# An Outer Bound for Multisource Multisink Network Coding With Minimum Cost Consideration

Xijin Yan, *Student Member, IEEE*, Jun Yang, *Member, IEEE*, and Zhen Zhang, *Fellow, IEEE*

***Abstract*—The max-flow min-cut bound is a fundamental result in the theory of communication networks, which characterizes the optimal throughput for a point-to-point communication network. The recent work of Ahlswede *et al.* extended it to single-source multisink multicast networks and Li *et al.* proved that this bound can be achieved by linear codes. Following this line, Erez and Feder as well as Ngai and Yeung proved that the max-flow min-cut bound remains tight in single-source two-sink nonmulticast networks. But the max-flow min-cut bound is in general quite loose (see Yeung, 2002). On the other hand, the admissible rate region of communication networks has been studied by Yeung and Zhang as well as Song and Yeung, but the bounds obtained by these authors are not explicit. In this work, we prove a new explicit outer bound for arbitrary multisource multisink networks and demonstrate its relation with the minimum cost network coding problem. We also determine the capacity region for a special class of three-layer networks.**

***Index Terms*—$K$-pairs transmission, max-flow min-cut bound, multisource multisink network, network coding, network sharing bound, side information, three-layer network.**

## I. Introduction

In recent few years, network coding has been a major focus of research in network information theory due to its potential applications in communication networks. Along with the fundamental theories that have been developed so far, more and more applications are emerging in practice. The intense studies in network coding theory have deeply influenced research areas such as information theory, coding theory, wireless communications, networking, and so on.

Among all these research topics, "What is the capacity region of a communication network?" and "How can we achieve the optimal throughput of a communication network?" are two fundamental problems. The earliest research dated back to the work by Yeung in [8], where he characterized the admissible rate region for a particular configuration of a diversity coding system subject to certain distortion criteria. A principle of superposition was proved on the two-level diversity coding system when the source consists of two independent data streams. Later on, Yeung and Zhang [9] generalized this result to an arbitrary number of levels in symmetrical multilevel diversity coding with independent data streams and showed that optimality can be achieved by superposition in general. Inspired by mobile satellite communication systems, a more general model than multilevel diversity coding, named distributed source coding system (Fig. 1), was developed by Yeung and Zhang in [6]. The inner and outer bounds of the admissible rate region were obtained in terms of $\Gamma_n^*$ and $\bar{\Gamma}_n^*$, respectively, which are fundamental regions of the entropy function [10]–[13].

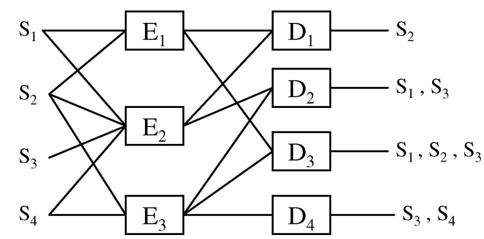


Fig. 1. An example of distributed source coding system.

Based on these works, Ahlswede, Cai, Li, and Yeung [1] studied the information flow in an arbitrary network with a single source whose data is multicast to a collection of destinations called sinks. They showed that the capacity of the network can be characterized by the max-flow min-cut bound, a generalization of the famous max-flow min-cut theorem for single-source single-sink network information flow. Following this line of research, Li, Yeung, and Cai [2] presented an explicit construction of linear codes for multicasting in a single-source network and proved that linear coding suffices to achieve the optimum. Subsequently, Koetter and Médard [14] developed an algebraic approach to the capacity region of network coding and found necessary and sufficient conditions for the feasibility of any given set of connections over a given network when restricted to linear coding. These results resolve the capacity problem for single-source multicast networks.

In the multisource network coding model considered in this paper, several mutually independent information sources are generated at possibly different nodes, and the data of each information source is multicast to a specific set of nodes. We call this model the multisource multisink network. If the data from all sources are sent to the same set of sinks, it is called the multisource multisink multicast case and has been treated as a trivial single-source multicast problem in [1]; otherwise, it is a multisource multisink network with different multicast requirement for each source, which we call the multisource multisink nonmulticast case. In this paper, we sometimes simply call

Manuscript received March 13, 2005; revised December 6, 2005. This work was supported in part by the National Science Foundation under Grant CCR 0326628. The material in this paper was presented in part at the First Workshop on Network Coding, Theory, and Applications, Riva del Garda, Italy, April 2005.

X. Yan and Z. Zhang are with the Communication Sciences Institute, Department of Electrical Engineering–Systems, University of Southern California, Los Angeles, CA 90089-2565 USA (e-mail: xyan@usc,edu; zzhang@commsci1.usc.edu).

J. Yang is with Wireless Communication and Networking Department, Philips Research, Briarcliff Manor, NY 10510 USA (e-mail: junyang@philips.com).

Communicated by N. Cai, Guest Editor.

Digital Object Identifier 10.1109/TIT.2006.874432

the second case the multisource case for convenience. Characterizing the capacity region for this case turns out not to be a simple extension of single-source multicast network coding. In the multisource case, the max-flow min-cut bounds cannot fully characterize the capacity region, they give only an outer bound. The best characterizations of the capacity region for multisource network coding which have been obtained so far in [7] make use of the regions $\Gamma_n^*$ and $\bar{\Gamma}_n^*$ developed in [6], which cannot be evaluated explicitly. For this reason, a linear programming bound (LP bound) is proposed as an outer bound and is proved to be tight for a class of special cases in multisource network coding. Although the LP bound can be evaluated, its evaluation is involved. Therefore, the search for more explicit bounds in the multisource multisink nonmulticast case has been a challenging task along this line of research.

As a first step toward this direction, Erez and Feder [3] as well as Ngai and Yeung [4] both proved that the max-flow min-cut bound remains tight in single-source two-sink nonmulticast networks. However, it can be quite loose in the more general multisource multisink networks [5]. In [15], Yan, Yang, and Zhang proved an improved outer bound for the three-layer network derived from the distributed source coding system. They showed, in certain cases, that their bound can bring significant improvement over the max-flow min-cut bound.

In this paper, we derive an improved outer bound over the max-flow min-cut bound for the capacity region for arbitrary acyclic multisource multisink networks, which is much more explicit and easier to understand than the LP bound.

This new bound is originally found by analyzing the role of the so-called side information at the decoders of a three-layer network. A three-layer network is a special network in which all the nodes are lined up in a three-layer architecture. It is a distributed source coding system in the network format. An example of a three-layer network which is equivalent to the distributed source coding system in Fig. 1 is shown in Fig. 2. In a three-layer network, each directed edge is assumed to be error free and thus is called an error-free channel. All channels except the coding channels (e.g., $(1,1'),(2,2'),(3,3')$ in Fig. 2) are considered straight connections, therefore, there is no constraint on the capacities of these channels. However, the information flow on each coding channel is limited by its capacity. The source nodes transmit independent messages to the sink nodes under the channel capacity constraints to meet the sink demands.

Suppose that the information sources in a three-layer network are $X_1, X_2, \ldots, X_K$. If the channel output of a coding channel is a function of the source data in a subset $\{X_i : i \in \alpha\}$ and it is available for a sink which is required to decode the source data in a subset $\{X_j : j \in \gamma\}$, then the output of this channel is said to be *side information* for the decoder if $\alpha \cap \gamma = \phi$. The following example explains the role of side information at the decoder.

We consider the classical example of network coding which consists of two sources and two sinks. A modified version $\mathcal{G}$ is shown in Fig. 3 where channel $(c_3, r_3)$ is disconnected. Each edge in $\mathcal{G}$ has unit capacity. Messages $X$ and $Y$ are generated at source nodes $S_1$ and $S_2$, respectively. Each of $X$ and $Y$ is of one bit. In order to decode $Y$ at sink node $T_1$ and $X$ at sink node $T_2$,

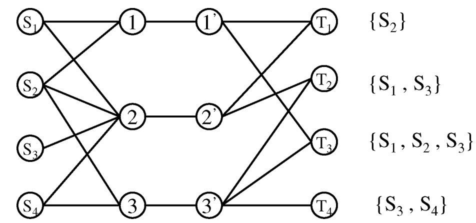


Fig. 2. An example of a three-layer network.

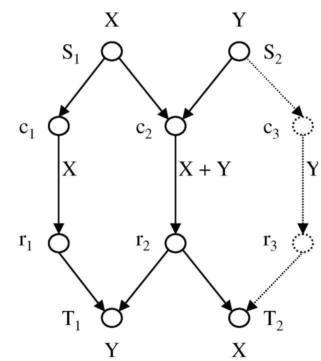


Fig. 3. An example of two-source two-sink network with partial side information.

$X+Y$ is sent on edge $(c_2, r_2)$ and $X$ is sent on edge $(c_1, r_1)$. We observe that $Y$ can be decoded at $T_1$ with the help of side information $X$, while $X$ cannot be decoded at $T_2$ because side information $Y$ is missing.

Now, we take a further look at the capacity region of this network. It is easy to see that the max-flow min-cut bound of the given network is

$$\begin{aligned} R_X &\le C(c_2, r_2) = 1 \\ R_Y &\le C(c_2, r_2) = 1 \\ R_X + R_Y &\le C(c_1, r_1) + C(c_2, r_2) = 2. \end{aligned}$$

However, the previous analysis indicates that $(R_X, R_Y) = (1,1)$ is not achievable in spite of satisfying the max-flow min-cut bound. The fact that sink node $T_2$ is missing side information suggests that a tighter outer bound may be obtained by analyzing the role of side information. Actually, we have

$$\begin{aligned} R_X &\le C(c_2, r_2) = 1 \\ R_Y &\le C(c_2, r_2) = 1 \\ R_X + R_Y &\le C(c_2, r_2) + \min\{C(c_1, r_1), C(c_3, r_3)\} \\ &= 1 + \min\{1, 0\} = 1. \end{aligned}$$

Obviously, this bound is tight. If we add channel $(c_3, r_3)$ to provide $Y$ at sink $T_2$ as in Fig. 4, then $X$ and $Y$ can be both decoded at $T_1$ and $T_2$, respectively. Hence, the availability of enough side information at decoders is essential for the decodability.

It needs to be pointed out that the analysis of the three-layer networks is essential since they serve as the building blocks for an arbitrary acyclic multisource multisink network when cuts are taken to provide constraints on the total amount of information in transmission. Therefore, it is an important tool to study general acyclic multisource multisink networks.

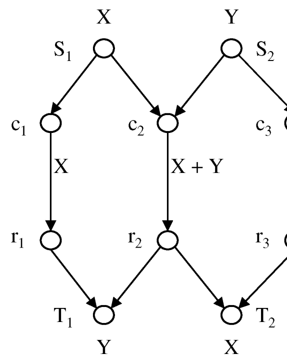


Fig. 4. An example of two-source two-sink network with full side information.

The rest of the paper is organized as follows: In Section II, we give a formal problem formulation and introduce the notions used throughout the paper. In Section III, we first formulate our new bound for a special type of networks known as a $K$-pairs transmission network (also called a multiple unicast network) which is the simplest case of a multisource network, then extend it to the general multisource network with arbitrary transmission requirements. The capacity region of a special class of networks, called degree-2 three-layer networks with $K$-pairs transmission requirements, is determined in the course of improving the bound. The corresponding proofs are given in Section IV. In Section V, we show by example that the new bound can offer significant improvement over the max-flow min-cut bound in some special cases but is still not tight in general. In Section VI, we discuss the implication of this new outer bound for the minimum-cost network coding problem. We define a region of the cost factors for which the minimum cost can be achieved by performing separate multicast coding for each source and show that some specific meaningful vectors of the cost factors belong to this region. Conclusions are drawn in Section VII.

## II. Network Model

### A. Some Notations for Directed Graphs

Let $\mathcal{G} = (\tilde{\mathcal{V}}, \mathcal{E})$ denote an acyclic directed graph, where $\tilde{\mathcal{V}}$ is a finite nonempty set whose elements are called nodes and $\mathcal{E}$ is a set of distinct ordered pairs of nodes called edges. We have

$$\mathcal{E} \subset \tilde{\mathcal{V}} \times \tilde{\mathcal{V}}.$$

A pair $(i,j) \in \mathcal{E}$ is an edge directed from node $i$ to node $j$. We assume that the graph has a special structure formulated as follows. Let

$$\tilde{\mathcal{V}} = \mathcal{S} \cup \mathcal{V} \cup \mathcal{T} = \{1, 2, \ldots, |\tilde{\mathcal{V}}|\} \tag{1}$$

where

$$\mathcal{V} = \{v_1, v_2, \ldots, v_{|\mathcal{V}|}\}$$

are the internal nodes,

$$\mathcal{S} = \{s_1, s_2, \ldots, s_{|\mathcal{S}|}\}$$

are the source nodes, and

$$\mathcal{T} = \{t_1, t_2, \ldots, t_{|\mathcal{T}|}\}$$

are the sink nodes. Let $\mathcal{A}_j$ be the set of nodes with an edge directed from node $j$ and $\mathcal{B}_j$ be the set of nodes with an edge directed into node $j$, i.e.,

$$\begin{aligned} \mathcal{A}_j &= \{i \in \tilde{\mathcal{V}}, (j,i) \in \mathcal{E}\} \\ \mathcal{B}_j &= \{i \in \tilde{\mathcal{V}}, (i,j) \in \mathcal{E}\}. \end{aligned} \tag{2}$$

We assume that for any source node $s_j$, $\mathcal{B}_{s_j} = \phi$, and for any sink node $t_j$, $\mathcal{A}_{t_j} = \phi$. We define a *path* $p(i_0, i_l)$ in $\mathcal{G}$ to be a sequence of distinct ordered nodes $(i_0, i_1, \ldots, i_l)$, where $(i_{j-1}, i_j) \in \mathcal{E}, 1 \le j \le l$. For acyclic networks, a path consists of distinct nodes. For an acyclic directed graph, define $i \prec j$ if there exists a path from $i$ to $j$. The relation $\prec$ is a *partial order*.

### B. A Model of Multisource Multisink Networks

In a multisource multisink network, several mutually independent information sources are generated at possibly different nodes, and each of the information sources is multicast to a specific set of sinks. We assume that there are no incoming edges directed into the sources and no outgoing edges directed from the sinks. This assumption essentially does not lose generality since in an arbitrary multisource multisink network, an intermediate node acting both as a source and a sink can be replicated to separate the multiple functions. That is, a copy of this node is made to act as a source and another copy is made to act as a sink. Both of the two copies are connected to the original node with links of unlimited capacity. For this reason, any multisource multisink network where each node can either act as a source, an internal node, or a sink can be studied in this manner. This is why we assume that the network studied in this paper has the special structure formulated in the previous subsection, although we do lose some generality when the network is assumed to be acyclic.

We denote a multisource multisink network by a directed graph $\mathcal{G} = (\tilde{\mathcal{V}}, \mathcal{E})$ defined above with a set of transmission requirements $\mathcal{M}$, which is specified by

1) $\mathcal{S}$, the set of source nodes;
2) $\mathcal{T}$, the set of sink nodes;
3) $\boldsymbol{C} = (C_{ij} : (i,j) \in \mathcal{E})$, where $C_{ij} > 0$ is the capacity constraint on edge $(i,j)$;
4) $\mathcal{D} = \{D_i : t_i \in \mathcal{T}\}$, where $D_i \in 2^S \setminus \phi$ specify the reconstruction requirements of the sink nodes.

The $j$th source, which is generated at source node $s_j$, is denoted by $\mathbf{X}_j = \{X_{jk}\}_{k=1}^{\infty}$. We assume that $\mathbf{X}_j, j = 1, \ldots, |\mathcal{S}|$ are independent, and $X_{jk}, k = 1, 2, \ldots$ are independent and identically distributed (i.i.d.) copies of a generic random variable $X_j$ with alphabet $\mathcal{X}_j$, where $|\mathcal{X}_j| < \infty$. Note from the source/channel coding separation property, it may not be necessary to use a source model as general as we use here. We may assume that the sources are i.i.d. binary-symmetric sources generated at the source nodes with data rates $(H(X_i) : s_i \in \mathcal{S})$.

We define a *cut* $\mathcal{U}$ in $\mathcal{G} = (\tilde{\mathcal{V}}, \mathcal{E})$ to be a subset $\mathcal{U} \subseteq \tilde{\mathcal{V}}$. For a cut $\mathcal{U}$ we define

$$\mathcal{E}_{\mathcal{U}} = \{(i,j) \in \mathcal{E} : i \in \mathcal{U}, j \in \tilde{\mathcal{V}} \setminus \mathcal{U}\}. \tag{3}$$

Suppose that a cut satisfies the condition that

$$\mathcal{S}_\alpha = \{s_i \in \mathcal{S} : i \in \alpha\} \subseteq \mathcal{U}$$

and

$$\mathcal{T}_\beta = \{t_i \in \mathcal{T} : i \in \beta\} \subseteq \tilde{\mathcal{V}} \setminus \mathcal{U};$$

then we say $\mathcal{U} \in \mathcal{U}_{\alpha,\beta}$, where $\mathcal{U}_{\alpha,\beta}$ represents the set of all $\mathcal{U}$ that have this property, and thus is called the set of cuts with respect to the source set $\alpha$ and sink set $\beta$. For cuts in this set, we call the set of edges across the cut $\mathcal{U}$ between the set of source nodes $\mathcal{S}_\alpha$ and the set of sink nodes $\mathcal{T}_\beta$ the *cut set* $\mathcal{E}_{\mathcal{U}}(\alpha,\beta)$ of $\mathcal{U}$, i.e.,

$$\mathcal{E}_{\mathcal{U}}(\alpha,\beta) = \{(i,j) \in \mathcal{E}(\alpha,\beta) : i \in \mathcal{U} \text{ and } j \notin \mathcal{U}\} \tag{4}$$

where

$$\mathcal{E}(\alpha,\beta) = \{(i,j) \in \mathcal{E} : \exists\, k \in \alpha \text{ such that } s_k \prec i, \exists\, m \in \beta \text{ such that } j \prec t_m\}. \tag{5}$$

For any cut $\mathcal{U} \in \mathcal{U}_{\alpha,\beta}$, define the capacity of the cut $\mathcal{U}$ to be the sum of the capacities of all the edges in $\mathcal{E}_{\mathcal{U}}(\alpha,\beta)$, i.e.,

$$C_{\mathcal{U}}(\alpha,\beta) = \sum_{(i,j) \in \mathcal{E}_{\mathcal{U}}(\alpha,\beta)} C_{ij}. \tag{6}$$

### C. Admissible Rate Region and Capacity Region for Multisource Multisink Networks

Let $\mathcal{G} = (\tilde{\mathcal{V}}, \mathcal{E})$ be a multisource multisink network with a set of transmission requirements $\mathcal{M}$, we define

$$d_i : \left(\prod_{j \in D_i} \mathcal{X}_j\right) \times \left(\prod_{j \in D_i} \mathcal{X}_j\right) \to \{0,1\}$$

as the Hamming distortion measure for $t_i \in \mathcal{T}$, i.e., for any $\mathbf{x}$ and $\mathbf{x}'$ in $(\prod_{j \in D_i} \mathcal{X}_j) \times (\prod_{j \in D_i} \mathcal{X}_j)$

$$d_i(\mathbf{x}, \mathbf{x}') = \begin{cases} 0, & \text{if } \mathbf{x} = \mathbf{x}' \\ 1, & \text{if } \mathbf{x} \neq \mathbf{x}'. \end{cases}$$

Let $X_j^n = (X_{j1}, \ldots, X_{jn})$. We consider an $(n, (\eta_{ij}, (i,j) \in \mathcal{E}), (\Delta_l, l \in \mathcal{T}))$ block code of block length $n$ consisting of

1) for each $(i,j) \in \mathcal{E}$, an encoding function

$$F_{ij} : \prod_{i' \in \mathcal{S} : (i',i) \in \mathcal{E}} \mathcal{X}_{i'}^n \times \prod_{i' \notin \mathcal{S} : (i',i) \in \mathcal{E}} \{0, 1, \ldots, \eta_{i'i} - 1\} \to \{0, 1, \ldots, \eta_{ij} - 1\}; \tag{7}$$

2) for each $l \in \mathcal{T}$, a decoding function

$$G_l : \prod_{l' \in \mathcal{S} : (l',l) \in \mathcal{E}} \mathcal{X}_{l'}^n \times \prod_{l' \notin \mathcal{S} : (l',l) \in \mathcal{E}} \{0, 1, \ldots, \eta_{l'l} - 1\} \to \prod_{k \in D_l} \mathcal{X}_k^n; \tag{8}$$

3) for each $l \in \mathcal{T}$, an average distortion

$$\Delta_l = n^{-1} E \sum_{k=1}^{n} d_l((X_{jk}, j \in D_l), (\hat{X}_{jk}, j \in D_l)) \tag{9}$$

where $(\hat{X}_j^n, j \in D_l)$ are the outputs of the decoder $G_l$ when the inputs are $X_j^n, j = 1, \ldots, |\mathcal{S}|$.

A rate vector $\boldsymbol{R} = (R_{ij}, (i,j) \in \mathcal{E})$ is admissible if for every $\epsilon > 0$, there exists for sufficiently large $n$ an

$$(n, (\eta_{ij}, (i,j) \in \mathcal{E}), (\Delta_l, l \in \mathcal{T}))$$

code such that

$$n^{-1} \log \eta_{ij} \leq R_{ij}, \forall (i,j) \in \mathcal{E}$$

and

$$\Delta_l \leq \epsilon, \forall l \in \mathcal{T}.$$

Let

$$\mathcal{R} = \{\boldsymbol{R} : \boldsymbol{R} \text{ is admissible}\} \tag{10}$$

be the admissible rate region of the sources. If the capacity vector $\boldsymbol{C} \in \mathcal{R}$, then we say that the transmission problem of the sources over the network with the given transmission requirements at the sinks is resolvable. In this paper, we are interested in characterizing the closure of the admissible rate region $\mathcal{R}$ for the sources under this formulation. Let the capacities of the channels be given and denoted by $\boldsymbol{C} = (C_{ij} : (i,j) \in \mathcal{E})$ as a capacity vector. It is apparent that whether or not $\boldsymbol{C} \in \mathcal{R}$ for a particular source is fully determined by the vector $\boldsymbol{H} = (H(X_k) : s_k \in \mathcal{S})$ of the entropy rates of the sources. Let $\mathcal{H}$ be the region consisting of all entropy vectors for which the capacity vector is admissible, i.e.,

$$\mathcal{H} = \{\boldsymbol{H} : \boldsymbol{C} \text{ is admissible}\}. \tag{11}$$

We say that the entropy vectors in this region are achievable and call this region the capacity region. The determination of the closure of $\mathcal{H}$ is an equivalent problem.

## III. Main Results

In this section, we first present our bound for a special class of networks called networks with $K$-pairs transmission requirements (or $K$ pairwise networks, or multiple unicast networks as called in some papers) in which the data from each source is decoded at only one sink for all $K$ sources and each sink is required to decode only one source. Then we generalize the result to the arbitrary transmission requirements case by using sink decomposition technique and by considering all possible $K$-pairs transmission subnetworks as we will define soon.

Let $\mathcal{G} = (\tilde{\mathcal{V}}, \mathcal{E})$ be a multisource multisink network with a set of transmission requirements $\mathcal{M}$. For any $(i,j) \in \mathcal{E}$, we define

$$\begin{aligned} \mathcal{S}_{ij} &= \{s_k \in \mathcal{S} : s_k \prec i\} \\ \mathcal{T}_{ij} &= \{t_l \in \mathcal{T} : j \prec t_l\} \end{aligned} \tag{12}$$

where $\prec$ is the partial order defined in Section II-A.

First, we assume that $|\mathcal{S}| = |\mathcal{T}| = K$ and $D_i = \{s_i\}$. This means that the $i$th source is decoded only at the $i$th sink for all $i$. In this special case, there exists a natural one-to-one correspondence between the source nodes and sink nodes. Therefore, to make the problem formulation more precise, we let $\mathcal{I}=\{1,2,\ldots,K\}$ be the index set of both $\mathcal{S}$ and $\mathcal{T}$. For convenience, we sometimes ignore the difference between $\mathcal{S}$, $\mathcal{T}$, and $\mathcal{I}$. For instance, we may view $\mathcal{S}_{ij}$ as a set of indices of source nodes rather than a set of source nodes themselves. This convention is particularly useful in this special $K$-pairs transmission case.

Suppose that $\gamma \subseteq \{1,2,\ldots,K\}, \gamma \neq \phi$ and $\prec_\gamma$ is a linear order in $\gamma$ (which may not be the natural order). Assuming that $\gamma$ consists of $\{i_1 \prec_\gamma i_2 \prec_\gamma \cdots \prec_\gamma i_{|\gamma|}\}$, we define

$$i_\beta \triangleq \min\{\beta \cap \gamma\} \tag{13}$$

where the minimum is taken with respect to the order $\prec_\gamma$

$$\gamma(\beta) \triangleq \{i \in \gamma : i \prec_\gamma i_\beta, i \neq i_\beta\} \tag{14}$$

and the set of cuts

$$\mathcal{U}_\gamma \triangleq \{\mathcal{U} : \mathcal{S}_\gamma \subseteq \mathcal{U}, \mathcal{T}_\gamma \subseteq \tilde{\mathcal{V}} \setminus \mathcal{U}\} \tag{15}$$

where $\mathcal{S}_\gamma = \{s_i : i \in \gamma\}, \mathcal{T}_\gamma = \{t_i : i \in \gamma\}, \beta \subseteq \mathcal{I}, \beta \neq \phi$.

*Theorem 1:* Given a multisource multisink network $\mathcal{G}=(\tilde{\mathcal{V}},\mathcal{E})$ with $K$-pairs transmission requirements (i.e., $\mathcal{I}=\{1,\ldots,K\}$ and $D_i=\{s_i\}, i \in \mathcal{I}$), if the transmission problem is resolvable, then for any nonempty subset $\gamma \subseteq \mathcal{I}$, any order $\prec_\gamma$ in $\gamma$, and any cut $\mathcal{U} \in \mathcal{U}_\gamma$

$$\sum_{k\in\gamma} H(\mathcal{X}_k) \le \sum_{\substack{(i,j)\in\mathcal{E}_\mathcal{U}:\ \mathcal{T}_{ij}\cap\gamma\neq\phi,\\ \mathcal{S}_{ij}\cap\gamma\not\subseteq\gamma(\mathcal{T}_{ij})}} C_{ij}. \tag{16}$$

From its proof, we will see that this bound is implied by the linear programming bound in [6]. For convenience, we temporarily call this new bound the network-sharing bound because it partially explains the role of side-information channels when several sources share a communication network.

For a $K$-pairs transmission network, this new bound is an improvement over the max-flow min-cut bound, i.e.,

$$\sum_{\substack{(i,j)\in\mathcal{E}_\mathcal{U}:\ \mathcal{T}_{ij}\cap\gamma\neq\phi,\\ \mathcal{S}_{ij}\cap\gamma\not\subseteq\gamma(\mathcal{T}_{ij})}} C_{ij} \le \sum_{\substack{(i,j)\in\mathcal{E}_\mathcal{U}:\ \mathcal{T}_{ij}\cap\gamma\neq\phi,\\ \mathcal{S}_{ij}\cap\gamma\neq\phi}} C_{ij}.$$

This result can be seen as follows. Given a set of nodes $\mathcal{U}$, let $\gamma$ denote the set of messages whose source lies in $\mathcal{U}$ and whose sink lies outside $\mathcal{U}$. If the elements of $\gamma$ are numbered $1,2,\ldots,|\gamma|$, the network sharing bound states that the combined transmission rate of the messages in $\gamma$ is bounded above by the combined capacity of the set consisting of all edges from $\mathcal{U}$ to $\tilde{\mathcal{V}}\setminus\mathcal{U}$ which lie on a path from the source of a message in $\gamma$ to the sink of the same message (Fig. 5(a)–(c)) or a lower-ordered message (Fig. 5(d)) in $\gamma$. However, in this formulation of

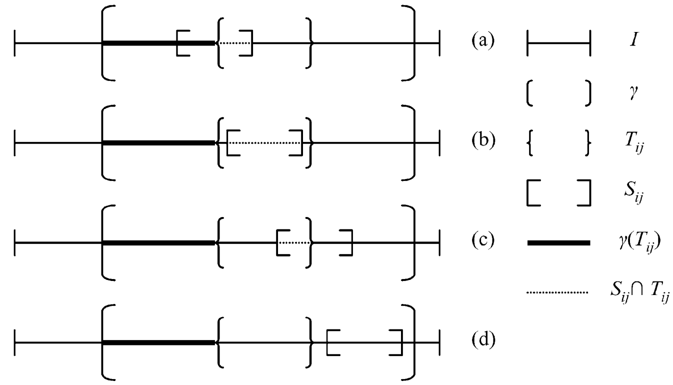


Fig. 5. Graphical description of the network sharing bound. The graph roughly shows possible cases of edges whose combined rate forms the network sharing bound for the message set $\gamma$.

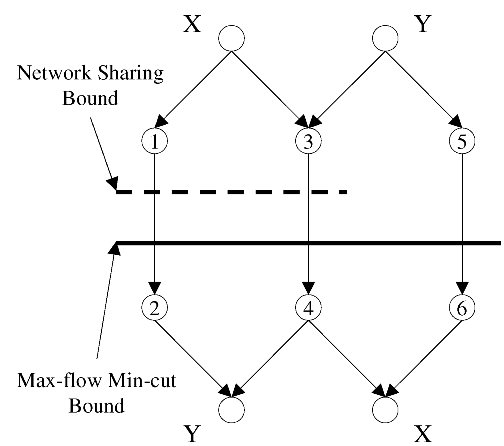


Fig. 6. An example of max-flow min-cut bound and network sharing bound for $\gamma = \{X \prec_\gamma Y\}$.

the bound, if we relax the constraint on "the same message or a lower-ordered message" to "any other message," we obtain the max-flow min-cut bound.

A simple example is the classical two-source two-sink network we gave in Section I, where the new outer bound and the max-flow min-cut bound are shown in Fig. 6. In the figure, the dotted line (labeled "network sharing bound") and the solid line (labeled "max-flow min- cut bound") designate two edge sets; in both cases, the specified bound asserts that the combined capacity of the edges in the designated set is an upper bound on the combined entropy of sources $X$ and $Y$.

In a general multisource multisink network, each sink can request messages from multiple sources. In such networks, our converse proof with $K$-pairs transmission requirements discussed in Section IV seems inapplicable. However, this problem can be easily solved by *sink decomposition*, i.e., we decompose each sink $t_i$ into $|D_i|$ copies, each of them has a single source reconstruction demand and has the same set of connections with other nodes as $t_i$. For example, by sink decomposition, the network in Fig. 2 can be transformed into the network in Fig. 7. Thus, any general multisource multisink network with arbitrary sink demands can be viewed as a

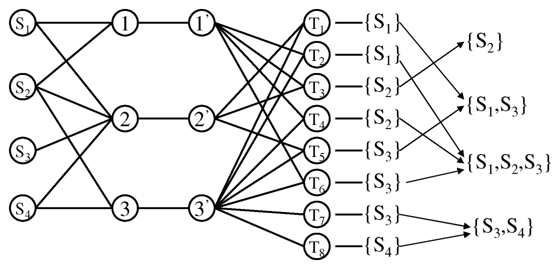


Fig. 7. A sink decomposition for the three-layer network shown in Fig. 2.

multisource multisink network with nonmulticast transmission requirements in which each sink decodes only one source.

By enumerating all possible $|\mathcal{S}|$-tuples in $\mathcal{T}_1 \times \mathcal{T}_2 \times \cdots \times \mathcal{T}_{|\mathcal{S}|}$ where $\mathcal{T}_i$ is the set of sink nodes at which the $i$th source is to be decoded, we get $\prod_{j\in\{1,\ldots,|\mathcal{S}|\}} |\mathcal{T}_j|$ $K$-pairs transmission subnetworks. Therefore, the intersection of our new outer bounds on the capacity regions with respect to these subnetworks gives an outer bound for the capacity region of the general multisource multisink network. Therefore, the sink node decomposition technique and the method in $K$-pairs transmission subnetwork help us to reduce the general multisource network problem to the simplest $K$-pairs transmission network case.

For any multisource multisink nonmulticast network $\mathcal{G}$, $\forall(j_1, j_2, \ldots, j_{|\mathcal{S}|})$, $j_i \in \mathcal{T}_i$ with nonmulticast transmission requirements discussed above, we define $G_{j_1,j_2,\ldots,j_{|\mathcal{S}|}}$ as a multisource multisink *subnetwork* of $\mathcal{G}$ with $K$-pairs transmission requirements by considering only the sinks with the given indices. Let $\mathcal{H}_{j_1,j_2,\ldots,j_{|\mathcal{S}|}}$ be the new outer bound on the entropy rates of the sources derived from this subnetwork. The following is a direct consequence of our main result.

*Corollary 1:* Given an arbitrary multisource multisink nonmulticast network $\mathcal{G} = (\tilde{\mathcal{V}}, \mathcal{E})$ with a set of transmission requirements $\mathcal{M}$, if the source transmission problem is resolvable, then

$$\mathcal{H} \subseteq \bigcap_{(j_1,j_2,\ldots,j_{|\mathcal{S}|})\in\prod_i \mathcal{T}_i} \mathcal{H}_{j_1,j_2,\ldots,j_{|\mathcal{S}|}}. \tag{17}$$

Examples have shown that the network-sharing bound is tight in some special networks and offers significant improvement over the max-flow min-cut bound in general. However, as we will see in Section V, it is also proved not tight in general even though the gap might be considerately small. In fact, the search for a tight bound in general multisource multisink networks is highly nontrivial. One observation that might reduce the complexity of the problem is to consider all possible $K$-pairs subnetworks as we did in the consequence. If this technique is "tight," then the multisource multisink network capacity problem can be reduced to the capacity problem of the class of multisource multisink networks with $K$-pairs transmission requirements discussed in this paper. This conjecture is true in the single-source multicast network case, where a similar idea was used to determine the capacity by using the max-flow min-cut bound to each sink and as long as the max-flow min-cut bound holds for all sinks, the multicast network transmission problem is resolvable. Whether or not this technique is "tight" in general is an unresolved problem. We tend to believe that it is not "tight" in general.

While trying to improving the network-sharing bound, we discovered a technique that can be used to determine the capacity region of a special class of three-layer networks with the following constraints: Let $(i, j)$ be a channel in the network, then

- either $\mathcal{S}_{ij} = \mathcal{T}_{ij}$ and $|\mathcal{S}_{ij}| \leq 2$,
- or $\mathcal{S}_{ij} \cap \mathcal{T}_{ij} = \phi$ and $|\mathcal{S}_{ij}| = |\mathcal{T}_{ij}| = 1$.

We call this class of networks *degree*-2 three-layer networks with $K$-pairs transmission requirements. Index the channels $(i, j)$ by $(\mathcal{S}_{ij}, \mathcal{T}_{ij})$. If two channels have the same indexing, then they are viewed as a single channel with capacity being the sum of the two capacities. Then the channel index is either of the form $(\{i, j\}, \{i, j\})$, $(i, i)$ or of the form $(i, j), i \neq j$ by the assumptions for this particular model. Let the information rate on the first kind of channels be $R_i$, $R_{ij}$, and the information rate on the second kind of channels be denoted by $R_i^j$, then we have the following theorem.

*Theorem 2:* For any degree-2 three-layer network with $K$-pairs transmission requirements, the capacity region is given by

$$\mathcal{R} = \left\{ \mathbf{R} : \forall\gamma, \sum_{i\in\gamma} H(X_i) \leq \sum_{i\in\gamma} R_i \right. \\ \left. + \sum_{\substack{i\in\gamma, j\in\gamma,\\ i\prec_\gamma j}} (R_{ij} + \min\{R_{ij}, R_i^j, R_j^i\}) + \sum_{i\in\gamma, j\notin\gamma} R_{ij} \right\}. \tag{18}$$

## IV. Proof of Main Result

### A. Proof of Theorem 1

Suppose a capacity vector $\boldsymbol{C} = (C_{ij}, (i, j) \in \mathcal{E})$ is admissible for the given network, then for every $\epsilon > 0$, there exists for sufficient large $n$ an $(n, (\eta_{ij}, (i, j) \in \mathcal{E}), (\Delta_l, l \in \mathcal{T}))$ code such that

$$n^{-1} \log \eta_{ij} \leq C_{ij}, \forall (i, j) \in \mathcal{E}$$

and

$$\Delta_l \leq \epsilon, \forall l \in \mathcal{T}.$$

*Lemma 1:* Define inductively

$$U_{ij} = F_{ij}(\{X_{i'}^n : i' \in \mathcal{S}, (i', i) \in \mathcal{E}\}, \\ \{U_{i'i} : i' \notin \mathcal{S}, (i', i) \in \mathcal{E}\}) \tag{19}$$

as the information flow on edge $(i, j) \in \mathcal{E}$. Then we have

1)
$$H(U_{ij}) \leq \log \eta_{ij} \leq nC_{ij}, \tag{20}$$

2)
$$H(U_{ij} | \{X_{i'}^n : i' \in \mathcal{S}, (i', i) \in \mathcal{E}\}, \\ \{U_{i'i} : i' \notin \mathcal{S}, (i', i) \in \mathcal{E}\}) = 0, \tag{21}$$

3) by Fano's inequality, $\exists\, \delta$ depending on $\epsilon$ such that $\delta \to 0$ as $\epsilon \to 0$ and for any $l \in \mathcal{T}$,

$$n^{-1} H(X_l^n | \{U_{il} : (i,l) \in \mathcal{E}\}) \le \delta, \tag{22}$$

4) furthermore

$$H(U_{ij} | \{X_{i'}^n : i' \in \mathcal{S}_{ij}\}) = 0, \tag{23}$$

5) for any $l \in \mathcal{T}$ and for any cut $\mathcal{U} \in \mathcal{U}_{\mathcal{I}}$,

$$n^{-1} H(X_l^n | \{U_{ij} : (i,j) \in \mathcal{E}_{\mathcal{U}}\}) \le \delta. \tag{24}$$

*Proof:*

Part 1) follows from the definition of the

$$(n, (\eta_{ij}, (i,j) \in \mathcal{E}), (\Delta_l, l \in \mathcal{T}))$$

code. Parts 2) and 3) are consequences of the definitions of the encoders and decoders. And parts 4) and 5) can be easily derived from Parts 2) and 3) by induction. □

Let $\gamma = \{i_1 \prec_\gamma i_2 \prec_\gamma \cdots \prec_\gamma i_{|\gamma|}\}$. Consider any cut $\mathcal{U} \in \mathcal{U}_\gamma$. For convenience of notation, we define a set

$$\Delta = \{U_{ij} : (i,j) \in \mathcal{E}_{\mathcal{U}}, \mathcal{T}_{ij} \cap \gamma \ne \phi, \mathcal{S}_{ij} \cap \gamma \nsubseteq \gamma(\mathcal{T}_{ij})\}$$

and another set $\Lambda = \{X_k^n : k \notin \gamma\}$.

Thus, we have

$$\begin{aligned}
&\sum_{(i,j)\in\Delta} nC_{ij} + \sum_{k\notin\gamma} H(X_k^n)\\
&\quad \ge \sum_{(i,j)\in\Delta} H(U_{ij}) + \sum_{k\notin\gamma} H(X_k^n)\\
&\quad \ge H(\Delta, \Lambda)\\
&\quad = H(\Delta, \{X_k^n : k \in \mathcal{S}\}) - H(\{X_k^n : k \in \gamma\} | \Delta, \Lambda)\\
&\quad \overset{(a)}{\ge} H(\{X_k^n : k \in \mathcal{S}\}) - n|\gamma|\delta\\
&\quad = \sum_{k\in\gamma} H(X_k^n) + \sum_{k\notin\gamma} H(X_k^n) - n|\gamma|\delta.
\end{aligned}$$

Therefore,

$$\sum_{k\in\gamma} n^{-1} H(X_k^n) \le \sum_{(i,j)\in\Delta} C_{ij} + |\gamma|\delta.$$

Let $\epsilon \to 0$ and $n \to \infty$, then $\delta \to 0$ and we complete the proof, where the key step (a) is proved as follows.

From the chain rule of entropy functions

$$H(\{X_k^n : k \in \gamma\} | \Delta, \Lambda) = \sum_{l=1}^{|\gamma|} H(X_{i_l}^n | X_{i_1}^n, \ldots, X_{i_{l-1}}^n, \Delta, \Lambda). \tag{25}$$

Thus, we only need to show that

$$\begin{aligned}
&H(X_{i_l}^n | X_{i_1}^n, \ldots, X_{i_{l-1}}^n, \Delta, \Lambda)\\
&\le H(X_{i_l}^n | X_{i_1}^n, \ldots, X_{i_{l-1}}^n, \{U_{ij} : (i,j) \in \mathcal{E}_{\mathcal{U}},\\
&\quad i_l \in \mathcal{T}_{ij}, \mathcal{S}_{ij} \cap \gamma \nsubseteq \gamma(\mathcal{T}_{ij})\}, \Lambda)\\
&\overset{(b)}{=} H(X_{i_l}^n | X_{i_1}^n, \ldots, X_{i_{l-1}}^n, \Lambda, \{U_{ij} : (i,j) \in \mathcal{E}_{\mathcal{U}},\\
&\quad i_{\mathcal{T}_{ij}} = i_t, i_l \in \mathcal{T}_{ij},\\
&\quad \mathcal{S}_{ij} \cap \gamma \nsubseteq \{i_1, \ldots, i_{t-1}\}\} : t = 1, \ldots, l)\\
&\overset{(c)}{\le} H(X_{i_l}^n | \{U_{ij} : (i,j) \in \mathcal{E}_{\mathcal{U}}, i_{\mathcal{T}_{ij}} = i_t, i_l \in \mathcal{T}_{ij},\\
&\quad \mathcal{S}_{ij} \cap \gamma \subseteq \{i_1, \ldots, i_{t-1}\}\} : t = 1, \ldots, l,\\
&\quad \{U_{ij} : (i,j) \in \mathcal{E}_{\mathcal{U}}, i_{\mathcal{T}_{ij}} = i_t, i_l \in \mathcal{T}_{ij},\\
&\quad \mathcal{S}_{ij} \cap \gamma \nsubseteq \{i_1, \ldots, i_{t-1}\}\} : t = 1, \ldots, l)\\
&= H(X_{i_l}^n | \{U_{ij} : (i,j) \in \mathcal{E}_{\mathcal{U}}, i_{\mathcal{T}_{ij}} = i_t, i_l \in \mathcal{T}_{ij}\}\\
&\quad : t = 1, \ldots, l)\\
&\overset{(d)}{=} H(X_{i_l}^n | U_{ij} : (i,j) \in \mathcal{E}_{\mathcal{U}}, i_l \in \mathcal{T}_{ij})\\
&\le n\delta.
\end{aligned}$$

The noted steps are explained as follows:

(b) $i_l \in \mathcal{T}_{ij}$ implies that $i_{\mathcal{T}_{ij}} \prec_\gamma i_l$;

(c) by fact 2), it follows that

$$H(U_{ij} : (i,j) \in \mathcal{E}_{\mathcal{U}}, \mathcal{S}_{ij} \cap \gamma \subseteq \{i_1, \ldots, i_{l-1}\} | X_{i_1}^n, \ldots, X_{i_{l-1}}^n, \Lambda) = 0$$

$\forall l \in \{1, \ldots, |\gamma|\}, \forall \mathcal{T}_{ij} \subseteq \{1, \ldots, |\mathcal{T}|\}, \mathcal{T}_{ij} \ne \phi$ and $H(Y|X) \le H(Y|g(X))$, for any function $g$;

(d) similar to (b).

Therefore, Theorem 1 is proved. □

An intuitive sketch of the above argument is that given all the messages in the complement of $\gamma$ (i.e., $\Lambda$) and given the data transmitted on edges in $\Delta$, one may decode all the messages in $\gamma$ as follows: first one may decode message 1 (in the given order) because every path from any source to the sink of message 1 either originates in $\Lambda$ or crosses $\Delta$; next one may decode message 2 because every path from any source to the sink of message 2 either originates in $\Lambda \cup \{s_1\}$ or crosses $\Delta$, and message 1 has already been decoded. Continuing in this way, one may decode all messages in $\gamma$ in the order $\prec_\gamma$ determined by their numbering.

### B. Proof of Theorem 2

*Proof of Achievability:* Let $\mathcal{H}$ be the capacity region, i.e.,

$$\mathcal{H} = \left\{ \mathbf{H} : \forall \gamma, \sum_{i\in\gamma} H(X_i) \le \sum_{i\in\gamma} R_i + \sum_{\substack{i\in\gamma, j\in\gamma,\\ i\prec_\gamma j}} (R_{ij} + \min\{R_{ij}, R_i^j, R_j^i\}) + \sum_{i\in\gamma, j\notin\gamma} R_{ij} \right\}. \tag{26}$$

This region is defined by $2^K - 1$ constraints, that is, for each nonempty subset $\gamma$ of $\mathcal{I}$, there is a constraint which will be simply called constraint $\gamma$. An extremal point of the region can be specified by at least $K$ active constraints (constraints with equality). We have the following result.

*Lemma 2:* For each extremal point of the region $\mathcal{H}$, there exists $K$ distinct active constraints $\gamma_i$: $i = 1, \ldots, K$ satisfying the condition that $\gamma_1 \subseteq \gamma_2 \subseteq \cdots \subseteq \gamma_K$ where $|\gamma_i| = i$.

*Proof:* Let $\mathbf{H}$ be an extremal point of $\mathcal{H} \subseteq R^K$.

*Observation:* Suppose at $\mathbf{H}$, there exist two sets $\alpha, \beta, \alpha \nsubseteq \beta, \beta \nsubseteq \alpha$, for which the corresponding constraints in (26) are active, i.e.,

$$\sum_{i\in\alpha} H(X_i) = \Omega_\alpha + \Gamma_\alpha - \Pi_\alpha$$
$$\sum_{i\in\beta} H(X_i) = \Omega_\beta + \Gamma_\beta - \Pi_\beta$$

where

$$\Omega_\gamma = \sum_{i\in\gamma} R_i$$
$$\Gamma_\gamma = \sum_{i\in\gamma}\sum_{j\in\mathcal{I}, j\neq i} R_{ij}$$
$$\Pi_\gamma = \sum_{\substack{i\in\gamma, j\in\gamma,\\ i\prec_\gamma j}} (R_{ij} - \min\{R_{ij}, R_i^j, R_j^i\}).$$

Then the two sets $\alpha\cup\beta$ and $\alpha\cap\beta$ are both active.

The observation is proved as follows. We have

$$\begin{aligned}\sum_{i\in\alpha\cup\beta} H(X_i) + \sum_{i\in\alpha\cap\beta} H(X_i) &= \sum_{i\in\alpha} H(X_i) + \sum_{i\in\beta} H(X_i)\\ &= \Omega_\alpha + \Omega_\beta + \Gamma_\alpha + \Gamma_\beta - \Pi_\alpha - \Pi_\beta\\ &\overset{(a)}{=} \Omega_{\alpha\cup\beta} + \Omega_{\alpha\cap\beta} + \Gamma_\alpha + \Gamma_\beta - \Pi_\alpha - \Pi_\beta\\ &\overset{(b)}{=} \Omega_{\alpha\cup\beta} + \Omega_{\alpha\cap\beta} + \Gamma_{\alpha\cup\beta} + \Gamma_{\alpha\cap\beta} - \Pi_\alpha - \Pi_\beta\\ &\overset{(c)}{\geq} \Omega_{\alpha\cup\beta} + \Omega_{\alpha\cap\beta} + \Gamma_{\alpha\cup\beta} + \Gamma_{\alpha\cap\beta} - \Pi_{\alpha\cup\beta} - \Pi_{\alpha\cap\beta}\end{aligned}$$

which together with constraints

$$\sum_{i\in\alpha\cup\beta} H(X_i) \leq \Omega_{\alpha\cup\beta} + \Gamma_{\alpha\cup\beta} - \Pi_{\alpha\cup\beta}$$
$$\sum_{i\in\alpha\cap\beta} H(X_i) \leq \Omega_{\alpha\cap\beta} + \Gamma_{\alpha\cap\beta} - \Pi_{\alpha\cap\beta}$$

implies the desired conclusion.

The noted steps are explained as follows:

(a) Define the indicator function of channel $(i,i)$ as

$$1_\gamma = \begin{cases} 1, & \text{if } i\in\gamma\\ 0, & \text{if } i\notin\gamma\end{cases}$$

then $1_\alpha + 1_\beta = 1_{\alpha\cup\beta} + 1_{\alpha\cap\beta}$, i.e., the number of counts of $(i,i)$ in $\alpha$ and $\beta$ is the same as in $\alpha\cup\beta$ and $\alpha\cap\beta$, thus the sum of rates of such channels should be equal.

(b) Let $Q_i = \sum_{j\in\mathcal{I}, j\neq i} R_{ij}$, thus, $\Gamma_\gamma = \sum_{i\in\gamma} Q_i$. Using a similar indicator function for $Q_i$, we have

$$1_\alpha + 1_\beta = 1_{\alpha\cup\beta} + 1_{\alpha\cap\beta}.$$

Therefore, $Q_i$ appears the same number of times in $\Gamma_\alpha + \Gamma_\beta$ and $\Gamma_{\alpha\cup\beta} + \Gamma_{\alpha\cap\beta}$.

(c) Define $\delta_{ij} = R_{ij} - \min\{R_{ij}, R_i^j, R_j^i\}$ as the excessive rate of $R_{ij}$ over the side information rate $R_i^j$ and $R_j^i$. To show the inequality holds, we list and compare the counts of $\delta_{ij}$ in $\Pi_\alpha + \Pi_\beta$ and $\Pi_{\alpha\cup\beta} + \Pi_{\alpha\cap\beta}$. To facilitate this procedure, we decompose the sets $\alpha, \beta$ and $\alpha\cup\beta, \alpha\cap\beta$ into common subregions as shown in Fig. 8(a), i.e., $\mathrm{I} = \alpha\setminus\beta$, $\mathrm{II} = \alpha\cap\beta$, $\mathrm{III} = \beta\setminus\alpha$, and $\mathrm{IV} = \alpha^c\cap\beta^c$, where $\gamma^c$ stands for the complement of $\gamma$. Let $A$ and $B$ be two subregions, by $A \Leftrightarrow B$, we mean the number of counts of terms with indices $i\in A$ and $j\in B$. We proceed as in the following table:

| $A \Leftrightarrow B$ | $\Pi_\alpha + \Pi_\beta$ | $\Pi_{\alpha\cup\beta} + \Pi_{\alpha\cap\beta}$ |
|---|---|---|
| I $\Leftrightarrow$ I | 1 | 1 |
| I $\Leftrightarrow$ II | 1 | 1 |
| I $\Leftrightarrow$ III | $\boxed{0}$ | $\boxed{1}$ |
| I $\Leftrightarrow$ IV | 0 | 0 |
| II $\Leftrightarrow$ II | 2 | 2 |
| II $\Leftrightarrow$ III | 1 | 1 |
| II $\Leftrightarrow$ IV | 0 | 0 |
| III $\Leftrightarrow$ III | 1 | 1 |
| III $\Leftrightarrow$ IV | 0 | 0 |
| IV $\Leftrightarrow$ IV | 0 | 0 |

Thus, in I $\Leftrightarrow$ III, $\delta_{ij}$ is counted in $\Pi_{\alpha\cup\beta} + \Pi_{\alpha\cap\beta}$ but not in $\Pi_\alpha + \Pi_\beta$. This asserts that $\Pi_\alpha + \Pi_\beta \leq \Pi_{\alpha\cup\beta} + \Pi_{\alpha\cap\beta}$, where the equality holds only if $\delta_{ij} = 0$, for all $(\{i,j\},\{i,j\})$ such that $i\in\mathrm{I},\ j\in\mathrm{III}$.

The observation is proved. Using this observation and induction, it is a simple matter to prove the conclusion of the lemma. That is, by replacing $\alpha, \beta$ with $\alpha\cap\beta$ and $\alpha\cup\beta$, we will eventually reach a set of active constraints satisfying the requirement of the lemma.

To prove the achievability of the region $\mathcal{H}$, we need only to prove the achievability of the extremal points. We simplify the exposition by assuming that sources are i.i.d. binary symmetric with rates per unit time being the entropies of the sources. This does not lose generality since i.i.d. sources with other alphabet size can be first compressed by binary source codes to "almost" i.i.d. binary data stream before transmission. By using block channel codes of block length $n$, without essential loss of generality, we assume that $n$ times of the coding channel capacities are integer valued.

We distinguish different coding strategies.

- Butterfly Coding

  Let $\tilde{R}_{ij} = \min\{R_{ij}, R_i^j, R_j^i\}$. By assuming that $m = n\tilde{R}_{ij}$ is an integer, we code two binary sequences $X_i^m = (X_i(1), \ldots, X_i(m))$ of length $m$ from source $s_i$ and $X_j^m = (X_j(1), \ldots, X_j(m))$ of length $m$ from source $s_j$ by transmitting $X_i^m$ and $X_j^m$ over side-information channels $(i,j)$ and $(j,i)$, respectively. Meanwhile, we

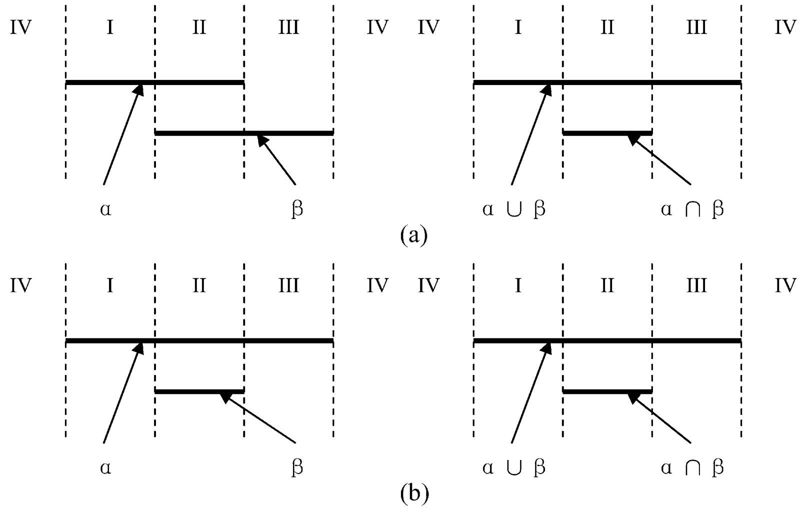


Fig. 8. Graphical representation of sets $\alpha, \beta$ and $\alpha \cup \beta, \alpha \cap \beta$, where region I represents $\alpha \setminus \beta$, II represents $\alpha \cap \beta$, III represents $\beta \setminus \alpha$ and IV represents $(\alpha \cup \beta)^c$. (a) $\alpha, \beta, \alpha \not\subseteq \beta, \beta \not\subseteq \alpha, \alpha \cap \beta \neq \phi$, (b) $\alpha \supseteq \beta$.

transmit $X_i^m \oplus X_j^m$ over the main channel $(\{i,j\},\{i,j\})$, where $\oplus$ stands for the modulo 2 sum. At sink $t_i$, both $X_i^m \oplus X_j^m$ and $X_j^m$ are received. Therefore, $X_i^m$ can be decoded. For the same reason, $X_j^m$ can be decoded at sink $t_j$.

- Transmission Without Coding

  If $\hat{R}_{ij} = R_{ij} - \tilde{R}_{ij} > 0$, by assuming that $l = n\hat{R}_{ij}$ is an integer, we can either transmit a binary sequence $X_i^l$ of length $l$ from source $s_i$ over the channel $(\{i,j\},\{i,j\})$ or a binary sequence $X_j^l$ of length $l$ from source $s_j$ over the channel $(\{i,j\},\{i,j\})$. In both cases, the transmitted data is received at sinks $t_i$ and $t_j$. This coding method is also used for channel $(i,i)$ where $nR_i$ bits can be transmitted from source $s_i$ to sink $t_i$.

*Remarks:*

- It is obvious that for channels $(\{i,j\},\{i,j\})$, $(i,j)$, and $(j,i)$, the two coding strategies can be performed simultaneously without violating the rate constraints for the channels.
- We always assume that the bits from the sources transmitted by distinct coding methods described above over channels are independent. Therefore, the total amount of data transmitted by several distinct methods is simply the summation of the amount of data transmitted by each of these methods.
- When use the transmission without coding over channel $(\{i,j\},\{i,j\})$, we have two possible choices. That is, we can either transmit a binary sequence from source $s_i$ or a binary sequence from source $s_j$. These two possible coding strategies are distinguished by saying that the rate $\hat{R}_{ij}$ is assigned to source $s_i$ when $X_i^l$ is transmitted over the channel or assigned to source $s_j$ otherwise.

The different ways of assigning rates $\hat{R}_{ij}$ for all possible $(i,j)$ to the sources give different information rate vectors for the sources. In the following, we describe a rate assignment strategy which proves the achievability of the extremal point with active constraints $\gamma_1 \subset \cdots \subset \gamma_K$, where $|\gamma_k| = k$ for all $k$. There exists a linear order of the index set $\mathcal{I} : i_1 \prec_\gamma \cdots \prec_\gamma i_K$ such that $\gamma_k = \{i_1, \ldots, i_k\}$. We assign the rate $\hat{R}_{ij}$ to $s_i$ if and only if $i \prec_\gamma j$ and assign to $s_j$ otherwise.

The total number of bits transmitted successfully from source $s_i$ to sink $t_i$ in $n$ time slots, i.e., $nH(X_i)$, is calculated as follows:

$$\begin{aligned} nH(X_i) &= nR_i + n\sum_j \min\{R_{ij}, R_i^j, R_j^i\} \\ &\quad + n\sum_{j:i\prec_\gamma j}(R_{ij} - \min\{R_{ij}, R_i^j, R_j^i\}) \\ &= n[R_i + \sum_{j:i\prec_\gamma j} R_{ij} + \sum_{j:j\prec_\gamma i} \min\{R_{ij}, R_i^j, R_j^i\}]. \end{aligned}$$

Therefore,

$$H(X_i) = R_i + \sum_{j:i\prec_\gamma j} R_{ij} + \sum_{j:j\prec_\gamma i} \min\{R_{ij}, R_i^j, R_j^i\}.$$

For any $k$

$$\begin{aligned} \sum_{i\in\gamma_k} H(X_i) &= \sum_{i\in\gamma_k} R_i + \sum_{i\in\gamma_k}\sum_{j:i\prec_\gamma j} R_{ij} + \\ &\quad + \sum_{i\in\gamma_k}\sum_{j:j\prec_\gamma i} \min\{R_{ij}, R_i^j, R_j^i\} \\ &= \sum_{i\in\gamma_k} R_i + \sum_{\substack{i,j\in\gamma_k,\\ i\prec_\gamma j}} (R_{ij} + \min\{R_{ij}, R_i^j, R_j^i\}) \\ &\quad + \sum_{i\in\gamma_k, j\notin\gamma_k} R_{ij}. \end{aligned}$$

This proves the achievability of the extremal point with active constraints $\gamma_1 \subset \gamma_2 \subset \cdots \subset \gamma_K$ and therefore the achievability of the region $\mathcal{H}$ by time sharing argument.

*Proof of Converse Part:* Suppose that the rate vector $\mathbf{R} = (R_{ij} : (i,j) \in \mathcal{E})$ is admissible for sources with entropies $H(X_i) : i \in \mathcal{I}$. Then, for any $\epsilon > 0$ for sufficiently large $n$, there exists an $(n, (\eta_{ij}, (i,j) \in \mathcal{E}), (\Delta_l : l \in \mathcal{T}))$ code such that

$$n^{-1}\log \eta_{ij} \le R_{ij}, \forall (i,j) \in \mathcal{E},$$

and

$$\Delta_l \leq \epsilon, \forall l \in \mathcal{T}.$$

Following the notation in Section III, let $U_{ij}$ be the information flow on channel $(\{i,j\},\{i,j\})$, $U_i$ be the information flow on channel $(i,i)$, and $U_i^j$ be the information flow on channel $(i,j) : i \neq j$. We have the following facts.

1)
$$H(X_i^n | U_i, \{U_{ij}, \forall j \neq i\}, \{U_k^i, \forall k \neq i\}) = 0; \quad (27)$$

2) the coded information rate meets the constraint of the channel, i.e.,
$$H(U_{ij}) \leq nR_{ij}, \quad H(U_i) \leq nR_i, \quad H(U_i^j) \leq nR_i^j. \quad (28)$$

Thus, we have

$$\begin{aligned}
&\sum_{i\in\gamma} nH(X_i)\\
&\overset{(a)}{=} \sum_{i\in\gamma} I(X_i^n; U_i, \{U_{ij} : j \neq i\}, \{U_k^i : k \neq i\})\\
&\overset{(b)}{=} \sum_{i\in\gamma} I(X_i^n; U_i, \{U_{ij} : j \neq i\} | \{U_k^i : k \neq i\})\\
&= \sum_{i\in\gamma} [I(X_i^n; \{U_{ij} : j \neq i\} | \{U_k^i : k \neq i\})\\
&\quad + I(X_i^n; U_i | \{U_{ij} : j \neq i\}, \{U_k^i : k \neq i\})]\\
&\overset{(c)}{\leq} \sum_{i\in\gamma} [H(U_i) + I(X_i^n; \{U_{ij} : j \neq i\} | \{U_k^i : k \neq i\})]\\
&= \sum_{i\in\gamma} [H(U_i) + \\
&\quad \sum_{j\neq i} I(X_i^n; U_{ij} | \{U_{il} : l < j\}, \{U_k^i : k \neq i\})]\\
&= \sum_{i\in\gamma} [H(U_i) + \sum_{j\neq i} H(U_{ij} | \{U_{il} : l < j\}, \{U_k^i : k \neq i\})\\
&\quad - \sum_{j\neq i} H(U_{ij} | X_i^n, \{U_{il} : l < j\}, \{U_k^i : k \neq i\})]\\
&\overset{(d)}{\leq} \sum_{i\in\gamma} [H(U_i) + \sum_{j\neq i} \{H(U_{ij}|U_j^i) - H(U_{ij}|X_i^n, U_j^i)\}]\\
&= \sum_{i\in\gamma} [H(U_i) + \sum_{j\neq i} I(U_{ij}; X_i^n | U_j^i)]\\
&= \sum_{i\in\gamma} H(U_i) + \sum_{i,j\in\gamma, i\prec_\gamma j} [I(U_{ij}; X_i^n | U_j^i)\\
&\quad + I(U_{ij}; X_j^n | U_i^j)] + \sum_{i\in\gamma, j\notin\gamma} I(U_{ij}; X_i^n | U_j^i)\\
&\overset{(e)}{\leq} \sum_{i\in\gamma} nR_i + n \sum_{i,j\in\gamma, i\prec_\gamma j} [R_{ij} + \min\{R_{ij}, R_i^j, R_j^i\}]\\
&\quad + n \sum_{i\in\gamma, j\notin\gamma} R_{ij}.
\end{aligned}$$

Dividing by $n$, Theorem 2 is proved.

In the proof, the marked steps are explained as follows:

(a) follows from fact 1).

(b) follows from the independence of $X_i^n$ and $U_k^i : k \neq i$.

(c) follows from the inequality
$$I(X_i^n; U_i | \{U_{ij} : j \neq i\}, \{U_k^i : k \neq i\}) \leq H(U_i).$$

(d) follows from the following two facts:
- $H(U_{ij} | \{U_{il} : l < j\}, \{U_k^i : k \neq i\}) \leq H(U_{ij}|U_j^i)$,
- and
$$\begin{aligned}&H(U_{ij} | X_i^n, \{U_{il} : l < j\}, \{U_k^i : k \neq i\})\\ &\quad = H(U_{ij} | X_i^n, U_j^i)\end{aligned}$$
because of the conditional independence of $U_{ij}$ and $\{U_{il} : l < j\}, \{U_k^i : k \neq i, k \neq j\}$ given $X_i^n$.

(e) in this step, we use fact 2) and the following inequality:
$$\begin{aligned}&I(U_{ij}; X_i^n | U_j^i) + I(U_{ij}; X_j^n | U_i^j)\\ &\quad \leq n(R_{ij} + \min\{R_{ij}, R_i^j, R_j^i\}).\end{aligned}$$

This inequality is proved as follows:
$$\begin{aligned}
&I(U_{ij}; X_i^n | U_j^i) + I(U_{ij}; X_j^n | U_i^j)\\
&= H(U_{ij}|U_j^i) + H(U_{ij}|U_i^j)\\
&\quad - H(U_{ij}|X_i^n, U_j^i) - H(U_{ij}|X_j^n, U_i^j)\\
&= H(U_{ij}|U_j^i) + I(U_{ij}; X_j^n; X_i^n, U_j^i | U_i^j)\\
&\leq H(U_{ij}) + I(U_{ij}; X_j^n; X_i^n, U_j^i | U_i^j).
\end{aligned}$$

Similarly, we can prove
$$\begin{aligned}&I(U_{ij}; X_i^n | U_j^i) + I(U_{ij}; X_j^n | U_i^j)\\ &\quad \leq H(U_{ij}) + I(U_{ij}; X_j^n, U_i^j; X_i^n | U_j^i).\end{aligned}$$

We have,
- $I(U_{ij}; X_j^n; X_i^n, U_j^i | U_i^j) \leq H(U_{ij}) \leq nR_{ij}$,
- $$\begin{aligned}&I(U_{ij}; X_j^n; X_i^n, U_j^i | U_i^j) \leq I(X_j^n; X_i^n, U_j^i | U_i^j)\\ &\quad = I(X_j^n; U_j^i | X_i^n, U_i^j) = H(U_j^i) \leq nR_j^i,\end{aligned}$$
- similarly, we have
$$I(U_{ij}; X_j^n, U_i^j; X_i^n | U_j^i) \leq nR_{ij}$$
and
$$I(U_{ij}; X_j^n, U_i^j; X_i^n | U_j^i) \leq nR_i^j.$$

These inequalities imply the desired result. □

## V. Examples

In this section, an example is given to illustrate the significant improvement of the network sharing bound over the max-flow min-cut bound in some special cases. However, the same example also proves that the network-sharing bound is still not tight in general. As an example of a degree-2 three-layer network with $K$-pairs transmission requirements, we further show how its capacity region can be determined by Theorem 2.

*Example 1:* Consider the three-layer network $\mathcal{G}$ in Fig. 9, where

$$\mathcal{I} = \{1,2,3\}$$
$$D_1 = \{1\}, D_2 = \{2\}, D_3 = \{3\}$$

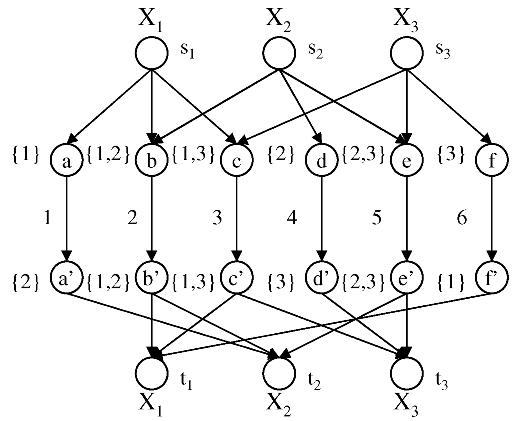


Fig. 9. An example of a three-pairs degree-2 three-layer network. For each edge $(i,i')$, the labeled sets next to nodes $i$ and $i'$ stand for $\mathcal{S}_{ii'}$ and $\mathcal{T}_{ii'}$, respectively.

$$C_{aa'} = C_{bb'} = C_{cc'} = C_{dd'} = C_{ee'} = C_{ff'} = 1$$
$$\mathcal{S}_{aa'} = \{1\}, \mathcal{S}_{bb'} = \{1,2\}, \mathcal{S}_{cc'} = \{1,3\}$$
$$\mathcal{S}_{dd'} = \{2\}, \mathcal{S}_{ee'} = \{2,3\}, \mathcal{S}_{ff'} = \{3\}$$
$$\mathcal{T}_{aa'} = \{2\}, \mathcal{T}_{bb'} = \{1,2\}, \mathcal{T}_{cc'} = \{1,3\}$$
$$\mathcal{T}_{dd'} = \{3\}, \mathcal{T}_{ee'} = \{2,3\}, \mathcal{T}_{ff'} = \{1\}.$$

The max-flow min-cut bound can be easily obtained as follows:

$$H(X_1) \le 2$$
$$H(X_2) \le 2$$
$$H(X_3) \le 2$$
$$H(X_1) + H(X_2) \le 4$$
$$H(X_1) + H(X_3) \le 4$$
$$H(X_2) + H(X_3) \le 4$$
$$H(X_1) + H(X_2) + H(X_3) \le 6.$$

Now, let us examine the network-sharing bound. Since the given network is symmetric, which induces only cyclic permutations of message set $\gamma$, we only need to examine two orders $\gamma = \{1 \prec 2 \prec 3\}$ and $\gamma = \{1 \prec 3 \prec 2\}$. Since the bounds of subsets of two or fewer source nodes are easy to be checked, here we just give the derivation for the bound with three sources. For each order, we examine each edge with the bound conditions in the first table at the bottom of the page. Thus, only edges $(b,b')$, $(c,c')$, $(e,e')$, and $(f,f')$ satisfy the bound condition, thus we obtain

$$H(X_1)+H(X_2)+H(X_3) \le C_{bb'}+C_{cc'}+C_{ee'}+C_{ff'} = 4.$$

Similarly, for $\gamma = \{1 \prec 3 \prec 2\}$, we have the values in the second table at the bottom of the page, and

$$H(X_1) + H(X_2) + H(X_3) \le C_{bb'} + C_{cc'} + C_{dd'} + C_{ee'} + C_{ff'} = 5.$$

Thus, by taking the minimum, we obtain the complete network-sharing bound

$$H(X_1) \le 2$$
$$H(X_2) \le 2$$
$$H(X_3) \le 2$$

| $\gamma = \{1 \prec 2 \prec 3\}$ | | | |
|---|---|---|---|
| $(i,i')$ | $\gamma(\mathcal{T}_{ii'})$ | $\mathcal{T}_{ii'} \cap \gamma \neq \phi$ | $\mathcal{S}_{ii'} \cap \gamma \not\subseteq \gamma(\mathcal{T}_{ii'})$ |
| $(a,a')$ | $\{1\}$ | $\surd$ | $\times$ |
| $(b,b')$ | $\phi$ | $\surd$ | $\surd$ |
| $(c,c')$ | $\phi$ | $\surd$ | $\surd$ |
| $(d,d')$ | $\{1,2\}$ | $\surd$ | $\times$ |
| $(e,e')$ | $\{1\}$ | $\surd$ | $\surd$ |
| $(f,f')$ | $\phi$ | $\surd$ | $\surd$ |

| $\gamma = \{1 \prec 3 \prec 2\}$ | | | |
|---|---|---|---|
| $(i,i')$ | $\gamma(\mathcal{T}_{ii'})$ | $\mathcal{T}_{ii'} \cap \gamma \neq \phi$ | $\mathcal{S}_{ii'} \cap \gamma \not\subseteq \gamma(\mathcal{T}_{ii'})$ |
| $(a,a')$ | $\{1\}$ | $\surd$ | $\times$ |
| $(b,b')$ | $\phi$ | $\surd$ | $\surd$ |
| $(c,c')$ | $\phi$ | $\surd$ | $\surd$ |
| $(d,d')$ | $\{1\}$ | $\surd$ | $\surd$ |
| $(e,e')$ | $\{1\}$ | $\surd$ | $\surd$ |
| $(f,f')$ | $\phi$ | $\surd$ | $\surd$ |

$$\begin{aligned}H(X_1)+H(X_2)&\le 3\\ H(X_1)+H(X_3)&\le 3\\ H(X_2)+H(X_3)&\le 3\\ H(X_1)+H(X_2)+H(X_3)&\le 4\end{aligned}$$

which suggests a significant improvement over the max-flow min-cut bound. However, it is not hard to see that the information rate triple $(2,1,1)$ is not achievable although it satisfies the network-sharing bound. In this case, $X_1$ is transmitting two bits $X_{11}$ and $X_{12}$ at each time unit. At sink $t_1$, the side information $X_3$ from edge $(f,f')$ is not enough to decode both $X_{11}$ and $X_{12}$. Thus, $C_{ff'}$ should not be included in the outer bound. In fact, to make the bound tight, the last inequality should be replaced by

$$\begin{aligned}&H(X_1)+H(X_2)+H(X_3)\\ &\le \sum_{i\in\{1,2,3\}} R_i + \sum_{\substack{i,j\in\{1,2,3\},\\ i\prec j}} (R_{ij}+\min\{R_{ij},R_i^j,R_j^i\})\\ &\quad+\sum_{\substack{i\in\{1,2,3\},\\ j\notin\{1,2,3\}}} R_{ij}\\ &=0+3+0\\ &=3,\end{aligned}$$

as implied by Theorem 2.

## VI. Minimum-Cost Network Coding

An important consequence of the network-sharing bound is the following observation. In a three-layer network, the network-sharing bound implies that

$$\inf\left\{\sum_{(i,j)\in\mathcal{F}} R_{ij} : \boldsymbol{R}\in\mathcal{R}\right\} = \sum_{j\in\mathcal{S}} H(X_j) \tag{29}$$

where $\mathcal{F}$ is the set of coding channels. This result means that, at least in the pairwise transmission case, coding among messages from different sources will not decrease the total rate of data flow in the network. If our goal is to minimize the total data rate of all channels in the network, coding messages from different sources has no benefit. This point can be seen from the following intuition: if we code messages from different sources, then the data rate for some of the side-information channels must be nonzero. This additional cost will not be compensable by the saving of coding. This is an important observation since it implies that coding in the single-source multicast case might be sufficient to achieve minimum total transmission cost.

Let the total rate of the side-information channels in a three-layer network be

$$R_s \triangleq \sum_{(i,j):\mathcal{S}_{ij}\cap\mathcal{T}_{ij}=\phi} R_{ij} \tag{30}$$

then we have the following.

*Corollary 2:*

$$\sum_{j\in\mathcal{S}} H(X_j) \le \sum_{(i,j)\in\mathcal{F}} R_{ij} - 2^{-|\mathcal{S}|}R_s. \tag{31}$$

*Proof:* Letting $\gamma=\mathcal{S}$ in Theorem 1, we obtain

$$\begin{aligned}&\sum_{j\in\mathcal{S}} H(X_j)\\ &\overset{(a)}{\le} \sum_{(i,j):\mathcal{S}_{ij}\cap\mathcal{T}_{ij}\neq\phi} R_{ij} + \min_{\prec} \sum_{\substack{(i,j):\mathcal{S}_{ij}\cap\mathcal{T}_{ij}=\phi,\\ \mathcal{S}_{ij}\nsubseteq\gamma(\mathcal{T}_{ij})}} R_{ij}\\ &\le \sum_{(i,j):\mathcal{S}_{ij}\cap\mathcal{T}_{ij}\neq\phi} R_{ij} + \frac{1}{|\mathcal{S}|!}\sum_{\prec} \sum_{\substack{(i,j):\mathcal{S}_{ij}\cap\mathcal{T}_{ij}=\phi,\\ \mathcal{S}_{ij}\nsubseteq\gamma(\mathcal{T}_{ij})}} R_{ij}\\ &\overset{(b)}{=} \sum_{(i,j):\mathcal{S}_{ij}\cap\mathcal{T}_{ij}\neq\phi} R_{ij}\\ &\quad+\frac{1}{|\mathcal{S}|!}\sum_{(i,j):\mathcal{S}_{ij}\cap\mathcal{T}_{ij}=\phi} |\mathcal{S}|!\left(1-\frac{1}{\binom{|\mathcal{S}_{ij}|+|\mathcal{T}_{ij}|}{|\mathcal{S}_{ij}|}}\right)R_{ij}\\ &\overset{(c)}{\le} \sum_{(i,j):\mathcal{S}_{ij}\cap\mathcal{T}_{ij}\neq\phi} R_{ij} + \sum_{(ij):\mathcal{S}_{ij}\cap\mathcal{T}_{ij}=\phi} (1-2^{-|\mathcal{S}|})R_{ij}\\ &= \sum_{(i,j)\in\mathcal{F}} R_{ij} - 2^{-|\mathcal{S}|}R_s\end{aligned}$$

where

(a) follows from the fact that $\mathcal{S}_{ij}\cap\mathcal{T}_{ij}\neq\phi$ implies $\mathcal{S}_{ij}\nsubseteq\gamma(\mathcal{T}_{ij})$. The minimization is taken over all possible linear orders in $\mathcal{S}$.

(b) follows from that for all $\mathcal{S}_{ij},\mathcal{T}_{ij}$ satisfying $|\mathcal{S}_{ij}|=a$, $|\mathcal{T}_{ij}|=b, \mathcal{S}_{ij}\cap\mathcal{T}_{ij}=\phi$, the total number of pairs $(\mathcal{S}_{ij},\mathcal{T}_{ij})$ is $\binom{|\mathcal{S}|}{a,b,|\mathcal{S}|-a-b}$; furthermore, for a fixed order $\prec$ in $\mathcal{S}$, the number of $(\mathcal{S}_{ij},\mathcal{T}_{ij})$ satisfying $|\mathcal{S}_{ij}|=a$, $|\mathcal{T}_{ij}|=b, \mathcal{S}_{ij}\cap\mathcal{T}_{ij}=\phi$, and $\mathcal{S}_{ij}\subseteq\gamma(\mathcal{T}_{ij})$ is $\binom{|\mathcal{S}|}{a+b}$. This is obtained when we choose $\mathcal{S}_{ij}$ and $\mathcal{T}_{ij}$ jointly with a fixed order, then there is only one pair satisfying $\mathcal{S}_{ij}\subseteq\gamma(\mathcal{T}_{ij})$. From the property of symmetry, a portion of

$$\frac{\binom{|\mathcal{S}|}{a+b}}{\binom{|\mathcal{S}|}{a,b,|\mathcal{S}|-a-b}} = \frac{1}{\binom{|\mathcal{S}_{ij}|+|\mathcal{T}_{ij}|}{|\mathcal{S}_{ij}|}} \tag{32}$$

should be excluded from the bound.

(c) follows from the fact

$$\binom{|\mathcal{S}_{ij}|+|\mathcal{T}_{ij}|}{|\mathcal{S}_{ij}|} \le 2^{|\mathcal{S}_{ij}|+|\mathcal{T}_{ij}|} \le 2^{|\mathcal{S}|}. \tag{33}$$

□

This result shows that once we code data from different sources, the rate increase caused by using side-information channels is greater than the savings achieved by coding. This problem becomes much more complicated in a multisource multisink network, even more so when costs are assigned to bandwidth for each edge. In [16], a measure of bandwidth saving is defined to determine the saving of bandwidth of network coding over routing for a given network and cost assignment. We extend the above measure as

$$S(\mathbf{a}) = 1-\frac{R_c(\mathbf{a})}{R_{nc}(\mathbf{a})} \tag{34}$$

where $R_c(\mathbf{a})$ is the minimum total required cost with coding data from different sources permitted and $R_{nc}(\mathbf{a})$ is the minimum total required cost without coding data from different sources, and $\mathbf{a}$ is a cost factor assignment used to calculate the total cost. Depending on the topology and the assigned cost factors of the given network, the cost saving $S(\mathbf{a})$ could be either positive or zero. That is, sometimes coding data from different sources does not bring any benefit over only coding data from the same source. Therefore, we are interested in defining such a cost factor region $\mathcal{A}$ for a given network which gives the set of cost factors for which minimum cost is achieved without coding data from different sources.

For a multisource multisink network $\mathcal{G} = (\tilde{\mathcal{V}}, \mathcal{E})$ with cost assignment $\mathbf{a} = (a_{ij} : (i,j) \in \mathcal{E})$, let

$$R(\mathbf{a}, \mathcal{R}) = \min_{\mathbf{r} \in \mathcal{R}} \sum_{(i,j) \in \mathcal{E}} a_{ij} r_{ij} \tag{35}$$

be the minimum total cost [17] of transmission in the code rate region $\mathcal{R}$. Let $\mathcal{R}_c$ and $\mathcal{R}_{nc}$ denote the rate region with and without coding data from different sources, respectively.

*Definition 1:* Given $\mathcal{G}$ with sources $\mathbf{X} = \{X_k\}_{k=1}^K$ and information rates $\{H(X_k)\}_{k=1}^K$, define

$$\mathcal{A}(\mathbf{X}) = \{\mathbf{a} : R(\mathbf{a}, \mathcal{R}_{nc}) = R(\mathbf{a}, \mathcal{R}_c)\} \tag{36}$$

as the cost factor region for which minimum cost is achieved without coding data from different sources.

Therefore, our goal is to characterizing $\mathcal{A}(\mathbf{X})$ for a given network with fixed sources $\mathbf{X}$ and information rates $\{H(X_k)\}_{k=1}^K$.

*Example 2:* Consider the three-layer network discussed in Section I. For clarity, we redraw its graph in Fig. 10. We assume $X$ and $Y$ are both of information rate of one bit and pose no constraint on the capacities except the coding channels. Let $\mathbf{a} = (a_{ij} : (i,j) \in \mathcal{F})$ be the cost assignments. Therefore, we have

$$\begin{aligned} R_c &= \min\{a_{12} + a_{34} + a_{56}, 2a_{34}\} \\ R_{nc} &= 2a_{34} \\ \Rightarrow \mathcal{A}(X,Y) &= \{\mathbf{a} : a_{34} \le a_{12} + a_{56}\}. \end{aligned}$$

Corollary 2 simply says that the all–1 vector belongs to the region $\mathcal{A}(\mathbf{X})$ for all sources.

## VII. Conclusion

In this paper, we proved an improved outer bound for the admissible rate region (as well as the capacity region) for an arbitrary acyclic multisource multisink network. Although the proposed network-sharing bound is still not tight in general, it provides significant improvement over the max-flow min-cut bound. As a consequence of this bound, we made progress on the minimum-cost network coding problem by defining a cost factor region for which minimum cost is achieved by restricting to strategies without coding data from different sources. The characterization of this region is worthy of further investigation.

A method of simplifying the network capacity problem is used in this paper when we derive an outer bound for the network by considering all possible subnetworks with pairwise (unicast) transmission requirements. Since this method gives a tight result in the single-source multicast case, it is interesting to know wether it still causes no losses in more general cases.

Besides the techniques used in the proof of the new outer bound, another proof technique was also proposed which leads to the determination of the capacity region for a class of simple networks. Although this method is not fully developed in this paper, we believe that it is a useful tool for further improving the outer bound.

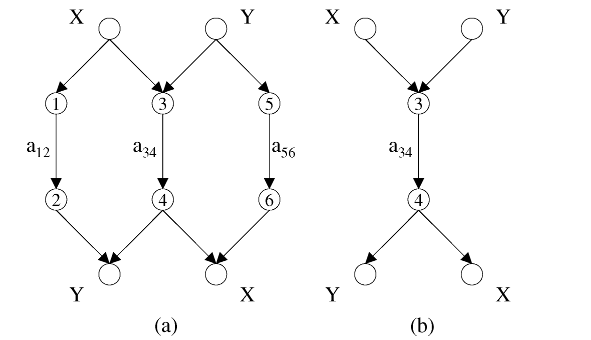


Fig. 10. Minimum cost transmission in a three-layer two-source two-sink network. (a) Coding among two sources. (b) No coding among two sources.

## Acknowledgment

The authors are indebted to the reviewers for their valuable comments. Special thanks should be given to reviewer A for the numerous constructive suggestions which improved the quality and presentation of this paper.

## References

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, “Network information flow,” *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.
[2] S.-Y. R. Li, R. W. Yeung, and N. Cai, “Linear network coding,” *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.
[3] E. Erez and M. Feder, “Capacity region and network codes for two receivers multicast with private and common data,” in *Proc. Workshop on Coding, Cryptography and Combinatorics*, Huangshen City, China, 2003.
[4] C. K. Ngai and R. W. Yeung, “Multisource network coding with two sinks,” in *Proc. Int. Conf. Communications, Circuits and Systems (ICCAS)*, vol. 1, Chengdu, China, Jun. 2004, pp. 34–37.
[5] R. W. Yeung, *A First Course in Information Theory*. New York: Kluwer/Plenum, 2002.
[6] R. W. Yeung and Z. Zhang, “Distributed source coding for satellite communications,” *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 1111–1120, May 1999.
[7] L. Song and R. W. Yeung, “Zero-error network coding for acyclic network,” *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3129–3139, Dec. 2003.
[8] R. W. Yeung, “Multilevel diversity coding with distortion,” *IEEE Trans. Inf. Theory*, vol. 41, no. 2, pp. 412–422, Mar. 1995.
[9] R. W. Yeung and Z. Zhang, “On symmetrical multilevel diversity coding,” *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 609–621, Mar. 1999.
[10] R. W. Yeung, “A framework for linear information inequalities,” *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1924–1934, Nov. 1997.
[11] Z. Zhang and R. W. Yeung, “A non-Shannon type conditional inequality of information quantities,” *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1982–1985, Nov. 1997.
[12] ——, “On characterization of entropy function via information inequalities,” *IEEE Trans. Inf. Theory*, vol. 44, no. 4, pp. 1440–1452, Jul. 1998.
[13] R. W. Yeung and Z. Zhang, “A class of non-Shannontype information inequalities and their applications,” *Commun. Inf. and Syst.*, vol. 1, pp. 87–100, Jan. 2001.
[14] R. Koetter and M. Médard, “An algebraic approach to network coding,” *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 782–795, Oct. 2003.
[15] X. Yan, J. Yang, and Z. Zhang, “An improved outer bound for multisource multisink network coding,” in *Proc. 1st Workshop on Network Coding,Theory, and Applications (NETCOD)*, Riva del Garda, Italy, Apr. 2005.
[16] R. W. Yeung, “Two approaches to quantifying the bandwidth advantage of network coding,” in *Proc. 2004 IEEE Information Theory Workshop*, San Antonio, TX, Oct. 2004.
[17] D. Lun, M. Médard, T. Ho, and R. Koetter, “Network coding with a cost criterion,” in *Proc. Int. Symp. Information theory and Its Applications*, Parma, Italy, Oct. 2004.