setminus (A \times A)$  and  $\Gamma_o(A) = \cup_{v \in A} \Gamma_o(v) \setminus (A \times A)$ , respectively, to represent the set of edges coming into and emerging from a set of vertices  $A \subseteq V$ . A *cut* is a subset of the vertex set  $V$ . For any cut  $C \subseteq V$  and

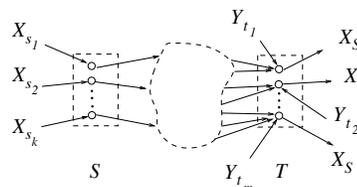


Fig. 1. A multicast network with side information at the sink

vertex  $v \in V$ ,

$$\mathcal{J}_C(v) \triangleq \begin{cases} 1 & \text{if } v \in C \\ 0 & \text{otherwise.} \end{cases}$$

We use  $S \subseteq V$  and  $T \subseteq V$  to denote the sets of source and sink nodes, respectively. Each source node  $s \in S$  has no input edges and one output edge; each sink node has no output edges. While there is no loss of generality in the first assumption, requiring sink nodes to be terminal nodes is a restrictive assumption when side information is available only at sink nodes. Each source node  $s \in S$  observes a source random process  $X_s \in \mathcal{X}_s$ , and each sink node  $t \in T$  observes a side-information random process  $Y_t \in \mathcal{Y}_t$ . For any index set  $I$ , we denote the ordered  $|I|$ -tuple  $(r_i : i \in I)$  by the short-hand notation  $r_I$ . Thus, for  $A \subseteq S$  and  $B \subseteq T$ , we use  $X_A = (X_s : s \in A)$  and  $Y_B = (Y_t : t \in B)$  to denote the vectors of source and side-information random variables respectively. Each intermediate node  $v \in V \setminus (S \cup T)$  has access only to the codewords received on the edges in  $\Gamma_i(v)$ .

The random process  $\{(X_S(i), Y_T(i))\}_{i=1}^{\infty}$  is drawn i.i.d. from known probability mass function  $P(\cdot)$ . We consider networks with multicast demands and receiver side information, as shown in Fig. 1. Thus each  $t \in T$  demands the complete collection of sources  $X_S$ ; the side information may differ from one sink to the next. Rate regions for multicast networks without side information appear in [1] for independent sources and [2] for dependent sources.

For any collection of rates  $R_E \geq 0$ , an  $(n, (2^{nR_e})_{e \in E})$

<sup>0</sup>This material is based upon work partially supported by NSF Grant No. CCF-0325324 and Caltech's Lee Center for Advanced Networking.

network code  $(f_E^{(n)}, g_T^{(n)})$  defines encoders

$$\begin{aligned} f_{(v,v')}^{(n)} &: \mathcal{X}_v^n \rightarrow \{1, \dots, 2^{nR_{(v,v')}}\} \quad \forall v \in S, (v, v') \in E \\ f_{(v,v')}^{(n)} &: \prod_{e \in \Gamma_i(v)} \{1, \dots, 2^{nR_e}\} \rightarrow \{1, \dots, 2^{nR_{(v,v')}}\} \\ &\quad \forall v \notin S, (v, v') \in E \end{aligned}$$

and decoders

$$g_t^{(n)}: \prod_{e \in \Gamma_i(t)} \{1, \dots, 2^{nR_e}\} \times \mathcal{Y}_t^n \rightarrow \prod_{s \in S} \mathcal{X}_s \quad \forall t \in T.$$

During transmission, the above maps are appropriately sequenced to ensure that at each node, the maps corresponding to the incoming edges are applied (and their outputs received) prior to applying the maps corresponding to the outgoing edges. We say that  $\{(f_E^{(n)}, g_T^{(n)})\}_{n=1}^\infty$  is a *valid* sequence of codes if the probability of an error at the receivers vanishes as  $n$  increases without bound; more precisely, if the random variable  $F_e^{(n)}$  denotes the codewords observed on the edge  $e$  corresponding to the input  $X_S^n$ , then

$$\lim_{n \rightarrow \infty} \Pr(g_t^{(n)}(F_{\Gamma_i(t)}^{(n)}, Y_t^n) \neq X_S^n) \rightarrow 0$$

for each  $t \in T$ . The set of *achievable* rate vectors  $\mathcal{R}$  for the network is all rates  $R_E$  for which valid sequences of  $(n, (2^{nR_e})_{e \in E})$  codes exist.

### III. MULTICAST WITH SIDE INFORMATION AT THE SINKS

The proofs of previous multicast results without side information use random binning for code design on independent sources [1] and random linear coding for code design on (both independent and) dependent sources [2]. Random linear code design is a form of random binning that adds additional structure to the random bin choices. This structure is useful in practice. Theorem 1 generalizes the proof of [2] to allow decoder side information.

To make the discussion precise, a  $(n, (2^{nR_e})_{e \in E})$  *linear code*  $(f_E^{(n)}, g_T^{(n)})$  is a set of mappings such that

$$f_{(v,v')}^{(n)} = \begin{cases} b_v^{(n)}(X_v^n) & v \in S, (v, v') \in E \\ \sum_{e \in \Gamma_i(v)} a_e f_e^{(n)} & v \notin S, (v, v') \in E \end{cases}$$

Here,  $b_s^{(n)}: \mathcal{X}_s^n \rightarrow \mathbb{F}_2^{nR_s}$  is an arbitrary (possibly non-linear) function and  $a_e \in \mathbb{F}_2^{nR_{in,e} \times nR_e}$  defines a linear map with  $R_{in,(v,v')} = \sum_{e' \in \Gamma_i(v)} R_{e'}$ . The decoder mappings  $(g_t^{(n)}: t \in T)$ 's are suitably chosen (often non-linear) functions. An  $(n, (2^{nR_e})_{e \in E})$  linear code  $(f_E^{(n)}, g_T^{(n)})$  is a *random linear code* if the coefficients  $\{a_e\}$  and mappings  $\{b_v^{(n)}\}$  are chosen independently and uniformly at random. Let  $\mathcal{R}$  and  $\mathcal{R}_L$  denote the set of rate vectors that are achievable through arbitrary codes and linear codes, respectively, on network  $\mathcal{N}$ . Let  $\mathcal{R}_C$  be the set of rate vectors that satisfy the cut-set bounds on  $\mathcal{N}$  [3], [2], i.e.,  $R_E \in \mathcal{R}_C$  if and only if for any  $C \subseteq V$ ,

$$\sum_{e \in \Gamma_o(C)} R_e \geq H(X_{C \cap S} | X_{S \setminus C}, Y_{T \setminus C}). \quad (1)$$

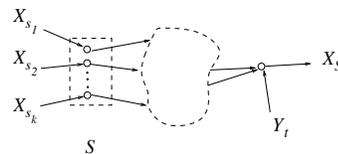


Fig. 2. The network  $\mathcal{N}_t$

Theorem 1 characterizes the rate region for multicast networks in the presence of side information at the sinks and also shows the sufficiency of linear codes for achieving this region.

*Theorem 1:*  $\mathcal{R} = \mathcal{R}_L = \mathcal{R}_C$ .

*Proof:* All rates achievable through random linear coding lie in  $\mathcal{R}_C$ . Thus,  $\mathcal{R}_L \subseteq \mathcal{R} \subseteq \mathcal{R}_C$  (cf. [3]). In the following, we show that  $\mathcal{R}_C \subseteq \mathcal{R}_L$ . Define  $\mathcal{N}_t$  to be the network obtained from  $\mathcal{N}$  by deleting all the sink nodes except the node  $t$ . The resulting network has the edge set  $E(t) \triangleq E \setminus \Gamma_i(T \setminus \{t\})$ , sources  $X_S$ , and side information  $Y_t$  at the only sink  $t$ . (See Fig. 2.) Since the side information is available at the sink, network  $\mathcal{N}_t$  is equivalent to a (multi-source) multicast problem with sources  $(X_S, Y_t)$ . Thus, random linear codes achieve any rate  $R_{E(t)}$  that satisfies

$$\sum_{e \in \Gamma_o(C)} R_e \geq H(X_{C \cap S} | X_{S \setminus C}, Y_t) \quad (2)$$

for all  $C \subseteq V$  [2, Theorem 6]. Denote by  $\mathcal{R}_{C,t}$  the set of rate vectors  $R_E$  such that  $R_{E(t)}$  satisfies (2).

Now, consider any  $R_E \in \cap_{t \in T} \mathcal{R}_{C,t}$ , and let  $\{(f_E^{(n)}, g_T^{(n)})\}$  be a code obtained by assigning random coefficients to a linear code at rate  $R_E$ . For this code,

$$\begin{aligned} &\Pr(g_t^{(n)}(F_{\Gamma_i(t)}^{(n)}, Y_t) \neq X_S \text{ for some } t \in T) \\ &\leq \sum_{t \in T} \Pr(g_t^{(n)}(F_{\Gamma_i(t)}^{(n)}, Y_t) \neq X_S) < |T|\epsilon \end{aligned}$$

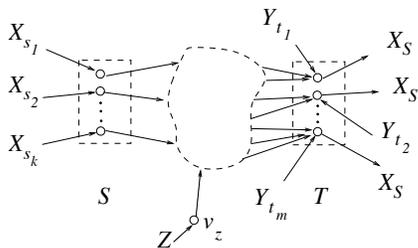
for sufficiently large  $n$  by the union bound since rate  $R_{E(t)}$  is achievable for each  $\mathcal{N}_t$  and the random linear encoding operation depends only on the input rates and output rates at each node. Therefore,  $R_E$  is achievable for the network  $\mathcal{N}$ .

Next, we show that  $\mathcal{R}_C = \cap_{t \in T} \mathcal{R}_{C,t}$ . Since all  $R_E \in \cap_{t \in T} \mathcal{R}_{C,t}$  are achievable,  $\cap_{t \in T} \mathcal{R}_{C,t} \subseteq \mathcal{R}_C$ . To see the reverse inclusion, let  $R_E \in \mathcal{R}_C$ . For any  $C \subseteq V$ , for which  $T \not\subseteq C$  let  $C_t = C \cup (T \setminus \{t\})$ , and let  $\tilde{C}_t = C \setminus (T \setminus \{t\})$ . Then,  $\Gamma_o(C_t) = \Gamma_o(\tilde{C}_t)$  and the cut-set inequality corresponding to the cut  $C_t$  in the network  $\mathcal{N}$  is

$$\sum_{e \in \Gamma_o(C_t)} R_e \geq H(X_{S \cap C_t} | X_{C_t^c}, Y_{T \setminus C_t}), \quad (3)$$

which is the same as the cut-set inequality corresponding to the cut  $\tilde{C}_t$  in  $\mathcal{N}_t$ . ■

An  $(n, (2^{nR_e})_{e \in E})$  code  $(f_E^{(n)}, g_T^{(n)})$  is generated by *random binning* if for each  $v \in S$ ,  $(v, v') \in E$ , and  $x^n \in \mathcal{X}_v^n$ ,  $f_{(v,v')}^{(n)}(x^n)$  is chosen uniformly at random from alphabet  $\{1, \dots, 2^{nR_{(v,v')}}\}$ ; and for each  $v \notin S$ ,  $(v, v') \in E$  and each  $i \in \prod_{e \in \Gamma_i(v)} \{1, 2, \dots, 2^{nR_e}\}$ ,  $f_{(v,v')}^{(n)}(i)$  is chosen uniformly

Fig. 3. Side information at a non-sink node  $v_z$ 

at random from alphabet  $\{1, \dots, 2^{nR(v,v')}\}$ . We use  $\mathcal{R}_B$  to denote the set of rate vectors that are achievable for a given network using random binning.

While random linear coding is a low-complexity method for achieving random binning, the codewords corresponding to different inputs are not necessarily independent of each other. Thus, Theorem 1 does not imply  $\mathcal{R} = \mathcal{R}_B = \mathcal{R}_C$ . Theorem 2 generalizes [1] first to dependent sources and then to allow side information, thereby giving an alternative proof that  $\mathcal{R} = \mathcal{R}_C$ . The proof of Theorem 2, which appears in the Appendix, is used in the proof of Theorem 3 in the next section.

*Theorem 2:*  $\mathcal{R} = \mathcal{R}_B = \mathcal{R}_C$ .

#### IV. MULTICAST WITH SIDE INFORMATION AT A NON-SINK NODE

Now suppose that side information random variable  $Z$  is present at node  $v_z \notin T$ . (See Fig. 3.) We again denote the sources by  $X_S$  and assume that each sink node  $t \in T$  has access to side information  $Y_t$  and wishes to reconstruct  $X_S$ . To obtain an achievability bound we first encode  $Z$  into a separate codeword for each subset of the sinks and then sequentially multicast each codeword to its corresponding sinks using the earlier codewords as side information. We then multicast  $X_S$  to all of the receivers using the codewords received by each sink  $t \in T$  as the side information for that sink. We calculate the rate required for each multicast using Theorem 1.

Let  $\mathcal{J} \triangleq \{\tau : \tau \subseteq T\}$  be the subsets of  $T$ . For any permutation  $\sigma$  of  $\mathcal{J}$ , any  $t \in T$ , and any  $\tau \in \mathcal{J}$  for which  $t \in \tau$  let  $\mathcal{J}(\sigma, t, \tau) \triangleq \{\tau' \subseteq T : t \in \tau', \sigma^{-1}(\tau') < \sigma^{-1}(\tau)\}$  be the indices of those codewords received by node  $t$  before the codeword for  $\tau$ . Finally, for any  $A \subseteq V$ , let  $\mathcal{J}(\sigma, A, \tau) \triangleq \cup_{t \in A} \mathcal{J}(\sigma, t, \tau)$ .

*Theorem 3:* Let  $U_{\mathcal{J}} = \{U_{\tau} : \tau \in \mathcal{J}\}$  be a set of random variables satisfying the Markov chains  $U_{\tau} \rightarrow Z \rightarrow (X_S, Y_T)$ . Then rate vector  $R_E$  is achievable if there exists a permutation  $\sigma$  of  $\mathcal{J}$  for which

$$\begin{aligned} & \sum_{e \in \Gamma_o(C)} R_e \\ & \geq \mathcal{J}_C(v_z) \sum_{\tau: \tau \not\subseteq C} [H(U_{\tau} | Y_{T \setminus C}, U_{\mathcal{J}(\sigma, T \setminus C, \tau)}) - H(U_{\tau} | Z)] \\ & \quad + H(X_{S \cap C} | X_{S \setminus C}, (U_{\tau} : \tau \not\subseteq C), Y_{T \setminus C}) \end{aligned} \quad (4)$$

for all  $C \subseteq V$ .

*Proof:* Again, the outline of our strategy is as follows. First, we encode the side information source  $Z$  into codewords  $U_{\mathcal{J}}$ , where for each  $\tau \in \mathcal{J}$ ,  $U_{\tau}$  is the codeword for subset  $\tau$  of the sinks. Next, we transmit each  $U_{\tau}$  to the sinks in  $\tau$  in the order described by permutation  $\sigma$ . We treat each transmission as a single-source multicast with side information  $U_{\mathcal{J}(\sigma, t, \tau)}$  at each receiver  $t \in \tau$ . Finally, we describe sources  $X_S$  to the receivers; by this time, the codewords available at receiver  $t$  are  $(U_{\tau} : t \in \tau)$ . We apply Theorem 1 to calculate the rate required for each such multicast.

The auxiliary random variable  $U_{\tau}$  captures information present in  $Z$  that is useful to all  $t \in \tau$ . We follow the approach of previously solved coded side-information problems (cf. [4], [5], [6]) in designing the codeword using each  $U_{\tau}$ .

For each  $\tau \subseteq T$ , we choose a random variable  $U_{\tau}$  that satisfies the given Markov chain condition. Next, we generate  $2^{n\tilde{R}_{\tau}}$  length- $n$  codewords  $U^{(1, \tau)}, \dots, U^{(2^{n\tilde{R}_{\tau}}, \tau)}$  such that each  $U_1^{(i, \tau)}, \dots, U_n^{(i, \tau)}$  is drawn i.i.d. according to the marginal of  $U_{\tau}$ . Define the encoder mapping  $\alpha_{\tau} : \mathcal{Z}^n \rightarrow \{1, \dots, 2^{n\tilde{R}_{\tau}}\}$ , where  $\alpha_{\tau}(z^n)$  is an index  $i$  for which  $(z^n, U^{(i, \tau)}) \in A_{\epsilon}^{(n)}(Z, U_{\tau})$ . Following the proof in [4], the existence of such an index occurs with probability approaching 1 provided  $\tilde{R}_{\tau} \geq I(Z; U_{\tau})$ . Let  $I_{\tau} = \alpha_{\tau}(Z^n)$ .

To transmit the indices  $I_{\mathcal{J}}$  to the respective sinks, we fix a permutation  $\sigma$  of  $\mathcal{J}$  and then for each  $i \in \{1, \dots, 2^{|\mathcal{J}|} - 1\}$  we multicast  $I_{\sigma(i)}$  with rate allocation  $R_E^{\sigma(i)}$  to the vertices in  $\sigma(i)$ . Note that the indices  $I_{\mathcal{J}(\sigma, t, \tau)}$  are available at the sink  $t$  earlier than the index  $I_{\tau}$ .

Thus, in order to achieve asymptotically vanishing error probability for decoding  $I_{\tau}$  at the sink  $t \in \tau$ , it suffices to perform random binning at each intermediate node while ensuring that the rate  $R_E^{\tau}$  satisfies

$$\sum_{e \in \Gamma_o(C)} R_e^{\tau} \geq \mathcal{J}_C(v_z) [I(U_{\tau}; Z) - I(U_{\tau}; Y_t, U_{\mathcal{J}(\sigma, t, \tau)})]$$

for each  $C \subseteq V$ . The above can be seen by combining the binning argument of [5] (see also [6]) and the proof from Theorem 2.

Replacing the right side of the above bound by its maximal value across all sinks  $t \in C^c$ , we find that any  $R_E^{\tau}$  satisfying

$$\sum_{e \in \Gamma_o(C)} R_e^{\tau} \geq \mathcal{J}_C(v_z) [I(U_{\tau}; Z) - \min_{t \in T \setminus C} I(U_{\tau}; Y_t, U_{\mathcal{J}(\sigma, t, \tau)})]$$

for each  $C \subseteq V$  is sufficient to make the error probability vanish asymptotically. Subtracting  $H(U_{\tau})$  from each of the above terms gives

$$\begin{aligned} \sum_{e \in \Gamma_o(C)} R_e^{\tau} & \geq \mathcal{J}_C(v_z) \left[ -H(U_{\tau} | Z) \right. \\ & \quad \left. + \max_{t \in T \setminus C} H(U_{\tau} | Y_t, U_{\mathcal{J}(\sigma, t, \tau)}) \right] \end{aligned}$$

Further simplification using an argument similar to the one that lets us obtain the region in (3) from the one given in (2)

in Theorem 1 equals the set of rate allocations  $R_E^\tau$  that satisfy

$$\sum_{e \in \Gamma_o(C)} R_e^\tau \geq \mathcal{J}_C(v_Z) [H(U_\tau | Y_{T \setminus C}, U_{\sigma, T \setminus C, \tau}) - H(U_\tau | Z)].$$

Adding the rates over all  $\tau \subseteq T$ , and finally, adding the rate required to multicast  $X_S$  with  $Y_t$  and  $(U_\tau : t \in \tau)$  present as side information at the sink node  $t$ , we obtain the achievability result given in (4).

## V. AN INNER BOUND ON THE RATE REGION WITH GENERAL DEMAND STRUCTURES

In this section, we use the result of the previous section to find an inner bound on the rate region for general demands. We again denote the sources by  $X_S$  and the demands by  $Y_T$ , where for each  $t \in T$ ,  $Y_t = X_{S(t)}$  for some  $S(t) \subseteq S$ . We obtain an achievable region by satisfying the demands  $Y_T$  through sequence of multicast sessions. For each multicast transmission, the demands met in previous multicast sessions are treated as side information for the current session. For  $s \in S$ , let  $T_s = \{t : s \in S(t)\}$ . Let  $\mathcal{S}_S = \{\sigma : \sigma \text{ is a permutation of } S\}$ . For  $\sigma \in \mathcal{S}_S$ , let  $\mathcal{R}_\sigma$  denote the set of rate vectors  $R_E$  satisfying the following inequality for all  $C \subseteq V$ :

$$\sum_{e \in \Gamma_o(C)} R_e \geq \sum_{k=1}^{|S|} \max_{t \in T \setminus C} H(X_{\{\sigma(k)\} \cap S(t)} | X_{\{\sigma(1), \dots, \sigma(k-1)\} \cap S(t)}), \quad (5)$$

and let  $\mathcal{R}_S$  denote the convex hull of  $\cup_{\sigma \in \mathcal{S}_S} \mathcal{R}_\sigma$ . The following theorem asserts the achievability of  $\mathcal{R}_S$ .

*Theorem 4:* Let  $\mathcal{R}$  denote the set of achievable rates for a network  $\mathcal{N}$ . Then,

$$\mathcal{R} \supseteq \mathcal{R}_S.$$

*Proof:* For any ordering  $\sigma \in \mathcal{S}_S$  of  $S$ , consider the sequence of multicast sessions such that the  $k$ -th session has multicast demands  $X_{\sigma(k)}$  at the sink nodes  $t$  for which  $\sigma(k) \in S(t)$ . By Theorem 1, the rate vector  $R_E^{(\sigma, k)}$  is sufficient to meet the demands for the  $k$ -th multicast session, if the following condition is satisfied:

$$\sum_{e \in \Gamma_o(C)} R_e^{(\sigma, k)} \geq \max_{t \in T \setminus C} H(X_{\{\sigma(k)\} \cap S(t)} | X_{\{\sigma(1), \dots, \sigma(k-1)\} \cap S(t)}). \quad (6)$$

Adding the rates required for each of the multicast sessions gives the achievability of the region  $\mathcal{R}_\sigma$  given in (5) for each  $\sigma$ . By the convexity of the rate region,  $\mathcal{R}_S$  is achievable. ■

While the closed-form expression for the above rate-region may be difficult to analyze, it is easily computable algorithmically. It should be noted, however, that the above rate region is not tight in general. In the following example,  $\mathcal{R}_S$  has no tight rate points.

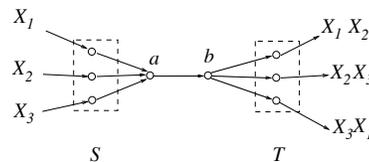


Fig. 4. Example 1

■ *Example 1:* Consider the network shown in Fig 4. Let  $X_1$  take values uniformly in  $\{0, 1, 2\}$ ,  $p_{X_2|X_1}(y|x) = 1/2$  for  $y \in \{x, (x+1) \pmod{3}\}$  and  $X_3 = (1 - X_2 - X_1) \pmod{3}$ . Then, for all possible  $\sigma \in \mathcal{S}_{\{1,2,3\}}$ , any  $R_E \in \mathcal{R}_\sigma$  satisfies:

$$\begin{aligned} R_{(a,b)} &\geq \log_2 3 + 1 + 1 \\ &= \log_2 3 + 2. \end{aligned}$$

On the other hand, for any achievable rate, the vector  $(X_1, X_2, X_3)$  is decodable at the node  $a$ . Thus, it suffices to ensure a rate of  $H(X_1, X_2, X_3) = \log_2 3 + 1$  on the link  $(a, b)$ . This proves that the rate point any rate point  $R_E \in \mathcal{R}_S$  is strictly suboptimal.

## VI. CONCLUSIONS

We generalize earlier multicast rate region bounds to allow side information at the decoders. We also generalize to networks with side information at one intermediate node in addition to the side information at the sinks. The generalization takes an approach similar to that used in the coded side information problem. The given bounds are interesting both on their own and for their applicability in proving other interesting bounds. For example, we can bound the rate region for a network with multiple multicasts by considering each multicast in turn and treating information received from “earlier” multicasts as side information available to the corresponding sinks.

## APPENDIX

*Proof of Theorem 2:* Since all rates in  $\mathcal{R}_B$  are achievable,  $\mathcal{R}_B \subseteq \mathcal{R}_C$ .

To prove  $\mathcal{R}_B \supseteq \mathcal{R}_C$ , we first show that  $\mathcal{R}_B$  is a convex set. Let  $R_{1,E}, R_{2,E} \in \mathcal{R}_B$ . Let  $\{f_{1,E}^{(n)}, g_{1,T}^{(n)}\}$  and  $\{f_{2,E}^{(n)}, g_{2,T}^{(n)}\}$  be two valid sequences of codes, henceforth called the component codes, that achieve the rates  $R_{1,E}$  and  $R_{2,E}$  respectively. Define a  $(n, (2^{\lceil n\lambda \rceil R_{1,e} + \lceil n(1-\lambda) \rceil R_{2,e}})_{e \in E})$  code  $(\tilde{f}_E^{(n)}, \tilde{g}_T^{(n)})$  by time-sharing  $f_{1,E}$  and  $f_{2,E}$  and appropriately defining the decoder functions  $\tilde{g}_T^{(n)}$  as follows:

$$\begin{aligned} \tilde{f}_e^{(n)} &= f_{1,e}^{\lceil \lambda n \rceil} * f_{2,e}^{\lceil (1-\lambda)n \rceil} \quad \forall e \in E \\ \tilde{g}_t^{(n)} &= g_{1,t}^{\lceil \lambda n \rceil} * g_{2,t}^{\lceil (1-\lambda)n \rceil} \quad \forall t \in T. \end{aligned}$$

For the composite code thus formed,

$$\begin{aligned} & \Pr(\tilde{g}_t^{(n)}(\tilde{F}_{\Gamma_i(t)}^{(n)}, Y_t^n) \neq X_S^n) \\ & \leq \Pr(g_{1,t}^{(\lceil \lambda n \rceil)}(F_{1,\Gamma_i(t)}^{(\lceil \lambda n \rceil)}, Y_t^{\lceil \lambda n \rceil}) \neq X_S^{\lceil \lambda n \rceil}) \\ & \quad + \Pr(g_{2,t}^{(\lfloor (1-\lambda)n \rfloor)}(F_{2,\Gamma_i(t)}^{(\lfloor (1-\lambda)n \rfloor)}, (Y_t^n)_{\lceil \lambda n \rceil+1}^n) \\ & \quad \neq (X_S^n)_{\lceil \lambda n \rceil+1}^n) \end{aligned}$$

Thus,  $\{(\tilde{f}_E^{(n)}, \tilde{g}_T^{(n)})\}$  is a valid sequence of codes. Since the component encoders are chosen independently and each of them is uniformly random mapping, it follows that the composite encoders are also uniform and independent, and hence, have the same distribution as a code that would have been formed by random binning. Further, the rate vector corresponding to the code thus constructed approaches  $\lambda R_{1,E} + (1-\lambda)R_{2,E}$  asymptotically. Thus,  $\lambda R_{1,E} + (1-\lambda)R_{2,E} \in \mathcal{R}_B$ . Therefore,  $\mathcal{R}_B$  is a convex set.

Now, let  $R_E$  be a boundary point of  $\mathcal{R}_C$ , that is  $R_E$  is a rate vector such that if  $\hat{R}_E \in \mathcal{R}_C$  and  $\hat{R}_e \leq R_e$  for some  $e \in E$ , then there exists an  $e' \in E$  for which  $\hat{R}_{e'} > R_{e'}$ . We claim that  $\sum_{e \in \Gamma_i(S)} R_e = H(X_S)$ . To see this, let us assume otherwise. By the achievability of  $\mathcal{R}_C$  proved in [2], there exists a sequence of valid  $(n, (2^{nR_e})_{e \in E})$  codes, say  $\{(f_E^{(n)}, g_T^{(n)})\}$ . Now, perform random binning on each  $X_S^r$  to obtain the network  $\mathcal{N}'$  which has the same set of vertices and edges as  $\mathcal{N}$ , while the  $r$ -dimensional sources  $X_S^r$  are replaced by sources  $B_S$ , where,  $B_s = (b_s(X_S^r) : s \in S)$ , where  $b_s : X_S^r \rightarrow \mathcal{X}_s^r$  is a random binning operation at rate  $\hat{R}_s^r$  satisfying  $x^r \in b_s^{-1}(\{x^r\}) \forall x^r \in b_s(\mathcal{X}_s^r)$ . The last condition can be ensured by appropriately relabeling the value of  $b_s(\cdot)$  in each bin. Since the code sequence  $\{(f_E^{(n)}, g_T^{(n)})\}$  is valid for a multicast with  $X_S$  as sources, it is valid for multicast with sources  $B_S$  too. Thus, sequence of codes  $\{(\tilde{f}_E^{(n)}, \tilde{g}_T^{(n)})\}$  is a valid code for the network, where the encoding functions  $\tilde{f}_E^{(n)}$  are constructed as

$$\tilde{f}_e^{(n)} = \begin{cases} b_s \circ f_e^{(n)} & e = e_s \text{ for some } s \in S \\ f_e^{(n)} & \text{otherwise} \end{cases}$$

and the decoding functions  $\tilde{g}_T$  are suitably defined. By Slepian-Wolf theorem ([7]),  $B_S$  is sufficient to reconstruct  $X_S^r$  with error probability vanishing asymptotically with  $r$ , as long as  $\sum_{s \in S} \hat{R}_s^r \geq rH(X_S)$ . This shows that there is an achievable rate  $\tilde{R}_E$  s.t.  $\tilde{R}_e \leq R_e \forall e \in E$  and  $\sum_{s \in S} \tilde{R}_{e_s} = H(X_S)$ , which contradicts the assumption that  $\mathbf{R}$  is a tight rate. Thus,  $\sum_{s \in S} R_{e_s} = H(X_S : s \in S)$ .

Define the network  $\mathcal{N}'_r$  with the same set of vertices and edges as  $\mathcal{N}$  and the sources  $X_S$  replaced by  $B_S$ , where each  $b_s$  is a random binning operation at rate  $R_s + \epsilon$ . Let  $\tilde{\mathcal{N}}_r$  be the network obtained by replacing  $B_S$  in  $\mathcal{N}'_r$  by  $\tilde{B}_S$ , where the sources  $\tilde{B}_S$  are independent of each other, but have the same first-order marginal distribution as  $B_S$ . By the proof used in the achievability result in [1] for the case of multicast with independent sources, the error probability for random binning codes on  $\tilde{\mathcal{N}}_r$  approaches zero asymptotically. Further, since  $R_E$  is a tight point,  $\sum_{s \in S} H(\tilde{B}_s) = \sum_{s \in S} H(B_s) <$

$H(B_S) + r|S|\epsilon$ . Since  $\epsilon$  is arbitrary, it follows that by using the same code on each link as  $\tilde{\mathcal{N}}_r$ , we see that random binning achieves the rate vector  $rR_E$  in  $\mathcal{N}'_r$ , and hence,  $R_E \in \mathcal{N}$ . To establish that the error probability for a code formed by random binning approaches 0, consider any sequence of codes  $\{(\tilde{f}_E^{(n)}, \tilde{g}_T^{(n)})\}$  that is valid for the network  $\tilde{\mathcal{N}}_r$ . By using the same codes on the network  $\hat{\mathcal{N}}_r$ , the error probability satisfies the following:

$$\begin{aligned} & \Pr(\tilde{g}_t^{(n)}(\tilde{F}_{\Gamma_i(t)}^{(n)}) \neq B_S) \\ & = \sum_{\substack{y^n \in (\prod_{s \in S} b_s(\mathcal{X}_s^r))^n \\ y^n : \tilde{g}_t^{(n)}(\tilde{f}_e : e \in \Gamma_i(t)) \neq y^n}} P_{B_S}(y^n) \\ & \leq \sum_{\substack{y^n \in (\prod_{s \in S} b_s(\mathcal{X}_s^r))^n \\ y^n : \tilde{g}_t^{(n)}(\tilde{f}_e : e \in \Gamma_i(t)) \neq y^n}} P_{\tilde{B}_S}(y^n) \\ & \quad + d_V(P_{B_S}, P_{\tilde{B}_S}), \end{aligned}$$

where  $d_V(p, q)$  denote the variational distance between the distributions  $p$  and  $q$ . The sequence of inequalities is furthered by the use of Pinsker's inequality as follows:

$$\begin{aligned} & \Pr(\tilde{g}_t^{(n)}(\tilde{F}_{\Gamma_i(t)}^{(n)}) \neq (B_S)^n) \\ & \leq \sum_{\substack{y^n \in (\prod_{s \in S} b_s(\mathcal{X}_s^r))^n \\ y^n : \tilde{g}_t^{(n)}(\tilde{f}_e : e \in \Gamma_i(t)) \neq y^n}} P_{\tilde{B}_S}(y^n) \\ & \quad + \sqrt{2nD(P_{B_S} || P_{\tilde{B}_S})} \\ & \leq \sum_{\substack{y^n \in \prod_{s \in S} B_s(\mathcal{X}_s^r)^n \\ y^n : \tilde{g}_t^{(n)}(\tilde{f}_e : e \in \Gamma_i(t)) \neq y^n}} P_{\tilde{B}_S}(y^n) + \sqrt{2n\epsilon} \end{aligned}$$

Since the choice of  $\epsilon$  is independent of  $n$ , we can make the second term vanish by choosing  $\epsilon = 1/n^2$ . The first term vanishes because the code  $(\tilde{f}_E, \tilde{g}_T)$  is a valid code for the network  $\tilde{\mathcal{N}}_r$ . Thus,  $R_E \in \mathcal{R}_B$ . ■

## REFERENCES

- [1] R. Ahlswede, N. Cai, S. Y. R. Li, and R. Yeung. Network information flow. *IEEE Transactions on Information Theory*, IT-46(4):1204–1216, July 2000.
- [2] T. Ho, R. Koetter, M. Médard, M. Effros, J. Shi, and D. Karger. A random linear network coding approach to multicast. *IEEE Transactions on Information Theory*, 52(10):4413–4430, October 2006.
- [3] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 1991.
- [4] R. Ahlswede and J. Körner. Source coding with side information and a converse for degraded broadcast channels. *IEEE Transactions on Information Theory*, IT-21(6):629–637, November 1975.
- [5] A. D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Transactions on Information Theory*, IT-22(1):1–10, January 1976.
- [6] T. Berger, K.B. Housewright, J.K. Omura, S. Ting, and J. Wolfowitz. An upper bound on the rate distortion function for source coding with partial side information at the decoder. *IEEE Transactions on Information Theory*, IT-25(6):664–666, November 1979.
- [7] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, IT-19:471–480, 1973.