On the multiple unicast capacity of 3-source, 3-terminal directed acyclic networks

Shurui Huang, *Student Member, IEEE* and Aditya Ramamoorthy, *Member, IEEE*
Department of Electrical and Computer Engineering
Iowa State University, Ames, Iowa 50011
Email: {hshurui, adityar}@iastate.edu

***Abstract*—We consider a directed acyclic network with unit-capacity edges and n source-terminal($s_i - t_i$) pairs that wish to communicate at unit rate via network coding. The connectivity between the $s_i - t_i$ pairs is quantified by means of a connectivity level vector, $[k_1\ k_2 \dots\ k_n]$ such that there exist k_i edge-disjoint paths between s_i and t_i. In this work we attempt to classify networks based on the connectivity level. Specifically, for case of $n = 3$, and a given connectivity level vector $[k_1\ k_2\ k_3]$, we aim to either present a constructive network coding scheme or an instance of a network that cannot support the desired rate. Certain generalizations for the case of higher n have also been found.**

I. Introduction

In a network that supports multiple unicast, there are several source terminal pairs; each source wishes to communicate with its corresponding terminal. Multiple unicast connections form bulk of the traffic over both wired and wireless networks. Thus, network coding schemes that can help improve network throughput for multiple unicasts are of considerable interest. However, it is well recognized that the design of constructive network coding schemes for multiple unicasts is a hard problem. Specifically, it is known that there are instances of networks where linear (whether scalar or vector) network coding is insufficient [1].

The multiple unicast problem has been examined for both directed acyclic networks [2][3][4] and undirected networks [5] in previous work. The work of [6], provides an information theoretic characterization for directed acyclic networks. However, this bound is not computable as there is no upper bound on the cardinality of the random variables involved in the characterization.

There have been attempts to find constructive solutions by leveraging linear network coding between pairs of flows (much like the butterfly structure). The work of [7] suggests back pressure algorithms for finding achievable rates using XOR operation between pairs of flows, while the work of [2] attempts to address it by packing butterfly networks within the original graph. The work of [3] (also see [8]), considers the multiple unicast problem in the case of two source-terminal pairs where each terminal wants to operate at unit-rate. Specifically, [3] presents a graph-theoretic condition that can be efficiently verified for checking the existence of a network coding solution for the unit-rate problem. The work of [9], also considers schemes for two-unicast, where one can trade-off rates between the different connections. For the case of two-unicast, [10] and [11] propose outer bounds on the capacity region.

The work of Das et al. [12], [13] considers multiple unicast for three source-terminal pairs using the idea of interference alignment that was proposed in the area of physical layer wireless communication [14]. This can be viewed as a scheme where the network operates via random linear network coding and the sources choose appropriate precoding matrices in order to satisfy the demands of each terminal. This is a potentially powerful technique that allows us to achieve half of the minimum cut of each source-terminal pair simultaneously. However, the scheme of [12], [13] requires several algebraic conditions to be satisfied in order for the technique to be applicable[1] and it is unclear whether this can be done efficiently. Some progress on this issue has been reported in [15].

In this work we consider linear network coding schemes for wired three-source, three-terminal directed acyclic networks with unit capacity edges. There are source-terminal pairs denoted $s_i - t_i, i = 1, \dots, 3$, such that the maximum flow from s_i to t_i is k_i. Each source contains a unit-entropy message that needs to be communicated to the corresponding terminal. In general, this is a hard problem as bounds or constructive schemes may depend heavily on the network topology. In this work, for a given connectivity level vector $[k_1\ k_2\ k_3]$ we attempt to either design a constructive scheme based on linear network codes or demonstrate an instance of a network where supporting unit-rate transmission is impossible. Our achievability schemes use a combination of random linear network coding and appropriate precoding. However, unlike [12], our schemes are not asymptotic. In particular, our solutions are based either scalar codes or vector codes that operate over two time units (i.e., two network uses). This is potentially useful, as one could arrive at multiple unicast schemes for arbitrary rates by packing unit-rate structures for which our achievability schemes apply.

A. Main Contributions

• For the case of three unicast sessions, we identify certain feasible and infeasible connectivity levels $[k_1\ k_2\ k_3]$. For the

This research was supported in part by NSF grant CCF-1018148.

[1] In the wireless physical layer setting, these can be assumed to hold under appropriate channel models.

feasible cases, we provide efficient schemes based on linear network coding. For the infeasible cases, we provide counter-examples, i.e., instances of graphs where the multiple unicast cannot be supported under any (potentially nonlinear) network coding scheme.

- We identify certain infeasible instances for two unicast sessions, where the source rates are not necessarily the same.

Owing to space limitations, proofs of several results are not provided. These can be found in the full version of the paper that is available at [16]. This paper is organized as follows. In Section II, we introduce the network coding model and problem formulation. Section III discusses infeasible instances, and Section IV discusses our achievable schemes for 3-source, 3-terminal multiple unicast networks. Section V concludes the paper with a discussion of future work.

II. Preliminaries

We represent the network as a directed acyclic graph $G = (V, E)$. Each edge $e \in E$ has unit capacity and can transmit one symbol from a finite field of size q per unit time (we are free to choose q large enough). If a given edge has higher capacity, it can be treated as multiple unit capacity edges. A directed edge e between nodes i and j is represented as (i, j), so that $head(e) = j$ and $tail(e) = i$. A path between two nodes i and j is a sequence of edges $\{e_1, e_2, \ldots, e_k\}$ such that $tail(e_1) = i, head(e_k) = j$ and $head(e_i) = tail(e_{i+1}), i = 1, \ldots, k-1$. The network contains a set of n source nodes s_i and n terminal nodes $t_i, i = 1, \ldots n$. Each source node s_i observes a discrete integer-entropy source, that needs to be communicated to terminal t_i. Without loss of generality, we assume that the source (terminal) nodes do not have incoming (outgoing) edges.

We now discuss the network coding model under consideration in this paper. For the sake of simplicity, suppose that each source has unit-entropy, denoted by X_i. In scalar linear network coding, the signal on an edge (i, j), is a linear combination of the signals on the incoming edges on i or the source signals at i (if i is a source). We shall only be concerned with networks that are directed acyclic and can therefore be treated as delay-free networks [17]. Let Y_{e_i} (such that $tail(e_i) = k$ and $head(e_i) = l$) denote the signal on edge $e_i \in E$. Then, we have

$$Y_{e_i} = \sum_{\{e_j | head(e_j) = k\}} f_{j,i} Y_{e_j} \text{ if } k \in V \backslash \{s_1, \ldots, s_n\}, \text{ and}$$

$$Y_{e_i} = \sum_{j=1}^{n} a_{j,i} X_j \quad \text{where } a_{j,i} = 0 \text{ if } X_j \text{ is not observed at } k.$$

The coefficients $a_{j,i}$ and $f_{j,i}$ are from the operational field. Note that since the graph is directed acyclic, it is equivalently possible to express Y_{e_i} for an edge e_i in terms of the sources X_j's. If $Y_{e_i} = \sum_{k=1}^{n} \beta_{e_i,k} X_k$ then we say that the global coding vector of edge e_i is $\boldsymbol{\beta}_{e_i} = [\beta_{e_i,1} \; \cdots \; \beta_{e_i,n}]$. We shall also occasionally use the term coding vector instead of global coding vector in this paper. We say that a node i (or edge e_i) is downstream of another node j (or edge e_j) if there exists a path from j (or e_j) to i (or e_i).

Vector linear network coding is a generalization of the scalar case, where we code across the source symbols in time, and the intermediate nodes can implement more powerful operations. Formally, suppose that the network is used over T time units. We treat this case as follows. Source node s_i now observes a vector source $[X_i^{(1)} \; \ldots \; X_i^{(T)}]$. Each edge in the original graph is replaced by T parallel edges. In this graph, suppose that a node j has a set of β_{inc} incoming edges over which it receives a certain number of symbols, and β_{out} outgoing edges. Under vector network coding, j chooses a matrix of dimension $\beta_{out} \times \beta_{inc}$. Each row of this matrix corresponds to the local coding vector of an outgoing edge from j.

Note that the general multiple unicast problem, where edges have different capacities and the sources have different entropies can be cast in the above framework by splitting higher capacity edges into parallel unit capacity edges, a higher entropy source into multiple, collocated unit-entropy sources; and the corresponding terminal node into multiple, collocated terminal nodes.

An instance of the multiple unicast problem is specified by the graph G and the source terminal pairs $s_i - t_i, i = 1, \ldots, n$, and is denoted $< G, \{s_i - t_i\}_1^n, \{R_i\}_1^n >$, where the rate R_i denotes the entropy of the i^{th} source. The $s_i - t_i$ connections will be referred to as sessions that we need to support.

The instance is said to have a scalar linear network coding solution if there exist a set of linear encoding coefficients for each node in V such that each terminal t_i can recover X_i using the received symbols at its input edges. Likewise, it is said to have a vector linear network coding solution with vector length T if the network employs vector linear network codes and each terminal t_i can recover $[X_i^{(1)} \; \ldots \; X_i^{(T)}]$. If the instance has either a scalar or a vector network coding solution, we say that it is feasible.

In a routing solution, each edge carries a copy of one of the sources, i.e., each coding vector is such that at most one entry takes the value 1, all others are 0. Scalar (vector) routing solutions can be defined in a manner similar to scalar (vector) network codes. We now define some quantities that shall be used throughout the paper.

Definition 2.1: Connectivity level. The connectivity level for source-terminal pair $s_i - t_i$ is said to be n if the maximum flow between s_i and t_i in G is n. The connectivity level of the set of connections $s_1 - t_1, \ldots, s_n - t_n$ is the vector [max-flow($s_1 - t_1$) max-flow($s_2 - t_2$) $\ldots$ max-flow($s_n - t_n$)].

We conclude this section by observing that a multiple unicast instance G with n $(s_i - t_i)$ pairs and connectivity level $[n \; n \; \ldots \; n]$ is always feasible. Specifically, we employ vector routing over n time units. Source s_i observes $[X_i^{(1)} \; \ldots \; X_i^{(n)}]$ symbols. Each edge e in the original graph G is replaced by n parallel edges, $e^1, e^2, \ldots, e^n$. Let G_α represent the subgraph of this graph consisting of edges with superscript α. It is evident that max-flow($s_\alpha - t_\alpha$) = n over G_α. Thus, we transmit $X_\alpha^{(1)}, \ldots, X_\alpha^{(n)}$ over G_α using routing, for all $\alpha = 1, \ldots, n$. demands of all the terminals. In general, though a network with the above connectivity level may not be able to support a scalar routing solution; an instance is shown in [18].

III. Network coding for three unicast sessions - Infeasible Instances

It is clear based on the discussion above that for three unicast sessions if the connectivity level is [3 3 3], then a vector routing solution always exists. We investigate counter-examples for certain connectivity levels in this section. This material has appeared in a previous paper by the authors and we only state the results here and refer the reader to [18] for details.

Lemma 3.1: [18] There exist multiple unicast instances with three unicast sessions, $< G, \{s_i - t_i\}_{i=1}^3, \{1, 1, 1\} >$ such that the connectivity levels [1 1 3], [2 2 2] and [2 2 3] are infeasible.

The [2 3] counter-example in [18] can be generalized to an instance with two unicast sessions with connectivity level $[n_1\ n_2]$ that cannot support rates $R_1 = n_1, R_2 = n_2 - 3n_1/2 + 1$ when $n_2 \geq 3n_1/2$ and $n_1 > 1$.

Theorem 3.2: For a directed acyclic graph G with two $s-t$ pairs, if the connectivity level for (s_1, t_1) is n_1, for (s_2, t_2) is n_2, where $n_2 \geq 3n_1/2$ and $n_1 > 1$, there exist instances that cannot support $R_1 = n_1$ and $R_2 = n_2 - 3n_1/2 + 1$.

Proof: Omitted owing to space limitations. ∎

IV. Network coding for three unicast sessions - Feasible Instances

In the discussion below, we show that all the instances with the connectivity levels [1 3 3], [2 2 4] and [1 2 5] are feasible using linear network codes. As pointed out by the work of [17], under linear network coding, the case of multiple unicast requires (a) the transfer matrix for each source-terminal pair to have a rank that is high enough, and (b) the interference at each terminal to be zero. Under random linear network coding, it is possible to assert that the rank of any given transfer matrix from a source s_i to a terminal t_j has w.h.p. a rank equal to the minimum cut between s_i and t_j; this in general is problematic for satisfying the zero-interference condition.

Our strategies rely on a combination of graph-theoretic and algebraic methods. Specifically, starting with the connectivity level of the graph, we use graph theoretic ideas to argue that the transfer matrices of the different terminals have certain relationships. The identified relationships then allow us to assert that suitable precoding matrices that allow each terminal to be satisfied can be found. We begin with the following definitions.

Definition 4.1: Minimality. Consider a multiple unicast instance $< G = (V, E), \{s_i - t_i\}_1^n >$, with connectivity level $[k_1\ k_2\ \dots\ k_n]$. The graph G is said to be minimal if the removal of any edge from E reduces the connectivity level. If G is minimal, we will also refer to the multiple unicast instance as minimal.

Clearly, given a non-minimal instance $G = (V, E)$, we can always remove the non-essential edges from it to obtain the minimal graph $G_{\min}$; this does not affect feasibility.

Definition 4.2: Overlap edge. An edge e is said to be an overlap edge for paths P_i and P_j in G, if $e \in P_i \cap P_j$.

Definition 4.3: Overlap segment. Consider a set of edges $E_{os} = \{e_1, \dots, e_l\} \subset E$ that forms a path. This path is called an overlap segment for paths P_i and P_j if

(i) $\forall k \in \{1, \dots, l\}$, e_k is an overlap edge for P_i and P_j,
(ii) none of the incoming edges into tail(e_1) are overlap edges for P_i and P_j, and
(iii) none of the outgoing edges leaving head(e_l) are overlap edges for P_i and P_j.

Our solution strategy is as follows. We first convert the original instance into another *structured* instance where each internal node has at most degree three (in-degree + out-degree). We then convert this new instance into a minimal one, and develop the code assignment algorithm. It will be evident that using this network code, one can obtain a network code for the original instance.

Following [19] we can efficiently construct a *structured* graph $\hat{G} = (\hat{V}, \hat{E})$ in which each internal node $v \in \hat{V}$ is of total degree at most three with the following properties.

(a) $\hat{G}$ is acyclic.
(b) For every source (terminal) in G there is a corresponding source (terminal) in $\hat{G}$.
(c) For any two edge disjoint paths P_i and P_j for one unicast session in G, there exist two *vertex* disjoint paths in $\hat{G}$ for the corresponding session in $\hat{G}$.
(d) A feasible network coding solution in $\hat{G}$ can be efficiently turned into a feasible network coding solution in G.

In all the discussions below, we will assume that the graph G is structured. It is clear that this is w.l.o.g. based on the previous arguments.

We state the result for connectivity level [1 3 3] here and refer the interested reader to the proof in [18].

Theorem 4.4: [18] A multiple unicast instance with three sessions, $< G, \{s_i - t_i\}_1^3, \{1, 1, 1\} >$ with connectivity level at least [1 3 3] is feasible.

A. Code Assignment Procedure For Instances With Connectivity Level [2 2 4]

We first investigate a two-unicast scenario with connectivity level [2 4] and rate requirement $\{2, 1\}$ and use that in conjunction with vector network coding to address three-unicast with connectivity level [2 2 4]. We use random linear coding and precoding at the sources to arrive at the result. Proving the existence of an appropriate precoding scheme requires us to explore the relation between the structure of the graph and the properties of the transfer matrices.

Lemma 4.5: A minimal multiple unicast instance $< G, \{s_1 - t_1, s_2 - t_2\}, \{2, 1\} >$ with connectivity level [2 4] is feasible.

Proof: Let $\mathcal{P}_1 = \{P_{11}, P_{12}\}$ denote two edge disjoint paths (also vertex disjoint due to the structured nature of G) from s_1 to t_1 and $\mathcal{P}_2 = \{P_{21}, P_{22}, P_{23}, P_{24}\}$ denote the four vertex disjoint paths from s_2 to t_2. Let the source messages at s_1 be denoted by X_1 and X_2, and the source message at s_2

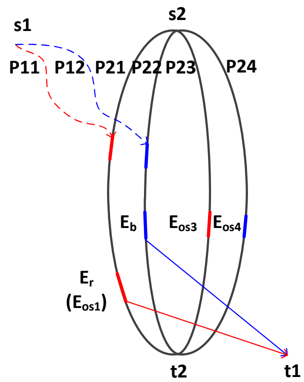


Fig. 1. A network with connectivity levels [2 4] and rate $\{2, 1\}$. A linear network coding solution can be found .

by X_3. We color the edges of the graph such that each edge on P_{11} is colored red, each edge on P_{12} is colored blue and each edge on a path in $\mathcal{P}_2$ is colored black.

As the paths in $\mathcal{P}_1$ and $\mathcal{P}_2$ are vertex-disjoint, it is clear that a node with an in-degree of two is such that its outgoing edge has two colors (either *(blue, black)* or *(red, black)*). The path further downstream continues to have two colors until it reaches a node of out-degree two.

Such an overlap segment with two colors will be referred to as a *mixed color overlap segment*. We shall also use the terms *red* or *blue overlap segment* to refer to segments with colors *(red, black)* and *(blue, black)* respectively. Note that by our naming convention path P_{ij} is a path that enters terminal t_i. Under the topological order in G we can identify the overlap segment on P_{ij} that is closest to t_i. In the discussion below this will be referred to as the last overlap segment with respect to path P_{ij}. Two overlap segments E_{os1} and E_{os2} are said to be neighboring with respect to P_{ij} if there are no overlap segments between them along P_{ij}.

Claim 4.6: Consider two neighboring mixed color overlap segments E_{os1} and E_{os2} with respect to path $P_{1i} \in \mathcal{P}_1$. Then E_{os1} and E_{os2} cannot lie on the same path $P_{2j} \in \mathcal{P}_2$.
proof: This follows from the minimality of graph. ■

Likewise, two neighboring mixed color overlap segments E_{os1} and E_{os2} with respect to P_{2i}, cannot lie on the same path P_{1j}.

To explain our coding scheme, we first denote the last red (blue) overlap segment with respect to P_{11} (P_{12}) by E_r (E_b). If there is no E_r, then X_1 can be transmitted along P_{11}. According to Lemma A.2, X_2 and X_3 can be transmitted to t_1 and t_2 respectively. A similar argument can be applied to the case when there is no E_b. Hence we assume that both E_r and E_b exist. Based on their locations in G, we distinguish the following two cases.

- *Case 1: E_r and E_b are on different paths $\in \mathcal{P}_2$.*

W.l.o.g. we assume that E_r and E_b are on paths P_{21} and P_{22}. If there are no mixed color overlap segments on either P_{23} or P_{24}, X_3 can be transmitted to t_2 through the overlap segment free path, and X_1, X_2 can be routed to t_1. Therefore, we focus on the case that there are mixed color overlap segments on both P_{23} and P_{24}. Let E_{osi} denote the last mixed color overlap segments with respect to P_{2i}, $i = 1, \ldots, 4$ (see example in Fig. 1).

Our coding scheme is as follows. Symbol X_i is transmitted over the outgoing edge from s_1 over $P_{1i}, i = 1, 2$; symbols $\theta_j X_3$ are transmitted over the outgoing edges of s_2 over P_{2j}, $j = 1, \ldots, 4$ respectively. The values of $\theta_j \in GF(q)$ will be chosen as part of the code assignment below. Let the coding vectors at each intermediate node be specified by indeterminates for now. The overall transfer matrix from the pair of sources $\{s_1, s_2\}$ to t_1 can be expressed as

$$[M_{11} \mid M_{12}] = \left[\begin{array}{cc|cccc} \alpha_1 & \beta_1 & \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} \\ \alpha_2 & \beta_2 & \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} \end{array}\right],$$

where M_{11} is a 2×2 matrix which contains the coefficients for X_1 and X_2 and M_{12} is a 2×4 matrix which contains the coefficients for $\theta_i X_3$, $i = 1, \ldots, 4$, i.e., the received vector at t_1 is $[M_{11} \mid M_{12}][X_1 \ X_2 \mid \theta_1 X_3 \ \ldots \ \theta_4 X_3]^T$. Recall that E_r and E_b are the last mixed color overlap segments with respect to P_{11} and P_{12}. Thus, they carry the same information as the incoming edges of t_1 which implies that the row vectors of $[M_{11} \mid M_{12}]$ are the coding vectors on E_r and E_b respectively. Similarly, the transfer matrix from $\{s_1, s_2\}$ to the edge set $\{E_r, E_b, E_{os3}, E_{os4}\}$ can be expressed as

$$[M_{21} \mid M_{22}] = \left[\begin{array}{cc|cccc} \alpha_1 & \beta_1 & \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} \\ \alpha_2 & \beta_2 & \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} \\ \alpha_3 & \beta_3 & \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} \\ \alpha_4 & \beta_4 & \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} \end{array}\right].$$

Note that the entries of the transfer matrices above are functions of the choice of local coding vectors in the network.

As there exist two edge disjoint paths from s_1 to $\{E_r, E_b\}$, the determinant of M_{11} is not identically zero. Similarly, since the edges E_r, E_b, E_{os3} and E_{os4} lie on different paths in $\mathcal{P}_2$, there are four edge disjoint paths from s_2 to $\{E_r, E_b, E_{os3}, E_{os4}\}$, and the determinant of M_{22} is not identically zero. This implies that their product is not identically zero. Hence, by the Schwartz-Zippel lemma, under random linear coding there exists an assignment of local coding vectors so that $rank(M_{11}) = 2$ and $rank(M_{22}) = 4$. We assume that the local coding vectors are chosen from a large enough field $GF(q)$ so that this is the case. For this choice of local coding vectors we propose a choice of $\underline{\theta} = [\theta_1 \ \theta_2 \ \theta_3 \ \theta_4]^T$ such that the decoding is simultaneously successful at both t_1 and t_2.

Decoding at t_1: As M_{11} is a square full-rank matrix, we only need to null the interference from s_2. Accordingly, we choose $\underline{\theta}$ from the null space of M_{12}, i.e.,

$$M_{12}\underline{\theta} = 0. \quad (1)$$

There are at least $q^2 - 1$ such non-zero choices for $\underline{\theta}$ as M_{12} is a 2×4 matrix.

Decoding at t_2: The primary issue is that one needs to demonstrate that the choice of $\underline{\theta}$ allows both terminals to simultaneously decode. Indeed, it may be possible that our choice of $\underline{\theta}$ along with a specific network topology may make it impossible to decode at t_2. The key argument that this does not happen requires us to leverage certain topological properties of the overlap segments, that we present below.

Claim 4.7: In G either one or both of the following statements hold. (i) E_r is the last overlap segment w.r.t. P_{21}. (ii) E_b is the last overlap segment w.r.t. P_{22}.

Proof: Assume that neither statement is true. This means that there is a blue overlap segment E'_b below E_r along P_{21}, and there is a red overlap segment E'_r below E_b along P_{22}. Thus, E'_r is upstream of E_r and E'_b is upstream of E_b. However, this means that edges E'_r, E_r, E'_b and E_b form a cycle, which is a contradiction. ∎

In the discussion below, w.l.o.g., we assume that E_r is the last overlap segment on P_{21}. The argument above allows us to identify three edges, E_r, E_{os3} and E_{os4} that carry the *same symbols* as those entering t_2. Our main idea is to cancel the X_1 and X_2 component using the information on E_{os3} and E_{os4} while retaining the X_3 component.

Let $\underline{\gamma}_i$ represent the vector $[\gamma_{i1}\ \gamma_{i2}\ \gamma_{i3}\ \gamma_{i4}]^T, i = 1, \ldots, 4$ in the discussion below. Note that if $[\alpha_3\ \beta_3]$ and $[\alpha_4\ \beta_4]$ are linearly independent, there exist δ_3 and δ_4 such that

$$[\alpha_1\ \beta_1] = \delta_3[\alpha_3\ \beta_3] + \delta_4[\alpha_4\ \beta_4],$$

where δ_3 and δ_4 are not both zero. Thus, t_2 can recover $[-\underline{\gamma}_1 + \delta_3\underline{\gamma}_3 + \delta_4\underline{\gamma}_4]^T\underline{\theta}X_3$. Note that $\underline{\gamma}_1^T\underline{\theta} = 0$, by the constraint on $\underline{\theta}$ above, thus we only need to pick $\underline{\theta}$ such that $[\delta_3\underline{\gamma}_3 + \delta_4\underline{\gamma}_4]^T\underline{\theta} \neq 0$. To see that this can be done, we note that M_{22} is full rank which implies that the matrix $[\underline{\gamma}_1\ \ \underline{\gamma}_2\ \ (\delta_3\underline{\gamma}_3+\delta_4\underline{\gamma}_4)]^T$ is full rank. Therefore, there exist at most q choices for $\underline{\theta}$ such that $[\underline{\gamma}_1\ \ \underline{\gamma}_2\ \ (\delta_3\underline{\gamma}_3+\delta_4\underline{\gamma}_4)]^T\underline{\theta} = 0$. Hence, there are at least $q^2 - q - 1 > 0$ non-zero choices for $\underline{\theta}$ that allow decoding at t_1 and t_2 simultaneously.

If $[\alpha_3\ \ \beta_3]$ and $[\alpha_4\ \ \beta_4]$ are dependent, decoding can be performed simply by working only with the received values over E_{os3} and E_{os4} using a similar argument as above.

• *Case 2: E_r and E_b are on the same path P_{2i}.*

We proceed by identifying the blue overlap segment E'_b that is a neighbor of E_b w.r.t. P_{12} and adapting the analysis above. The details can be found in [16]. ∎

By using the result of Lemma 4.5 and vector network coding over two time units, we have the following theorem when the connectivity level is [2 2 4].

Theorem 4.8: A multiple unicast instance with three sessions, $< G, \{s_i - t_i\}_1^3, \{1,1,1\} >$ with connectivity level at least [2 2 4] is feasible.

B. Code Assignment Procedure For Instances With Connectivity Level [1 2 5]

We now consider the network code assignment for networks where the connectivity level is [1 2 5]. The code assignment in this case requires somewhat different techniques. In particular, the idea of using a two-session unicast result along with vector network coding does not work unlike the cases considered previously. At the top level, we still use random network coding followed by appropriate precoding to align the interference seen by the terminals. However, as we shall see below, we will need to depart from a purely random linear code in the network in certain situations.

As before, we consider a minimal structured graph G and let X_i be the source symbol at source node s_i for $i = 1, \ldots, 3$ and $\mathcal{P}_1 = \{P_{11}\}$ denote the path from s_1 to t_1, $\mathcal{P}_2 = \{P_{21}, P_{22}\}$ denote the edge disjoint paths from s_2 to t_2, $\mathcal{P}_3 = \{P_{31}, P_{32}, P_{33}, P_{34}, P_{35}\}$ denote the edge disjoint paths from s_3 to t_3.

Our scheme operates as follows: X_1 is transmitted over the outgoing edge from s_1 along P_{11} , $\xi_i X_2$ are transmitted over the outgoing edges of s_2 along P_{2i}, $i = 1, 2$, and $\theta_j X_3$ are transmitted over the outgoing edges of s_3 along P_{3j}, $j = 1, \ldots, 5$ where $\underline{\xi} = [\xi_1\ \xi_2]^T$ and $\underline{\theta} = [\theta_1\ \ldots\ \theta_5]^T$ are precoding vectors chosen from a finite field with size q.

Let M_i denote the transfer matrix from $\{s_1, s_2, s_3\}$ to terminal t_i. The matrix M_i is partitioned into blocks so that $M_i = [M_{i1} \mid M_{i2} \mid M_{i3}]$, and each M_{ij} corresponds to the transformation from source s_j to terminal t_i, i.e., the number of columns in M_{ij} is $1, 2$ and 5 for $j = 1, 2$ and 3 respectively. Similarly, the number of rows in M_{ij} is $1, 2$ and 5 for $i = 1, 2$ and 3 respectively.

In the discussion below we will need to refer to the individual entries of M_1 and M_2. Accordingly, we express these matrices explicitly as follows.

$$\begin{aligned} M_1 &= [M_{11} \mid M_{12} \mid M_{13}] = \left[\alpha_1 \mid \underline{\beta}^T \mid \underline{\gamma}^T\ \right] \\ &= [\alpha_1 \mid \beta_1\ \beta_2 \mid \gamma_1\ \gamma_2\ \gamma_3\ \gamma_4\ \gamma_5]\,, \\ M_2 &= [M_{21} \mid M_{22} \mid M_{23}] = \left[\begin{array}{c|c|c} \alpha'_1 & \underline{\beta'}^T_1 & \underline{\gamma'}^T_1 \\ \alpha'_2 & \underline{\beta'}^T_2 & \underline{\gamma'}^T_2 \end{array}\right] \\ &= \left[\begin{array}{c|cc|ccccc} \alpha'_1 & \beta'_{11} & \beta'_{12} & \gamma'_{11} & \gamma'_{12} & \gamma'_{13} & \gamma'_{14} & \gamma'_{15} \\ \alpha'_2 & \beta'_{21} & \beta'_{22} & \gamma'_{21} & \gamma'_{22} & \gamma'_{23} & \gamma'_{24} & \gamma'_{25} \end{array}\right]. \end{aligned}$$

We are given that $min{-}cut(s_1{-}t_1) = 1, min{-}cut(s_2{-}t_2) = 2$ and $min - cut(s_3 - t_3) = 5$. This implies that $\det(M_{ii})$ is not identically zero for $i = 1, \ldots, 3$, and furthermore that their product $\det(M_{11})\det(M_{22})\det(M_{33})$ is not identically zero.

We first identify a minimal structured subgraph G' of G that has the following properties.

(i) There exists a path P'_{11}, from s_1 to t_1,
(ii) vertex disjoint paths P'_{21} and P'_{22} from s_2 to t_2,
(iii) path $P'_{1\to 2}$ from s_1 to t_2 and
(iv) path $P'_{2\to 1}$ from s_2 to t_1.

Here again, G' is said to be minimal if the removal of any edge from it causes one of the above properties to fail. We note that it is possible that there do not exist any paths from s_1 to t_2 or from s_2 to t_1 in G. These situations are considered below.

Our analysis depends on the following topological properties of G'.

Case 1: The graph G' is such that

- there is no path from s_1 to t_2 in G', i.e., $P'_{1\to 2} = \emptyset$ (this happens only if there is no path from s_1 to t_2 in G), or

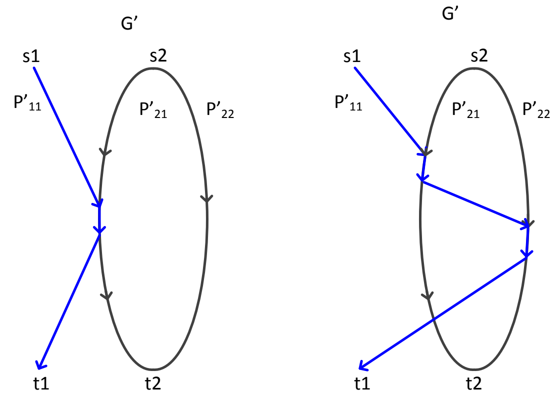
- there is no path from s_2 to t_1 in G', i.e., $P'_{2\to1} = \emptyset$ (this happens only if there is no path from s_2 to t_1 in G), or
- there are paths $P'_{1\to2}$ and $P'_{2\to1}$ in G', and there are overlap segments between P'_{11} and $P'_{21} \cup P'_{22}$.

Case 2: The graph G' is such that

- there are paths $P'_{1\to2}$ and $P'_{2\to1}$ in G', and P'_{11} does not overlap with either P'_{21} or P'_{22}.

We emphasize that the condition of Case 2 is the logical negation of the conditions in Case 1.

Theorem 4.9: A multiple unicast instance with three sessions, $< G, \{s_i - t_i\}_1^3, \{1,1,1\} >$, with connectivity level [1 2 5] is feasible.



(a) (b)

Fig. 2. (a) Subgraph G' when P'_{11} overlaps with P'_{21}. (b) Subgraph G' when P'_{11} overlaps with both P'_{21} and P'_{22}.

Proof: We break up the proof into two parts based on type of the subgraph G' that we can find in G.

Proof when there exists a subgraph G' that satisfies the conditions of Case 1:

We perform random linear coding over the graph G over a large enough field. In the discussion below, we will leverage the fact that multivariate polynomials that are not identically zero, evaluate to a non-zero value with high probability (w.h.p.) under a uniformly random choice of the variables. This is needed at several places. By using standard union bound techniques, we can claim that our strategy works with high probability.

In particular, in the discussion below, we assume that the matrices $M_{ii}, i = 1, \ldots, 3$ are full rank and design appropriate precoding vectors $\underline{\xi}$ and $\underline{\theta}$.

Decoding at t_1: For t_1 to decode X_1, we need to have, $\alpha_1 \neq 0$ and the precoding constraints.

$$[\beta_1\ \beta_2]\underline{\xi} = 0 \tag{2}$$

$$[\gamma_1\ \gamma_2\ \gamma_3\ \gamma_4\ \gamma_5]\underline{\theta} = 0 \tag{3}$$

There are at least $q-1$ non-zero vectors $\underline{\xi}$ and q^4-1 non-zero vectors $\underline{\theta}$ that can be selected from the field of size q such that eq. (2) and eq. (3) are satisfied.

Decoding at t_2:

We begin by noting that since $rank(M_{22}) = 2$, $M_{22}\underline{\xi} \neq 0$, as long as $\xi \neq 0$. Next, we argue according to the topological structure of G'. The following possibilities can occur.

(i) *There is no path from s_1 to t_2 in G', i.e., $P'_{1\to2} = \emptyset$.* This implies that $\alpha'_1 = \alpha'_2 = 0$ and in G, interference at t_2 only exists due to s_3. Next, at least one component of $M_{22}\underline{\xi}$ will be non-zero, based on the argument above; w.l.o.g. assume that it is the first component. We choose $\underline{\theta}$ to satisfy

$$\underline{\gamma}_1'^T\underline{\theta} = 0 \tag{4}$$

It is evident that there are at least q^3-1 non-zero choices of $\underline{\theta}$ that satisfy the required constraints on $\underline{\theta}$ (eqs. (3) and (4)). Hence t_2 can decode.

(ii) *There exists a path $P'_{1\to2}$ from s_1 to t_2, i.e., $P'_{1\to2} \neq \emptyset$.*. This means that M_{21} is not identically zero. Here, we first align the interference from s_3 within the span of interference from s_1 by selecting an appropriate $\underline{\theta}$. We have the following lemma.

*Lemma 4.10:*If $M_{21} \neq 0$, there exist at least q^4-1 choices for $\underline{\theta}$ such that

$$M_{23}\underline{\theta} = cM_{21} \tag{5}$$

where c is some constant.

Proof: First, w.l.o.g., we assume $\alpha'_2 \neq 0$. Hence, there exists a full rank 2×2 upper triangular matrix U such that $UM_{21} = [0\ \alpha'_2]^T$. Next, define

$$[1\ 0]UM_{23} = \underline{\tilde{\gamma}}_1'^T \tag{6}$$

and choose $\underline{\theta}$ to satisfy $\underline{\tilde{\gamma}}_1'^T\underline{\theta} = 0$ and set $c = \underline{\gamma}_2'^T\underline{\theta}/\alpha'_2$. Upon inspection, it can be verified that this implies that $UM_{23}\underline{\theta} = cUM_{21}$. As U is invertible, and there is only one linear constraint on $\underline{\theta}$, we have the required conclusion. ∎

Thus, under this choice of $\underline{\theta}$, the interference from s_3 is aligned within the span of the interference from s_1 at t_2. Let $\underline{X} = [X_1\ X_2\ X_3]^T$. The received signal at t_2 is given by

$$[M_{21} \mid M_{22}\underline{\xi} \mid M_{23}\underline{\theta}]\underline{X} = [M_{21}\ M_{22}\underline{\xi}]\begin{bmatrix} X_1 + cX_3 \\ X_2 \end{bmatrix} \tag{7}$$

To conclude the decoding argument for t_2, we need the claim below.

*Claim 4.11:*If M_{21} is not identically zero, under random linear coding w.h.p., there exists a $\underline{\xi}$ such that $rank[M_{21}\ \ M_{22}\underline{\xi}] = 2$ and $[\beta_1\ \beta_2]\underline{\xi} = 0$.

Proof: We will show that there exists an assignment of local coding vectors such that $\det[M_{21}\ \ M_{22}\underline{\xi}] \neq 0$. This will imply that w.h.p. under random linear coding, this property continues to hold.

Suppose that there is no path from s_2 to t_1 in G, i.e., $P'_{2\to1} = \emptyset$ and $[\beta_1\ \beta_2]$ is identically zero. This does not impose any constraint on $\underline{\xi}$. Next, M_{22} is full rank w.h.p. Hence, we can choose a $\underline{\xi}$ such that required condition is satisfied.

If there exists a path $P'_{2\to1}$ from s_2 to t_1 in G', $[\beta_1\ \ \beta_2]$ is not identically zero. W.l.o.g., we assume that β_1 is not identically zero. By Lemma A.3, proving that $\det[M_{21}\ \ M_{22}\underline{\xi}] \neq 0$, is equivalent to checking that the determinant in (11) is not identically zero. Now we demonstrate that there exists a set of local coding vectors such that the determinant in (11) is non-zero. We consider the subgraph $G' = P'_{11} \cup P'_{21} \cup P'_{22} \cup P'_{1\to 2} \cup P'_{2\to 1}$ (identified above) - our choice of the coding vectors on all the other edges will be assigned to the zero vector. As both $P'_{1\to 2} \neq \emptyset$ and $P'_{2\to 1} \neq \emptyset$, we only consider the case where P'_{11} overlaps with $P'_{21} \cup P'_{22}$. We distinguish the following cases.

1) *P'_{11} overlaps with either P'_{21} or P'_{22}.* W.l.o.g., assume it is P'_{21}. First note that when P'_{11} overlap with one of P'_{21} and P'_{22} in G', there is a path from s_1 to t_2 and a path from s_2 to t_1 in $P'_{11} \cup P'_{21} \cup P'_{22}$. Hence, G' can be completely represented by $P'_{11} \cup P'_{21} \cup P'_{22}$. This is shown in Fig. 2(a). It is evident that we can choose coding coefficients such that

$$[\beta_1 \quad \beta_2] = [1\ 0]$$

$$[M_{21} \quad M_{22}] = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{8}$$

By substituting them into eq. (11), the determinant of $[M_{21} \quad M_{22}\underline{\xi}]$ is not zero.

2) *P'_{11} overlaps with both P'_{21} and P'_{22}.* Using a similar argument as above, G' can be completely represented by $P'_{11} \cup P'_{21} \cup P'_{22}$ if P'_{11} overlaps with both P'_{21} and P'_{22}. Next, by Lemma A.1, there will be one overlap between P'_{11} and each of P'_{21} and P'_{22}. This is shown in Fig. 2(b). Assume P'_{11} overlap with P'_{21} first. We can find a set of coding coefficients such that

$$[\beta_1 \quad \beta_2] = [1\ 1]$$

$$[M_{21} \quad M_{22}] = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \tag{9}$$

By substituting them into eq. (11), the determinant of $[M_{21} \quad M_{22}\underline{\xi}]$ is not zero.

In both cases, therefore the required condition with hold w.h.p. under random linear coding. ■

Terminal t_2 can decode since it can solve the system of equations specified by eq. (7).

Decoding at t_3: At t_3, we need to decode X_3 in the presence of the interference from s_1 and s_2. The prior constraints on $\underline{\theta}$ (in the discussion above), namely (3) and (4) for case (i), or (3) and (5) for case (ii) allow at least $q^3 - 1$ choices for it. As M_{33} is full-rank, this implies that there are at least $q^3 - 1$ corresponding distinct $M_{33}\underline{\theta}$ vectors. Next, for t_3 to decode X_3, from Lemma A.4, we need to have

$$M_{33}\underline{\theta} \notin span([M_{31} \quad M_{32}\underline{\xi}]). \tag{10}$$

Since there are at most q^2 vectors in $span([M_{31} \quad M_{32}\underline{\xi}])$, there are at least $q^3 - q^2 - 1 > 0$ choices for $\underline{\theta}$ such that all the required constraints on $\underline{\theta}$ are satisfied.

Proof when there exists a subgraph G' that satisfies the conditions of Case 2:

As before, our overall strategy will be to use random linear network coding, however in certain cases we will need to make modifications to the code assignment. We argue based on the properties of the minimal structured subgraph G'. Specifically, through a sequence of arguments (for details see [16]), it is possible to show that G' is topologically equivalent to one of the graphs shown in Figs. 3(a), 3(b) and 3(c).

For the class of G' that fall in Fig. 3(a), it suffices to use the approach in the proof of Theorem 4.9. Namely, we use random linear coding in the network and precoding at sources s_2 and s_3. As in this case $M_{21} \neq 0$, one needs to argue that $rank[M_{21}\ M_{22}\underline{\xi}] = 2$. Following the line of argument used previously, we can do this by demonstrating a choice of local coding coefficients such that $[\beta_1 \ \ \beta_2] = [1\ 0]$ and $[M_{21} \ \ M_{22}] = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. However, such an approach does not work when the subgraph G' belongs to the class of graphs shown in Figs. 3(b) and 3(c). For instance, it is easy to observe that if we use random coding on Fig. 3(b), and precoding to cancel the X_2 component at t_1, then t_2 will receive a linear combination of X_1 and X_2 w.h.p., i.e., decoding X_2 at t_2 will fail. Accordingly, when G' looks like Fig. 3(b) or 3(c), we need to use a different network coding scheme that we now present.

Modified random coding for cases in Fig 3(b) and Fig 3(c).
It is clear that the strategy of random linear network coding and precoding at the sources fails since the determinant of the matrix $[M_{21}\ M_{22}\underline{\xi}]$ is identically zero for the cases in Fig. 3(b) and 3(c). Thus, at the top level our approach is to modify the original graph G by removing certain edges and identifying a special node in G that is upstream of t_2. The transfer matrix on the two incoming edges of this special node can be expressed as $[\tilde{M}_{21}\ \tilde{M}_{22}\ \tilde{M}_{23}]$ such that the determinant of $[\tilde{M}_{21}\ \tilde{M}_{22}\underline{\xi}]$ is not identically zero. Thus, at this node it becomes possible to remove the effect of X_1 via deterministic coding. Accordingly, our strategy is to first perform random linear coding at all nodes except the special node and those that are downstream of the special node. Following this, we perform deterministic coding at the special node to cancel the effect of X_1, and random linear coding downstream of it. Finally, we argue based on the precoding constraints that each terminal can decode its desired message. Owing to space limitations we are unable to include a detailed proof here (it can be found in [16]). ■

V. Conclusions and Future Work

In this work we mainly considered three-source, three-terminal network coding based multiple unicast for directed acyclic networks with unit capacity edges. Our focus was on characterizing the feasibility of achieving unit-rate transmission for each session based on the knowledge of the connectivity level vector. For the infeasible instances we have demonstrated specific network topologies where communicating at unit-rate is impossible, while for the feasible instances we have designed constructive linear network coding schemes that satisfy the demands of each terminal. Our schemes are non-asymptotic and require vector network coding over at most two time units. Our work leaves out one specific connectivity level vector, namely $[1\ 2\ 4]$ for which we have been unable to

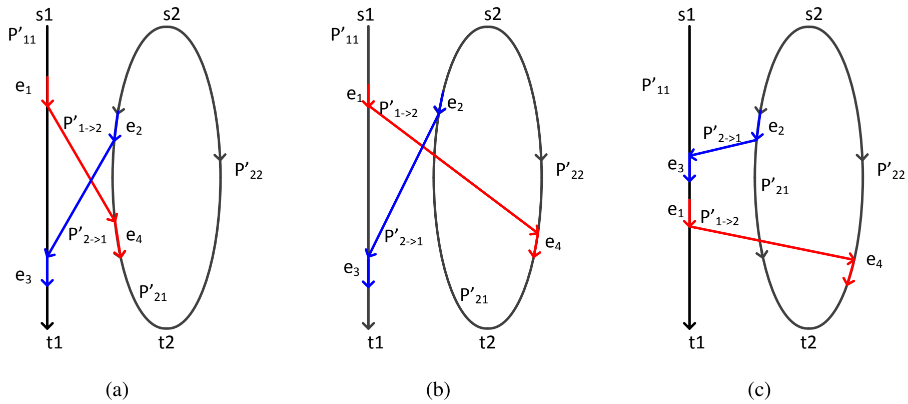


Fig. 3. (a) (b) (c) Subgraph G' when P'_{11} does not overlap with either P'_{21} or P'_{22}.

provide either a feasible network code or a network topology where communicating at unit rate is impossible. Our work is potentially useful for arriving at multiple unicast schemes for arbitrary rates as in these cases one could pack unit-rate structures for which our constructive schemes apply.

Appendix

The statements here are stated without proof owing to space limitations. Detailed arguments can be found in [16].

Lemma A.1: Consider a minimal multiple unicast instance, $< G, \{s_1-t_1, s_2-t_2\} >$ with connectivity level $[1\ m]$. Denote the $s_1 - t_1$ path by P_1 and the set of edge disjoint $s_2 - t_2$ paths by $\{P_{21}, \ldots, P_{2m}\}$. There can be at most one overlap segment between P_1 and each $P_{2i}, i = 1, \ldots, m$.

Lemma A.2: A minimal multiple unicast instance $< G, \{s_1-t_1, s_2-t_2\}, \{1, m\} >$ with connectivity level $[1\ m+1]$ is always feasible.

Lemma A.3: If $\beta_1 \neq 0$, $det([M_{21} \quad M_{22}\underline{\xi}])$ can be represented by

$$\frac{\xi_2}{\beta_1}\det\left[\begin{array}{cc} \alpha'_1 & -\beta_2\beta'_{11} + \beta_1\beta'_{12} \\ \alpha'_2 & -\beta_2\beta'_{21} + \beta_1\beta'_{22} \end{array}\right]. \quad (11)$$

where $\underline{\xi}$ satisfies $[\beta_1\ \beta_2]\underline{\xi} = 0$.

Lemma A.4: Consider a system of equations $Z = H_1X_1 + H_2X_2$, where X_1 is a vector of length l_1 and X_2 is a vector of length l_2 and $Z \in span([H_1 \quad H_2])^2$. The matrix H_1 has dimension $z_t \times l_1$, and rank $l_1 - \sigma$, where $0 \leq \sigma \leq l_1$. The matrix H_2 is full rank and has dimension $z_t \times l_2$ where $z_t \geq (l_1 + l_2 - \sigma)$. Furthermore, the column spans of H_1 and H_2 intersect only in the all-zeros vectors, i.e. $span(H_1) \cap span(H_2) = \{0\}$. Then there exists a unique solution for X_2.

[2] $span(A)$ refers to the column span of A.

References

[1] R. Dougherty, C. Freiling, and K. Zeger, "Insufficiency of linear coding in network information flow," *IEEE Trans. on Info. Th.*, vol. 51, no. 8, pp. 2745 – 2759, 2005.

[2] D. Traskov, N. Ratnakar, D. Lun, R. Koetter, and M. Medard, "Network Coding for Multiple Unicasts: An Approach based on Linear Optimization," in *IEEE Intl. Symp. on Info. Th.*, 2006, pp. 1758–1762.

[3] C.-C. Wang and N. B. Shroff, "Pairwise Intersession Network Coding on Directed Networks," *IEEE Trans. on Info. Th.*, vol. 56, no. 8, pp. 3879–3900, 2010.

[4] N. Harvey, R. Kleinberg, and A. Lehman, "On the capacity of information networks," *IEEE Trans. on Info. Th.*, vol. 52, no. 6, pp. 2345–2364, 2006.

[5] Z. Li and B. Li, "Network coding: the Case of Multiple Unicast Sessions," in *42st Allerton Conference on Communication, Control, and Computing*, 2004.

[6] X. Yan, R. W. Yeung, and Z. Zhang, "The Capacity Region for Multi-source Multi-sink Network Coding," in *IEEE Intl. Symp. on Info. Th.*, 2007, pp. 116–120.

[7] T. Ho, Y. Chang, and K. J. Han, "On constructive network coding for multiple unicasts," in *44th Allerton Conference on Communication, Control and Computing*, 2006.

[8] S. Shenvi and B. K. Dey, "A Simple Necessary and Sufficient Condition for the Double Unicast Problem," in *IEEE Intl. Conf. Comm.*, 2010, pp. 1–5.

[9] E. Erez and M. Feder, "Improving the Multicommodity Flow Rate with Network Codes for Two Sources," *IEEE J. Select. Areas Comm.*, vol. 27, no. 5, pp. 814–824, 2009.

[10] S. U. Kamath, D. N. C. Tse, and V. Anantharam, "Generalized Network Sharing Outer Bound and the Two-Unicast Problem," in *Proceedings of Netcod*, 2011, pp. 1–6.

[11] J. Price and T. Javidi, "Network Coding Games with Unicast Flows," *IEEE J. Select. Areas Comm.*, vol. 26, no. 7, pp. 1302–1316, 2008.

[12] A. Das, S. Vishwanath, S. A. Jafar, and A. Markopoulou, "Network Coding for Multiple Unicasts: An Interference Alignment Approach," in *IEEE Intl. Symp. on Info. Th.*, 2010, pp. 1878 – 1882.

[13] A. Ramakrishnan, A. Das, H. Maleki, A. Markopoulou, S. Jafar, and S. Vishwanath, "Network coding for three unicast sessions: Interference alignment approaches," in *48th Allerton Conference on Communication, Control and Computing*, 2010, pp. 1054 – 1061.

[14] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the K user interference channel," *IEEE Trans. on Info. Th.*, vol. 54, no. 8, pp. 3425–3441, 2008.

[15] J. Han, C.-C. Wang, and N. Shroff, "Analysis of Precoding-based Intersession Network Coding and The Corresponding 3-Unicast Interference Alignment Scheme," in *49th Allerton Conference on Communication, Control and Computing*, 2011, pp. 1033–1040.

[16] S. Huang and A. Ramamoorthy, "On the multiple unicast capacity of 3-source, 3-terminal directed acyclic networks," *preprint*, 2012 [Online] *Available: http://www.ece.iastate.edu/˜adityar/publications.htm*.

[17] R. Koetter and M. Médard, "An Algebraic approach to network coding," *IEEE/ACM Trans. on Netw.*, vol. 11, no. 5, pp. 782–795, 2003.

[18] S. Huang and A. Ramamoorthy, "A note on the multiple unicast capacity of directed acyclic networks," in *IEEE Intl. Conf. Comm.*, 2011, pp. 1–6.

[19] M. Langberg, A. Sprintson, and J. Bruck, "The encoding complexity of network coding," *IEEE Trans. on Info. Th.*, vol. 52, no. 6, pp. 2368–2397, 2006.