On the Generalized Network Sharing bound and edge-cut bounds for network coding

Sudeep Kamath and David N.C. Tse
Dept. of EECS
University of California Berkeley
Email: sudeep,dtse@eecs.berkeley.edu

Abstract—We consider sum-rate edge-cut bounds on network coding rates for the multiple unicast problem. We first show that the Generalized Network Sharing (GNS) bound is equivalent to a functional dependence bound in the literature. After defining a notion of profile of an edge-cut, we show that the only profiles for which, every edge-cut with the said profile leads to a fundamental bound on network coding rates, are the so-called GNS profiles and further, we quantify with a tight constant factor, the amount by which network coding can potentially beat edge-cuts associated with other profiles. Finally, we show that the problem of computing the GNS bound is NP-complete, even for two-unicast networks.

I. INTRODUCTION

Network Coding deals with the study of capacity regions of the simplest class of networks - wireline networks - where links between nodes of the network are noise-free and orthogonal. Rather than routing information as a commodity, having nodes perform coding operations has significant potential to improve rates [1], [2], [3]. However, recent results suggest that characterizing capacity of a network is a hard problem - so-called Non-Shannon inequalities are important for characterizing capacity [4] and linear coding schemes do not achieve capacity in general [5].

Nonetheless, it may often be useful to give guarantees on how far the performance of some coding scheme is from capacity. For this purpose, it is useful to develop outer bounds on the capacity region. The simplest and oldest such bound is the cutset bound [6], [7]. This bound however, is often quite loose. The tightest known explicit bound is the so-called LP bound [8] that harnesses the full power of the basic or so-called Shannon inequalities. It is however, computationally intractable since the linear program has size exponential in that of the network.

The literature also has bounds derived from the graph-theoretic structure of the network. These edge-cut bounds have conventionally served as outer bounds to commodity flow problems. Indeed, such commodity flow bounds derived from edge-cuts are not in general, fundamental, i.e. they can potentially be beaten by network coding [9]. It is of interest to study these edge-cut bounds because they tend to be simpler and more intuitive than the LP bound while also being tighter than the cutset bound. Different works have studied what makes edge-cuts fundamental. [10] proposed the Network Sharing bound which was subsequently improved to the Generalized Network Sharing (GNS) bound in [11], [9], [12] study bounds derived from functional dependence graphs and [3] studies bounds derived from information dominance. Recently, there has also been some progress in studying weighted sum-rate edge-cut bounds. [13] exhibits such bounds for multimeessage multicast problems while [14] produces bounds for multiple unicast problems based on a class of Shannon inequalities and a knowledge of some graph structure.

In this paper, we focus on sum-rate edge-cut bounds for multiple-unicast networks. We show that for multiple unicast networks, the GNS bound is equivalent to the more sophisticated functional dependence bound derived from functional dependence graphs [12]. Next, we study edge-cut bounds on network coding capacity based purely on what we define as the ‘profile’ of the edge-cut, which is simply the knowledge of the residual connectivity between sources and destinations after a set of edges have been removed from the graph. We show that the only edge-cut profiles for which, every edge-cut with the said profile always leads to a fundamental bound on network coding rates, are the profiles of GNS-cuts. Furthermore, we provide the tight constant associated with every edge-cut profile up to which network coding may potentially outperform the associated edge-cut. Finally, we consider the problem of computation of the GNS bound. We show that this problem is NP-complete, even for two-unicast networks.

The rest of the paper is organized as follows. We describe notation and preliminaries in Section II. We show the equivalence between the GNS bound and the functional dependence bound of [12] in Section III. We study bounds from edge-cuts based purely on source-destination connectivity in Section IV. We prove the NP-completeness of the GNS-cut in Section V. Finally, we conclude with some future directions in Section VI.

II. PRELIMINARIES

We briefly describe some notation that will be used throughout the rest of the paper.

Definition 1: A $k$-unicast uncaptacitated network $\mathcal{G}$ is a directed graph with vertex set $V(\mathcal{G})$, edge set $E(\mathcal{G})$ and containing source-destination pairs $\{(s_i; d_i)\}_{i = 1}^k$. For each $i \in \{1, 2, \ldots, k\}$, $s_i$ has independent information to be communicated to $d_i$ at rate $R_i$.

Definition 2: The uncaptacitated network $\mathcal{G}$ can be converted to a capacitated network by assigning non-negative capacities $C_e := (C_e : e \in E(\mathcal{G})) \in \mathbb{R}_{\geq 0}^{|E(\mathcal{G})|}$ to the edges of $\mathcal{G}$.
Definition 3: Given an uncapacitated network \( G \) and a fixed assignment of capacities \( C_G \), the set of rate tuples achievable by routing solutions is called the commodity flow region denoted by \( F(G,C_G) \) and the closure of the set of rate tuples achievable (upto vanishing error) by network coding is called the capacity region denoted by \( C(G,C_G) \).

Edge-cut bounds have traditionally been studied in context of commodity flow problems since they are simple outer bounds on the commodity flow region. To define edge-cut bounds, fix a \( k \)-unicast uncapacitated network \( G \), and an edge set \( E \subseteq E(G) \). Define the disconnectivity of the edge-cut derived from \( E \), denoted \( D_{G,E} \), to be the subset of \( \{1,2,\ldots,k\} \) where index \( j \in D_{G,E} \) if and only if there is no path from \( s_j \) to \( d_j \) in \( G \setminus E \). Consider Statements 1 and 2 below.

Statement 1: For any assignment of capacities \( C_G \in [0,\infty)^E \), and any rate tuple \( (R_1,R_2,\ldots,R_k) \in F(G,C_G) \),
\[
\sum_{j \in D_{G,E}} R_j \leq \sum_{e \in E} C_e. \tag{1}
\]

Statement 2: For any assignment of capacities \( C_G \in [0,\infty)^E \), and any rate tuple \( (R_1,R_2,\ldots,R_k) \in C(G,C_G) \),
\[
\sum_{j \in D_{G,E}} R_j \leq \sum_{e \in E} C_e. \tag{2}
\]

Inequality (1) is the edge-cut bound derived from \( E \). Statement 1 can be seen to be obviously true. However, Statement 2 is not true in general, as evidenced by the butterfly counterexample in Fig. 1.

![Fig. 1. (a) shows a butterfly network example with all edges having unit capacity. \( E = \{e\} \) yields edge-cut bound \( R_1 + R_2 \leq 1 \) which is violated by \( R_1 = R_2 = 1 \) achieved by coding scheme in (b).](image)

We say that the edge-cut derived from \( E \) is fundamental if Statement 2 holds, i.e. if Inequality (2) holds for all capacity assignments and all rate tuples in the corresponding capacity region. We shall be interested in what kinds of edge-cuts are fundamental. We define the Generalized Network Sharing (GNS) bound here since it is the focus of this paper.

Definition 4: Given a \( k \)-unicast uncapacitated network \( G \), an edge-cut derived from \( E \subseteq E(G) \) is called a GNS-cut for the \( k \)-unicast network if there exists a permutation \( \pi : \{1,2,\ldots,k\} \mapsto \{1,2,\ldots,k\} \) such that for any \( i,j \) there are no paths from \( s_i \) to \( d_j \) in \( G \setminus E \) whenever \( \pi(i) \geq \pi(j) \).

Note that for \( k = 1 \), GNS-cuts are just vertex bipartition cuts that feature in the cutset bound. The usefulness of GNS-cuts arises from the following theorem.

Theorem 1: (from [11]) GNS-cuts are fundamental.

Given a \( k \)-unicast uncapacitated network \( G \), one can set out to build a collection of fundamental edge-cut bounds as follows:

- Fix a non-empty subset \( D \subseteq \{1,2,\ldots,k\} \) and consider the \( |D| \)-unicast network where for each \( j \in D \), \( s_j \) communicates to \( d_j \) at rate \( R_j \).
- For every edge cut derived from a set of edges \( E \subseteq E(G) \) that forms a GNS-cut for the \( |D| \)-unicast problem, we include the edge-cut bound \( \sum_{j \in D} R_j \leq \sum_{e \in E} C_e \). (It’s easy to see that these are also going to be fundamental bounds for the original \( k \)-unicast problem.)
- Repeat for all choices of non-empty subsets \( D \).

This collection of fundamental edge-cut bounds will be called the GNS edge-cut bound collection for \( G \).

Through most of this paper, we shall only consider complete edge-cuts, namely edge-cuts which disconnect all sources from their respective destinations so that \( D_{G,E} = \{1,2,\ldots,k\} \). This is without loss of generality, since if we have a non-complete but fundamental edge-cut, one only needs to prove the necessary bound by considering a complete edge-cut for a suitable \( |D| \)-unicast problem with \( D = D_{G,E} \subseteq \{1,2,\ldots,k\} \).

III. EQUIVALENCE TO THE FUNCTIONAL DEPENDENCE BOUND

The problem of identifying edge-cut bounds that are fundamental has been approached using different techniques. These include the following:

- PdE bound [9]
- Information Dominance bound [3]
- Functional Dependence bound [12]

We show that the GNS bound is equivalent to the functional dependence bound [12]. Connection of the GNS bound to the PdE bound [9] and the information dominance bound [3] will be explored in a future work.

Theorem 2: For multiple-unicast networks, the GNS edge-cut bound collection is equivalent to the functional dependence bound [12].

Proof: It is easy to check that the GNS bound is a special case of the functional dependence bound [12]. Now, given a \( k \)-unicast uncapacitated network \( G \), the functional dependence bound [12] says that the inequality
\[
\sum_{i=1}^k R_i \leq \sum_{e \in E} C_e \tag{3}
\]
holds for all capacity assignments \( C_G \) and all \( (R_1,R_2,\ldots,R_k) \in C(G,C_G) \), for a set of edges \( E \) that correspond to a so-called maximal irreducible set (defined below). We will show that such a set of edges always yields a GNS-cut and that will complete the proof.
We describe the construction from [12] of the functional dependence graph (FDG) denoted by $Z$. Corresponding to the information message of source $s_i$, introduce a (so-called) pseudo-variable $Y_i$ and corresponding to each edge $e$, introduce a pseudo-variable $U_e$. For each $e$, draw incoming edges in to $U_e$ from each of the pseudo-variables associated with all incoming sources and edges incident on the tail of $e$. For each destination $d_i$, draw incoming edges in to $Y_i$ from each of the pseudo-variables associated with all incoming edges and sources incident on $d_i$. This completes the construction of $Z$. In the network $G$, each source $s_i$ must have at least one path to its own destination $d_i$. So, we have that the FDG $Z$ is cyclic (in notation of [12]). A maximal irreducible set is a subset of vertices $A$ of the FDG $Z$ with the property that after one removes all edges outgoing from vertices in $A$ and successively removes all vertices and edges with no incoming edges and vertices respectively, then no vertex in $Z$ remains. It must also be that no proper subset of $A$ has the same property but we will not need this latter condition.

We start with a maximal irreducible set $A$ that has none of the source variables $Y_1$, say $A = \{U_e : e \in E\}$. Consider vertices and edges being removed by this procedure one at a time. Since the process ends with all the $Y_i$’s removed, let the order in which they get removed be given by a permutation $\pi$, i.e. let the order be $Y_{\pi(1)}, Y_{\pi(2)}, \ldots, Y_{\pi(k)}$. Then, none of the sources $s_1, s_2, \ldots, s_k$ have a path to $d_{\pi(1)}$ in $G \setminus E$. Further, none of the sources with the possible exception of $s_{\pi(1)}$ can have a path to $d_{\pi(2)}$ in $G \setminus E$. Continuing this chain of reasoning, we find that the edge-cut derived from $E$ is a GNS-cut for the network $G$ with permutation $\pi$.

IV. EDGE-CUT BOUNDS BASED ONLY ON SOURCE-DESTINATION CONNECTIVITY

Fundamentality of an edge-cut bound is a purely graph-theoretic property. A simple way to classify different edge-cuts is to look at connectivity from all sources to all destinations. For a $k$-unicast uncapacitated network $G$, and a subset of edges $E \subseteq \mathcal{E}(G)$ which yield a complete edge-cut, we define the profile of the edge-cut derived from $E$, denoted $\mathcal{P}_{G,E}$, to be a directed graph with nodes having labels $s_1, s_2, \ldots, s_k, d_1, d_2, \ldots, d_k$ with $s_i$’s having only outgoing edges, $d_i$’s having only incoming edges and an edge from $s_i$ to $d_j$ if and only if there is a path from $s_i$ to $d_j$ in $G \setminus E$. If the edge-cut derived from $E$ is a GNS-cut, then we call the corresponding profile a GNS profile. Fig. 2 shows all possible profiles of complete edge-cuts for a 2-unicast network.

From Theorem 1, all edge-cuts with a GNS profile result in fundamental bounds. A natural question to ask is whether there are other edge-cut profiles for which it is also true that all edge-cuts with that profile result in fundamental bounds. Furthermore, it is of interest to provide some bounds in the case of an edge-cut profile that does not necessarily give fundamental bounds for all networks. Both these issues are addressed by Theorem 3. As an example, the profile in Fig. 2(d) is a non-GNS profile and this profile happens to not give fundamental bounds in all networks as seen by the example in Fig. 1.

To state the main result of this section, we need one more definition. Given a profile $P$ of a complete edge-cut for a $k$-unicast network, we define a specific capacitated network - its canonical network $N(P)$ - an index coding instance, as follows. Take the directed graph represented by the profile $P$ and add two nodes $u$ and $v$. Add edges from all the $s_i$’s to $u$, from $v$ to all the $d_i$’s and from $u$ to $v$. All edges have infinite capacity except the edge from $u$ to $v$ which has capacity 1 unit. For each $i$, $s_i$ has independent information to be communicated to $d_i$. Let the sum-capacity of this network be denoted by $\rho(P)$. Fig. 3 shows two examples of profiles of edge-cuts and their corresponding canonical networks.

Theorem 3: Fix an edge-cut profile $P$. For any $k$-unicast uncapacitated network $G$, and any complete edge-cut derived from edge set $E \subseteq \mathcal{E}(G)$ with $\mathcal{P}_{G,E} = P$, we have the inequality

$$\sum_{j=1}^{k} R_j \leq \rho(P) \sum_{e \in E} C_e,$$  \hspace{1cm} (4)$$

for any assignment of capacities $C_e \in \mathbb{R}_{\geq 0}$, and any rate tuple within capacity, $(R_1, R_2, \ldots, R_k) \in \mathcal{C}(G, C_G)$. Moreover, the constant $\rho(P)$ in Inequality (4) cannot be improved upon and satisfies $\rho(P) \geq 1$, with equality if and only if $P$ is a GNS-profile.

Proof: Suppose we have a $k$-unicast capacitated network $(G, C_G)$, and a complete edge-cut derived from edge set $E$, whose profile is $P$. We will perform modifications to the network and its capacities which can only enhance its capacity region. For each directed edge $(x, y)$ in $E$, add an edge from each of the sources to $x$ and from $y$ to each of the destinations.
the sum capacity of this enhanced network is $N$ which one can set example, Fig. 3(a), (b) are both 3-unicast non-GNS profiles for combinations of these symbols on the edge $(u,v)$, where

$$
\rho(N) = 3.
$$

Now, assign infinite capacities to all edges of this network that do not belong to $E$.

Now all source messages can be assumed to be present in their entirety at the tails of each edge in $E$ and all destinations are connected with an infinite capacity path to the heads of each edge in $E$, and therefore, any coding scheme operating on this network can be translated to a coding scheme on a $(\sum_{e \in E} C_e)$-scaled copy of $N(P)$ and vice versa. Therefore, the sum capacity of this enhanced network is $\rho(P)\sum_{e \in E} C_e$ and so, for any rate tuple $(R_1, R_2, \ldots, R_k) \in C(G, C_e)$, we have the desired inequality (4). The constant $\rho(P)$ cannot be improved upon since $N(P)$ is an example of a network with edge-cut derived from $E = \{u,v\}$ having the desired profile $P$ and for which Inequality (4) is tight by the definition of $\rho(P)$.

$\rho(P) \geq 1$ is obvious from the definition since commodity flow can achieve a sum-rate of 1 in $N(P)$. For a GNS profile $P$, Theorem 1 gives $\rho(P) \leq 1$. We only need to show that $\rho(P) > 1$ for any non-GNS profile $P$. It is easy to show that for any non-GNS profile $P$, one can find a sequence of $t \geq 2$ distinct indices $i_1, i_2, \ldots, i_t \in \{1, 2, \ldots, k\}$ such that in the directed graph represented by $P$, we have that $s_{i_t}$ has an edge to $d_{i_t}$ for $r = 1, 2, \ldots, t−1$ and $s_{i_t}$ has an edge to $d_{i_t}$. (For example, Fig. 3(a), (b) are both 3-unicast non-GNS profiles for which one can set $i_1 = 1, i_2 = 2, i_3 = 3$.) We now propose a coding scheme for $N(P)$ which achieves a sum-rate of $\frac{1}{t−1}$, thus showing $\rho(P) \geq \frac{1}{t−1} > 1$.

Assume that only the sources $s_{i_0}, \alpha = 1, 2, \ldots, t$ wish to deliver one message symbol from a finite field $\mathbb{F}$ to their respective destinations and that the edge $(u,v)$ can carry one finite symbol per time slot. We will accomplish this task in $t−1$ time slots. Node $u$ receives all the $t$ finite field message symbols, say $X_1, X_2, \ldots, X_t$. It sends out $t−1$ random linear combinations of these symbols on the edge $(u,v)$, where the co-efficient are uniformly chosen from $\mathbb{F}$ independent across the different symbols and across time. Each destination $d_{i_0}, \alpha = 1, 2, \ldots, t$ receives all of these $t−1$ symbols and also has one message symbol of side information from the source directly connected to it. Standard calculations similar to those in [2] can then be used to show that each destination can recover its intended message with high probability as the size of the finite field $\mathbb{F}$ goes to infinity. Thus, there exists some coding scheme that delivers the desired performance.

**Remark 1:** We note that $\rho(P)$ may be quite hard to compute, especially for large $k$. However, once computed for a profile for a specific $k$, it gives useful bounds for all $k$-unicast networks with no restrictions on the size of such networks. Recent work in [15] provides inner bounds on the entire capacity region for the index coding problem. In particular, their bounds are tight for up to five-node networks which would allow us to evaluate the sum-capacity $\rho(P)$ exactly for all canonical networks $N(P)$ with $k \leq 5$.

V. NP-COMPLETENESS OF MINIMUM GNS-CUT

The works of [3], [9], [12] provide algorithms to check if their approach can deduce the fundamentality of a given edge-cut. However, the number of edge-cuts is exponential in the size of the network and so listing all of them and checking if they provide fundamental bounds is computationally intractable. For a single-unicast problem, we know that Ford-Fulkerson’s algorithm reveals the mincut efficiently inspite of there being exponentially many edge-cuts. Given a capacitated $k$-unicast network, can we have any algorithm that efficiently finds, among all complete edge-cuts $E$ that are GNS-cuts, the one that has the smallest value of $\sum_{e \in E} C_e$? Theorem 4 will show unfortunately that we cannot, even for $k = 2$, unless P=NP. Let us define the following decision problem.

**MIN 2-GNS-CUT**

**Instance:** A two-unicast uncapacitated network $G$ and an assignment of non-negative capacities $C_e$ to the edges.

**Question:** Is there a set of edges $E \subseteq E(G)$ with $\sum_{e \in E} C_e \leq K$ such that the edge-cut derived from $E$ is a GNS-cut?

**Theorem 4:** MIN 2-GNS-CUT is NP-complete.

**Proof:** It is clear that MIN 2-GNS-CUT is in NP. We give a polynomial transformation from the multiterminal cut problem for three terminals which is known to be NP-complete [16]. In the multiterminal cut problem, we are given a number $K$ and an unweighted undirected graph $H$ with three special vertices or “terminals” $x, y, z$. We are asked whether there is a subset of edges $F$ of the graph $H$ with $|F| \leq K$ such that $H \setminus F$ has no paths between any two of $x, y, z$. Given $(H, K)$, we construct a corresponding instance of MIN 2-GNS-CUT as follows. Let the number of edges of $H$ be $N$ with $K \leq N$.

The two-unicast capacitated network $G$ is obtained by replacing each undirected edge $(u,v)$ of $H$ with a gadget as shown in Fig. 4. The gadget introduces two new vertices $u, u'$ and constitutes five edges, the one central edge having
unit capacity and four flank edges each having capacity $N + 1$ units. Finally, $s_1$ is identified with terminal $x$, $d_2$ with terminal $y$ and both $s_2$ and $d_1$ with terminal $z$.

We show that $\mathcal{G}$ has a GNS-cut derived from a set of edges $E$ with $\sum_{e \in E} C_e \leq K$ if and only if $\mathcal{H}$ has a set of edges $F$ forming a multiterminal cut with $|F| \leq K$.

Suppose that in the undirected graph $\mathcal{H}$, there is a multiterminal cut $F$ with at most $K$ edges. Then, picking the central edge of the gadgets corresponding to the edges in $F$ gives a GNS-cut in $\mathcal{G}$ with edge set $E$ such that $\sum_{e \in E} C_e = |F| \leq K$.

Conversely, suppose there is a GNS-cut in $\mathcal{G}$ derived from an edge set $E$ which satisfies $\sum_{e \in E} C_e \leq K$. As $s_2$ and $d_1$ are identified, the GNS-cut must have the profile shown in Fig. 2 (c). Moreover, as $K \leq N$, the edge set $E$ cannot contain any flank edge and must consist exclusively of central edges of gadgets. Choosing the undirected edges of $\mathcal{H}$ corresponding to the gadgets whose central edges lie in $E$ gives an edge set $F$ of $\mathcal{H}$ that has at most $K$ edges and is a multiterminal cut in $\mathcal{H}$.

VI. DISCUSSION

Theorem 3 does not say that an edge-cut with a non-GNS profile must necessarily not be fundamental. Consider the example in Fig. 5(a). The edge-cut derived from $E = \{e_1, e_2\}$ is not a GNS-cut for the two-unicast network shown, yet $R_1 + R_2 \leq C_{e_1} + C_{e_2}$ is a fundamental edge-cut bound. The reason of course, is that $R_1 \leq C_{e_2}$ and $R_2 \leq C_{e_1}$ follow from the cutset bound. Fig. 5(b) shows each edge assigned unit capacity and a specific coding scheme. This coding scheme makes it clear why functional dependence or information dominance do not capture this bound for the 2-unicast problem: The information flowing on $\{e_1, e_2\}$ does not dominate all the source messages.

Although this example is somewhat daft, the general question is not. Is there a fundamental sum-rate edge-cut bound for a $k$-unicast network that is not implied by the GNS edge-cut bound collection? The answer is No for $k = 1$ (by Max-Flow-Min-Cut theorem) and was shown to also be No for $k = 2$ in Theorem 5 of [11]. The question is open for $k \geq 3$.

It is also of interest to explore weighted sum-rate edge-cut bounds as studied in [14]. It would be useful to determine other classes of Shannon (or non-Shannon) inequalities and other information about the graph structure of the network that can help in deriving such bounds.

VII. ACKNOWLEDGEMENTS

We would like to thank Satish Rao for fruitful discussions on the GNS bound and for suggesting the NP-completeness reduction used in the paper.

REFERENCES