Network Coding Capacity Region of Line Networks with Node and Edge Constraints

S. M. Sadegh Tabatabaei Yazdi*, Serap A. Savari† and Gerhard Kramer‡

Abstract

The network coding capacity region of line networks is established when there are both node and edge constraints. Cut set bounds and progressive d-separating edge set bounds provide outer bounds while a linear network coding scheme achieves capacity.

1. INTRODUCTION

Network coding lets network processors decode received messages and re-encode them together with their own messages before sending packets to adjacent processors. We represent a network as a graph where nodes represent the processors and edges represent communication channels between processors. The outgoing symbol from every node, at any particular time instant n, can be any function of the incoming symbols to that node at earlier time instants $1, 2, \cdots, n-1$ and/or its own messages at the present and earlier time instants. It is well known [1] that for some networks the use of network coding increases throughput compared to routing where processors only store and forward their generated and received messages to other processors. Ahlswede et al. [1] have shown that network coding achieves the min-cut rate for a multicast session from a source node to several sink nodes. By min-cut we mean the minimum total capacity of the edges that disconnect the source node from at least one sink node. Li et al. [14] showed that linear network coding suffices to achieve the optimal rate. In this paper we study the benefits of network coding in line networks, that appear in many different communication systems such as peer to peer communication networks, and wireless ad-hoc and sensor networks (see, e.g., [12]). Line networks are also the basic elements in more complicated network topologies such as tree networks. The graph of a line network with only

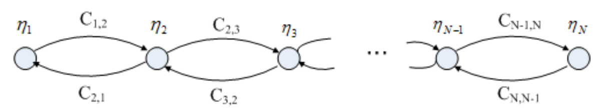

Figure 1. A simple model for a line network with N communication units.

edge constraints can be depicted as a cascaded series of nodes where every pair of consecutive nodes share two oppositely directed edges (see Fig. 1). In this model, the only constraints on the data transmission rate between different nodes are the edge capacities indicated by $C_{i,i+1}$ for the edge from node i to $i+1$, $1 \leq i \leq N-1$, and $C_{j,j-1}$ for the edge from node j to $j-1$, $2 \leq j \leq N$. By applying the cut set bounds introduced in [4], [7] on the edges of this graph, one finds that routing is throughput-optimal for the transmission of multiple multicast sessions with independent message sources, where all nodes can simultaneously have multicast sessions to all subsets of the remaining nodes. In practical networks, however, it often happens that, in addition to the edge constraints, the nodes are also restricted in terms of the maximum amount of data that they are able to process in a certain amount of time. We refer to these latter restrictions as *node constraints*, and they might be due to the limited bus bandwidth between different processing or memory units and/or the limited speed of processors.

One way to model a node constraint is to convert it into edge constraints on a new network topology (see [5], [6]) in order to be able to use well-known analysis techniques for edge constrained networks. The process is illustrated in Fig. 2: Split node η_j in the original network into two nodes I_j and O_j in the new network. Connect all incoming edges to η_j in the original network to I_j and replace all outgoing edges from η_j in the original network with outgoing edges from O_j in the revised network graph. Next, in the new network create an edge directed from I_j to O_j with capacity C_j to model the node capacity for processing information at node η_j. Assume that all messages generated at η_j in the original network are generated at I_j for the revised network and that all messages that are originally decoded at η_j are now decoded at O_j. Note that all messages which are processed at node η_j pass through the edge between I_j and O_j

*S. M. S Tabatabaei Yazdi is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor MI 48109, USA, sadegh@umich.edu

†S. A. Savari is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor MI 48109, USA, savari@eecs.umich.edu

‡G. Kramer is with Bell Labs, Alcatel-Lucent, Murray Hill NJ 07974 USA, gkr@bell-labs.com

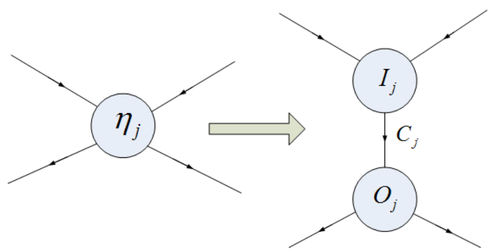

Figure 2. Converting a node constraint to an edge constraint.

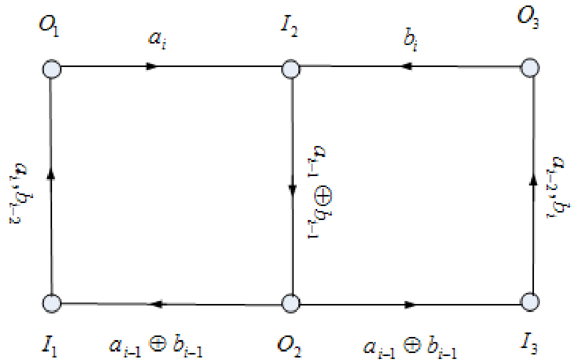

Figure 3. Network coding example in a line network with three processors.

in the revised network without affecting other parts of the network model (alternative models have the messages originating at O_j or decoded at I_j; the choice of model depends on the application). In the following example, we demonstrate that the capacity region of a line network subject to both node and edge constraints is in general larger than its routing capacity region. Later in the paper we establish that a linear network coding scheme achieves any point in the capacity region.

Example 1.1 *Consider the graph corresponding to a line network with three processors which is depicted in Fig. 3. Suppose that node I_1 wants to send the bit sequence $\mathbf{a} = [a_1, a_2, a_3, \cdots]$ with a rate of one bit per second to destination O_3 and node I_3 wants to send the independent bit sequence $\mathbf{b} = [b_1, b_2, b_3, \cdots]$ with a rate of one bit per second to destination O_1. Both nodes begin transmission at time instant zero. In a routing solution for this communication problem, the bit sequence of* **a** *passes through the path $I_1, O_1, I_2, O_2, I_3, O_3$. Similarly, the bit sequence* **b** *passes through the path $I_3, O_3, I_2, O_2, I_1, O_1$. Thus, in a routing solution all node capacities should be at least two bits per second since the nodes are used by both flows, and the remaining edge capacities on the graph should be at least one bit per second since they are used by either one of the flows. We next show that routing is not optimal for this setting. Consider a network coding scheme in which edge (I_2, O_2) passes the bitwise XOR of its incoming bit sequences, i.e., $[a_1 \oplus b_1, a_2 \oplus b_2, \cdots]$. Further, node O_2 passes its incoming bit sequence along edges (O_2, I_1) and (O_2, I_3). The scheme is shown on the graph of Fig. 3. We assume that the delay between I_i and O_i for transferring information is negligible and thus I_1 and I_3 start to receive the bit sequence $[a_1 \oplus b_1, a_2 \oplus b_2, \cdots]$ from node O_2 with one second delay with respect to the start time of zero. Next I_1 and I_3 respectively decode the bit sequences* **b** *and* **a** *by using the received bit sequence and their own messages. Finally I_1 and I_3 respectively send the bit sequences* **b** *and* **a** *to O_3 and O_1 via the edges (I_1, O_1) and (I_3, O_3). This network coding scheme shows that we can save one bit per second of capacity on edge (I_2, O_2) over the routing solution while keeping the other capacities fixed. In Sections III and IV we will prove that this network coding scheme is optimal in the sense that the capacities can not be reduced further for these demands.*

We review earlier work on line network topologies. In [13] the capacity of bidirected ring networks with multiple unicast sessions is derived. Since finite-length line networks are special cases of ring networks, the multiple unicast session problem for bidirected (or undirected) line networks is also solved. In [2] the authors investigate the network coding capacity of unidirected line networks with edge constraints only for several cases of independent and dependent message sources. In their analysis, they decompose the network into several components with a single demand node in each component and assume that all previous node demands are satisfied in each component. Then they show that the sum of feasible rates of the components is achievable in the parent network for a broad class of demands. The multimessage multicast capacity of the line networks that they consider is derived as a special case of more general results. For independent message sources one can show that the bidirected cut set bounds introduced in [4], [7] establish that routing achieves capacity for multimessage multicast sessions on undirected or bidirected line networks. In [10], [11] the authors consider a cascade of Discrete Memoryless Channels (DMCs) with identical capacities and discuss the network coding benefits when intermediate nodes can only process fixed length information blocks; these papers provide the relationship between the code block length and the size of the network for a constant end-to-end rate. In [12] the authors show that network coding schemes with a finite field size can achieve the maximum possible network coding capacity for cascaded erasure channels with a single source and a single destination.

The rest of this paper is organized as follows. In Section II we formally define the problem and the mathematical model. We consider line networks with both node and

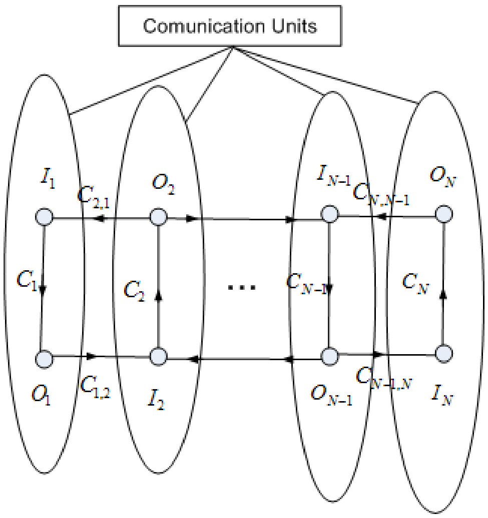

Figure 4. A general line network with N nodes.

edge constraints and allow every node to send multicast messages to every other nonempty collection of nodes in the network. In Section III we use information theoretic tools to find outer bounds on the network coding capacity region. Finally, in Section IV we propose a network coding scheme that achieves the outer bounds of Section III.

2. PROBLEM DEFINITION

Consider a cascaded series of N nodes $\eta_1, \eta_2, \cdots, \eta_N$ forming a line network as in Fig. 1. We convert a line network with node constraints into one with only edge constraints as illustrated by Fig. 4. The node constraint for node η_j is converted into an edge of capacity C_j between nodes I_j and O_j, and we call this pair of nodes and the edge connecting them *communication unit* j, $j \in \{1, 2, \cdots, n\}$. Communication unit j shares Discrete Memoryless Channels (DMCs) with units $j-1$ and $j+1$; i.e., in the revised network node O_j has outgoing edges to nodes I_{j-1} and I_{j+1} with capacities $C_{j,j-1}$ and $C_{j,j+1}$, respectively. All units can potentially communicate multicast messages to multiple subsets of the $N-1$ remaining units on the line. Let $A_k = i_k \rightarrow \{j_{k_1}, j_{k_2}, \cdots, j_{k_{L_k}}\}$ denote a traffic session and W_{A_k} its corresponding message; here W_{A_k}, is generated at node I_{i_k} and is destined for nodes in the set $\{O_{j_{k_1}}, O_{j_{k_2}}, \cdots, O_{j_{k_{L_k}}}\}$. Let $\hat{W}^j_{A_k}$ represent the estimate of message W_{A_k} at one of its destinations O_j. Assume that $j_{k_1} < j_{k_2} < \cdots < j_{k_{L_k}}$. Let $\mathscr{R}_{A_k}$ be the rate of session A_k.

We wish to characterize all feasible rate vectors in a line network with N units with subject to the node and edge constraints. In the next section we employ the graphical method of the progressive d-separating edge set (PdE) bounds introduced in [8], [9] and the cut set bounds [4], [7] to the network of Fig. 4 to find outer bounds on the set of feasible rate vectors. In the last section, we develop a simple linear coding scheme on the binary field which achieves these bounds.

3. NETWORK CODING BOUNDS

In this section, we find upper bounds on the achievable rates in line networks. We first use the PdE edge set bounds to find upper bounds on the feasible rates subject to the node constraints. We next use cut set bounds to find upper bounds on the feasible rates subject to the edge constraints.

3.1. Bounds from the Node Constraints

Consider the revised graph of a general line network in Fig. 4, and let $e_{j,j+1}$ and $e_{j,j-1}$ denote the edges from node O_j to nodes I_{j+1} and I_{j-1}, respectively. Let e_j denote the edge on the graph from node I_j to O_j. Like the cut set bounds, PdE bounds begin with a set of edges ε_d; however, they also need to satisfy a verification procedure. We choose to study $\varepsilon_d = e_i$ and the following set of traffic sessions S_d

- $S_{d_1} = \{A_k : i \in \{j_{k_1}, \cdots, j_{k_L}\}\}$
- $S_{d_2} = \{A_k : i \notin \{i_k, j_{k_1}, \cdots, j_{k_L}\}, i_k < i < j_{k_L}\}$
- $S_{d_3} = \{A_k : i = i_k\}$
- $S_d = S_{d_1} \cup S_{d_2} \cup S_{d_3}$
- $S_d^C = \{A_k : A_k \notin S_d\}$.

The PdE algorithm also takes as input an ordering of the sessions in S_d. We first assign three arbitrarily chosen ordering functions to the subsets of sessions $S_{d_1}, S_{d_2}, S_{d_3}$, so that we have:

$$S_{d_1} = \{S_{d_{1,1}}, S_{d_{1,2}}, \cdots, S_{d_{1,|S_{d_1}|}}\}$$
$$S_{d_2} = \{S_{d_{2,1}}, S_{d_{2,2}}, \cdots, S_{d_{2,|S_{d_2}|}}\}$$
$$S_{d_3} = \{S_{d_{3,1}}, S_{d_{3,2}}, \cdots, S_{d_{3,|S_{d_3}|}}\}.$$

We consider the multicast sessions in S_d in the order

$$S_{d_{1,1}}, S_{d_{1,2}}, \cdots, S_{d_{1,|S_{d_1}|}}, S_{d_{2,1}}, S_{d_{2,2}}, \cdots, S_{d_{2,|S_{d_2}|}},$$
$$S_{d_{3,1}}, S_{d_{3,2}}, \cdots, S_{d_{3,|S_{d_3}|}}.$$

Example 3.1 *We use a small line network with 3 communication units to help illustrate the PdE procedure below. We choose to study $\varepsilon_d = e_2$ and the following set of traffic sessions S_d*

- $S_{d_1} = \{A_k : 2 \in \{j_{k_1}, \cdots, j_{k_L}\}\}$
- $S_{d_2} = \{A_k : 2 \notin \{i_k, j_{k_1}, \cdots, j_{k_L}\}, i_k < 2 < j_{k_L}\}$
- $S_{d_3} = \{A_k : 2 = i_k\}$
- $S_d = S_{d_1} \cup S_{d_2} \cup S_{d_3}$
- $S_d^C = \{A_k : A_k \notin S_d\}$.

For our network and node constraint, the steps of the PdE procedure are as follows (see [9]):

1. (Initialization) Consider the Functional Dependence Graph (FDG) of the network illustrated in Fig. 5, which is the line graph (see, e.g., [3] for a definition) of the network with the addition of nodes representing the messages, their estimates, and noise. In this graph, we represent the inputs to the edges $e_{j,j-1}, e_{j,j+1}$ and e_j of the original network by the random variables $X_{j,j-1}, X_{j,j+1}$ and X_j respectively, and their outputs by the random variables $Y_{j,j-1}, Y_{j,j+1}$ and Y_j. We assume that each edge output is a function of the edge input and the corresponding noise random variable; we represent the noise random variables $Z_{j,j-1}, Z_{j,j+1}$, and Z_j on the FDG with solid circles. We simplify the FDG of Fig. 5 by collecting at every communication unit j, all messages that belong to sessions in $S_{d_1}, S_{d_2}, S_{d_3}$ and S_d^C and respectively representing them by $W_{j,1}, W_{j,2}, W_{j,3}$ and W_j^C. We likewise use the notation $\hat{W}_{j,1}, \hat{W}_{j,2}, \hat{W}_{j,3}$ and $\hat{W}_j^C$ for the respective estimates of the messages. More specifically, we define

$$W_{j,l} = \{W_{A_k} : j = i_k, A_k \in S_{d_l}\}, \qquad l = 1,2,3$$
$$W_j^C = \{W_{A_k} : j = i_k, A_k \in S_d^C\}$$
$$\hat{W}_{j,l} = \{\hat{W}_{A_k}^j : j \in \{j_{k_1}, \cdots, j_{k_{L_k}}\}, A_k \in S_{d_l}\}, \quad l = 1,2,3$$
$$\hat{W}_j^C = \{\hat{W}_{A_k}^j : j \in \{j_{k_1}, \cdots, j_{k_{L_k}}\}, A_k \in S_d^C\}.$$

(The FDG for Example 3.1 is illustrated in Fig. 10.)

- We remove all nodes and edges in the FDG except those encountered when moving backward one or more edges starting from any of the nodes representing: (1) Y_i and Z_i, (2) all messages W_{A_k} for all sessions, and (3) any choice of nonempty subset of $\{\hat{W}_{A_k}^p : p = j_{k_1}, j_{k_2}, \cdots, j_{k_{L_k}}\}$ for all $A_k \in S_d$; we choose this subset to be $\hat{W}_{i,1}$, $\hat{W}_{j,2}$ for $j = i+1, i+2, \cdots, N$ and $\hat{W}_{j,3}$ for $j = 1,2,\cdots,i-1,i+1,\cdots,N$. This choice guarantees that for every session $A_k \in S_d$ we have chosen at least one of its destinations, because (1) if $A_k \in S_{d_1}$, then $i \in \{j_{k_1}, \cdots, j_{k_{L_k}}\}$, and hence $\hat{W}_{A_k}^i \in \hat{W}_{i,1}$ is nonempty, (2) if $A_k \in S_{d_2}$, then there is an r with $r \in \{j_{k_1}, \cdots, j_{k_{L_k}}\}$ and $i < r$, thus $\hat{W}_{A_k}^r \in \{\hat{W}_{i+1,2}, \hat{W}_{i+2,2}, \cdots, \hat{W}_{N,2}\}$ is nonempty, and (3) for every $A_k \in S_{d_3}$, then there is a t with $t \in \{j_{k_1}, \cdots, j_{k_{L_k}}\}$ and $t \neq i$; thus $\hat{W}_{A_k}^t = \{\hat{W}_{1,3}, \cdots, \hat{W}_{i-1,3}, \hat{W}_{i+1,3}, \cdots, \hat{W}_{N,3}\}$ is nonempty. The resulting FDG is shown in Fig. 6. (The resulting FDG for Example 3.1 is illustrated in Fig. 11.)
- Further we remove the edges coming out of the nodes on the FDG representing Y_i, Z_i and W_j^C for $j = 1,2,\cdots,N$, and successively remove all edges coming out of the nodes which are disconnected from any source nodes in a directed sense. The resulting graph is shown in Fig. 7. (The resulting graph for Example 3.1 is illustrated in Fig. 12.)

2. (Iterations) Since for all $A_k \in \{S_{d_{1,1}}, S_{d_{1,2}}, \cdots, S_{d_{1,|S_{d_1}|}}\}$ W_{A_k} is disconnected from all of its estimates,
   - we remove the edges coming out of $W_{j,1}$ for $j = 1,2,\cdots,N$, and
   - successively remove all edges coming out of the nodes which are disconnected from any source nodes in a directed sense. The resulting graph is depicted in Fig. 8. (The resulting graph for Example 3.1 is illustrated in Fig. 13.)

3. Since for all $A_k \in \{S_{d_{2,1}}, S_{d_{2,2}}, \cdots, S_{d_{2,|S_{d_2}|}}\}$ W_{A_k} is disconnected from all of its estimates,
   - we remove the edges coming out of $W_{j,2}$ for $j = 1,2,\cdots,N$, and
   - we successively remove all edges coming out of the nodes which are disconnected from any source nodes in a directed sense. The resulting graph is depicted in Fig. 9. (The resulting graph for Example 3.1 is illustrated in Fig. 14.)

4. Finally, since for all $A_k \in \{S_{d_{3,1}}, S_{d_{3,2}}, \cdots, S_{d_{3,|S_{d_3}|}}\}$ W_{A_k} is disconnected from all of its estimates, we obtain the following PdE bound:

$$\sum_{A_k \in S_d} \mathscr{R}_{A_k} \leq C_i, \tag{1}$$

or equivalently,

$$\sum_{A_j \in S_{d_1}} \mathscr{R}_{A_j} + \sum_{A_k \in S_{d_2}} \mathscr{R}_{A_k} + \sum_{A_l \in S_{d_3}} \mathscr{R}_{A_l} \leq C_i. \tag{2}$$

Observe that we can use symmetry to establish another PdE bound corresponding to the sets of sessions $S'_{d_1} = S_{d_1}, S'_{d_2} =$

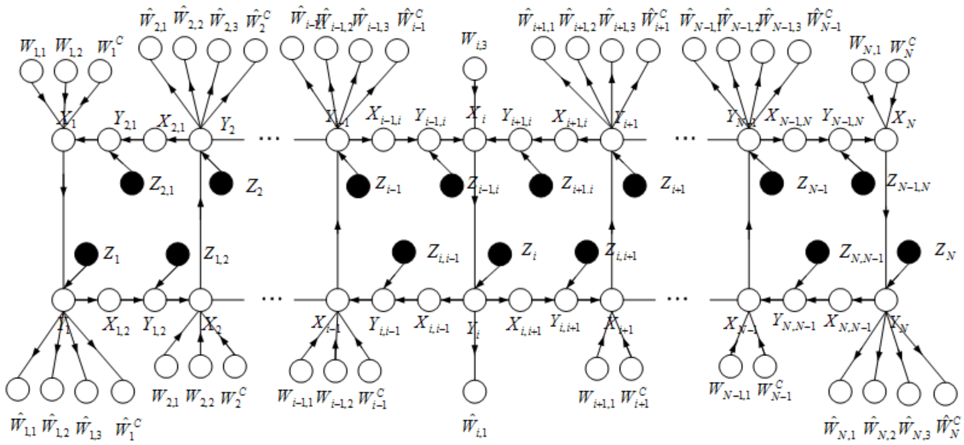

Figure 5. The Functional Dependence Graph (FDG) for the network in Fig. 4.

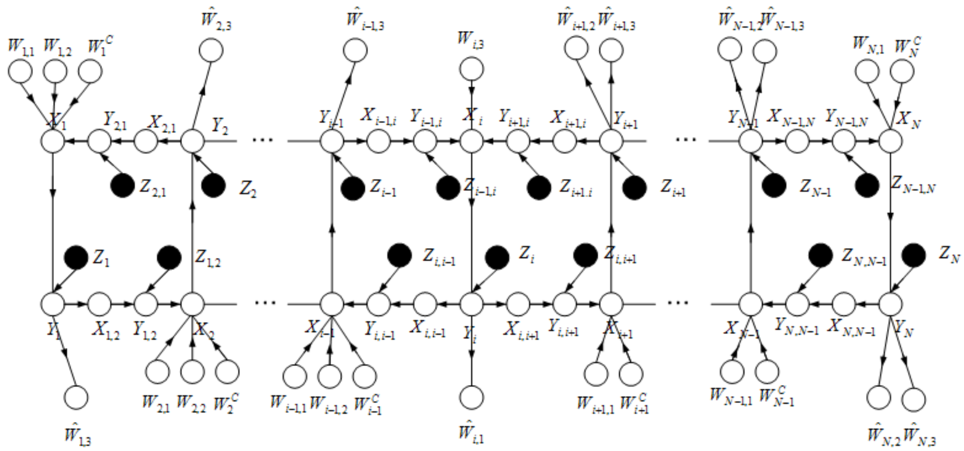

Figure 6. The resulting graph after the first step of the initialization for the graph of Fig. 5.

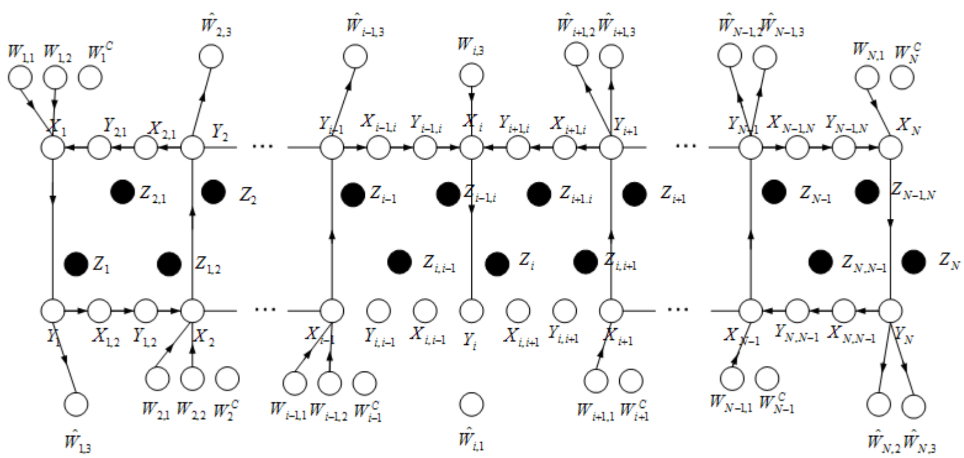

Figure 7. The resulting graph after the second step of the initialization for the graph of Fig. 6.

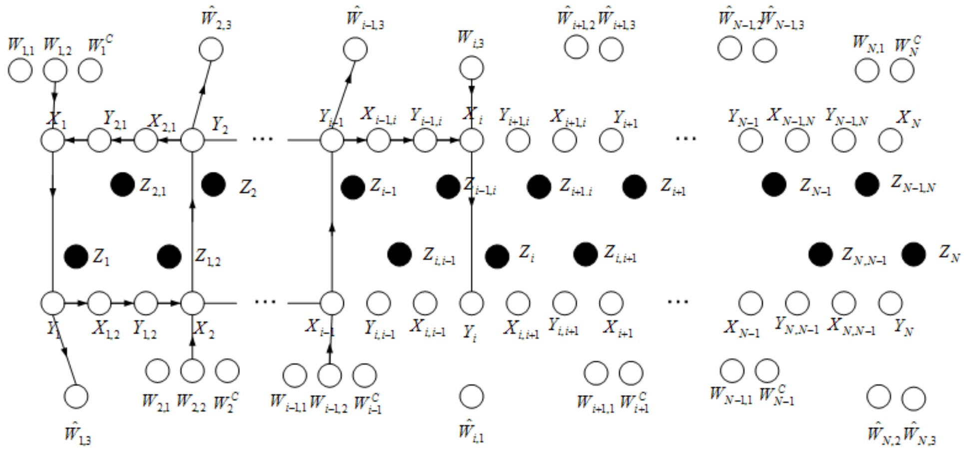

Figure 8. The resulting graph after the first iteration of the PdE procedure for the graph of Fig. 7.

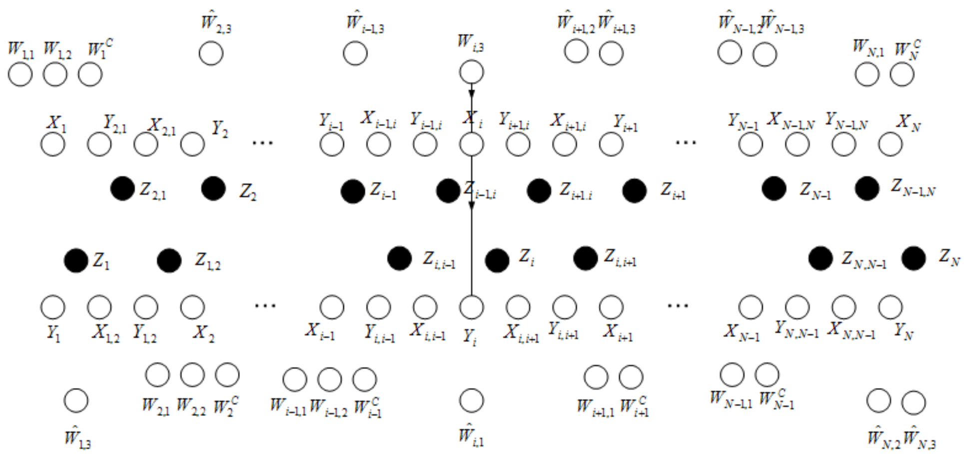

Figure 9. The resulting graph after the second iteration of the PdE procedure for the graph of Fig. 8.

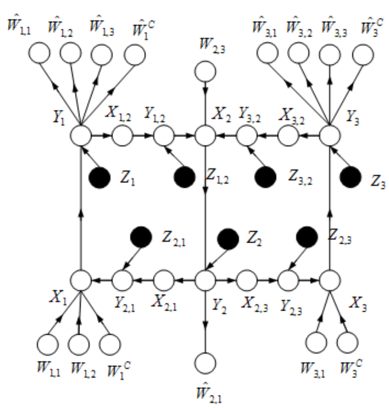

Figure 10. The Functional Dependence Graph (FDG) for the network in Fig. 4, with $N = 3$ and $i = 2$.

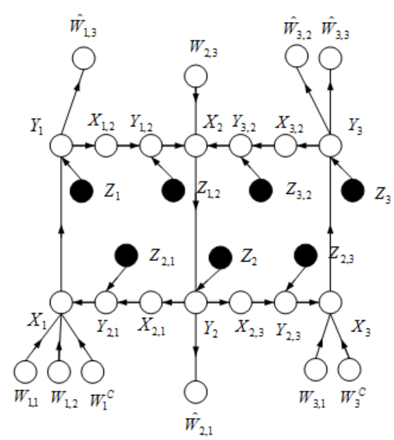

Figure 11. The resulting graph after the first step of the initialization for the graph of Fig. 10.

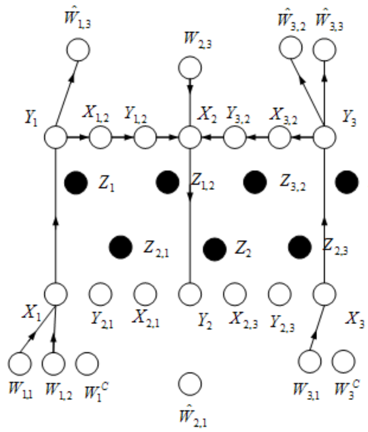

Figure 12. The resulting graph after the second step of the initialization for the graph of Fig. 11.

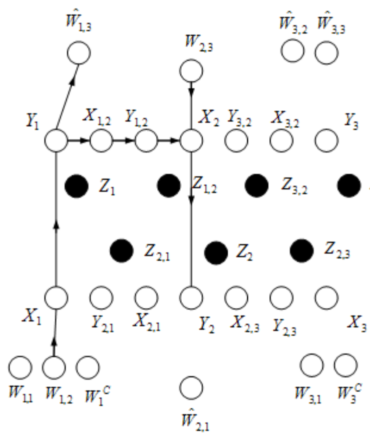

Figure 13. The resulting graph after the first iteration of the PdE procedure for the graph of Fig. 12.

$\{A_k : i \notin \{j_{k_1}, j_{k_2}, \cdots, j_{k_{L_k}}\}, j_{k_1} < i < i_k\}, S'_{d_3} = S_{d_3}, S'_d = S'_{d_1} \cup S'_{d_2} \cup S'_{d_3}$:

$$\sum_{A_j \in S'_{d_1}} \mathscr{R}_{A_j} + \sum_{A_k \in S_{d'_2}} \mathscr{R}_{A_k} + \sum_{A_l \in S_{d'_3}} \mathscr{R}_{A_l} \leq C_i. \tag{3}$$

The bounds (2) and (3) imply

$$\sum_{A_j \in S_{d_1}} \mathscr{R}_{A_j} + \max\{\sum_{A_k \in S_{d_2}} \mathscr{R}_{A_k}, \sum_{A_k \in S_{d'_2}} \mathscr{R}_{A_k}\} + \sum_{A_l \in S_{d_3}} \mathscr{R}_{A_l} \leq C_i. \tag{4}$$

3.2. Bounds from the Edge Constraints

The cut set bound starts with a subset ε of directed edges of the network graph that partition the set of nodes such that any path between two nodes in different subsets includes

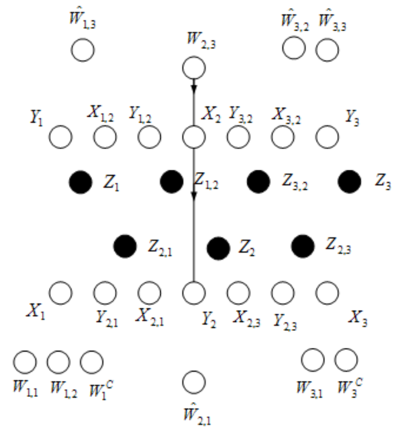

Figure 14. The resulting graph after the second iteration of the PdE procedure for the graph of Fig. 13.

at least one edge in ε. Let $C_{U \to U^c}$ denote the sum of the capacities of the edges in the cut that are directed from node subset U to its complement. For any subset S of nodes, the corresponding cut set bounds are:

$$\sum_{i_k \in S, \{j_{k_1}, \cdots, j_{k_{L_k}}\} \cap S^c \neq \phi} \mathscr{R}_{A_k} \leq C_{S \to S^c} \tag{5}$$

$$\sum_{i_k \in S^c, \{j_{k_1}, \cdots, j_{k_{L_k}}\} \cap S \neq \phi} \mathscr{R}_{A_k} \leq C_{S^c \to S}. \tag{6}$$

We select $\varepsilon = \{e_{i,i-1}, e_{i-1,i}\}$ for some i with $2 \leq i \leq N$ and obtain the following two cut set bounds:

$$\sum_{i_k \geq i, j_{k_1} < i} \mathscr{R}_{A_k} \leq C_{i,i-1}, \quad 2 \leq i \leq N \tag{7}$$

$$\sum_{i_k < i, j_{k_{L_k}} \geq i} \mathscr{R}_{A_k} \leq C_{i-1,i}, \quad 2 \leq i \leq N. \tag{8}$$

It is convenient to rewrite (8) as

$$\sum_{i_k \leq i, j_{k_{L_k}} > i+1} \mathscr{R}_{A_k} \leq C_{i,i+1}, \quad 1 \leq i \leq N-1. \tag{9}$$

Many of the results here generalize to undirected networks by using bidirected cut set bounds introduced in [7]. It should also be interesting to study networks with undirected edge constraints and other models for source and sink placements.

4. NETWORK CODING SCHEME

In this section we provide a network coding scheme to achieve the bounds of the last section. We begin by introducing the following notation for different collections of multicast or unicast messages with respect to some fixed communication unit i.

- Let W_i^n be the binary representation of the set of multicast messages originating at node I_i at time instant n with at least one destination O_{j_1} with $j_1 < i$ and at least one destination O_{j_2} with $j_2 > i$.

- Let $W_i^{L,n}$ be the binary representation of the set of multicast or unicast messages originating at node I_i at time instant n with all destination nodes O_j satisfying $j < i$.

- Let $W_i^{R,n}$ be the binary representation of the set of multicast or unicast messages originating at node I_i at time instant n with all destination nodes O_j having the property $j > i$.

- Let L_i^n be the binary representation of the set of all multicast messages generated at nodes I_j with $j < i$ which should be decoded at O_i and at least one node O_k with $k > i$, such that the message from I_j has been generated at time instant $n-(i-j)$.

- Let R_i^n be the binary representation of the set of all multicast messages generated at nodes I_j with $j > i$ which should be decoded at O_i and at least one node O_k with $k < i$, such that the message from I_j has been generated at time instant $n-(j-i)$.

- Let $\tilde{L}_i^n$ be the binary representation of the set of all multicast or unicast messages generated at nodes I_j with $j < i$ which should be decoded in at least one node O_k with $k > i$ but not at O_i, such that the message from I_j has been generated at time instant $n-(i-j)$.

- Let $\tilde{R}_i^n$ be the binary representation of the set of all multicast or unicast messages generated at nodes I_j with $j > i$ which should be decoded in at least one node O_k with $k < i$ but not at O_i, such that the message from I_j has been generated at time instant $n-(j-i)$.

- Let $\hat{L}_i^n$ be the binary representation of the set of all multicast or unicast messages generated at nodes I_j with $j < i$ which should be decoded at I_i but not any other node O_k with $k > i$, such that the message from I_j has been generated at time instant $n-(i-j)$.

- Let $\hat{R}_i^n$ be the binary representation of the set of all multicast or unicast messages generated at nodes I_j with $j > i$ which should be decoded at I_i but not any other node O_k with $k < i$, such that the message from I_j has been generated at time instant $n-(j-i)$.

Example 4.1 *Consider Fig. 15, which depicts a line network with four communication units. According to the above definitions we have the following set of messages with respect to communication unit 2 for the general multi-message multicast problem. Let $W_{A_k}^t$ represent the message corresponding to session A_k generated at time instant t.*

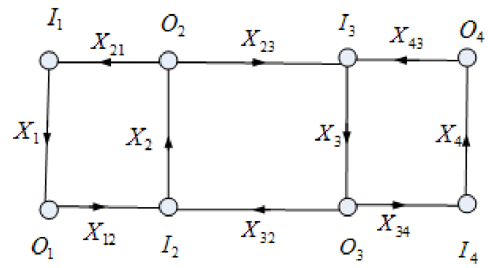

Figure 15. A line network with 4 communication units.

- $W_2^n = [W_{2\to\{1,3\}}^n, W_{2\to\{1,4\}}^n, W_{2\to\{1,3,4\}}^n]$
- $W_2^{L,n} = [W_{2\to\{1\}}^n]$
- $W_2^{R,n} = [W_{2\to\{3\}}^n, W_{2\to\{4\}}^n, W_{2\to\{3,4\}}^n]$
- $L_2^n = [W_{1\to\{2,3\}}^{n-1}, W_{1\to\{2,4\}}^{n-1}, W_{1\to\{2,3,4\}}^{n-1}]$
- $R_2^n = [W_{3\to\{2,1\}}^{n-1}, W_{3\to\{2,1,4\}}^{n-1}, W_{4\to\{2,1\}}^{n-2}, W_{4\to\{2,1,3\}}^{n-2}]$
- $\tilde{L}_2^n = [W_{1\to\{3\}}^{n-1}, W_{1\to\{4\}}^{n-1}, W_{1\to\{3,4\}}^{n-1}]$
- $\tilde{R}_2^n = [W_{3\to\{1\}}^{n-1}, W_{3\to\{1,4\}}^{n-1}, W_{4\to\{1\}}^{n-2}, W_{4\to\{1,3\}}^{n-2}]$
- $\hat{L}_2^n = [W_{1\to\{2\}}^{n-1}]$
- $\hat{R}_2^n = [W_{3\to\{2\}}^{n-1}, W_{3\to\{2,4\}}^{n-1}, W_{4\to\{2\}}^{n-2}, W_{4\to\{2,3\}}^{n-2}]$

To represent our network coding scheme we introduce two binary vector operators. Let $\mathbf{a} = [a_1, a_2, \cdots, a_{n_a}]$ and $\mathbf{b} = [b_1, b_2, \cdots, b_{n_b}]$ be two arbitrary binary vectors of lengths n_a and n_b respectively. Then we define:

$$\mathbf{a} \oplus \mathbf{b} = [a_1 \oplus b_1, a_2 \oplus b_2, \cdots, a_{n_a} \oplus b_{n_a}] \tag{10}$$

if $n_a \leq n_b$, and

$$\mathbf{a} \oplus \mathbf{b} = [a_1 \oplus b_1, \cdots, a_{n_b} \oplus b_{n_b}, a_{n_b+1}, \cdots, a_{n_a}] \tag{11}$$

if $n_a > n_b$. Here $a_i \oplus b_i$ is the XOR of bits a_i and b_i. Furthermore we define $\mathbf{a} \otimes \mathbf{b}$ as follows:

$$\mathbf{a} \otimes \mathbf{b} = \mathbf{a} \oplus \mathbf{b}, \qquad \text{if } n_a \geq n_b \tag{12}$$

$$\mathbf{a} \otimes \mathbf{b} = \mathbf{b} \oplus \mathbf{a}, \qquad \text{if } n_a < n_b. \tag{13}$$

Observe that the dimension of $\mathbf{a} \oplus \mathbf{b}$ is n_a while the dimension of $\mathbf{a} \otimes \mathbf{b}$ is $\max\{n_a, n_b\}$.

Suppose that all source nodes start transmitting at time $n=0$ and messages at negative time instants are assumed to take the value zero. Our network coding scheme consists of three parts, which respectively describe the vectors $X_i^n, X_{i,i-1}^n$ and $X_{i,i+1}^n$ for any communication unit i:

$$X_i^n = [\tilde{R}_i^n \otimes \tilde{L}_i^n, R_i^n, L_i^n, \hat{R}_i^n, \hat{L}_i^n, W_i^n, W_i^{R,n}, W_i^{L,n}] \quad (14)$$

$$X_{i,i-1}^n = [\tilde{R}_i^n \oplus \tilde{L}_i^n, R_i^n, W_i^{L,n}, W_i^n] \quad (15)$$

$$X_{i,i+1}^n = [\tilde{L}_i^n \oplus \tilde{R}_i^n, L_i^n, W_i^{R,n}, W_i^n]. \quad (16)$$

Next we demonstrate that for any node on the network, the outgoing messages from that node are some functions of the incoming messages to that node at earlier time instants and the messages generated at that node at earlier or present time instants. First consider X_i^n, which is the outgoing message from node I_i at time instant n. The messages available to node I_i at time instant n are $[X_{i-1,i}^t]_{t=1}^{n-1}, [X_{i+1,i}^t]_{t=1}^{n-1}, [W_i^t, W_i^{L,t}, W_i^{R,t}]_{t=1}^n$. As an intermediate step in our proof we will use induction to show that at time instant n node I_i is able to decode the following vectors of messages:

$$[\tilde{R}_i^n, \hat{R}_i^n, R_i^n] \quad (17)$$

$$[\tilde{L}_i^n, \hat{L}_i^n, L_i^n]. \quad (18)$$

Observe that at $n=0$, both (17), (18) are zero vectors and I_i has nothing to decode. For the inductive step assume that up to time instant $n-1$ node I_i has access to vectors $[\tilde{R}_i^t, \hat{R}_i^t, R_i^t]_{t=1}^{n-1}$ and $[\tilde{L}_i^t, \hat{L}_i^t, L_i^t]_{t=1}^{n-1}$. Our definitions imply the following relationships at time instant t:

$$[\tilde{R}_i^t, \hat{R}_i^t, R_i^t] \quad \text{is a permutation of} \quad [\tilde{R}_{i+1}^{t-1}, R_{i+1}^{t-1}, W_{i+1}^{t-1}, W_{i+1}^{L,t-1}] \quad (19)$$

$$[\tilde{L}_i^t, \hat{L}_i^t, L_i^t] \quad \text{is a permutation of} \quad [\tilde{L}_{i-1}^{t-1}, L_{i-1}^{t-1}, W_{i-1}^{t-1}, W_{i-1}^{R,t-1}]. \quad (20)$$

To see this, observe that the left hand side of (19) is the binary representation of the set of all messages generated at all nodes I_j with $j>i$ at time instant $t-(j-i)$ which have a destination at some node O_k with $k \le i$. This set of messages are either generated at I_j with $j > i+1$ at time instant of $t-(j-i)$ or at I_{i+1} at time instant $t-1$. By definition the former group of messages is identical to $[\tilde{R}_{i+1}^{t-1}, R_{i+1}^{t-1}]$, and the latter group of messages is identical to $[W_{i+1}^{t-1}, W_{i+1}^{L,t-1}]$; these together form the right hand side of (19). An analogous argument holds for (20). By setting $t=n$ in (19) and (20) we see that in order for I_i to be able to decode (17) and (18), it is sufficient for it to decode $[\tilde{R}_{i+1}^{n-1}, R_{i+1}^{n-1}, W_{i+1}^{n-1}, W_{i+1}^{L,n-1}]$ and $[\tilde{L}_{i-1}^{n-1}, L_{i-1}^{n-1}, W_{i-1}^{n-1}, W_{i-1}^{R,n-1}]$. At time instant n node I_i has access to the vectors $X_{i-1,i}^{n-1}$ and $X_{i+1,i}^{n-1}$. (15) and (16) imply

$$X_{i+1,i}^{n-1} = [\tilde{R}_{i+1}^{n-1} \oplus \tilde{L}_{i+1}^{n-1}, R_{i+1}^{n-1}, W_{i+1}^{L,n-1}, W_{i+1}^{n-1}] \quad (21)$$

$$X_{i-1,i}^{n-1} = [\tilde{L}_{i-1}^{n-1} \oplus \tilde{R}_{i-1}^{n-1}, L_{i-1}^{n-1}, W_{i-1}^{R,n-1}, W_{i-1}^{n-1}]. \quad (22)$$

Hence I_i can extract messages $[R_{i+1}^{n-1}, W_{i+1}^{n-1}, W_{i+1}^{L,n-1}]$ and $[L_{i-1}^{n-1}, W_{i-1}^{n-1}, W_{i-1}^{R,n-1}]$ directly from the received messages $X_{i-1,i}^{n-1}$ and $X_{i+1,i}^{n-1}$. The two remaining messages that I_i needs to decode are $\tilde{R}_{i+1}^{n-1}$ and $\tilde{L}_{i-1}^{n-1}$, and we next describe the process to do this.

By our inductive hypothesis at $t=n-2$ node I_i knows the message vectors $[\tilde{R}_i^{n-2}, \hat{R}_i^{n-2}, R_i^{n-2}]$ and $[\tilde{L}_i^{n-2}, \hat{L}_i^{n-2}, L_i^{n-2}]$. Observe that node I_i knows message vector $[W_i^{n-2}, W_i^{L,n-2}, W_i^{R,n-2}]$ at time instant n. Therefore I_i knows the vectors $[\tilde{R}_i^{n-2}, R_i^{n-2}, W_i^{n-2}, W_i^{L,n-2}]$ and $[\tilde{L}_i^{n-2}, L_i^{n-2}, W_i^{n-2}, W_i^{R,n-2}]$ at time instant n. Therefore by (19) and (20) it follows that at time instant $t=n$, node I_i knows vectors $[\tilde{R}_{i-1}^{n-1}, \hat{R}_{i-1}^{n-1}, R_{i-1}^{n-1}]$ and $[\tilde{L}_{i+1}^{n-1}, \hat{L}_{i+1}^{n-1}, L_{i+1}^{n-1}]$. From these I_i obtains $\tilde{R}_{i-1}^{n-1}$ and $\tilde{L}_{i+1}^{n-1}$. Since by (21) and (22) I_i can extract $\tilde{R}_{i+1}^{n-1} \oplus \tilde{L}_{i+1}^{n-1}$ and $\tilde{L}_{i-1}^{n-1} \oplus \tilde{R}_{i-1}^{n-1}$ at time instant n from $X_{i+1,i}^{n-1}$ and $X_{i-1,i}^{n-1}$, respectively, it can decode $\tilde{R}_{i+1}^{n-1}$ and $\tilde{L}_{i-1}^{n-1}$. Thus I_i can decode $[\tilde{R}_{i+1}^{n-1}, R_{i+1}^{n-1}, W_{i+1}^{n-1}, W_{i+1}^{L,n-1}]$ and $[\tilde{L}_{i-1}^{n-1}, L_{i-1}^{n-1}, W_{i-1}^{n-1}, W_{i-1}^{R,n-1}]$ or equivalently $[\tilde{R}_i^n, \hat{R}_i^n, R_i^n]$ and $[\tilde{L}_i^n, \hat{L}_i^n, L_i^n]$ at time instant n, as desired. Therefore, by (14), I_i may transmit X_i^n at time instant n.

We next wish to show that $X_{i,i+1}^n$ and $X_{i,i-1}^n$ are functions of the incoming messages to node O_i until time instant n. We assume that the delay between I_i and O_i for transferring information is negligible, and hence the outgoing message of O_i can be any function of $[X_i^t]_{t=1}^n$. This assumption is reasonable as a communication unit models a single processor with small internal delays. By (14), (15) and (16) we see that I_i only needs to construct $\tilde{R}_i^n \oplus \tilde{L}_i^n$ and $\tilde{L}_i^n \oplus \tilde{R}_i^n$ to be able to transmit $X_{i,i-1}^n$ and $X_{i,i+1}^n$. Observe that $\tilde{R}_i^n \oplus \tilde{L}_i^n$ and $\tilde{L}_i^n \oplus \tilde{R}_i^n$ can be obtained from $\tilde{R}_i^n \otimes \tilde{L}_i^n$, which is a component of X_i^n.

We next must show that the receiver node, O_i, at communication unit i is able to decode all messages with destination η_i in the original network successfully. Node O_i has access to messages $[X_i^t]_{t=1}^n = [\tilde{R}_i^t \otimes \tilde{L}_i^t, R_i^t, L_i^t, \hat{R}_i^t, \hat{L}_i^t, W_i^t, W_i^{R,t}, W_i^{L,t}]_{t=1}^n$, at time instant n. Observe that $[R_i^n, L_i^n, \hat{R}_i^n, \hat{L}_i^n]$ is the part of X_i^n that includes all messages with destination O_i; if the message originates at source I_j then it is generated at time instant $n-|j-i|$. Therefore every message with destination O_i will be decoded at O_i with a constant delay depending on the distance between its source and O_i in the network.

We have claimed that our network coding scheme is optimal in terms of bandwidth consumption. We demonstrate

this fact by proving that the entropies of random variables $X_i^n, X_{i,i-1}^n$ and $X_{i,i+1}^n$ satisfy the bounds in Section III with equality. As each of these vectors have components which are independent and uniformly distributed binary random variables, their entropies are equal to their lengths. We therefore use the notation $H(\cdot)$ for either the entropy or length of a random vector.
By (14) we have:

$$H(X_i^n) = H(\tilde{R}_i^n \otimes \tilde{L}_i^n) + H(R_i^n) + H(L_i^n) + H(\hat{R}_i^n) + H(\hat{L}_i^n) + H(W_i^n) + H(W_i^{R,n}) + H(W_i^{L,n}). \quad (23)$$

It follows from our earlier definitions that $H(\tilde{R}_i^n) = \sum_{a \in S'_{d_2}} \mathscr{R}_a$, $H(\tilde{L}_i^n) = \sum_{a \in S_{d_2}} \mathscr{R}_a$, $H(R_i^n) + H(L_i^n) + H(\hat{R}_i^n) + H(\hat{L}_i^n) = \sum_{a \in S_{d_1}} \mathscr{R}_a$, $H(W_i^n) + H(W_i^{R,n}) + H(W_i^{L,n}) = \sum_{a \in S_{d_3}} \mathscr{R}_a$. Since $H(\mathbf{a} \otimes \mathbf{b}) = \max\{H(\mathbf{a}), H(\mathbf{b})\}$, (23) gives (4).
By (15) we have

$$H(X_{i,i-1}^n) = H(\tilde{R}_i^n \oplus \tilde{L}_i^n) + H(R_i^n) + H(W_i^{L,n}) + H(W_i^n). \quad (24)$$

Since $H(\tilde{R}_i^n \oplus \tilde{L}_i^n) = H(\tilde{R}_i^n)$ we obtain the same bound as (7). With a similar argument we can achieve (9) for $X_{i,i+1}^n$.

5. ACKNOWLEDGMENTS

The work of S. M. S. Tabatabaei Yazdi was supported by NSF Grant No. CCF-0430201. The work of S. A. Savari was partially sponsored by the same grant. The work of G. Kramer was partially supported by the Board of Trustees of the University of Illinois Subaward No. 04-217 under NSF Grant No. CCR-0325673.

References

[1] R. Ahlswede, N. Cai, S.-Y .R. Li, and R. W. Yeung,"Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1204–1216, July 2000.

[2] M. Bakshi, M. Effros, W. Gu, and R. Koetter "On Network Coding of Independent and Dependent Sources in Line Networks," *Proc. 2007 IEEE Int. Symp. Inform. Theory*, (Nice, France), pp. 1096-1100, Jun. 24-29 2007.

[3] A. Brandstädt, V. B. Le, and J. P. Spinrad, *Graph Classes: A Survey.* SIAM Monographs on Discrete Mathematics and Applications, Society for Industrial & Applied Mathematics, 1999.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

[5] L. R. Ford and D. R. Fulkerson, *Flows in Networks*, Princeton: Ponceton University Press, 1962.

[6] A. Frank, B. Shepherd, V. Tandon, and Z. Vegh,"Node-capacitated ring routing," *Mathematics of Operations Research*, vol. 27, no. 2, pp. 372-383, May 2002.

[7] G. Kramer and S. A. Savari, "On networks of two-way channels," in A. Ashikhmin, A. Barg, eds., *DIMACS Workshop on Algebraic Coding Theory and Information Theory, (December 15-18, 2003, Rutgers University, Piscataway, NJ)*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, vol. 68, pp. 133-143, American Mathematical Society, Rhode Island, 2005.

[8] G. Kramer and S. A. Savari, "Progressive d-separating edge set bounds on network coding rates," *Proc. 2005 IEEE Int. Symp. Inform. Theory*, (Adelaide, Australia), pp. 1588-1592, Sept. 4-9 2005.

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

[10] U. Niesen, C. Fragouli and D. Tuninetti "On the Capacity of Line Networks," Submitted to *IEEE Trans. Inform. Theory*, 2005.

[11] U. Niesen, C. Fragouli and D. Tuninetti "On Cascaded Channels with Finite Complexity Processing at Intermediate Nodes", *Proc. 43th Annual Allerton Conf. on Communication, Control and Computing*, (Monticello, IL), September, 2005.

[12] P. Pakzad, C. Fragouli and A. Shokrollahi "Coding Schemes for Line Networks," *Proc. 2005 IEEE Int. Symp. Inform. Theory*, (Adelaide, Australia), Sept. 4-9 2005.

[13] S. A. Savari and G. Kramer, "The Multimessage Unicast Capacity Region for Bidirectional Ring Networks," *Proc. 2006 IEEE Int. Symp. Inform. Theory*, Seattle, WA, July 2006.

[14] R. W. Yeung S-Y. R. Li and N. Cai. "Linear network coding," *IEEE Trans. Inform. Theory*, vol. 49, pp. 371-381, February 2003.