DESIGN AND ANALYSIS OF LOW COMPLEXITY NETWORK CODING SCHEMES

A Dissertation

by

SEYED MOHAMMADSADEGH TABATABAEI YAZDI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2011

Major Subject: Electrical Engineering

DESIGN AND ANALYSIS OF LOW COMPLEXITY NETWORK CODING SCHEMES

A Dissertation

by

SEYED MOHAMMADSADEGH TABATABAEI YAZDI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Serap A. Savari
Committee Members,	Tie Liu
	Alexander Sprintson
	Catherine Yan
Head of Department,	Costas N. Georghiades

August 2011

Major Subject: Electrical Engineering

ABSTRACT

Design and Analysis of Low Complexity Network Coding Schemes. (August 2011)

Seyed Mohammadsadegh Tabatabaei Yazdi, B.S., Sharif University of Technology;

M.S., University of Michigan Ann Arbor

Chair of Advisory Committee: Dr. Serap A. Savari

In classical network information theory, information packets are treated as commodities, and the nodes of the network are only allowed to duplicate and forward the packets. The new paradigm of network coding, which was introduced by Ahlswede et al., states that if the nodes are permitted to combine the information packets and forward a function of them, the throughput of the network can dramatically increase. In this dissertation we focused on the design and analysis of low complexity network coding schemes for different topologies of wired and wireless networks.

In the first part we studied the routing capacity of wired networks. We provided a description of the routing capacity region in terms of a finite set of linear inequalities. We next used this result to study the routing capacity region of undirected ring networks for two multimessage scenarios. Finally, we used new network coding bounds to prove the optimality of routing schemes in these two scenarios.

In the second part, we studied node–constrained line and star networks. We derived the multiple multicast capacity region of node–constrained line networks based on a low complexity binary linear coding scheme. For star networks, we examined the

multiple unicast problem and offered a linear coding scheme. Then we made a connection between the network coding in a node–constrained star network and the problem of index coding with side information.

In the third part, we studied the linear deterministic model of relay networks (LDRN). We focused on a unicast session and derived a simple capacity-achieving transmission scheme. We obtained our scheme by a connection to the submodular flow problem through the application of tools from matroid theory and submodular optimization theory. We also offered polynomial-time algorithms for calculating the capacity of the network and the optimal coding scheme.

In the final part, we considered the multicasting problem in an LDRN and proposed a new way to construct a coding scheme. Our construction is based on the notion of flow for a unicast session in the third part of this dissertation. We presented randomized and deterministic polynomial-time versions of our algorithm.

To my parents Maryam and Mohsen

ACKNOWLEDGMENTS

I wish to express my most sincere gratitude to my advisor, Professor Serap Savari, for her constant and generous support and guidance throughout my graduate studies. Her support began long before I started my graduate studies in the United States and continued during my Master's studies at Michigan and my PhD studies at Texas A&M. Her mind provoking questions and thoughtful discussions have greatly influenced my thought processes during the completion of my doctoral thesis. I am deeply thankful for the freedom that she gave me to explore among a variety of problems and choose the one that best fit my interests. Professor Savari taught me the power of a good example and to keep searching for something bigger. She has taught me the true meaning of scientific research and her philosophy about life is always an inspiration for me.

I would also like to thank profusely Professor Gerhard Kramer for sharing his knowledge and experience with me. I have learned truly innovative research approaches, as well as many information theoretical concepts from him. I have been quite fortunate to have Professor Tie Liu as a mentor and a member of my thesis committee. I have enjoyed our many chats, especially since he has always been happy to share his magnificent views on more practical issues. I would also like to thank Professor Alex Sprintson for many exciting research discussions that we have had. His discussions and comments have greatly improved the quality of my work. I was also very lucky to have Professor Huafei Yan on my exam committee. Her research experience and style of teaching have always inspired my work.

Several of the results in this thesis were obtained in collaboration with some wonderful coauthors, including Gerhard Kramer, Ali Kakhbod, Farzad Farnoud and Kelli Talaska. Thanks for all of your kindness, patience, and hard work.

I would also like to thank the National Science Foundation for the generous support of my research by NSF Grant of No. CCF-0430201 and NSF Grant of No. CCF-0832810.

Finally, I would like to extend my deepest gratitude, love and affection to my beloved family, for loving me, believing in me and wishing the best for me. I owe all my achievements to their pure hearts, hearts so pure that anything they dream for me comes true.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

CHAPTER I

INTRODUCTION

A. Motivation

Communication networks are inseparable from modern life. Given the cellular networks which support the communication among cell phones, the fiber optic networks which enable the transmission of data over the internet, and sensor and satellite networks, our everyday lives are deeply affected by the operation of communication networks. Network information theory studies the fundamental limits of transmission over communication over communication networks and aims at the design of transmission scenarios in which the avaialble resourses, including hardware resources, energy resources, and time resources are optimally used.

A key question in network information theory is to find the *capacity* of a network, which is the maximum concurrent rate of transmission between different users over the network, under fixed constraints over the available resources. Although this topic has been studied for more than four decades there remain many important open problems; for example, the capacity of a simple network consisting of a source node, a destination node, and a relay node is unknown. One of the major advances towards the capacity characterization of wired networks was made by the pioneering work of Ahlswede et al. [2], who introduced the notion of *network coding* in 2000. In classical network theory wired networks with noiseless connections among different nodes was studied by treating information packets as "commodities." As a consequence, nodes of the network were only permitted to *route*;

The journal model is *IEEE Transactions on Information Theory*.

i.e., to duplicate the received information packets and forward them to adjacent nodes. The paradigm of network coding allows nodes to combine information packets to form new packets and forward them to adjacent nodes.

Ahlswede et al. [2] have shown that network coding is necessary to achieve the capacity of a *multicast* network where a source wishes to send the same information to several destinations. Subsequent works have also discovered several instances of networking scenarios for which network coding outperforms routing schemes. While it is a promising idea, our knowledge about the network coding capacity of general networks and optimal network coding schemes is still very limited, and except for a few simple structures the capacity is unknown. On the other hand, due to the constraints on the power resources and the maximum tolerable delay, only low complexity network coding schemes can in practice be implemented by the nodes. Therefore it is of great importance to study the perfomance of low-complexity network coding schemes and to study the conditions under which these schemes can be optimal and capacity achieving. This will be the focus of the present dissertation. For several basic topologies of wired and wireless networks we will analyze the performance of low-complexity network coding schemes and will provide guarantees on their optimality.

In this dissertation we will present communication networks with graphs of directed or undirected edges. A *session* refers to the communication from a source vertex to a set of destination vertices. A *unicast* session has a single destination, a multicast session has at least two destinations, and a *broadcast* session is the special case of a multicast session where all of the vertices in the network except the source are destinations. We will respectively refer to the *network coding capacity region* or *routing capacity region* of a network as the set of achievable rates among the concurrent sessions using network coding or routing protocols.

B. Background, Contributions, and Related Work

1. Undirected ring networks

We begin this dissertation by studying the routing capacity region of *undirected ring networks* from a network coding point of view. An undirected ring network is a mathematical model consisting of an undirected graph with the topology of a cycle; the vertices of the graph communicate via edges, and the sum of the flow along the two directions of an edge is bounded by its capacity. In the past two to three decades, the increasing need for high-bandwidth, reliable, and potentially long-distance communication systems caused by various high-demand and real-time applications and services resulted in the extensive deployment of communication networks based on SONET/SDH rings (see, e.g., [11], [22], [79], [90]). Because of their commercial importance, ring networks have been widely studied. The routing capacity region of multiple unicast sessions in undirected ring networks was first derived by Okamura and Seymour [64] as the special case of a more general result for planar graphs and later by Vachani et al. in [90] with a different method. However the routing capacity of the general multiple multicast sessions remains unknown.

Finding the routing capacity region of a network is equivalent to solving the problem of fractionally packing *Steiner* trees in the network graph; a Steiner tree is a tree which connects the source of a session to all destinations of that session. The routing capacity region is an inner bound to the network coding capacity region for the same communication problem. Li et al. [55] considered *undirected networks* in which communication links are bidirectional, and the total flow in both directions is limited by the capacity of the link. They showed that for a single multicast session the "Steiner strength" of an undirected network provides an upper bound to the network coding capacity which is at most twice the routing capacity for the same problem. This bound does not extend to graphs with multiple multicast sessions, and there is often a gap between the routing capacity region and the best information theoretic outer bounds on the network coding capacity region such as those offered by progressive d-separating edge set (*PdE*) bounds [48], [49].

In this dissertation we introduce a new method for characterizing the routing capacity region of arbitrary networks. Our method is based on two steps. First we generalize the "Japanese" theorem of [35], [65] that is a description of the routing capacity of networks with multiple unicast sessions in terms of an infinite set of inequalities to the case of the multiple multicast sessions. Next we use a novel technique to reduce the infinite set of inequalities into a finite and minimal set. We apply our method on two scenarios of communication over undirected ring networks:

- the source and destination vertices of each communication session form a string of adjacent vertices, and
- each session is either a broadcast or a unicast session.

In these cases, via a geometric argument, our method results in a simple description of the routing capacity region. Next we prove the optimality of routing by providing tight upper bounds on the network coding capacity of the undirected ring networks. These upper bounds match our finite description of the routing capacity region in both cases.

While our focus is on the derivation of capacity regions, earlier work has considered other aspects of deploying network coding in ring networks. For example the authors [28] investigated the benefits of network coding for saving energy in a number of broadcast wireless network topologies including rings. They showed that low complexity network coding schemes double the energy efficiency of ring networks. A different aspect of ring networks is considered in [74], which studies packet-switched wavelength division multiplexing (WDM) on unidirectional and bidirectional ring networks. In this model the total capacity of the ring is divided into different wavelengths, and each node has access to a specific wavelength for receiving or sending packets. The authors of [74] consider a des-

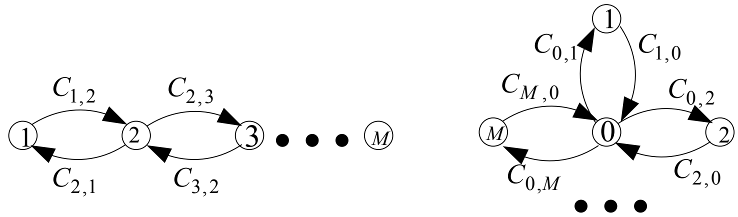

Fig. 1. Basic models of line and star networks.

tination stripping protocol, where packets are removed from the ring by their destinations upon the completion of transmission. The authors investigate various statistics of the routing capacity region for a probabilistic multiple multicast problem in which the probability that a particular session is among the set of sessions to be supported is proportional to its number of destinations.

2. Node constrained line and star networks

Lines and stars (see Figure 1) are simple topologies that appear as sub–networks in many different communication networks, e.g., wireless ad–hoc networks, sensor networks, peer–to–peer networks, and networks of optical fibers and tree and mesh networks. In this dissertation we are interested in *node–constrained* line and star networks. Given a network, the classic approach [26, 30] to produce the corresponding node–constrained network is to split each node j in the original network into two nodes I_j and O_j in the new network. Each incoming edge to j in the original network corresponds to an incoming edge to I_j in the new network, and each outgoing edge from j in the original network corresponds to an outgoing edge from O_j in the new network. There is an edge directed from I_j to O_j with capacity C_j to model the constraint on the capacity of node j. We assume that all messages generated at j in the original network are generated at I_j for the revised network and that all messages that are originally decoded at j are now decoded at O_j . Note that all

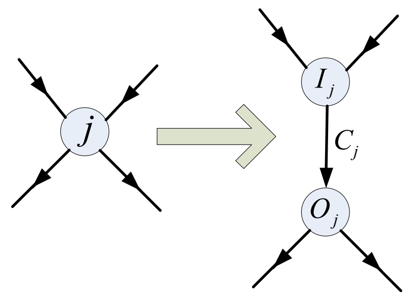

Fig. 2. Converting a node constraint to an edge constraint.

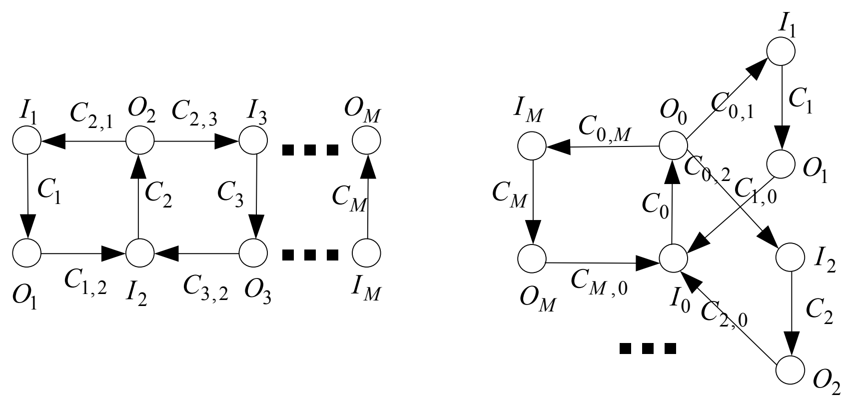

Fig. 3. The model of line and star networks with node constraints.

messages which are processed at node j pass through the edge between I_j and O_j in the revised network without affecting other parts of the network model (see Figure 2). If we apply this transformation to the networks of Figure 1, we obtain the node–constrained line and star networks of Figure 3. It is simple to verify that for a multiple multicast problem in the line and star networks of Figure 1 a routing scheme in which messages can only be duplicated and forwarded at the nodes is sufficient to achieve the maximum information theoretic throughput. However this is not true for the models of Figure 3 and network coding techniques [2] are required to achieve the maximum throughput.

As we will see later, there are several reasons for studying node-constrained line and star networks. Node-constrained networks are an interesting class of networks in their own right. They represent the limited memory resources, limited processor speed and limited bus bandwith in communication units. Node constrained networks can also model the broadcasting nature of wireless medium. We will see later two applications of the node-constrained models in wireless line networks and in the problem of index coding with side information [10].

In this dissertation we study the network coding problem for node-constrained line networks with multiple multicast sessions. We use cut bounds and entropy arguments [49, 15] based on edge cuts to find an outer bound on the network coding capacity region. We subsequently propose a linear coding scheme that achieves the outer bound. Then we consider node-constrained star networks and use entropy arguments to provide outer bounds on the multiple unicast capacity region. We also offer a coding scheme that is optimal for a broad class of problems and is based on the problem of packing edge disjoint cycles in directed graphs. Finally, we make connection between the network coding problem in star networks and the index coding problem with side information.

Line and star networks have received considerable attention in the networking literature. The authors of [73] study the capacity of edge-constrained bidirected ring networks with multiple unicast sessions. Since finite-length line networks are special cases of ring networks, [73] also treats the multiple unicast session problem for edge-constrained bidirected (or undirected) line networks. The authors of [9] investigate the network coding capacity of unidirected line networks with edge constraints for several cases of independent and dependent message sources. The multimessage multicast capacity of such line networks is derived as a special case of more general results. The authors of [63, 62] consider a cascade of Discrete Memoryless Channels (DMCs) with identical capacities and the network coding benefits when intermediate nodes can process only fixed length information blocks. They provide the relationship between the code block length and the size of the network for a constant end-to-end rate. The paper [66] shows that network coding schemes with a finite field size achieve network coding capacity for cascaded erasure channels with a single source and a single destination. For star networks, a number of routing and packet forwarding algorithms have been proposed, and these have been optimized in terms of query time [41], echo delay, error probability, scalability, failure tolerance and reliability (see, e.g., [14, 61]). Wireless line networks with broadcast have been treated in [45, 50].

3. Linear deterministic relay networks

The linear deterministic model of relay networks (LDRN) was put forward by Avestimehr, Diggavi, and Tse in [7] as an attempt to gain insight into the flow of information over wireless networks. Relay channels [15, §15.7] are a class of networks in which there is a source, a unique destination and at least one intermediary transmitter-receiver node which may be employed to assist in the communication between the source and the destination. They are not well-understood in general, and even the capacity of a Gaussian relay channel remains an outstanding open problem in network information theory. LDRNs have recently attracted considerable attention because they capture certain physical aspects of wireless communication, such as broadcasting and interference, and they are discrete and deterministic like traditional wireline network models. Avestimehr, Diggavi, and Tse use this model and a random coding transmission scheme for it to respectively approximate the capacity of specific wireless relay channels with Gaussian noise and to devise coding scheme for them [8]. Recent work [25, 59] respectively connects the linear deterministic model and the algorithm of [7] to the capacity of other types of communication channels and to the design of near-optimal coding schemes for them.

An LDRN is a wireless networking model which can be visualized as a layered directed network $\mathcal{N} = (V, E)$ with set of "nodes" $V = \bigcup_{i=1}^{M} V_i$, where V_i denotes the set of

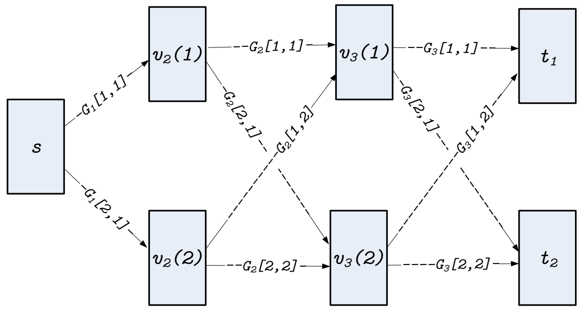

Fig. 4. An LDRN with four layers. Here $t_1 = v_4(1)$ and $t_2 = v_4(2)$.

nodes in layer i, and set of "edges" E. Let $V_i = \{v_i(1), \cdots, v_i(m_i)\}$, where m_i denotes the number of nodes in layer i. The first layer consists of a single node $s = v_1(1)$ called the source node. There are g destination nodes denoted by $t_l \triangleq v_{K_l}(d_l), l \in \{1, \cdots, g\}$, distributed in layers $K_1, K_2, \cdots, K_g$. If $g = 1$ the communication session is unicast and if $g > 1$ it is multicast. There is an "edge" from every node in V_i to every node in V_{i+1} which corresponds to the *transfer matrix* between the two nodes. Figure 4 is an example of an LDRN with four layers and two destination nodes.

During one use of the communication channel between layers i and $i+1$, $v_i(j)$ transmits a predetermined length vector $\mathbf{x}_i[j]$ to the nodes in layer $i+1$ and $v_{i+1}(k)$ receives a predetermined length vector $\mathbf{y}_{i+1}[k]$ given by

$$\mathbf{y}_{i+1}[k] = \sum_{j=1}^{m_i} G_i[k,j]\mathbf{x}_i[j],$$

where $G_i[k,j]$ is a predetermined transfer matrix of the edge $(v_i(j), v_{i+1}(k)) \in E$. Note that we can set $G_i[k,j]$ to be the all-zero matrix if there is no connection from $v_i(j)$ to

$v_{i+1}(k)$. All vectors and matrices are over a fixed finite field $\mathbb{F}$. One can define

$$\mathbf{x}_i = \begin{bmatrix} \mathbf{x}_i[1] \\ \vdots \\ \mathbf{x}_i[m_i] \end{bmatrix}, \mathbf{y}_{i+1} = \begin{bmatrix} \mathbf{y}_{i+1}[1] \\ \vdots \\ \mathbf{y}_{i+1}[m_{i+1}] \end{bmatrix}$$

and the block matrix $G_i = [G_i[k,j]]\,, 1 \leq k \leq m_{i+1}, 1 \leq j \leq m_i$. Then the received vectors at layer $i+1$ are related to the transmitted vectors at layer i by the following relationship

$$\mathbf{y}_{i+1} = G_i \mathbf{x}_i.$$

The capacity of an LDRN for a single unicast or multicast session from source s to the destinations $t_1, \cdots, t_g$ is derived in [7]. Define a cut between the source node s and a destination node t_j as a partition of nodes V into two sets A and B, with $s \in A$ and $t_j \in B$. The capacity of the cut is defined as the rank of the transfer matrix from the transmitted vectors of the nodes in A to the received vectors of the nodes in B. [7] shows that the minimum capacity of the cuts between s and t_j is the capacity of a unicast session between s and t_j. Furthermore the multicast capacity of the network between source s and destinations $t_1, \cdots, t_g$ is the minimum of the min–cut capacities between the source and each destination. The capacity–achieving scheme in [7] is a random linear coding scheme that is asymptotically optimal when the network is used for multiple rounds.

In this dissertation we construct low-complexity and capacity-achieving coding schemes for a single unicast session and a single multicast session over an LDRN. In Chapter IV we consider a unicast session. Our coding scheme for a unicast session can achieve the capacity in one use of the network. The coding operation for the relay nodes is forwarding a subset of the received symbols from the previous layer to the nodes in the next layer. In that sense our coding scheme is the counterpart of routing scheme for wired networks. We use a multitude of tools from matroid theory and submodular optimization to analyze our scheme and to obtain polynomial-time algorithms for constructing it.

In Chapter V we consider the transmission of a multicast session over an LDRN. Our coding scheme for the multicast session is obtained by combining coding schemes for unicast sessions from the source to each individual destination. In particular our scheme can be regarded as a generalization of the network coding scheme for the transmission of a multicast session over wired networks by Jaggi et al. [37]. The analysis tools here are mainly linear algebraic and probabilistic methods. We demonstrate that such a coding scheme can be constructed in polynomial time and $\lceil \log(g+1) \rceil$ uses of the network suffice to achieve the capacity.

Earlier work [5, 71] obtained capacity results for a different type of deterministic relay network in which the nodes broadcast data but the signals are received without interference. The paper [4] considers the same problem we address here, but restricts the linear model to the binary field. The approach of [4] is based on a path augmentation argument similar to the Ford-Fulkerson algorithm (see, e.g., [27]). This work was later extended to arbitrary finite fields [19]. In an independent work, Goemans et al. [33] study the flow in linear deterministic relay networks as the special case of a more general model of flow in networks based upon linking systems [75]. They use matroid partitioning and matroid intersection algorithms to obtain a capacity-achieving flow in the network.

For the case of a single multicast session, there have been multiple recent attempts to devise deterministic and efficient algorithms for constructing capacity–achieving coding schemes. In [18], Ebrahimi and Fragouli developed an algebraic framework for vector network coding and used this framework to devise a multicast transmission scheme over an LDRN. Our scheme has a lower complexity of construction and needs fewer uses of the network to achieve capacity. Erez et al. [23] offer a different construction by progressing through the network according to a topological order and maintaining the linear independence of certain subsets of coding vectors along the processing. However, the proposed algorithm does not appear to have a polynomial running time. Kim and Médard [43] generalized the algebraic framework of Koetter and Médard [44] for classical network coding to LDRNs and devised an algebraic algorithm for constructing multicast codes. Again, the proposed algorithm does not appear to have a polynomial running time. Khojastepour and Keshavarz-Haddad [42] proposed an algorithm using rotational codes to asymptotically achieve the multicast capacity of LDRN networks. Rotational codes have some built–in advantages as they are easy to implement at the relay nodes. However, the existence of deterministic polynomial–time algorithms for the construction of efficient rotational codes for multicast transmission over an LDRN remains unknown.

C. Dissertation Outline

This dissertation is organized as follows. In Chapter II we discuss the routing capacity of general wired networks and the network coding capacity of undirected ring networks. In Chpater III we present our results on the network coding capacity of node–constrained line and star networks. In Chpater IV we study a construction of a coding scheme for a unicast session in linear deterministic relay networks and in Chapter V we study a construction of a coding scheme for a multicast session over linear deterministic relay networks.

CHAPTER II

ON THE MULTIMESSAGE CAPACITY REGION FOR UNDIRECTED RING NETWORKS

A. Introduction

In this chapter we develop a new technique which leads to the tight characterization of the routing capacity region of an arbitrary network. The routing capacity region of networks with multiple sessions can be formulated as a system of linear inequalities in the (total) rates and the partial rates; each partial rate is the portion of the flow of a session that is routed along a specific Steiner tree. This initial formulation is not the solution to our problem because we do not want the partial rates as part of our description. Fourier-Motzkin elimination [76] is a procedure to project the set of solutions of a general set of linear inequalities to a subset of the variables; this can in principle be applied to the initial formulation of our problem to obtain the routing capacity region, but this approach would be complex. Our strategy is different. The "Japanese" theorem of [35], [65] describes the routing capacity region of networks with multiple unicast sessions and no multicast sessions as an infinite set of inequalities. Each inequality corresponds to a different vector of "distances" assigned to each edge in the network. Each edge distance can be chosen as an arbitrary non-negative integer, and this is why there are initially infinitely many inequalities to consider. We extend the Japanese theorem to networks supporting multiple multicast sessions, and this again results in an infinite description of the capacity region. We next consider the boundary points of the polyhedral solution and develop a novel algorithmic technique to find the finite set of necessary and sufficient inequalities among the infinite set of Japanese theorem inequalities. More specifically, our "inequality elimination" technique checks the redundancy of any inequality in defining the routing capacity region.

For the ring networks the work of Okamura and Seymour [64] and Vachani et al. [90] imply that the bounds corresponding to the Japanese theorem inequalities with exactly two non-zero edge distances, both of which are equal to one are the necessary and sufficient conditions for a collection of multiple unicast sessions to be feasible by routing. In different words, the feasibility condition for multiple unicast sessions is for the total rate across every cut in the network to be bounded from above by the capacity of the cuts. On the other hand it is known [34, 47, 73] that the cut set bounds are outer bounds on the network coding capacity of networks. As a result, the network coding capacity region of undirected ring networks in the case of multiple unicast sessions is completely characterized by the cut set bounds and is equal to the routing capacity rgion. Here we are interested in multiple multicast capacity regions in undirected ring networks and we focus on the two special cases where

- the source and destination vertices of each communication session form a string of adjacent vertices, and
- each session is either a broadcast or a unicast session.

In these cases we derive the routing capacity region and use a new argument to show that routing is rate-optimal; i.e., the network coding capacity region is no larger than the routing capacity region. We use our inequality elimination technique to prove that for the two special cases of the multiple multicast problem that we study here, we can restrict our attention to edge distances in the set $\{0, 1\}$. The next step of our analysis is to show that the network coding capacity region of each of these communication problems is identical to its routing capacity region. Our outer bounds on the network coding capacity region are based on a new analysis which extracts common information from edge cuts in order to increase some of the coefficients of the rates that appear in the inequalities.

The remainder of this chapter is organized as follows: In Section B we generalize the

Japanese theorem to multiple multicast networks. We next develop our elimination technique to reduce the infinite description of the routing capacity region into a finite one. In Section C, we consider two classes of communication problems on undirected ring networks and prove that in these cases we need only consider edge distances in $\{0, 1\}$. In Section D, we establish that the routing bounds also apply when network coding is permitted and conclude that routing is rate-optimal.

B. Routing Capacity Region in Networks

1. The Japanese theorem

Consider an undirected or directed network $G(V, E)$ in which the edge set is $E = \{e_1, \cdots, e_{|E|}\}$, the vertex set is $V = \{v_1, \cdots, v_{|V|}\}$ where for every set A, $|A|$ denotes the cardinality of the set. Let S denote the set of multicast sessions. Let C_e represent the capacity of edge $e \in E$, i.e., the maximum flow that can pass through edge e. In this chapter we assume that all capacities are rational. A multicast session $s \in S$ with rate R_s is defined by a source vertex $\nu_s \in V$ and a set of destination vertices $D_s \subset V$ each of which receives the messages in session s. The set of trees that span $\nu_s \cup D_s$ in G is denoted by $\mathcal{T}_s$. A feasible routing solution assigns to each spanning tree $T \in \bigcup_s \mathcal{T}_s$ a partial rate $r_T \geq 0$ that satisfies the following two conditions:

1. $\sum_{T \in \mathcal{T}_s} r_T = R_s$ for every $s \in S$
2. $\sum_{s \in S} \sum_{T \in \mathcal{T}_s : e \in T} r_T \leq C_e$ for every $e \in E$.

We call the rate vector $R = (R_1, \cdots, R_{|S|})$ *routing-feasible* if there is a feasible routing solution for it. The "Japanese" theorem of [35], [65] characterizes the set of all routing-feasible rate vectors for an arbitrary network with multiple unicast sessions with an infinite set of linear constraints. We start by the extending the Japanese theorem to the multiple multicast case. Let $\mathbf{a} = (a_1, \cdots, a_{|E|})$ denote a "distance" vector that assigns edge e a non-negative integral distance a_e. Then for any path or tree T we define its length $L_{\mathbf{a}}(T)$ to be the sum of the distances of the edges in T. Let $\ell_{\mathbf{a}}(s) = \min_{T \in \mathcal{T}_s} L_{\mathbf{a}}(T)$. The Japanese theorem of [35], [65] is as follows:

Theorem B.1 (The Japanese theorem) *If S is a set of unicast sessions, the polytope $P \subset \mathbb{R}^{|S|}$ of all routing-feasible rate vectors $R = (R_1, \cdots, R_{|S|})$ is:*

$$P = \{R \in \mathbb{R}^{|S|} : 0 \leq R_i, \sum_{s \in S} \ell_{\mathbf{a}}(s) R_s \leq \sum_{e \in E} a_e C_e$$
$$\textit{for all non-negative integral distance vectors } \mathbf{a}\} \qquad (2.1)$$

The following result generalizes Theorem B.1 to include multisession multicast routing.

Theorem B.2 (The extended Japanese theorem) *If S is a set of multicast sessions, the polytope $P \subset \mathbb{R}^{|S|}$ of all routing-feasible rate vectors $R = (R_1, \cdots, R_{|S|})$ is also determined by* (2.1).

Theorems B.1 and B.2 are both consequences of Farkas' Lemma (see, e.g., [100, §1.4]). We prove Theorem B.2 in Appendix A.

Because Fourier-Motzkin elimination results in a finite description of the routing capacity region, it follows that the infinite set of inequalities in Theorem B.2 contains infinitely many redundant constraints. We next introduce a method to eliminate the redundant constraints.

2. The reduced set of inequalities

An inequality in (2.1) is said to be *redundant* if it is implied by other inequalities in (2.1). A *minimal* set of inequalities that defines P is then a subset of inequalities in (2.1) with no redundant inequality. For distance vector $\mathbf{a}$, we say that rate vector R is on the hyperplane corresponding to $\mathbf{a}$ if $\sum_{s \in S} \ell_{\mathbf{a}}(s) R_s = \sum_{e \in E} a_e C_e$. We have the following result:

Lemma B.3 *A minimal set of inequalities that defines P is unique up to the multiplication of inequalities by positive scalars. Furthermore, if* $\mathbf{a}$ *and* $\mathbf{b}$ *are two distance vectors such that every* routing-feasible *rate vector R on the hyperplane corresponding to* $\mathbf{a}$ *is also on the hyperplane corresponding to* $\mathbf{b}$*, then the inequality corresponding to* $\mathbf{a}$ *in* (2.1) *is redundant.*

Proof See Appendix B.

$\square$

Lemma B.4 *The routing-feasible rate vector $R \in P$ is on the hyperplane corresponding to* $\mathbf{a}$ *if and only if*

1. *For each session $s \in S$ and every $T \in \mathcal{T}_s$, $r_T = 0$ if $L_{\mathbf{a}}(T) > \ell_{\mathbf{a}}(s)$; i.e., session s is routed only along the shortest paths and trees determined by the distance vector $\mathbf{a} = (a_1, \cdots, a_{|E|})$, and*
2. *$\sum_{s \in S} \sum_{T \in \mathcal{T}_s : e \in T} r_T = C_e$ for every $e \in E$ with $a_e > 0$; i.e., every edge with a non-zero distance is fully utilized.*

Proof To establish necessity, assume that the rate vector $R = (R_1, \cdots, R_{|S|})$ is routable and is on the hyperplane $\sum_{s \in S} \ell_{\mathbf{a}}(s) R_s = \sum_{e \in E} a_e C_e$. For any edge e in the network, the sum of all flows passing through it is at most C_e. By multiplying both sides of this inequality by a_e and summing the resulting inequalities over all edges $e \in E$ we find that

$$\sum_{s \in S} \sum_{T \in \mathcal{T}_s} L_{\mathbf{a}}(T) r_T \leq \sum_{e \in E} a_e C_e. \tag{2.2}$$

A lower bound for the left-hand side of the preceding inequality is obtained when all sessions are routed along their shortest spanning trees:

$$\sum_{s \in S} \ell_{\mathbf{a}}(s) R_s \leq \sum_{s \in S} \sum_{T \in \mathcal{T}_s} L_{\mathbf{a}}(T) r_T \leq \sum_{e \in E} a_e C_e.$$

As we assume that the rate vector is on the hyperplane given by

$$\sum_{s\in S} \ell_{\mathbf{a}}(s)R_s = \sum_{e\in E} a_e C_e,$$

it follows that Condition 1) holds. To arrive at a contradiction, suppose next that Condition 2) is invalid. Hence the rate-tuple $R = (R_1, \cdots, R_{|S|})$ is also routing-feasible in network G with link capacities C'_e for $e \in E$ in which $C'_e \le C_e$ for all e with strict inequality for at least one value of e with $a_e > 0$. The extended Japanese theorem (2.1) implies that

$$\sum_{s\in S} \ell_{\mathbf{a}}(s)R_s \le \sum_{e\in E} a_e C'_e < \sum_{e\in E} a_e C_e, \tag{2.3}$$

which contradicts $\sum_{s\in S} \ell_{\mathbf{a}}(s)R_s = \sum_{e\in E} a_e C_e$. Thus Condition 2) holds.

To establish sufficiency, consider a routing-feasible rate vector which satisfies Conditions 1) and 2). The argument for constraint (2.2) applies for any routing-feasible point, and Condition 2) implies that (2.2) can be replaced by the equality $\sum_{s\in S}\sum_{T\in\mathcal{T}_s} L_{\mathbf{a}}(T)r_T = \sum_{e\in E} a_e C_e$. By Condition 1) we know that

$$\sum_{s\in S} \ell_{\mathbf{a}}(s)R_s = \sum_{s\in S}\sum_{T\in\mathcal{T}_s} L_{\mathbf{a}}(T)r_T = \sum_{e\in E} a_e C_e. \tag{2.4}$$

Hence the rate vector R is on the hyperplane corresponding to $\mathbf{a}$, completing the proof.

$\square$

As we will next see, the true significance of the vector of edge distances in the extended Japanese theorem lies in the collection of shortest routing paths and trees for that distance vector used by the various unicast and multicast sessions; this can be viewed as a variation of Wardrop's principle [92].

Proposition B.5 *Consider two distance vectors,* $\mathbf{a} = (a_1, \cdots, a_{|E|})$ *and* $\mathbf{b} = (b_1, \cdots, b_{|E|})$*. If*

1. *for every edge* $e \in E$*,* $a_e = 0$ *implies* $b_e = 0$*, and*

2. *for every session* $s \in S$ *and tree* $T \in \mathcal{T}_s$, $L_{\mathbf{a}}(T) = \ell_{\mathbf{a}}(s)$ *implies* $L_{\mathbf{b}}(T) = \ell_{\mathbf{b}}(s)$,

then the inequality corresponding to distance vector $\mathbf{a}$ *is redundant in defining polytope* P *given the inequality corresponding to distance vector* $\mathbf{b}$.

Before we prove this result, we will discuss an example of it. Consider an undirected ring network with $V = \{1,\ 2,\ 3\}$ which supports all possible unicast and multicast sessions. We represent session s as $\nu_s \to D_s$. Then our set of sessions is given by $S = \{1 \to 2, 2 \to 1, 2 \to 3, 3 \to 2, 3 \to 1, 1 \to 3, 1 \to \{2,3\}, 2 \to \{1,3\}, 3 \to \{1,2\}\}$. Let $e_1 = \{1,2\}, e_2 = \{2,3\}, e_3 = \{3,1\}$. Suppose $C_1 = C_2 = C_3 = 1$ and $\mathbf{a} = (2,1,3)$. It is straightforward to confirm that

- $\ell_{\mathbf{a}}(1 \to 2) = \ell_{\mathbf{a}}(2 \to 1) = 2$ and the shortest path is e_1,
- $\ell_{\mathbf{a}}(2 \to 3) = \ell_{\mathbf{a}}(3 \to 2) = 1$ and the shortest path is e_2,
- $\ell_{\mathbf{a}}(3 \to 1) = \ell_{\mathbf{a}}(1 \to 3) = 3$ and both paths are shortest, and
- $\ell_{\mathbf{a}}(1 \to \{2,3\}) = \ell_{\mathbf{a}}(2 \to \{1,3\}) = \ell_{\mathbf{a}}(3 \to \{1,2\}) = 3$ and the shortest tree is $\{e_1, e_2\}$.

Therefore, the halfspace resulting from distance vector $\mathbf{a}$ is

$$\begin{aligned} &2(R_{1\to2} + R_{2\to1}) + (R_{2\to3} + R_{3\to2}) + 3(R_{3\to1} + R_{1\to3}) \\ &+3(R_{1\to\{2,3\}} + R_{2\to\{1,3\}} + R_{3\to\{1,2\}}) \le 2C_1 + C_2 + 3C_3 = 6. \end{aligned} \tag{2.5}$$

Next take $\mathbf{b} = (1,0,1)$. Observe that the shortest paths and shortest trees for each session under distance vector $\mathbf{a}$ continue to be shortest paths and shortest trees for the sessions under $\mathbf{b}$, although $\mathbf{b}$ has a second shortest path for unicast sessions $1 \to 2$ and $2 \to 1$ and a second shortest tree for the multicast sessions. The halfspace resulting from $\mathbf{b}$ is

$$(R_{1\to2}+R_{2\to1})+(R_{3\to1}+R_{1\to3})+(R_{1\to\{2,3\}}+R_{2\to\{1,3\}}+R_{3\to\{1,2\}}) \le C_1+C_3 = 2. \tag{2.6}$$

The proposition stipulates that (2.5) is redundant for specifying the routing capacity region given (2.6). The proof establishes that every routing-feasible rate-tuple like $R_{1\to 2} = R_{2\to 3} = R_{3\to 1} = 1,\ R_s = 0,\ s \notin \{1 \to 2,\ 2 \to 3,\ 3 \to 1\}$ which satisfies (2.5) with equality necessarily satisfies (2.6) with equality. The rate-tuple $R_{1\to\{2,3\}} = 2,\ R_s = 0,\ s \notin \{1 \to \{2, 3\}\}$ exemplifies a *routing infeasible* rate-tuple which satisfies (2.5) with equality; notice that four units of capacity are needed to route two units of multicast traffic, while the network has only three units of capacity.

Proof Consider a routing-feasible rate vector R on the hyperplane defined by $\mathbf{a}$. By Lemma B.4, Condition 1), every session is routed only along the shortest paths and trees for $\mathbf{a}$, and hence by assumption only along the shortest paths and trees for $\mathbf{b}$. Furthermore note that any edge e with $b_e > 0$ must have $a_e > 0$ by assumption, and hence this edge must be fully utilized by Lemma B.4, Condition 2). By Lemma B.4 it follows that the routable point also is on the hyperplane defined for distance vector $\mathbf{b}$. Therefore by Lemma B.3, the bound corresponding to $\mathbf{a}$ is redundant given the inequality corresponding to $\mathbf{b}$.

$\square$

We offer an alternate algebraic proof for Proposition B.5 in Appendix C. Proposition B.5 provides a powerful algorithmic technique for deriving the minimal set of inequalities for describing polytope P, and we next apply it to two communication problems on undirected ring networks.

C. The Routing Rate Region in Ring Networks

In this section we focus on the ring network $G(V, E)$, with set of vertices $V = \{1, \cdots, n\}$, and set of edges $E = \{1, \cdots, n\}$, as illustrated in Figure 5. As an additional notation for rings, let $L_{\mathbf{a}}(p, q)$ denote the distance in the clockwise direction between vertices p and q assuming distance vector $\mathbf{a}$. As a first step in understanding the general multiple multicast

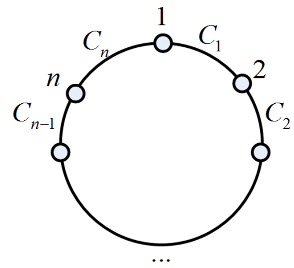

Fig. 5. An undirected ring network with n vertices.

problem on undirected ring networks, we focus here on the analysis of two special cases. In the first case we consider a ring network in which each session $s \in S$ is a *line* session; i.e., the source ν_s and all destinations D_s, form a sequence (in any order) of adjacent vertices in the network. In the second case we study the ring network problem with multiple unicast and broadcast sessions. These special cases already require new techniques, and we are unaware of similar analyses in the literature. Our approach in each case is to show that for an arbitrary non-trivial distance vector $\mathbf{a} = (a_1, \cdots, a_n)$ we can construct a non-trivial distance vector $\mathbf{b} = (b_1, \cdots, b_n)$ with the following properties:

- **Property 1:** $b_i = 0$ or 1 for all i,
- **Property 2:** $b_i = 0$ whenever $a_i = 0$,
- **Property 3:** for every session s, if $T \in \mathcal{T}_s$ and $\ell_{\mathbf{a}}(s) = L_{\mathbf{a}}(T)$, then $\ell_{\mathbf{b}}(s) = L_{\mathbf{b}}(T)$.

Hence Proposition B.5 implies that distance vector $\mathbf{a}$ can be eliminated by $\mathbf{b}$. It then follows that we can restrict our attention to distance vectors in the set $\{0, 1\}^n$.

Remark 1 *For ring networks it is sometimes more convenient to restate Property 3 in terms of the* complementary trees *of each session. Let $\mathcal{T}_s$ denote the set of routing trees of session s. Then the complementary tree $\overline{T}$ of a tree $T \in \mathcal{T}_s$ is the tree that remains after removing the edges and internal vertices of T from G. Let $\bar{\mathcal{T}}_s$ denote the set of all complementary trees* *corresponding to session s. For distance vector $\mathbf{a}$ let $c_{\mathbf{a}}$ denote the sum of all edge distances in the network. Given $\mathbf{a}$ and tree T, it follows that $L_{\mathbf{a}}(T) = c_{\mathbf{a}} - L_{\mathbf{a}}(\bar{T})$. Therefore, $\ell_{\mathbf{a}}(s) = \min_{T\in\mathcal{T}_s} L_{\mathbf{a}}(T) = c_{\mathbf{a}} - \max_{\hat{T}\in\bar{\mathcal{T}}_s} L_{\mathbf{a}}(\hat{T})$. We conclude that tree $T \in \mathcal{T}_s$ satisfies $\ell_{\mathbf{a}}(s) = L_{\mathbf{a}}(T)$ if and only if $L_{\mathbf{a}}(\bar{T}) = \max_{\hat{T}\in\bar{\mathcal{T}}_s} L_{\mathbf{a}}(\hat{T})$. Therefore Property 3 is equivalent to:*

- **Property 3':** *for every session s, if $T \in \bar{\mathcal{T}}_s$ and $L_{\mathbf{a}}(T) = \max_{\hat{T}\in\bar{\mathcal{T}}_s} L_{\mathbf{a}}(\hat{T})$, then $L_{\mathbf{b}}(T) = \max_{\hat{T}\in\bar{\mathcal{T}}_s} L_{\mathbf{b}}(\hat{T})$.*

Finally consider session s with $\nu_s = o$ and $D_s = \{d_1, \cdots, d_K\}$, where $d_1 < \cdots < d_i < o < d_{i+1} < \cdots < d_K$. Let us denote the path from vertex j to vertex k in the clockwise direction on the ring G by $G(j,k)$. Then it is easy to verify that

$$\bar{\mathcal{T}}_s = \{G(d_1,d_2), \cdots, G(d_i,o), G(o,d_{i+1}), \cdots, G(d_{K-1},d_K), G(d_K,d_1)\}.$$

We next provide an algorithm for constructing $\mathbf{b} \in \{0,1\}^n$ for a given $\mathbf{a}$.

1. Algorithm for constructing a binary distance vector $\mathbf{b}$ for distance vector $\mathbf{a}$ for the case of line sessions

Consider a set of line sessions S. Let $A = \max\{a_1, \cdots, a_n\} > 0$ and take $m_1 < \cdots < m_N$ to be the set of indices of all edges of maximum distance in $\mathbf{a}$, so that $a_{m_1} = a_{m_2} = \cdots = a_{m_N} = A$. Without loss of generality, we can assume that $m_1 = 1$ and so $a_1 = A > 0$. We abuse notation somewhat and write $L_{\mathbf{a}}(v_1, v_2) = \sum_{i=v_1}^{v_2-1} a_i$ as the length of the clockwise path from vertex v_1 to vertex v_2. The following algorithm shows that every distance vector can be reduced to a binary distance vector.

1. Set $b_j = 0$ for all $j \in \{1, \cdots, n\}$ with $a_j = 0$.

2. Set $i = 1$.

3. Complete the following steps:

   (a) Set $b_i = 1$.

   (b) Search for an index j such that $L_{\mathbf{a}}(i+1, j) < A$, but $L_{\mathbf{a}}(i+1, j+1) \geq A$, and $j \leq n$. If such a j exists, it must be unique. In this case let $i = j$ and return to Step 3. If no such j exists, go on to Step 4).

4. Set each remaining edge distance in $\mathbf{b}$ to 0.

We illustrate the algorithm above with an example:

Example 2 *Consider distance vector* $\mathbf{a} = (3, 1, 3, 0, 1, 2)$. *Then* $A = 3$ *and we set* $b_4 = 0$. *We initialize* $i = 1$ *and set* $b_1 = 1$. *Since* $L_{\mathbf{a}}(2, 3) < A$ *and* $L_{\mathbf{a}}(2, 4) \geq A$, *we next set* $i = 3$ *and* $b_3 = 1$. *Because* $L_{\mathbf{a}}(4, 6) < A$ *and* $L_{\mathbf{a}}(4, 1) \geq A$ *we set* $i = 6$, $b_6 = 1$. *As we can not further increase* i *we next set* $b_2 = b_5 = 0$. *The output of the algorithm will be* $\mathbf{b} = (1, 0, 1, 0, 0, 1)$.

2. Proof of the algorithm performance

It is clear that the algorithm satisfies Properties 1 and 2; we next show that it also satisfies Property 3. Fix a line session $s \in S$. For convenience, we relabel the vertices of the ring so that the source and all destinations for the session s are all on a line starting at vertex 1 and ending at vertex $|D_s| + 1$. (Note that we may now have $m_1 \neq 1$.)

Consider the set of complementary trees for session s:

$$\bar{\mathcal{T}}_s = \{G(1, 2), G(2, 3), \cdots, G(|D_s|, |D_s| + 1), G(|D_s| + 1, 1)\}.$$

Observe that

$$\max_{T \in \bar{\mathcal{T}}_s} L_{\mathbf{a}}(T) = \max\{a_1, a_2, \cdots, a_{|D_s|}, L_{\mathbf{a}}(|D_s| + 1, 1)\}.$$

To show that Property 3' holds we must verify that if $T \in \bar{\mathcal{T}}_s$ and $L_{\mathbf{a}}(T) = \max_{\hat{T} \in \bar{\mathcal{T}}_s} L_{\mathbf{a}}(\hat{T})$, then $L_{\mathbf{b}}(T) = \max_{\hat{T} \in \bar{\mathcal{T}}_s} L_{\mathbf{b}}(\hat{T})$. We begin with some lemmas.

Lemma C.1 *The vector* $\mathbf{b}$ *produced by the algorithm satisfies* $b_{m_1} = b_{m_2} = \cdots = b_{m_N} = 1$.

Proof By Step 3) of the algorithm we have $b_{m_1} = 1$. To arrive at a contradiction, for some $i > m_1$ suppose $b_i = 1$, j is the next smallest integer for which $b_j = 1$, and there is some $k > 1$ with $i < m_k < j$ and $b_{m_k} = 0$. Then Step 3) of the algorithm implies that $L_{\mathbf{a}}(i+1, j) < A$. However, since $m_k < j$, it follows that $L_{\mathbf{a}}(i+1, j) \geq L_{\mathbf{a}}(i+1, m_k+1) \geq a_{m_k} = A$, which is a contradiction. Therefore $b_{m_k} = 1$.

$\square$

Lemma C.2 *If* $L_{\mathbf{a}}(i, j) < A$*, then there is at most one edge* $k \in \{i, \cdots, j-1\}$ *for which* $b_k = 1$.

Proof Suppose instead that there are $k, l \in \{i, \cdots, j-1\}$ with $k < l$, $b_k = b_l = 1$, and $b_{k+1} = \cdots = b_{l-1} = 0$. Then Step 3b) implies that $L_{\mathbf{a}}(k+1, l) < A$ and $L_{\mathbf{a}}(k+1, l+1) \geq A$. However, it then follows that $L_{\mathbf{a}}(i, j) \geq L_{\mathbf{a}}(k+1, l+1) \geq A$, which contradicts the assumption that $L_{\mathbf{a}}(i, j) < A$.

$\square$

Lemma C.3 *If* $L_{\mathbf{a}}(i, j) = A$*, then there is exactly one edge* $k \in \{i, \cdots, j-1\}$ *for which* $b_k = 1$.

Proof Suppose first that there is no edge $k \in \{i, \cdots, j-1\}$ with $b_k = 1$. Let $r < i$ be the largest integer for which $b_r = 1$. Then $L_{\mathbf{a}}(r+1, j) < A$. However, since $A = L_{\mathbf{a}}(i, j) \leq L_{\mathbf{a}}(r+1, j)$ this can not happen. Next suppose that there are k and $l \in \{i, \cdots, j-1\}$ with

$b_k = b_l = 1$ and $b_{k+1} = \cdots = b_{l-1} = 0$. By Step 1), $b_k = 1$ implies $a_k > 0$. By Step 3b), since $b_{k+1} = \cdots = b_{l-1} = 0$ and $b_l = 1$, it follows that $L_{\mathbf{a}}(k+1, l+1) \geq A$. Observe that $A = L_{\mathbf{a}}(i,j) \geq a_k + L_{\mathbf{a}}(k+1, l+1) > A$, which is a contradiction.

□

Lemma C.4 *If $L_{\mathbf{a}}(i,j) > A$, then there exists $k \in \{i, \cdots, j-1\}$ such that $b_k = 1$.*

Proof Suppose that there is no such k with $b_k = 1$, and let r be the largest integer less than i with $b_r = 1$. Then $L_{\mathbf{a}}(r+1, j) < A$. This contradicts the fact that $A < L_{\mathbf{a}}(i,j) \leq L_{\mathbf{a}}(r+1, j)$.

□

Since $\max_{T \in \bar{\mathcal{T}}_s} L_{\mathbf{a}}(T) \geq \max\{a_1, \cdots, a_n\}$, there are two possibilities for $\max_{T \in \bar{\mathcal{T}}_s} L_{\mathbf{a}}(T)$.

1. $\max_{T \in \bar{\mathcal{T}}_s} L_{\mathbf{a}}(T) = A$
2. $\max_{T \in \bar{\mathcal{T}}_s} L_{\mathbf{a}}(T) > A$

We consider the following three cases; the first two correspond to $\max_{T \in \bar{\mathcal{T}}_s} L_{\mathbf{a}}(T) = A$, and the third corresponds to $\max_{T \in \bar{\mathcal{T}}_s} L_{\mathbf{a}}(T) > A$. In each case we find the maximum length complementary trees with respect to $\mathbf{a}$ and show that they are also maximum length with respect to $\mathbf{b}$.

- First suppose that $\max_{T \in \bar{\mathcal{T}}_s} L_{\mathbf{a}}(T) = A$ and $L_{\mathbf{a}}(|D_s|+1, 1) < A$. Then by Lemma C.1 we have $b_{m_1} = b_{m_2} = \cdots = b_{m_N} = 1$, and by Lemma C.2 at most one among the edges in $\{|D_s|+1, |D_s|+2, \cdots, n\}$ will have a unit edge distance in $\mathbf{b}$. Therefore $\max_{T \in \bar{\mathcal{T}}_s} L_{\mathbf{b}}(T) = 1$. Also, if for $T \in \bar{\mathcal{T}}_s$, $L_{\mathbf{a}}(T) = \max_{\hat{T} \in \bar{\mathcal{T}}_s} L_{\mathbf{a}}(\hat{T}) = A$, then $T = G(m_i, m_i+1)$ for some $m_i \in \{1, \cdots, |D_s|\}$. Since $b_{m_i} = \max_{T \in \bar{\mathcal{T}}_s} L_{\mathbf{b}}(T) = 1$, it follows that Property 3' is satisfied.

- Next suppose that $\max_{T\in\bar{\mathcal{T}}_s} L_{\mathbf{a}}(T) = A$ and $L_{\mathbf{a}}(|D_s|+1,1) = A$. Then by Lemma C.1 we have $b_{m_1} = b_{m_2} = \cdots = b_{m_N} = 1$, and by Lemma C.3, exactly one among the edges in $\{|D_s|+1, |D_s|+2, \cdots, n\}$ should have a unit edge distance in $\mathbf{b}$. Hence $L_{\mathbf{b}}(|D_s|+1,1) = 1$ and $\max_{T\in\bar{\mathcal{T}}_s} L_{\mathbf{b}}(T) = 1$. Also, if for $T \in \bar{\mathcal{T}}_s$, $L_{\mathbf{a}}(T) = \max_{\hat{T}\in\bar{\mathcal{T}}_s} L_{\mathbf{a}}(\hat{T}) = A$, then either $T = G(m_i, m_i+1)$ for some $m_i \in \{1, \cdots, |D_s|\}$ or $T = G(|D_s|+1, n)$, and since $b_{m_i} = L_{\mathbf{b}}(|D_s|+1,1) = \max_{T\in\bar{\mathcal{T}}_s} L_{\mathbf{b}}(T) = 1$, it follows that Property 3' is satisfied.

- Finally, suppose $L_{\mathbf{a}}(|D_s|+1,1) > A$. Then by Lemma C.4, at least one among the edges in $\{|D_s|+1, |D_s|+2, \cdots, n\}$ should have a unit edge distance in $\mathbf{b}$, and hence $\max_{T\in\bar{\mathcal{T}}_s} L_{\mathbf{b}}(T) = L_{\mathbf{b}}(|D_s|+1,1)$. Also, if for $T \in \bar{\mathcal{T}}_s$, $L_{\mathbf{a}}(T) = \max_{\hat{T}\in\bar{\mathcal{T}}_s} L_{\mathbf{a}}(\hat{T})$, then $T = G(|D_s|+1, n)$. Since $\max_{T\in\bar{\mathcal{T}}_s} L_{\mathbf{b}}(T) = L_{\mathbf{b}}(|D_s|+1,1)$, we have that Property 3' is satisfied.

We have now shown that for any line session, our algorithm generates a binary distance vector $\mathbf{b}$ that satisfies Properties 1, 2, and 3. Thus, each distance vector with a distance greater than one can be reduced to a binary distance vector. Therefore by Proposition B.5 the routing capacity region can be determined by all binary distance vectors.

3. Algorithm for constructing a binary distance vector $\mathbf{b}$ for distance vector $\mathbf{a}$ for the case of unicast and broadcast sessions

We next consider the routing capacity region of a ring with a set of sessions S such that $|D_s| = 1$ or $|D_s| = n-1$ for all $s \in S$. We again prove that binary distance vectors suffice for describing polytope P by constructing a binary vector $\mathbf{b}$ which eliminates a given distance vector $\mathbf{a}$.

We first assume all edge distances in $\mathbf{a}$ are positive and subsequently extend the algorithm to general distance vectors with some zero elements. The heart of the algorithm is the following

Basic Generation Procedure:

1. If all a_i are equal, then set $b_i = 1$ for $i \in \{1, \cdots, n\}$. Otherwise, proceed to the next step.

2. Draw a circle $\mathcal{C}$, with points on its perimeter corresponding to each vertex of the ring so that the length of the arc between two adjacent points on the circle is proportional to the corresponding edge distance in $\mathbf{a}$.

3. From each point on the perimeter of $\mathcal{C}$ draw a diameter originating from that point.

4. If the arc corresponding to an edge on $\mathcal{C}$ intersects at least one diameter, then set the corresponding edge distance in $\mathbf{b}$ to one; otherwise set it to zero.

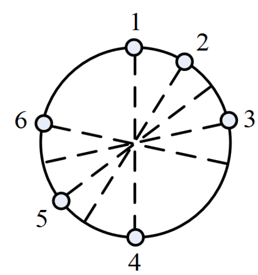

Fig. 6. An instance of applying the *Basic Generation Procedure* to a ring network in which $\mathbf{a} = (1, 2, 4, 2, 2, 3)$.

Example 3 *Consider a ring network with 6 vertices and a distance vector* $\mathbf{a} = (1, 2, 4, 2, 2, 3)$. *We wish to find the corresponding binary distance vector* $\mathbf{b} = (b_1, \cdots, b_6)$ *according to the Basic Generation Procedure. We first draw the circle* $\mathcal{C}$ *and all diameters for* $\mathbf{a}$ *according to Steps 2) and 3) (see Figure 6). Since edges* $2, 3, 4$*, and* 5 *are intersected by at least one diameter we set* $b_2 = b_3 = b_4 = b_5 = 1$ *and* $b_1 = b_6 = 0$. *The resulting binary distance vector is* $\mathbf{b} = (0, 1, 1, 1, 1, 0)$.

The Basic Generation Procedure sometimes needs a correction to result in a $\mathbf{b}$ with the desired Properties; this depends on the path lengths of the different unicast sessions. We will see the appropriate method of constructing $\mathbf{b}$ for different cases and a proof of validity for each case:

First we categorize positive distance vectors $\mathbf{a}$ based on the path lengths of the different pairs of vertices into three types :

- **Type 1:** There is no pair of vertices with equal clockwise and counterclockwise routing path lengths by distance vector $\mathbf{a}$.
- **Type 2:** There is exactly one pair of vertices with equal clockwise and counterclockwise routing path lengths by distance vector $\mathbf{a}$.
- **Type 3:** There are multiple pairs of vertices with equal clockwise and counterclockwise routing path lengths by distance vector $\mathbf{a}$.

Theorem C.5 *If positive distance vector* $\mathbf{a}$ *is of Type 1, then the Basic Generation Procedure generates a distance vector* $\mathbf{b}$ *that satisfies Properties 1, 2, and 3.*

Proof See Appendix D.

□

Theorem C.6 *If positive distance vector* $\mathbf{a}$ *is of Type 2, then either*

1. *the Basic Generation Procedure*

   or

2. *the Basic Generation Procedure followed by the change of a particular edge distance from 0 to 1 in* $\mathbf{b}$

results in a legitimate vector $\mathbf{b}$ *that satisfies Properties 1,2, and 3.*

Proof See Appendix E.

□

Theorem C.7 *If positive distance vector* $\mathbf{a}$ *is of Type 3, then* $\mathbf{a}$ *can be decomposed into several subvectors of Type 2 such that a combination of the binary distance subvectors corresponding to the subvectors of* $\mathbf{a}$ *results in a distance vector* $\mathbf{b}$ *that satisfies Properties 1,2, and 3.*

Proof See Appendix F.

□

To complete the algorithm, consider distance vectors $\mathbf{a}$ with at least one element being equal to zero. By Proposition B.5 for all i we set $b_i = 0$ if $a_i = 0$. Form a shorter distance vector, $\mathbf{a}' = (a'_1, a'_2, \cdots, a'_{n'})$, which is $\mathbf{a}$ without its zero elements. Observe that the length of a path between two vertices by $\mathbf{a}$ in a specific direction is equal to the length of the path between a corresponding pair of vertices for $\mathbf{a}'$ in the same direction. Thus given a suitable binary distance vector $\mathbf{b}'$ for $\mathbf{a}'$, we can find $\mathbf{b}$ by appropriately inserting zero edge distances into $\mathbf{b}'$. Clearly $\mathbf{b}$ preserves the shortest paths and broadcast trees for $\mathbf{a}$, completing the algorithm.

4. Concluding remarks on the Extended Japanese theorem

The algorithms that we provided in this section reduce a given distance vector into a binary distance vector in linear time in the size of the ring. The advantage of the reduction is that it provides a simple and finite characterization of the routing capacity region as opposed to the infinite set of inequalities. The following example illustrates the routing capacity region of a 3 vertex ring network.

Example 4 *Consider a ring network supporting the unicast and broadcast sessions* $S = \{1 \to 2, 1 \to 3, 2 \to \{1,3\}\}$. *The routing capacity region of this problem can be derived by considering all binary distance vectors of length* 3 *and their corresponding inequalities as follows:*

- *the distance vector* $(1,1,0)$ *results in the inequality* $R_{1\to 2} + R_{2\to\{1,3\}} \leq C_1 + C_2$*,*
- *the distance vector* $(1,0,1)$ *results in the inequality* $R_{1\to 2} + R_{1\to 3} + R_{2\to\{1,3\}} \leq C_1 + C_3$*,*
- *the distance vector* $(0,1,1)$ *results in the inequality* $R_{1\to 3} + R_{2\to\{1,3\}} \leq C_2 + C_3$*,*
- *the distance vector* $(1,1,1)$ *results in the inequality* $R_{1\to 2} + R_{1\to 3} + 2R_{2\to\{1,3\}} \leq C_1 + C_2 + C_3$*.*

Observe that binary distance vectors with all zeroes or a single one result in trivial inequalities since the shortest trees have length zero.

To conclude this section we point out that the bounds corresponding to binary distance vectors are not in general sufficient to characterize the routing capacity region of undirected rings with multiple multicast sessions. For example, we have the following lemma:

Lemma C.8 *For an undirected ring* G *with* $n \geq 5$ *vertices supporting all multicast sessions, the distance vector* $\mathbf{a} = (x, 1, \cdots, 1)$ *for* $2 \leq x \leq \lfloor \frac{n-2}{2} \rfloor$ *can not be reduced to any* $\mathbf{b} = (b_1, \cdots, b_n)$ *with* $\max\{b_1, \cdots, b_n\} < x$*.*

Proof To arrive at a contradiction, assume that we have found a $\mathbf{b}$ with $\max\{b_1, \cdots, b_n\} < x$ that satisfies the conditions of Proposition B.5. Let s be an arbitrary multicast session with $\nu_s = o$ and $D_s = \{d_1, d_2, \cdots, d_K\}$, where $d_1 < \cdots < d_i < o < d_{i+1} < \cdots < d_K$. Note that the set of complementary trees for session s is

$$\bar{\mathcal{T}}_s = \{G(d_1, d_2), \cdots, G(d_i, o), G(o, d_{i+1}), G(d_{i+1}, d_{i+2}), \cdots, G(d_{K-1}, d_K), G(d_K, d_1)\}.$$

Thus to satisfy the conditions of Proposition B.5, the longest trees in $\bar{\mathcal{T}}_s$ with respect to $\mathbf{a}$ should remain longest under $\mathbf{b}$.

Consider the multicast session from 1 to $\{2, x+2, x+3, \cdots, n\}$. Here among the complementary trees $G(1,2), G(2,x+2), \cdots, G(n,1)$ there are two longest trees $G(1,2)$ and $G(2,x+2)$ under $\mathbf{a}$, and hence they should remain longest under $\mathbf{b}$. Thus $b_1 = \sum_{i=2}^{x+1} b_i$. Likewise consider the multicast sessions from 1 to $\{2,3,x+3,\cdots,n\}$, from 1 to $\{2,3,4,x+4,\cdots,n\}$, $\cdots$, from 1 to $\{2,3,\cdots,n-x,n\}$, and from 1 to $\{2,3,\cdots,n-x+1\}$ to obtain the constraints:

$$\begin{aligned} b_1 &= b_2 + b_3 + \cdots + b_{x+1} \\ b_1 &= b_3 + b_4 + \cdots + b_{x+2} \\ &\vdots \\ b_1 &= b_{n-x+1} + b_{n-x+2} + \cdots + b_n. \end{aligned} \tag{2.7}$$

Therefore,

$$b_2 = b_{x+2},\ b_3 = b_{x+3},\ \cdots,\ b_{n-x} = b_n. \tag{2.8}$$

Next consider the multicast sessions from 1 to $\{3, x+4, \cdots, n\}$, from 1 to $\{3,4,x+5,\cdots,n\}$, $\cdots$, from 1 to $\{3,4,\cdots,n-x-1,n\}$, and from 1 to $\{3,4,\cdots,n-x\}$. The constraints maintaining the longest complementary trees with respect to $\mathbf{a}$ results in the following set of equalities:

$$\begin{aligned} b_1 + b_2 &= b_3 + b_4 + \cdots + b_{x+3} \\ b_1 + b_2 &= b_4 + b_5 + \cdots + b_{x+4} \\ &\vdots \\ b_1 + b_2 &= b_{n-x} + b_{n-x+1} + \cdots + b_n \end{aligned} \tag{2.9}$$

Hence,

$$b_3 \;=\; b_{x+4},\; b_4 \;=\; b_{x+5},\; \cdots,\; b_{n-x-1} \;=\; b_n. \tag{2.10}$$

Since $n - x \geq x + 2$, (2.8) and (2.10) imply:

$$b_2 = b_3 \cdots = b_n \doteq b \tag{2.11}$$

and

$$b_1 = xb. \tag{2.12}$$

Since distance vector $\mathbf{b} \neq \mathbf{0}$, it follows that $b \neq 0$. Hence $b_1 = xb$ should be an integer bounded below by x, which is a contradiction.

□

Although we were able above to characterize the exact routing capacity region for two special cases, it appears difficult to apply our tools to arbitrary collections of multiple multicast sessions and/or arbitrary networks. However, these ideas offer some insights that help to further characterize the minimal set of distance vectors that define routing capacity region for general multiple multicast networks. Indeed, in a recent work [39] we provide upper and lower bounds to show that the maximum edge distance needed for multiple multicast sessions in an undirected network grows exponentially with the size of the largest cycle of the network. The lower bound was obtained by demonstrating that a particular distance vector can not be reduced to another distance vector with a smaller maximal element. For the upper bound, observe that distance vectors are characterized by their shortest trees for the various sessions. Therefore for a given (and feasible) set of shortest trees, one can solve an integer programming problem to determine a distance vector with the same set of minimal trees. By investigating the size and complexity of the integer program and applying theorems of integer linear programming (see [76, Ch.10]), we establish that the integer program always has distance vector solutions with elements that are exponentially large in the size of the largest cycle of the network.

D. Network Coding Bounds

The set of routing bounds corresponding to binary distance vectors provide an inner bound to the network coding capacity region. To show that these bounds are tight for our two problems, we next present an information theoretic argument to establish the same bounds as outer bounds to the network coding capacity region. We say that a session is of *Type 1* if it is a line session and is of *Type 2* if it is a unicast or broadcast session. For $i \in \{1, 2\}$ we say the set S of sessions is of Type i when every $s \in S$ is of Type i. We first prove the following useful lemma:

Lemma D.1 *For a ring with n vertices supporting a set of sessions of Type 1 or of Type 2, every routing bound corresponding to a binary distance vector with m ones, $m \leq n$, is equivalent to a routing bound for a ring with m vertices, where the distance vector for this latter network is the all-ones vector.*

Proof For the ring with n vertices and a binary distance vector, create a possibly smaller ring by successively replacing the vertices u and v with one vertex if u and v have zero distance between them. The routing bound corresponding to the original binary distance vector is clearly the same as the routing bound for the new ring with an all-ones distance vector.

□

The proof that this routing bound is a network coding bound is developed next. It is easy to verify that the sessions of Type 1 or 2 will still be of the same type for the smaller network. Therefore we hereafter only consider distance vectors of the form $\mathbf{b}_n = (1, \cdots, 1)$ for a ring G with $n \geq 2$ vertices.

First consider a ring with $n = 2$; here there are only two sessions, s_1 from 1 to 2 and s_2 from 2 to 1. The routing bound for this case, i.e., for $\mathbf{b}_2$ is $R_{s_1} + R_{s_2} \leq C_1 + C_2$, and this can easily be derived as a network coding bound using cut set bounds on edges 1 and 2.

Next consider a ring with $n = 3$ and distance vector $\mathbf{b}_3$. Here all multicast sessions are always both of Type 1 and of Type 2. The routing bound for this case is as follows:

$$\sum_{\{s:|D_s|=1\}} R_s + \sum_{\{s:|D_s|=2\}} 2R_s \leq C_1 + C_2 + C_3. \tag{2.13}$$

To show that (2.13) holds for network coding, we use the bidirected cut set bounds from [47] . Decompose each of the undirected capacities C_1, C_2, and C_3 into two unidirectional capacities: $C_1 = C_{12} + C_{21}$, $C_2 = C_{23} + C_{32}$, and $C_3 = C_{13} + C_{31}$, so that C_{pq} denotes the portion of the edge capacity which is directed from vertex p to vertex q. Then (2.13) can be obtained by summing the three bidirected cut set bounds derived from the pairs of directed edges $\big((1,2),(1,3)\big)$, $\big((2,1),(2,3)\big)$, and $\big((3,1),(3,2)\big)$.

For distance vector $\mathbf{b}_n$, $n \geq 4$, it turns out that the bidirected cut set bounds can not provide us with tight enough bounds for the two types of sessions. In this case we obtain our results for network coding via another set of bounds which are derived by using the data processing inequality and the chain rule for mutual information.

Theorem D.2 *Consider the ring with four vertices illustrated in Figure 7. Then for network coding:*

$$\sum_{\nu_s \in \{1,3\}} R_s + \sum_{\nu_s=2,\ 4 \in D_s} R_s + \sum_{\nu_s=4,\ 2 \in D_s} R_s + \sum_{\{2,4\} \subseteq D_s} R_s \leq C_{12} + C_{32} + C_{14} + C_{34}. \tag{2.14}$$

Proof See Appendix G.

$\square$

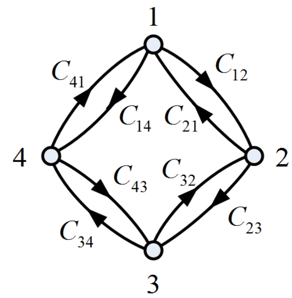

Fig. 7. A ring with four vertices.

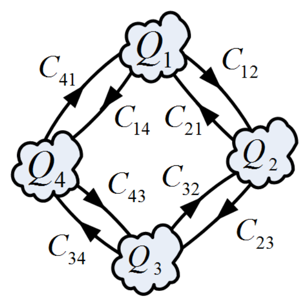

Fig. 8. A general ring with four sets of vertices.

For larger ring networks, a similar relationship can be established when vertices $1,\ 2,\ 3,$ and 4 are respectively replaced by four sets of neighboring vertices $Q_1,\ Q_2,\ Q_3,$ and Q_4.

Proposition D.3 *For the ring in Figure 8 we have the inequality*

$$\sum_{s\in U_1} R_s + \sum_{s\in U_2} R_s + \sum_{s\in U_3} R_s + \sum_{s\in U_4} R_s \leq C_{12} + C_{32} + C_{14} + C_{34} \tag{2.15}$$

where:

- $U_1 = \Big\{ s : \nu_s \in Q_1, D_s \cap (Q_2 \cup Q_3 \cup Q_4) \neq \emptyset \Big\} \cup \Big\{ \nu_s \in Q_3, D_s \cap (Q_1 \cup Q_2 \cup Q_4) \neq \emptyset \Big\}$
- $U_2 = \Big\{ s : \nu_s \in Q_2, D_s \cap Q_4 \neq \emptyset \Big\}$
- $U_3 = \Big\{ s : \nu_s \in Q_4, D_s \cap Q_2 \neq \emptyset \Big\}$
- $U_4 = \Big\{ s : \nu_s \in (Q_1 \cup Q_3), D_s \cap Q_2 \neq \emptyset, D_s \cap Q_4 \neq \emptyset \Big\}$

and C_{ij} denotes the portion of the edge capacity directed from Q_i to Q_j.

Proof Consider a network in which the number of vertices and the capacities C_{12}, C_{21}, C_{14}, C_{41}, C_{32}, C_{23}, C_{34}, and C_{43} are the same as in our original network in Figure 7, but all other edge capacities are infinite. Assume the same traffic demands in the new network as in the original ring. A network coding solution that achieves the demands of the original ring is also a solution for this new ring. On the other hand, all capacities are infinite within any four groups of vertices Q_1, Q_2, Q_3, or Q_4, so each group can be treated as a single supervertex. Hence, our previous bound (2.14) for the ring of four vertices continues to hold for this modified ring.

□

We next use inequality (2.15) to show that any network code must satisfy the routing bound corresponding to $\mathbf{b}_n$ in an undirected ring with $n \geq 4$ vertices. Consider inequality (2.15), and let $E(i, j)$ denote the inequality derived by setting Q_2 and Q_4 to be two vertices i and j. For the values of i or j not between 1 and n, we consider their value modulo n in $E(i, j)$. We separately study the two cases of interest.

1. Proof of the network coding bound for line sessions for $n \geq 4$

Consider a line session s. By using (2.15) we wish to find the coefficient of R_s in $E(i, j)$. First suppose that $|D_s \cap \{i, j\}| = 0$. By (2.15), s does not belong to U_2, U_3, or U_4. Furthermore since all source and destination vertices of s are adjacent, s is not in U_1. Therefore the coefficient of R_s in this case is zero. Next suppose that $|D_s \cap \{i, j\}| = 1$. In this case, s belongs to one of the sets $U_1, U_2,$ or U_3, but not two of them together. Therefore the coefficient of R_s in this case is one. Finally suppose that $|D_s \cap \{i, j\}| = 2$. In this case s belongs to both U_1 and U_4, and therefore the coefficient of R_s is two. As a summary of these cases $E(i,j)$ can be written as follows:

$$\sum_{s \text{ is of Type 1}} |D_s \cap \{i,j\}| R_s \leq C_{(i-1)(i)} + C_{(i+1)(i)} + C_{(j-1)(j)} + C_{(j+1)(j)}. \tag{2.16}$$

Next we derive the network coding bound for this case and show that it is the same as the routing bound. Suppose that $n = 2m$ is an even integer. Then consider the sum $\sum_{i=1}^{m} E(i, m+i)$, which can be expanded as follows:

$$\begin{aligned} \sum_{i=1}^{m} \sum_{s \text{ is of Type 1}} |D_s \cap \{i, i+m\}| R_s = \\ \sum_{s \text{ is of Type 1}} \Big(\sum_{i=1}^{m} |D_s \cap \{i, i+m\}| \Big) R_s \leq \\ \sum_{i=1}^{m} \Big(C_{(i-1)(i)} + C_{(i+1)(i)} + C_{(i+m-1)(i+m)} + C_{(i+m+1)(i+m)} \Big). \end{aligned} \tag{2.17}$$

Every directed capacity appears exactly once on the right hand side of (2.17), and thus it is equal to $\sum_{i=1}^{n} C_i$. Furthermore $\sum_{i=1}^{m} |D_s \cap \{i, i+m\}| = |D_s|$. Therefore (2.17) results in the following inequality:

$$\sum_{s \text{ is of Type 1}} |D_s| R_s \leq \sum_{i=1}^{n} C_i. \tag{2.18}$$

Since for a line session s the source and destination vertices form a string of adjacent vertices, it follows that $\ell_{\mathbf{b}_n}(s)$ is the number of destination vertices of s on the ring. Hence we obtain a network coding bound which is the same as the routing inequality for this case.

Next, suppose that $n = 2m+1$ is an odd number and consider the sum $0.5\big(\sum_{i=1}^{n} E(i, m+$

$i)\big)$. Using (2.16), this sum can be expanded as follows.

$$0.5\Big(\sum_{i=1}^{n}\sum_{s \text{ is of Type 1}} |D_s \cap \{i, i+m\}| R_s\Big) = \tag{2.19}$$

$$0.5\Big(\sum_{s \text{ is of Type 1}} \big(\sum_{i=1}^{n} |D_s \cap \{i, i+m\}|\big) R_s\Big) \leq \tag{2.20}$$

$$0.5\Big(\sum_{i=1}^{n} \big(C_{(i-1)(i)} + C_{(i+1)(i)} + C_{(i+m-1)(i+m)} + C_{(i+m+1)(i+m)}\big)\Big). \tag{2.21}$$

Every directed capacity $C_{(i+1)(i)}$ appears in two terms of the summation of (2.21) which are the terms corresponding to $E(i, i+m)$ and $E(i-m, i)$. Therefore (2.21) is $\sum_{i=1}^{n} C_i$. Next consider that in $\big(\sum_{i=1}^{n} |D_s \cap \{i, i+m\}|\big)$ in the expression (2.20), every destination $d_s \in D_s$ is counted twice, namely in the terms corresponding to $E(d_s, d_s + m)$ and $E(d_s - m, d_s)$. Therefore (2.20) is $0.5\Big(\sum_{s \text{ is of Type 1}} \big(2|D_s|\big) R_s\Big) = \sum_{s \text{ is of Type 1}} \ell_{\mathbf{b}_n}(s) R_s$, and hence the final result follows.

2. Proof of the network coding bound for unicast and broadcast sessions for $n \geq 4$

Consider again the cases $n = 2m$ and $n = 2m + 1$ separately. Let $E_s(i, j)$ denote the coefficient of R_s on the left hand side of $E(i, j)$ when s is of Type 2. First notice that by definition, for a unicast session s with source vertex ν_s and destination vertex d_s, if ν_s and d_s are on two sides of the ring which are separated by vertices i and j, or if d_s is i or j, then $E_s(i, j)$ is one; otherwise it is zero. Furthermore, since a broadcast session s is a special case of a line session it follows that $E_s(i, j)$ is $|D_s \cap \{i, j\}|$.

For $n = 2m$ we consider the sum $\sum_{i=1}^{m} E(i, m+i)$. Our previous arguments for line sessions implies that the right hand side of this summation is $\sum_{i=1}^{n} C_i$. The following is the left hand side of the summation:

$$\sum_{i=1}^{m} \sum_{s \text{ is broadcast}} |D_s \cap \{i, i+m\}| R_s + \sum_{i=1}^{m} \sum_{s \text{ is unicast}} E_s(i, i+m) R_s =$$

$$\sum_{s \text{ is broadcast}} \ell_{\mathbf{b}_n}(s) R_s + \sum_{s \text{ is unicast}} \big(\sum_{i=1}^{m} E_s(i, i+m)\big) R_s. \tag{2.22}$$

The coefficient of a broadcast session in (2.22) follows from the argument for line sessions. For a unicast session s consider $\sum_{i=1}^{m} E_s(i, i+m)$ and without loss of generality assume that $\nu_s = 1$ and $d_s \leq m$. Then the nonzero terms of the summation are $E_s(2, m+2) = \cdots = E_s(d_s, m+d_s) = 1$. Since $\ell_{\mathbf{b}_n}(s) = d_s - 1$, the coefficient of R_s will be $\ell_{\mathbf{b}_n}(s)$ and therefore the network coding inequality of the form

$$\sum_{s \text{ is of Type 2}} \ell_{\mathbf{b}_n}(s) R_s \leq \sum_{i=1}^{n} C_i \tag{2.23}$$

is obtained in this setting, which is the same as the corresponding routing bound.

For $n = 2m+1$, we can obtain the same inequality by instead considering

$$0.5\big(\sum_{i=1}^{n} E(i, m+i)\big).$$

We consider the counterpart of (2.19)-(2.21) for this case. By the argument for line sessions, the right hand side of this summation is $\sum_{i=1}^{n} C_i$. Next we expand the left hand side of this summation:

$$0.5\Big(\sum_{i=1}^{n} \sum_{s \text{ is broadcast}} |D_s \cap \{i, i+m\}| R_s\Big) + 0.5\Big(\sum_{i=1}^{n} \sum_{s \text{ is unicast}} E_s(i, i+m) R_s\Big) =$$

$$\sum_{s \text{ is broadcast}} \ell_{\mathbf{b}_n}(s) R_s + 0.5\Big(\sum_{s \text{ is unicast}} \big(\sum_{i=1}^{n} E_s(i, i+m)\big) R_s\Big). \tag{2.24}$$

For a unicast session s with $\nu_s = 1$ and $d_s \leq m+1$ consider $\sum_{i=1}^{n} E_s(i, i+m)$. The nonzero terms of this summation are $E_s(2, m+2) = \cdots = E_s(d_s, m+d_s) = 1$ and $E_s(2, m+3) = \cdots = E_s(d_s, d_s+m+1) = 1$. Since $\ell_{\mathbf{b}_n}(s) = d_s - 1$, $\sum_{i=1}^{n} E_s(i, i+m) =$

$2\ell_{\mathbf{b}_n}(s)$ and therefore the network coding inequality of the form

$$\sum_{s \text{ is of Type 2}} \ell_{\mathbf{b}_n}(s) R_s \leq \sum_{i=1}^{n} C_i \tag{2.25}$$

follows for this setting and is the same as the corresponding routing inequality.

CHAPTER III

NETWORK CODING IN NODE–CONSTRAINED LINE AND STAR NETWORKS

A. Introduction

In this chapter we study the network coding capacity of node-constrained line and star networks as depicted in Figure 3. There are several reasons for studying node-constrained line and star networks.

1. Node-constrained networks are an interesting class of networks in their own right.

2. *Limited processing capacity:* In some applications, network throughput is constrained by the limited processing capacity of the nodes, e.g., the limited bus bandwidth between different processing or memory units and/or the limited speed of processors and/or limited communication ports. We refer to such limitations generically as node constraints. Node constraints arise in sensor or satellite networks where the energy resources are limited and low complexity and/or low power processors are needed. They may also arise in optical networks where the edges have very high capacity and the rates are restricted by the speed of the node processors. For example a single optical fiber can carry information at many terabits per second over hundreds or thousands of kilometers [3, 24, 98], while typical processing units process data with rates of up to perhaps hundreds of gigabits per second.

   We may model the limitations of processing data at some node j by adding a node capacity C_j as in Figure 2. In this model, all traffic that originates at node j, all traffic that is destined for node j, and all traffic that is relayed by node j are limited by the capacity C_j. Of course, this model is a simplification of real processors for which these three traffic classes might face different routes and different capacity constraints. To capture these effects, there are alternative models for node–

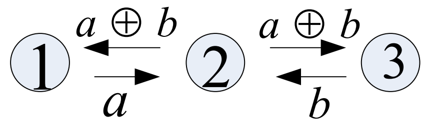

Fig. 9. Nodes 1 and 3 exchange their information through node C.

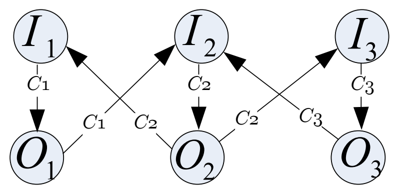

Fig. 10. Wired model of wireless network in Figure 9.

constrained networks where only a part of the traffic is limited by the node constraint [26, 30]. Our network coding techniques can be extended to these models with little change. For instance, a variation of our model in which only the relayed traffic and the traffic that is originated at node j pass through the edge with capacity C_j is discussed in the Appendix H. We remark that the above processing model does not limit the complexity of the network coding scheme and may therefore not properly represent the nodes' true computational limitations. However, as we will observe later, all our coding schemes are low complexity schemes over the binary field. Furthermore, our codes are one–shot schemes that operate only on one generation of messages at each time instant and therefore have the same memory requirement as routing schemes. We thus expect that some choice of capacity C_j is a reasonable measure of the available node resources.

3. *Network coding in wireless line networks:* Network coding can be beneficial in wireless networks [40, 45, 50]. A standard example is the wireless network in Figure

9 where nodes 1 and 3 wish to exchange the bits a and b through the relay node 2. Because of the broadcast nature of the wireless medium, nodes can receive the data transmitted from neighboring senders. If the nodes deploy routing strategies, four transmissions are needed; with network coding only three transmissions are needed by using the code shown in Figure 9. One way to apply the results from the network coding literature to wireless networks is to transform a wireless network into a wired network. Wu et al. [97] consider such a transformation: each node j in the *wireless* network corresponds to a node I_j and a virtual node O_j in the *wired* model. The virtual node plays the role of a bottleneck that carries the traffic sent by node j. If $\mathcal{N}_j$ denotes the set of outgoing neighbors (or descendants) of node j in the wireless network, then in the wired network there is a link of capacity C_j that connects I_j to O_j and there are links each of capacity C_j connecting O_j to all nodes I_k corresponding to nodes k in $\mathcal{N}_j$. The capacity C_j is chosen based on "physical layer" considerations. The wired network corresponding to network of Figure 9 is depicted in Figure 10. If we apply the procedure to a line of M wireless nodes, we end up with the node–constrained line network of Figure 3, where $C_{j,k} = C_j$ for all j and k. The benefits of using network coding for the model of Figure 3 have been studied in [95].

While the network coding scheme in [95] is similar to our scheme, our work is different in several ways. First, [95] considers the broadcast scenario where the information that originates at each sender will be decoded by all receivers. Our scheme, on the other hand, supports *multiple multicast*. Second, the model of [95] assumes unit capacity edges while we assume that the capacities of edges can be arbitrary positive integers. Finally, we point out that in applying our network coding scheme to the wired model of wireless networks we have to take some care. One issue is that the virtual node O_j in the wired model is not a real node and thus it can not perform any network coding operation and may only forward data to the nodes in $\mathcal{N}_j$. The other issue is that in the model of [95] node I_j is both the sender and receiver of messages but in our model node I_j is the sender of messages and node O_j is the receiver of messages. We revisit these issues in the Appendix H.

4. *Index coding with side information:* Consider a wireless broadcast network where a sender node s has access to a vector of bits $[b_1, \cdots, b_n]$. At each time instant, s can broadcast a bit as a function of $[b_1, \cdots, b_n]$ to n users $d_1, d_2, \cdots, d_n$. Each user $i \in \{1, \cdots, n\}$ is interested in bit b_i and knows a fixed subset of bits in $[b_1, \cdots, b_n]$. The *index coding problem with side information* asks for the minimum number of transmissions such that after all transmissions are completed, each node d_i can decode the bit b_i from the transmitted information and the bits that it knows. The authors of [10] introduced the index coding problem with side information and made connections with some earlier communication problems such as the *"coding on demand by an informed source"* problem. They further proposed a subset of linear coding schemes that includes the optimal linear scheme in terms of the number of transmissions. The paper [56] disproved a conjecture in [10] on the optimality of linear index coding by constructing an instance of the index coding problem in which any linear scheme over any field is suboptimal. Finally, the paper [21] established that a general multiple unicast network coding problem in an acyclic network can be reduced to an instance of the index coding problem.

In this chapter we establish connections between node–constrained star networks and index coding with side information. Our proposed scheme is optimal for certain classes of problems. However, we also show that the network coding problem in a node–constrained star network with only a node constraint at the central node and all other constraints removed is a special case of the index coding with side information problem. Using this connection, we give an example for which the linear coding scheme proposed for the general index coding problem improves upon our coding schemes for node–constrained star networks.

This chapter is organized as follows. In Section B we discuss the general network coding problem in line networks with multiple multicast sessions. We use cut bounds and entropy arguments [49, 15] based on edge cuts to find an outer bound on the network coding capacity regions. We subsequently propose a linear coding scheme that achieves the outer bound. In Section C we consider star networks and provide upper bounds on their multiple unicast capacity region. We describe a coding and decoding scheme for each network node that is obtained by finding the maximum number of edge disjoint cycles in a graph called the *demand graph* corresponding to the communication problem. We demonstrate that our coding scheme is optimal for a broad class of demand graphs. We next consider the relationships between the network coding problem in star networks and the index coding problem with a side information [10]. In the Appendix H we discuss another variation of the node–constrained model for line networks and characterize the capacity region and a capacity–achieving network coding scheme.

B. Network Coding in Line Networks

We represent a line network with the corresponding directed graph $G(V, E)$ as in Figure 3 where $V = \{I_1, \cdots, I_M, O_1, \cdots, O_M\}$. The outgoing symbol from every node, at any particular time instant t, can be any function of the incoming symbols to that node at earlier time instants $1, 2, \cdots, t-1$ and/or its own messages at the present and earlier time instants. As we will see later in this section, our scheme depends on a small number of previous messages at each node and therefore has a low encoding and decoding complexity.

Consider a network coding session which runs from time $t = 1$ to $t = T$. We introduce the following notation:

- $X_i^{(t)}$ represents the binary vector transmitted on edge (I_i, O_i) at time t.
- $X_{i,j}^{(t)}$ represents the binary vector transmitted on edge (O_i, I_j) at time t.
- S is the set of sessions in the network and for every $s \in S$, there is a corresponding source node $\nu_s \in \{I_1, \cdots, I_M\}$ and a set of destination nodes $\mathcal{D}_s$, where $\mathcal{D}_s \subset \{O_1, \ldots, O_M\}$ and $O_j \notin \mathcal{D}_s$ if $\nu_s = I_j$. We generally represent session s by $\nu_s \to \mathcal{D}_s$.
- $W_s^{(t)}$ is the message of session s at time t.
- R_s is an integer that represents the rate of session s, i.e., $H(W_s^{(t)}) = R_s$ for all t.
- C_i is the capacity of the edge (I_i, O_i) and $C_{i,j}$ is the capacity of the edge (O_i, I_j). In other words, we require $H(X_i^{(t)}) \leq C_i$ and $H(X_{i,j}^{(t)}) \leq C_{i,j}$ for all t.
- $Y^t = \left[Y^{(1)}, \cdots, Y^{(t)}\right]$.
- $[W_{s_1}, \cdots, W_{s_k}]$ the vector formed by concatenation of the messages $W_{s_1}, \cdots, W_{s_k}$.

Our main result for line networks with node constraints is summarized in the following theorem.

Theorem 5 *A non–negative rate tuple $(R_s : s \in S)$ is achievable in the node-constrained line network of Figure 3 if and only if it satisfies the following bounds:*
For every $i \in \{1, \cdots, M\}$

$$\sum_{\nu_s = I_i} R_s + \sum_{O_i \in \mathcal{D}_s} R_s + \max\left\{\sum_{s \in U_1} R_s, \sum_{s \in U_2} R_s\right\} \leq C_i \tag{3.1}$$

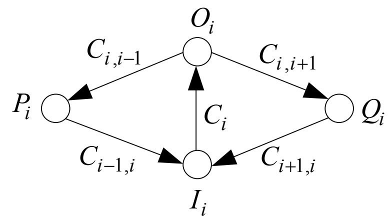

Fig. 11. Network G' made from network G.

where U_1 *and* U_2 *are defined as follows:*

$$U_1 = \{s : \nu_s = I_j, j < i\} \bigcap \{s : O_i \notin \mathcal{D}_s\} \bigcap \{s : \mathcal{D}_s \cap \{O_{i+1}, \cdots, O_M\} \neq \emptyset\}$$
$$U_2 = \{s : \nu_s = I_j, j > i\} \bigcap \{s : O_i \notin \mathcal{D}_s\} \bigcap \{s : \mathcal{D}_s \cap \{O_1, \cdots, O_{i-1}\} \neq \emptyset\},$$

and for every $i \in \{1, \cdots, M\}$

$$\sum_{s:\nu_s \in \{I_j : j \geq i\}, \mathcal{D}_s \cap \{O_j : j < i\} \neq \emptyset,} R_s \leq C_{i,i-1}, \tag{3.2}$$
$$\sum_{s:\nu_s \in \{I_j : j \leq i\}, \mathcal{D}_s \cap \{O_j : j > i\} \neq \emptyset,} R_s \leq C_{i,i+1}.$$

The proof of Theorem 5 has two parts. We first determine upper bounds on the set of achievable rates and then show that the upper bounds are tight by proposing an achievable scheme.

1. Capacity upper bounds for line networks

We begin by analyzing the constraint of the edges $(I_i, O_i), i = 1, 2, \cdots, M$. As the first step, we form network G' from the line network G by increasing the capacities of all edges but (I_i, O_i), (O_{i-1}, I_i), (O_{i+1}, I_i), (O_i, I_{i-1}) and (O_i, I_{i+1}) in G to infinity and then collecting all nodes on the left–hand side of the edge (I_i, O_i) as a single node P_i and all nodes on the right–hand side of the edge (I_i, O_i) as a single node Q_i. More formally, $P_i \triangleq$

$\{I_1, I_2, \ldots, I_{i-1}, O_1, O_2, \ldots, O_{i-1}\}$ and $Q_i \triangleq \{I_{i+1}, I_{i+2}, \ldots, I_M, O_{i+1}, O_{i+2}, \ldots, O_M\}$. G' is illustrated in Figure 11 and is the result of this process. G' has a set of sessions S'. Let $\mathcal{D}'_{s'}$ be the set of all nodes in the graph G that form the elements of the set $\mathcal{D}_{s'}$. A session $s' \in S'$ has a source $\nu_{s'} \in \{P_i, Q_i, \{I_i\}\}$ and a set of destinations $\mathcal{D}_{s'} \subseteq \{P_i, Q_i, \{O_i\}\}$ and its corresponding message $W_{s'}$ is a concatenation of the messages corresponding to sessions in the set

$$S_{s'} = \{s : \nu_s \in \nu_{s'}, \mathcal{D}_s \cap \mathcal{D}'_{s'} \neq \emptyset\}.$$

It is obvious that the upper bounds on the achievable rates in the network G' are also upper bounds on the achievable rates in the network G if we replace the rate of a session s' in G' by $\sum_{s \in S_{s'}} R_s$ in G. If there is a session s in G for which the set $\nu_s \cup \mathcal{D}_s$ is a subset of P_i or Q_i, then session s clearly will not be part of any session in G'. Next we find an upper bound on the achievable rates of the network G' in terms of the capacity C_i. Consider the vector of the middle edge X_i^T:

$$\begin{aligned}
& H\left(X_i^T, \left\{W_{s'}^T : \nu_{s'} = P_i, O_i \notin \mathcal{D}_{s'}\right\}\right) \\
& \overset{a}{\geq} H\left(\left\{W_{s'}^T : \nu_{s'} = P_i\right\}, \left\{W_{s'}^T : O_i \in \mathcal{D}_{s'}\right\}, X_{i,i-1}^T, X_{i,i+1}^T\right) \\
& \overset{b}{\geq} H\left(\left\{W_{s'}^T : O_i \in \mathcal{D}_{s'}\right\}, \left\{W_{s'}^T : P_i \in \mathcal{D}_{s'}\right\}, X_{i,i+1}^T\right) \\
& \overset{c}{=} H\left(\left\{W_{s'}^T : \nu_{s'} = Q_i\right\}, \left\{W_{s'}^T : O_i \in \mathcal{D}_{s'}\right\}, \left\{W_{s'}^T : P_i \in \mathcal{D}_{s'}\right\}, X_{i,i+1}^T\right) \\
& \overset{d}{\geq} H\left(\left\{W_{s'}^T : O_i \in \mathcal{D}_{s'}\right\}, \left\{W_{s'}^T : P_i \in \mathcal{D}_{s'}\right\}, \left\{W_{s'}^T : Q_i \in \mathcal{D}_{s'}\right\}\right) \\
& \overset{e}{=} H\left(\left\{W_{s'}^T : s' \in S'\right\}\right).
\end{aligned} \tag{3.3}$$

In (3.3), a holds because we assume that messages are perfectly decoded at their destinations and because the decoded messages and outgoing flows at O_i are functions of X_i^T. b holds because the decoded messages at P_i are functions of $X_{i,i-1}^T$ and $\left\{W_{s'}^T : \nu_{s'} = P_i\right\}$ (recall that we ignore all messages originating in P_i and destined for nodes in P_i only, and similarly for Q_i). c holds because the set of messages with the source at Q_i is a subset of the set of messages with a destination at O_i or P_i. d holds because the decoded messages at Q_i are functions of $X_{i,i+1}^T$ and $\left\{W_{s'}^T : \nu_{s'} = Q_i\right\}$. Finally, e holds since the previous expression includes all messages in the network. By the entropy inequality $H\left(X_i^T, \left\{W_{s'}^T : \nu_{s'} = P_i, O_i \notin \mathcal{D}_{s'}\right\}\right) \leq H\left(X_i^T\right) + H\left(\left\{W_{s'}^T : \nu_{s'} = P_i, O_i \notin \mathcal{D}_{s'}\right\}\right)$ and (3.3), we have:

$$\begin{aligned} TC_i &\geq H\left(X_i^T\right) \\ &\geq H\left(\left\{W_{s'}^T : s' \in S'\right\}\right) - H\left(\left\{W_{s'}^T : \nu_{s'} = P_i, O_i \notin \mathcal{D}_{s'}\right\}\right) \\ &= T\left(\sum_{\nu_{s'}=I_i} R_{s'} + \sum_{\nu_{s'}=Q_i} R_{s'} + \sum_{\nu_{s'}=P_i, O_i \in \mathcal{D}_{s'}} R_{s'}\right). \end{aligned} \tag{3.4}$$

By the symmetry of the network with respect to P_i and Q_i we likewise have:

$$C_i \geq \sum_{\nu_{s'}=I_i} R_{s'} + \sum_{\nu_{s'}=P_i} R_{s'} + \sum_{\nu_{s'}=Q_i, O_i \in \mathcal{D}_{s'}} R_{s'}. \tag{3.5}$$

Since $\sum_{v_{s'}=Q_i} R_{s'} = \sum_{v_{s'}=Q_i, O_i \in \mathcal{D}_{s'}} R_{s'} + \sum_{v_{s'}=Q_i, O_i \notin \mathcal{D}_{s'}} R_{s'}$, if we replace $R_{s'}$ by $\sum_{s \in S_{s'}} R_s$ in the bounds in (3.4), (3.5), we obtain the bound (3.1) in the original network G.

Next we will use cut set bounds to derive bounds on the capacity of edges the (O_i, I_{i-1}) and (O_i, I_{i+1}). The cut set bound starts with a subset $U \subseteq V$ of the nodes of network graph G. 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 U of nodes, the corresponding cut set bound is:

$$\sum_{s : \nu_s \in U, \mathcal{D}_s \cap U^c \neq \emptyset} R_s \leq C_{U \to U^c}.$$

For our purpose we use the partition $U_1 = \{O_j, I_j : j \geq i\}$ for the bound on the edge (O_i, I_{i-1}) and we use the partition $U_2 = \{O_j, I_j : j \leq i\}$ for the bound on the edge (O_i, I_{i+1}).

Remark 6 *The bounds (3.1) are progressive d-separating edge cut bounds by choosing the* *edge set $\mathcal{E} = \{I_i, O_i\}$ and the source set $\mathcal{S}$ in [49] as one of $\mathcal{S}_1 = \{s : \nu_s = I_i\} \bigcup \{s : O_i \in \mathcal{D}_s\} \bigcup U_1$ and $\mathcal{S}_2 = \{s : \nu_s = I_i\} \bigcup \{s : O_i \in \mathcal{D}_s\} \bigcup U_2$.*

Remark 7 *The bounds (3.1) and (3.2) are the ones we will need to establish the capacity region. However, one might also guess that applying the cut set bound to the graph G' will recover (3.1). A quick check on Figure 11 shows that this is not the case. In other words, the bound (3.1) is not a cut set bound for either G or G' in general.*

2. Network coding scheme for line networks

a. Network coding scheme

Next we provide a network coding scheme that achieves the bounds in Theorem 5. This proves that those bounds describe the network coding capacity of line networks. We begin by introducing the following notation for different collections of multicast or unicast messages with respect to some fixed node $i \in \{1, \cdots, M\}$.

Let $L_i \triangleq \{1, \cdots, i-1\}$ and $R_i \triangleq \{i+1, \cdots, M\}$ respectively denote the set of nodes on the left hand side and right hand side of node i. Let $A, B, C \in \{i, L_i, R_i\}$ and define

- $W^{(t)}_{A\to B,C} \triangleq \left[W^{(t')}_s : \nu_s \in A, \mathcal{D}_s \cap B \neq \emptyset, \mathcal{D}_s \cap C \neq \emptyset, t' = t - |\nu_s - i|\right],$
- $W^{(t)}_{A\to \overline{B},C} \triangleq \left[W^{(t')}_s : \nu_s \in A, \mathcal{D}_s \cap B = \emptyset, \mathcal{D}_s \cap C \neq \emptyset, t' = t - |\nu_s - i|\right],$
- $W^{(t)}_{A\to B,\overline{C}} \triangleq \left[W^{(t')}_s : \nu_s \in A, \mathcal{D}_s \cap B \neq \emptyset, \mathcal{D}_s \cap C = \emptyset, t' = t - |\nu_s - i|\right].$

It will become clear from our network coding scheme that the order in which the messages of different sessions appear in each of the preceding vectors does not matter as long as it is fixed throughout the scheme described below.

As an example $W^{(t)}_{i\to L_i,\bar{R}_i}$ denotes the concatenation of messages generated at time t at node i with at least one destination k with $k < i$ and with no destination l with $l > i$.

Also $W^{(t)}_{L_i \to i, R_i}$ denotes the concatenation of messages with source j with $j < i$, generated at time instant $t - (i - j)$ with destination at i and at least one destination k with $k > i$.

One should also notice that in a network coding scheme the outgoing message from every node at time instant t is some function of its incoming messages at times $1, 2, \cdots, t-1$ and its own messages at times $1, 2, \cdots, t$ if it is a sender node. Further, to simplify our notation we assume that inside a single node i, the delay between its sender and receiver (which are indicated by I_i and O_i for node i) is negligible but the delay of transmission for any messages between the receiver of a node and the sender of its neighbor node is one unit of time.

Remark 8 *Although the analysis here assumes zero delay in the node processors, with few changes in the time indices the results can be shown to remain valid for the case of positive delays in the node processors.*

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 binary vectors of lengths n_a and n_b respectively. We define:

$$\mathbf{a} \oplus \mathbf{b} = \begin{cases} [a_1 \oplus b_1, a_2 \oplus b_2, \cdots, a_{n_a} \oplus b_{n_a}] & \text{if } n_a \leq n_b, \\ [a_1 \oplus b_1, a_2 \oplus b_2, \cdots, a_{n_b} \oplus b_{n_b}, a_{n_b+1}, \cdots, a_{n_a}] & \text{if } n_a > n_b. \end{cases}$$

Here $a_i \oplus b_i$ is the bitwise XOR of the ith bits of $\mathbf{a}$ and $\mathbf{b}$. Furthermore we define:

$$\mathbf{a} \otimes \mathbf{b} = \begin{cases} \mathbf{b} \oplus \mathbf{a} & \text{if } n_a \leq n_b, \\ \mathbf{a} \oplus \mathbf{b} & \text{if } n_a > n_b. \end{cases}$$

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

Suppose that all source nodes start transmitting at time $t = 0$ and messages at negative

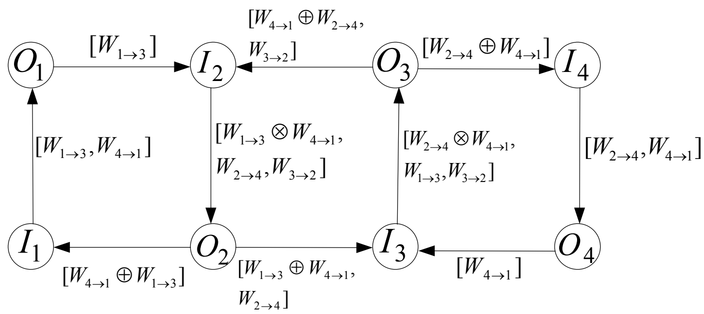

Fig. 12. A network code for a 4 node line network with sessions $1 \to 3, 2 \to 4, 3 \to 2$, and $4 \to 1$.

time instants are assumed to take the value zero. Our network coding scheme consists of three parts, which respectively describe the vectors $X_i^{(t)}, X_{i,i-1}^{(t)}$ and $X_{i,i+1}^{(t)}$ for any node i:

$$X_i^{(t)} \triangleq [W_{R_i\to\bar{i},L_i}^{(t)} \otimes W_{L_i\to\bar{i},R_i}^{(t)}, W_{R_i\to i,L_i}^{(t)}, W_{L_i\to i,R_i}^{(t)}, W_{R_i\to i,\bar{L}_i}^{(t)}, W_{L_i\to i,\bar{R}_i}^{(t)}, W_{i\to L_i,R_i}^{(t)}, W_{i\to\bar{L}_i,R_i}^{(t)}, W_{i\to L_i,\bar{R}_i}^{(t)}], \tag{3.6}$$

$$X_{i,i-1}^{(t)} \triangleq \left[W_{R_i\to\bar{i},L_i}^{(t)} \oplus W_{L_i\to\bar{i},R_i}^{(t)}, W_{R_i\to i,L_i}^{(t)}, W_{i\to L_i,\bar{R}_i}^{(t)}, W_{i\to L_i,R_i}^{(t)}\right], \tag{3.7}$$

$$X_{i,i+1}^{(t)} \triangleq \left[W_{L_i\to\bar{i},R_i}^{(t)} \oplus W_{R_i\to\bar{i},L_i}^{(t)}, W_{L_i\to i,R_i}^{(t)}, W_{i\to\bar{L}_i,R_i}^{(t)}, W_{i\to L_i,R_i}^{(t)}\right]. \tag{3.8}$$

For an example of our network code for multiple unicast sessions see Figure 12. In order for (3.6), (3.7), and (3.8) to define a valid network coding scheme we need to demonstrate that for any node on the network, the outgoing flows from that node are some functions of the incoming flows to that node at earlier time instants and the messages generated at that node at earlier or present time instants. For this purpose we define two auxiliary vectors of messages as follows

$$F_i^{(t)} \triangleq \left[W^{(t)}_{R_i\to\bar{i},L_i}, W^{(t)}_{R_i\to i,\bar{L}_i}, W^{(t)}_{R_i\to i,L_i}\right], \tag{3.9}$$

$$G_i^{(t)} \triangleq \left[W^{(t)}_{L_i\to\bar{i},R_i}, W^{(t)}_{L_i\to i,\bar{R}_i}, W^{(t)}_{L_i\to i,R_i}\right]. \tag{3.10}$$

We notice that the information available to node I_i at time instant t is $X^{t-1}_{i-1,i}$, $X^{t-1}_{i+1,i}$, $W^t_{i\to L_i,R_i}$, $W^t_{i\to\bar{L}_i,R_i}$, $W^t_{i\to L_i,\bar{R}_i}$ and the information available to node O_i at time instant t is X_i^{t-1}. Next we will prove the following result:

Theorem 9 *At time instant t, vectors $F_i^{(t)}$ and $G_i^{(t)}$ as defined in (3.9) and (3.10) respectively and $X_i^{(t)}$ as defined in (3.6) are functions of the information available to node I_i at time instant t and $X^{(t)}_{i,i-1}$ and $X^{(t)}_{i,i+1}$ as defined in (3.7) and (3.8) respectively, are functions of the information available to node O_i at time instant t.*

Proof We use induction on time instant t to prove our claim. For $t = 0$ all vectors in Theorem 9 are zero vectors and the claim holds trivially. Suppose that for all time instants $t \leq n-1$ and all nodes $i \in \{1, \cdots, M\}$ Theorem 9 holds. Next we prove it for time instant $t = n$. As the first step we show the following equivalence of vectors for any time instant t:

$$\begin{aligned} F_i^{(t)} &= \left[W^{(t)}_{R_i\to\bar{i},L_i}, W^{(t)}_{R_i\to i,\bar{L}_i}, W^{(t)}_{R_i\to i,L_i}\right] \text{ is a permutation of} \\ &\quad \left[W^{(t-1)}_{R_{i+1}\to\overline{i+1},L_{i+1}}, W^{(t-1)}_{R_{i+1}\to i+1,L_{i+1}}, W^{(t-1)}_{i+1\to L_{i+1},R_{i+1}}, W^{(t-1)}_{i+1\to L_{i+1},\bar{R}_{i+1}}\right], \end{aligned} \tag{3.11}$$

$$\begin{aligned} G_i^{(t)} &= \left[W^{(t)}_{L_i\to\bar{i},R_i}, W^{(t)}_{L_i\to i,\bar{R}_i}, W^{(t)}_{L_i\to i,R_i}\right] \text{ is a permutation of} \\ &\quad \left[W^{(t-1)}_{L_{i-1}\to\overline{i-1},R_{i-1}}, W^{(t-1)}_{L_{i-1}\to i-1,R_{i-1}}, W^{(t-1)}_{i-1\to L_{i-1},R_{i-1}}, W^{(t-1)}_{i-1\to R_{i-1},\bar{L}_{i-1}}\right]. \end{aligned} \tag{3.12}$$

To see this, observe that the left hand side of (3.11) is the 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 $k \leq i$. This set of messages is either generated at I_j with $j > i+1$ at time instant $t-(j-i)$ or at I_{i+1} at time instant $t-1$. By definition the former group of messages is identical to $\left[W^{(t-1)}_{R_{i+1}\to\overline{i+1},L_{i+1}}, W^{(t-1)}_{R_{i+1}\to i+1,L_{i+1}}\right]$, and the latter group of messages is identical to $\left[W^{(t-1)}_{i+1\to L_{i+1},R_{i+1}}, W^{(t-1)}_{i+1\to L_{i+1},\bar{R}_{i+1}}\right]$; these together form the right hand side of (3.11). An analogous argument holds for (3.12).

By setting $t = n$ in (3.11) and (3.12) we see that $F_i^{(n)}$ and $G_i^{(n)}$ are functions of the information available to I_i at time instant n if and only if the right hand sides of (3.11) and (3.12) are functions of the information available to I_i at time instant n. At time instant n node I_i has access to the vectors $X^{(n-1)}_{i-1,i}$ and $X^{(n-1)}_{i+1,i}$. Our induction hypothesis and equations (3.7) and (3.8) imply

$$X^{(n-1)}_{i+1,i} = \tag{3.13}$$
$$\left[W^{(n-1)}_{R_{i+1}\to\overline{i+1},L_{i+1}} \oplus W^{(n-1)}_{L_{i+1}\to\overline{i+1},R_{i+1}}, W^{(n-1)}_{R_{i+1}\to i+1,L_{i+1}}, W^{(n-1)}_{i+1\to L_{i+1},\bar{R}_{i+1}}, W^{(n-1)}_{i+1\to L_{i+1},R_{i+1}}\right],$$
$$X^{(n-1)}_{i-1,i} = \tag{3.14}$$
$$\left[W^{(n-1)}_{L_{i-1}\to\overline{i-1},R_{i-1}} \oplus W^{(n-1)}_{R_{i-1}\to\overline{i-1},L_{i-1}}, W^{(n-1)}_{L_{i-1}\to i-1,R_{i-1}}, W^{(n-1)}_{i-1\to\bar{L}_{i-1},R_{i-1}}, W^{(n-1)}_{i-1\to L_{i-1},R_{i-1}}\right].$$

Hence I_i can extract messages

$$\left[W^{(n-1)}_{R_{i+1}\to i+1,L_{i+1}}, W^{(n-1)}_{i+1\to L_{i+1},R_{i+1}}, W^{(n-1)}_{i+1\to L_{i+1},\bar{R}_{i+1}}\right],$$
$$\left[W^{(n-1)}_{L_{i-1}\to i-1,R_{i-1}}, W^{(n-1)}_{i-1\to L_{i-1},R_{i-1}}, W^{(n-1)}_{i-1\to\bar{L}_{i-1},R_{i-1}}\right].$$

directly from the received messages $X^{(n-1)}_{i-1,i}$ and $X^{(n-1)}_{i+1,i}$. The two remaining messages that I_i needs to decode are $W^{(n-1)}_{R_{i+1}\to\overline{i+1},L_{i+1}}$ and $W^{(n-1)}_{L_{i-1}\to\overline{i-1},R_{i-1}}$, and we next describe the process to do this. By our inductive hypothesis at $t = n-2$ node I_i knows the message vectors

$$F_i^{(n-2)} = \left[W^{(n-2)}_{R_i\to\bar{i},L_i}, W^{(n-2)}_{R_i\to i,\bar{L}_i}, W^{(n-2)}_{R_i\to i,L_i}\right],$$
$$G_i^{(n-2)} = \left[W^{(n-2)}_{L_i\to\bar{i},R_i}, W^{(n-2)}_{L_i\to i,\bar{R}_i}, W^{(n-2)}_{L_i\to i,R_i}\right].$$

Observe that node I_i knows message vector

$$\left[W^{(n-2)}_{i\to L_i,R_i}, W^{(n-2)}_{i\to \bar{L}_i,R_i}, W^{(n-2)}_{i\to L_i,\bar{R}_i}\right]$$

at time instant n. Therefore I_i knows the vectors

$$\begin{aligned}&\left[W^{(n-2)}_{R_i\to \bar{i},L_i}, W^{(n-2)}_{R_i\to i,L_i}, W^{(n-2)}_{i\to L_i,R_i}, W^{(n-2)}_{i\to L_i,\bar{R}_i}\right],\\&\left[W^{(n-2)}_{L_i\to \bar{i},R_i}, W^{(n-2)}_{L_i\to i,R_i}, W^{(n-2)}_{i\to L_i,R_i}, W^{(n-2)}_{i\to \bar{L}_i,R_i}\right]\end{aligned}$$

at time instant n. Set t to $n-1$ and i to $i-1$ in (3.11), and set t to $n-1$ and i to $i+1$ in (3.12). Then (3.11) and (3.12) imply that the preceding vectors are permutations of the following vectors and thus at time instant n they are available to node I_i:

$$\begin{aligned}F^{(n-1)}_{i-1} &= \left[W^{(n-1)}_{R_{i-1}\to \overline{i-1},L_{i-1}}, W^{(n-1)}_{R_{i-1}\to i-1,\bar{L}_{i-1}}, W^{(n-1)}_{R_{i-1}\to i-1,L_{i-1}}\right],\\G^{(n-1)}_{i+1} &= \left[W^{(n-1)}_{L_{i+1}\to \overline{i+1},R_{i+1}}, W^{(n-1)}_{L_{i+1}\to i+1,\bar{R}_{i+1}}, W^{(n-1)}_{L_{i+1}\to i+1,R_{i+1}}\right].\end{aligned}$$

From $F^{(n-1)}_{i-1}$ and $G^{(n-1)}_{i+1}$, I_i obtains $W^{(n-1)}_{R_{i-1}\to \overline{i-1},L_{i-1}}$ and $W^{(n-1)}_{L_{i+1}\to \overline{i+1},R_{i+1}}$. Since by (3.13) and (3.14) I_i can extract

$$W^{(n-1)}_{R_{i+1}\to \overline{i+1},L_{i+1}} \oplus W^{(n-1)}_{L_{i+1}\to \overline{i+1},R_{i+1}}$$

and

$$W^{(n-1)}_{L_{i-1}\to \overline{i-1},R_{i-1}} \oplus W^{(n-1)}_{R_{i-1}\to \overline{i-1},L_{i-1}}$$

at time instant n from $X^{(n-1)}_{i+1,i}$ and $X^{(n-1)}_{i-1,i}$, respectively, it can decode $W^{(n-1)}_{R_{i+1}\to \overline{i+1},L_{i+1}}$ and $W^{(n-1)}_{L_{i-1}\to \overline{i-1},R_{i-1}}$. Thus I_i can decode

$$\begin{aligned}&\left[W^{(n-1)}_{R_{i+1}\to \overline{i+1},L_{i+1}}, W^{(n-1)}_{R_{i+1}\to i+1,L_{i+1}}, W^{(n-1)}_{i+1\to L_{i+1},R_{i+1}}, W^{(n-1)}_{i+1\to L_{i+1},\bar{R}_{i+1}}\right],\\&\left[W^{(n-1)}_{L_{i-1}\to \overline{i-1},R_{i-1}}, W^{(n-1)}_{L_{i-1}\to i-1,R_{i-1}}, W^{(n-1)}_{i-1\to L_{i-1},R_{i-1}}, W^{(n-1)}_{i-1\to \bar{L}_{i-1},R_{i-1}}\right]\end{aligned}$$

or equivalently

$$F_i^{(n)} = \left[W^{(n)}_{R_i\to\bar{i},L_i}, W^{(n)}_{R_i\to i,\bar{L}_i}, W^{(n)}_{R_i\to i,L_i}\right],$$
$$G_i^{(n)} = \left[W^{(n)}_{L_i\to\bar{i},R_i}, W^{(n)}_{L_i\to i,\bar{R}_i}, W^{(n)}_{L_i\to i,R_i}\right],$$

at time instant n, as desired. From (3.6) it is simple to check that $X_i^{(n)}$ is a function of $F_i^{(n)}$ and $G_i^{(n)}$ and the messages generated at I_i at time instant n. Therefore I_i may transmit $X_i^{(n)}$ at time instant n.

We next wish to show that $X^{(n)}_{i,i+1}$ and $X^{(n)}_{i,i-1}$ 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^n. This assumption is reasonable as a node in the original network models a single processor with small internal delays. By (3.6), (3.7) and (3.8) we see that I_i only needs to construct $W^{(n)}_{R_i\to\bar{i},L_i} \oplus W^{(n)}_{L_i\to\bar{i},R_i}$ and $W^{(n)}_{L_i\to\bar{i},R_i} \oplus W^{(n)}_{R_i\to\bar{i},L_i}$ for O_i to be able to transmit $X^{(n)}_{i,i-1}$ and $X^{(n)}_{i,i+1}$. Observe that $W^{(n)}_{R_i\to\bar{i},L_i} \oplus W^{(n)}_{L_i\to\bar{i},R_i}$ and $W^{(n)}_{L_i\to\bar{i},R_i} \oplus W^{(n)}_{R_i\to\bar{i},L_i}$ can be obtained from $W^{(n)}_{R_i\to\bar{i},L_i} \otimes W^{(n)}_{L_i\to\bar{i},R_i}$, which is a component of $X_i^{(n)}$.

□

We next must show that the receiver node, O_i, is able to successfully decode all messages destined for node i in the original network. Node O_i has access to vector X_i^n at time instant $t = n$. Observe that

$$\left[W^{(n)}_{R_i\to i,L_i}, W^{(n)}_{L_i\to i,R_i}, W^{(n)}_{R_i\to i,\bar{L}_i}, W^{(n)}_{L_i\to i,\bar{R}_i}\right]$$

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.

b. Optimality of the network coding scheme

We demonstrate the optimality of our scheme by proving that any non–negative rate tuple $(R_s : s \in S)$ that satisfies the bounds in Theorem 5 is achievable in the network by implementing our network coding scheme; i.e., any rate in the achievable region results in certain $X_i^{(t)}, X_{i,i-1}^{(t)}$ and $X_{i,i+1}^{(t)}$ which have the properties $H\left(X_i^{(t)}\right) \leq C_i, H\left(X_{i,i-1}^{(t)}\right) \leq C_{i,i-1}$ and $H\left(X_{i,i+1}^{(t)}\right) \leq C_{i,i+1}$ and thus can be supported on this network.

As each of the vectors $X_i^{(t)}, X_{i,i-1}^{(t)}$ and $X_{i,i+1}^{(t)}$, 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 (3.6) we have:

$$\begin{aligned} H\left(X_i^{(t)}\right) = &H\left(W_{R_i \to \bar{i}, L_i}^{(t)} \otimes W_{L_i \to \bar{i}, R_i}^{(t)}\right) + H\left(W_{R_i \to i, L_i}^{(t)}\right) + H\left(W_{L_i \to i, R_i}^{(t)}\right) \\ &+ H\left(W_{R_i \to i, \bar{L}_i}^{(t)}\right) + H\left(W_{L_i \to i, \bar{R}_i}^{(t)}\right) + H\left(W_{i \to L_i, R_i}^{(t)}\right) \\ &+ H\left(W_{i \to \bar{L}_i, R_i}^{(t)}\right) + H\left(W_{i \to L_i, \bar{R}_i}^{(t)}\right). \end{aligned} \tag{3.15}$$

It follows from our earlier definitions that

$$\begin{aligned} \sum_{\nu_s = I_i} R_s &= H\left(W_{i \to L_i, R_i}^{(t)}\right) + H\left(W_{i \to \bar{L}_i, R_i}^{(t)}\right) + H\left(W_{i \to L_i, \bar{R}_i}^{(t)}\right), \\ \sum_{O_i \in D_s} R_s &= H\left(W_{R_i \to i, L_i}^{(t)}\right) + H\left(W_{L_i \to i, R_i}^{(t)}\right) + H\left(W_{R_i \to i, \bar{L}_i}^{(t)}\right) + H\left(W_{L_i \to i, \bar{R}_i}^{(t)}\right), \\ \sum_{s \in U_1} R_s &= H\left(W_{L_i \to \bar{i}, R_i}^{(t)}\right), \\ \sum_{s \in U_2} R_s &= H\left(W_{R_i \to \bar{i}, L_i}^{(t)}\right). \end{aligned}$$

Since $H(\mathbf{a} \otimes \mathbf{b}) = \max\{H(\mathbf{a}), H(\mathbf{b})\}$, (3.15) gives

$$H\left(X_i^{(t)}\right) = \sum_{\nu_s = I_i} R_s + \sum_{O_i \in \mathcal{D}_s} R_s + \max\left\{\sum_{s \in U_1} R_s, \sum_{s \in U_2} R_s\right\} \leq C_i,$$

where U_1 and U_2 are defined in Theorem 5.

By (3.7) we have

$$\begin{aligned} H\left(X_{i,i-1}^{(t)}\right) = {} & H\left(W_{R_i \to \bar{i}, L_i}^{(t)} \oplus W_{L_i \to \bar{i}, R_i}^{(t)}\right) + H\left(W_{R_i \to i, L_i}^{(t)}\right) \\ & + H\left(W_{i \to L_i, \bar{R}_i}^{(t)}\right) + H\left(W_{i \to L_i, R_i}^{(t)}\right). \end{aligned} \tag{3.16}$$

Since $H\left(W_{R_i \to \bar{i}, L_i}^{(t)} \oplus W_{L_i \to \bar{i}, R_i}^{(t)}\right) = H\left(W_{R_i \to \bar{i}, L_i}^{(t)}\right)$ we obtain

$$H\left(X_{i,i-1}^{(t)}\right) = \sum_{s : \nu_s \in \{I_j : j \geq i\}, \mathcal{D}_s \cap \{O_j : j < i\} \neq \emptyset,} R_s \leq C_{i,i-1}.$$

With a similar argument we obtain

$$H\left(X_{i,i+1}^{(t)}\right) = \sum_{s : \nu_s \in \{I_j : j \leq i\}, \mathcal{D}_s \cap \{O_j : j > i\} \neq \emptyset,} R_s \leq C_{i,i+1}.$$

Remark 10 *While our network coding scheme is based on integer rates, rational values of rates that satisfy our network coding bounds can also be supported by our network coding scheme if we run it over multiple rounds. Let $(R_s : s \in S)$ be a rational rate tuple in the capacity region of a line network G, where every non–zero R_s can be written as $\frac{p_s}{q_s}$ with p_s and q_s relatively prime non–negative integers. Let q be the least common multiple of $\{q_s : s \in S\}$. Notice that the integer rate tuple $(qR_s : s \in S)$ is in the capacity region of the line network G' which is made from G by increasing the capacity of each edge e from C_e to qC_e. The transmission scenario of network G' can be achived in network G by using each edge for q consecutive times. This implies that the rate $(\frac{1}{q} qR_s : s \in S) = (R_s : s \in S)$ is achievable in network G. This consideration is also valid for star networks as is discussed in the next section.*

C. Network Coding in Star Networks

For unicast sessions over the node constrained star networks depicted by Figure 3, we introduce the following notation.

- $X_i^{(t)}$ is the binary vector transmitted from I_i to O_i at time instant t for $i \in \{0, \cdots, M\}$.
- $X_{i,0}^{(t)}$ and $X_{0,i}^{(t)}$ respectively denote the binary vectors at time t from O_i to I_0 and from O_0 to I_i.
- $W_{i\to j}^{(t)}$ is the message from node i to node j at time t for $i, j \in \{0, \cdots, M\}$.
- $C_i, C_{i,0}$, and $C_{0,i}$ respectively denote the capacities of the edges $(I_i, O_i), (O_i, I_0)$, and (O_0, I_i).
- $R_{i\to j}$ is an integer that represents the rate of transmission from node i to node j.
- T is the final time instant of transmission.
- $Y^t = \left[Y^{(1)}, \cdots, Y^{(t)}\right]$.

Therefore, by the definitions it follows that $H\left(X_i^{(t)}\right) \le C_i$, $H\left(X_{i,0}^{(t)}\right) \le C_{i,0}$, $H\left(X_{0,i}^{(t)}\right) \le C_{0,i}$, and $H\left(W_{i\to j}^{(t)}\right) = R_{i\to j}$. We follow the convention that $Y = \left[Y^{(1)}, \cdots, Y^{(T)}\right]$ and $W = \{W_{i\to j} : i, j \in \{0, \cdots, M\}\}$.

A permutation $\Pi : \{1, \cdots, M\} \to \{1, \cdots, M\}$ is any one–to–one function from the set $\{1, \cdots, M\}$ to itself. We have the following upper bounds on the achievable rates of a star network:

Theorem 11 *The set of achievable unicast rate tuples $(R_{i\to j} : i, j \in \{1, \cdots, M\})$ in the node-constrained star network of Figure 3 satisfies the following bounds:*
For any permutation $\Pi(\cdot)$ of the set $\{1, \cdots, M\}$

$$\sum_{i,j\in\{0,\cdots,M\}} R_{i\to j} - \sum_{i=1}^{M}\sum_{j=i+1}^{M} R_{\Pi(i)\to\Pi(j)} \le C_0, \tag{3.17}$$

and for every $i \in \{1, \cdots, M\}$

$$\sum_{j \neq i} \left(R_{i \to j} + R_{j \to i}\right) \leq C_i, \tag{3.18}$$

$$\sum_{j \neq i} R_{i \to j} \leq C_{i,0}, \tag{3.19}$$

$$\sum_{j \neq i} R_{j \to i} \leq C_{0,i}. \tag{3.20}$$

In the next subsection we prove Theorem 11. In Subsection 2 we will discuss a network coding scheme and we will describe classes of problems for which it is optimal in Subsection b.

1. Upper bounds on the achievable rates in star networks

First we prove the validity of (3.17). Define the following sets of messages:

$$W_{\Pi(i)} = \left\{W_{\Pi(i) \to \Pi(i+1)}, W_{\Pi(i) \to \Pi(i+2)}, \cdots, W_{\Pi(i) \to \Pi(M)}\right\},$$
$$W'_{\Pi(i)} = \left\{W_{\Pi(i) \to \Pi(i-1)}, W_{\Pi(i) \to \Pi(i-2)}, \cdots, W_{\Pi(i) \to \Pi(1)}\right\}.$$

We split set W into $4M$ disjoint sets of messages according to permutation Π as follows:

$$W = \bigcup_{i=1}^{M} W_{\Pi(i)} \bigcup_{i=1}^{M} W'_{\Pi(i)} \bigcup_{i=1}^{M} W_{i \to 0} \bigcup_{i=1}^{M} W_{0 \to i}.$$

Recall that $X_0 = \left[X_0^{(1)}, \cdots, X_0^{(T)}\right]$. Since $X_0^{(t)}$ for any time instant t is a function of W, so is X_0. In the following expansion of the entropy term $H\left(X_0, W\right)$ we use the fact that $H\left(X_0 | W\right) = 0$, the independence of the different messages $W_{i \to j}$, and the chain rule for entropy.

$$
\begin{aligned}
H(W) &= H(X_0, W) \\
&= H(X_0) + H\left(\bigcup_{i=1}^{M} W_{i\to 0}\middle| X_0\right) \\
&+ \sum_{i=1}^{M} H\left(W_{\Pi(i)}\middle| X_0, \bigcup_{i=1}^{M} W_{i\to 0}, \bigcup_{j=1}^{i-1} W_{0\to\Pi(j)}, \bigcup_{j=1}^{i-1} W_{\Pi(j)}, \bigcup_{j=1}^{i-1} W'_{\Pi(j)}\right) \\
&+ \sum_{i=1}^{M} H\left(W'_{\Pi(i)}\middle| X_0, \bigcup_{i=1}^{M} W_{i\to 0}, \bigcup_{j=1}^{i-1} W_{0\to\Pi(j)}, \bigcup_{j=1}^{i} W_{\Pi(j)}, \bigcup_{j=1}^{i-1} W'_{\Pi(j)}\right) \\
&+ \sum_{i=1}^{M} H\left(W_{0\to\Pi(i)}\middle| X_0, \bigcup_{i=1}^{M} W_{i\to 0}, \bigcup_{j=1}^{i-1} W_{0\to\Pi(j)}, \bigcup_{j=1}^{i} W_{\Pi(j)}, \bigcup_{j=1}^{i} W'_{\Pi(j)}\right). \qquad (3.21)
\end{aligned}
$$

To simplify (3.21), we first observe that messages decoded at the hub's output node O_0 are a function of X_0, and so $H\left(\bigcup_{i=1}^{M} W_{i\to 0}\middle| X_0\right) = 0$.

Next consider any term in the second summation of (3.21). For each i, $W'_{\Pi(i)}$ is a union over $j < i$ of the messages $W_{\Pi(i)\to\Pi(j)}$. $W_{\Pi(i)\to\Pi(j)}$ is decoded at $O_{\Pi(j)}$ as a function of $X_{\Pi(j)}$. We observe that $X_{\Pi(j)}$ is a function of $X_0, W_{\Pi(j)}, W'_{\Pi(j)}$, and $W_{\Pi(j)\to 0}$. Since our entropy term is conditional on all such random variables for $j < i$, it is therefore equal to zero.

Next consider any term in the third summation of (3.21). Since $W_{0\to\Pi(i)}$ is decoded as a function of $X_0, W_{\Pi(i)}, W'_{\Pi(i)}$, and $W_{\Pi(i)\to 0}$ and since our entropy term is conditional on all these random variables, our entropy term is equal to zero. Finally, to bound the terms in the first summation of (3.21), observe that

$$
H\left(W_{\Pi(i)}\middle| X_0, \bigcup_{i=1}^{M} W_{i\to 0}, \bigcup_{j=1}^{i-1} W_{0\to\Pi(j)}, \bigcup_{j=1}^{i-1} W_{\Pi(j)}, \bigcup_{j=1}^{i-1} W'_{\Pi(j)}\right) \le H\left(W_{\Pi(i)}\right).
$$

As a consequence of these arguments, we obtain the following bound on $H(X_0)$:

$$
H(X_0) \ge H(W) - \sum_{i=1}^{M} H\left(W_{\Pi(i)}\right). \qquad (3.22)
$$

To establish (3.17), we replace the entropy terms in (3.22) with their corresponding rates and divide both sides of the resulting inequality by T.

Since (3.17) is valid for every choice of permutation Π, the most restrictive bound on the rates corresponds to that Π which minimizes $\sum_{i=1}^{M}\sum_{j=i+1}^{M} R_{\Pi(i)\to\Pi(j)}$. In the next part we will discuss a coding scheme which sometimes solves this optimization problem, and we will discuss cases in which our outer bound on the capacity region is tight.

Remark 12 *The bound (3.17) can alternatively be written as* $\sum_{i=2}^{M}\sum_{j=1}^{i-1} R_{\Pi(i)\to\Pi(j)} \leq C_0$.

Remark 13 *Notice that by Theorem 11,*

$$C_0 \geq \sum_{i=1}^{M}\sum_{j=1}^{M} R_{i\to j} - \min_{\Pi}\sum_{i=1}^{M}\sum_{j=i+1}^{M} R_{\Pi(i)\to\Pi(j)},$$

where C_0 is the capacity of the central node. Also for any set of rates $\{R_{i\to j}\}_{i,j}$, we have

$$\min_{\Pi}\sum_{i=1}^{M}\sum_{j=i+1}^{M} R_{\Pi(i)\to\Pi(j)} \leq 0.5\sum_{i=1}^{M}\sum_{j=1}^{M} R_{i\to j}.$$

Therefore $C_0 \geq 0.5\sum_{i=1}^{M}\sum_{j=1}^{M} R_{i\to j}$. Observe that for the routing scheme in which all messages are separately sent through the central link, the total capacity consumption of the central link is $\sum_{i=1}^{M}\sum_{j=1}^{M} R_{i\to j}$. Hence, the capacity used by the central node in any network coding scenario can never be less than half of the capacity consumption of the routing scheme.

Remark 14 *The bound (3.17) also follows by the PdE argument in [49] by choosing $\mathcal{E} = \{I_0, O_0\}$ and $\mathcal{S} = \bigcup_{i=2}^{M}\bigcup_{j=1}^{i-1}\{s : \nu_s = I_i, D_s = j\}$.*

We next discuss the outer bounds (3.18), (3.19), and (3.20) on the capacity region which involve the terms $C_i, C_{i,0}$, and $C_{0,i}$ for $i \geq 1$ respectively. We proceed by constructing another network from the star network depicted in Figure 3; the new network will have a

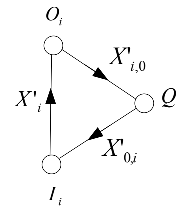

Fig. 13. Modified star network.

capacity region containing the achievable rate region of the star network in Figure 3. To create the new network, fix the index i and increase to infinity the capacity of all links except the ones used for the transmission of random variables $X_i, X_{i,0}$, and $X_{0,i}$. For this new network, we will focus on the rate bounds involving only terms corresponding to a source at I_i or a destination at O_i; i.e., we implicitly set other rate terms to 0. Since for $j, k \neq i$ the path from I_j to O_k has infinite capacity, we will merge all nodes except for I_i and O_i into a single supernode which we label Q; Figure 13 depicts the modified network. We point out that the network coding bounds for this network are discussed in [49] and the results are obtained using progressive d-separating edge set (*PdE*) bounds. We here use entropy bounds directly.

Observe that in the modified network we have a unicast session from O_i to Q with rate $\sum_{j\neq i} R_{i\to j}$, and we have a second unicast session from Q to I_i with rate $\sum_{j\neq i} R_{j\to i}$. Let us respectively denote the corresponding messages by $W_{i\to Q}$ and $W_{Q\to i}$. As illustrated in Figure 13, the counterpart vectors in the modified network for $X_i, X_{i,0}$, and $X_{0,i}$ are respectively called $X'_i, X'_{i,0}$, and $X'_{0,i}$. Since X'_i is a function of $W_{i\to Q}$ and $W_{Q\to i}$, the chain rule for entropy implies

$$\begin{aligned} H\left(W_{i\to Q}, W_{Q\to i}\right) &= H\left(X'_i, W_{i\to Q}, W_{Q\to i}\right) \\ &= H\left(X'_i\right) + H\left(W_{Q\to i}|X'_i\right) + H\left(W_{i\to Q}|X'_i, W_{Q\to i}\right). \end{aligned} \tag{3.23}$$

Since the messages destined for O_i are decoded using the vector X_i', it follows that

$$H\left(W_{Q\to i}|X_i'\right) = 0.$$

Furthermore, the messages destined for Q are decoded from $X_{i,0}'$ and $W_{Q\to i}$. Observe that the structure of the network implies that $X_{i,0}'$ is a function of X_i'. Therefore, we have $H\left(W_{i\to Q}|X_i', W_{Q\to i}\right) = 0$. These entropy terms and (3.23) imply

$$H\left(X_i'\right) = H\left(W_{i\to Q}, W_{Q\to i}\right) \tag{3.24}$$

By dividing both sides of (3.24) by T, we obtain

$$\sum_{j\neq i}\left(R_{i\to j} + R_{j\to i}\right) \leq C_i. \tag{3.25}$$

Note that (3.25) clearly holds for the original network as well.

We next apply a cut-set bound argument from [15, Ch. 14] to the triangular network in Figure 13 in order to obtain bounds on achievable rates involving capacities $C_{i,0}$ and $C_{0,i}$. First, consider the cut set $S = \{I_i, O_i\}$ and apply the cut-set arguments of [15, Ch. 14] to show that

$$\sum_{j\neq i} R_{i\to j} \leq C_{i,0}. \tag{3.26}$$

Next consider $S^c = \{I_i, O_i\}$ and apply the cut-set arguments of [15, Ch. 14] to show that

$$\sum_{j\neq i} R_{j\to i} \leq C_{0,i}. \tag{3.27}$$

2. Network coding scheme for star networks

a. Network coding scheme

We describe the encoding and decoding processes used at every node. We assume throughout that there is a unit delay associated with the channel from an output node to the input node of another processor. With a few changes in the indices of different messages our analysis is valid for the case of processors with non–zero delays.

Encoding and decoding at I_0: At time t, I_0 has access to $X_{i,0}^{t-1}$ for $i \in \{1, \cdots, M\}$. Therefore its output $X_0^{(t)}$ is a function of this information as well as its messages $W_{0\to j}^{t}, j \neq 0$, to other nodes. We only use the most recent $X_{i,0}^{(t-1)}$, which is a function of $W_{i\to j}^{(t-1)}, j \neq i$. We claim that all messages $W_{i\to j}^{(t-1)}$ are decodable from $X_{i,0}^{(t-1)}$; this assertion will become clear when we describe the encoding rules for O_i. Node I_0 transmits

$$\left[W_{1\to 0}^{(t-1)}, \cdots, W_{M\to 0}^{(t-1)}, W_{0\to 1}^{(t)}, \cdots, W_{0\to M}^{(t)}\right]$$

within $X_0^{(t)}$ to O_0. It also performs a series of operations on the messages $W_{i\to j}^{(t-1)}, i, j \neq 0$.

1. I_0 forms a directed multigraph called a demand graph, which is the multigraph $D(V, E)$ with $V = \{1, \cdots, M\}$ and with $R_{i\to j}$ parallel edges from node i to j if the rate of $W_{i\to j}^{(t-1)}$ is $R_{i\to j}$ bits/use of the network for $i, j \neq 0$. Notice that there is a one to one correspondence between the edges of D and the bits of $W_{i\to j}^{(t-1)}, i, j \neq 0$, in binary form.

2. I_0 extracts a set of edge disjoint cycles $\{\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_P\}$ from D. A cycle is a directed path of edges which starts and ends at the same node and in between visits other nodes, each one at most once.

3. Let $\oplus$ denote the XOR operation and $|\mathcal{C}_k|$ be the number of edges in cycle $\mathcal{C}_k$. For $k \in \{1, \cdots, P\}$, I_0 encodes the bits corresponding to the edge of cycle $\mathcal{C}_k$ into the following binary vector and sends this to O_0:

$$\left[b_1^k \oplus b_2^k, b_1^k \oplus b_3^k, \cdots, b_1^k \oplus b_{|\mathcal{C}_k|}^k\right] \tag{3.28}$$

   Here the ordering of bits $b_1^k, \cdots, b_{|\mathcal{C}_k|}^k$ corresponds to their ordering along cycle $\mathcal{C}_k$ starting at an arbitrary edge.

4. The remaining bits on D which correspond to the edges in $E \setminus \bigcup_{k=1}^{P} \mathcal{C}_k$ are sent to O_0 without encoding.

The length of $X_0^{(t)}$ is therefore $\left[\sum_{i,j} R_{i\to j}\right] - P$ bits.

Encoding and decoding at O_0: By receiving $X_0^{(t)}$, O_0 has to decode the messages

$$\left[W_{1\to 0}^{(t-1)}, \cdots, W_{M\to 0}^{(t-1)}\right]$$

which are destined for O_0. This can be done directly, since $W_{i\to 0}^{(t-1)}$ is a vector included in $X_0^{(t)}$. Then O_0 encodes the vector $X_{0,i}^{(t)}$ as a function of $X_0^{(t)}$ for $i \in \{1, \cdots, M\}$. The encoding of $X_{0,i}^{(t)}$ is done in the following steps:

1. The vector $W_{0\to i}^{(t)}$ is part of $X_{0,i}^{(t)}$ which is directly available from $X_0^{(t)}$.

2. For each cycle $\mathcal{C}_k, k \in \{1, \cdots, P\}$, O_0 determines if node i is a vertex along the cycle. If so, then it lets b_j^k and b_{j+1}^k denote the bits entering and exiting node i along cycle $\mathcal{C}_k$ (using the convention $b_{|\mathcal{C}_k+1|}^k = b_1^k$) and it forms and transmits the bit $b_j^k \oplus b_{j+1}^k = (b_1^k \oplus b_j^k) \oplus (b_1^k \oplus b_{j+1}^k)$.

3. The bits of the messages with destination at O_i that belong to $E \setminus \bigcup_{k=1}^{P} \mathcal{C}_k$ are sent without encoding as part of $X_{0,i}^{(t)}$.

Encoding and decoding at I_i: By receiving $X_{0,i}^{(t)}$, I_i first decodes all messages $W_{j\to i}^{(t-1)}, j \neq 0, W_{0\to i}^{(t)}$ that have destination O_i:

1. The vector $W_{0\to i}^{(t)}$ is directly decodable as a part of $X_{0,i}^{(t)}$.

2. If a bit of a message $W_{j\to i}^{(t-1)}, j \neq 0$, is part of a cycle $\mathcal{C}_k$ and is represented by b_l^k, then the bit $b_l^k \oplus b_{l+1}^k$ is part of $X_{0,i}^{(t)}$, and since b_{l+1}^k is part of a message with source I_i then $X_{0,i}^{(t)}$ extracts b_l^k by using $(b_l^k \oplus b_{l+1}^k) \oplus b_{l+1}^k$.

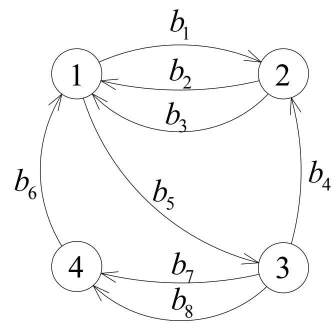

Fig. 14. Demand graph of example 15.

3. If a bit of a message $W_{j\to i}^{(t-1)}, j \neq 0$, is part of $E \setminus \bigcup_{k=1}^{P} \mathcal{C}_k$, then it is directly available from $X_{0,i}^{(t)}$.

After the decoding process, I_i chooses the following vector as $X_i^{(t+1)}$ and sends it to node O_i:

$$\left[W_{1\to i}^{(t-1)}, \cdots, W_{i-1\to i}^{(t-1)}, W_{i+1\to i}^{(t-1)}, \cdots, W_{M\to i}^{(t-1)}, W_{0\to i}^{(t)}, W_{i\to 0}^{(t+1)}, \cdots, W_{i\to M}^{(t+1)}\right].$$

Encoding and decoding at O_i: Node O_i directly decodes the messages

$$\left[W_{1\to i}^{(t-1)}, \cdots, W_{i-1\to i}^{(t-1)}, W_{i+1\to i}^{(t-1)}, \cdots, W_{M\to i}^{(t-1)}, W_{0\to i}^{(t)}\right]$$

with destination O_i and encodes $X_{i,0}^{(t+1)}$ as the vector $\left[W_{i\to 0}^{(t+1)}, \cdots, W_{i\to M}^{(t+1)}\right]$. This vector also includes all of the information that is needed for the encoding process at I_0. We conclude our description of the encoding and decoding processes in the node-constrained star network with the following simple example:

Example 15 *Assume that the following rates are the demands in a star network with $M = 4$ branches: $R_{1\to 2} = 1, R_{2\to 1} = 2, R_{1\to 3} = 1, R_{3\to 4} = 2, R_{4\to 1} = 1$, and $R_{3\to 2} = 1$. Figure 14 shows the demand graph corresponding to the rates. Here, $b_k, k \in \{1, \cdots, 8\}$, denotes the bits to be transmitted. We describe here part of the operations of nodes I_0,*

$O_0, I_1,$ and O_1. Node I_0 after receiving the demand graph information from $O_1, \cdots, O_4$, uses the cycles $\mathcal{C}_1 = \{b_1, b_2\}$ and $\mathcal{C}_2 = \{b_6, b_5, b_7\}$ for the coding scheme, and forms $X_0^{(t)} = [b_3, b_4, b_8, b_1 \oplus b_2, b_7 \oplus b_5, b_7 \oplus b_6]$, and sends it to O_0. Now we concentrate on node 1 and its corresponding vectors. Node 1 is the source of b_1 and b_5, and is the destination of $b_2, b_3,$ and b_6. After receiving $X_0^{(t)}$, O_0 forms $X_{0,1}^{(t)} = [b_3, b_1 \oplus b_2, b_5 \oplus b_6]$ and sends it to I_1. Since I_1 knows b_1 and b_5, it forms $X_1^{(t+1)} = [b_3, b_2, b_6]$ and sends it to O_1 which is the destination of these bits. Notice that in order to transmit $X_0^{(t)}$, C_0 must be at least 6 bits/use of the network. To show that this requirement on C_0 is tight for all encoding schemes, we compute the tightest bound of the form (3.17)*. In this case the permutation $\Pi(1) = 1, \Pi(2) = 2, \Pi(3) = 4,$ and $\Pi(4) = 3$ results in the tightest bound, which is $C_0 \geq 6$. Since our scheme always achieves the upper bounds on the capacity of all edges except the hub edge, it is optimal for this example. In the next section we will prove that our coding scheme is optimal for a broad class of demand graphs.*

b. Optimality of the network coding scheme

We first establish that our network coding scheme is optimal for all edges except the one from I_0 to O_0; i.e., the vectors $X_i^{(t)}, X_{i,0}^{(t)}$, and $X_{0,i}^{(t)}$ corresponding to any rate tuple $(R_{i\to j} : i, j \in \{0, \cdots, M\})$ that satisfies the bounds (3.18), (3.19), and (3.20), satisfy the capacity constraints $H(X_i^{(t)}) \leq C_i, H(X_{i,0}^{(t)}) \leq C_{i,0}$, and $H(X_{0,i}^{(t)}) \leq C_{0,i}$ for any $i \in \{1, \cdots, M\}$. Recall that

$$X_i^{(t+1)} = \left[W_{1\to i}^{(t-1)}, \cdots, W_{i-1\to i}^{(t-1)}, W_{i+1\to i}^{(t-1)}, \cdots, W_{M\to i}^{(t-1)}, W_{0\to i}^{(t)}, W_{i\to 0}^{(t+1)}, \cdots, W_{i\to M}^{(t+1)}\right].$$

The total length of $X_i^{(t+1)}$ is $H(X_i^{(t+1)}) = \sum_{j\neq i} (R_{i\to j} + R_{j\to i}) \leq C_i$. Next consider

$$X_{i,0}^{(t+1)} = \left[W_{i\to 0}^{(t+1)}, \cdots, W_{i\to M}^{(t+1)}\right].$$

The total length of $X_{i,0}^{(t+1)}$ is $H(X_{i,0}^{(t+1)}) = \sum_{j \neq i} R_{i \to j} \leq C_{i,0}$. Finally from the encoding process at node O_0 we know that there is one bit corresponding to every bit of messages $W_{j \to i}^{(t-1)}, j \neq 0$, and $W_{0 \to i}^{(t)}$, and therefore the total length of $X_{0,i}^{(t)}$ is $H(X_{0,i}^{(t)}) = \sum_{j \neq i} R_{j \to i} \leq C_{0,i}$.

Next we focus on the performance of the hub edge. We will establish that for some special cases of the demand graphs, if the rate tuple $(R_{i \to j} : i, j \in \{0, \cdots, M\})$ satisfies the bound (3.17), then the corresponding vector of $X_0^{(t)}$ obtained from our network coding scheme satisfies the condition $H(X_0^{(t)}) \leq C_0$. Along with the optimality results on the other edges, this shows that the upper bounds on the achievable rates in Theorem 11 exactly characterize the capacity region for these demand graphs.

Recall that we have chosen $X_0^{(t)}$ to be a function of the bits along P edge disjoint cycles of the demand graph and that $X_0^{(t)}$ has $\sum_{i,j} R_{i \to j} - P$ bits. Suppose that the rates are constrained by (3.17); i.e., $\sum_{i,j} R_{i \to j} - \sum_{i=1}^{M} \sum_{j=i+1}^{M} R_{\Pi(i) \to \Pi(j)} \leq C_0$. For our coding scheme the size of the smallest possible vector $X_0^{(t)}$ can be found from the maximum number P^* of edge disjoint cycles in the demand graph D. Let Π^* represent a permutation that minimizes $\sum_{i=1}^{M} \sum_{j=i+1}^{M} R_{\Pi(i) \to \Pi(j)}$. We will show a class of demand graphs for which $P^* = \sum_{i=1}^{M} \sum_{j=i+1}^{M} R_{\Pi^*(i) \to \Pi^*(j)}$. In these cases it follows $H(X_0^{(t)}) = \sum_{i,j} R_{i \to j} - P^* = \sum_{i,j} R_{i \to j} - \sum_{i=1}^{M} \sum_{j=i+1}^{M} R_{\Pi^*(i) \to \Pi^*(j)} \leq C_0$, and our coding and decoding schemes are optimal. The following theorem provides the maximum number of edge disjoint cycles in planar directed multigraphs:

Theorem 16 *[Lucchesi, 1976[57]] Let $D(V, E)$ be a planar directed multigraph. We define a set $S \subseteq E$ to be a feedback edge set if graph $(V, E \setminus S)$ is an acyclic, directed multigraph. The maximum number P^* of edge disjoint cycles of D is equal to the number of elements of the minimal feedback edge set S^*.*

For example, in Figure 14 the set $S = \{b_1, b_5\}$ is a minimal feedback edge set, where we

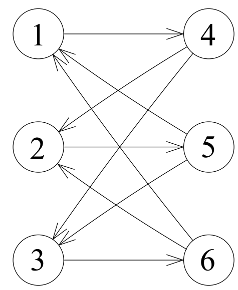

Fig. 15. Directed $\mathcal{K}_{3,3}$ graph.

have labeled the edges with the bits they carry. Thus, the maximal number of edge-disjoint cycles in Figure 14 is two.

Notice that for every feedback edge set S, the multigraph $(V, E \setminus S)$ is acyclic, and hence there is a permutation Π of nodes in $(V, E \setminus S)$ such that the edges from $\Pi(i)$ to $\Pi(j)$ corresponding to $R_{\Pi(i)\to\Pi(j)}$ are in $(V, E \setminus S)$ if and only if $i > j$; i.e., the edges corresponding to $R_{\Pi(i)\to\Pi(j)}$ are in S if and only if $i < j$. This implies that S^* corresponds to the permutation Π^* that minimizes $\sum_{i=1}^{M}\sum_{j=i+1}^{M} R_{\Pi(i)\to\Pi(j)}$. It follows from Theorem 16 that $|S^*| = P^*$, and hence our scheme is optimal for planar demand graphs.

Since [57] also establishes that the minimal feedback edge set can be computed by a polynomial time algorithm, our coding algorithm is a polynomial time algorithm. Furthermore, Theorem 16 has been extended to a broader class of directed graphs (see, e.g., [91]). The smallest directed graph that violates Theorem 16 is the nonplanar complete bipartite graph $\mathcal{K}_{3,3}$ depicted in Figure 15. We can potentially improve the rate of our coding scheme by coding over multiple copies of the demand graph; we consider a demand graph corresponding to the set of rates $nR_{i\to j}$ for each i and j. The capacity needed to transmit the new values of the rates will be normalized by n to give the capacity needed per use of the network. For the example of $\mathcal{K}_{3,3}$ graph observe that if we replace each edge in the graph with n parallel edges, the number of disjoint cycles is at most $1.5n$ because we have $3n$ edges directed from left hand side nodes to the right hand side nodes and each cycle uses at least two of these edges. On the other hand, for even values of n, $1.5n$ disjoint cycles exist. They are formed by nodes $1, 4, 2, 5, 1$ and $2, 5, 3, 6, 2$ and $1, 4, 3, 6, 1$ and use $.5n$ cycles from each form. However, it is easy to verify that when $n = 1$, the maximum number of disjoint cycles is 1. Hence, our scheme requires $C_0 = \frac{9n-1.5n}{n} = 7.5$ bits/use of the network. However, the best bound from (3.22) is that $C_0 \geq 7$ bits/use of the network, and this corresponds to the permutation function $\Pi^* = (1, 3, 5, 2, 4, 6)$. Yet for this case there is a linear code with $C_0 = 7$ bits/use of the network. This code will be described in Subsection 3.

3. Star networks and index coding

In this section we first review the problem of index coding with side information and the optimal linear coding scheme for it, and then we will discuss the relationship between the index coding problem and the network coding problem in star networks with node constraints. We restrict our attention to linear codes over the binary field and we assume that every edge is used only once, namely we set $T = 1$. However it is straightforward to generalize most of the results to the case of coding over multiple time instants with $T > 1$.

Definition 17 *(Index coding with side information [10]) Consider a broadcast network N with a source node s and destination nodes $d_1, d_2, \cdots, d_n$. The source is connected to the destinations via a broadcast link and all destinations have access to all of the broadcasted data. The source has an input of n bits $\mathbf{b} = [b_1, \cdots, b_n]^T$ and each d_i needs the bit b_i and knows a subset of the bits in $\mathbf{b}$. The subset of bits known by each destination is specified by a side information graph $I(V, E)$ that is a directed graph with n nodes such that $(i, j) \in E(I)$ if node d_i knows the bit b_j. The problem is to find the optimal communication scheme where s maps $\mathbf{b}$ into a binary codeword $\mathbf{c} = [c_1, \cdots, c_l]^T$ such that every node d_i can recover b_i from the vector $\mathbf{c}$ and the subset of bits in $\mathbf{b}$ that it knows. The goal is to minimize the*

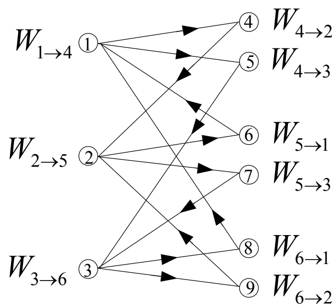

Fig. 16. Line graph $L_{\mathcal{K}_{3,3}}$.

length of the broadcast codeword.

In the following we establish a connection between the network coding problem in node constrained star networks and index coding with side information. For the directed graph $D(V, E)$ define its line graph *$L_D(V', E')$ as the directed graph with node set $V' = E$ and edges $(e_i, e_j) \in E'$ if the head of e_i and the tail of e_j coincide in graph D. For example, Figure 16 is the line graph corresponding to $\mathcal{K}_{3,3}$.*

Lemma 18 *The network coding problem in a star network with all capacities other than C_0 being infinite, and in which the rates $R_{0\to i}$ and $R_{i\to 0}$ are zero for $i \in \{1, \cdots, M\}$, is equivalent to the index coding problem in which the side information graph I is a line graph L_D of some directed graph D.*

Proof Consider a star network with only a hub node constraint and with the demand graph $D(V, E)$. Since the only constraint is C_0, we merge nodes O_i and I_i together and consider each node $i \neq 0$ as a single unit. Without loss of generality we can assume that each node i sends its uncoded messages to the node I_0 and the coding is performed at the node I_0. The length of the coded message will be $H(X_0) = C_0$. Also without loss of generality in our coding scheme, O_0 forwards the whole vector X_0 to all nodes $i \neq 0$. At node i the messages with destination at node i are decoded from the received vector X_0 and the messages from node i to other nodes. We decompose the decoder at node i into several decoders where each decodes one bit from the $\sum_j R_{j\to i}$ messages with destination at node i and each decoder has access to the vector X_0 and $\sum_j R_{i\to j}$ bits of messages with source at node i. This corresponds to the index coding problem where there are $\sum_{i,j} R_{i\to j}$ destination nodes and the broadcast vector, X_0, is received by all destinations. Each node in the side information graph $I(V', E')$ corresponds to a bit of a message or an edge of the graph $D(V, E)$. Let d_i denote the destination of bit $e_i \in E$. If we let $e_i = (u, v)$ where $u, v \in V$, then by the definition of a demand graph, d_i knows all bits that have a source at v. Therefore in graph I, e_i is connected to e_j if and only if e_j is of the form (v, w) for some node $w \in V$. This shows that I is the line graph of the graph D. Conversely if the side information graph I is of the form L_D, then the network coding problem in the star network with demand graph D corresponds to the index coding with side information graph I.

□

Next we discuss the optimal linear codes for the index coding problem.

Definition 19 *Consider a general index coding problem with a side information graph $I(V', E')$ for which $V' = \{1, \cdots, n\}$. An $n \times n$ matrix A in the binary field is said to* fit I *if $A_{i,i} = 1$ for $i \in \{1, \cdots, n\}$ and $A_{i,j} = 0$ if $(i, j) \notin E', i \neq j$.*

Given a matrix A that fits I and the message vector $\mathbf{b}$ we form vector $\mathbf{c}' = [c'_1, \cdots, c'_n]^T$ by the transformation $\mathbf{c}' = A \cdot \mathbf{b}$. We will argue that every destination node d_i will be able to decode the message bit b_i from the vector $\mathbf{c}'$ and its side information bits. Upon receiving $\mathbf{c}'$, node d_i first extracts bit c'_i from it. Since A fits graph I, c'_i can be written as follows:

$$c'_i = \sum_j A_{i,j} b_j = b_i + \sum_{j:d_i \text{ knows } b_j} A_{i,j} b_j.$$

Then d_i successfully recovers b_i from the equation $b_i = c'_i - \sum_{j:d_i \text{ knows } b_j} A_{i,j} b_j$ since c'_i, $\{b_j : d_i \text{ knows } b_j\}$, and A are known at d_i.

Next suppose that $\text{rank}(A) = r$ and without loss of generality let the first r row of A be linearly independent. This implies that vector $[c'_1, \cdots, c'_r]$ suffices for reconstructing the whole vector $\mathbf{c}'$. We choose our transmitted vector to be $\mathbf{c} = [c'_1, \cdots, c'_r]$. Then we have the following result:

Theorem 20 *[Optimal linear code [10]] The length of the optimal linear code for the index coding problem with side information graph I is minrk$(I) \triangleq \min\{$rank$(A) : A$ fits $I\}$.*

The preceding discussion also suggests that an optimal linear code can be constructed using a matrix A that fits I and satisfies $\text{rank}(A) = \text{minrk}(I)$. However, the calculation of $\text{minrk}(I)$ and finding a corresponding matrix can, in general, be NP-Hard [10].

Remark 21 *The authors of [10] consider the index coding problem for multiple unicast sessions. However, Theorem 20 can easily be extended to the multiple multicast case. It then will correspond to the optimal linear coding scheme of the multiple multicast problem in node-constrained star networks with only a hub constraint.*

Example 22 *By applying Theorem 20 we find the optimal network coding scheme for the star network with demand graph $\mathcal{K}_{3,3}$ and with only node constraint C_0. The corresponding index coding problem has the side information graph $L_{\mathcal{K}_{3,3}}$ as depicted in Figure 16. By* *definition, the matrices that fit $L_{\mathcal{K}_{3,3}}$ have the following form:*

$$A = \begin{pmatrix} 1 & 0 & 0 & y_1 & y_2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & y_3 & y_4 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & y_5 & y_6 \\ 0 & y_7 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & y_8 & 0 & 1 & 0 & 0 & 0 & 0 \\ y_9 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & y_{10} & 0 & 0 & 0 & 1 & 0 & 0 \\ y_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & y_{12} & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

where each $y_j, j \in \{1, \cdots, 12\}$ can either be 0 or 1. We find that minrk$(L_{\mathcal{K}_{3,3}}) = 7$, which is achieved by setting $y_j = 1, j \in \{1, \cdots, 12\}$. We see that this matches the lower bound of 7 on C_0 from the previous section and this shows that linear network coding is optimal for this problem.

CHAPTER IV

A MAX-FLOW/MIN-CUT ALGORITHM FOR LINEAR DETERMINISTIC RELAY NETWORKS

A. Introduction

Our focus in this chapter is on illustrating some connections between the flow of information in linear deterministic relay networks (LDRNs) and the *submodular flow* problem. Submodular flow was first introduced by Edmonds and Giles [20]. It generalizes the classical model of network flow and includes several combinatorial optimization problems as special cases. In the classical model of flow, each edge has a capacity that constrains the flow through that edge. The submodular flow model replaces these edge constraints with constraints on the total flow passing through each subset of nodes. These constraints satisfy a certain submodularity property. Frank [29] proved a max–flow/min–cut theorem for the submodular flow model and polynomial-time algorithms [31] exist for finding the maximum flow.

The new connection that we make in this chapter to submodular flow will provide us with powerful tools from submodular optimization theory to study the linear deterministic model of relay networks. We will use this framework to prove the optimality of a simple, flow–based coding scheme for LDRNs and to provide polynomial–time algorithms for the scheme.

In Section B we review the LDRN model and some basic results from [7]. The network model is a layered graph with the first layer consisting of the source node, the last layer consisting of the destination node, and each intermediate layer consisting of one or more relay nodes. The channel model from one layer to the next layer of the network is described by a transfer function which is a linear function described by matrix multiplication in some fixed finite field. We will begin our development in Section C by introducing a notion of *flow* from one layer to the next layer. Our flow corresponds to a non–singular submatrix of the transfer function between two consecutive layers, and it preserves the information sent from one layer to the next layer. In order to analyze our transmission scheme between two consecutive layers and to find the set of achievable flows from one layer to the next layer, we apply the Rado–Hall theorem [68, 93, Ch. 7] from *transversal theory* [38, 58, 67, 93, Ch. 7]. Transversal theory studies independent structures in matroids. In Section D, we extend our notion of flow from one layer to the next layer of a network to the flow over the whole network. This corresponds to a transmission scheme from the source to the destination of an LDRN. In this section we prove that our flow is a special case of submodular flow and demonstrate that it is capacity–achieving by using the max–flow/min–cut theorem for submodular flow.

In Section E, we discuss the algorithmic aspects and computational complexity of finding the capacity of an LDRN and an optimal transmission scheme for it based on our flow algorithm. Our algorithms are based on polynomial–time algorithms for obtaining independent sets for the Rado–Hall theorem [78, Ch. 41] and submodular minimization [77] as well as algorithms for maximum submodular flow [31, 36].

B. The Linear Deterministic Model of Relay Networks

In this section we describe the LDRN and the corresponding max–flow/min–cut result for it from [7]. An LDRN is a layered directed network $\mathcal{N}$ with set of nodes $V = \bigcup_{i=1}^{M} V_i$, where V_i denotes the set of nodes in layer i, and set of edges E. Let $V_i = \{v_i(1), \cdots, v_i(m_i)\}$, where m_i denotes the number of nodes in layer i. The first layer consists of a single node $s = v_1(1)$ called the source node, and the last layer consists of a single node $t = v_M(1)$ called the destination node. There is an "edge" from every node in V_i to every node in

V_{i+1} which corresponds to a matrix we will define shortly. Observe that the study of an arbitrary directed network can also be placed in this framework if one instead considers its time-expanded representation of the network [7].

During one use of the communication channel between layers i and $i+1$, $v_i(j)$ transmits vector $\mathbf{x}_i[j]$ to the nodes in layer $i+1$ and $v_{i+1}(k)$ receives a vector $\mathbf{y}_{i+1}[k]$ given by

$$\mathbf{y}_{i+1}[k] = \sum_{j=1}^{m_i} G_i[k,j]\mathbf{x}_i[j],$$

where $G_i[k,j]$ is a predetermined "transfer function" of the edge $(v_i(j), v_{i+1}(k)) \in E$. Note that we can set $G_i[k,j]$ to be the all-zero matrix if there is no connection from $v_i(j)$ to $v_{i+1}(k)$. When Avestimehr, Diggavi and Tse [7] approximate a Gaussian relay network with an LDRN they set $G_i[k,j]$ to the $(n_0 - n_i^{k,j})$th power of matrix $\mathbf{J}$, where

- $\mathbf{J}$ is a shift matrix of size $n_0 \times n_0$, i.e., the element in row i and column j of $\mathbf{J}$ is one if $i = j+1$ and is zero otherwise,
- $n_i^{k,j} = \left\lceil \frac{1}{2}\log \left| h_i^{k,j} \right|^2 \right\rceil$,
- $n_0 = \max_{i,j,k} n_i^{k,j}$,
- $h_i^{j,k}$ represents the fading channel from node $v_i(j)$ to node $v_{i+1}(k)$.

Here we consider the case where $G_i[k,j]$ has a general form. All vectors and matrices are over a fixed finite field $\mathbb{F}_q$. One can define

$$\mathbf{x}_i = \begin{bmatrix} \mathbf{x}_i[1] \\ \vdots \\ \mathbf{x}_i[m_i] \end{bmatrix}, \mathbf{y}_{i+1} = \begin{bmatrix} \mathbf{y}_{i+1}[1] \\ \vdots \\ \mathbf{y}_{i+1}[m_{i+1}] \end{bmatrix}$$

and the block matrix $G_i = [G_i[k,j]]\,, 1 \le k \le m_{i+1}, 1 \le j \le m_i$. Then the received vectors at layer $i+1$ are related to the transmitted vectors at layer i by the following relationship

$$\mathbf{y}_{i+1} = G_i \cdot \mathbf{x}_i.$$

In general, in one session, each node can successively transmit and/or receive n vectors . Let the superscript $1 \leq h \leq n$ denote the hth transmitted or received vector. Node s successively sends the vectors $\mathbf{x}_s^1, \cdots, \mathbf{x}_s^n$ to the nodes in V_2. Likewise, each node $v_i(j)$ receives all of the vectors $\mathbf{y}_i^1[j], \cdots, \mathbf{y}_i^n[j]$ and subsequently transmits the vectors $\mathbf{x}_i^1[j], \cdots, \mathbf{x}_i^n[j]$. The session ends when node t receives vectors $\mathbf{y}_t^1, \cdots, \mathbf{y}_t^n$. The received vectors at layer $i+1$ are related to the transmitted vectors at layer i as follows

$$\mathbf{y}_{i+1}^h = G_i \cdot \mathbf{x}_i^h.$$

Node $v_i(j)$ may in general choose the transmitted vector $\mathbf{x}_i^h[j]$ to be any function of its received vectors $\mathbf{y}_i^1[j], \cdots, \mathbf{y}_i^n[j]$.

The capacity or maximum rate of reliable information transfer between s and t in the limit when n is very large can be characterized in terms of the *cuts* of the network. A cut Ω is a subset of the nodes V, and Ω is said to *separate* s and t if $s \in \Omega$ and $t \in \overline{\Omega} \triangleq V \setminus \Omega$. For $J \subseteq \{1, \cdots, m_i\}$ and $K \subseteq \{1, \cdots, m_{i+1}\}$ define $G_i[K, J]$ as the submatrix of G_i which includes the blocks $G_i[k, j]$ with $j \in J$ and $k \in K$. Let $A = \{a : v_i(a) \in \Omega\}$ and $B = \left\{b : v_{i+1}(b) \in \overline{\Omega}\right\}$. In an abuse of notation we define

$$G_i[\Omega] \triangleq G_i[B, A].$$

For any cut Ω, let the *cut function* $C(\Omega)$ be given by

$$C(\Omega) = \sum_{i=1}^{M-1} \text{rank}(G_i[\Omega]),$$

where the rank function is computed in the field $\mathbb{F}_q$. Avestimehr, Diggavi, and Tse [7] proved

Theorem 23 *The capacity of the network $\mathcal{N}$ is given by $C = \min_{\Omega \text{ separates } s \text{ and } t} C(\Omega)$.*

The minimum of the cut functions considered in Theorem 23 is shown in [7] to be an upper bound on the network capacity C by means of an information-theoretic cut–set bound [15, 15.10.1]. The capacity–achieving scheme is over a session with n consecutive transmissions at every node. First, node s encodes its message $\omega \in \{1, \cdots, q^{nR}\}$ as a vector in $\mathbb{F}_q^{nR}$ denoted by $\mathbf{y}_s(\omega)$. Relay node $v_i(j)$ selects $\mathbf{x}_i^h[j]$ for every $1 \leq h \leq n$ to be a random linear function of its received vectors $\mathbf{y}_i^1[j], \cdots, \mathbf{y}_i^n[j]$.

It is shown in [7] that if $R < C$ and n is sufficiently large, then with probability approaching one node t receives nR linearly independent linear combinations of the message vector $\mathbf{y}_s(\omega)$ from which it will be able to decode message ω.

In contrast to the randomized scheme with large number of blocks proposed in [7], we offer a deterministic, capacity–achieving scheme in which each node has only one transmission, i.e., $n = 1$. Our scheme, which can be found in polynomial time, is a closer counterpart to the flow–based schemes in traditional directed networks than is the scheme of [7].

C. Transmission from One Layer to the Next Layer

In this section we consider a notion of flow from the nodes in V_i to the nodes in V_{i+1} for $1 \leq i \leq M-1$. Since in our algorithm each node has only one transmission, we hereafter drop the time superscript.

For a matrix G label the rows of G with elements from a set P and the columns of G with elements from a set Q. For $p \in P$ and $q \in Q$ let $G(p,q)$ denote the element in row p and column q. For $A \subseteq P$ and $B \subseteq Q$ let $G(A,B)$ denote the submatrix of G consisting of the rows in A and the columns in B. Next consider a partition of the row indices as $P = \bigcup_{k=1}^{m} P[k]$ and a partition of the column indices as $Q = \bigcup_{j=1}^{n} Q[j]$. Given

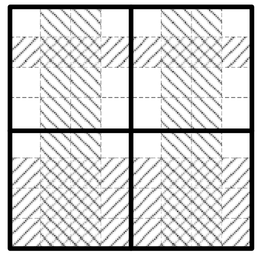

Fig. 17. An example of a matrix flow for vector $\mathbf{d} = (2, 2; 1, 3)$ in a matrix G with four blocks. Each small square is an entry of G and the full–rank submatrix is the intersection of dashed rows with dashed columns.

these partitions, we say that matrix G is a block matrix with $m \times n$ blocks and we use the notation $G[k, j]$ to denote the block $G(P[k], Q[j])$. Notice that the matrix G_i defined in the previous section is a block matrix with $m_{i+1} \times m_i$ blocks.

Definition 24 *Let $\mathbf{d} = (h_1, \cdots, h_n; g_1, \cdots, g_m)$ be a two–part vector of non–negative integers which satisfies $R \triangleq \sum_{j=1}^{n} h_j = \sum_{i=1}^{m} g_i$. We say that matrix G* supports *flow $\mathbf{d}$ if there exists a full rank $R \times R$ submatrix $\hat{G}$ of G such that*

$$\hat{G} = G\left(\bigcup_{k=1}^{m} \hat{P}[k], \bigcup_{j=1}^{n} \hat{Q}[j]\right)$$

with $|\hat{P}[k]| = g_k$, $|\hat{Q}[j]| = h_j$, $\hat{P}[k] \subseteq P[k]$, and $\hat{Q}[j] \subseteq Q[j]$ for all $1 \leq k \leq m$ and $1 \leq j \leq n$. Furthermore, we say that such a submatrix $\hat{G}$ is a solution *for flow $\mathbf{d}$ and that it has rate R.*

Figure 17 illustrates a matrix flow. For the physical interpretation of flow, suppose that rows of matrix G_i are indexed by elements of a set P and its columns are indexed by elements of a set Q. We consider a partition of $P = \bigcup_{k=1}^{m_{i+1}} P[k]$ and $Q = \bigcup_{j=1}^{m_i} Q[j]$ such that $G_i[k, j] = G_i(P[k], Q[j])$. Furthermore, we index the elements of each vector $\mathbf{x}_i[j]$ with elements of the set $Q[j]$ for $1 \leq j \leq m_i$ and the elements of $\mathbf{y}_{i+1}[k]$ with elements of the set $P[k]$ for $1 \leq k \leq m_{i+1}$ so that the indices of the elements of $\mathbf{x}_i$ and $\mathbf{y}_{i+1}$ are respectively consistent with the indices of the corresponding columns and rows of G_i. For some vector $\mathbf{e}$ with indices of elements in some set H, we let $\mathbf{e}(z)$ for $z \in H$ be the element with index z and $\mathbf{e}(\hat{H})$ for $\hat{H} \subseteq H$ be the subvector of $\mathbf{e}$ with elements with indices in $\hat{H}$.

Next, suppose that G_i supports a flow $\mathbf{d}_i$ with rate R which is given by

$$\hat{G}_i = G_i \left(\bigcup_{k=1}^{m_{i+1}} \hat{P}[k], \bigcup_{j=1}^{m_i} \hat{Q}[j] \right)$$

with $\hat{P}[k] \subseteq P[k]$ and $\hat{Q}[j] \subseteq Q[j]$ for all $1 \leq k \leq m_{i+1}$ and $1 \leq j \leq m_i$. Set every element of $\mathbf{x}_i[j]$ with index in $Q[j] \setminus \hat{Q}[j]$ to zero. Let

$$\hat{\mathbf{x}}_i = \begin{bmatrix} \mathbf{x}_i(\hat{Q}[1]) \\ \vdots \\ \mathbf{x}_i(\hat{Q}[m_i]) \end{bmatrix}, \hat{\mathbf{y}}_{i+1} = \begin{bmatrix} \mathbf{y}_{i+1}(\hat{P}[1]) \\ \vdots \\ \mathbf{y}_{i+1}(\hat{P}[m_{i+1}]) \end{bmatrix}.$$

Then our construction implies that

$$\hat{\mathbf{y}}_{i+1} = \hat{G}_i \cdot \hat{\mathbf{x}}_i.$$

Notice that $\mathbf{x}_i(\hat{Q}[j])$ has length h_j for $1 \leq j \leq m_i$, $\mathbf{y}_{i+1}(\hat{P}[k])$ has length g_k for $1 \leq k \leq m_{i+1}$, and the vectors $\hat{\mathbf{x}}_i$ and $\hat{\mathbf{y}}_{i+1}$ each have length R. Since $\hat{G}_i$ is a full–rank matrix, it is possible to uniquely decode the information $\hat{\mathbf{x}}_i$ from $\hat{\mathbf{y}}_{i+1}$. Hence this scheme enables R units of information to flow from the nodes in V_i to the nodes in V_{i+1} during a transmission.

Next we provide the necessary and sufficient conditions for matrix G to support a flow $\mathbf{d} = (h_1, \cdots, h_n; g_1, \cdots, g_m)$. Suppose that for $W \subseteq \{1, \cdots, m\}$ and $U \subseteq \{1, \cdots, n\}$, $G[W, U]$ denotes the submatrix $G(\bigcup_{k \in W} P[k], \bigcup_{j \in U} Q[j])$. Then

Theorem 25 *Matrix G supports flow $\mathbf{d}$ if and only if for every $W \subseteq \{1, \cdots, m\}$ and $U \subseteq \{1, \cdots, n\}$,*

$$\mathit{rank}(G[W,U]) \geq \sum_{k \in W} g_k + \sum_{j \in U} h_j - R. \tag{4.1}$$

This combinatorial property of matrices is, to our knowledge, new and may be of independent interest in the theory of matrices as well as in the study of independent structures. Theorem 25 holds for matrices with entries from an arbitrary field and is therefore more general than its application to this relay problem. The rest of this section is devoted to the proof of Theorem 25. A different proof appears in [85].

1. Matroids and Transversal theory

In this section we provide some basic results from matroid theory that we will use later in our proofs.

First we introduce matroids (see [93, Ch. 1]). Suppose that for a set H, 2^H denote the set of all its subsets.

Definition 26 *Given a set E and a function $r : 2^E \to \mathbb{N}$ we say that the pair (E, r) is a matroid if:*

1. $r(A) \leq |A|$ for all $A \subseteq E$.
2. If $A, B \subseteq E$ with $A \subseteq B$, then $r(A) \leq r(B)$.
3. For any $A, B \subseteq E$, we have $r(A \cup B) + r(A \cap B) \leq r(A) + r(B)$.

Consider a matrix G with set of indices of rows P and set of indices of columns Q. Let $E_0 \triangleq P \cup Q$ denote the disjoint union of the indices of rows and columns of the matrix G. We next define the function $r_0 : 2^{E_0} \to \mathbb{N}$. For every $A \subseteq P$ and every $B \subseteq Q$ we let

$$r_0(A \cup B) \triangleq \text{rank}(G(A, Q \backslash B)) + |B|. \tag{4.2}$$

Theorem 27 *(Kung [51])(E_0, r_0) as defined above is a matroid.*

The following is a corollary of Theorem 27 that will be particularly useful in our proofs:

Lemma 28 *Given matrix G with sets of indices of rows and columns respectively denoted by P and Q. For every $A_1, A_2 \subseteq P, B_1, B_2 \subseteq Q$ we have*

$$\begin{aligned} &\textit{rank}(G(A_1, Q\backslash B_1)) + \textit{rank}(G(A_2, Q\backslash B_2)) \geq \\ &\textit{rank}(G(A_1 \cap A_2, Q\backslash(B_1 \cap B_2))) + \textit{rank}(G(A_1 \cup A_2, Q\backslash(B_1 \cup B_2))). \end{aligned} \tag{4.3}$$

Proof By Property 3 of a matroid

$$r_0(A_1 \cup B_1) + r_0(A_2 \cup B_2) \geq r_0((A_1 \cap A_2) \cup (B_1 \cap B_2)) + r_0((A_1 \cup A_2) \cup (B_1 \cup B_2)).$$

By expanding function r_0 we have

$$\begin{aligned} &\text{rank}(G(A_1, Q\backslash B_1)) + \text{rank}(G(A_2, Q\backslash B_2)) + |B_1| + |B_2| \geq \\ &\text{rank}(G(A_1 \cap A_2, Q\backslash(B_1 \cap B_2))) + \text{rank}(G(A_1 \cup A_2, Q\backslash(B_1 \cup B_2))) \\ &+ |B_1 \cap B_2| + |B_1 \cup B_2|. \end{aligned}$$

Since $|B_1| + |B_2| = |B_1 \cap B_2| + |B_1 \cup B_2|$ the result follows.

□

Finally we will use the following transversal theorem for matroids. In every matroid (E, r), a set $A \subseteq E$ is an *independent* set if $r(A) = |A|$.

Theorem 29 *(Rado–Hall [68, 93, Ch. 7]) Let (E, r) be a matroid and $A_1, \cdots, A_n \subseteq E$. Given non–negative integers $\ell_1, \cdots, \ell_n$, there exists disjoint subsets $\hat{A}_1 \subseteq A_1, \cdots, \hat{A}_n \subseteq A_n$ with $|\hat{A}_i| = \ell_i$ and $\hat{A}_1 \cup \cdots \cup \hat{A}_n$ an independent set if and only if for every subset*

$I \subseteq \{1, \cdots, n\}$ *the following holds*

$$r\left(\bigcup_{i\in I} A_i\right) \geq \sum_{i\in I} \ell_i.$$

2. Proof of theorem 25

We next apply the Rado–Hall theorem to the matroid structure of matrices in Theorem 27 to prove Theorem 25.

Consider matrix G with rows and columns indices from P and Q respectively. Define r_0 as in (4.2). By Theorem 27, $(P \cup Q, r_0)$ is a matroid. Next consider the partition of $P \cup Q = P[1] \cup \cdots \cup P[m] \cup Q[1] \cup \cdots \cup Q[n]$ as in the setting of Definition 24 and Theorem 25. Assign to each $P[k]$ a non–negative integer g_k and to each $Q[j]$ a non–negative integer h_j. By the Rado–Hall theorem, there exist disjoint subsets $\hat{P}[1] \subseteq P[1], \cdots, \hat{P}[m] \subseteq P[m]$ and $\tilde{Q}[1] \subseteq Q[1], \cdots, \tilde{Q}[n] \subseteq Q[n]$ such that $|\hat{P}[k]| = g_k$, $|\tilde{Q}[j]| = |Q[j]| - h_j$, and $\bigcup_{i=1}^{m} \hat{P}[k] \bigcup_{j=1}^{n} \tilde{Q}[j]$ is an independent set in $(P \cup Q, r_0)$ if and only if for every $W \subseteq \{1, \cdots, m\}$ and $J \subseteq \{1, \cdots, n\}$ we have

$$r_0\left(\bigcup_{k\in W} P[k] \bigcup_{j\in J} Q[j]\right) \geq \sum_{k\in W} g_k + \sum_{j\in J}(|Q[j]| - h_j). \tag{4.4}$$

By using the definition of the rank function r_0 from (4.2), condition (4.4) is equivalent to

$$\text{rank}\left(G\left(\bigcup_{k\in W} P[k], Q\backslash\left(\bigcup_{j\in J} Q[j]\right)\right)\right) + |\bigcup_{j\in J} Q[j]| \geq \sum_{k\in W} g_k + \sum_{j\in J}(|Q[j]| - h_j)$$

or

$$\text{rank}\left(G\left(\bigcup_{k\in W} P[k], Q\backslash\left(\bigcup_{j\in J} Q[j]\right)\right)\right) \geq \sum_{k\in W} g_k - \sum_{j\in J} h_j.$$

Let $U = \{1, \cdots, n\} \backslash J$. Then the preceding condition is equivalent to having for every $W \subseteq \{1, \cdots, m\}$ and for every $U \subseteq \{1, \cdots, n\}$,

$$\text{rank}\left(G\left(\bigcup_{k\in W} P[k], \bigcup_{j\in U} Q[j]\right)\right) \geq \sum_{k\in W} g_k + \sum_{j\in U} h_j - \sum_{k=1}^{n} h_k = \sum_{k\in W} g_k + \sum_{j\in U} h_j - R.$$

Next let $\hat{Q}[j] = Q[j]\backslash\tilde{Q}[j]$. By definition, $\bigcup_{k=1}^{m} \hat{P}[k] \bigcup_{j=1}^{n} \tilde{Q}[j]$ is an independent set of the matroid if

$$r_0\left(\bigcup_{k=1}^{m} \hat{P}[k] \bigcup_{j=1}^{n} \tilde{Q}[j]\right) = |\bigcup_{k=1}^{m} \hat{P}[k] \bigcup_{j=1}^{n} \tilde{Q}[j]| = \sum_{k=1}^{m} g_k + \sum_{j=1}^{n}(|Q[j]| - h_j).$$

By the definition of r_0 from (4.2), the preceding condition is equivalent to

$$\text{rank}\left(G\left(\bigcup_{k=1}^{m} \hat{P}[k], \bigcup_{j=1}^{n} \hat{Q}[j]\right)\right) + \sum_{j=1}^{n}(|Q[j]| - h_j) = \sum_{k=1}^{m} g_k + \sum_{j=1}^{n}(|Q[j]| - h_j).$$

Therefore $\bigcup_{k=1}^{m} \hat{P}[k] \bigcup_{j=1}^{n} \tilde{Q}[j]$ is an independent set of the matroid if

$$\text{rank}\left(G\left(\bigcup_{k=1}^{m} \hat{P}[k], \bigcup_{j=1}^{n} \hat{Q}[j]\right)\right) = \sum_{k=1}^{m} g_k = R. \tag{4.5}$$

Observe that since $G\left(\bigcup_{k=1}^{m} \hat{P}[k], \bigcup_{j=1}^{n} \hat{Q}[j]\right)$ has R rows and R columns the condition above holds if and only if $G\left(\bigcup_{k=1}^{m} \hat{P}[k], \bigcup_{j=1}^{n} \hat{Q}[j]\right)$ has full rank. Our argument therefore implies that there exists a full–rank submatrix $G\left(\bigcup_{k=1}^{m} \hat{P}[k], \bigcup_{j=1}^{n} \hat{Q}[j]\right)$ with $\hat{P}[1] \subseteq P[1], \cdots, \hat{P}[m] \subseteq P[m]$ and $\hat{Q}[1] \subseteq Q[1], \cdots, \hat{Q}[n] \subseteq Q[n]$ such that $|\hat{P}[k]| = g_k$ and $|\hat{Q}[j]| = h_j$ if and only if for every $W \subseteq \{1, \cdots, m\}$ and for every $U \subseteq \{1, \cdots, n\}$,

$$\text{rank}\,(G[W, U]) \geq \sum_{k \in W} g_k + \sum_{j \in U} h_j - R.$$

This proves Theorem 25.

D. Flow in the Network

In this section we generalize the notion from Section C of the flow from a layer to the next layer to the flow in the entire network. This notion of flow underlies our transmission scheme for LDRNs. We will show the connection between our flow and submodular flow and will prove that our transmission scheme is capacity–achieving by applying known results on maximum submodular flow in networks.

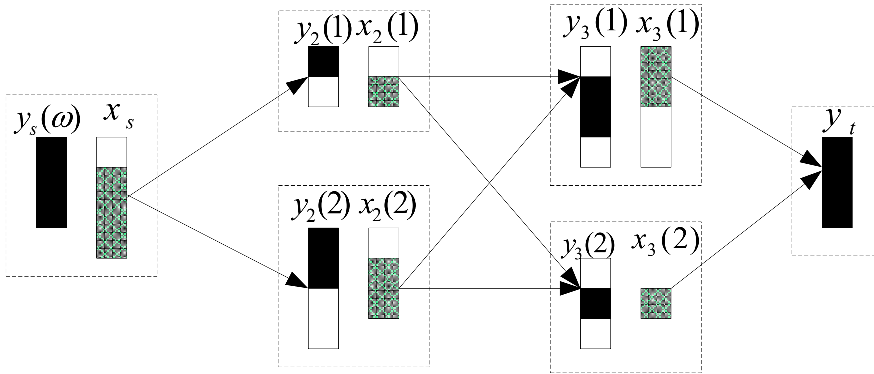

Fig. 18. A schematic of a rate–3 flow in a relay network with 6 nodes. Here we have $\ell_1(1) = 3, \ell_2(1) = 1, \ell_2(2) = 2, \ell_3(1) = 2, \ell_3(2) = 1$, and $\ell_4(1) = 3$.

Definition 30 *Suppose non–negative integers, $\ell_i(j), 1 \leq i \leq M, 1 \leq j \leq m_i$ are given such that they satisfy $\sum_{j=1}^{m_i} \ell_i(j) = R$ for every $1 \leq i \leq M$. We say that vector $\mathbf{d} = (\ell_i(j) : 1 \leq i \leq M, 1 \leq j \leq m_i)$ is a rate–R* flow *supported by network $\mathcal{N}$ if for every $1 \leq i \leq M-1$ the vector*

$$\mathbf{d}_i = (\ell_i(1), \cdots, \ell_i(m_i); \ell_{i+1}(1), \cdots, \ell_{i+1}(m_{i+1}))$$

is a rate–R flow supported by matrix G_i. Figure 18 illustrates a flow in an LDRN.

Suppose that for each $1 \leq i \leq M-1$ the row and column indices of matrix G_i are respectively from elements of the sets P_i and Q_i. Consider the partitions $P_i = \bigcup_{k=1}^{m_{i+1}} P_i[k]$ and $Q_i = \bigcup_{j=1}^{m_i} Q_i[j]$, and let $G_i[k,j] = G_i(P_i[k], Q_i[j])$ for every $1 \leq k \leq m_{i+1}$ and $1 \leq j \leq m_i$. We also use the elements of P_i and Q_i to respectively index the elements of the vectors $\mathbf{y}_{i+1}$ and $\mathbf{x}_i$ to be consistent with the indices for the rows and columns of G_i; i.e., $\mathbf{y}_{i+1}[k] = \mathbf{y}_{i+1}(P_i[k])$ for every $1 \leq k \leq m_{i+1}$ and $\mathbf{x}_i[j] = \mathbf{x}_i(Q_i[j])$ for every $1 \leq j \leq m_i$.

Suppose that network $\mathcal{N}$ supports flow $\mathbf{d} = (\ell_i(j) : 1 \le i \le M, 1 \le j \le m_i)$; i.e., every matrix G_i supports a rate–R flow

$$\mathbf{d}_i = (\ell_i(1), \cdots, \ell_i(m_i); \ell_{i+1}(1), \cdots, \ell_{i+1}(m_{i+1}))\,.$$

Let $\hat{G}_i = G_i\left(\bigcup_{k=1}^{m_{i+1}} \hat{P}_i[k], \bigcup_{j=1}^{m_i} \hat{Q}_i[j]\right)$, where $\hat{P}_i[k] \subseteq P_i[k]$ and $\hat{Q}_i[j] \subseteq Q_i[j]$ denote the solution of the flow for G_i. Then by the definition of a flow $|\hat{P}_i[k]| = \ell_{i+1}(k)$ for every $1 \le k \le m_{i+1}$ and $|\hat{Q}_i[j]| = \ell_i(j)$ for every $1 \le j \le m_i$. Define

$$\hat{\mathbf{x}}_i = \begin{bmatrix} \mathbf{x}_i(\hat{Q}_i[1]) \\ \vdots \\ \mathbf{x}_i(\hat{Q}_i[m_i]) \end{bmatrix}, \hat{\mathbf{y}}_{i+1} = \begin{bmatrix} \mathbf{y}_{i+1}(\hat{P}_i[1]) \\ \vdots \\ \mathbf{y}_{i+1}(\hat{P}_i[m_{i+1}]) \end{bmatrix}.$$

A rate–R flow supported by network $\mathcal{N}$ corresponds to the following simple rate–R transmission scheme:

Transmission scheme:

Given the length–R encoded vector $\mathbf{y}_s(\omega)$, node s generates vector $\mathbf{x}_1$ by setting $\hat{\mathbf{x}}_1 = \mathbf{y}_s(\omega)$ and the other entries of $\mathbf{x}_1$ to zero. The transformation at every relay node $v_i(j)$ is similar: after receiving vector $\mathbf{y}_i(j)$, node $v_i(j)$ encodes vector $\mathbf{x}_i(j)$ by setting

$$\mathbf{x}_i(\hat{Q}_i[j]) = \mathbf{y}_i(\hat{P}_{i-1}[j]) \tag{4.6}$$

and setting

$$\mathbf{x}_i(Q_i[j] \setminus \hat{Q}_i[j]) = \mathbf{0}.$$

Notice that the dimension of the subvectors on both sides of (4.6) are equal since both have length $\ell_i(j)$. Finally node t first extracts subvector $\hat{\mathbf{y}}_M$ from the received vector $\mathbf{y}_M$ and then decodes the encoded message $\mathbf{y}_s(\omega)$ as follows. Observe that for every i,

$$\hat{\mathbf{y}}_{i+1} = \hat{G}_i \cdot \hat{\mathbf{x}}_i.$$

Since $\hat{G}_i$ is a full–rank matrix, we have $\hat{\mathbf{x}}_i = \hat{G}_i^{-1} \cdot \hat{\mathbf{y}}_{i+1}$. Furthermore, from (4.6) we have $\hat{\mathbf{x}}_i = \hat{\mathbf{y}}_i$. These imply that $\mathbf{y}_s(\omega) = G_1^{-1} G_2^{-1} \cdots G_{M-1}^{-1} \hat{\mathbf{y}}_M$. Since the matrices $\hat{G}_i$ are nonsingular, the decoding operation is well defined.

In this section we prove the following result:

Theorem 31 *A network $\mathcal{N}$ with capacity C has a rate–R flow if and only if $R \leq C$.*

This result shows that our coding scheme can achieve the capacity of the network. A different proof from the one which follows appears in [85]. To analyze our flow, we begin with some basic definitions and results from the theory of submodular flow:

1. Submodular flow

Here we introduce submodular flow and the corresponding max–flow/min–cut theorem from [60, §9.3].

Let $\mathcal{G}(\mathcal{V}, \mathcal{A})$ be a directed graph with node set $\mathcal{V}$ and edge set $\mathcal{A}$. For every $a \in \mathcal{A}$, let $\partial^+ a$ be the tail of a and $\partial^- a$ be the head of a. Therefore we can write $a = (\partial^+ a, \partial^- a)$. For each node $v \in U$ define

$$\delta^+ v = \left\{a : a \in \mathcal{A}, \partial^+ a = v\right\},$$
$$\delta^- v = \left\{a : a \in \mathcal{A}, \partial^- a = v\right\}.$$

For $H \subseteq \mathcal{V}$ define

$$\Delta^+ H = \left\{a : a \in \mathcal{A}, \partial^+ a \in H, \partial^- a \in \mathcal{V} \setminus H\right\},$$
$$\Delta^- H = \left\{a : a \in \mathcal{A}, \partial^- a \in H, \partial^+ a \in \mathcal{V} \setminus H\right\}.$$

In words $\Delta^+ H$ is the set of edges that are leaving H and $\Delta^- H$ is the set of edges that are entering H.

A function $\rho : 2^{\mathcal{V}} \to \mathbb{R} \cup \{+\infty\}$ is called a *submodular set function* if for every $X, Y \subseteq \mathcal{V}$ it satisfies

$$\rho(X) + \rho(Y) \geq \rho(X \cap Y) + \rho(X \cup Y).$$

In a *submodular flow* problem we are given a graph $\mathcal{G}(\mathcal{V}, \mathcal{A})$, an upper capacity function $c_u : \mathcal{A} \to \mathbb{R} \cup \{+\infty\}$, a lower capacity function $c_l : \mathcal{A} \to \mathbb{R} \cup \{-\infty\}$, and a submodular set function $\rho : 2^{\mathcal{V}} \to \mathbb{R} \cup \{+\infty\}$, where $c_u(a) \geq c_l(a)$ for every $a \in A$ and $\rho(\emptyset) = \rho(\mathcal{V}) = 0$. A *feasible submodular flow* means a function $\phi : \mathcal{A} \to \mathbb{R}$ that satisfies

1. for every $a \in \mathcal{A}, c_l(a) \leq \phi(a) \leq c_u(a)$,

2. for every $H \subseteq \mathcal{V}$,
$$\sum_{a \in \Delta^+ H} \phi(a) - \sum_{a \in \Delta^- H} \phi(a) \leq \rho(H).$$

The submodular flow problem is *feasible* if it admits a feasible flow.

The *maximum submodular flow problem* is to find a feasible flow ϕ that maximizes $\phi(a_0)$ for a specific edge $a_0 \in \mathcal{A}$. Formally we have

Maximum submodular flow problem

$$\text{Maximize } \phi(a_0)$$

$$\text{subject to } c_l(a) \leq \phi(a) \leq c_u(a), \quad \text{for every } a \in \mathcal{A} \tag{4.7}$$

$$\sum_{a \in \Delta^+ H} \phi(a) - \sum_{a \in \Delta^- H} \phi(a) \leq \rho(H), \quad \text{for every } H \subseteq \mathcal{V} \tag{4.8}$$

$$\phi(a) \in \mathbb{R}, \quad \text{for every } a \in \mathcal{A}. \tag{4.9}$$

A max–flow/min–cut theorem holds for this problem:

Theorem 32 *(Max–flow/min–cut theorem for submodular flow [29]) For a feasible maximum submodular flow problem*

$$\max\{\phi(a_0) : \textit{(4.7), (4.8), (4.9)}\} =$$
$$\min\left[c_u(a_0), \min_{X\subseteq\mathcal{V}}\left\{\sum_{a\in\Delta^-X} c_u(a) - \sum_{a\in\Delta^+X\backslash\{a_0\}} c_l(a) + \rho(X) : a_0 \in \Delta^+X\right\}\right]. \quad (4.10)$$

If c_l, c_u and ρ are integer valued and (4.10) is finite, then there exists an integer–valued maximum flow $\phi : \mathcal{A} \to \mathbb{Z}$.

2. Proof of theorem 31

In this section we first construct an auxiliary graph $\mathcal{G}(\mathcal{V}, \mathcal{A})$ for the LDRN $\mathcal{N}$. Then we define a submodular flow problem on the graph $\mathcal{G}$ which is equivalent to finding a flow in our LDRN. The proof of Theorem 31 will follow from an application of Theorem 32 to this submodular flow problem on graph $\mathcal{G}$.

The auxiliary graph: Define $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ by

$$\mathcal{V} = \{\nu_i(j) : 2 \le i \le M-1, 1 \le j \le m_i\} \cup$$
$$\{\upsilon_i(j) : 2 \le i \le M-1, 1 \le j \le m_i\} \cup \{\upsilon_1(1), \nu_M(1)\}$$

and

$$\mathcal{A} = \{a_i(j) : 2 \le i \le M-1, 1 \le j \le m_i\} \cup \{a_0\}$$

where $a_i(j) = (\nu_i(j), \upsilon_i(j))$ and $a_0 = (\nu_M(1), \upsilon_1(1))$. (See Figure 19 for an example of graph $\mathcal{G}$.)

The submodular flow problem on graph $\mathcal{G}$: Fix a positive integer R_0. Let $c_u(a) = +\infty$ for $a \in \mathcal{A}\backslash\{a_0\}$, $c_u(a_0) = R_0$, and $c_l(a) = 0$ for every edge $a \in \mathcal{A}$. To define the function

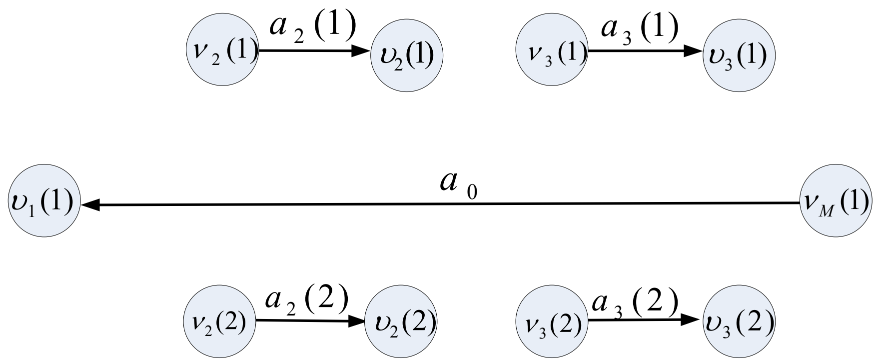

Fig. 19. Graph $\mathcal{G}$ corresponding to the LDRN in Figure 18.

$\rho : 2^{\mathcal{V}} \to \mathbb{R} \cup \{+\infty\}$ we need some definitions. For any $H \subseteq \mathcal{V}$ and $1 \le i \le M-1$ let

$$J_i(H) \triangleq \{j : \upsilon_i(j) \notin H\},$$

$$K_i(H) \triangleq \{k : \nu_{i+1}(k) \in H\}.$$

Consider the correspondence of the indices of the node set $\mathcal{V}$ and the node set V and define

$$\rho_i(H) \triangleq \text{rank}(G_i[K_i(H), J_i(H)])$$

and ρ by

$$\rho(H) \triangleq \sum_{i=1}^{M-1} \rho_i(H).$$

Lemma 33 *The function ρ defined above is submodular and satisfies $\rho(\emptyset) = \rho(\mathcal{V}) = 0$.*

Proof To establish the submodularity of ρ, we need to show that for every $H_1, H_2 \subseteq \mathcal{V}$,

$$\rho(H_1) + \rho(H_2) \ge \rho(H_1 \cap H_2) + \rho(H_1 \cup H_2).$$

Fix $1 \le i \le M-1$, and for $H \subseteq \mathcal{V}$ let

$$F_i(H) = \{j : \upsilon_i(j) \in H\}.$$

Notice that $F_i(H_1 \cap H_2) = F_i(H_1) \cap F_i(H_2)$, $K_i(H_1 \cap H_2) = K_i(H_1) \cap K_i(H_2)$, $F_i(H_1 \cup H_2) = F_i(H_1) \cup F_i(H_2)$, and $K_i(H_1 \cup H_2) = K_i(H_1) \cup K_i(H_2)$. Let $Q_i = \{1, \cdots, m_i\}$. Then

$$\rho_i(H) = \text{rank}(G_i[K_i(H), Q_i \backslash F_i(H)]),$$

and

$$\begin{aligned}
\rho_i(H_1) + \rho_i(H_2) &= \text{rank}(G_i[K_i(H_1), Q_i \backslash F_i(H_1)]) + \text{rank}(G_i[K_i(H_2), Q_i \backslash F_i(H_2)]) \\
&\overset{(a)}{\geq} \text{rank}(G_i[K_i(H_1) \cap K_i(H_2), Q_i \backslash (F_i(H_1) \cap F_i(H_2))]) \\
&\quad + \text{rank}(G_i[K_i(H_1) \cup K_i(H_2), Q_i \backslash (F_i(H_1) \cup F_i(H_2))]) \\
&= \rho_i(H_1 \cap H_2) + \rho_i(H_1 \cup H_2),
\end{aligned}$$

where (a) follows by Lemma 28. From the preceding inequality and the definition of ρ it follows that

$$\begin{aligned}
\rho(H_1) + \rho(H_2) &= \sum_{i=1}^{M-1} (\rho_i(H_1) + \rho_i(H_2)) \\
&\geq \sum_{i=1}^{M-1} (\rho_i(H_1 \cap H_2) + \rho_i(H_1 \cup H_2)) \\
&= \rho(H_1 \cap H_2) + \rho(H_1 \cup H_2),
\end{aligned}$$

which is the desired inequality.

Next observe that $K_i(\emptyset) = \emptyset$ and $J_i(\mathcal{V}) = \emptyset$ for every $1 \leq i \leq M-1$. Thus $\rho_i(\emptyset) = \rho_i(\mathcal{V}) = 0$ for every $1 \leq i \leq M-1$. Therefore $\rho(\emptyset) = \rho(\mathcal{V}) = 0$.

□

The functions c_l, c_u, and ρ defined above are the setting for a submodular flow problem. The submodular flow problem is feasible because the flow $\phi(a) = 0$ for every $a \in \mathcal{A}$ is always a feasible flow.

Let Φ be the set of all integer functions $\phi : \mathcal{A} \to \mathbb{Z}$ with $\phi(a_0) \leq R_0$. Also let $\mathbf{D}$ be the set of all integer vectors $\mathbf{d} = (\ell_i(j) : 1 \leq i \leq M, 1 \leq j \leq m_i)$ with $\ell_1(1) = \ell_M(1) \leq R_0$. Define the bijection $T : \Phi \to \mathbf{D}$ by $T(\phi) = (\ell_i(j) : 1 \leq i \leq M, 1 \leq j \leq m_i)$, where

$$\ell_i(j) = \phi\left(a_i(j)\right), \quad 2 \leq i \leq M-1, 1 \leq j \leq m_i$$
$$\ell_1(1) = \ell_M(1) = \phi(a_0).$$

Lemma 34 *The flow $\phi \in \Phi$ is a feasible submodular flow for graph $\mathcal{G}$ if and only if the vector $\mathbf{d} = T(\phi)$ is a feasible flow in the corresponding network $\mathcal{N}$.*

Notice that since for any feasible flow $\mathbf{d} = (\ell_i(j) : 1 \leq i \leq M, 1 \leq j \leq m_i)$ in network $\mathcal{N}$ with rate R we have $\ell_1(1) = \ell_M(1) = R$, the set $\mathbf{D}$ includes all feasible flow vectors for $\mathcal{N}$. Furthermore, the rate R for a feasible flow $\mathbf{d} = (\ell_i(j) : 1 \leq i \leq M, 1 \leq j \leq m_i)$ is the value of $\phi(a_0)$, where $\phi = T^{-1}(\mathbf{d})$. Next we prove Lemma 34.

Proof First suppose that $\phi \in \Phi$ is a feasible submodular flow. We begin by showing that $\mathbf{d} = T(\phi)$ is a feasible flow in $\mathcal{N}$. Let $\mathbf{d} = (\ell_i(j) : 1 \leq i \leq M, 1 \leq j \leq m_i)$. First notice that $\mathbf{d}$ is an integer vector with non–negative elements since $0 = c_l(a) \leq \phi(a)$. Next we show that for every $2 \leq i \leq M-1$,

$$\sum_{j=1}^{m_i} \ell_i(j) = \ell_1(1) = \ell_M(1).$$

Fix a value of $1 \leq i \leq M-1$ and let $H = \{\upsilon_i(1), \cdots, \upsilon_i(m_i)\} \cup \{\nu_{i+1}(1), \cdots, \nu_{i+1}(m_{i+1})\}$. Then $K_j(H) = \emptyset$ for $j \neq i$ and $J_i(H) = \emptyset$. This implies that for every $1 \leq j \leq M-1$, $\rho_j(H) = 0$, and hence $\rho(H) = 0$. Now by the second condition of a feasible submodular flow and the fact that $\rho(H) = 0$ we have

$$\sum_{a \in \Delta^+ H} \phi(a) \leq \sum_{a \in \Delta^- H} \phi(a).$$

For $i = 1$ the above condition is

$$\sum_{j=1}^{m_2} \phi\left(a_2(j)\right) \leq \phi(a_0). \tag{4.11}$$

For $2 \leq i \leq M-2$ the condition is

$$\sum_{j=1}^{m_{i+1}} \phi\left(a_{i+1}(j)\right) \leq \sum_{k=1}^{m_i} \phi\left(a_i(k)\right). \tag{4.12}$$

For $i = M-1$ the condition is

$$\phi(a_0) \leq \sum_{j=1}^{m_{M-1}} \phi\left(a_{M-1}(j)\right). \tag{4.13}$$

Inequalities (4.11), (4.12), (4.13) imply that

$$\sum_{j=1}^{m_i} \phi(a_i(j)) = \phi(a_0)$$

for $2 \leq i \leq M-2$, which proves our claim.

Let $R = \phi(a_0)$. Next we show that

$$\mathbf{d}_i = (\ell_i(1), \cdots, \ell_i(m_i); \ell_{i+1}(1), \cdots, \ell_{i+1}(m_{i+1}))$$

is a rate–R flow for matrix G_i for $1 \leq i \leq M-1$. Fix $1 \leq i \leq M-1$, $K \subseteq \{1, \cdots, m_{i+1}\}$, and $J \subseteq \{1, \cdots, m_i\}$. Let $H = \{\upsilon_i(j) : j \notin J\} \cup \{\nu_{i+1}(k) : k \in K\}$. Then for every $l \neq i$, we have $K_l(H) = \emptyset$ and hence $\rho_l(H) = 0$. Also, $J_i(H) = J$ and $K_i(H) = K$, and hence $\rho_i(H) = \text{rank}(G_i[K, J])$. Therefore $\rho(H) = \rho_i(H) = \text{rank}(G_i[K, J])$. The second condition of a feasible submodular flow implies that

$$\sum_{a \in \Delta^+ H} \phi(a) - \sum_{a \in \Delta^- H} \phi(a) \leq \text{rank}(G_i[K, J]).$$

We have $\Delta^+ H = \{a_{i+1}(k) : k \in K\}$ and $\Delta^- H = \{a_i(j) : j \notin J\}$. Thus $\sum_{a\in\Delta^+ H} \phi(a) = \sum_{k\in K} \ell_{i+1}(k)$ and $\sum_{a\in\Delta^- H} \phi(a) = R - \sum_{j\in J} \ell_i(j)$. Therefore

$$\sum_{j\in J} \ell_i(j) + \sum_{k\in K} \ell_{i+1}(k) - R \leq \text{rank}(G_i[K, J])$$

for every $K \subseteq \{1, \cdots, m_{i+1}\}$ and $J \subseteq \{1, \cdots, m_i\}$. Therefore $\mathbf{d}_i$ is a feasible flow for G_i and $\mathbf{d}$ is a feasible flow for network $\mathcal{N}$.

Our next step is to show that if $\mathbf{d} = (\ell_i(j) : 1 \leq i \leq M, 1 \leq j \leq m_i)$ is a feasible rate–R flow in network $\mathcal{N}$ and $\mathbf{d} \in \mathbf{D}$, then $\phi = T^{-1}(\mathbf{d})$ is a feasible submodular flow in $\mathcal{G}$. Notice that the first property of a feasible submodular flow is satisfied because $\mathbf{d}$ is a non–negative vector and hence for any $a_i(j) \in \mathcal{A}\setminus \{a_0\}$, $\phi(a_i(j)) = \ell_i(j)$ is a non–negative integer. Furthermore, $\phi(a_0) = \ell_1(1)$ and $0 \leq \phi(a_0) \leq R_0$. Next fix $H \subseteq \mathcal{V}$. For the second property we have to show that:

$$\sum_{a\in\Delta^+ H} \phi(a) - \sum_{a\in\Delta^- H} \phi(a) \leq \rho(H).$$

Let $H_i = H \cap (\{\upsilon_i(1), \cdots, \upsilon_i(m_i)\} \cup \{\nu_{i+1}(1), \cdots, \nu_{i+1}(m_{i+1})\})$ for $1 \leq i \leq M - 1$. We notice that H_i's for $1 \leq i \leq M - 1$ form a partition of H, i.e., $\bigcup_{i=1}^{M-1} H_i = H$ and $H_i \cap H_j = \emptyset$ for $i \neq j$. This implies that

$$\sum_{a\in\Delta^+ H} \phi(a) - \sum_{a\in\Delta^- H} \phi(a) = \sum_{i=1}^{M-1} \left(\sum_{a\in\Delta^+ H_i} \phi(a) - \sum_{a\in\Delta^- H_i} \phi(a) \right). \tag{4.14}$$

To arrive at (4.14), observe that each $a \in \Delta^+ H$ corresponds to exactly one i such that $a \in \Delta^+ H_i$ and each $a \in \Delta^- H$ corresponds to exactly one i such that $a \in \Delta^- H_i$. Next suppose that for some i, $a \in \Delta^+ H_i$ but $a \notin \Delta^+ H$. This implies that there exists exactly one j such that $a \in \Delta^- H_j$. Therefore the terms $\phi(a)$ and $-\phi(a)$ cancel each other out on the right hand side. Similarly if for some i, $a \in \Delta^- H_i$ but $a \notin \Delta^- H$, then there exists exactly one j such that $a \in \Delta^+ H_j$. Therefore the terms $-\phi(a)$ and $+\phi(a)$ cancel each other out on the right–hand side. Therefore the right–hand side of (4.14) is equal to its left–hand side.

Next fix $1 \leq i \leq M-1$. We have $K_i(H) = K_i(H_i)$ and $J_i(H) = J_i(H_i)$, and thus $\rho_i(H) = \rho_i(H_i)$. Therefore

$$\rho(H) = \sum_{i=1}^{M-1} \rho_i(H) = \sum_{i=1}^{M-1} \rho_i(H_i). \tag{4.15}$$

Since $\mathbf{d}$ is a feasible flow in $\mathcal{N}$,

$$\mathbf{d}_i = (\ell_i(1), \cdots, \ell_i(m_i); \ell_{i+1}(1), \cdots, \ell_{i+1}(m_{i+1}))$$

is a feasible flow for G_i for $1 \leq i \leq M-1$. Let $J = J_i(H_i)$ and $K = K_i(H_i)$. By the feasibility of $\mathbf{d}_i$ we have

$$\sum_{j \in J} \ell_i(j) + \sum_{k \in K} \ell_{i+1}(k) - R \leq \text{rank}(G_i[K, J]).$$

By substituting $\text{rank}(G_i[K, J]) = \rho_i(H_i)$, $\sum_{k \in K} \ell_{i+1}(k) = \sum_{a \in \Delta^+ H_i} \phi(a)$, and $R - \sum_{j \in J} \ell_i(j) = \sum_{a \in \Delta^- H_i} \phi(a)$ we find that

$$\sum_{a \in \Delta^+ H_i} \phi(a) - \sum_{a \in \Delta^- H_i} \phi(a) \leq \rho_i(H_i). \tag{4.16}$$

Summing the preceding inequalities for $1 \leq i \leq M-1$ and by using (4.14) and (4.15) we find that ϕ satisfies the second property of a submodular flow.

$\square$

Next we prove the following lemma:

Lemma 35 *The maximum feasible submodular flow* $\phi(a_0)$ *in graph* $\mathcal{G}$ *is* $\min\{R_0, C\}$.

Proof We use Theorem 32 to prove this result. We need to find the value of

$$\gamma = \min \left[c_u(a_0), \min_{X \subseteq \mathcal{V}} \left\{ \sum_{a \in \Delta^- X} c_u(a) - \sum_{a \in \Delta^+ X \setminus \{a_0\}} c_l(a) + \rho(X) : a_0 \in \Delta^+ X \right\} \right].$$

Since $c_u(a_0) = R_0$ and $c_l(a) = 0$ for every $a \in \mathcal{A}$, we have

$$\gamma = \min\left[R_0, \min_{X \subseteq \mathcal{V}}\left\{\sum_{a \in \Delta^- X} c_u(a) + \rho(X) : a_0 \in \Delta^+ X\right\}\right].$$

Furthermore, since for every $a \in \mathcal{A} \setminus \{a_0\}$, $c_u(a) = +\infty$, we can restrict the minimization to the subsets X such that $\Delta^- X = \emptyset$. Therefore if for $2 \leq i \leq M-1$, $\nu_i(j) \notin X$, we should also have $\upsilon_i(j) \notin X$. In addition, since $a_0 \in \Delta^+ X$, we have $\nu_M(1) \in X$ and $\upsilon_1(1) \notin X$. Next suppose that for $X \subseteq \mathcal{V}$ we have $\Delta^- X = \emptyset$, $a_0 \in \Delta^+ X$, and for some $2 \leq i \leq M-1$ and $1 \leq j \leq m_i$ we have $\nu_i(j) \in X$ and $\upsilon_i(j) \notin X$. Set $X' = X \cup \{\upsilon_i(j)\}$. We still have $\Delta^- X' = \emptyset$ and $a_0 \in \Delta^+ X'$. Also for every $l \neq i$, $K_l(X) = K_l(X')$ and $J_l(X') = J_l(X)$. Thus, $\rho_l(X') = \rho_l(X)$. But we have $J_i(X') = J_i(X) \setminus \{j\}$ and $K_i(X') = K_i(X)$. Therefore, $\text{rank}(G_i[K_i(X'), J_i(X')]) \leq \text{rank}(G_i[K_i(X), J_i(X)])$ and $\rho_i(X') \leq \rho_i(X)$. Consequently $\rho(X') \leq \rho(X)$. This shows that for computing γ, X can be ignored given the set X'. In this way we can consider the minimization only over the sets $X \subseteq \mathcal{V}$ such that $a_0 \in \Delta^+ X$ and for every $2 \leq i \leq M-1$ and $1 \leq j \leq m_i$, either $\nu_i(j)$ and $\upsilon_i(j)$ are both in X or are both out of X. Call the set of all such sets $\mathbf{X}$ and let $\mathbf{M}$ be the set of all cuts in network $\mathcal{N}$ that separate s from t. Consider the bijection $B : \mathbf{X} \to \mathbf{M}$ in which $B(X) = \{s\} \bigcup \{\upsilon_i(j) : 2 \leq i \leq M-1, 1 \leq j \leq m_i, \upsilon_i(j) \notin X\}$. In words, if for $2 \leq i \leq M-1$ and $1 \leq j \leq m_i$, the nodes $\nu_i(j)$ and $\upsilon_i(j)$ are both out of X we include $\upsilon_i(j)$ in $B(X)$ and if are both in X we exclude $\upsilon_i(j)$ from X. Furthermore, we let $s \in B(X)$ and $t \notin B(X)$. In this way every separating cut in network $\mathcal{N}$ corresponds to a set $X \in \mathbf{X}$ and vice versa. Observe that $G_i[K_i(X), J_i(X)]$ is the transfer function from the nodes $\upsilon_i(j) \notin X$ to the nodes $\nu_{i+1}(k) \in X$ in graph $\mathcal{G}$, which is the same as the transfer function from the nodes $\upsilon_i(j) \in B(X)$ to the nodes $\upsilon_{i+1}(k) \notin B(X)$ in $\mathcal{N}$. Therefore

$G_i[K_i(X), J_i(X)] = G_i[B(X)]$. Thus

$$\rho(X) = \sum_{i=1}^{M-1} \rho_i(X) = \sum_{i=1}^{M-1} \text{rank}(G_i[K_i(X), J_i(X)])$$
$$= \sum_{i=1}^{M-1} \text{rank}(G_i[B(X)]) = C(B(X)).$$

Therefore for every $X \in \mathbf{X}$, $\rho(X) = C(B(X))$, and the result follows by

$$\gamma = \min\left\{R_0, \min_{X\in\mathbf{X}} \rho(X)\right\} = \min\left\{R_0, \min_{\Omega\in\mathbf{M}} C(\Omega)\right\} = \min\left\{R_0, C\right\}.$$

$\square$

We have chosen c_l and $\rho(H)$ to be integer valued. Also $c_u(a)$ can be selected as a very large integral constant for all edges in $\mathcal{A}\setminus\{a_0\}$ and is an integer for a_0 with $c_u(a_0) = R_0$. Therefore the second part of Theorem 32 implies that a submodular flow with $\phi(a_0) = \min\{R_0, C\}$ is feasible in $\mathcal{G}$ with integer values for every integer $R_0 \geq 0$. By Lemma 34, all flows with rate $R_0 \leq C$ are achievable in $\mathcal{N}$, proving Theorem 31.

E. Algorithmic Discussion

In this section we offer a two–part discussion on the complexity of constructing a transmission scheme for network $\mathcal{N}$. We begin by finding a solution for a feasible flow for a matrix.

Lemma 36 *Given matrix G as in Definition 30, a feasible flow $\mathbf{d} = (h_1, \cdots, h_n; g_1, \cdots, g_m)$ can be found in time $O(d^6)$, where d is the maximum number of rows and columns of G.*

Proof We use some basic facts and results from matroid theory. In Subsection 1 we saw one definition of a matroid in terms of the rank function r. Next consider the following equivalent definition of a matroid which focuses upon its independent sets:

Definition 37 *Given a set E and a family of its subsets $\mathcal{I}$, we say that the pair $(E,\mathcal{I})$ is a matroid if:*

1. $\emptyset \in \mathcal{I}$.

2. If $A \in \mathcal{I}$, then $B \in \mathcal{I}$ for every $B \subseteq A$.

3. If $X, Y \in \mathcal{I}$ and $|X| > |Y|$, then there exists $x \in X$ such that $x \notin Y$ and $Y \cup \{x\} \in \mathcal{I}$.

$\mathcal{I}$ is the set of all independent sets of the matroid.

For our purpose we construct a matroid as follows. Given a set E and a family of disjoint subsets $A_1, \cdots, A_n$ of E, let $\ell_1, \cdots, \ell_n$ be non–negative integers. Let

$$\mathcal{I} = \left\{ X : X = \bigcup_{i=1}^{n} X_i, X_i \subseteq A_i, |X_i| \leq \ell_i \right\}. \tag{4.17}$$

We next show that $\mathcal{M}_2 = (E,\mathcal{I})$ is a matroid with the set $\mathcal{I}$ of independent sets by considering the conditions of Definition 37. The first and second properties are easy to check from the definition. For the third property, consider two subsets $X, Y \in \mathcal{I}$ with $|X| > |Y|$. Suppose that $X = \bigcup_{i=1}^{n} X_i$ and $Y = \bigcup_{i=1}^{n} Y_i$, where $X_i, Y_i \subseteq A_i$, $|X_i| \leq \ell_i, |Y_i| \leq \ell_i$. Since $|X| > |Y|$, there exists $1 \leq i \leq n$ such that $|X_i| > |Y_i|$. Therefore $X_i \backslash Y_i$ is non–empty. Choose an element $x \in X_i$ such that $x \notin Y_i$, and let $Z_i = Y_i \cup \{x\}$. Since $Z_i \subseteq A_i$ and $|Z_i| = |Y_i| + 1 \leq |X_i| \leq \ell_i$ the set $Y \cup \{x\} = Y_1 \cup \cdots \cup Y_{i-1} \cup Z_i \cup Y_{i+1} \cup \cdots \cup Y_n \in \mathcal{I}$, and therefore the third condition is satisfied.

Let $\mathcal{M}_1 = (E, r)$ be another matroid with rank function r defined on the set E. Let $\mathcal{I}_1 = \{X : X \subseteq E, |X| = r(X)\}$ represent the set of all independent sets of $\mathcal{M}_1$. In the *maximum–size common independent set problem* (see [78, Ch. 41]) we look for a set $M \in \mathcal{I} \cap \mathcal{I}_1$ of maximum cardinality, i.e., $|M| = \max_{I \in \mathcal{I} \cap \mathcal{I}_1} |I|$. Suppose that there exists disjoint subsets $\hat{A}_1 \subseteq A_1, \cdots, \hat{A}_n \subseteq A_n$ with $|\hat{A}_i| = \ell_i$ and $\hat{A} = \hat{A}_1 \cup \cdots \cup \hat{A}_n$ an independent set in $\mathcal{M}_1$. We observe that in this case $\hat{A}$ is a solution of the maximum–size common independent set problem. There are polynomial–time algorithms for solving the maximum–size common independent set problem; for instance, the "cardinality matroid intersection algorithm" of [78, §41.2] has a complexity of $O(\ell^2|E|(\ell + t))$ where in our setting $\ell = \ell_1 + \cdots + \ell_n$, and t is the maximum time needed to evaluate $r(A)$ for any subset $A \subseteq E$.

Returning to the problem of finding a solution for a feasible flow, suppose that matrix G admits a flow $\mathbf{d} = (h_1, \cdots, h_n; g_1, \cdots, g_m)$. As we have shown in the proof of Theorem 25, finding a solution $\hat{G}$ to the flow $\mathbf{d}$ is equivalent to finding subsets $\hat{P}[1] \subseteq P[1], \cdots, \hat{P}[m] \subseteq P[m]$ and $\tilde{Q}[1] \subseteq Q[1], \cdots, \tilde{Q}[n] \subseteq Q[n]$ such that $|\hat{P}[k]| = g_k$ and $|\tilde{Q}[j]| = |Q[j]| - h_j$ and $\bigcup_{k=1}^{m} \hat{P}[k] \bigcup_{j=1}^{n} \tilde{Q}[j]$ is an independent set in the matroid $(P \cup Q, r_0)$. Let t be the time needed to evaluate r_0 for any subset of $P \cup Q$, where r_0 is defined in (4.2), and let d be the maximum number of rows and columns of G. By the previous argument, the approach of applying the cardinality matroid intersection algorithm of [78, §41.2] has complexity $O(\ell^2|P \cup Q|(\ell + t))$, where

$$\ell = \sum_{i=1}^{m} g_i + \sum_{j=1}^{n}(|Q[j]| - h_j) = \sum_{j=1}^{n} |Q[j]| = |Q| \leq d.$$

To evaluate r_0 we need to find the rank of a submatrix of the matrix G, which has a complexity of at most $O(d^3)$ by the Gaussian elimination technique [6]. Also since $|P \cup Q| \leq 2d$, we find that the total complexity of this problem is $O(\ell^2|P \cup Q|(\ell + t)) = O(d^6)$.

$\square$

We next discuss the complexity of finding a rate–R flow in network $\mathcal{N}$ when $R \leq C$.

Lemma 38 *Given rate R, the complexity of finding a flow $\mathbf{d}$ in network $\mathcal{N}$ with rate $R \leq C$ is $O(|V|^9 d^3 M)$.*

Proof As we discussed in the proof of Theorem 31, finding a flow vector $\mathbf{d}$ in $\mathcal{N}$ with rate R is equivalent to finding a maximum submodular flow in graph $\mathcal{G}$. There are several polynomial–time algorithms for the latter problem. Here we use the one which is described in [31], which has a running time of $O(|\mathcal{V}|^3 t_m)$. The notation $|\mathcal{V}|$ denotes the number of nodes in graph $\mathcal{G}$, and t_m represents the time needed to evaluate the following minimization problem for some two fixed nodes $u, v \in \mathcal{V}$, and for some fixed function $\eta : \mathcal{V} \to \mathbb{R}$:

$$\min \left\{ \rho(X) - \sum_{x \in X} \eta(x) : X \subseteq \mathcal{V}, v \in X, u \notin X \right\}.$$

This problem is a "submodular minimization problem" [36] and can be solved in time $t_m = O(|\mathcal{V}|^6 t_\rho)$ by the submodular minimization algorithm in [77], where t_ρ is the time of evaluating function ρ for any subset of $\mathcal{V}$. Since $\rho(X) = \sum_{i=1}^{M-1} \rho_i(X)$, there are $M-1$ rank evaluations needed to calculate the function ρ. In each step we have to evaluate the rank of a submatrix of a matrix G_i, which if we assume an upper bound of d on each of its dimensions, requires $O(d^3)$ time by the Gaussian elimination technique [6]. Therefore $t_\rho = O(d^3 M)$ and $t_m = O(|\mathcal{V}|^6 d^3 M)$. Finally $O(|\mathcal{V}|^3 t_m) = O(|\mathcal{V}|^9 d^3 M) = O(|V|^9 d^3 M)$ is the complexity of finding a flow vector of rate R in network $\mathcal{N}$.

$\square$

The algorithm in Lemma 38 finds the flow vector $\mathbf{d}$ and consequently every $\mathbf{d}_i$ for $1 \leq i \leq M-1$. Then by Lemma 36, for each G_i we can find a solution $\hat{G}_i$ in time $O(d^6)$, $1 \leq i \leq M-1$, leading to a total complexity for this part of $O(d^6 M)$. The total complexity for constructing our coding scheme will be $O(|V|^9 d^3 M) + O(d^6 M)$.

In order to most effectively use Lemma 38, we need to know the capacity of the network $\mathcal{N}$.

Lemma 39 *The capacity of the network $\mathcal{N}$ can be found in time $O(|V|^6 M d^3)$.*

Proof Let $\mathbf{\Omega}_s$ be the set of all cuts that separate s from t in the network $\mathcal{N}$. We first show that the function $C(\Omega) : \mathbf{\Omega}_s \to \mathbb{Z}$ is a submodular set function. Let $\Omega_1, \Omega_2 \in \mathbf{\Omega}_s$. We observe that $\Omega_3 = \Omega_1 \cap \Omega_2$ and $\Omega_4 = \Omega_1 \cup \Omega_2$ are also in $\mathbf{\Omega}_s$. Fix $i \in \{1, \cdots, M-1\}$ and let $P = \{1, \cdots, m_{i+1}\}$. For $j \in \{1, 2, 3, 4\}$ define $A_j = \{a : v_i(a) \in \Omega_j\}$ and $B_j = \{b : v_{i+1}(b) \in \Omega_j\}$. By definition, $G_i[\Omega_j] = G_i[P \setminus B_j, A_j]$. Also $A_1 \cap A_2 = A_3$, $B_1 \cap B_2 = B_3$, $A_1 \cup A_2 = A_4$, and $B_1 \cup B_2 = B_4$. By applying Lemma 28 to the transpose of matrix G_i it follows that

$$
\begin{aligned}
&\text{rank}(G_i[\Omega_1]) + \text{rank}(G_i[\Omega_2]) \\
&= \text{rank}(G_i[P \setminus B_1, A_1]) + \text{rank}(G_i[P \setminus B_2, A_2]) \\
&\geq \text{rank}(G_i[P \setminus (B_1 \cap B_2), (A_1 \cap A_2)]) + \text{rank}(G_i[P \setminus (B_1 \cup B_2), (A_1 \cup A_2)]) \\
&= \text{rank}(G_i[P \setminus B_3, A_3]) + \text{rank}(G_i[P \setminus B_4, A_4]) \\
&= \text{rank}(G_i[\Omega_3]) + \text{rank}(G_i[\Omega_4]).
\end{aligned}
$$

Summing the preceding inequality over all values of $i \in \{1, \cdots, M-1\}$ we find that

$$C(\Omega_1) + C(\Omega_2) \geq C(\Omega_3) + C(\Omega_4),$$

which is the submodularity relationship.

Since the capacity C is the minimum value of the set $\{C(\Omega) : \Omega \in \mathbf{\Omega}_s\}$ we can use the submodular function minimization algorithm of [77] to find C in time $O(|V|^6 t_C)$, where t_C is the time needed for calculating function $C(\Omega)$ for any $\Omega \in \mathbf{\Omega}_s$. This calculation involves $M-1$ rank calculation each of a submatrix of matrix G_i for $1 \leq i \leq M-1$. As we discussed in the proof of Lemma 38, this has a total time complexity of $O(Md^3)$. Therefore we can find the capacity in time $O(|V|^6 Md^3)$.

□

CHAPTER V

A DETERMINISTIC POLYNOMIAL–TIME ALGORITHM FOR CONSTRUCTING A MULTICAST CODING SCHEME FOR LINEAR DETERMINISTIC RELAY NETWORKS

A. Introduction

In this chapter we build upon our work in [85, 84] to design a simple and low complexity transmission scheme for a multicast session over an LDRN. Our scheme will be constructed by progressively combining the coding schemes for unicast sessions from the source to each destination. In many ways our scheme is similar to and is a generalization of the scheme in [37] for a multicast session in wired networks. We will offer both randomized and deterministic versions of our algorithm and show that when there are g destinations, $\lceil \log(g+1) \rceil$ uses of the network suffice to achieve capacity, which resembles the result for wired networks [37].

We will next review our earlier results on a single unicast session [85, 84] in Section B and then discuss our coding construction for a multicast session in Section C.

B. A Single Unicast Session

In this section we briefly explain the coding scheme for a single unicast session from [85, 84]. This will be the building block of our multicast coding scheme. The setting of the network in this chapter is similar to the setting of the network in Chapter IV by noticing that here we consider g destination nodes of $t_l \triangleq v_{K_l}(d_l), l \in \{1, \cdots, g\}$, distributed in layers $K_1, K_2, \cdots, K_g$.

Recall that for each $i \in \{1, \cdots, M-1\}$ the transmitted vector of layer i and the received vector of layer $i+1$ are related through matrix G_i by $\mathbf{y}_{i+1} = G_i \mathbf{x}_i$.

For each layer $i \in \{1, \cdots, M\}$ label the indices of the vector $\mathbf{y}_i$ with the elements of a set P_i and label the indices of the vector $\mathbf{x}_i$ with the elements of a set Q_i. We choose all sets P_i and Q_i to be disjoint for different values of i. For any $A \subseteq P_i$, let $\mathbf{y}_i(A)$ denote the subvector of $\mathbf{y}_i$ corresponding to indices with labels from set A. Similarly, for any $B \subseteq Q_i$, let $\mathbf{x}_i(B)$ denote the subvector of $\mathbf{x}_i$ associated with indices with labels from set B. Next partition each set P_i into subset $P_i = \cup_{j=1}^{m_i} P_i[j]$ and Q_i into subsets $Q_i = \cup_{j=1}^{m_i} Q_i[j]$ such that $P_i[j]$ is the subset of indices of $\mathbf{y}_i$ that belong to the subvector $\mathbf{y}_i[j]$ and $Q_i[j]$ is the subset of indices of $\mathbf{x}_i$ that belong to the subvector $\mathbf{x}_i[j]$. Therefore we have $\mathbf{y}_i[j] = \mathbf{y}_i(P_i[j])$ and $\mathbf{x}_i[j] = \mathbf{x}_i(Q_i[j])$ for any $j \in \{1, \cdots, m_i\}$. For any $i \in \{1, \cdots, M-1\}$ we will use the sets P_{i+1} and Q_i to label the rows and the columns of the matrix G_i such that for each $p \in P_{i+1}$ the row of G_i corresponding to the element $\mathbf{y}_{i+1}(p)$ is labeled with p and for each $q \in Q_i$ the column of G_i corresponding to the element $\mathbf{x}_i(q)$ is labeled with q. For $p \in P_{i+1}$ and $q \in Q_i$ let $G_i(p, q)$ denote the element in row p and column q of matrix G_i. For $A \subseteq P_{i+1}$ and $B \subseteq Q_i$ let $G_i(A, B)$ denote the submatrix of G_i consisting of the rows in A and the columns in B. Our labeling implies that $G_i(P_{i+1}[k], Q_i[j]) = G_i[k, j]$ for any $j \in \{1, \cdots, m_i\}$ and $k \in \{1, \cdots, m_{i+1}\}$.

If node s holds a column vector message $\mathbf{w} \in \mathbb{F}^{R \times 1}$ and we are looking at a linear coding scheme, then at each layer $i \in \{1, \cdots, M\}$, each element of vectors $\mathbf{x}_i$ and $\mathbf{y}_i$ will be a linear transformation of the vector $\mathbf{w}$. We represent the "global coding vector" (see [37]) for the element $\mathbf{x}_i(q), q \in Q_i$, with row vector $\boldsymbol{x}_i(q) \in \mathbb{F}^{1 \times R}$ such that $\mathbf{x}_i(q) = \boldsymbol{x}_i(q)\mathbf{w}$ and the global coding vector for the element $\mathbf{y}_i(p), p \in P_i$, with row vector $\boldsymbol{y}_i(p) \in \mathbb{F}^{1 \times R}$ such that $\mathbf{y}_i(p) = \boldsymbol{y}_i(p)\mathbf{w}$. For subsets $B \subseteq Q_i$ and $A \subseteq P_i$ we use the notation $\boldsymbol{x}_i(B)$ and $\boldsymbol{y}_i(A)$ to respectively denote the matrices that are formed by the vectors $\boldsymbol{x}_i(q)$ and $\boldsymbol{y}_i(p)$ for $q \in B$ and $p \in A$. Therefore we have $\mathbf{x}_i(B) = \boldsymbol{x}_i(B)\mathbf{w}$ and $\mathbf{y}_i(A) = \boldsymbol{y}_i(A)\mathbf{w}$.

Suppose that the network supports a rate–R unicast connection between source node s and a destination node $t = v_K(d)$ for $K \leq M$ and $d \in \{1, \cdots, m_K\}$. The main result of

[85, 84] can be summarized in the following theorem:

Theorem 40 *For each $1 \leq i \leq K$ and for each $1 \leq j, k \leq m_i$ there exist subsets $\hat{Q}_i[j] \subseteq Q_i[j]$ and $\hat{P}_i[k] \subseteq P_i[k]$ such that the following hold*

1. $|\hat{P}_i[j]| = |\hat{Q}_i[j]|$ *for* $i \in \{1, \cdots, K\}, j \in \{1, \cdots, m_i\}$,
2. $\sum_{j=1}^{m_i} |\hat{P}_i[j]| = \sum_{j=1}^{m_i} |\hat{Q}_i[j]| = R,$ *for* $i \in \{1, \cdots, K-1\}$,
3. $|\hat{P}_K[d]| = R$ *and* $|\hat{P}_K[k]| = 0$ *for* $k \neq d$,
4. $G_i(\bigcup_{k=1}^{m_{i+1}} \hat{P}_{i+1}[k], \bigcup_{j=1}^{m_i} \hat{Q}_i[j])$ *is a nonsingular matrix for* $i \in \{1, \cdots, K-1\}$.

Furthermore such subsets can be found by an algorithm that runs in a time that is polynomial in the size of the network $\mathcal{N}$.

We call the subsets $\hat{Q}_i[j] \subseteq Q_i[j]$ and $\hat{P}_i[k] \subseteq P_i[k]$ for $i \in \{1, \cdots, M\}$ and $j, k \in \{1, \cdots, m_i\}$ a *flow* of rate R in the LDRN from the source node s to the destination node t.

The four properties of a flow in Theorem 40 depend on $G_1, ..., G_{M-1}$ and do not depend on the specific choice of the set $\hat{P}_1[1]$ among all subsets of $P_1[1]$ with size R. Therefore, if there exists a rate–R flow, we can set $\hat{P}_1[1]$ to be any subset of $P_1[1]$ of size R.

Notice that the existence of a flow of rate R implies the following simple and low complexity coding scheme of rate R from the source s to the destination t: To send message $\mathbf{w} \in \mathbb{F}^{R \times 1}$, source node $s = v_1(1)$ sets $\mathbf{y}_1(\hat{P}_1[1]) = \mathbf{w}$ and $\mathbf{y}_1(P_1[1] \setminus \hat{P}_1[1]) = \mathbf{0}$. Next, any node $v_i(j), i \in \{1, \cdots, M\}, j \in \{1, \cdots, m_i\}$, in the network forms the vector $\mathbf{x}_i[j]$ by setting

$$\mathbf{x}_i(\hat{Q}_i[j]) = \mathbf{y}_i(\hat{P}_i[j]).$$

We say that element $p \in \hat{P}_i[j]$ is "matched" with element $q \in \hat{Q}_i[j]$ when $\mathbf{x}_i(q)$ is set to $\mathbf{y}_i(p)$ through the preceding equation (see Figure 20 for an example of a flow). We further

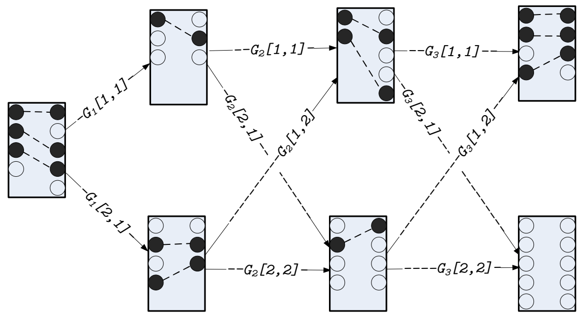

Fig. 20. An example of a rate–3 flow from the source node s to the destination node t_1. Here the matched elements of flow are connected together through dashed lines.

let $\mathbf{x}_i(Q_i[j] \setminus \hat{Q}_i[j]) = \mathbf{0}$. From the properties of flow it follows that at the destination $t = v_K(d)$

$$\mathbf{x}_K(\hat{Q}_K[d])$$
$$= G_{K-1}(\hat{P}_K[d], \bigcup_{j=1}^{m_{K-1}} \hat{Q}_{K-1}[j]) \cdots G_2(\bigcup_{k=1}^{m_3} \hat{P}_3[k], \bigcup_{j=1}^{m_2} \hat{Q}_2[j]) G_1(\bigcup_{k=1}^{m_2} \hat{P}_2[k], \hat{Q}_1[1])\mathbf{w}.$$

Since each matrix $G_i(\bigcup_{k=1}^{m_{i+1}} \hat{P}_{i+1}[k], \bigcup_{j=1}^{m_i} \hat{Q}_i[j])$ is nonsingular, node t can recover vector $\mathbf{w}$ from the received vector $\mathbf{x}_K(\hat{Q}_K[d])$ through a linear transformation.

C. A Coding Scheme for a Multicast Session

Assume that there are g destination nodes $t_1, \cdots, t_g$ in the network and the min–cut capacity from the source node s to each destination is at least R. We are interested in a multicast coding scheme in which all destinations can simultaneously receive the message $\mathbf{w} \in \mathbb{F}^{R\times 1}$ of the source. Our scheme will be designed by combining the flows of rate R from the source to each destination.

Suppose that $t_l = v_{K_l}(d_l)$ for $l \in \{1, \cdots, g\}$. From Theorem 40 for each $t_l, l \in \{1, \cdots, g\}$, there exists a flow with subsets $P_i^l[k] \subseteq P_i[k]$ and $Q_i^l[j] \subseteq Q_i[j]$ for $1 \leq i \leq K_l$

and $1 \leq j, k \leq m_i$ such that:

1. $|P_i^l[j]| = |Q_i^l[j]|$ for $i \in \{1, \cdots, K_l\}, j \in \{1, \cdots, m_i\}$,

2. $\sum_{j=1}^{m_i} |P_i^l[j]| = \sum_{j=1}^{m_i} |Q_i^l[j]| = R$, for $i \in \{1, \cdots, K_l - 1\}$,

3. $|P_{K_l}^l[d_l]| = R$ and $|P_{K_l}^l[k]| = 0$ for $k \neq d_l$,

4. $G_i(\bigcup_{k=1}^{m_{i+1}} P_{i+1}^l[k], \bigcup_{j=1}^{m_i} Q_i^l[j])$ is a nonsingular matrix for $i \in \{1, \cdots, K_l - 1\}$.

Since $P_1^l[1], l \in \{1, \cdots, g\}$, can be any subset of $P_1[1]$ of size R, we set all subsets $P_1^l[1], l \in \{1, \cdots, g\}$, to be the same subset of $P_1[1]$.

Our design criterion for a multicast coding scheme is that for each destination $t_l, l \in \{1, \cdots, g\}$, at each layer $i \in \{1, \cdots, K_l\}$, the global coding vectors correponding to the elements of the vectors $\mathbf{y}_i(P_i^l[j])$ for $j \in \{1, \cdots, m_i\}$ must be linearly independent vectors and hence the length–R vector

$$\mathbf{y}_i\left(\bigcup_{j=1}^{m_i} P_i^l[j]\right)$$

can uniquely determine the message vector $\mathbf{w}$. In other words we require for each destination t_l and each layer $i \in \{1, \cdots, K_l\}$:

- Condition (*): the matrix $\boldsymbol{y}_i\left(\bigcup_{j=1}^{m_i} P_i^l[j]\right)$ must be nonsingular.

The destination node $t_l = v_{K_l}(d_l)$ will receive the length–R vector

$$\mathbf{y}_{K_l}(P_{K_l}^l[d_l]) = \boldsymbol{y}_{K_l}(P_{K_l}^l[d_l])\mathbf{w}.$$

Since $\boldsymbol{y}_{K_l}(P_{K_l}^l[d_l])$ is a nonsingular matrix, t_l will be able to decode message $\mathbf{w}$.

Notice that at each node $v_i(j)$ for $i \in \{2, \cdots, M\}$ we only have control over the design of the coding vectors $\boldsymbol{x}_i(q)$ for $q \in Q_i[j]$ which can be a linear function of the coding vectors $\{\boldsymbol{y}_i(p) : p \in P_i[j]\}$. The coding vectors $\boldsymbol{y}_i(p)$ for $p \in P_i[j]$ are determined from the coding vectors of the previous layer and matrix G_{i-1}. In our design we will assign coding vectors layer by layer, starting from the first layer. At each layer i we fix an arbitrary order on the elements of the set Q_i and assign the coding vectors $\boldsymbol{x}_i(q)$ in this order.

Initialization: We start from the first layer. Since $P_1^l[1]$ is the same subset for every $l \in \{1, \cdots, g\}$ we set $\boldsymbol{y}_1(P_1^l[1]) = I_{R\times R}$, i.e., the $R \times R$ identity matrix, and set $\boldsymbol{y}_1(P_1[1] \setminus P_1^l[1]) = \mathbf{0}$ for every $l \in \{1, \cdots, g\}$. In other words we set $\mathbf{y}_1(P_1^l[1]) = \mathbf{w}$ and $\mathbf{y}_1(P_1[1] \setminus P_1^l[1]) = \mathbf{0}$ for every $l \in \{1, \cdots, g\}$. Therefore Condition (*) will be satisfied for all destinations in the first layer.

Inductive Step: We continue our coding construction inductively. Suppose that the Condition (*) holds for layer i and for all destinations $t_l = v_{K_l}(d_l)$ with $K_l \geq i$. Next we will design the coding vectors $\boldsymbol{x}_i(q)$ for $q \in Q_i$ one by one and in the order of the elements of Q_i so that at the end the Condition (*) holds for layer $i+1$ and all destinations $t_l = v_{K_l}(d_l)$ with $K_l \geq i+1$.

At this step of the algorithm for each destination t_l with $K_l \geq i+1$ we maintain two matrices. One is the matrix A_l which is initially

$$A_l = \boldsymbol{y}_i\left(\bigcup_{j=1}^{m_i} P_i^l[j]\right),$$

and is updated throughout the algorithm. The other matrix is

$$F_l = G_i(\bigcup_{k=1}^{m_{i+1}} P_{i+1}^l[k], Q_l'),$$

where initially $Q_l' = \bigcup_{j=1}^{m_i} Q_i^l[j]$ and it is updated throughout the algorithm. Throughout the algorithm we maintain the invariance that the product $F_l A_l$ is a nonsingular matrix for every destination t_l with $K_l \geq i+1$. We will also verify that after all of the elements of Q_i are processed, for every destination t_l with $K_l \geq i+1$ we will have $F_l A_l = \boldsymbol{y}_{i+1}\left(\bigcup_{j=1}^{m_{i+1}} P_{i+1}^l[j]\right)$, which is sufficient for Condition (*) to hold at layer $i+1$.

A_l is initially invertible since Condition (*) holds for layer i. Matrix F_l is also initially nonsingular by the definition of a flow to destination t_l given in Theorem 40. Therefore the product F_lA_l is initially nonsingular. Next we will explain the design of the coding vector $\boldsymbol{x}_i(q)$ for $q \in Q_i$ and describe the updating process of F_l and A_l for every destination t_l with $K_l \geq i+1$. We consider two cases:

1. If q is part of the flow for destination t_l, i.e., $q \in Q_i^l[j]$ for some $j \in \{1, \cdots, m_i\}$, then update matrix A_l by replacing row $\boldsymbol{y}_i(p_l)$ with $\boldsymbol{x}_i(q)$, which we will later explain how to design starting with the Analysis of Case 1 below. Here $p_l \in P_i^l[j]$ is the unique element that is matched with $q \in Q_i^l[j]$ in the flow for destination t_l. There is no change needed for matrix F_l.

2. If q is not part of the flow for destination t_l, then update A_l by adding a new row $\boldsymbol{x}_i(q)$ to it and insert a column $G_i(\bigcup_{k=1}^{m_{i+1}} P_{i+1}^l[k], \{q\})$ into F_l so that the set of column indices grows from Q_l' to $Q_l' \cup \{q\}$. In this step we place $\boldsymbol{x}_i(q)$ in the row of A_l counting from the top which is the same as the position of the new column $G_i(\bigcup_{k=1}^{m_{i+1}} P_{i+1}^l[k], \{q\})$ in the updated F_l counting from the left. The constraints on $\boldsymbol{x}_i(q)$ will be discussed in the Analysis of Case 2.

When we have gone through all of the elements of Q_i, F_l would be $G_i(\bigcup_{k=1}^{m_{i+1}} P_{i+1}^l[k], Q_i)$ and A_l would be the matrix $\boldsymbol{x}_i(Q_i)$. Therefore we have

$$F_lA_l = G_i(\bigcup_{k=1}^{m_{i+1}} P_{i+1}^l[k], Q_i)\boldsymbol{x}_i(Q_i) = \boldsymbol{y}_{i+1}\left(\bigcup_{j=1}^{m_{i+1}} P_{i+1}^l[j]\right)$$

where the second equation holds since G_i is the transfer matrix from $\boldsymbol{x}_i(Q_i) = \boldsymbol{x}_i$ to $\boldsymbol{y}_{i+1}$. This equation guarantees that $\boldsymbol{y}_{i+1}\left(\bigcup_{j=1}^{m_{i+1}} P_{i+1}^l[j]\right)$ is nonsingular, as desired.

Next we analyze each case and find the condition that $\boldsymbol{x}_i(q)$ needs to satisfy in order for F_lA_l to remain nonsingular:

1. Analysis of case 1

Without loss of generality suppose that $\boldsymbol{x}_i(q)$ is the first row of A_l and that matrix A_l after the update is of the form

$$A_l = \begin{bmatrix} \boldsymbol{x}_i(q) \\ A'_l \end{bmatrix}.$$

Therefore A_l before the update is of the form $\begin{bmatrix} \boldsymbol{y}_i(p_l) \\ A'_l \end{bmatrix}$. We require that the matrix $F_l A_l$ be nonsingular. We write

$$F_l = \begin{bmatrix} \boldsymbol{\alpha} & F'_l \end{bmatrix},$$

where $\boldsymbol{\alpha} \in \mathbb{F}^{R\times 1}$ is the first column of F_l. Let us define $H = \begin{bmatrix} \boldsymbol{\alpha} & F'_l \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_i(p_l) \\ A'_l \end{bmatrix}$, which is the matrix $F_l A_l$ resulting from the previous step and is nonsingular by the inductive assumption.

Theorem 41 *Let $\boldsymbol{\gamma}_l = H^{-1}\boldsymbol{\alpha}$. In Case 1, if $\boldsymbol{x}_i(q)$ is chosen such that*

$$1 + (\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l))\boldsymbol{\gamma}_l \neq 0, \tag{5.1}$$

then after the updating process $F_l A_l$ remains nonsingular.

Proof Using standard matrix calculus, $F_l A_l$ after the updating process can be written as

$$\begin{aligned} F_l A_l &= \begin{bmatrix} \boldsymbol{\alpha} & F'_l \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_i(q) \\ A'_l \end{bmatrix} \\ &= \boldsymbol{\alpha}\boldsymbol{x}_i(q) + F'_l A'_l. \end{aligned}$$

We can write

$$F'_l A'_l = H - \boldsymbol{\alpha}\boldsymbol{y}_i(p_l)$$

and therefore

$$F_l A_l = H + \boldsymbol{\alpha}(\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l)).$$

For the moment suppose that $F_l A_l$ is singular. This means that there exist a non–zero column vector $\boldsymbol{\beta} \in \mathbb{F}^{R\times 1}$ with $F_l A_l \boldsymbol{\beta} = \mathbf{0}$. This implies that

$$H\boldsymbol{\beta} + \boldsymbol{\alpha}(\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l))\boldsymbol{\beta} = \mathbf{0}. \tag{5.2}$$

Since $\boldsymbol{\alpha} = H\boldsymbol{\gamma}_l$, (5.2) can be rewritten as

$$H\boldsymbol{\beta} + H\boldsymbol{\gamma}_l(\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l))\boldsymbol{\beta} = H(\boldsymbol{\beta} + \boldsymbol{\gamma}_l(\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l))\boldsymbol{\beta}) = \mathbf{0}.$$

Since H is nonsingular, the identity holds if and only if

$$\boldsymbol{\beta} + \boldsymbol{\gamma}_l(\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l))\boldsymbol{\beta} = \mathbf{0}.$$

If we premultiply the vectors from both sides of the preceding vector equation by $\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l)$, we find that

$$\begin{aligned}&(\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l))\boldsymbol{\beta} + (\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l))\boldsymbol{\gamma}_l(\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l))\boldsymbol{\beta}\\ &= (1 + (\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l))\boldsymbol{\gamma}_l)(\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l))\boldsymbol{\beta} = 0.\end{aligned}$$

The expression above is product of two numbers $(1 + (\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l))\boldsymbol{\gamma}_l)$ and $(\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l))\boldsymbol{\beta}$. We argue that $(\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l))\boldsymbol{\beta}$ is not zero. Observe that if this number was zero, then equation (5.2) and the nonsingularity of H would imply that $H\boldsymbol{\beta}$ and $\boldsymbol{\beta}$ are both zero vectors, contradicting our assumption that $\boldsymbol{\beta}$ is a non–zero vector. Therefore

$$1 + (\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l))\boldsymbol{\gamma}_l = 0.$$

This argument implies that for F_lA_l to be nonsingular it is sufficient to have the following inequality:

$$1 + (\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l))\boldsymbol{\gamma}_l \neq 0.$$

□

2. Analysis of case 2

The analysis is very similar to Case 1. Without loss of generality assume that the new row is added to the bottom of A_l and the new column is added to the right of F_l. After the update A_l is of the form

$$A_l = \begin{bmatrix} A'_l \\ \boldsymbol{x}_i(q) \end{bmatrix}.$$

Here A'_l represents matrix A_l before the update. Also matrix F_l after the update is of the form

$$F_l = \begin{bmatrix} F'_l & \boldsymbol{\alpha} \end{bmatrix},$$

where $\boldsymbol{\alpha} \in \mathbb{F}^{R\times 1}$ is the new column added to F'_l, which is the matrix F_l before the update. Our inductive assumption implies that $H = F'_l A'_l$ is nonsingular. We again let $\boldsymbol{\gamma}_l = H^{-1}\boldsymbol{\alpha}$.

Theorem 42 *In Case 2, if $\boldsymbol{x}_i(q)$ is chosen such that*

$$1 + \boldsymbol{x}_i(q)\boldsymbol{\gamma}_l \neq 0, \tag{5.3}$$

then after the updating process F_lA_l remains nonsingular.

Proof We can write

$$F_lA_l = H + \boldsymbol{\alpha}\boldsymbol{x}_i(q).$$

F_lA_l is singular if there exists a non–zero vector $\boldsymbol{\beta}$ such that

$$F_lA_l\boldsymbol{\beta} = H\boldsymbol{\beta} + \boldsymbol{\alpha}\boldsymbol{x}_i(q)\boldsymbol{\beta} = \mathbf{0}.$$

Since $\boldsymbol{\alpha} = H\boldsymbol{\gamma}_l$, the preceding equations can be rewritten

$$F_l A_l \boldsymbol{\beta} = H\boldsymbol{\beta} + H\boldsymbol{\gamma}_l \boldsymbol{x}_i(q)\boldsymbol{\beta} = H(\boldsymbol{\beta} + \boldsymbol{\gamma}_l \boldsymbol{x}_i(q)\boldsymbol{\beta}) = \mathbf{0}. \tag{5.4}$$

Since H is nonsingular, (5.4) implies

$$\boldsymbol{\beta} + \boldsymbol{\gamma}_l \boldsymbol{x}_i(q)\boldsymbol{\beta} = \mathbf{0}.$$

If we premultiply both sides of the preceding equation by $\boldsymbol{x}_i(q)$ we obtain

$$\boldsymbol{x}_i(q)\boldsymbol{\beta} + \boldsymbol{x}_i(q)\boldsymbol{\gamma}_l \boldsymbol{x}_i(q)\boldsymbol{\beta} = \boldsymbol{x}_i(q)\boldsymbol{\beta}(1 + \boldsymbol{x}_i(q)\boldsymbol{\gamma}_l) = 0.$$

The previous equality holds if either $\boldsymbol{x}_i(q)\boldsymbol{\beta} = 0$ or if $1 + \boldsymbol{x}_i(q)\boldsymbol{\gamma}_l = 0$. If $\boldsymbol{x}_i(q)\boldsymbol{\beta} = 0$ then by (5.4) $H\boldsymbol{\beta} = \mathbf{0}$, which together with the invertibility of H implies that $\boldsymbol{\beta} = \mathbf{0}$. But $\boldsymbol{\beta} \neq \mathbf{0}$ by assumption. Therefore if $F_l A_l$ is a singular matrix, we have

$$1 + \boldsymbol{x}_i(q)\boldsymbol{\gamma}_l = 0.$$

The preceding argument implies that $F_l A_l$ is nonsingular if

$$1 + \boldsymbol{x}_i(q)\boldsymbol{\gamma}_l \neq 0.$$

$\square$

3. A randomized algorithm and the existence of a solution

Let us summarize the analysis up to this point. Here we fix a node $q \in Q_i$ and concentrate on the design of the coding vector $\boldsymbol{x}_i(q)$. The coding vector $\boldsymbol{x}_i(q)$ can be assigned in a way that meets our requirements if

$$\tau \triangleq \prod_{t_l : q \in Q_i^l[j], j \in \{1, \cdots, m_i\}} (1 + (\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l))\boldsymbol{\gamma}_l) \prod_{t_l : q \notin Q_i^l[j], j \in \{1, \cdots, m_i\}} (1 + \boldsymbol{x}_i(q)\boldsymbol{\gamma}_l) \neq 0.$$

In the preceding equation t_l is restricted to the destinations for which $K_l \geq i+1$, and the vectors $\boldsymbol{\gamma}_l$ and $\boldsymbol{y}_i(p_l)$ are specified in the analyses of Cases 1 and 2.

One other constraint is that $\boldsymbol{x}_i(q)$ can only be a linear combination of the vectors $\{\boldsymbol{y}_i(p) : p \in P_i[j]\}$. Let us set $\boldsymbol{x}_i(q) = \sum_{p \in P_i[j]} \theta_p \boldsymbol{y}_i(p)$. The following theorem through a probabilistic argument guarantees the existence of a valid solution for the coefficients θ_p when $|\mathbb{F}| > g$:

Theorem 43 *If $\boldsymbol{x}_i(q) = \sum_{p \in P_i[j]} \theta_p \boldsymbol{y}_i(p)$ and the coefficients $\{\theta_p : p \in P_i[j]\}$ are chosen from the uniform distribution over the field $\mathbb{F}$ and $|\mathbb{F}| > g$, then with positive probability $\tau \neq 0$.*

Proof For each destination t_l with $K_l \geq i+1$ define ϕ_l as the event that the corresponding term in the defining product of τ is zero. Then we have

$$\Pr(\tau = 0) = \Pr(\bigvee_{t_l : K_l \geq i+1} \phi_l) \leq \sum_{t_l : K_l \geq i+1} \Pr(\phi_l).$$

Now consider a destination t_l with $K_l \geq i+1$. If $q \in Q_i^l[j]$ and $p_l \in P_i^l[j]$ is matched with q, we need to have $1 + (\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l))\boldsymbol{\gamma}_l \neq 0$. There exist $\omega_0 \in \mathbb{F}, \omega_p \in \mathbb{F}, p \in P_i[j]$, which are determined by $\boldsymbol{y}_i(p), p \in P_i[j]$, and $\boldsymbol{\gamma}_l$ and satisfy

$$1 + (\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l))\boldsymbol{\gamma}_l = \omega_0 + \sum_{p \in P_i[j]} \omega_p \theta_p$$

There are two cases to consider. First, if $\omega_p = 0$ for all $p \in P_i[j]$, then $\omega_0 + \sum_{p \in P_i[j]} \omega_p \theta_p = \omega_0$ is a constant independent of $\theta_p, p \in P_i[j]$. Furthermore by setting $\theta_{p_l} = 1$ and $\theta_p = 0$ for $p \in P_i[j]$ and $p \neq p_l$ so that $\boldsymbol{x}_i(q) = \boldsymbol{y}_i(p_l)$, we find that

$$\omega_0 = 1 + (\boldsymbol{x}_i(q) - \boldsymbol{y}_i(p_l))\boldsymbol{\gamma}_l = 1.$$

Therefore in this case $\Pr(\phi_l) = 0$. Next if there exists some $p \in P_i[j]$ for which $\omega_p \neq 0$ then $\omega_0 + \sum_{p \in P_i[j]} \omega_p \theta_p$ depends on $\theta_p, p \in P_i[j]$. Since $\theta_p, p \in P_i[j]$, are uniformly distributed random variables over $\mathbb{F}$, $\omega_0 + \sum_{p \in P_i[j]} \omega_p \theta_p$ is likewise uniformly distributed over $\mathbb{F}$. In this case $\Pr(\phi_l) = \frac{1}{|\mathbb{F}|}$.

Next suppose that $q \notin Q_i^l[j]$. Here we need to have $1 + \boldsymbol{x}_i(q)\boldsymbol{\gamma}_l \neq 0$. Following the preceding argument, there exist $\omega_0 \in \mathbb{F}, \omega_p \in \mathbb{F}, p \in P_i[j]$, which are determined by $\boldsymbol{y}_i(p), p \in P_i[j]$, and $\boldsymbol{\gamma}_l$ and satisfy

$$1 + \boldsymbol{x}_i(q)\boldsymbol{\gamma}_l = \omega_0 + \sum_{p \in P_i[j]} \omega_p \theta_p.$$

If $\omega_p = 0$ for all $p \in P_i[j]$, then $\omega_0 + \sum_{p \in P_i[j]} \omega_p \theta_p = \omega_0$ is a constant independent of $\theta_p, p \in P_i[j]$. By setting $\theta_p = 0$ for all $p \in P_i[j]$ so that $\boldsymbol{x}_i(q) = \boldsymbol{0}$, we obtain $\omega_0 = 1 + \boldsymbol{x}_i(q)\boldsymbol{\gamma}_l = 1$, and $\Pr(\phi_l) = 0$. If there is some $p \in P_i[j]$ for which $\omega_p \neq 0$, then an analogous argument to our earlier one implies that $\Pr(\phi_l) = \frac{1}{|\mathbb{F}|}$.

As a result, for any destination t_l with $K_l \geq i + 1$, we have $\Pr(\phi_l) \leq \frac{1}{|\mathbb{F}|}$. Therefore

$$\Pr(\tau = 0) \leq \sum_{t_l : K_l \geq i+1} \Pr(\phi_l) \leq \frac{g}{|\mathbb{F}|}.$$

Since we are interested in the event that $\tau \neq 0$, we have

$$\Pr(\tau \neq 0) \geq 1 - \frac{g}{|\mathbb{F}|}.$$

Therefore if $|\mathbb{F}| > g$, then $\Pr(\tau \neq 0) > 0$ and there is at least one valid solution for $\boldsymbol{x}_i(q)$. This also yields a randomized algorithm with probability of success of at least $1 - \frac{g}{|\mathbb{F}|}$. If we take the size of the field to be $|\mathbb{F}| \geq 2g$ then the probability of success will be at least $1 - \frac{g}{2g} = \frac{1}{2}$.

$\square$

4. A deterministic polynomial time algorithm

We next describe a deterministic algorithm with polynomial running time for finding the vectors $\boldsymbol{x}_i(q), q \in Q_i[j]$. For each $q \in Q_i$ we seek a vector $\boldsymbol{u} = \boldsymbol{x}_i(q)$ which is a linear combination of the vectors in $\{\boldsymbol{y}_i(p) : p \in P_i[j]\}$ such that for any destination t_l with $K_l \geq i+1$, if $q \in Q_i^l[j]$ and $p_l \in P_i^l[j]$ is matched with q, then $1 + (\boldsymbol{u} - \boldsymbol{y}_i(p_l))\boldsymbol{\gamma}_l \neq 0$, and if $q \notin Q_i^l[j]$ then $1 + \boldsymbol{u}\boldsymbol{\gamma}_l \neq 0$.

Define the subset of indices of destinations W as

$$W = \left\{ l \in \{1, \cdots, g\} : K_l \geq i+1, q \in Q_i^l[j] \text{ for some } j \in \{1, \cdots, m_i\}, \boldsymbol{y}_i(p_l)\boldsymbol{\gamma}_l \neq 0 \right\}.$$

We can express the conditions that $\boldsymbol{u}$ needs to satisfy as

$$\begin{cases} 1 + (\boldsymbol{u} - \boldsymbol{y}_i(p_l))\boldsymbol{\gamma}_l \neq 0, & \text{for } l \in W, \\ 1 + \boldsymbol{u}\boldsymbol{\gamma}_l \neq 0, & \text{for } l \notin W \text{ and } K_l \geq i+1. \end{cases} \tag{5.5}$$

In the following two steps we are going to find a vector $\boldsymbol{u}$ that is a linear combination of the vectors $\{\boldsymbol{y}_i(p) : p \in P_i[j]\}$ and satisfies the above conditions:

Step 1: Find a vector $\boldsymbol{w}$ which is a linear combination of the vectors $\{\boldsymbol{y}_i(p_l) : l \in W\}$ and satisfies $\boldsymbol{w}\boldsymbol{\gamma}_l \neq 0$ for every $l \in W$.

Step 2: Set $\boldsymbol{u} = \sigma\boldsymbol{w}$ for some $\sigma \in \mathbb{F}$ to meet the constraints of (5.5).

To demonstrate the validity of this procedure we begin by using a result from [37] to show that Step 1 is feasible if $|\mathbb{F}| \geq g$:

Lemma 44 *([37, Lemma 8]) Let $n \leq |\mathbb{F}|$. Let $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n \in \mathbb{F}^{1\times R}$ and $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n \in \mathbb{F}^{R\times 1}$ with $\boldsymbol{a}_i\boldsymbol{b}_i \neq 0, i \in \{1, \cdots, n\}$. There exists a linear combination $\boldsymbol{c}$ of $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n$ such that $\boldsymbol{c}\boldsymbol{b}_i \neq 0, i \in \{1, \cdots, n\}$. Such a vector $\boldsymbol{c}$ can be found in time $O(n^2R)$.*

By applying the preceding lemma, if $g \leq |\mathbb{F}|$ then $|W| \leq |\mathbb{F}|$, and we can find a vector $\boldsymbol{w} \in \mathbb{F}^{1\times R}$ such that $\boldsymbol{w}$ is a linear combination of the vectors in $\{\boldsymbol{y}_i(p_l) : l \in W\}$ and for every $l \in W$, we have that $\boldsymbol{w}\boldsymbol{\gamma}_l \neq 0$. Next given $g < |\mathbb{F}|$ we prove the feasibility of Step 2:

Lemma 45 *Given vector $\boldsymbol{w}$ from Step 1, there exists $\sigma \in \mathbb{F}$ such that $\boldsymbol{u} = \sigma\boldsymbol{w}$ satisfies (5.5).*

Proof For $l \in W$, we require $1 + (\sigma\boldsymbol{w} - \boldsymbol{y}_i(p_l))\boldsymbol{\gamma}_l \neq 0$. Therefore

$$\sigma \neq \frac{\boldsymbol{y}_i(p_l)\boldsymbol{\gamma}_l - 1}{\boldsymbol{w}\boldsymbol{\gamma}_l}. \tag{5.6}$$

For $l \notin W$ and $K_l \geq i+1$ we need to have $1 + \sigma\boldsymbol{w}\boldsymbol{\gamma}_l \neq 0$. If $\boldsymbol{w}\boldsymbol{\gamma}_l = 0$ then this condition holds for all values of σ. Otherwise we need to have

$$\sigma \neq \frac{-1}{\boldsymbol{w}\boldsymbol{\gamma}_l}. \tag{5.7}$$

There are at most g constraints of the form (5.6) and (5.7) on σ. Therefore if the size of field $\mathbb{F}$ is greater than the number of destinations g, this deterministic approach will find at least one σ that is not in the discriminating set by considering at most g elements of $\mathbb{F}$.

$\square$

Theorem 46 *The overall complexity of constructing the multicast coding scheme by using the deterministic algorithm in Steps 1 and 2 is*

$$O(gmMr(mrR + R^3 + gR + (mr)^2 \log mr)).$$

Proof If we are given the set of vectors $\boldsymbol{\gamma}_l$, then it takes $O(gR)$ steps to form the set W. Given set W and $g \leq |\mathbb{F}|$, Lemma 44 implies that we can find a vector $\boldsymbol{w} \in \mathbb{F}^{1\times R}$ in time $O(g^2R)$. By adding the time $O(gR)$ needed to produce set W, we need a total time of $O(g^2R + gR) = O(g^2R)$ to find vector $\boldsymbol{w}$. The complexity of finding an appropriate σ by forming the discriminating set in Lemma 45 is $O(g)$ and therefore the complexity of finding vector $\boldsymbol{u}$ is $O(g + g^2R) = O(g^2R)$. To find the overall complexity of finding the vector

$\boldsymbol{x}_i(q)$, we need to evaluate the complexity of finding vector $\boldsymbol{\gamma}_l$ for every $l \in \{1, \cdots, g\}$ with $K_l \geq i+1$. From the analyses of Cases 1 and 2, $\boldsymbol{\gamma}_l = H^{-1}\boldsymbol{\alpha}$, where matrix H is $F_l A_l$ from the previous step of the algorithm. Since matrix F_l has size $R \times L$ and matrix A_l has size $L \times R$ for some $R \leq L \leq |Q_i|$, computing H needs $O(R|Q_i|)$ operations. Evaluating H^{-1} also needs $O(R^3)$ steps and so there are a total of $O(R|Q_i| + R^3)$ operations for evaluating $\boldsymbol{\gamma}_l$. Since there are at most g different $l \in \{1, \cdots, g\}$ with $K_l \geq i+1$, we will have $O(gR|Q_i|+gR^3)$ as the total complexity of evaluating different values of $\boldsymbol{\gamma}_l$ for any specific $q \in Q_i$. Therefore the total complexity of evaluating $\boldsymbol{x}_i(q)$ will be $O(gR|Q_i|+gR^3+g^2R)$. Let us assume that the number of nodes m_i at each layer $i \in \{1, \cdots, M\}$ is at most m. Furthermore assume that the size of transmitted and received signals at each node is at most r. Therefore the total complexity of evaluating each $\boldsymbol{x}_i(q)$ will be $O(gRmr + gR^3 + g^2R)$. Since there are at most mMr different $\boldsymbol{x}_i(q)$ to be evaluated, if we assume that the unicast flows from source to each destination is provided, the total complexity of our algorithm is $O(gRm^2Mr^2 + gR^3mMr + g^2RmMr) = O(gRmMr(mr + R^2 + g))$. The complexity of computing a unicast flow to a destination by the algorithm given in [33] is $O(M(mr)^3 \log mr)$. Since we have g destinations, the total complexity of computing the unicast flows will be $O(gM(mr)^3 \log mr)$. If we add this running time to the running time of our algorithm, the total running time will be $O(gmMr(mrR+R^3+gR+(mr)^2 \log mr))$.

$\square$

We can compare the running time of our algorithm to the running time of the algoithm given in [18] which is $O(g(r^2mM+R)^3 \log(r^2mM+R)+r^2mM(r^2mM+R)^2+(g \log gRM)^3)$ and see that our proposed algorithm is considerably faster.

5. Number of network uses to achieve capacity

We have shown that it is sufficient for the size of the field of operation $\mathbb{F}$ of the LDRN to be greater than g to guarantee the existence of a multicast coding solution. In general however, the network operates over some fixed field which is usually $\mathbb{F}_p$ for some prime number p. In order to achieve a greater field size, we will use multiple rounds of the network. Here we will argue that if we use the network for k rounds, it is equivalent to an LDRN with field of operation $\mathbb{F} = \mathbb{F}_p^k$. This implies that in order to have a field size at least $g+1$, it is sufficient to use the network for $k = \lceil \log_p(g+1) \rceil$ rounds.

Suppose that the network is used for k rounds and we use the superscript $0 \leq t \leq K-1$ to denote the time index that a vector is received or sent. For each $i \in \{1, \cdots, M-1\}$ we have

$$\mathbf{y}_{i+1}^t = G_i \mathbf{x}_i^t, \qquad 0 \leq t \leq K-1.$$

Observe we can use a dummy variable D as the unit delay operator and represent the preceding k equations as a single equation

$$\sum_{t=0}^{k-1} \mathbf{y}_{i+1}^t D^t = G_i \sum_{t=0}^{k-1} \mathbf{x}_i^t D^t.$$

Next, notice that $\sum_{t=0}^{k-1} \mathbf{y}_{i+1}^t D^t$ and $\sum_{t=0}^{k-1} \mathbf{x}_i^t D^t$ can be regarded as new vectors in the extension field $\mathbb{F}_p^k$ and we can assume that the network is operating in the extension field $\mathbb{F}_p^k$. Since the transfer matrix between the layers i and $i+1$ is still G_i and has not changed in the new field, the existence of the unicast flow over the original field implies the existence of flow over the extended field. Therefore our analysis is valid over any field $\mathbb{F}_p^k$.

CHAPTER VI

CONCLUSIONS

In this dissertation we considered different topologies of wired and wireless networks and proposed network coding schemes with low-complexity encoding and decoding at the intermediate nodes of the network. We used a variety of tools and ideas from combinatorics, graph theory, linear algebra, and submodular optimization theory to study the optimality of our coding schemes and to offer low-complexity algorithms to design the coding schemes.

In the first part we gave a new characterization of the routing capacity region of wired networks. We also offered an elimination technique that reduces the initial characterization from an infinite set of inequalities into a finite and minimal set of inequalities. Our technique led to a simple characterization of the routing capacity region of undirected ring networks in two different scenarios. In a related work [39], we study the new characterization from a computational complexity point of view. For future work it is interesting to study the design of efficient algorithms for computing the set of minimal inequalities. Another direction of research is to quantify the approximation of the routing capacity region with a subset of the inequalities in the minimal set. For the study of undirected ring networks we have also devised new information theoretic bounds that are stronger than the known cut set bounds. We believe that such bounds can be applied to a larger class of networks to obtain tighter outer bounds on the network coding capacity region of networks.

In the second part we studied the network coding capacity region of node-constrained line and star networks. As another direction of research one can study more complex network structures such as tree networks or lattice networks. We believe that our coding constructions can be extended to such structures to obtain optimal or near-optimal transmission schemes.

In the last part we studied the linear deterministic relay model of wireless networks from a combinatorial point of view. In particular, we applied tools from matrix theory, matroid theory, and submodular optimization theory to devise polynomial-time algorithms for a unicast connection or a multicast connection in the network. Our algorithm is similar to a routing scheme for wired networks in the case of a unicast connection and is similar to the algorithm of Jaggi et al. [37] in the case of a multicast connection. We also offered polynomial-time algorithms for finding the transmission schemes. For future work it would be interesting to extend our techniques to the multiple unicast connections. If this is possible it may lead to new results for the transversal theory of matrices. Another direction of study would be to investigate if our coding techniques could be used to construct codes for Gaussian relay networks. The recent paper [69] offers initial results along these lines.

REFERENCES

[1] B. W. Abeysundara, and A. E. Kamal, “High-speed local area networks and their performance: a survey,” *ACM Computing Surveys (CSUR)*, vol. 23, pp. 221–264, June 1991.

[2] 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.

[3] M. S. Alfiad, D. van den Borne, T. Wuth, M. Kuschnerov, B. Lankl, C. Weiske, E. de Man, A. Napoli, H. de Waardt, “111 Gb/s POLMUX-RZ-DQPSK Transmission over 1140 km of SSMF with 10.7 Gb/s NRZ-OOK Neighbours,” in Proc. *ECOC 2008,* Brussels, Belgium, 21-25 September 2008, pp. 117–118.

[4] A. Amaudruz, and C. Fragouli, “Combinatorial algorithms for wireless information flow,” in Proc. *of ACM-SIAM Symposium on Discrete Algorithms (SODA09)*, New York, January 2009, pp. 555–564.

[5] M. R. Aref, *Information Flow in Relay Networks*, Ph.D. dissertation, Stanford Univ., Palo Alto, CA, 1980.

[6] K. A. Atkinson, *An Introduction to Numerical Analysis*, 2nd ed., New York, John Wiley & Sons, 1989.

[7] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, “Wireless network information flow: A deterministic approach,” *IEEE Trans. Inform. Theory*, vol. 57, pp. 1872–1905, April 2011.

[8] S. Avestimehr, S. Diggavi, and D. Tse, “Approximate capacity of Gaussian relay networks,” in Proc. *IEEE Symposium on Information Theory (ISIT08),* Toronto, June 2008, pp. 474–478.

[9] M. Bakshi, M. Effros, W. Gu, and R. Koetter "On network coding of independent and dependent sources in line networks," in Proc. *2007 IEEE Int. Symp. Inform. Theory*, Nice, France, Jun. 24-29 2007, pp. 1096-1100.

[10] Z. Bar-Yossef, Y. Birk, T. S. Jayram, T. Kol, "Index coding with side information," in Proc. *47th Annual IEEE Symposium on Foundations of Computer Science*, Berkeley, CA, Oct. 2006, pp. 197–206.

[11] U. Black, *Optical Networks: Third Generation Transport Systems*, Upper Saddle River, NJ, Prentice Hall PTR, 2002.

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

[13] T. Chan and A. Grant, "Dualities between entropy functions and network codes," *IEEE Trans. Inform. Theory*, vol. 54, pp. 4470–4487, October 2008.

[14] C. Chen and J. Chen, "Nearly optimal one-to-many parallel routing in star networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 8, no. 12, pp.1196–1202, Dec. 1997.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., New York: Wiley, 2006.

[16] A. Dasylva, and R. Srikant, "Optimal WDM schedules for optical star networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 7, no. 3, pp. 446–456, June 1999.

[17] R. Diestel, *Graph Theory*, 3rd ed., Berlin, Springer, 2005.

[18] J. Ebrahimi, and C. Fragouli, “Vector network coding algorithms”, in Proc. *IEEE Symposium on Information Theory (ISIT2010),* Austin, TX, June 2010, pp. 2408–2412.

[19] J. Ebrahimi, and C. Fragouli, “Combinatorial algorithms for wireless information flow,” arXiv:0909.4808v1 [cs.DS], pp. 1–36.

[20] J. Edmonds and E. Giles, “A min-max relation for submodular functions on graphs,” *Annals of Discrete Mathematics,* vol. 1, pp. 185–204, 1977.

[21] S. El Rouayheb, A. Sprintson, C. Georghiades, “On the relation between the index coding and the network coding problems,” in Proc. *IEEE International Symposium on Information Theory*, Toronto, Canada, Jul. 2008, pp. 1823–1827.

[22] R. C. Elsenpeter and T. J. Velte, *Optical Networking: A Beginner’s Guide*, Berkeley, CA: McGraw-Hill/Osborne, 2002.

[23] E. Erez, Y. Xu and E. M. Yeh, “Coding for the deterministic network model,” in Proc. *48th Annual Allerton Conference on Communication, Control, and Computing (Allerton),* Allerton, IL, Sept. 29 2010-Oct. 1 2010, pp. 1534–1541.

[24] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, “Capacity limits of optical fiber networks,” *IEEE/OSA J. Lightwave Techn.*, vol. 28, no. 4, pp. 662-701, February 2010.

[25] R. Etkin, D. Tse, and H. Wang, “Gaussian interference channel capacity to within one bit,” *IEEE Trans. Inform. Theory*, vol. 54, pp. 5534–5562, December 2008.

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

[27] L. R. Ford and D. R. Fulkerson, “Maximal flow through a network,” *Canadian Journal of Mathematics,* vol. 8, pp. 399–404, 1956.

[28] C. Fragouli, J. Widmer and J. Y. Le Boudec, “Efficient broadcasting using network coding,” *IEEE/ACM Transactions on Networking*, vol. 16, no. 2, pp. 450–463, Apr. 2008.

[29] A. Frank, “Finding feasible vectors of Edmonds–Giles polyhedra,” *Journal of Combinatorial Theory, Ser. B,* vol. 36, pp. 221–239, 1984.

[30] 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.

[31] S. Fujishige and X. Zhang, “New algorithms for the intersection problem of submodular systems,” *Japan J. Indust. Appl. Math.,* vol. 9, pp. 369–382, 1992.

[32] R. G. Gallager, *Information Theory and Reliable Communication*, New York, John Wiley & Sons, Inc., 1968.

[33] M. Goemans, S. Iwata and R. Zenklusen, “An algorithmic framework for wireless information flow,” in Proc. *47th Annual Allerton Conference on Communication, Control and Computing,* Monticello, IL, Sept. 29- Oct. 2, 2009, pp. 294–300.

[34] N. J. A. Harvey, R. Kleinberg and A. R. Lehman, “On the capacity of information networks,” *IEEE/ACM Transactions on Networking*, vol. 14, pp. 2345–2364, Jun. 2006.

[35] M. Iri, “On an extension of the maximum-flow minimum-cut theorem to multicommodity flows,” *J. Operations Research Soc. of Japan*, vol. 13, no. 3, pp. 129–135, Jan. 1970.

[36] S. Iwata, S. T. McCormick, M. Shigeno, "A fast cost scaling algorithm for submodular flow," *Information Processing Letters,* vol. 74, no. 3-4, pp. 123–128, May 2000.

[37] S. Jaggi, P. Sanders, P. A. Chou, M. Effros, S. Egner, K. Jain, and L. Tolhuizen, "Polynomial time algorithms for multicast network code construction," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 1973–1982, June 2005.

[38] D. Jungnickel, "Transversaltheorie: Ein überblick," *Bayreuther Math. Schriften*, vol. 21, pp. 122–155, 1986.

[39] A. Kakhbod, S. A. Savari, and S. M. S. Tabatabaei Yazdi, "A study of the routing capacity regions of networks", in Proc. *Asilomar Conference on Signals, Systems and Computers,* Pacific Grove, CA, Oct. 2008, pp. 2226–2230.

[40] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "XORs in the air: practical wireless network coding," *IEEE/ACM Transactions on Networking,* vol. 16, no. 3, pp. 497–510, June 2008.

[41] L. Kerschberg, P. D. Ting, and S. B. Yao, "Query optimization in star computer networks," *ACM Transactions on Database Systems (TODS)*, vol. 7, no. 4, pp. 678–711, December 1982.

[42] M. Khojastepour and A. Keshavarz-Haddad, "Rotational coding achieves multicast capacity of deterministic wireless networks," in *47th Annual Allerton Conference on Communication, Control, and Computing,* Monticello, IL, Sept. 2009, pp. 1330–1337.

[43] M. Kim and M. Médard, "Algebraic network coding approach to deterministic wireless relay networks," in Proc. *Allerton Conference on Communications, Control, and Computing,* Monticello, IL, Sept. 2010, pp. 1518–1525.

[44] R. Koetter and M. Médard, “An algebraic approach to network coding,” *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 782–795, Oct. 2003.

[45] G. Kramer, “Communication on line networks with deterministic or erasure broadcast channels,” in Proc. *2009 IEEE Information Theory Workshop (ITW2009)*, Taormina, Italy, Oct. 2009, pp. 404–405.

[46] G. Kramer, M. Gastpar, and P. Gupta, “Cooperative strategies and capacity theorems for relay networks,” *IEEE Trans. Inform. Theory*, vol. 51, no. 9, pp. 3037-3063, Sept. 2005.

[47] 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, Providence, RI, 2005.

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

[49] 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.

[50] G. Kramer, S. M. S. Tabatabaei Yazdi, and S. A. Savari, “Network coding on line networks with broadcast,” in Proc. *2008 Conference on Information Sciences and Systems,* Princeton, NJ, March 19-21, 2008, pp. 781–784.

[51] J. P. S. Kung, “Bimatroids and invariants,” *Advances in Mathematics,* vol. 30, pp. 238–249, 1978.

[52] N. Laneman, D. Tse, and G. Wornell, “Cooperative diversity in wireless networks: Efficient protocols and outage behavior,” *IEEE Trans. Inform. Theory*, vol. 50, no. 11, November 2004.

[53] S. Y. R. Li, R. W. Yeung, and N. Cai, “Linear network coding,” *IEEE Transactions on Information Theory*, vol. 49, pp. 371-381, 2003.

[54] Z. Li and B. Li, “Network coding: The case of multiple unicast sessions,” in Proc. *42nd Annual Allerton Conf. on Communication, Control and Computing*, Monticello, IL, September, 2004.

[55] Z. Li, B. Li, D. Jiang, and L. C. Lau, “On achieving optimal end-to-end throughput in data networks: Theoretical and empirical studies, ” *ECE Technical Report*, University of Toronto, Toronto, Canada, 2004.

[56] E. Lubetzky, and U. Stav, “Non-linear index coding outperforming the linear optimum,” in Proc. *48th Annual IEEE Symposium on Foundations of Computer Science*, Providence, RI, Oct. 2007, pp. 161–168.

[57] C.L. Lucchesi. *A Minmax Equality for Directed Graphs*, Ph.D. dissertation, University of Waterloo, Ontario, Canada, 1976.

[58] L. Mirsky, *Transversal Theory: An Account of Some Aspects of Combinatorial Mathematics*, Mathematics in Science and Engineering, vol. 75, London, Academic Press, 1971.

[59] S. Mohajer, S. Diggavi, C. Fragouli, and D. Tse, “Transmission techniques for relay–interference networks,” in Proc. *Allerton Conference on Communication, Control, and Computing,* Monticello, IL, September 2008, pp. 467–474.

[60] K. Murota, *Discrete Convex Analysis,* SIAM Monographs on Discrete Mathematics and Applications, Philadelphia, PA, Society for Industrial and Applied Mathematics (SIAM), 2003.

[61] A. Natarajan, M. Motani, B. de Silva, K. Yap, and K. C. Chua, "Network architectures and support: Investigating network architectures for body sensor networks," in Proc. *1st ACM SIGMOBILE International Workshop on Systems and Networking Support for Healthcare and Assisted Living Environments*, San Juan, Puerto Rico, June 2007, pp. 19–24.

[62] U. Niesen, C. Fragouli and D. Tuninetti "On cascaded channels with finite complexity processing at intermediate nodes," in Proc. *43th Annual Allerton Conf. on Communication, Control and Computing*, Monticello, IL, September, 2005.

[63] U. Niesen, C. Fragouli and D. Tuninetti "On the capacity of line networks," *IEEE Trans. Inform. Theory*, vol. 53, no. 11, pp. 4039–4058, Nov. 2007.

[64] H. Okamura and P. D. Seymour, "Multicommodity flows in planar graphs," *J. Combin. Theory*, Series B, vol. 31, pp. 75–81, 1981.

[65] K. Onaga and O. Kakusho, "On feasibility conditions of multicommodity flows in networks," *IEEE Trans. Circuit Theory*, vol. CT-18, no. 4, pp. 425-429, July 1971.

[66] P. Pakzad, C. Fragouli and A. Shokrollahi, "Coding schemes for line networks," in Proc. *2005 IEEE Int. Symp. Inform. Theory*, Adelaide, Australia, Sept. 4-9 2005, pp. 1853–1857.

[67] H. Perfect, "A generalization of Rado's Theorem on independent transversals," in Journal *Mathematical Proceedings of the Cambridge Philosophical Society,* vol. 66, pp. 513–515, 1969.

[68] R. Rado, "A theorem on independence relations," *Quarterly Journal of Mathematics*, Oxford, vol. 13, pp. 83–89, 1942.

[69] A. Raja and P. Viswanath, "Compress-and-forward scheme for a relay network: Approximate optimality and connection to algebraic flows," submitted to *IEEE Transactions on Information Theory,* January 2011.

[70] V. Ramachandran, "A minmax arc theorem for reducible flow graphs," *SIAM J. Discrete Math*, vol. 3, pp. 554–560, 1990.

[71] N. Ratnakar and G. Kramer, "The multicast capacity of deterministic relay networks with no interference," *IEEE Transactions on Information Theory,* vol. 52, no. 6, pp. 2425-2432, 2006.

[72] S. A. Savari and G. Kramer, "A study of ring networks," Invited presentation, in *2006 IEEE Communication Theory Workshop,* Dorado, Puerto Rico, May 24, 2006.

[73] S. A. Savari and G. Kramer, "The multimessage unicast capacity region for bidirectional ring networks," in Proc. *2006 IEEE Int. Symp. Inform. Theory*, Seattle, WA, July 2006, pp. 763–767.

[74] M. Scheutzow, M. Reisslein, M. Maier, and P. Seeling, "Multicast capacity of packet-switched ring WDM networks," *IEEE Trans. Inform. Theory*, vol. 54, no. 2, pp. 623-644, Feb. 2008.

[75] A. Schrijver, "Matroids and linking systems," *Journal of Combinatorial Theory, Series B,* vol. 26, pp. 349–369, 1979.

[76] A. Schrijver, *Theory of Linear and Integer Programming*, New York: Wiley, 1998.

[77] A. Schrijver, “A combinatorial algorithm minimizing submodular functions in strongly polynomial time,” *Journal of Combinatorial Theory, Series B*, vol. 80, pp. 346-355, 2000.

[78] A. Schrijver, *Combinatorial Optimization*, Berlin, Springer, 2003.

[79] A. Schrijver, P. Seymour, and P. Winkler, “The ring loading problem,” *SIAM Review*, vol. 41, no. 4, pp. 777-791, 1999.

[80] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July and October, 1948.

[81] J.L. Szwarcfiter, “On minimum cuts of cycles and maximum disjoint cycles,” In *Graphs and Algorithms (Boulder, CO, 1987)*, Contemp. Math., pp. 153-166. Amer. Math. Soc., Providence, RI, 1989.

[82] S. M. S. Tabatabaei Yazdi and S. A. Savari, “A combinatorial study of linear deterministic relay networks,” *arXiv:0904.2401v1 [cs.IT],* pp. 1–24, April 2009.

[83] S. M. S. Tabatabaei Yazdi and S. A. Savari, “A combinatorial result on block matrices,” *arXiv:0908.0042v1 [math.CO],* pp. 1–4, August 2009.

[84] S. M. S. Tabatabaei Yazdi and S. A. Savari, “A combinatorial study of linear deterministic relay networks,” in Proc. *IEEE Information Theory Workshop on Information Theory (ITW 2010),* Cairo, Egypt, January 6-8, 2010, pp. 1–5.

[85] S. M. S. Tabatabaei Yazdi and S. A. Savari, “A max-flow/min-cut algorithm for a class of wireless networks,” in *ACM-SIAM Symposium on Discrete Algorithms (SODA 2010),* Austin, Texas, January 17- 19, 2010, pp. 1209–1226.

[86] S. M. S. Tabatabaei Yazdi, S. A. Savari, K. Carlson(Talaska), and G. Kramer, “The capacity region of a collection of multicast sessions in an undirected ring network,” in

Proc. *Int. Workshop on Network Coding (IWNC) as part of the 36th Int. Conf. Parallel Proc. (ICPP),* Xian, China, Sept. 10-14, 2007, pp. 40–47.

[87] S. M. S. Tabatabaei Yazdi, S. A. Savari, F. Farnoud, and G. Kramer, "A multimessage capacity region for undirected ring networks," in Proc. *2007 IEEE Int. Symp. Inform. Theory*, Nice, France, June 24-29, 2007, pp. 1091–1095.

[88] S. M. Tabatabaei Yazdi, S. A. Savari, and G. Kramer, "Network coding capacity region of line networks with node and edge constraints," in *45th Annual Allerton Conf. on Commun., Control and Comp.*, Monticello, IL, Sept. 26-28, 2007, pp. 949-958.

[89] S. M. S. Tabatabaei Yazdi, S. A. Savari, and G. Kramer, "Network coding in star networks," in Proc. *2008 IEEE International Symposium on Information Theory,* Toronto, Canada, July 6-11, 2008, pp. 325-329.

[90] R. Vachani, A. Shulman, P. Kubat, and J. Ward, "Multicommodity flows in ring networks," *INFORMS J. Computing*, vol. 8, no. 3, pp. 235-242, Summer 1996.

[91] C. Wang, E. L. Floyd and M. L Soffa, "Feedback vertex sets and cyclically reducible graphs," *Journal of the ACM (JACM)* vol. 32 , no. 2, pp. 296 - 313, April 1985.

[92] J. G. Wardrop, "Some theoretical aspects of road traffic research," in Proc. *Institute of Civil Engineers,* prt. II, vol.1, pp. 325-378, 1952.

[93] D. J. A. Welsh, *Matroid Theory*, London, Academic Press, 1976.

[94] J. Wilson, *Sensor Technology Handbook,* Burlington, MA, Elsevier, December 2004.

[95] Y. Wu, P. A. Chou, and S.-Y Kung, "Information exchange in wireless networks with network coding and physical-layer broadcast," in *Conference on Information Sciences and Systems,* The Johns Hopkins University, Baltimore, MD, March 16–18, 2005.

[96] Y. Wu, P. A. Chou, and S.-Y. Kung, “Minimum-energy multicast in mobile ad hoc networks using network coding,” *IEEE Transactions on Communications,* vol. 53, no. 11, pp. 1906–1918, Nov. 2005.

[97] Y. Wu, P. Chou, Q. Zhang, K. Jain, W. Zhu, and S. Kung, “Network planning in wireless ad hoc networks: A cross-layer approach,” *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 136–150, Jan. 2005.

[98] S. Yao, *Polarization in Fiber Systems: Squeezing Out More Bandwidth*, Pittsfield, MA, Laurin Publishing, 2003.

[99] J. W. Yeh and W. Siegmund, “Local network architectures,” in Proc. *3rd ACM SIGSMALL Symposium and the First SIGPC Symposium on Small Systems SIGSMALL,* New York, NY, September 1980, pp. 10–14.

[100] G. M. Ziegler, *Lectures on Polytopes*, New York: Springer-Verlag, 1995.

APPENDIX A

Here we prove Theorem B.2 from Chapter II. A rate vector $R = (R_1, \cdots, R_{|S|})$ is routable in a network $G(V, E)$ if and only if the following linear program has a solution for $\{r_T\}$:

1. $\sum_{T \in \mathcal{T}_s} r_T \geq R_s$ for every $s \in S$
2. $\sum_{s \in S} \sum_{T \in \mathcal{T}_s : e \in T} r_T \leq C_e$ for every $e \in E$.
3. $0 \leq r_T$, for every $T \in \mathcal{T}_s$ and every $s \in S$.

Notice that the first set of inequalities can be changed to equalities, but it is equivalent and more convenient here to work with inequalities.

Label the elements of E and S from 1 to $|E|$ and from 1 to $|S|$ respectively. Also label the elements of $\mathcal{T}_s$, by $T_s^1, \cdots, T_s^{|\mathcal{T}_s|}$. For $e \in \{1, \cdots, |E|\},\ s \in \{1, \cdots, |S|\},\ T_s^j \in \{T_s^1, \cdots, T_s^{|\mathcal{T}_s|}\}$, let

$$\delta^s_{e, T_s^j} = \begin{cases} 1, & e \in E(T_s^j) \\ 0, & \text{otherwise.} \end{cases}$$

Define

$$\mathbf{r} = \left(r_{T_1^1}, \cdots, r_{T_1^{|\mathcal{T}_1|}}, \cdots, r_{T_{|S|}^1}, \cdots, r_{T_{|S|}^{|\mathcal{T}_{|S|}|}} \right)^T \tag{A.1}$$

$$\mathbf{c} = \left(C_1, \cdots, C_{|E|}, -R_1, \cdots, -R_{|S|}, 0, \cdots, 0 \right)^T \tag{A.2}$$

and matrix **M** as follows:

$$
\mathbf{M} = \left(\begin{array}{ccc|ccc|ccc}
\delta^1_{1,T_1^1} & \cdots & \delta^1_{1,T_1^{|\mathcal{T}_1|}} & & & & \delta^{|S|}_{1,T^1_{|S|}} & \cdots & \delta^{|S|}_{1,T_{|S|}^{|\mathcal{T}_{|S|}|}} \\
\vdots & \ddots & \vdots & & \cdots & & \vdots & \ddots & \vdots \\
\delta^1_{|E|,T_1^1} & \cdots & \delta^1_{|E|,T_1^{|\mathcal{T}_1|}} & & & & \delta^{|S|}_{|E|,T^1_{|S|}} & \cdots & \delta^{|S|}_{|E|,T_{|S|}^{|\mathcal{T}_{|S|}|}} \\
\hline
-1 & \cdots & -1 & & & & 0 & \cdots & 0 \\
0 & \cdots & 0 & & \cdots & & \vdots & \ddots & \vdots \\
\vdots & \ddots & \vdots & & & & 0 & \cdots & 0 \\
0 & \cdots & 0 & & & & -1 & \cdots & -1 \\
\hline
-1 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
 & \ddots & & & & & & & \\
 & & -1 & & & & & & \\
 & & & -1 & & & & & \\
\vdots & & \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\
 & & & & & -1 & & & \\
 & & & & & & -1 & & \\
 & & & & & & & \ddots & \\
0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & -1
\end{array}\right). \tag{A.3}
$$

Then a routing-feasible assignment of $\{r_T : T \in \mathcal{T}_s, s \in S\}$ satisfies the following matrix inequality:

$$
\mathbf{Mr} \leq \mathbf{c}. \tag{A.4}
$$

Farkas' lemma (see, e.g., [100, §1.4]) provides necessary and sufficient conditions for the feasibility of a system of linear inequalities. The following lemma applies Farkas' lemma to inequality (A.4):

Lemma .1 (Farkas) *There exists a solution to* (A.4) *if and only if every row vector* $\mathbf{v}^T$ *with*

v $\geq$ ***0*** *and* $\mathbf{v}^T\boldsymbol{M} = \boldsymbol{0}$ *satisfies* $\mathbf{v}^T\boldsymbol{c} \geq 0$.

We define

$$\mathbf{v}^T = \left(v_1, \cdots, v_{|E|}, v_{|E|+1}, \cdots, v_{|E|+|S|}, v_{|E|+|S|+1}, \cdots, v_{|E|+|S|+\sum_{s\in S}|\mathcal{T}_s|}\right). \tag{A.5}$$

Note that the steps of Fourier-Motzkin elimination maintain rational or integral rate coefficients throughout the procedure; this is why we need not consider irrational edge distances. Equation $\mathbf{v}^T\mathbf{M} = \mathbf{0}$ implies:

$$\sum_{e\in E} v_e \delta^s_{e,T^j_s} - v_{|E|+s} - v_{|E|+|S|+z^j_s} = 0,\ s \in S, j \in \{1, \cdots, |\mathcal{T}_s|\}, \tag{A.6}$$

where

$$z^j_s = \begin{cases} |\mathcal{T}_1| + |\mathcal{T}_2| + \cdots + |\mathcal{T}_{s-1}| + j, & s \in S, s > 1, j \in \{1, \cdots, |\mathcal{T}_s|\} \\ j, & s = 1, j \in \{1, \cdots, |\mathcal{T}_1|\} \end{cases}$$

Therefore by Lemma .1, for every $\mathbf{v} \geq \mathbf{0}$ satisfying (A.6), the inequality $\mathbf{v}^T\mathbf{c} \geq 0$ must hold. It can be written as

$$v_{|E|+1}R_1 + v_{|E|+2}R_2 + \cdots + v_{|E|+|S|}R_{|S|} \leq \sum_{e\in E} v_e C_e. \tag{A.7}$$

Fix a distance vector $\mathbf{a} = (a_1, \cdots, a_{|E|})$ and let

$$\mathbf{v_a} = \{\mathbf{v} \geq \mathbf{0} : (v_1, \cdots, v_{|E|}) = (a_1, \cdots, a_{|E|})\}$$

. Then for $\mathbf{v} \in \mathbf{v_a}$ inequality (A.6) can be written as

$$L_{\mathbf{a}}(T^j_s) - v_{|E|+s} - v_{|E|+|S|+z^j_s} = 0,\ s \in S, j \in \{1, \cdots, |\mathcal{T}_s|\}, \tag{A.8}$$

and inequality (A.7) can be written as

$$v_{|E|+1}R_1 + v_{|E|+2}R_2 + \cdots + v_{|E|+|S|}R_{|S|} \leq \sum_{e\in E} a_e C_e. \tag{A.9}$$

Since $v_{|E|+|S|+z_s^j} \geq 0$, then by (A.8), $v_{|E|+s} \leq L_{\mathbf{a}}(T_s^j)$ for every s and j. Therefore for every $\mathbf{v} \in \mathbf{v_a}$, $v_{|E|+s}$ can be bounded from above by $\min_j L_{\mathbf{a}}(T_s^j) = \ell_{\mathbf{a}}(s)$. Observe that it is possible to choose $v_{|E|+s} = \ell_{\mathbf{a}}(s)$ for every s by setting $v_{|E|+s} = \ell_{\mathbf{a}}(s)$ and $v_{|E|+|S|+z_s^j} = L_{\mathbf{a}}(T_s^j) - \ell_{\mathbf{a}}(s)$. Next notice that the left hand side of (A.9) is maximized among vectors in $\mathbf{v_a}$ when the values of $v_{|E|+s}$ are maximized; i.e., when $v_{|E|+s} = \ell_{\mathbf{a}}(s)$. Equivalently (A.9) holds if and only if the following inequality is satisfied:

$$\sum_{s \in S} \ell_{\mathbf{a}}(s) R_s \leq \sum_{e \in E} a_e C_e; \tag{A.10}$$

this is the routing bound corresponding to the distance vector **a**.

APPENDIX B

Here we prove Lemma B.3 from Chapter II. We begin by introducing some terminology and a result from [76, Ch. 8].

Let E be an arbitrary nonempty subset of the inequalities from (2.1) that define the polytope P of all routing-feasible rate vectors. Let F represent the collection of rate-vectors in P that satisfy each inequality in E with equality. If F is nonempty, it is called a *face* of the polytope P. A face F of P is said to be a *facet* of P if there is no face $F' \neq F$ of P for which $F \subset F'$. The following result from [76, §8.4] is central to the proof of Lemma B.3:

Theorem .2 ([76], Theorem 8.2) *Suppose polytope P has no inequality which is always satisfied by equality. Further assume that $Ax \leq b$ is a minimal set of inequalities that define P. Let A_i^T denote the i^{th} row of A and let b_i denote the i^{th} element of column vector b. For each row i, there is a one-to-one correspondence between the defining halfspace $A_i^T x \leq b_i$ and a facet F_i of P given by $F_i = \{R \in P : A_i^T R = b_i\}$, where we represent rate-tuples as column vectors. Furthermore $Ax \leq b$ is the unique minimal representation of P up to the multiplication of inequalities by positive scalars.*

To apply the preceding theorem we must first establish that there is no inequality in (2.1) that is always satisfied with equality. We assume that there is at least one Steiner tree in the network corresponding to each session $s \in S$ so that there is at least one rate-tuple with $R_s > 0$ for each $s \in S$. Next consider the inequality corresponding to the distance vector $\mathbf{a} \neq \mathbf{0}$. Let $F_{\mathbf{a}} = \{R \in P : \sum_{s \in S} \ell_{\mathbf{a}}(s) R_s = \sum_{e \in E} a_e C_e\}$. If R is a rate-tuple on the face $F_{\mathbf{a}}$, then R is not the all-zero vector. Furthermore, for any $0 < \epsilon < 1$, ϵR will be a feasible rate-tuple not on $F_{\mathbf{a}}$. Therefore, this inequality can not be always satisfied with equality.

To complete the proof of the lemma, observe that if for two distance vectors $\mathbf{a}$ and $\mathbf{b}$ the face $F_{\mathbf{a}}$ is included in $F_{\mathbf{b}}$, then $F_{\mathbf{a}}$ is not a facet of P. By the previous theorem, the inequality corresponding to distance vector $\mathbf{a}$ cannot be part of a minimal representation of P. Hence distance vector $\mathbf{a}$ is redundant in the presence of distance vector $\mathbf{b}$.

APPENDIX C

Here we provide an algebraic proof for Proposition B.5 from Chapter II. Given distance vector $\mathbf{a} = (a_1, \cdots, a_{|E|})$, let $M_s,\ s \in \{1, \dots, |S|\}$, denote the number of shortest routing trees for session s. Assume without loss of generality that for a fixed $s,\ L_{\mathbf{a}}(T_j^s)$ is non-decreasing with j so that

$$L_{\mathbf{a}}(T_1^s) = \cdots = L_{\mathbf{a}}(T_{M_s}^s) = \ell_{\mathbf{a}}(s).$$

Suppose that $(R_1, \dots, R_{|S|})$ is a routable rate-tuple lying on the hyperplane $\sum_{s \in S} \ell_{\mathbf{a}}(s) R_s = \sum_{e \in E} a_e C_e$. By Condition 1) of Lemma B.4 it follows that

$$\sum_{s \in S} \ell_{\mathbf{a}}(s) R_s = \sum_{s \in S} \ell_{\mathbf{a}}(s) \sum_{j=1}^{M_s} r_{T_j^s} = \sum_{s \in S} \sum_{j=1}^{M_s} L_{\mathbf{a}}(T_j^s) r_{T_j^s} = \sum_{e \in E} a_e C_e.$$

Define $\tilde{\delta}_{e,(s,j)}$ to be one if edge e is in the j^{th} shortest routing tree for session s with respect to distance vector $\mathbf{a}$ and 0 otherwise. Then by writing each $L_{\mathbf{a}}(T_j^s)$ as the sum of the edge distances a_e for which edge e occurs in the j^{th} shortest routing tree for session s, we see that

$$\ell_{\mathbf{a}}(s) = L_{\mathbf{a}}(T_j^s) = \sum_{e \in E} \tilde{\delta}_{e,(s,j)} a_e,\ s \in S, j \in \{1, \cdots, M_s\}. \tag{C.1}$$

Condition 1) of Lemma B.4 states that edge e is used by session s only if e is on a shortest routing tree for s. Therefore the partial flow of session s through routing tree T_j^s is $\tilde{\delta}_{e,(s,j)} r_{T_j^s}$, which is zero for $j > M_s$. Condition 2) of Lemma B.4 states that every edge with a non-zero distance is fully utilized. Therefore if $a_e > 0$ for $e \in E$, then

$$\sum_{s \in S} \sum_{j=1}^{M_s} \tilde{\delta}_{e,(s,j)} r_{T_j^s} = C_e. \tag{C.2}$$

Set $\mathcal{S}$ to be the $|E|\times|E|$ diagonal matrix with diagonal entries $\sigma_e = \sum_{s\in S}\sum_{j=1}^{M_s}\tilde{\delta}_{e,(s,j)} r_{T_j^s} - C_e$, and set D to be the $|E|\times|E|$ diagonal matrix with diagonal entries $d_e = \frac{b_e}{a_e}$ when $a_e \neq 0$ and $d_e = 0$ when $a_e = 0$. It follows from (C.2) that $\mathcal{S}\mathbf{a}^T = \mathbf{0}$, and hence $D\mathcal{S}\mathbf{a}^T = \mathbf{0}$ as well. Diagonal matrices commute, so $\mathcal{S}D\mathbf{a}^T = \mathbf{0}$. Consider any edge $e \in E$. If $a_e \neq 0$ then $d_e a_e = b_e$. If $a_e = 0$, then $d_e a_e = 0$, and Condition 1) of Proposition B.5 implies that $b_e = 0$. Therefore $d_e a_e = b_e$ in this case as well. Thus $D\mathbf{a}^T = \mathbf{b}^T = (b_1, \cdots, b_{|E|})^T$ and it follows that $\mathcal{S}\mathbf{b}^T = \mathbf{0}$. Therefore,

$$0 = \sum_{e\in E}\sigma_e b_e = \sum_{s\in S}\sum_{j=1}^{M_s}\left(\sum_{e\in E}\tilde{\delta}_{e,(s,j)} b_e\right) r_{T_j^s} - \sum_{e\in E} b_e C_e. \tag{C.3}$$

By Condition 2) of Proposition B.5, $L_{\mathbf{b}}(T_1^s) = \cdots = L_{\mathbf{b}}(T_{M_s}^s) = \ell_{\mathbf{b}}(s)$ for each $s \in S$, and the counterpart to (C.1) is

$$\ell_{\mathbf{b}}(s) = L_{\mathbf{b}}(T_j^s) = \sum_{e\in E}\tilde{\delta}_{e,(s,j)} b_e,\; s \in S, j \in \{1, \cdots, M_s\}. \tag{C.4}$$

Substituting (C.4) into (C.3) we obtain

$$0 = \sum_{s\in S}\sum_{j=1}^{M_s}\ell_{\mathbf{b}}(s) r_{T_j^s} - \sum_{e\in E} b_e C_e = \sum_{s\in S}\ell_{\mathbf{b}}(s) R_s - \sum_{e\in E} b_e C_e,$$

and so $(R_1 \cdots, R_{|S|})$ does lie on the hyperplane $\sum_{s\in S}\ell_{\mathbf{b}}(s) R_s = \sum_{e\in E} b_e C_e$.

Thus, if a routable rate-tuple $(R_1, \cdots, R_{|S|})$ is on the hyperplane corresponding to $\mathbf{a}$, then it is also on the hyperplane corresponding to $\mathbf{b}$. Since the routing rate region can be described in terms of its defining hyperplanes, the bound given by the hyperplane for $\mathbf{a}$ is redundant assuming we already have the bound given by the hyperplane given by $\mathbf{b}$.

APPENDIX D

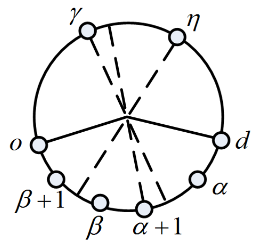

Fig. 21. An instance of a distance vector of type 1 on a ring network.

Here we prove Theorem C.5 from Chapter II. Properties 1 and 2 are trivially satisfied in this case. Next we show that Property 3 also holds. We begin our discussion with unicast sessions. Consider two arbitrary vertices o and d of the network as in Figure 21 and the unicast session s from o to d. Recall that for any two vertices i and j and distance vector $\mathbf{a}$, $L_{\mathbf{a}}(i,j)$ represents the length of the clockwise path from i to j on the ring with respect to the vector $\mathbf{a}$. Suppose $L_{\mathbf{a}}(d,o) < L_{\mathbf{a}}(o,d)$. We wish to show that $L_{\mathbf{b}}(d,o) \leq L_{\mathbf{b}}(o,d)$ so that the shortest path between two vertices remains shortest for $\mathbf{b}$. We first discuss the case for which there are at least two edges on the clockwise path from d to o with edge distance of 1 in $\mathbf{b}$. Consider an arbitrary edge α on the clockwise path from d to o such that $b_\alpha = 1$. Let β denote the next edge after α on the path from d to o in the clockwise direction with $b_\beta = 1$. Let $\mathcal{C}(o,d)$ denote the clockwise path from o to d. Since $b_\alpha = b_\beta = 1$, there must be at least two distinct vertices, say γ and η, on $\mathcal{C}(o,d)$ for which the diameter starting from these points will intersect the arcs corresponding to edges α and β (see Figure 21). Next consider the diameter starting from point corresponding to vertex $\alpha + 1$. This diameter should intersect $\mathcal{C}(o,d)$ at an edge between vertices η and γ. Therefore, for each pair of successive edges on the clockwise path from d to o with unit edge distances in $\mathbf{b}$

there is an edge on the clockwise path from o to d with unit edge distance in $\mathbf{b}$. Thus $L_{\mathbf{b}}(d,o) - 1 \le L_{\mathbf{b}}(o,d)$. Furthermore, the two diameters starting from vertices o and d intersect $\mathcal{C}(o,d)$, and will produce two more unit edge distances in $\mathbf{b}$ that we have not yet counted. Thus we have $L_{\mathbf{b}}(d,o) \le L_{\mathbf{b}}(o,d)$. To complete this argument, consider the case with $L_{\mathbf{b}}(d,o) = 0$; in this case apparently $0 = L_{\mathbf{b}}(d,o) \le L_{\mathbf{b}}(o,d)$. Finally, if $L_{\mathbf{b}}(d,o) = 1$, we know that at least one of the diameters starting from o or d will produce a unit edge distance on $\mathcal{C}(o,d)$, and hence $1 = L_{\mathbf{b}}(d,o) \le L_{\mathbf{b}}(o,d)$.

Next we show that Property 3 holds for broadcast sessions. Since the trees for routing broadcast sessions are the collection of paths consisting of all but one edge in the network, a shortest tree for a broadcast session corresponds to omitting an edge with maximal edge distance. Since $b_k \in \{0,\ 1\}$ for all k, we have to show that if a_i is a maximal edge distance in $\mathbf{a}$ then $b_i = 1$. To arrive at a contradiction, assume that $b_i = 0$. Then it follows that there is another edge j for which the diameters starting from vertices i and $i+1$ both intersect the arc corresponding to edge j. Hence $a_j > a_i$, which contradicts the maximality of a_i.

APPENDIX E

Here we prove Theorem C.6 from Chapter II. To construct $\mathbf{b}$ for a Type 2 of distance vector $\mathbf{a}$, we use the Basic Generation Procedure, or the Basic Generation Procedure followed by the change of a particular edge distance from 0 to 1. Observe that Property 1 is trivially satisfied. Since we are only considering positive distance vectors $\mathbf{a}$, then Property 2 also holds. We next discuss Property 3. Assume that $\{p, q\}$ is the unique pair of vertices satisfying $L_{\mathbf{a}}(p, q) = L_{\mathbf{a}}(q, p)$. There are two subcases to consider:

1. Suppose the Basic Generation Procedure produces a vector $\mathbf{b}$ with $L_{\mathbf{b}}(p, q) = L_{\mathbf{b}}(q, p)$. Then in this case we do not need to make any changes to vector $\mathbf{b}$. Let us study a unicast session between two vertices o and d and show that Property 3 holds for it. If the clockwise path from p to q includes the clockwise path from o to d, then $L_{\mathbf{a}}(o, d) < L_{\mathbf{a}}(d, o)$. Similarly, $L_{\mathbf{b}}(o, d) \leq L_{\mathbf{b}}(p, q) = L_{\mathbf{b}}(q, p) \leq L_{\mathbf{b}}(d, o)$, which shows that Property 3 holds for the unicast session between o and d. A symmetric argument holds for the case where o and d are both located on the counterclockwise path from p to q.

   Next assume that o is located on the clockwise path from p to q and d is located on the counterclockwise path from p to q and that $L_{\mathbf{a}}(o, d) < L_{\mathbf{a}}(d, o)$ as depicted in Figure 22. (By the assumption, $L_{\mathbf{a}}(o, d) \neq L_{\mathbf{a}}(d, o)$.) By the same argument as in the proof of Theorem C.5, we can show that for each pair of successive edges on the clockwise path from o to q with unit distance in $\mathbf{b}$, there is an edge with unit distance on the clockwise path from d to p. Hence $L_{\mathbf{b}}(o, q) - 1 \leq L_{\mathbf{b}}(d, p)$. Observe that the diameter originating at vertex o will intersect an edge between d and p, and

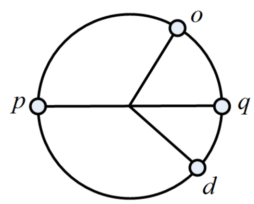

Fig. 22. An instance of case 1 on a ring network.

will produce another unit edge distance which we have not yet counted. (Note that vertex q can not produce any extra unit distance edge as the diameter originating at this vertex intersects the circle at p.) Thus, $L_{\mathbf{b}}(o,q) \leq L_{\mathbf{b}}(d,p)$. By symmetry, $L_{\mathbf{b}}(q,d) \leq L_{\mathbf{b}}(p,o)$. By summing these two inequalities we obtain

$$L_{\mathbf{b}}(o,d) = L_{\mathbf{b}}(o,q) + L_{\mathbf{b}}(q,d) \leq L_{\mathbf{b}}(d,p) + L_{\mathbf{b}}(p,o) = L_{\mathbf{b}}(d,o)$$

which shows that Property 3 holds for the unicast session between o and d.

The argument for broadcast sessions from Theorem C.5 also holds for this case without any change, and so Property 3 is satisfied for broadcast sessions in this case as well.

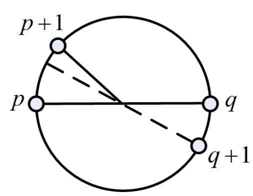

Fig. 23. An instance of case 2 on a ring network.

2. Next suppose the lengths of the different paths between p and q do not remain the same for vector $\mathbf{b}$. We first describe how to modify distance vector $\mathbf{b}$ and we then show that the resulting distance vector $\mathbf{b}$ satisfies Property 3. Figure 23 depicts the circle corresponding to the edge distances in $\mathbf{a}$. Observe that $a_p \neq a_q$ since the

unicast session between $p+1$ and $q+1$ do not have two equal length routing paths. Assume without loss of generality that $a_p > a_q$. Consider the clockwise path from $q+1$ to p and the clockwise path from p to q. With an argument similar to that for Type 1 distance vectors we can show that corresponding to every pair of successive edges with unit distance in $\mathbf{b}$ on the first path, there is an edge with unit distance on the second path. Thus we have $L_{\mathbf{b}}(q+1,p)-1 \leq L_{\mathbf{b}}(p,q)$. Observe that the diameter originating at vertex $q+1$ intersects the second path between vertices p and $p+1$ and produces another unit edge distance which we have not yet counted. Furthermore, edge q will not be intersected by any diameter and thus $b_q = 0$. We have the following relationship:

$$L_{\mathbf{b}}(q,p) = L_{\mathbf{b}}(q+1,p) \leq L_{\mathbf{b}}(p,q).$$

Now consider Figure 24 in which we have moved the point corresponding to vertex

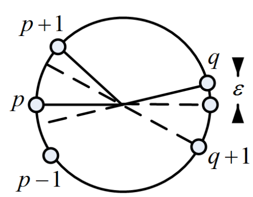

Fig. 24. The situation in case 2 after moving q on $\mathcal{C}$.

q an arbitrarily small distance ϵ in the counterclockwise direction. If we apply the Basic Generation Procedure on this new set of edge distances and call the resulting binary set of distance vector $\mathbf{b}'$, then $b'_i = b_i$ for all edges except for edge q which is intersected by the diameter originating at vertex p, and possibly for edge $p-1$ which is now intersected by the diameter originating at vertex q; i.e., $b'_{p-1} = b'_q = 1$. If we use the argument for Type 1 distance vectors here, we find that for every pair of successive unit edge distances in $\mathbf{b}'$ for the edges on the clockwise path from p to q

there is another edge with distance one on the clockwise path from q to p, and thus $L_{\mathbf{b}'}(p,q) - 1 \leq L_{\mathbf{b}'}(q,p)$. On the other hand note that the two diameters originating at vertices p and q both produce two edges with unit distance in $\mathbf{b}'$ which we have not yet counted. Thus $L_{\mathbf{b}'}(p,q) + 1 \leq L_{\mathbf{b}'}(q,p)$. Using the relationship of elements in $\mathbf{b}'$ and $\mathbf{b}$ we have $L_{\mathbf{b}'}(p,q) = L_{\mathbf{b}}(p,q)$ and $L_{\mathbf{b}'}(q,p) = L_{\mathbf{b}}(q,p) + 1 + (1 - b_{p-1})$. Thus we have

$$L_{\mathbf{b}}(p,q) \leq L_{\mathbf{b}}(q,p) + (1 - b_{p-1}).$$

Recall that $L_{\mathbf{b}}(q,p) \leq L_{\mathbf{b}}(p,q)$. Therefore the condition $L_{\mathbf{b}}(q,p) \neq L_{\mathbf{b}}(p,q)$ implies $b_{p-1} = 0$, and in this case we have $L_{\mathbf{b}}(p,q) = L_{\mathbf{b}}(q,p) + 1$. In order to obtain $L_{\mathbf{b}}(q,p) = L_{\mathbf{b}}(p,q)$ we manually change the value of b_q from zero to one in $\mathbf{b}$. We next show that Property 3 holds for distance vector $\mathbf{b}$.

First we consider the unicast sessions. For all unicast sessions where both the source and destination are on the clockwise or counterclockwise path from p to q we use the argument from the preceding case. Next consider the unicast session between two vertices o and d, where o is located on the clockwise path from p to q and d is located on the counterclockwise path from p to q on the ring. First assume that the shortest

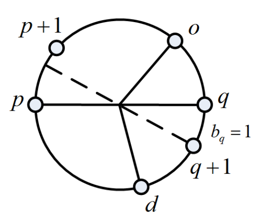

Fig. 25. The situation in case 2 with the shorter path from o to d in the clockwise direction.

path between o and d by $\mathbf{a}$ is the clockwise path from o to d (see Figure 25) . Using the argument for Type 1 distance vectors, we conclude that for every pair of successive unit edge distances on the clockwise path from o to q there is another unit edge distance on the clockwise path from d to p, and for every pair of successive unit edge distances on the clockwise path from $q+1$ to d there is another edge with distance of one on the clockwise path from p to o. Since we have reset $b_q = 1$, we obtain $L_{\mathbf{b}}(o,q) - 1 \leq L_{\mathbf{b}}(d,p)$ and $L_{\mathbf{b}}(q,d) - 2 \leq L_{\mathbf{b}}(p,o)$. Thus $L_{\mathbf{b}}(o,d) - 3 \leq L_{\mathbf{b}}(d,o)$. Now if the diameters starting at $q+1$ and d generate two distinct unit edge distances in $\mathbf{b}$, these two along with the one generated by the diameter starting at o will add three more ones to $L_{\mathbf{b}}(d,o)$ and result in $L_{\mathbf{b}}(o,d) \leq L_{\mathbf{b}}(d,o)$.

However it might happen that the diameters starting at $q+1$ and d both intersect edge p on $\mathcal{C}$. In this case we use the following argument. We have $L_{\mathbf{b}}(p,q) = L_{\mathbf{b}}(q,p)$ as the result of the modifications. Therefore $1 + L_{\mathbf{b}}(p+1,q) = 1 + L_{\mathbf{b}}(q+1,p)$ and hence $L_{\mathbf{b}}(p+1,q) = L_{\mathbf{b}}(q+1,p)$. However $L_{\mathbf{b}}(q+1,p) = L_{\mathbf{b}}(d,p)$ because there is no diameter intersecting the clockwise path between $q+1$ and d. To arrive at a contradiction, suppose that there is at least one diameter intersecting the clockwise path between $q+1$ and d. By assumption the diameters starting at $q+1$ and d both intersect edge p on $\mathcal{C}$, and it follows that any diameter that intersects the clockwise path between $q+1$ and d must originate at a vertex between vertices p and $p+1$. However there is no vertex between p and $p+1$, and so there is no diameter intersecting the clockwise path between $q+1$ and d. Therefore $L_{\mathbf{b}}(q+1,p) = L_{\mathbf{b}}(d,p)$. Thus $L_{\mathbf{b}}(p+1,q) = L_{\mathbf{b}}(d,p)$. Since $L_{\mathbf{b}}(p+1,q) + 1 = L_{\mathbf{b}}(p+1,d)$ and $L_{\mathbf{b}}(d,p) + 1 = L_{\mathbf{b}}(d,p+1)$, we have $L_{\mathbf{b}}(p+1,d) = L_{\mathbf{b}}(d,p+1)$. Finally,

$$L_{\mathbf{b}}(o,d) \leq L_{\mathbf{b}}(p+1,d) = L_{\mathbf{b}}(d,p+1) \leq L_{\mathbf{b}}(d,o)$$

Next suppose the shortest path between o and d is the counterclockwise path (see Figure 26). Using the argument for Type 1 distance vectors as in the previous instance, we conclude that $L_{\mathbf{b}}(d,p) - 1 \leq L_{\mathbf{b}}(o,q)$ and $L_{\mathbf{b}}(p,o) - 1 \leq L_{\mathbf{b}}(q,d)$. By accounting for the two unit edge distances produced by the diameters originating at o and d which we have not yet counted and by summing the two inequalities we get

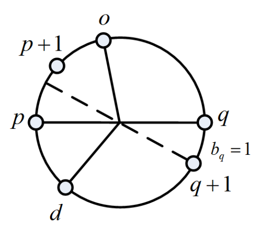

Fig. 26. The situation in case 2 with the shorter path from o to d in the counterclockwise direction.

$L_{\mathbf{b}}(d, o) \leq L_{\mathbf{b}}(o, d)$. Hence the shortest path for $\mathbf{a}$ will remain shortest for $\mathbf{b}$ and Property 3 holds for unicast sessions.

To complete our proof we need to show that Property 3' holds for broadcast sessions. Consider a broadcast session s and the set of its complementary trees. Since each routing tree for a broadcast session is the total ring after removing a single edge from it, then the set of complementary trees will be the set of all trees formed by single edges and their end vertices. Then to satisfy Property 3', we need to show that if edge e_0 satisfies $a_{e_0} = \max_{e \in E} a_e$, then $b_{e_0} = \max_{e \in E} b_e = 1$. First notice that as we saw in the argument for Type 1 distance vectors, the Basic Generation Procedure results in every edge with largest edge distance having a unit distance in $\mathbf{b}$. Moreover, the modifications for this case increase one edge distance from zero to one. Therefore all maximum length complementary trees with respect to $\mathbf{a}$ will have length 1 under $\mathbf{b}$, and hence Property 3' will be satisfied.

APPENDIX F

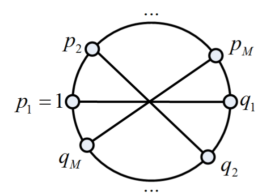

Fig. 27. The situation where there are several pairs of vertices with equal length clockwise and counterclockwise paths.

Here we prove Theorem C.7 from Chapter II. Consider the case where there are exactly M pairs of vertices, say $\{p_1, q_1\}, \{p_2, q_2\}, \cdots, \{p_M, q_M\}$ for which $L_{\mathbf{a}}(p_i, q_i) = L_{\mathbf{a}}(q_i, p_i)$ for all $i \in \{1, \cdots, M\}$. Without loss of generality we assume that $1 = p_1 < p_2 < \cdots < p_M < q_1 < q_2 < \cdots < q_M \leq n$, as depicted in Figure 27. To construct $\mathbf{b}$, we first decompose $\mathbf{a}$ into M subvectors $\mathbf{a}_1, \cdots, \mathbf{a}_M$, with

$$\mathbf{a}_i = (a_{p_i}, a_{p_i+1}, \cdots, a_{p_{i+1}-1}, a_{q_i}, a_{q_i+1}, \cdots, a_{q_{i+1}-1})$$

for $1 \leq i \leq M-1$ and $\mathbf{a}_M = (a_{p_M}, a_{p_M+1}, \cdots, a_{q_1-1}, a_{q_M}, a_{q_M+1}, \cdots, a_n)$. From Figure 27 it is easy to see that $L_{\mathbf{a}}(p_i, p_{i+1}) = L_{\mathbf{a}}(q_i, q_{i+1})$ for $1 \leq i \leq M-1$ and $L_{\mathbf{a}}(p_M, q_1) = L_{\mathbf{a}}(q_M, p_1)$. Therefore if we form the circle corresponding to subvector $\mathbf{a}_i$ and relabel the vertices on it from 1 to $p_{i+1} - p_i + q_{i+1} - q_i$, we obtain $L_{\mathbf{a}_i}(1, p_{i+1} - p_i + 1) = L_{\mathbf{a}_i}(p_{i+1} - p_i + 1, 1)$ for $1 \leq i \leq M-1$ and $L_{\mathbf{a}_M}(1, q_1 - p_M + 1) = L_{\mathbf{a}_M}(q_1 - p_M + 1, 1)$. Observe that each subvector $\mathbf{a}_i$, $1 \leq i \leq M$, in isolation corresponds to a circle with $|\mathbf{a}_i|$ vertices which has exactly one pair of vertices with equal length clockwise and counterclockwise paths between them, namely vertex 1 and vertex $p_{i+1} - p_i + 1$; if there were more than one such pair, then there would be another pair p' and q' of vertices on the original ring such that $p_i < p' < p_{i+1}$, $q_i < q' < q_{i+1}$, and $L_{\mathbf{a}}(p', q') = L_{\mathbf{a}}(q', p')$, contradicting our initial assumption. Hence $\mathbf{a}_i$ is a Type 2 distance vector.

For every $\mathbf{a}_i$ we use the argument for Type 2 distance vectors to construct a binary distance vector $\mathbf{b}_i$. We obtain $\mathbf{b} = (b_1, \cdots, b_n)$ from the relationships

$$(b_{p_i}, b_{p_i+1}, \cdots, b_{p_{i+1}-1}, b_{q_i}, b_{q_i+1}, \cdots, b_{q_{i+1}-1}) = \mathbf{b}_i$$

for $1 \le i \le M-1$ and

$$(b_{p_M}, b_{p_M+1}, \cdots, b_{q_1-1}, b_{q_M}, b_{q_M+1}, \cdots, b_n) = \mathbf{b}_M.$$

Obviously $\mathbf{b}$ satisfies Properties 1 and 2. Next we prove that it also satisfies Property 3.

We first consider unicast sessions. The construction of $\mathbf{b}$ results in the following relationships:

$$\begin{aligned} L_{\mathbf{b}}(p_i, p_{i+1}) &= L_{\mathbf{b}}(q_i, q_{i+1}), \quad 1 \le i \le M-1, \\ L_{\mathbf{b}}(p_M, p_1) &= L_{\mathbf{b}}(q_M, q_1). \end{aligned} \tag{F.1}$$

Consider a pair o and d of vertices on the ring and the unicast session s between them.

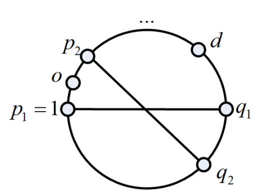

Fig. 28. The case where $d < q_1$.

Assume without loss of generality that $p_1 \le o \le p_2$ and $L_{\mathbf{a}}(o, d) \le L_{\mathbf{a}}(d, o)$, so that $o < d \le q_2$. We must establish that $L_{\mathbf{b}}(o, d) \le L_{\mathbf{b}}(d, o)$. To show this, first assume that $d < q_1$ (see Figure 28) . In this case it follows from (F.1) that

$$L_{\mathbf{b}}(o, d) \le L_{\mathbf{b}}(p_1, q_1) = L_{\mathbf{b}}(q_1, p_1) \le L_{\mathbf{b}}(d, o). \tag{F.2}$$

Next let $q_1 \leq d \leq q_2$ (see Figure 29). Then by assumption

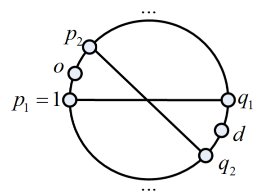

Fig. 29. The case where $q_1 \leq d \leq q_2$.

$$\begin{aligned} L_{\mathbf{a}}(o,d) = \quad & L_{\mathbf{a}}(o,p_2) + L_{\mathbf{a}}(p_2,p_3) + \cdots + L_{\mathbf{a}}(p_M,q_1) + L_{\mathbf{a}}(q_1,d) \leq \\ & L_{\mathbf{a}}(d,q_2) + L_{\mathbf{a}}(q_2,q_3) + \cdots + L_{\mathbf{a}}(q_M,p_1) + L_{\mathbf{a}}(p_1,o) = \quad L_{\mathbf{a}}(d,o). \end{aligned} \tag{F.3}$$

It follows from (F.3) that

$$L_{\mathbf{a}}(o,p_2) + L_{\mathbf{a}}(q_1,d) \leq L_{\mathbf{a}}(d,q_2) + L_{\mathbf{a}}(p_1,o). \tag{F.4}$$

Since $\mathbf{b}_1$ satisfies Property 3 for $\mathbf{a}_1$ it follows from (F.4) and the definition of $\mathbf{b}$ that

$$L_{\mathbf{b}}(o,p_2) + L_{\mathbf{b}}(q_1,d) \leq L_{\mathbf{b}}(d,q_2) + L_{\mathbf{b}}(p_1,o). \tag{F.5}$$

Therefore (F.5) and (F.1) imply

$$\begin{aligned} L_{\mathbf{b}}(o,d) = \quad & L_{\mathbf{b}}(o,p_2) + L_{\mathbf{b}}(p_2,p_3) + \cdots + L_{\mathbf{b}}(p_M,q_1) + L_{\mathbf{b}}(q_1,d) \leq \\ & L_{\mathbf{b}}(d,q_2) + L_{\mathbf{b}}(q_2,q_3) + \cdots + L_{\mathbf{b}}(q_M,p_1) + L_{\mathbf{b}}(p_1,o) = \quad L_{\mathbf{b}}(d,o). \end{aligned} \tag{F.6}$$

Hence distance vector $\mathbf{b}$ satisfies Property 3 for the unicast sessions. To prove that this is also true for broadcast sessions, we have to show that all edges with largest edge distances for $\mathbf{a}$ have unit edge distances for $\mathbf{b}$. Let $p_i \leq k \leq p_{i+1}$ be an edge distance with largest distance for $\mathbf{a}$. Then it must correspond to an edge with largest distance for $\mathbf{a}_i$. Hence our earlier argument for Type 2 distance vectors establishes that the position corresponding to b_k in $\mathbf{b}_i$ is equal to one, and this proves our claim.

APPENDIX G

Here we prove the network coding bounds for a ring with four vertices. Consider a ring network with four vertices and replace every edge in the network with two oppositely directed edges and obtain the directed graph $G(V, E)$ of Figure 7. Suppose that the network is clocked, i.e., a universal clock ticks N times. For this network we introduce the following notation for the network coding setting from time 1 to N.

- Let W_s denote the message of session s.
- Let $X_{ij}^{(t)}$ denote the bitstream of edge from i to j at time t and $X_{ij}^{k} = \left[X_{ij}^{(1)}, \cdots, X_{ij}^{(k)}\right]$.

Vertex i transmits the bitstream $X_{ij}^{(t)}$, $(i, j) \in E$ after clock tick $t - 1$ and before clock tick t for $t = 1, \cdots, N$ and vertex j receives bitstream $X_{ij}^{(t)}$ at clock tick t. In a network coding solution, $X_{ij}^{(t)}$ is a function of $\left\{W_s : \nu_s = i\right\}$ and $\left\{X_{(i+1)(i)}^{t-1}, X_{(i-1)(i)}^{t-1}\right\}$. After time N, at every vertex i the received messages $\left\{W_s : i \in D_s\right\}$ with destination i can be decoded as a function of $\left\{W_s : \nu_s = i\right\}$ and $\left\{X_{(i+1)(i)}^{N}, X_{(i-1)(i)}^{N}\right\}$. For this setting we prove the following lemma:

Lemma .3 *For any network and any network coding solution, there is a one to one correspondence between the set of messages in the network* $\left\{W_s : s \in S\right\}$*, and the set of bitstreams of the edges* $\left\{X_{ij}^{N} : (i, j) \in E\right\}$.

Proof Since the encoding functions at vertices are deterministic, a set of messages uniquely determine a set of bitstreams. Next suppose that there are two different realizations of $\left\{W_s : s \in S\right\}$, say $U = \left\{u_s : s \in S\right\}$ and $V = \left\{v_s : s \in S\right\}$ corresponding to a realization of bitstreams $\left\{X_{ij}^{N} : (i, j) \in E\right\}$, say $X = \left\{x_{ij}^{N} : (i, j) \in E\right\}$. It means that there is at least one session s_0 for which $u_{s_0} \neq v_{s_0}$. Next we show that F which is another realization of messages and is a combination of messages in U and V as the following

$$F = \left\{f_s : s \in S\right\}, f_s = \begin{cases} u_s & \text{if} \quad \nu_s \neq \nu_{s_0}, \\ v_s & \text{if} \quad \nu_s = \nu_{s_0}, \end{cases} \tag{G.1}$$

also results in the bitstream X. We use induction over time instances. For every vertex i, X_{ij}^1 is a function of $\left\{W_s : \nu_s = i\right\}$. For $i \neq \nu_{s_0}$, the realization of X_{ij}^1 corresponding to F is the same as its realization for U, and for $i = \nu_{s_0}$, the realization of X_{ij}^1 corresponding to F is the same as its realization for V. Therefore by assumption X_{ij}^1 will have the same realization for U, V, and F. As the indution step, suppose that for time instance k, the realization of X_{ij}^k is the same for U, V, and F. Next, $X_{ij}^{(k+1)}$ is a function of $\left\{W_s : \nu_s = i\right\}$ and $\left\{X_{ji}^k, (j,i) \in E\right\}$. Therefore, the realization of $X_{ij}^{(k+1)}$ corresponding to F is equal to its realization corresponding to U if $i \neq \nu_{s_0}$ and to its realization corresponding to V if $i = \nu_{s_0}$. Thus, by the induction hypothesis, $X_{ij}^{(k+1)}$ will have the same realization for U, V, and F, which completes our induction.

Next consider a vertex $d \in D_{s_0}$. First notice that since $d \neq \nu_{s_0}$, the realization of $\left\{X_{jd}^N, (j,d) \in E\right\}$ and $\left\{W_s : \nu_s = d\right\}$ is the same for set of messages U and F. Therefore d will decode the same messages for all sessions with destination d in both cases. But this contradicts the fact that W_{s_0} has two different realizations for U and F. Therefore every set of bitstreams $\left\{X_{ij}^N : (i,j) \in E\right\}$ corresponds to a unique set of messages $\left\{W_s : s \in S\right\}$.

$\square$

We apply the result of Lemma .3 to the ring network of Figure 7. Furthermore in the ring network of Figure 7 every realization of messages $\left\{W_s : \nu_s \in \{2,4\}\right\}$ and bitstreams $\left\{X_{ij}^N : i \in \{1,3\}, j \in \{2,4\}\right\}$ uniquely determines a realization of bitstreams $\left\{X_{ij}^N : i \in \{2,4\}, j \in \{1,3\}\right\}$. Therefore together with Lemma .3 we conclude that every realiztion of $\left\{W_s : \nu_s \in \{2,4\}\right\}$ and $\left\{X_{ij}^N : i \in \{1,3\}, j \in \{2,4\}\right\}$ uniquely determines a realization of $\{W_s : s \in S\}$ and vice versa. Therefore the following equation holds:

$$H\Big(\{W_s : s \in S\}\Big) = H\Big(\{W_s : \nu_s \in \{2,4\}\}, \{X_{ij}^N : i \in \{1,3\}, j \in \{2,4\}\}\Big). \quad \text{(G.2)}$$

By expanding the right hand side of (G.2) we obtain the following equation:

$$\begin{aligned} H\Big(\{W_s : s \in S\}\Big) &= H\Big(\{W_s : \nu_s = 2\}, X_{12}^N, X_{32}^N\Big) + H\Big(\{W_s : \nu_s = 4\}, X_{14}^N, X_{34}^N\Big) \\ &\quad - I\Big(\underbrace{\{W_s : \nu_s = 2\}, X_{12}^N, X_{32}^N}_{U_1}; \underbrace{\{W_s : \nu_s = 4\}, X_{14}^N, X_{34}^N}_{U_2}\Big). \quad \text{(G.3)} \end{aligned}$$

Next we find a lower bound for $I(U_1; U_2)$. First notice that U_1 consists of the set of messages originating at vertex 2 and the bitstreams received by vertex 2. Therefore, this set uniquely determines the messages destined for vertex 2. Hence, this set can be used to obtain the set of messages $M = \{W_s : \nu_s = 2,\ 4 \in D_s\} \cup \{W_s : \nu_s = 4,\ 2 \in D_s\} \cup \{W_s : \{2,4\} \subseteq D_s\}$. By an analogous argument for vertex 4 it follows that the set of messages M is a function of U_2. Thus the mutual information term in (G.3) can be written as $I(U_1, M; U_2, M)$. By expanding this term and using the data processing inequality we have,

$$I(U_1, M; U_2, M) \geq I(M; M) = H(M).$$

We combine the preceding bound with (G.3) to establish:

$$\begin{aligned} &H\left(\{W_s : s \in S\}\right) \\ &+ H\left(\{W_s : \nu_s = 2,\ 4 \in D_s\}, \{W_s : \nu_s = 4,\ 2 \in D_s\}, \{W_s : \{2,4\} \subseteq D_s\}\right) \\ &\leq H\left(\{W_s : \nu_s = 2\}, X_{12}^N, X_{32}^N\right) + H\left(\{W_s : \nu_s = 4\}, X_{14}^N, X_{34}^N\right) \\ &\leq H\left(\{W_s : \nu_s = 2\}\right) + H\left(\{W_s : \nu_s = 4\}\right) \\ &+ H\left(X_{12}^N\right) + H\left(X_{32}^N\right) + H\left(X_{14}^N\right) + H\left(X_{34}^N\right). \quad \text{(G.4)} \end{aligned}$$

By the independence of messages of different sessions and (G.4) we have:

$$\begin{aligned} &H\left(\{W_s : \nu_s \in \{1,3\}\}\right) \\ &+ H\left(\{W_s : \nu_s = 2,\ 4 \in D_s\}, \{W_s : \nu_s = 4,\ 2 \in D_s\}, \{W_s : \{2,4\} \subseteq D_s\}\right) \\ &\leq H\left(X_{12}^N\right) + H\left(X_{32}^N\right) + H\left(X_{14}^N\right) + H\left(X_{34}^N\right). \end{aligned} \tag{G.5}$$

Notice that $H\Big(X_{ij}^N\Big) \leq NC_{ij}$ and $H\Big(W_s\Big) = NR_s$. It follows that for a ring with four vertices:

$$\sum_{s:\nu_s \in \{1,3\}} R_s + \sum_{s:\nu_s = 2,\ 4 \in D_s} R_s + \sum_{s:\nu_s = 4,\ 2 \in D_s} R_s + \sum_{s:\{2,4\} \subseteq D_s} R_s \leq C_{12} + C_{32} + C_{14} + C_{34}. \tag{G.6}$$

APPENDIX H

Here we extend our results for node–constrained line networks in Section B to a slightly different setting where the source of messages generated at node i in the original network and the destination of messages decoded at node i in the original network are both node I_i in the node–constrained model. In other words, the set of destination nodes $\mathcal{D}_s$ of session s is a subset of $\{I_1, ..., I_M\}$ rather than $\{O_1, ..., O_M\}$. We can characterize the capacity region of the network as follows.

Theorem 47 *A rate tuple $(R_s : s \in S)$ is achievable in the node-constrained line network with both source and destination at node I_i, if and only if it is nonnegative and it satisfies the following bounds:*
For every $i \in \{1, \cdots, M\}$

$$\sum_{\nu_s = I_i} R_s + \max\left\{\sum_{s\in U_1} R_s, \sum_{s \in U_2} R_s\right\} \leq C_i \tag{H.1}$$

where U_1 and U_2 are defined as follows:

$$U_1 = \{s : \nu_s = I_j, j < i\} \bigcap \{s : \mathcal{D}_s \cap \{I_{i+1}, \cdots, I_M\} \neq \emptyset\}$$
$$U_2 = \{s : \nu_s = I_j, j > i\} \bigcap \{s : \mathcal{D}_s \cap \{I_1, \cdots, I_{i-1}\} \neq \emptyset\},$$

and for every $i \in \{1, \cdots, M\}$

$$\sum_{s:\nu_s \in \{I_j : j \geq i\}, \mathcal{D}_s \cap \{I_j : j < i\} \neq \emptyset,} R_s \leq C_{i,i-1}, \tag{H.2}$$
$$\sum_{s:\nu_s \in \{I_j : j \leq i\}, \mathcal{D}_s \cap \{I_j : j > i\} \neq \emptyset,} R_s \leq C_{i,i+1}.$$

Proof We will closely follow the argument of the proofs of Section B from Chapter III. First we prove that bounds (H.1) and (H.2) provide upper bounds on the set of achievable rate tuples. Construct network G' as explained in Section B (see Figure 11). We have the following entropy inequalities:

$$\begin{aligned}
&H\left(X_i^T, \left\{W_{s'}^T : \nu_{s'} = P_i\right\}, W_{Q_i \to I_i}^T\right)\\
&\overset{a}{\geq} H\left(\left\{W_{s'}^T : \nu_{s'} = P_i\right\}, W_{Q_i \to I_i}^T, X_{i,i-1}^T, X_{i,i+1}^T\right)\\
&\overset{b}{=} H\left(\left\{W_{s'}^T : P_i \in \mathcal{D}_{s'}\right\}, \left\{W_{s'}^T : \nu_{s'} = P_i\right\}, W_{Q_i \to I_i}^T, X_{i,i-1}^T, X_{i,i+1}^T\right)\\
&\overset{c}{\geq} H\left(\left\{W_{s'}^T : \nu_{s'} = P_i\right\}, \left\{W_{s'}^T : \nu_{s'} = Q_i\right\}, X_{i,i-1}^T, X_{i,i+1}^T\right)\\
&\overset{d}{\geq} H\left(\left\{W_{s'}^T : \nu_{s'} = P_i\right\}, \left\{W_{s'}^T : \nu_{s'} = Q_i\right\}, \left\{W_{s'}^T : \nu_{s'} = I_i\right\}\right)\\
&\overset{e}{=} H\left(\left\{W_{s'}^T : s' \in S'\right\}\right). \qquad \text{(H.3)}
\end{aligned}$$

In (H.3) a holds because $X_{i,i-1}^T$ and $X_{i,i+1}^T$ are functions of X_i^T. b holds because the message set $\left\{W_{s'}^T : P_i \in \mathcal{D}_{s'}\right\}$ is a function of $\left\{W_{s'}^T : \nu_{s'} = P_i\right\}$ and $X_{i,i-1}^T$. c holds because $\left\{W_{s'}^T : \nu_{s'} = Q_i\right\}$ is a subset of $\left\{W_{Q_i \to I_i}^T\right\} \cup \left\{W_{s'}^T : P_i \in \mathcal{D}_{s'}\right\}$. d holds because if a message in $\left\{W_{s'}^T : \nu_{s'} = I_i\right\}$ is decoded at P_i then it is a function of $\left\{W_{s'}^T : \nu_{s'} = P_i\right\}, X_{i,i-1}^T$ and if it is decoded at Q_i then it is a function of $\left\{W_{s'}^T : \nu_{s'} = Q_i\right\}, X_{i,i+1}^T$. Finally e holds because $\left\{W_{s'}^T : \nu_{s'} = P_i\right\} \cup \left\{W_{s'}^T : \nu_{s'} = Q_i\right\} \cup \left\{W_{s'}^T : \nu_{s'} = I_i\right\}$ is the set of all messages in the network.

By (H.3) we have

$$H\left(X_i^T, \left\{W_{s'}^T : \nu_{s'} = P_i\right\}, W_{Q_i \to I_i}^T\right) \geq H\left(\left\{W_{s'}^T : s' \in S'\right\}\right).$$

From the preceding ineqality and

$$H\left(X_i^T\right) + H\left(\left\{W_{s'}^T : \nu_{s'} = P_i\right\}, W_{Q_i \to I_i}^T\right) \geq H\left(X_i^T, \left\{W_{s'}^T : \nu_{s'} = P_i\right\}, W_{Q_i \to I_i}^T\right)$$

we obtain

$$H\left(X_i^T\right) \geq H\left(\left\{W_{s'}^T : s' \in S'\right\}\right) - H\left(\left\{W_{s'}^T : \nu_{s'} = P_i\right\}, W_{Q_i \to I_i}^T\right).$$

The right hand side of the preceding inequality is

$$T\left(\sum_{s' \in S'} R_{s'} - \sum_{s' : \nu_{s'} = P_i} R_{s'} - R_{Q_i \to I_i}\right) = T\left(\sum_{s' \in S' \setminus (\{s' : \nu_{s'} = P_i\} \cup \{Q_i \to I_i\})} R_{s'}\right).$$

Notice that $S' \setminus (\{s' : \nu_{s'} = P_i\} \cup \{Q_i \to I_i\})$ includes all sessions with source at I_i and all sessions with source at Q_i that have P_i as a destination. Thus, the right hand side of the preceding inequality is $T\left(\sum_{\nu_{s'} = I_i} R_{s'} + \sum_{\nu_{s'} = Q_i, P_i \in \mathcal{D}_{s'}} R_{s'}\right)$. Therefore we have

$$\begin{aligned} TC_i &\geq H\left(X_i^T\right) \\ &\geq T\left(\sum_{\nu_{s'} = I_i} R_{s'} + \sum_{\nu_{s'} = Q_i, P_i \in \mathcal{D}_{s'}} R_{s'}\right). \end{aligned} \tag{H.4}$$

By the symmetry of the network with respect to P_i and Q_i we likewise have:

$$C_i \geq \sum_{\nu_{s'} = I_i} R_{s'} + \sum_{\nu_{s'} = P_i, Q_i \in \mathcal{D}_{s'}} R_{s'}. \tag{H.5}$$

If we replace $R_{s'}$ by $\sum_{s \in S_{s'}} R_s$ in the bounds in (H.4), (H.5), we obtain the bound (H.1) in the original network G.

The bounds (H.2) are cut set bounds analogous to the bounds (3.2) in Section B.

Next we describe the network coding scheme. We write

$$\begin{aligned} W_{L_i \to R_i}^{(t)} &= \left[W_{L_i \to i, R_i}^{(t)}, W_{L_i \to \bar{i}, R_i}^{(t)}\right], \\ W_{R_i \to L_i}^{(t)} &= \left[W_{R_i \to i, L_i}^{(t)}, W_{R_i \to \bar{i}, L_i}^{(t)}\right]. \end{aligned}$$

The network coding scheme is indicated by the vectors $X_i^{(t)}, X_{i,i-1}^{(t)}$, and $X_{i,i+1}^{(t)}$ as follows:

$$X_i^{(t)} \triangleq \left[W_{R_i \to L_i}^{(t)} \otimes W_{L_i \to R_i}^{(t)}, W_{i \to L_i, R_i}^{(t)}, W_{i \to \bar{L}_i, R_i}^{(t)}, W_{i \to L_i, \bar{R}_i}^{(t)}\right], \tag{H.6}$$

$$X_{i,i-1}^{(t)} \triangleq \left[W_{R_i \to L_i}^{(t)} \oplus W_{L_i \to R_i}^{(t)}, W_{i \to L_i, \bar{R}_i}^{(t)}, W_{i \to L_i, R_i}^{(t)}\right], \tag{H.7}$$

$$X_{i,i+1}^{(t)} \triangleq \left[W_{L_i \to R_i}^{(t)} \oplus W_{R_i \to L_i}^{(t)}, W_{i \to \bar{L}_i, R_i}^{(t)}, W_{i \to L_i, R_i}^{(t)}\right]. \tag{H.8}$$

Next we will verify the validity of our network coding scheme. Recall from Section B that

$$F_i^{(t)} = \left[W_{R_i \to L_i}^{(t)}, W_{R_i \to i, \bar{L}_i}^{(t)}\right], \tag{H.9}$$

$$G_i^{(t)} = \left[W_{L_i \to R_i}^{(t)}, W_{L_i \to i, \bar{R}_i}^{(t)}\right]. \tag{H.10}$$

We prove that at time instant t, vectors $F_i^{(t)}$ and $G_i^{(t)}$ as defined in (H.9) and (H.10) respectively and $X_i^{(t)}$ as defined in (H.6) are functions of the information available to node I_i at time instant t and $X_{i,i-1}^{(t)}$ and $X_{i,i+1}^{(t)}$ as defined in (H.7) and (H.8) respectively, are functions of the information available to node O_i at time instant t. For $t = 0$ the claim holds trivially. Suppose that for all time instants $t \leq n-1$ and all nodes $i \in \{1, \cdots, M\}$ our claim holds. Next consider time instant $t = n$. Recall the following relationships from Section B:

$$\begin{aligned} F_i^{(t)} &= \left[W_{R_i \to L_i}^{(t)}, W_{R_i \to i, \bar{L}_i}^{(t)}\right] \text{ is a permutation of} \\ &\left[W_{R_{i+1} \to L_{i+1}}^{(t-1)}, W_{i+1 \to L_{i+1}, R_{i+1}}^{(t-1)}, W_{i+1 \to L_{i+1}, \bar{R}_{i+1}}^{(t-1)}\right], \end{aligned} \tag{H.11}$$

$$\begin{aligned} G_i^{(t)} &= \left[W_{L_i \to R_i}^{(t)}, W_{L_i \to i, \bar{R}_i}^{(t)}\right] \text{ is a permutation of} \\ &\left[W_{L_{i-1} \to R_{i-1}}^{(t-1)}, W_{i-1 \to L_{i-1}, R_{i-1}}^{(t-1)}, W_{i-1 \to R_{i-1}, \bar{L}_{i-1}}^{(t-1)}\right]. \end{aligned} \tag{H.12}$$

We show how I_i can decode the right hand side of (H.11) and (H.12) at time instant n. By the inductive hypothesis, at time instant n node I_i has access to the vectors $X_{i-1,i}^{(n-1)}$ and

$X_{i+1,i}^{(n-1)}$,

$$X_{i+1,i}^{(n-1)} = \left[W_{R_{i+1}\to L_{i+1}}^{(n-1)} \oplus W_{L_{i+1}\to R_{i+1}}^{(n-1)}, W_{i+1\to L_{i+1},\bar{R}_{i+1}}^{(n-1)}, W_{i+1\to L_{i+1},R_{i+1}}^{(n-1)}\right], \tag{H.13}$$

$$X_{i-1,i}^{(n-1)} = \left[W_{L_{i-1}\to R_{i-1}}^{(n-1)} \oplus W_{R_{i-1}\to L_{i-1}}^{(n-1)}, W_{i-1\to \bar{L}_{i-1},R_{i-1}}^{(t)}, W_{i-1\to L_{i-1},R_{i-1}}^{(t)}\right], \tag{H.14}$$

and can extract the messages

$$\left[W_{i+1\to L_{i+1},R_{i+1}}^{(n-1)}, W_{i+1\to L_{i+1},\bar{R}_{i+1}}^{(n-1)}\right],$$
$$\left[W_{i-1\to L_{i-1},R_{i-1}}^{(n-1)}, W_{i-1\to \bar{L}_{i-1},R_{i-1}}^{(n-1)}\right].$$

The two remaining messages that I_i needs to decode are $W_{R_{i+1}\to L_{i+1}}^{(n-1)}$ and $W_{L_{i-1}\to R_{i-1}}^{(n-1)}$, and we next describe the process to do this. By our inductive hypothesis, at $t = n-2$ node I_i knows the vectors

$$F_i^{(n-2)} = \left[W_{R_i\to L_i}^{(n-2)}, W_{R_i\to i,\bar{L}_i}^{(n-2)}\right],$$
$$G_i^{(n-2)} = \left[W_{L_i\to R_i}^{(n-2)}, W_{L_i\to i,\bar{R}_i}^{(n-2)}\right].$$

Observe that node I_i knows message vector

$$\left[W_{i\to L_i,R_i}^{(n-2)}, W_{i\to \bar{L}_i,R_i}^{(n-2)}, W_{i\to L_i,\bar{R}_i}^{(n-2)}\right]$$

at time instant n. Therefore I_i knows the vectors

$$\left[W_{R_i\to L_i}^{(n-2)}, W_{i\to L_i,R_i}^{(n-2)}, W_{i\to L_i,\bar{R}_i}^{(n-2)}\right],$$
$$\left[W_{L_i\to R_i}^{(n-2)}, W_{i\to L_i,R_i}^{(n-2)}, W_{i\to \bar{L}_i,R_i}^{(n-2)}\right]$$

at time instant n. Set t to $n-1$ and i to $i-1$ in (H.11), and set t to $n-1$ and i to $i+1$ in (H.12). Then (H.11) and (H.12) imply that the preceding vectors are permutations of the following vectors and thus at time instant n they are available to node I_i:

$$F_{i-1}^{(n-1)} = \left[W_{R_{i-1}\to L_{i-1}}^{(n-1)}, W_{R_{i-1}\to i-1,\bar{L}_{i-1}}^{(n-1)}\right],$$
$$G_{i+1}^{(n-1)} = \left[W_{L_{i+1}\to R_{i+1}}^{(n-1)}, W_{L_{i+1}\to i+1,\bar{R}_{i+1}}^{(n-1)}\right].$$

From $F_{i-1}^{(n-1)}$ and $G_{i+1}^{(n-1)}$, I_i obtains $W_{R_{i-1}\to L_{i-1}}^{(n-1)}$ and $W_{L_{i+1}\to R_{i+1}}^{(n-1)}$. Since by (H.13) and (H.14) I_i can extract $W_{R_{i+1}\to L_{i+1}}^{(n-1)} \oplus W_{L_{i+1}\to R_{i+1}}^{(n-1)}$ and $W_{L_{i-1}\to R_{i-1}}^{(n-1)} \oplus W_{R_{i-1}\to L_{i-1}}^{(n-1)}$ at time instant n from $X_{i+1,i}^{(n-1)}$ and $X_{i-1,i}^{(n-1)}$, respectively, it can decode $W_{R_{i+1}\to L_{i+1}}^{(n-1)}$ and $W_{L_{i-1}\to R_{i-1}}^{(n-1)}$. Thus I_i can decode

$$\left[W_{R_{i+1}\to L_{i+1}}^{(n-1)}, W_{i+1\to L_{i+1},R_{i+1}}^{(n-1)}, W_{i+1\to L_{i+1},\bar{R}_{i+1}}^{(n-1)}\right],$$
$$\left[W_{L_{i-1}\to R_{i-1}}^{(n-1)}, W_{i-1\to L_{i-1},R_{i-1}}^{(n-1)}, W_{i-1\to \bar{L}_{i-1},R_{i-1}}^{(n-1)}\right],$$

or equivalently

$$F_i^{(n)} = \left[W_{R_i\to L_i}^{(n)}, W_{R_i\to i,\bar{L}_i}^{(n)}\right],$$
$$G_i^{(n)} = \left[W_{L_i\to R_i}^{(n)}, W_{L_i\to i,\bar{R}_i}^{(n)}\right],$$

at time instant n, as desired. From $(H.6)$ it is simple to check that $X_i^{(n)}$ is a function of $F_i^{(n)}$ and $G_i^{(n)}$ and the messages generated at I_i at time instance n. Therefore 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 X_i^n. By (H.6), (H.7) and (H.8) we see that O_i only needs to construct $W_{R_i\to L_i}^{(n)} \oplus W_{L_i\to R_i}^{(n)}$ and $W_{L_i\to R_i}^{(n)} \oplus W_{R_i\to L_i}^{(n)}$ to be able to transmit $X_{i,i-1}^{(n)}$ and $X_{i,i+1}^{(n)}$. Observe that $W_{R_i\to L_i}^{(n)} \oplus W_{L_i\to R_i}^{(n)}$ and $W_{L_i\to R_i}^{(n)} \oplus W_{R_i\to L_i}^{(n)}$ can be obtained from $W_{R_i\to L_i}^{(n)} \otimes W_{L_i\to R_i}^{(n)}$, which is a component of $X_i^{(n)}$.

We next must show that the receiver node, I_i, is able to successfully decode all messages destined for node i in the original network. Observe that $\left[W_{R_i\to i,L_i}^{(n)}, W_{R_i\to i,\bar{L}_i}^{(n)}\right]$ is a part of vector $F_i^{(n)}$ and $\left[W_{L_i\to i,R_i}^{(n)}, W_{L_i\to i,\bar{R}_i}^{(n)}\right]$ is a part of vector $G_i^{(n)}$. $F_i^{(n)}$and $G_i^{(n)}$ can be decoded at I_i at time instant n. Therefore if a message originates at source I_j at time instant $n-|j-i|$ with a destination at I_i, it can be decoded at time instant n at I_i.

Next we demonstrate that any non–negative rate tuple $(R_s : s \in S)$ that satisfies the bounds in Theorem 47 results in $X_i^{(t)}, X_{i,i-1}^{(t)}$ and $X_{i,i+1}^{(t)}$ which have the properties $H\left(X_i^{(t)}\right) \le C_i, H\left(X_{i,i-1}^{(t)}\right) \le C_{i,i-1}$, and $H\left(X_{i,i+1}^{(t)}\right) \le C_{i,i+1}$, and thus can be supported on this network.

By (H.6) we have:

$$H\left(X_i^{(t)}\right) = H\left(W_{R_i\to L_i}^{(t)} \otimes W_{L_i\to R_i}^{(t)}\right) + H\left(W_{i\to L_i,R_i}^{(t)}\right) + H\left(W_{i\to \bar{L}_i,R_i}^{(t)}\right) + H\left(W_{i\to L_i,\bar{R}_i}^{(t)}\right) \tag{H.15}$$

It follows from our earlier definitions that

$$\sum_{\nu_s=I_i} R_s = H\left(W_{i\to L_i,R_i}^{(t)}\right) + H\left(W_{i\to \bar{L}_i,R_i}^{(t)}\right) + H\left(W_{i\to L_i,\bar{R}_i}^{(t)}\right),$$
$$\sum_{s\in U_1} R_s = H\left(W_{L_i\to R_i}^{(t)}\right),$$
$$\sum_{s\in U_2} R_s = H\left(W_{R_i\to L_i}^{(t)}\right).$$

Since $H(\mathbf{a}\otimes\mathbf{b}) = \max\{H(\mathbf{a}), H(\mathbf{b})\}$, (H.15) gives

$$H\left(X_i^{(t)}\right) = \sum_{\nu_s=I_i} R_s + \max\left\{\sum_{s\in U_1} R_s, \sum_{s\in U_2} R_s\right\} \le C_i.$$

By (H.7) we have

$$H\left(X_{i,i-1}^{(t)}\right) = H\left(W_{R_i\to L_i}^{(t)} \oplus W_{L_i\to R_i}^{(t)}\right) + H\left(W_{i\to L_i,\bar{R}_i}^{(t)}\right) + H\left(W_{i\to L_i,R_i}^{(t)}\right). \tag{H.16}$$

Since $H\left(W_{R_i\to L_i}^{(t)} \oplus W_{L_i\to R_i}^{(t)}\right) = H\left(W_{R_i\to L_i}^{(t)}\right)$ we obtain

$$H\left(X_{i,i-1}^{(t)}\right) = \sum_{s:\nu_s\in\{I_j:j\ge i\},\mathcal{D}_s\cap\{I_j:j<i\}\ne\emptyset,} R_s \le C_{i,i-1}.$$

With a similar argument we obtain

$$H\left(X_{i,i+1}^{(t)}\right) = \sum_{s:\nu_s \in \{I_j : j \le i\}, \mathcal{D}_s \cap \{I_j : j > i\} \neq \emptyset,} R_s \le C_{i,i+1}.$$

$\square$

Notice that in the wired model of the wireless line described in the introduction, node O_i is required to only forward the data to the other nodes and not to perform any network coding operation on the incoming flow to produce the outgoing flow. This requirement is satisfied by our network coding scheme since the vectors $X_{i,i-1}^{(n)}$ and $X_{i,i+1}^{(n)}$ are both parts of the vector $X_i^{(n)}$. Therefore our scheme is applicable to wireless network coding.

VITA

Seyed Mohammadsadegh Tabatabaei Yazdi received the B.S. degree in Electrical Engineering from Sharif University of Technology, Tehran, Iran, in 2006. He received the M.S. degree in Electrical Engineering from University of Michigan, Ann Arbor, MI, in 2007, and the Ph.D. degree in Electrical Engineering from Texas A&M University in 2011.

His research mainly focuses on network coding, wireless networks, and combinatorial optimization.

He may be reached at 231B WERC, Texas A&M University, College Station, TX, 77843. His email is sadegh.tamu@gmail.com.

The typist for this thesis was Seyed Mohammadsadegh Tabatabaei Yazdi.