Generalized Network Sharing Outer Bound and the Two-Unicast Problem

Sudeep U. Kamath, David N. C. Tse and Venkat Anantharam
Wireless Foundations, Dept of EECS, University of California at Berkeley, Berkeley, California 94720, USA
{sudeep, dtse, ananth}@eecs.berkeley.edu

***Abstract*— We describe a simple improvement over the Network Sharing Bound [1] for the multiple unicast problem. We call this the Generalized Network Sharing (GNS) Outer Bound. We note two properties of this bound with regard to the two-unicast problem: a) it is the tightest bound that can be realized using only edge-cut bounds and b) it is tight in the special case when all edges except those from a so-called minimal GNS set have sufficiently large capacities. Finally, we present an example showing that the GNS outer bound is not tight for the two-unicast problem.**

Keywords: Two-unicast problem, Edge-cut bounds, Generalized Network Sharing outer bound, GNS set

I. Introduction

Recent results in network coding due to Dougherty, Freiling, Zeger suggest that characterizing the capacity region of a general multi-source multi-sink network is hard: scalar-linear solvability of a general network is equivalent to the solvability of a general polynomial collection [2], linear coding is insufficient to achieve capacity [3], non-Shannon information inequalities can strictly improve outer bounds on the capacity region of a network obtained by Shannon information inequalities alone [4]. Further, Chan and Grant show in [5] that the problem of determining the achievable rate pairs (R_0, R_1) in a network with two messages with collocated sources but many destinations, each requesting either the common message or both messages, is equivalent to the problem of characterizing the set of all *almost entropic functions,* $\bar{\Gamma}^*$. The networks presented as "counterexamples" in these works have three or more sources or three or more destinations. A natural question to ask is whether having fewer sources and destinations will lead to a more amenable problem.

We are led to study the problem of two sources and two destinations - each source with an independent message for its own destination - i.e. the two-unicast problem, as a possible fruitful direction. The only network coding results in the literature dealing exclusively with the two-unicast networks are [6] and [7] which provide necessary and sufficient conditions for achieiving $(1, 1)$ in a two-unicast network with all links having integer capacities. This result unfortunately, relies heavily on the assumption of integer link capacities, and hence cannot give us necessary and sufficient conditions for achieving other points such as $(2, 2)$ or $(3, 3)$ by scaling of link capacities. Characterizing the capacity region of a given two-unicast network seems like an interesting direction, which is the motivation for this work.

We provide an outer bound for the multiple unicast problem that is a simple improvement over the Network Sharing outer bound [1], which we call the Generalized Network Sharing (GNS) outer bound. We observe two interesting properties of this bound related to the two-unicast problem - properties that suggest that the bound may be tight for all two-unicast networks. Unfortunately we find that this is not the case and conclude the paper with a two-unicast "counterexample".

II. Network Model

A network $\mathcal{N}$ consists of a directed acyclic graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ along with a link-capacity vector $\underline{\mathrm{C}} = (C_e)_{e \in \mathcal{E}(\mathcal{G})}$ with $C_e \in \mathbb{R}_{\geq 0} \cup \{\infty\}$ $\forall e \in \mathcal{E}(\mathcal{G})$. An n-unicast network $(n \geq 1)$ has n distinguished vertices $s_1, s_2, \ldots, s_n$ called sources and n distinguished vertices $t_1, t_2, \ldots, t_n$ called destinations, where each source s_i has independent information to be communicated to destination t_i.

For edge $e = (v, v') \in \mathcal{E}(\mathcal{G})$, define $\mathsf{tail}(e) := v$ and $\mathsf{head}(e) := v'$, the edge being directed from the tail to the head. For $v \in \mathcal{V}(\mathcal{G})$, let $\mathsf{In}(v)$ and $\mathsf{Out}(v)$ denote the edges entering into and leaving v respectively.

For $S \subseteq \mathcal{E}(\mathcal{G})$, define $C(S) := \sum_{e \in S} C_e$. For disjoint non-empty $A, B \subseteq \mathcal{V}(\mathcal{G})$, we say $S \subseteq \mathcal{E}(\mathcal{G})$ is an $A - B$ cut if there is no directed path from any vertex in A to any vertex in B in the graph $\mathcal{G} \setminus S$. Define the *mincut* from A to B by $c(A; B) := \min \{C(S) : S \text{ is an } A - B \text{ cut}\}$.

We say that the rate tuple $(R_1, R_2, \ldots, R_n)$ is *achievable* for the n-unicast network $\mathcal{N} = (\mathcal{G}, \underline{\mathrm{C}})$, if there exists a positive integer N (called block length), a finite alphabet $\mathcal{A}$ and encoding functions

- For $e \in \mathsf{Out}(s_i)$, $f_e : \mathcal{A}^{\lceil NR_i \rceil} \mapsto \mathcal{A}^{\lfloor NC_e \rfloor}$, $1 \leq i \leq n$
- For $e \in \mathsf{Out}(v)$, $v \neq s_i, 1 \leq i \leq n$, $f_e : \Pi_{e' \in \mathsf{In}(v)} \mathcal{A}^{\lfloor NC_{e'} \rfloor} \mapsto \mathcal{A}^{\lfloor NC_e \rfloor}$

and decoding functions $f_{t_i} : \Pi_{e' \in \mathsf{In}(t_i)} \mathcal{A}^{\lfloor NC_{e'} \rfloor} \mapsto \mathcal{A}^{\lceil NR_i \rceil}$, $1 \leq i \leq n$, so that $\forall (m_1, m_2, \ldots, m_n) \in \Pi_{j=1}^n \mathcal{A}^{\lceil NR_j \rceil}$, we have $g_{t_i}(m_1, m_2, \ldots, m_n) = m_i$, $\forall i, 1 \leq i \leq n$ where $g_{t_i} : \Pi_{j=1}^n \mathcal{A}^{\lceil NR_j \rceil} \mapsto \mathcal{A}^{\lceil NR_i \rceil}$ are functions induced inductively by $\{f_e : e \in \mathcal{E}(\mathcal{G})\}$ and $f_{t_i}, 1 \leq i \leq n$.

The capacity region for an n-unicast network $\mathcal{N}$, denoted $\mathcal{C}(\mathcal{N}) = \mathcal{C}(\mathcal{G}, \underline{\mathrm{C}})$, is defined as the closure of the set of achievable rate tuples. The closure of the set of achievable rate tuples over choice of $\mathcal{A}$ as any finite field and all functions being linear operations on N-dimensional vector spaces over the finite field, is called the vector linear coding capacity region $\mathcal{C}_{\mathsf{vector}}$. If we further have $N = 1$, then the convex closure of achievable rate tuples is called the scalar linear coding capacity region $\mathcal{C}_{\mathsf{scalar}}$. We consider only two-unicast networks in this paper.

III. Generalized Network Sharing Outer Bound

We first describe the Network Sharing outer bound and the Generalized Network Sharing (GNS) outer bound for the case of a two-unicast network.

Theorem 1: (Network Sharing outer bound [1]) Fix $(i,j) = (1,2)$ or $(2,1)$. For a two-unicast network $\mathcal{N} = (\mathcal{G}, \underline{\mathrm{C}})$, if $T \subseteq \mathcal{E}(\mathcal{G})$ is an $s_1, s_2 - t_1, t_2$ cut and if $S \subseteq T$ such that for each edge $e \in T \setminus S$, we have that $\mathsf{tail}(e)$ is reachable from s_i but not from s_j and $\mathsf{head}(e)$ can reach t_j but not t_i, then we have $R_1 + R_2 \leq C(S)\ \forall (R_1, R_2) \in \mathcal{C}(\mathcal{N})$.

We define a set $S \subseteq \mathcal{E}(\mathcal{G})$ to be a *GNS set* if

- $\mathcal{G} \setminus S$ has no paths from s_1 to t_1, s_2 to t_2 and s_2 to t_1 OR
- $\mathcal{G} \setminus S$ has no paths from s_1 to t_1, s_2 to t_2 and s_1 to t_2.

Theorem 2: (GNS outer bound) For a two-unicast network $\mathcal{N} = (\mathcal{G}, \underline{\mathrm{C}})$ and a GNS set $S \subseteq \mathcal{E}(\mathcal{G})$, we have $R_1 + R_2 \leq C(S)\ \forall\ (R_1, R_2) \in \mathcal{C}(\mathcal{N})$.

Note that Theorem 2 implies Theorem 1.

Proof: Consider a scheme of block length N achieving rate pair (R_1, R_2) over alphabet $\mathcal{A}$. Let W_1, W_2 be independent and distributed uniformly over the sets $\mathcal{A}^{\lceil NR_1 \rceil}$ and $\mathcal{A}^{\lceil NR_2 \rceil}$ respectively. For each edge e, define X_e as the concatenated evaluation of the functions specified by the scheme for edge e.

Let $X_S := (X_e)_{e \in S}$. Then, $H(W_1, W_2 | X_S) = H(W_1 | X_S) + H(W_2 | W_1, X_S)$. But, $H(W_1 | X_S) = 0$ because $\mathcal{G} \setminus S$ has no paths from s_1 or s_2 to t_1. And $H(W_2 | W_1, X_S) = 0$ because $\mathcal{G} \setminus S$ has no paths from s_2 to t_2. Thus, $H(W_1, W_2 | X_S) = 0$. So, $N \cdot \log |\mathcal{A}| \cdot (R_1 + R_2) \leq H(W_1) + H(W_2) = H(W_1, W_2) \leq H(X_S) \leq N \cdot \log |\mathcal{A}| \cdot C(S)$.

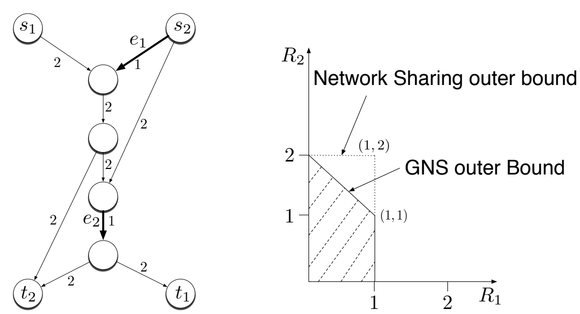
As the inequality holds for every achievable rate pair, it also holds for every point in the closure of the set of achievable rate pairs. ∎

For a two-unicast network $\mathcal{N}$, let the *GNS sum-rate bound* $c_{\mathsf{gns}}(s_1, s_2; t_1, t_2)$ be defined as $c_{\mathsf{gns}}(s_1, s_2; t_1, t_2) := \min\{C(S) : S \subseteq \mathcal{E}(\mathcal{G})$ is a GNS set$\}$. The *GNS outer bound* is defined as the region $\{(R_1, R_2) : R_1 \leq c(s_1; t_1), R_2 \leq c(s_2; t_2), R_1 + R_2 \leq c_{\mathsf{gns}}(s_1, s_2; t_1, t_2)\}$. For a given two-unicast network, the GNS sum-rate bound is a number while the GNS outer bound is a region. The following generalization of Theorem 2 may be proved similarly.

Theorem 3: Consider an n-unicast network $\mathcal{N} = (\mathcal{G}, \underline{\mathrm{C}})$. For non-empty $I \subseteq \{1, 2, \ldots, n\}$ and $S \subseteq \mathcal{E}(\mathcal{G})$, suppose there exists a bijection $\pi : I \mapsto \{1, 2, \ldots, |I|\}$ such that $\forall\ i, j \in I$, $\mathcal{G} \setminus S$ has no paths from source s_i to destination t_j whenever $\pi(i) \geq \pi(j)$. Then,

$$\sum_{i \in I} R_i \leq C(S)\ \forall (R_1, R_2, \ldots, R_n) \in \mathcal{C}(\mathcal{N}).$$

The GNS outer bound is a special case of the edge-cut bounds in [8]. However, we will show in Theorem 5 that it is the tightest possible outer bound resulting from edge-cut bounds for two-unicast networks and is thus, equivalent to the bound in [8]. The GNS outer bound is also a special case of the LP bound in [9], which is the tighest outer bound obtainable using Shannon information inequalities. In Section IV-C, we will show that the LP bound is in general tighter than the GNS outer bound for two-unicast networks. However, the GNS outer bound can be strictly better than the Network Sharing outer bound [1] as shown in Fig. 1.



(a) Grail with variable capacities. e_1, e_2 have unit capacity and all other edges have capacity 2 units.

(b) Network Sharing outer bound and GNS outer bound for network in (a)

Fig. 1. The GNS outer bound can be strictly better than the Network Sharing outer bound. $\{e_1, e_2\}$ is a GNS set.

IV. Properties of the GNS outer bound for two-unicast networks

A. Tightest outer bound resulting from edge-cut bounds

For an uncapacitated two-unicast network $\mathcal{G}$, an inequality of the form $\alpha_1 R_1 + \alpha_2 R_2 \leq C(S)$, with $\alpha_1, \alpha_2 \in \{0, 1\}, S \subseteq \mathcal{E}(\mathcal{G})$ is called an *edge-cut bound* if the inequality holds for all $(R_1, R_2) \in \mathcal{C}(\mathcal{G}, \underline{\mathrm{C}})$, for each choice of $\underline{\mathrm{C}}$. The cutset outer bound, the Network Sharing outer bound and the GNS outer bound are all outer bounds resulting from edge-cut bounds. Further, the Network Sharing outer bound is an improvement over the cutset bound and the GNS outer bound is an improvement over the Network Sharing outer bound. In Theorem 5, we show that it is impossible to improve on the GNS outer bound using edge-cut bounds for two-unicast networks. First, we will state and prove a useful result.

Theorem 4: (Two-Multicast Theorem) For a two-multicast network $\mathcal{N} = (\mathcal{G}, \underline{\mathrm{C}})$ with sources s_1 and s_2 multicasting independent messages at rates R_1 and R_2 respectively to both the destinations t_1 and t_2, (R_1, R_2) is an achievable rate pair if and only if

$$R_1 \leq \min\{c(s_1; t_1), c(s_1; t_2)\}$$
$$R_2 \leq \min\{c(s_2; t_1), c(s_2; t_2)\}$$
$$R_1 + R_2 \leq \min\{c(s_1, s_2; t_1), c(s_1, s_2; t_2)\}$$

Proof: The necessity of these conditions is obvious. For proving sufficiency, fix a rate pair (R_1, R_2) that satisfies these conditions and consider a new network $\tilde{\mathcal{N}}$ obtained by adding a super-source s with two outgoing edges to s_1 and s_2 with link capacities R_1 and R_2 respectively. We use the single source multicast result ([10], [11]) on $\tilde{\mathcal{N}}$ to infer the existence of a scheme for s multicasting at rate $R_1 + R_2$ to the destinations t_1 and t_2. This allows us to construct a two-multicast scheme in the original network $\mathcal{N}$ achieving desired rate pair. ∎

Theorem 5: Let $\mathcal{G}$ be an uncapacitated two-unicast network, and let $S \subseteq \mathcal{E}(\mathcal{G})$ such that $R_1 + R_2 \leq C(S)$ is an edge-cut bound. If S is not a GNS set, then $c(s_1; t_1) + c(s_2; t_2) \leq C(S)$ for all choice of $\underline{\mathrm{C}}$.

Remark: The cutset bounds provide all possible edge-cut bounds on the individual rates. Theorem 5 says that the GNS sets together provide all possible *non-trivial* edge-cut bounds on the sum rate.

Proof: Suppose $R_1 + R_2 \leq C(S)$ holds for all $(R_1, R_2) \in \mathcal{C}(\mathcal{G}, \underline{\mathrm{C}})$ for all choices of $\underline{\mathrm{C}}$. Then, it must be that S is an $s_i - t_i$ cut for $i = 1, 2$, so that $\mathcal{G} \setminus S$ has no paths from s_1 to t_1 or s_2 to t_2. If $\mathcal{G} \setminus S$ has no paths from s_1 to t_2 also, then S is a GNS set and the outer bound follows from Theorem 2. Likewise if $\mathcal{G} \setminus S$ has no paths from s_2 to t_1.

So, suppose that $\mathcal{G} \setminus S$ has no paths from s_1 to t_1 or s_2 to t_2 but it has paths from s_1 to t_2 and s_2 to t_1. Define $C_i(S) := \min\{C(T) : T \subseteq S,\ T \text{ is an } s_i - t_i \text{ cut}\}$ for $i = 1, 2$. Fix any choice of non-negative reals $\{c_e : e \in S\}$. Consider the following choice of link capacities: $C_e = c_e \ \forall e \in S$ and $C_e = \infty \ \forall e \notin S$. Note that for this choice of link capacities, $c(s_i; t_i) = C_i(S), i = 1, 2$. By Theorem 4, (R_1, R_2) is achievable for two-multicast from s_1, s_2 to t_1, t_2 if and only if $R_1 \leq C_1(S)$ and $R_2 \leq C_2(S)$, since $c(s_1, s_2; t_1) \geq c(s_2; t_1) = \infty, c(s_1, s_2; t_2) \geq c(s_1; t_2) = \infty$. Thus, $(C_1(S), C_2(S))$ is achievable for two-multicast and hence, also for two-unicast. Since $R_1 + R_2 \leq C(S)$ holds for all $(R_1, R_2) \in \mathcal{C}(\mathcal{G}, \underline{\mathrm{C}})$, we must have

$$C_1(S) + C_2(S) \leq C(S) \quad \forall\{c_e : e \in S\}. \qquad (1)$$

This is a purely graph theoretic property about the structure of the set of edges S relative to the uncapacitated network $\mathcal{G}$. Now, for an arbitrary assignment of link capacities $\underline{\mathrm{C}}$, we have by definition, $c(s_1; t_1) \leq C_1(S)$ and $c(s_2; t_2) \leq C_2(S)$. Using (1) gives us $c(s_1; t_1) + c(s_2; t_2) \leq C(S)$. ∎

B. Tightness in special cases

The next theorem shows that any minimal GNS set, i.e. a GNS set with no proper GNS subset, provides an outer bound that is not "obviously loose".

Theorem 6: For a given two-unicast graph $\mathcal{G}$, let $S \subseteq \mathcal{E}(\mathcal{G})$ be a minimal GNS set. Choose an arbitrary collection of non-negative reals $\{c_e : e \in S\}$. Consider the following link-capacity-vector $\underline{\mathrm{C}} : C_e = c_e \ \forall e \in S, \quad C_e = \infty \ \forall e \notin S$. Then, for the two-unicast network $(\mathcal{G}, \underline{\mathrm{C}})$, the GNS outer bound is identical to the capacity region $\mathcal{C}(\mathcal{G}, \underline{\mathrm{C}})$, i.e. the GNS outer bound is tight.

Remark: Theorem 6 does not say that a sum rate of $C(S)$ is achievable, only that all rate pairs in $\{(R_1, R_2) : R_1 \leq c(s_1; t_1), R_2 \leq c(s_2; t_2), R_1 + R_2 \leq c_{\mathsf{gns}}(s_1, s_2; t_1, t_2)\}$ are achievable. A sum rate of $C(S)$ is achievable only when $C(S) \leq c(s_1; t_1) + c(s_2; t_2)$ for the choice of capacities.

Proof: Define $C_i(S) := \min\{C(T) : T \subseteq S,\ T \text{ is an } s_i - t_i \text{ cut}\}$ for $i = 1, 2$ as before. As S is a minimal GNS set, the GNS outer bound for $(\mathcal{G}, \underline{\mathrm{C}})$ is given by

$$R_1 \leq C_1(S), R_2 \leq C_2(S), R_1 + R_2 \leq C(S). \qquad (2)$$

We will assume that c_e is an integer for each $e \in S$ and describe scalar linear coding schemes over the binary field $\mathbb{F}_2$ with block length $N = 1$ achieving the GNS outer bound. Having done this, it is easy to see that the theorem would also hold for choice of non-negative rational and thus, also non-negative real choice of $c_e, e \in S$. Henceforth, we will imagine a link of capacity c_e as having c_e unit capacity edges connected in parallel. This change could be made in the graph and in this proof, we will use $\mathcal{G}$ to denote the graph with all edges having unit capacity, possibly having multiple edges in parallel connecting two vertices.

Note that a given GNS set S is minimal if and only if for each $e \in S$, we have that $S \setminus e$ is not a GNS set. This allows us to partition the edges in S by their connectivity in $\mathcal{G} \setminus \{S \setminus e\}$ as $S_1^1 \cup S_1^2 \cup S_1^{12} \cup S_2^1 \cup S_2^2 \cup S_2^{12} \cup S_{12}^1 \cup S_{12}^2 \cup S_{12}^{12}$ where $e \in S$ lies in S_x^y if, in the graph $\mathcal{G} \setminus \{S \setminus e\}$, $\mathsf{tail}(e)$ is reachable only from source indices x and $\mathsf{head}(e)$ is capable of reaching only destination indices y. Eg. S_{12}^2 contains edge e in S if and only if in $\mathcal{G} \setminus \{S \setminus e\}$, we have that $\mathsf{tail}(e)$ is reachable from s_1, s_2 and $\mathsf{head}(e)$ can reach t_2, but cannot reach t_1.

Define $\hat{S}_1 := S_1^1 \cup S_1^{12} \cup S_{12}^1 \cup S_{12}^{12}$ and $\hat{S}_2 := S_2^2 \cup S_2^{12} \cup S_{12}^2 \cup S_{12}^{12}$. Thus, $\hat{S}_i$, for $i = 1, 2$ is the set of edges in S which have their tails reachable from s_i by a path of infinite capacity and their heads reaching t_i by a path of infinite capacity. We will show $C_i(S) = C(\hat{S}_i) + c_{\mathcal{G} \setminus \hat{S}_i}(s_i; t_i)$, for $i = 1, 2$. By the Max Flow Min Cut Theorem, there exists a flow of value $C_i(S)$ from s_i to t_i

in $\mathcal{G}$. At most $C(\hat{S}_i)$ of the flow goes through edges in $\hat{S}_i$. Thus, there exists a flow of value at least $C_i(S)-C(\hat{S}_i)$ in $\mathcal{G}\setminus\hat{S}_i$. So, $c_{\mathcal{G}\setminus\hat{S}_i}(s_i;t_i) \geq C_i(S)-C(\hat{S}_i)$. Now, consider $T_i \subseteq S$ in $\mathcal{G}$ such that T_i is an $s_i - t_i$ cut and $C(T_i) = C_i(S)$. Then, since $\hat{S}_i \subseteq T_i$, we have that $T_i \setminus \hat{S}_i$ is an $s_i - t_i$ cut in $\mathcal{G}\setminus\hat{S}_i$. Thus, $c_{\mathcal{G}\setminus\hat{S}_i}(s_i;t_i) \leq C(T_i\setminus\hat{S}_i) = C(T_i) - C(\hat{S}_i) = C_i(S) - C(\hat{S}_i)$.

Case I: S is a minimal GNS set such that $\mathcal{G}\setminus S$ has no paths from either of s_1, s_2 to t_1, t_2. In this case, $S_1^2, S_2^1 = \emptyset$ by minimality of S. Thus, $C_1(S)+C_2(S) \geq C(\hat{S}_1)+C(\hat{S}_2) = C(S)+C(S_{12}^{12}) \geq C(S)$. So, in this case, the GNS outer bound (2) is a pentagonal region and we have to show achievability of the two corner points $(C_1(S), C(S)-C_1(S))$ and $(C(S)-C_2(S), C_2(S))$.

Consider the following scheme. Edges in $S_1^1, S_1^{12}, S_{12}^1, S_{12}^{12}$ forward s_1's message bits to t_1 and edges in $S_2^2, S_{12}^2, S_2^{12}$ forward s_2's message bits to t_2. This achieves

$$R_1 = C(\hat{S}_1) = C(S_1^1) + C(S_1^{12}) + C(S_{12}^1) + C(S_{12}^{12}),$$
$$R_2 = C(S_2^2) + C(S_{12}^2) + C(S_2^{12}).$$

Note that we have $R_1 + R_2 = C(S)$ for this rate pair. Now, we will increase R_1 up to $C_1(S)$ while preserving this sum rate. Construct $c_{\mathcal{G}\setminus\hat{S}_1}(s_1;t_1)$ unit capacity "edge-disjoint" paths from s_1 to t_1 in $\mathcal{G}\setminus\hat{S}_1$. This gives us $c_{\mathcal{G}\setminus\hat{S}_1}(s_1;t_1)$ paths in $\mathcal{G}$ such that none of them use any edge in $\hat{S}_1$. Any such path encounters a first finite capacity edge from S_{12}^2 and a last finite capacity edge from S_2^{12}. The intermediate finite capacity edges may be assumed to lie in S_2^2 only. If intermediate finite capacity edges lie in S_{12}^2 or S_2^{12}, we can modify the path so that this is not the case, while preserving the edge-disjointness property. Now, a simple XOR coding scheme as shown in Fig. 2(a) improves R_1 by one bit and reduces R_2 by one bit as s_2 has to set $b_1 \oplus b_2 \oplus b_3 = 0$ to allow t_1 to decode a. In the general case, we have an arbitrary number of finite capacity edges from S_2^2 along the path, for which we perform a similar XOR scheme. When this is carried out for each of the $c_{\mathcal{G}\setminus\hat{S}_1}(s_1;t_1)$ paths, we have a scheme achieving $(C_1(S), C(S)-C_1(S))$. Similarly, the other corner point $(C(S)-C_2(S), C_2(S))$ may be shown to be achievable.

Case II: S is a minimal GNS set such that $\mathcal{G}\setminus S$ has no paths from s_1 to t_1, s_2 to t_2, or s_2 to t_1 but it has paths from s_1 to t_2. As S is a minimal GNS set, we have $S_1^2 = \emptyset$. In this case, the GNS outer bound (2) is not necessarily a pentagonal region. We first show achievability of the rate pair $R_1 = C_1(S), R_2 = \min\{C_2(S), C(S)-C_1(S)\}$.

Stage I - Basic Scheme: It is easy to see that we can achieve the rate pair given by

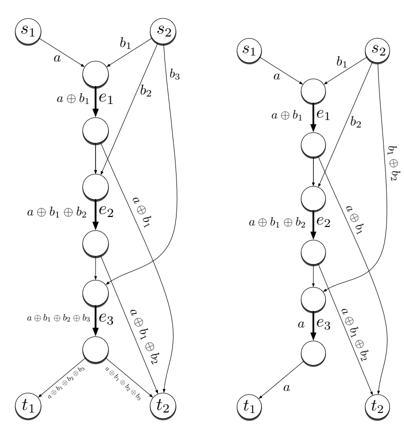
$$R_1 = C(\hat{S}_1) = C(S_1^1) + C(S_{12}^1) + C(S_1^{12}) + C(S_{12}^{12}),$$
$$R_2 = C(S_2^2) + C(S_2^{12}) + C(S_{12}^2) + \min\{C(S_2^1), C(S_{12}^{12})\},$$

by a routing + butterfly coding approach as follows.

- Edges in $S_1^1, S_1^{12}, S_{12}^1$ forward s_1's message bits to t_1 and edges in $S_2^2, S_2^{12}, S_{12}^2$ forward s_2's message bits to t_2.
- Edges in S_{12}^{12} and S_2^1 along with an infinite capacity path from s_1 to t_2 perform "preferential routing for s_1 with butterfly coding for s_2," i.e.
  - if $C(S_2^1) < C(S_{12}^{12})$, then an amount of $C(S_{12}^{12})-C(S_2^1)$ of the capacity of edges in S_{12}^{12} is used for routing s_1's message bits, while the rest is used for butterfly coding, i.e. an XOR operation is performed over $C(S_2^1)$ bits from source s_1 with $C(S_2^1)$ bits from source s_2 to be transmitted over the edges in S_{12}^{12}. Edges in S_2^1 provide $C(S_2^1)$ bits of side-information from s_2 to t_1, while the infinite capacity path from s_1 to t_2 provides side-information to t_2.
  - if $C(S_2^1) \geq C(S_{12}^{12})$, then all of the capacity of edges in S_{12}^{12} is used for butterfly coding.

Stage II - Improving R_1 up to $C_1(S)$: We know $c_{\mathcal{G}\setminus\hat{S}_1}(s_1;t_1) = C_1(S) - C(\hat{S}_1)$. Find $c_{\mathcal{G}\setminus\hat{S}_1}(s_1;t_1)$ unit capacity "edge-disjoint" paths from s_1 to t_1 in $\mathcal{G}$ such that none of them use any edge in $\hat{S}_1$. Each such unit capacity path from s_1 to t_1 in $\mathcal{G}$ starts with a first finite capacity edge in S_{12}^2, ends with the last finite capacity edge in S_2^{12} or S_2^1 and with all intermediate edges lying, without loss of generality, in S_2^2. Whenever the capacity of all edges in S_2^1 is used up, we would have reached a sum rate of $C(S)$, as all edges are carrying independent linear combinations of message bits. In that case, we will increase R_1 by one bit and reduce R_2 by one bit. Else, we will increase R_1 by one bit while not altering R_2.

- If the last finite capacity edge lies in S_2^{12}, perform coding as in Fig. 2(a). If the capacity of S_2^1 edges is not fully used, use free unit capacity of some edge $e \in S_2^1$ to relay the XOR value of $b_1 \oplus b_2 \oplus b_3$ from s_2 to t_1. Use the infinite capacity path from s_1 to t_2 to send the symbol a. If there is no free edge in S_2^1, then s_2 sets $b_1 \oplus b_2 \oplus b_3 = 0$. This increases R_1 by one bit and reduces R_2 by one bit.



(a) Coding Performed in Case I. Also used in Case II, Stage II - Last finite capacity edge in S_2^{12}

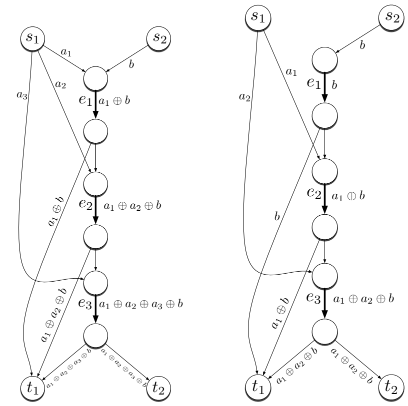
(b) Case II, Stage II - Last finite capacity edge in S_2^1

Fig. 2. Improving R_1 up to $C_1(S)$

- Suppose the last finite capacity edge, call it e_3, lies in S_2^1. Suppose there is a free edge $e \in S_2^1$. If e_3 is being used, it must be used as a conduit for side-information to t_1, as part of the butterfly coding. Use e to relay that side-information to t_1. So, we can assume e_3 is free. Now, perform coding as in Fig. 2(b). Use the infinite capacity path from s_1 to t_2 to relay the symbol a. This improves R_1 by one bit while R_2 remains unchanged. If there is no free edge in S_2^1, then we must have achieved a sum rate of $C(S)$. Edge e_3 now relays a to t_1 improving R_1 by one bit. However, the edge e_3 must have been assisting in butterfly coding using some edge in S_{12}^{12} and the infinite capacity $s_1 - t_2$ path. Now, the edge e_3 can no longer provide side-information to t_1. So, the corresponding unit capacity in some edge in S_{12}^{12} now performs routing of s_1's message bit as opposed to XOR mixing of one bit of s_1's message and one bit of s_2's message. This reduces R_2 by one bit.

Stage III - Improving R_2 **up to** $\min\{C(S) - C_1(S), C_2(S)\}$: If the capacity of S_2^1 edges is all used up, we have achieved a sum rate of $R_1 + R_2 = C(S)$ and so, $R_2 = C(S) - C_1(S)$. If not, we have $R_1 = C_1(S), R_2 = C(\hat{S}_2)$. We have $C_2(S) = C(\hat{S}_2) + c_{\mathcal{G}\setminus\hat{S}_2}(s_2; t_2)$. Similar to before, we find $c_{\mathcal{G}\setminus\hat{S}_2}(s_2; t_2)$ unit capacity "edge-disjoint" paths from s_2 to t_2 in $\mathcal{G}$ such that the paths don't use any edge in $\hat{S}_2$. Each such unit capacity path encounters a first finite capacity edge from S_{12}^1 or S_2^1 and a last finite capacity edge from S_1^{12} while all intermediate finite capacity edges may be assumed to lie in S_1^1. Note that edges in $S_1^1, S_1^{12}, S_{12}^1$ are all performing pure routing of s_1's message. At any point, if the capacity of S_2^1 edges is fully used, we have reached $R_1 = C_1(S), R_2 = C(S) - C_1(S)$. If the capacity is not fully used, perform the modification as described below.

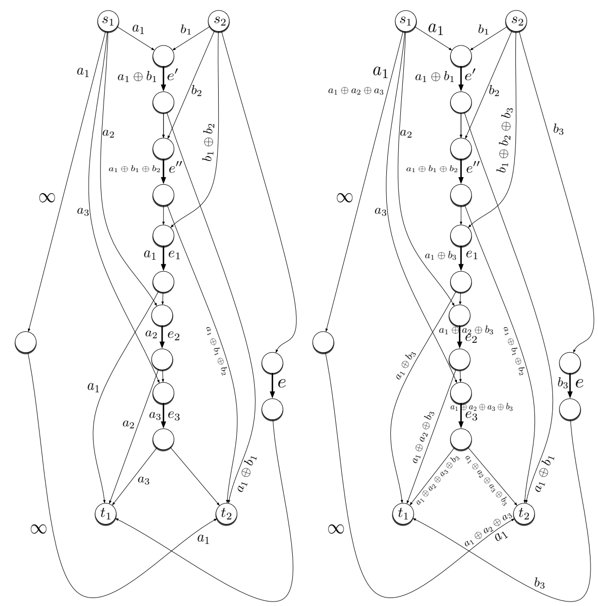
- If the first finite capacity edge lies in S_{12}^1, perform coding as in Fig. 3(a). Use unit capacity of a free edge in S_2^1 to relay symbol b from s_2 to t_1 and use the s_1 to t_2 infinite capacity path to send the XOR value of $a_1 \oplus a_2 \oplus a_3$ to t_2. This leaves R_1 unaffected and improves R_2 by one bit.
- Suppose the first finite capacity edge, call it e_1, lies in S_2^1. If e_1 is not being used, perform coding as in Fig. 3(b). Use unit capacity of edge $e_1 \in S_2^1$ to send a symbol b from s_2 to t_1. The infinite capacity s_1 to t_2 path is used to send $a_1 \oplus a_2$ from s_1 to t_2. This allows t_2 to decode b and improves R_2 by one bit while leaving R_1 unaffected. If e_1 is being used for sending side-information to t_1 as part of the butterfly coding, then pick some free edge $e \in S_2^1$ for the transfer of side-information freeing up e_1 and allowing us to use the coding described in Fig. 3(b). If e_1 is being used but not for sending side-information as part of butterfly coding, it must have gotten used in Stage II as a means of improving R_1. In this case, we use some free edge $e \in S_2^1$ and superimpose scheme shown in Fig. 3(b) with already existing scheme Fig. 2(b). This modification is shown via Fig. 4(a) and Fig. 4(b). This improves R_2 by one bit while R_1 remains unchanged.



(a) Case II, Stage III - First finite capacity edge in S_{12}^1

(b) Case II, Stage III - First finite capacity edge in S_2^1

Fig. 3. Improving R_2 up to $\min\{C(S) - C_1(S), C_2(S)\}$



(a) Case II, Stage III - e', e'', e_1 are being used in Stage II. e_2, e_3 serve to route s_1's bits to t_1.

(b) Case II, Stage III - Chosen s_2-t_2 path uses edges e_1, e_2, e_3. Modified scheme uses some free edge $e \in S_2^1$.

Fig. 4. Improving R_2 up to $\min\{C(S) - C_1(S), C_2(S)\}$ in the case when e_1 was already being used in Stage II.

This stage terminates achieving $R_1 = C_1(S), R_2 = \min\{C_2(S), C(S) - C_1(S)\}$. Because the GNS set is not symmetric in indices 1 and 2, we also have to show achievability of the rate pair $R_1 = \min\{C_1(S), C(S) - C_2(S)\}, R_2 = C_2(S)$. This can be shown similarly.

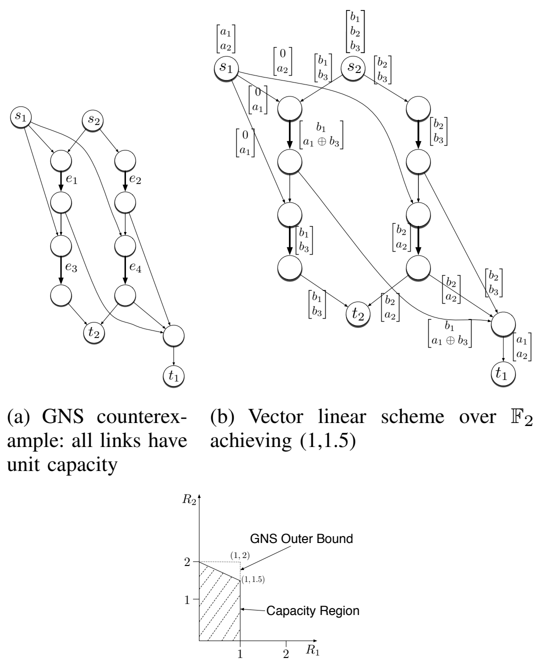
Case III: S is a minimal GNS set such that $\mathcal{G} \setminus S$ has no paths from s_1 to t_1, s_2 to t_2, or s_1 to t_2 but it has paths from s_2 to t_1. This case is identical to Case II. ∎

C. GNS outer bound is not tight

We now provide an example of a two-unicast network in Fig. 5(a) showing that,

- the GNS outer bound is not tight, so edge-cut bounds do not suffice to characterize the capacity region,
- the trade-off between rates on the boundary of the capacity region need not be 1:1,
- the capacity region may have a non-integral corner point even if all links have integer capacity and thus,
- scalar linear coding is not sufficient to achieve capacity.

Fig. 5(b) shows a two time step vector linear coding scheme over $\mathbb{F}_2$ that achieves $(1, 1.5)$.



(a) GNS counterexample: all links have unit capacity

(b) Vector linear scheme over $\mathbb{F}_2$ achieving (1,1.5)

(c) Capacity region

Fig. 5. Counterexample to tightness of the GNS outer bound

We will prove the inequality $R_1 + 2R_2 \leq 4$ for any rate pair (R_1, R_2) in the capacity region of this network. Consider a scheme of block length N over alphabet $\mathcal{A}$ achieving the rate pair (R_1, R_2). Let W_1, W_2 be independent and distributed uniformly over the sets $\mathcal{A}^{\lceil NR_1 \rceil}$ and $\mathcal{A}^{\lceil NR_2 \rceil}$ respectively. For edge $e = e_1, e_2, e_3, e_4$, define X_e as the concatenated evaluation of the functions specified by the scheme for edge e.

$$
\begin{aligned}
&H(W_1) \\
&= I(X_{e_1}, X_{e_2}, X_{e_4}; W_1) + H(W_1 | X_{e_1}, X_{e_2}, X_{e_4}) && (3)\\
&= I(X_{e_1}, X_{e_2}; W_1) + I(X_{e_4}; W_1 | X_{e_1}, X_{e_2}) + 0 && (4)\\
&I(X_{e_1}, X_{e_2}; W_1) \\
&= I(X_{e_1}, X_{e_2}; W_1, W_2) - I(X_{e_1}, X_{e_2}; W_2 | W_1) && (5)\\
&= H(X_{e_1}, X_{e_2}) - H(W_2 | W_1) + H(W_2 | W_1, X_{e_1}, X_{e_2}) && (6)\\
&= H(X_{e_1}, X_{e_2}) - H(W_2) + 0 && (7)\\
&I(X_{e_4}; W_1 | X_{e_1}, X_{e_2}) \\
&= I(X_{e_4}; W_1, X_{e_1}, X_{e_2}) - I(X_{e_4}; X_{e_1}, X_{e_2}) && (8)\\
&\leq H(X_{e_4}) - I(X_{e_4}; W_2) && (9)\\
&= H(X_{e_4}) - I(X_{e_3}, X_{e_4}; W_2) + I(X_{e_3}; W_2 | X_{e_4}) && (10)\\
&\leq H(X_{e_4}) - H(W_2) + H(X_{e_3} | X_{e_4}) && (11)\\
&= H(X_{e_3}, X_{e_4}) - H(W_2) && (12)
\end{aligned}
$$

(4) follows from $\{e_1, e_2, e_4\}$ being an $s_1, s_2 - t_1$ cut, (7) follows from $\{e_1, e_2\}$ being an $s_2 - t_2$ cut.

Thus, we have $N \cdot \log |\mathcal{A}| \cdot (R_1 + 2R_2) \leq H(W_1) + 2H(W_2) \leq H(X_{e_1}, X_{e_2}) + H(X_{e_3}, X_{e_4}) \leq 4N \cdot \log |\mathcal{A}|$. Thus, the network has a capacity region as shown in Fig. 5(c).

V. Discussion

Let $\mathcal{C}_{\mathsf{LP}}$ denote the LP bound in [9] and $\mathcal{C}_{\mathsf{GNS}}$ denote the GNS outer bound for the n-unicast problem that can be obtained from Theorem 3. We have

$$\mathcal{C}_{\mathsf{scalar}} \subseteq \mathcal{C}_{\mathsf{vector}} \subseteq \mathcal{C} \subseteq \mathcal{C}_{\mathsf{LP}} \subseteq \mathcal{C}_{\mathsf{GNS}}.$$

[3] shows that $\mathcal{C}_{\mathsf{vector}} \subsetneq \mathcal{C}$ and [4] shows $\mathcal{C} \subsetneq \mathcal{C}_{\mathsf{LP}}$ for general n-unicast networks. The network in Fig. 5(a) shows that for two-unicast networks, $\mathcal{C}_{\mathsf{scalar}} \subsetneq \mathcal{C}_{\mathsf{vector}}$ and $\mathcal{C}_{\mathsf{LP}} \subsetneq \mathcal{C}_{\mathsf{GNS}}$ in general. It would be interesting to know whether or not

- $\mathcal{C}_{\mathsf{vector}} \subsetneq \mathcal{C}$
- $\mathcal{C} \subsetneq \mathcal{C}_{\mathsf{LP}}$

for a general two-unicast network.

Acknowledgment

Research support from the ARO MURI grant W911NF-08-1-0233, "Tools for the Analysis and Design of Complex Multi-Scale Networks", from the NSF grant CNS-0910702, from the NSF Science & Technology Center grant CCF-0939370, "Science of Information", from Marvell Semiconductor Inc., and from the U.C. Discovery program is gratefully acknowledged.

References

[1] X. Yan, J. Yang, and Z. Zhang, "An outer bound for multisource multisink network coding with minimum cost consideration", *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2373–2385, June 2006.

[2] R. Dougherty, C. Freiling, and K. Zeger, "Linear network codes and systems of polynomial equations", *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2303–2316, May 2008.

[3] R. Dougherty, C. Freiling, and K. Zeger, "Insufficiency of linear coding in network information flow", *IEEE Transactions on Information Theory*, vol. 51, no. 8, pp. 2745–2759, August 2005.

[4] R. Dougherty, C. Freiling, and K. Zeger, "Networks, matroids and non-shannon information inequalities", *IEEE Transactions on Information Theory*, vol. 53, no. 6, pp. 1949–1969, June 2007.

[5] T. Chan and A. Grant, "Mission impossible: Computing the network coding capacity region", in *Proc. of IEEE ISIT*, Toronto, Canada, July 2008.

[6] Chih-Chun Wang and Ness B. Shroff, "Pairwise intersession network coding on directed networks", *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3879–3900, August 2010.

[7] Kai Cai, K.B. Letaief, Pingyi Fan, and Rongquan Feng, "On the solvability of 2-pair unicast networks: A cut-based characterization", *arXiv:1007.0465 [cs.IT]*, July 2010.

[8] G. Kramer and S. Savari, "Edge-cut bounds on network coding rates", *Journal of Network and Systems Management*, vol. 14, no. 1, pp. 49–67, March 2006.

[9] R.W. Yeung and Z. Zhang, "Distributed source coding for satellite communications", *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1111–1120, May 1999.

[10] R. Ahlswede, N. Cai, S.-Y.R. Li, and R.W. Yeung, "Network information flow", *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.

[11] R. Koetter and M. Médard, "An algebraic approach to network coding", *IEEE/ACM Transactions on Networking*, 2003.