Analysis and design of multi-session network coding

Xu, Xiaoli

2015

Xu, X. (2015). Analysis and design of multi-session network coding. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/62567>

<https://doi.org/10.32657/10356/62567>

ANALYSIS AND DESIGN OF MULTI-SESSION NETWORK CODING

XU Xiaoli

School of Electrical and Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

2015

Acknowledgements

I owe my deepest gratitude to my Ph.D. supervisor Prof. Yong Liang Guan for his exemplary guidance, consistent patience, encouragements and supports. His invaluable advise on technical as well as professional matters shaped me as a researcher. The way that he manage the multitude of demands on his time always amaze and inspire me. I have been extremely lucky to have a supervisor who cared so much about my work, and who can always provided the most insightful comments and constructive suggestions quickly. Without his persistent help, this thesis would not have been possible.

I would like to express my heartfelt gratitude and thanks to Prof. Satyajit Thakor from Indian Institute of Technology Mandi, for his willing to collaborate and insightful comments and suggestions on my research work. The work on reduced FDG and refined FD bound are motivated by his celebrated work “Network coding capacity: a functional dependence bound” published in ISIT 2009. Prof. Thakor has provided consistent guidance on my research work since we met in NetCod 2011. His insightful comments and hard questions have very much steered the directions of my work and made my work possible.

I would also like to sincerely thank Prof. Tracy Ho from California Institute of Technology for the helpful discussions and suggestions on my work, especially on the topic “linear network coding schemes for double-unicast networks”. I have benefited immensely from her constructive criticism on my work and her solid knowledge on network coding.

My gratitude also goes to Prof. Dina Katabi from Massachusetts Institute of Technology and Prof. Tiffany Jing Li from Lehigh University for their helpful discussions and guidance on my research work when they visited NTU. I would also like to thank my colleagues in INFINITUS lab at NTU, Dr. Zilong Liu, Dr. Xinmei Huang and Dr. Guojun Han for the contributive discussions now and then.

Special thanks go to my collaborator Mr. Meenakshi Sundaram Gandhi Praveen Kumar. His solid knowledge in hardware and continuous hardworking made our network coding testbed possible. It is a pleasure experience to work with him.

Last but not least, I wish to express my heartfelt thanks to my husband Dr. Zeng Yong for his continuous support over the years. He is not only my soul mate in life but also a good collaborator for my research. Without him, this thesis would certainly have been impossible.

Xu Xiaoli

August 2014, Singapore

Abstract

By enabling intermediate nodes to combine the received messages before forwarding, network coding is able to achieve higher network throughput than conventional routing. For single-session networks, the network coding capacity is given by the celebrated max-flow min-cut theorem and achieved by simple linear network coding techniques. However, characterizing and approaching the network coding capacity for general multi-session networks is still an open problem. Many existing results suggest that this problem is extremely hard and hence studying network coding for multi-session networks relies heavily on analyzing the outer and inner bounds of network coding capacity. This thesis is devoted to the analysis and design of multi-session network coding, targeting research problems such as the simplification of outer bounds based on the functional dependence graph (FDG), the derivation of improved outer bounds for multi-session network coding capacity, the design of network code for special multi-session networks, and the throughput analysis of erasure networks.

FDG is an important class of directed graph that captures the functional dependence relationships of a set of random variables of a network, which has been frequently used in characterizing network coding capacity bounds. Since the computational complexity of network coding capacity bounds generally grows exponentially with the order of the FDG of a network, it is desirable to find an FDG with the smallest size possible. In the first part of this thesis, some systematic graph reduction techniques are presented to significantly reduce the FDG size. The resulting reduced FDGs are proven to give the same network coding capacity regions/bounds as that obtained by using the original FDGs, but with much less computations.

Most of the practical outer bounds of network coding capacity, such as the cut-set bound, the network sharing bound and the functional dependence (FD)

bound, are developed based on identifying the information bottlenecks in the network, hence they are usually referred to as “edge-cut bounds”. One major limitation of such edge-cut bounds is that every source is treated “equally”, in the sense that the same weight is assigned for each source rate. As a consequence, edge-cut bounds are usually quite loose even for simple double-unicast networks. In the second part of this thesis, we derive a new lower bound on the mutual information between any two subsets of source/edge random variables in a network. The lower bound on the mutual information is obtained based on the functional dependence relationships among the source/edge random variables in the network, hence it is applicable for any networks, regardless of whether the receivers have overlapping demands. Based on the developed lower bound on the mutual information, a systematic approach is presented to obtain tighter FD bounds than those in the literature.

While various outer bounds of network coding capacity have been proposed, only very limited results on the inner bounds are available for general networks. Inner bounds, or achievable rate regions, are challenging to obtain in general, as network code design depends heavily on the network topologies. In this thesis, network codes are designed for double-unicast networks assuming random linear network coding are performed at the intermediate nodes. The main approach is to optimize the network-layer precoding and decoding at the source and sink nodes, respectively, to maximize the achievable rate region. The obtained achievable region is shown to be strictly larger than the existing achievable regions in the literature. Furthermore, we show that this maximum achievable region which is obtained with the assumption of random linear network coding can be further enlarged by appropriately incorporating routing at the intermediate nodes.

In the last part of this thesis, we study the throughput of erasure networks where each link has an independent erasure probability. To achieve the capacity of such erasure networks, spatial-temporal network coding, i.e., coding across symbols

received from different network edges as well as at different time slots, with infinite temporal coding length is generally required. However, in practice, only finite temporal coding length is feasible due to constraints such as tolerable delay, buffer size, and acceptable coding complexity. A practical question to answer is thus what minimum temporal coding length is needed so as to achieve a targeted percentage of the network capacity. To this end, we derive the throughput of the single-session multicast erasure networks applied with spatial-temporal network coding as a function of the temporal coding length. Numerical examples demonstrate a very good match between the theoretical and simulation results. The analytical result is also shown to be applicable for multi-source single-sink erasure networks.

Abbreviations

BATS	BATched Sparse Codes.
BP	Belief Propagation.
FD	Functional Dependence.
FDG	Functional Dependence Graph.
IA	Interference Alignment.
IC	Interference Channel.
i.i.d.	independent and identically distributed.
J-R-RLNC	Joint Routing and Random Linear Network Coding.
LP	Linear Programming.
MIMO	Multiple Input Multiple Output.
PdE	Progressive d -separating Edge-set.
RLNC	Random Linear Network Coding.

Symbols

Throughout this thesis, scalars are denoted by italic letters. Boldface lower- and upper-case letters denote vectors and matrices, respectively.

$\mathbb{R}$	The field of real numbers.
$\mathbb{F}_q$	Finite field of order q .
$\mathbb{F}_q^n$	The space of $n \times 1$ vectors in a finite field of order q .
$\mathbb{F}_q^{n \times m}$	The space of $n \times m$ matrices in a finite field of order q .
$\oplus$	Exclusive OR.
Γ^*	The set of all entropic vectors.
Γ	The set of all polymatroidal pseudo-entropy vectors.
$G(V, E)$	Graph with set of nodes V and set of edges E .
$e = (u, v)$	Edge from node u to node v .
C_e	The capacity of edge e .
$\text{tail}(e)$	Tail node of edge e .
$\text{head}(e)$	Head node of edge e .
$\text{In}(v)$	The set of edges entering node v .
$\text{Out}(v)$	The set of edges leaving node v .
$\text{Up}(v)$	The set of immediate upstream nodes of node v .
$\text{Down}(v)$	The set of immediate downstream nodes of node v .
$\mapsto$	Mapping to.
$\text{proj}_B(A)$	The projection of A on the coordinates of B .
$\text{con}(A)$	The convex hull of A .

$\overline{A}$	The closure of A .
$\bar{G}(\bar{V}, \bar{E})$	A functional dependence graph for a set of random variables $\bar{V}$.
Y_s	Source random variable.
U_e	Edge random variable.
$ A $	Cardinality of set A .
X_A	A set of variables whose subscripts are in set A , i.e., $\{X_a : a \in A\}$.
$H(\cdot)$	Entropy function.
$H(\cdot \cdot)$	Conditional entropy.
$I(\cdot;\cdot)$	Mutual information.
$A \rightarrow B$	A determines B .
$\phi(A)$	The largest set of nodes with $A \rightarrow \phi(A)$.
A^c	The complementary set of A .
$\mathbf{A}^T$	The transpose of matrix $\mathbf{A}$.
$\mathbf{A}^{-1}$	The inverse of matrix $\mathbf{A}$.
$\text{rank}(\mathbf{A})$	The rank of matrix $\mathbf{A}$.
$\otimes$	Kronecker product.
$\mathbf{I}_n$	An identity matrix of size $n \times n$.
$\mathbf{0}_{n \times m}$	A zero matrix of size $n \times m$.
$\mathcal{R}(\mathbf{A})$	The range (or column space) of matrix $\mathbf{A}$.
$\mathcal{N}(\mathbf{A})$	The null space of matrix $\mathbf{A}$.
$(x)^+$	$(x)^+ \triangleq \max\{0, x\}$.
$\mathbb{E}[\cdot]$	Expectation.
$\Pr(\cdot)$	Probability of an event.
$\mathcal{N}(\mu, \sigma^2)$	Gaussian distribution with mean μ and variance σ^2 .
$\mathcal{B}(M, p)$	Binomial distribution with parameters M and p .

Contents

Acknowledgements	iii
Abstract	v
Abbreviations	viii
Symbols	xi
List of Figures	xvii
List of Tables	xix
1 Introduction	1
1.1 Motivation and Objectives	1
1.1.1 Outer Bounds for Network Coding Capacity	3
1.1.2 Network Code Construction for Multi-session Networks . . .	4
1.1.3 Network Coding for Erasure Networks	5
1.2 Major Contribution of the Thesis	5
1.2.1 Reduced Functional Dependence Graph	6
1.2.2 Multi-session Network Coding Capacity Bounds	7
1.2.3 Linear Network Coding Schemes for Special Classes of Multi-session Networks	7
1.2.4 Performance Analysis of Finite-Length Spatial-Temporal Net- work Coding	8
1.3 Organization of the Thesis	9
2 An Overview on Network Coding	11
2.1 Network Model	11
2.2 Network Coding Capacity	12
2.3 Linear Programming Bound	14
2.4 Functional Dependence Graph and Functional Dependence Bound .	15
2.5 Linear Network Coding	18
2.5.1 An Algebraic Formulation of Linear Network Coding	19
2.5.2 Random Linear Network Coding	20
3 Reduced Functional Dependence Graphs	25

3.1	Backgrounds	27
3.1.1	Hypernetwork	27
3.1.2	Basic Network Simplifications	27
3.2	Main Results	29
3.2.1	Edge Elimination in the Network	29
3.2.2	Edge Reduction in the FDG	32
3.2.3	An Iterative Algorithm for Reduced FDG	33
3.2.4	Extensions of Proposed Reduction Techniques	34
3.3	Further FDG Reduction for Linear Network Coding	36
3.4	Performance Evaluation	38
3.5	Application of Reduced FDG	40
3.5.1	Complexity Reduction for LP Bound Computation	41
3.5.2	Complexity Reduction in Algebraic Formulation of Linear Network Coding	43
3.6	Chapter Summary	45
4	Refined Functional Dependence Bounds for Network Coding Ca- pacity	47
4.1	Related Works	48
4.1.1	Refined Edge-Cut Bound	49
4.1.2	Weighted Sum-rate FD Bound	50
4.2	Main Result	50
4.2.1	A Lower Bound on Mutual Information	51
4.2.2	Refined FD Bound	53
4.3	An Algorithm for the Refined FD Bounds	56
4.3.1	Guidelines for Determining A and B	56
4.3.2	Determining the Capacities of C and D	60
4.3.3	A Summarized Algorithm	61
4.4	Application on Degraded Multicast Networks	62
4.5	Chapter Summary	64
5	Linear Network Coding Schemes for Multi-Session Networks	67
5.1	Achievable Region for Double-Unicast Networks with Linear Network Coding	67
5.1.1	System Model	70
5.1.2	A Linear Network Coding Scheme	71
5.1.3	An Achievable Rate Region in Terms of Min-cuts	85
5.2	Joint Routing and RLNC	91
5.2.1	J-R-RLNC for Degraded 2-message Multicast Networks	92
5.2.2	J-R-RLNC for Double-Unicast Networks	97
5.2.3	Extension to General Multi-session Networks	99
5.3	Chapter Summary	99
6	Spatial-Temporal Network Coding for Erasure Networks	101

6.1	Throughput Analysis of Finite-Length Spatial -Temporal Network Coding	102
6.1.1	Illustrative Example: Line Network	103
6.1.2	Single-Session Erasure Networks	104
6.1.3	Throughput Analysis	106
6.1.4	A Simple Lower Bound	112
6.1.5	Numerical Examples	113
6.1.6	An Achievable Scheme	114
6.2	Multi-Source Single-Sink Erasure Networks	117
6.2.1	System Model	118
6.2.2	A Random Linear Network Coding Approach	119
6.2.3	Throughput with Finite-Length Spatial-Temporal Network Coding	122
6.3	Chapter Summary	125
7	Conclusions and Future Work	127
7.1	Conclusions	127
7.2	Future Work	129
	Author's Publications	131
	Bibliography	133
A	A Counter-Example of Corollary 3.10 for Nonlinear Network Coding	145
B	Proof of Lemma 5.6	149
C	Detailed Steps of the Fourier-Motzkin Elimination	153
C.1	List of Facts	161
D	Detailed Steps for Deriving (6.3)	165
E	Number Partitions	167

List of Figures

1.1	(a) The bottleneck link is used to route x_1 . (b) The bottleneck link is used to route x_2 . (c) Network coding is applied.	2
1.2	(a) A single-session multicast network. (b) A multi-session multicast network.	2
2.1	(a) A 2-unicast butterfly network. (b) FDG of the butterfly network.	16
2.2	(a) A simple network constructed in [1]. (b) The corresponding directed line graph.	21
2.3	A single-session multicast network with RLNC.	22
3.1	An example of hypernetwork.	27
3.2	(a) Two Y-network with shared source and sink nodes. (b) Two source nodes communicating with a common sink. (c) A Y-network.	29
3.3	(a) The original network G . (b) The reduced network $\{G - e\}$	29
3.4	(a) The FDG of the original network. (b) The FDG of the reduced network.	30
3.5	(a) An example network. (b) The reduced network after edge elimination with Lemma 3.2. (c) The FDG of the original network. (d) The FDG of the reduced network.	31
3.6	(a) The reduced FDG from Figure 5.2(d). (b) The corresponding reduced network.	33
3.7	(a) A component of an FDG. (b) The reduced component of an FDG.	35
3.8	(a) A 3-unicast network. (b) The reduced FDG from Algorithm 1. (c) The FDG with further reduction by applying Corollary 3.10. . .	38
3.9	Performance evaluation of the proposed reduction techniques on randomly generated networks.	40
3.10	Performance evaluation of the proposed reduction techniques for linear network coding.	41
3.11	(a) The butterfly network. (b) The corresponding FDG. (c) The reduced FDG with reduction techniques proposed in Corollary 3.3. (d) The reduced FDG with extended reduction techniques proposed in Proposition 3.8.	43
3.12	A network that is solvable only over fields with characteristic 2 . . .	45
4.1	A network with overlapping demands.	49
4.2	A 2-unicast network with unit edge capacity.	55
4.3	(a) A network example. (b) The corresponding reduced FDG. . . .	59

4.4	3-user degraded multicast network.	63
4.5	3-user degraded multicast combinational network.	65
4.6	Compare the edge-cut bound with the refined FD bound for a 3-user degraded multicast combinational network.	66
5.1	The proposed achievability scheme with linear precoding and decoding at s_1 and t_1 , respectively.	71
5.2	(a) An example of double-unicast networks. (b) The achievable rate regions.	83
5.3	(a) The grail network. (b) The achievable regions.	86
5.4	3-user degraded 2-message multicast network.	92
5.5	(a) A 3-user degraded 2-message multicast network example. (b) The achievable regions obtained by RLNC and J-R-RLNC.	95
5.6	(a) An example of double-unicast networks. (b) The achievable region with RLNC and J-R-RLNC.	99
6.1	A two-hop line network with packet loss rate p	103
6.2	An intermediate node v with spatial-temporal network coding.	105
6.3	The Butterfly network.	109
6.4	Normalized throughput ρ_M versus network edge erasure probability p for various temporal coding length M	114
6.5	Normalized throughput ρ_M versus network edge erasure probability p for random networks.	115
6.6	The structure of BATS.	115
6.7	Compare the achieved throughput of BATS code with the analytical throughput for $M = 8$ in the butterfly network.	116
6.8	Compare the achieved throughput of BATS code with the analytical throughput for various M in two-hop line network.	117
6.9	A multi-source single-sink network.	118
6.10	The channel model of (6.26).	121
6.11	A 2-source single-sink erasure network.	124
6.12	Compare the analytical throughput with the simulation results.	124
6.13	A 2-source single-sink erasure network.	125
6.14	Compare the analytical throughput with the simulation results.	126
A.1	An counter example showing that Corollary 3.10 cannot be extended to nonlinear network coding. All the edges are of unit capacity and the independent message a, b, c, d, e are available at the source nodes $1 - 3, 11 - 12$. The sink nodes $37 - 46, 49$ each demand one of the messages, as indicated.	146
A.2	(a) A portion of the network shown in Fig. A.1. (b) The conventional FDG. (c) The reduced FDG by Corollary 3.10.	147
A.3	Coding for lossless computing with side information.	147

List of Tables

4.1	A list of refined FD bounds	62
6.1	List of all possible combinations of $\{\{\mathbf{Y}_i, \mathbf{Y}_j\}, 1 \leq i < j \leq 6\}$	109

Chapter 1

Introduction

1.1 Motivation and Objectives

In traditional communications, network messages are viewed as physical commodities which can only be routed throughout the network without alteration. In contrast, with the emerging technology of network coding [2, 3], network messages are treated as information and *allowed to be coded at the intermediate nodes*. The potential advantages of network coding over routing include the enhancement in network throughput [4–6], reliability [7–9], energy efficiency [10, 11], security [12] and others [13–16]. Due to these promising benefits, network coding has found a wide range of applications in ad hoc networks [17–19], wireless mesh networks [6, 20–23], peer-to-peer communication systems [24–26] and distributed data storage [27].

Network coding achieves higher throughput than routing by communicating more information via fewer packet transmissions. The most famous example for illustrating the throughput benefit of network coding is known as the butterfly network, which was constructed in [3] and reproduced in Figure 1.1. In this butterfly network, the source wants to send two symbols x_1 and x_2 simultaneously to two sink nodes t_1 and t_2 . With routing, the highlighted link $e = (v_3, v_4)$ in Figure 1.1 becomes a bottleneck of this network, which can either send x_1 or x_2 in one time slot. Hence, at least one of the sink node cannot recover both x_1 and x_2 . In contrast, with network coding, this bottleneck link can send a coded version of these two messages, i.e., $x_1 \oplus x_2$ as shown in Figure 1.1(c). Since t_1 has already received x_1 from s_1 directly, it can obtain x_2 by subtracting x_1 from the coded packet $x_1 \oplus x_2$. Meanwhile, t_2 can also recover x_1 in a similar manner.

FIGURE 1.1: (a) The bottleneck link is used to route x_1 . (b) The bottleneck link is used to route x_2 . (c) Network coding is applied.

The butterfly network indicates that network coding can increase the throughput of single-session multicast networks. Generally, a single-session multicast network, as shown in Figure 1.2(a), has a common source node and multiple sink nodes that demand the same set of information from the source. Hence, there is no interference within this network. On the other hand, a multi-session network, as shown in Figure 1.2(b), has more than one source nodes and different information may be wanted by different sink node. The unwanted information is known as inter-session interference.

FIGURE 1.2: (a) A single-session multicast network. (b) A multi-session multicast network.

For single-session multicast networks, the network capacity is given by the celebrated max-flow min-cut theorem and it is achieved by linear network coding [28, 29]. However, the max-flow min-cut bound is not tight for multi-session networks. Actually, characterizing the capacity region for general multi-session networks is a long standing open problem. An exact expression of capacity region for multi-session networks is given in terms of the entropy region Γ^* in [30]. Unfortunately, characterization of Γ^* is still open if there are more than three random variables involved. It was proved in [31] that characterizing the network coding capacity is as difficult as determining the solution of an infinite set of information inequalities [32]. Furthermore, linear network coding is not sufficient for achieving the capacity of general multi-session networks [33]. Due to the difficulty in explicit capacity characterization, analyzing the outer and inner bounds of the capacity is crucial for understanding network coding in multi-session networks.

1.1.1 Outer Bounds for Network Coding Capacity

By maximizing the weighted sum rates subject to the encoding/decoding constraints and the set of elemental Shannon type inequalities, a computable outer bound for the network coding, known as linear programming (LP) bound, can be obtained [34]. However, the complexity for computing the LP bound increases double exponentially with the network size, which hinders the LP bound to be used in most practical scenarios. On the other hand, most of easily-computable outer bounds are obtained based on graph theories, i.e., identifying the set of network edges that forms the bottleneck of the network. The most well-known graph theoretic bounds include the edge-cut bound [35], the network sharing bound [36], the progressive d -separating edge-set (PdE) bound [37] and the functional dependence (FD) bound [38]. In particular, the edge-cut bound is the direct extension of the max-flow min-cut bounds for single-session multicast networks. The network sharing bound is defined for a special type of multiple-unicast networks (the information generated by each source is demanded by exactly one sink node only) by utilizing the notion of information dominance introduced in [39]. The PdE bound is obtained by exhaustively searching all subsets of edges that disconnect the source nodes and the corresponding sink nodes in an ordered manner. In FD bound, the sum of the source rates is bounded by the capacity of the maximal irreducible sets in the functional dependence graph (FDG) of the network.

One major limitation of these graph-theoretic bounds is that every source is treated “equally”, in the sense that the same weight is assigned for each source rate. As a consequence, they are usually quite loose, even for simple networks [40]. In this thesis, we aim to improve these graph theoretic-bounds by introducing some weighted bounds and reduce the computational complexity for the aforementioned outer bounds.

1.1.2 Network Code Construction for Multi-session Networks

Unlike single-session multicast networks where simple random linear network coding (RLNC) [41] is capacity-achieving, constructing network code for multi-session networks is much more challenging. When there are multiple sessions sharing the network, we need to achieve a balance between maximizing the transmission rate for each session and minimizing the inter-session interference. One practical and simple approach is to restrict network coding to only within each session [42], which is known as *intra-session network coding*. However, in general, achieving the optimal rates requires *inter-session network coding*, i.e., combining the information symbols from different sessions.

It remains unknown how to construct the optimal network code for multi-session networks. In general, linear network coding is inadequate, as evident from a multi-session network constructed in [33] based on Fano and non-Fano matroids. However, due to complexity considerations, linear network coding is preferred in most practical scenarios, which thus motivates some suboptimal linear network coding schemes for multi-session networks. A “poison-antidote” approach was proposed in [43–45], which searches for the butterfly structures in the network using linear optimization method. Besides, Wang et al. introduced a pairwise inter-session network coding approach, where the coding operations is restricted to two symbols [46] to reduce the inter-session interference. Recently, Heindlmaier et al. proposed a scheme for inter-session network coding by creating virtual multicast sessions and applying RLNC for this setup [47].

All the aforementioned approaches are based on solving optimization problems whose complexity grows exponentially with the network size. Explicit construction of inter-session network coding for general multi-session networks is not available yet. However, if we restrict the number of sources and sinks, network code construction

can be made much simpler. For instance, RLNC with network precoding was shown to be optimal for networks with single source node and two sink nodes [48, 49]. For double-unicast networks, an achievable rate region can be obtained by “rate-exchange” method [50]. Furthermore, interference alignment was shown to achieve half of the min-cut for some three-unicast networks [51, 52]. In this thesis, we aim to construct explicit linear network code for special multi-session networks, such as double-unicast networks and 2-message degraded broadcast networks, that outperforms these existing schemes.

1.1.3 Network Coding for Erasure Networks

Although the concept of network coding was initially proposed for lossless networks, it has also been shown to be able to improve the performance of lossy networks with packet erasures [7, 53–55]. In particular, RLNC achieves the capacity of single-session unicast/multicast erasure networks, which is given by the capacity of the bottleneck edges subject to erasures [56]. To achieve network capacity, not only packets coming from different links, but also those coming from the same link at different time slots, need to be combined together. This technique is known as *spatial-temporal* network coding, which is more general than the conventional purely spatial network coding scheme. Besides, the length of the temporal network coding should be sufficiently large to ensure that the percentage of erased packets converges to the erasure probability at each link. However, in practice, the temporal coding length is usually limited to a small-to-moderate number due to constraints such as tolerable communication delay, available buffer size, and/or coding/decoding complexity. A practical network code for erasure networks has been proposed in [57] by utilizing a joint fountain code and random linear network code. In this thesis, we will investigate the optimal throughput of erasure networks as a function of the temporal coding length.

1.2 Major Contribution of the Thesis

The main contributions of this thesis are summarized as follows.

1.2.1 Reduced Functional Dependence Graph

Functional dependence graph (FDG) is an important class of directed graph that captures the functional dependence relationships of a set of random variables and it is frequently used in characterizing network coding capacity bounds. Since the computational complexity of such bounds in general grows exponentially with the order of the FDG, it is desirable to find an FDG with the minimum size. To this end, some systematic graph reduction techniques are proposed. It is proved that the reduced FDGs give the same network coding capacity regions/bounds as the original FDGs, yet much less computation is required to compute the regions/bounds.

We identify non-essential edges in the original network and show that eliminating these edges results in a hypernetwork which has the same capacity as the original network. Edge elimination in the network is equivalent to node reduction in the corresponding FDG. Besides, we show that certain edges in the FDG may also be removed without affecting the functional dependence relationships of the random variables. Then, an iterative algorithm has been designed in this thesis to utilize all the proposed reduction techniques and the complexity of this algorithm is discussed. We show that the FDG can be further reduced if *linear* network coding capacity is considered, as the case in many practical scenarios due to the simplicity of linear network coding schemes. However, such reduction cannot be extended to non-linear network coding, as proved by a counter-example we constructed. Extensive simulations have been performed to show that with the proposed techniques, substantial reduction can be achieved for the FDG of general networks.

The main results related to these reduced FDG works have been presented in the following conference and journal papers:

- **X. Xu**, S. Thakor, and Y. L. Guan, “Reduced Functional Dependence Graph and its Applications,” *IEEE International Symposium on Network Coding (NetCod)*, (Boston, USA), Jun. 2012.
- **X. Xu**, S. Thakor, and Y. L. Guan, “Reduced Functional Dependence Graphs,” *IET Networks*, vol. 3, no. 2, Jun. 2014.

1.2.2 Multi-session Network Coding Capacity Bounds

Most existing bounds treat all sources “equally”, hence it only gives sum-rate bounds which is usually quite loose for multi-session networks. In this thesis, we present a lower bound on the mutual information between two sets of source/edge random variables in a general multi-session network by exploring their functional dependence relationships. Based on the proposed lower bounds of mutual information, a new class of information-theoretic outer bounds for the network coding capacity is derived. These new outer bounds improve over the original functional dependence bounds by including different rate weights and/or improved sum-rate bounds.

The results for this part have been published in the following conference papers:

- **X. Xu**, S. Thakor, and Y. L. Guan, “Weighted Sum-rate Functional Dependence Bound for Network Coding Capacity,” in *International Symposium on Information Theory and its Applications (ISITA)* (Hawaii, USA), Oct., 2012.
- **X. Xu**, S. Thakor, Y. L. Guan, “On the Mutual Information between Random Variables in Networks”, *Information Theory Workshop*, (Seville, Spain), Sep., 2013.

1.2.3 Linear Network Coding Schemes for Special Classes of Multi-session Networks

First, we present an achievable rate region for double-unicast networks in which the intermediate nodes perform RLNC, and the source and sink nodes optimize their strategies in the form of network precoding and decoding to maximize the achievable region. Such a setup can be modeled as a deterministic interference channel, whose capacity region has been obtained in [58]. For the particular class of *linear* deterministic interference channels of our interest, in which the outputs and interference are *linear* deterministic functions of the inputs, we show that the known capacity region can be achieved by *linear* strategies, which has much lower complexity compared to the non-linear superposition random code in [58]. As a result, for any RLNC performed at the intermediate nodes, the proposed linear precoding and decoding for the source and sink nodes will give the maximum achievable rate region for double-unicast networks.

We further derive a suboptimal but easy-to-compute rate region that is independent of the network coding coefficients used at the intermediate nodes; instead it is specified by the min-cuts of the network. It is found that even this suboptimal region is strictly larger than the existing achievable rate regions reported in the literature.

Although the proposed source and sink node processing is optimal for double-unicast networks, RLNC alone may not always be the optimal strategy for the intermediate nodes in general, as the desired information and the interference are coupled to the largest extent. To further improve the achievable region, a joint routing and random linear network coding (J-R-RLNC) approach is introduced, where some source information is transmitted using routing, to reduce the inter-session interference, and the rest is randomly network coded at the intermediate nodes with proper network precoding at the source nodes. The path used for routing and the corresponding routing rate can be determined using Fourier-Motzkin eliminations. This concept of J-R-RLNC can be extended to general multi-session networks where an achievable region with RLNC can be obtained.

The main results for this part have been published in the following conference and journal papers:

- **X. Xu** and Y. L. Guan, “Joint Routing and Random Network Coding for Multi-session Networks”, *IEEE International Conference on Networks*, (Singapore), Dec., 2013.
- **X. Xu**, Y. Zeng, Y. L. Guan, and T. Ho “An Achievable Region for Double-Unicast Networks with Linear Network Coding,” *IEEE Transactions on Communications*, vol. 62, no. 10, pp. 3621 - 3630, Aug. 2014.

1.2.4 Performance Analysis of Finite-Length Spatial-Temporal Network Coding

To achieve the capacity of erasure networks, spatial-temporal network coding, i.e., coding for symbols both from different links as well as at different time slots, with infinite temporal length is generally required. However, in practice, only finite temporal coding length M is feasible due to considerations such as delay, buffer size, and coding complexity constraints. A practical question to answer is thus what minimum M is needed in order to achieve a target percentage of the network

capacity. To this end, we derive the throughput of single-session erasure networks with spatial-temporal network coding as a function of M . The obtained result is given in terms of the min-cut value k , the number of minimum cuts N , and the number of critical edges L , of the network. For low-complexity computation, we also derive a lower-bound of the expected throughput in terms of k and L only. Numerical examples demonstrate a very good match between the theoretical and simulated results. The batched sparse code (BATS code) proposed in [57] is shown by simulation to achieve the analytical throughput.

Furthermore, we show that the capacity of multi-source single-sink erasure networks can be achieved by random linear network coding with sufficiently large temporal coding length. For the case of finite temporal coding length M , the expected throughput of the multi-source single-sink networks is derived based on the analytical result obtained for the single-session erasure networks.

The main results have been published in the following journal paper:

- **X. Xu**, Y. Zeng, and Y. L. Guan, “Performance Analysis of Finite-Length Spatial-Temporal Network Coding,” *IEEE Communications Letters*, vol. 18, no. 7, pp. 1163-1166, Jul. 2014.

1.3 Organization of the Thesis

This thesis is organized as follows.

Chapter 1, the present chapter, gives the motivation, objectives, main research contributions, and the organization of this thesis.

Chapter 2 presents the network model and the associated terminologies. Furthermore, some backgrounds on network coding are introduced, such as implicit network coding capacity, LP bound, FD bound and RLNC.

Chapter 3 develops reduction techniques for FDG. With the reduced FDGs, various bounds for network coding capacity can be computed much more efficiently without any loss in tightness.

Chapter 4 presents a lower bound on the mutual information between two sets of source/edge random variables in a general multi-session network by exploring their functional dependence relationships. This bound is shown to be useful to

derive a new class of tighter information-theoretic upper bounds on network coding capacity given existing edge-cut bounds.

Chapter 5 proposes a linear precoding and decoding scheme for the source and sink nodes of double-unicast networks by assuming that RLNC is performed at each intermediate node. The achievable region with the proposed scheme is derived, based on which a easily computable region in terms of min-cut values is deduced. Furthermore, a joint routing and RLNC scheme is proposed those multi-session networks where an achievable region with RLNC can be obtained.

Chapter 6 derives the analytical throughput of erasure networks with finite-length spatial-temporal network coding.

Chapter 7 concludes the thesis and points out possible future working directions.

Chapter 2

An Overview on Network Coding

In this chapter, we give the network model used throughout this thesis and present related works in the field of network coding. This chapter serves to provide an overview of network coding technologies and facilitates coherent development and presentation of the results given in the subsequent chapters.

This chapter is organized as follows. In Section 2.1, the network model and the associated terminologies are introduced. Then, the formal definition of network code and an implicit expression for network coding capacity derived in [34] are reproduced in Section 2.2. An LP bound for network coding capacity is then introduced in Section 2.3, which is obtained by replacing the set of entropy vectors in the capacity expression with the set of polymatroid pseudo-entropy vectors. As a relaxation of LP bound, the functional dependence bound is then discussed in Section 2.4. Finally, we briefly describe an algebraic formulation of linear network coding and introduce random linear network coding (RLNC) in Section 2.5.

2.1 Network Model

Throughout this thesis, we consider networks that can be represented by directed acyclic graph $G = (V, E)$, where V is the set of nodes and E is the set of edges. For an edge $e = (u, v) \in E$, with $u, v \in V$, define $\text{tail}(e) = u$ and $\text{head}(e) = v$. The capacity of edge e is denoted by C_e . Furthermore, for $v \in V$, the set of edges entering into v and leaving from v are denoted by $\text{In}(v)$ and $\text{Out}(v)$, respectively. Hence, the *indegree* and the *outdegree* of node v are respectively given by $|\text{In}(v)|$ and $|\text{Out}(v)|$.

Let $S \subset V$ denote the set of source nodes that generate independent information and $T \subset V$ represent the set of sink nodes where the information is demanded. Without loss of generality, we assume that there are no incoming edges to the source nodes and no outgoing edges from the sink nodes, i.e., $\forall s \in S$ and $\forall t \in T$, $\text{In}(s) = \text{Out}(t) = \emptyset$. Furthermore, denote by $\beta(t)$ the set of sources demanded by the sink node $t \in T$. If each sink node demands exactly one source, i.e., $|\beta(t_i)| = 1$ and the demands of different sink nodes are disjoint, i.e., $\beta(t_i) \neq \beta(t_j)$, for $i \neq j$, the multi-session networks are known as *multiple-unicast networks*.

2.2 Network Coding Capacity

Let Y_s denote the random variable associated with the information source s and U_e denote the random variable associated with the codeword sent on edge e . For brevity, the set of source random variables $\{Y_s, s \in S\}$ and the set of edge random variables $\{U_e, e \in E\}$ will be denoted by Y_S and U_E , respectively. Furthermore, let the support of source random variable Y_s and edge random variable U_e be denoted by $\mathcal{Y}_s, s \in S$ and $\mathcal{U}_e, e \in E$, respectively. For a random variable X , we use $H(X)$ to denote its entropy, i.e.,

$$H(X) = - \sum_x p(x) \log_2 p(x), \quad (2.1)$$

where $p(x)$ is the probability mass function of X .

If there is a network code that achieves the rate tuple $(R_s, s \in S)$, the corresponding source and edge random variables defined by this network code should satisfy the following constraints

$$H(Y_s) \geq R_s, \quad (2.2)$$

$$H(Y_S) = \sum_{s \in S} H(Y_s), \quad (2.3)$$

$$H(U_{\text{Out}(s)} | Y_s) = 0, \forall s \in S, \quad (2.4)$$

$$H(U_{\text{Out}(v)} | U_{\text{In}(v)}) = 0, \forall v \in V \setminus (S \cup T), \quad (2.5)$$

$$H(Y_{\beta(t)} | U_{\text{In}(t)}) = 0, \forall t \in T, \quad (2.6)$$

$$H(U_e) \leq C_e, \forall e \in E. \quad (2.7)$$

The conditions given in (2.2)-(2.7) can be understood as follows. The inequality (2.2) states that the entropy of the source random variable Y_s should be greater or equal to the information rate of source s . The equality (2.3) specifies that the sources are mutually independent¹. Equation (2.4) says that $U_{\text{Out}(s)}$ is a function of Y_s for $s \in S$ and (2.5) signifies that $U_{\text{Out}(v)}$ is a function of $U_{\text{In}(v)}$. Note that these two equalities correspond to the encoding constraints of network coding, i.e., the information sent out by the source node s must be a function of the information available at s and that sent out by a non-source node v must be a function of the information received at v . The equality (2.6) corresponds to the decoding requirements, i.e., the sink node t must be able to decode the desired source information $Y_{\beta(t)}$ based on the received information $U_{\text{In}(t)}$. Lastly, the inequality (2.7) means that the information conveyed by any edge should not exceed the edge capacity.

Equation (2.2)-(2.7) defines a set of necessary conditions for a valid network code. Let $\mathcal{H}_N$ be the Euclidean space of dimension $2^N - 1$, with each dimension labeled by a unique non-empty subset of $\{Y_s : s \in S; U_e, e \in E\}$ and $N = |S| + |E|$. For example, the space for random variable $\{Y_1, U_1, U_2\}$ has dimension 7, which are labeled by $H(Y_1), H(U_1), H(U_2), H(Y_1, U_1), H(Y_1, U_2), H(U_1, U_2)$ and $H(U_1, U_2, U_3)$, respectively. Then, each equation in (2.2)-(2.7) defines a sub-region in $\mathcal{H}_N$ [34], which are:

$$\begin{aligned}\mathcal{L}_1 &= \left\{ \mathbf{h} \in \mathcal{H}_N : h_{Y_S} = \sum_{s \in S} h_{Y_s} \right\}, \\ \mathcal{L}_2 &= \left\{ \mathbf{h} \in \mathcal{H}_N : h_{U_{\text{Out}(s)}|Y_s} = 0, s \in S \right\}, \\ \mathcal{L}_3 &= \left\{ \mathbf{h} \in \mathcal{H}_N : h_{U_{\text{Out}(v)}|U_{\text{In}(v)}} = 0, v \in V \setminus (S \cup T) \right\}, \\ \mathcal{L}_4 &= \left\{ \mathbf{h} \in \mathcal{H}_N : h_{Y_{\beta(t)}|U_{\text{In}(t)}} = 0, t \in T \right\}, \\ \mathcal{L}_5 &= \{ \mathbf{h} \in \mathcal{H}_N : h_{U_e} \leq C_e, e \in E \}.\end{aligned}$$

Furthermore, the entropy function, as defined in (2.1), is also constrained by a set of conditions. Denote by Γ_N^* the subregion of $\mathcal{H}_N$ that contains all possible entropic vectors², i.e.,

$$\Gamma_N^* = \{ \mathbf{h} \in \mathcal{H}_N : \mathbf{h} \text{ is entropic.} \}$$

¹Networks with correlated sources have been studied in [59, 60].

²Please refer to Chapter 13 of [34] for the definition of entropic vectors.

Theorem 2.1. [34] *The network coding capacity region for a given multi-session networks is*

$$\mathcal{R} = \wedge \left(\text{proj}_{Y_S} \left(\overline{\text{con}(\Gamma_N^* \cap \mathcal{L}_1 \cap \mathcal{L}_2 \cap \mathcal{L}_3)} \cap \mathcal{L}_4 \cap \mathcal{L}_5 \right) \right). \quad (2.8)$$

where

- $\text{proj}_{Y_S}(A)$ is the projection of the set A on the coordinates h_{Y_S} ,
- $\wedge(A) = \{\mathbf{h} \in \mathcal{H}_N : 0 \leq \mathbf{h} \leq \mathbf{h}' \text{ for some } \mathbf{h}' \in A\}$,
- $\text{con}(A)$ is the convex hull of A and
- $\overline{A}$ is the closure of A .

2.3 Linear Programming Bound

Note that the network coding capacity given by (2.8) is not computable in general since Γ_N^* is still unknown for $N \geq 3$. Let Γ_N represent the subregion of $\mathcal{H}_N$ defined by the set of almost-entropic vectors, i.e., vectors that satisfy the basic entropy inequalities, including the non-negativity of entropy, conditional entropy and conditional mutual information. Since $\Gamma_N^* \subset \Gamma_N$, by replacing Γ_N^* in (2.8) with Γ_N , a computable outer bound can be obtained as

$$\mathcal{R}_{LP} = \wedge \left(\text{proj}_{Y_S} \left(\overline{\text{con}(\Gamma_N \cap \mathcal{L}_1 \cap \mathcal{L}_2 \cap \mathcal{L}_3)} \cap \mathcal{L}_4 \cap \mathcal{L}_5 \right) \right). \quad (2.9)$$

Since Γ_N and $\mathcal{L}_i, i = 1, 2, 3$, are closed and convex regions, (2.9) can be simplified as

$$\mathcal{R}_{LP} = \wedge \left(\text{proj}_{Y_S} (\Gamma_N \cap \mathcal{L}_1 \cap \mathcal{L}_2 \cap \mathcal{L}_3 \cap \mathcal{L}_4 \cap \mathcal{L}_5) \right) \quad (2.10)$$

It was proved in [34] that the minimal set of inequalities implying all basic entropy inequalities for the set of random variables $\{Y_S, U_E\}$ is given by:

$$H(A|\{Y_S, U_E\} \setminus A) \geq 0, A \in \{Y_S, U_E\} \quad (2.11)$$

$$\begin{aligned} I(A; B|C) &\geq 0, \quad A \neq B, A, B \in \{Y_S, U_E\}, \\ C &\subseteq \{Y_S, U_E\} \setminus \{A, B\}. \end{aligned} \quad (2.12)$$

Hence, the LP bound can be computed by solving the following linear programming problem with varied source weights w_s :

$$\begin{aligned} \max \quad & \sum_{s \in S} w_s H(Y_s) \\ \text{subject to} \quad & (2.3) \text{ to } (2.7), (2.11) \text{ to } (2.12). \end{aligned} \tag{2.13}$$

For example, the sum-rate bound can be obtained by solving (2.13) with $w_s = 1, \forall s \in S$.

The LP bound is tight for most special cases of multi-session networks whose capacity regions are known [36, 61, 62]. However, its computational complexity is prohibitive for practical networks in general, since for any given set of source weights $\{w_s, s \in S\}$, the dimension of the LP problem (2.13) grows exponentially with the network size. In order to reduce the computational complexity, some graph-theoretical bounds, such as the edge-cut [35] bound and the FD bound [38], have been extensively studied.

2.4 Functional Dependence Graph and Functional Dependence Bound

Functional dependence graph (FDG) is a class of directed graph that captures the functional dependence relationships among a set of random variables. FDG was first used in [63] to establish the conditional independence among random variables involved in a communication system, and it was later shown to be quite useful in characterizing the capacity bounds for networks with edge capacity constraints.

Definition 2.2 (Functional dependence graph [38]). Let $\bar{V}$ be a set of random variables. A directed graph $\bar{G} = (\bar{V}, \bar{E})$ is called a functional dependence graph for $\bar{V}$ if and only if for all $v \in \bar{V}$,

$$H(v|\{u : (u, v) \in \bar{E}\}) = 0. \tag{2.14}$$

In other words, each node in the FDG corresponds to a random variable, which is a function of its parent nodes. For a network represented by $G = (V, E)$, an FDG can be constructed for the set of source and edge random variables $\{Y_s, s \in S; U_e, e \in E\}$, with the edges in the FDG representing the functional dependence relationships of

these random variables due to the encoding and decoding requirements specified in (2.4)-(2.6). The FDG for a communication network is usually a cyclic graph since the source variables are decodable at the sink nodes. The order of the FDG, which is given by $|\bar{V}| = |S| + |E|$, is usually much larger than that of the original network.

Example 2.1. Consider the 2-unicast butterfly network shown in Figure 2.1, where the source nodes s_1 and s_2 want to send independent messages to the sink node t_1 and t_2 , respectively. Let Y_i represent the random variable associated with the information source at $s_i, i = 1, 2$, and U_i represent the random variable associated with the codeword sent over edge $e_i, i = 1, \dots, 7$. The corresponding FDG is shown in Figure 2.1(b), where Y_i is connected to U_j if e_j is an outgoing edge of source s_i in the original network, U_i and U_j is connected if $\text{head}(e_i) = \text{tail}(e_j)$, and finally U_i is connected to Y_j if e_i is an incoming edge of the destination t_j .

FIGURE 2.1: (a) A 2-unicast butterfly network. (b) FDG of the butterfly network.

For two sets of nodes A and B in an FDG, we say that A determines B , denoted as $A \rightarrow B$, if no elements in B are left after deleting all outgoing edges from A and subsequently deleting all the nodes and edges with no incoming edges or nodes, respectively. Let $\phi(A)$ denote the largest set of nodes that are eliminated during the above procedures. Clearly, $\phi(A)$ gives the largest set of nodes with $A \rightarrow \phi(A)$. For example, let $A = \{U_2, U_3\}$ in the FDG of the butterfly network shown in Figure 2.1(b), then we have $\phi(A) = \{U_2, U_3, U_5, U_6, U_7\}$.

Lemma 2.3. [38] Let $\bar{G} = (\bar{V}, \bar{E})$ be a functional dependence graph. Then for two subset of variables $A, B \subset \bar{V}$, we have

$$A \rightarrow B \implies H(B|A) = 0.$$

Lemma 2.3 can be proved by recursively applying the definition of FDG. For $\bar{v} \in \bar{V}$, let $\text{Up}(\bar{v})$ and $\text{Down}(\bar{v})$ respectively denote the set of immediate upstream and downstream nodes of $\bar{v}$ in the FDG, i.e., $\text{Up}(\bar{v}) = \{\bar{u} : (\bar{u}, \bar{v}) \in \bar{E}\}$ and $\text{Down}(\bar{v}) = \{\bar{u} : (\bar{v}, \bar{u}) \in \bar{E}\}$. It thus follows that $\text{Up}(\bar{v}) \rightarrow \bar{v}$ and hence $H(\bar{v}|\text{Up}(\bar{v})) = 0$.

The above results are useful in determining whether a set of variables in an FDG is irreducible, which is defined as

Definition 2.4. [38] A set of nodes A in an FDG is *irreducible* if there is no proper subset of A determines A , i.e., $\nexists B \subsetneq A$ such that $B \rightarrow A$.

An irreducible set A is *maximal* if $A \rightarrow \bar{V}$, which is equivalent to $\phi(A) = \bar{V}$. The set of all possible maximal irreducible sets in an FDG can be found using the algorithm presented in [38].

According to Lemma 2.3, if A is a maximal irreducible set, we have $H(Y_S|A) = 0$. Hence, the sum rate of the sources are bounded as $\sum_s R_s \leq H(Y_S) \leq H(A)$. If A contains only the edge variables, by invoking the sub-additivity of entropy and edge capacity bounds, we can obtain an explicit outer bound for the sum rate as $\sum_s R_s \leq \sum_{e:U_e \in A} C_e$, which is known as FD bound. Formally, FD bound is defined in the following theorem.

Theorem 2.5 (Functional Dependence Bound [38]). *Let $G = (V, E)$ be a given network with source and edge random variables $Y_s, s \in S$ and $U_e, e \in E$, respectively. Let $\bar{G} = (\bar{V}, \bar{E})$ be its corresponding FDG. Denote the collection of all maximal irreducible sets by $\mathcal{M}$. Then for any $W \subseteq S$, we have*

$$\sum_{s \in W} R_s \leq \min_{A: \{U_A, Y_{W^c}\} \in \mathcal{M}} \sum_{e \in A} C_e, \quad (2.15)$$

where $W^c \triangleq S \setminus W$ and C_e is the capacity of edge e .

Example 2.2. Consider the FDG of the butterfly network shown in Figure 2.1(b). The collection of all maximal irreducible sets are

$$\begin{aligned} &\{Y_1, Y_2\}, \{Y_1, U_3\}, \{Y_1, U_5\}, \{Y_1, U_6\}, \{Y_2, U_2\}, \{Y_1, U_5\}, \{Y_1, U_7\}, \{U_1, U_2, U_3\} \\ &\{U_1, U_2, U_6\}, \{U_1, U_5\}, \{U_1, U_6, U_7\}, \{U_2, U_3, U_4\}, \{U_3, U_4, U_7\}, \{U_4, U_5\}, \{U_4, U_6, U_7\}. \end{aligned}$$

Assume that the edges in the butterfly network have unit capacity, i.e., $C_e = 1, \forall e \in E$.

E. The FD bound given in (2.15) can then be expressed as

$$\begin{aligned} R_1 &\leq 1, \\ R_2 &\leq 1, \\ R_1 + R_2 &\leq 2, \end{aligned} \tag{2.16}$$

which matches with the capacity region of the butterfly network.

The details in deriving the above bound are given as follows. First, the bound for rate R_1 of source s_1 can be obtained based on the maximal irreducible set $\{Y_2, U_2\}$ as

$$\begin{aligned} H(Y_1) + H(Y_2) &\stackrel{(a)}{=} H(Y_1, Y_2) \stackrel{(b)}{\leq} H(Y_2, U_2) \\ &\stackrel{(c)}{\leq} H(Y_2) + H(U_2) \\ \Rightarrow R_1 &\leq H(Y_1) \leq H(U_2) \leq 1. \end{aligned}$$

where (a) follows from the independency of the information sources, (b) follows from the property of the maximal irreducible set, i.e., $\{Y_2, U_2\} \rightarrow \{Y_1, Y_2\}$, and (c) follows from the subadditivity of entropy.

Similarly, the bound for R_2 can be obtained by using the maximal irreducible set $\{Y_1, U_3\}$. Furthermore, the sum rate bound is obtained based on the maximal irreducible set $\{U_1, U_5\}$ as

$$\begin{aligned} R_1 + R_2 &\leq H(Y_1) + H(Y_2) = H(Y_1, Y_2) \\ &\leq H(U_1, U_5) \\ &\leq H(U_1) + H(U_5) \\ &\leq 2. \end{aligned}$$

2.5 Linear Network Coding

Linear network coding has been shown to be capacity-achieving for single-session multicast networks [28]. Though sub-optimal for general multi-session networks, linear network coding is of great practical interest due to its simplicity. For networks employing linear network coding, an algebraic formulation for the solvability of any given rate tuple has been introduced in [1]. This formulation will be reviewed

in this section, and a special class of linear network coding called random linear network coding (RLNC) [41] will also be reviewed.

2.5.1 An Algebraic Formulation of Linear Network Coding

Consider a network $G(V, E)$ with a set of source nodes $S \subset V$ and a set of sink nodes $T \subset V$. Assume that the source node $s_i \in S$ generates k_i packets, denoted by $\mathbf{x}_i^1, \dots, \mathbf{x}_i^{k_i}$ and the sink node t_j wants to decode the information sent by a subset of sources $\beta(t_j) \subset S$. Define $K_n = \sum_{i=1}^n k_i, n \leq |S|$. Let $\mathbf{y}_e$ denote the packets carried by edge e . With linear network coding, the packet carried by an edge originating from the source node is restricted to a *linear* function of the input packets. The packet carried by an outgoing edge from a non-source node $v \in V$ is restricted to a *linear* combination of the packets received by node v . Mathematically, we have

$$\mathbf{y}_e = \begin{cases} \sum_{l=1}^{k_i} \alpha_{l,e} \mathbf{x}_i^l, & e \in \text{Out}(s_i), s_i \in S \\ \sum_{e' \in \text{In}(v)} \gamma_{e',e} \mathbf{y}_{e'}, & e \in \text{Out}(v), v \notin S, \end{cases} \quad (2.17)$$

where $\alpha_{s,e}$ and $\gamma_{e',e}$ are local encoding coefficients. Note that (2.17) corresponds to the encoding constraints specified in (2.4)-(2.5), for linear encodings.

Definition 2.6. (Directed Line Graph [1]) For a network $G(V, E)$, the directed line graph is defined as $\tilde{G}(\tilde{V}, \tilde{E})$ with node set $\tilde{V} = E$ and edge set $\tilde{E} = \{(e, e') \in E^2 : \text{head}(e) = \text{tail}(e')\}$. An edge $\tilde{e} = (e, e') \in \tilde{E}$ is labeled with the corresponding coding coefficient $\gamma_{e,e'}$.

Let $\mathbf{F}$ denote the adjacent matrix of the directed line graph and $\mathbf{A} \in \mathbb{F}_q^{K_{|S|} \times |E|}$ be a matrix whose (i, j) th element is defined as

$$\mathbf{a}_{i,j} = \begin{cases} \alpha_{l,e_j}, & e_j \in \text{Out}(s_p), i = K_{p-1} + l, s_p \in S \\ 0, & \text{otherwise.} \end{cases} \quad (2.18)$$

Similarly, let $\mathbf{B} \in \mathbb{F}_q^{K_{|S|} \times |E|}$ be a matrix capturing the decoding process with the (i, j) th element defined as

$$\mathbf{b}_{i,j} = \begin{cases} \varepsilon_{e_j,l}, & e_j \in \text{In}(t_p), s_u \in \beta(t_p), i = K_{u-1} + l \\ 0, & \text{otherwise} \end{cases} \quad (2.19)$$

Then, it was shown in [1] that the input-output relationship of this network can be expressed as

$$\mathbf{Y} = \mathbf{A}(\mathbf{I}_{|E|} - \mathbf{F})^{-1} \mathbf{B}^T \mathbf{X}, \quad (2.20)$$

where $\mathbf{I}_{|E|} \in \mathbb{F}_q^{|E| \times |E|}$ is an identity matrix and $\mathbf{X}$ is the collection of input packets, i.e., $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^1 \\ \vdots \\ \mathbf{x}_{|S|}^{k_{|S|}} \end{bmatrix}$

Example 2.3. Consider the simple network shown in Figure 2.2(a). The corresponding directed line graph is constructed in Figure 2.2(b) and its adjacent matrix is given by

$$\mathbf{F} = \begin{bmatrix} 0 & 0 & \gamma_{e_1, e_4} & \gamma_{e_1, e_5} & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma_{e_2, e_4} & \gamma_{e_2, e_5} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \gamma_{e_3, e_6} & \gamma_{e_3, e_7} \\ 0 & 0 & 0 & 0 & 0 & \gamma_{e_4, e_6} & \gamma_{e_4, e_7} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (2.21)$$

Furthermore, assume that the source generates three packets, then, the input and output processes are captured by the following matrices $\mathbf{A}$ and $\mathbf{B}$, respectively:

$$\mathbf{A} = \begin{bmatrix} \alpha_{1, e_1} & \alpha_{1, e_2} & \alpha_{1, e_3} \\ \alpha_{2, e_1} & \alpha_{2, e_2} & \alpha_{2, e_3} \\ \alpha_{3, e_1} & \alpha_{3, e_2} & \alpha_{3, e_3} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \varepsilon_{e_5, 1} & \varepsilon_{e_6, 1} & \varepsilon_{e_7, 1} \\ \varepsilon_{e_5, 2} & \varepsilon_{e_6, 2} & \varepsilon_{e_7, 2} \\ \varepsilon_{e_5, 3} & \varepsilon_{e_6, 3} & \varepsilon_{e_7, 3} \end{bmatrix}. \quad (2.22)$$

Hence, the transition matrix is given by $\mathbf{A}(\mathbf{I}_7 - \mathbf{F})^{-1} \mathbf{B}^T$ and the network is solvable if there exists a set of variables $\{\alpha, \gamma, \varepsilon\}$ such that the resulting transition matrix equals an identity matrix.

2.5.2 Random Linear Network Coding

One special class of linear network coding is known as random linear network coding (RLNC), where the coding coefficients in (2.17) are randomly chosen from $\mathbb{F}_q$. RLNC [41] was shown to achieve the capacity of single-session multicast networks with high probability [64, 65]. Compared with deterministic linear network coding schemes

FIGURE 2.2: (a) A simple network constructed in [1]. (b) The corresponding directed line graph.

[29], RLNC is more robust for dynamic networks [15]. Besides, it is an efficient scheme to correct network erasures [53]. Due to its simplicity, RLNC has also been used to construct low-complexity network codes for multi-session networks [50, 66–68].

Consider a single-session multicast network where the source generates n packets, denoted by $\mathbf{x}_1, \dots, \mathbf{x}_n$. Suppose a sink node $t \in T$ receives n coded packets from edges $e_1, \dots, e_n$. With RLNC, the received packets can be expressed with matrix relationship as

$$\begin{bmatrix} \mathbf{y}_{e_1} \\ \vdots \\ \mathbf{y}_{e_n} \end{bmatrix} = \begin{bmatrix} g_{e_1}^1 & \cdots & g_{e_1}^n \\ \vdots & \ddots & \vdots \\ g_{e_n}^1 & \cdots & g_{e_n}^n \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} = \mathbf{G} \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}.$$

In the ideal case with full rank transition matrix $\mathbf{G}$, t can recover the source information by matrix inversion as

$$\begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} = \mathbf{G}^{-1} \begin{bmatrix} \mathbf{y}_{e_1} \\ \vdots \\ \mathbf{y}_{e_n} \end{bmatrix}.$$

With sufficiently large field size q , the decoding will succeed with high probability if n does not exceed the max-flow min-cut bound.

Lemma 2.7. [41] Consider a single-session multicast network with min cut n and a network code with the network coding coefficients $\alpha_{s,e}$ and $\beta_{e',e}$ uniformly chosen from a finite field $\mathbf{F}_q$. The resulting global encoding matrix $\mathbf{G}$ has rank n if the field size q is sufficiently large.

To obtain the transition matrix $\mathbf{G}$ at the sink nodes in a non-coherent manner, an identity matrix is usually tagged in front of the n source packets. Specifically, the packet $\mathbf{x}_i$ is tagged with a unit vector of length n , which has non-zero entry only at the i th position. The overhead symbols are coded in the same way as other data symbols at the intermediate nodes of the network. The sink node can obtain the encoding coefficients from the overhead part of the received packets.

Example 2.4. Consider the single-session multicast network shown in Figure 2.3. Assume that each edge has unit capacity. The min cut of this network is 2 and hence the source can simultaneously send 2 packets to all the sink nodes. With RLNC, the augmented packets tagged with coding overhead are sent out and the coding coefficients at each link are uniformly selected from a finite field.

FIGURE 2.3: A single-session multicast network with RLNC.

One realization of RLNC in $\mathbb{F}_7$ is shown in Figure 2.3. Note that t_1 receives the following two packets:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} 1 & 3 & \mathbf{x}_1 + 3\mathbf{x}_2 \\ 2 & 0 & 2\mathbf{x}_1 \end{bmatrix},$$

Based on the coefficients obtained from the overhead part, the source packets can be decoded by matrix inversion as

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 2 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$$

Note that the overhead of RLNC is linearly proportional to the total number of packets to be coded together. For fixed number of packets with sufficiently large packet length, the coding overhead is negligible. However, in practice, the packet length is limited due to the frame size constraint [69]. To reduce the coding overhead, some compression techniques have been introduced in [70][71].

Chapter 3

Reduced Functional Dependence Graphs

FDG plays an important role in studying the capacity for networks with edge capacity constraints. Besides the progressive d -separating edge-set bound [37] and the functional dependence bound [38] that are directly obtained using FDGs, many other capacity bounds are also closely related to the FDGs, such as the LP bound and the bounds based on information dominance as presented in [39]. The random variables and the functional dependence constraints that are important in computing the LP bound of a given network are directly reflected by its corresponding FDG. The complexity of computing the network coding outer bounds usually grows exponentially with the order of the FDG, which is normally much larger than the order of the network. Hence, it is desirable to reduce the order of the FDG for studying a network, which thus motivates our work presented in this chapter.

Specifically, we introduce some systematic graph processing techniques to reduce the order of the FDG. The resulting FDG is referred to as the *reduced FDG*, which is of smaller size but preserves all the essential properties of the original network such that the capacity of the original network is completely determined by the reduced FDG. As a result, the reduced FDGs can be used to compute the network coding capacity bounds with less computational complexity. Furthermore, since the directed line graph introduced in Section 2.5.1 is a subgraph of the FDG, the reduced FDG can also be used to simplify the algebraic formulation of linear network coding schemes with significantly reduced number of variables and much smaller adjacency matrix. Besides the aforementioned applications, the reduced FDG can

also be used to simplify the network code constructions based on the remaining nodes, without compromising the optimality.

The equivalence between the reduced FDG and original FDG is established by proving the equivalence between their corresponding networks. Network equivalence was previously studied in [72] and [73] with the focus on bounding the capacity of noisy wireless networks by the capacity of certain error-free networks. Inspired by [72] and [73], the authors in [74] have investigated the equivalence between different network topologies, based on which the network graph can be simplified by replacing certain components of the network with simpler structures. The FDG of the simplified network in [74] is of smaller size as compared with that of the original network, and thus the capacity of the network can be more efficiently computed. However, the techniques introduced in [74] can only be applied to networks with very specific structures, and thus have limited applicability in practical scenarios with arbitrary network topology.

In this chapter, with the notion of *hypernetwork*, we show that many networks and their corresponding FDGs can be further reduced significantly. First, some “non-essential” edges in the network are identified and *eliminated*, which results in a “hypernetwork”. The equivalence between the resultant hypernetwork and the original network is established by showing that their capacity regions are the same. Note that edge elimination in the original network is equivalent to node reduction in the corresponding FDG. We further show that some edges of the FDG may be *removed* without affecting the functional dependence relationships among those random variables represented by its nodes. Edge removal in the FDG creates new opportunities for further node reduction of the FDG. Compared with network reduction techniques introduced in the literature [74], the techniques proposed in this chapter achieves additional reduction efficiently.

Finally, we show that the networks can be further reduced if *linear* network coding capacity is considered, as in the case for many practical network coding scenarios due to the simplicity of linear network coding. To the best of our knowledge, this is the first work that simplifies network capacity computation for linear codes.

FIGURE 3.1: An example of hypernetwork.

3.1 Backgrounds

3.1.1 Hypernetwork

The notion of *hypernetwork*, introduced in [75], generalizes the model of conventional network by allowing edges with more than one tail and/or head nodes. A directed hypernetwork can be represented by a directed hypergraph $H = (V, E)$, where V is the set of nodes and E is the set of directed *hyperedges*. A directed hyperedge is an ordered pair of disjoint subsets of nodes, i.e., $e = (V_i, V_j)$, where $V_i, V_j \subseteq V$. Similar to that for the conventional network, we define $\text{tail}(e) = V_i$ and $\text{head}(e) = V_j$. Note that the conventional network is a special case of the hypernetwork with $|\text{tail}(e)| = |\text{head}(e)| = 1$ for any $e \in E$.

Example 3.1. An example of hypernetwork is illustrated in Figure 3.1, which has 12 nodes connected by the following 4 hyperedges:

$$e_1 = (\{1, 2\}, \{4, 5, 6\}), e_2 = (\{3, 4\}, \{7, 8\}), e_3 = (6, \{9, 10\}), e_4 = (\{10, 11\}, 12).$$

3.1.2 Basic Network Simplifications

In this subsection, we present some basic simplifications of the network. A given network $G = (V, E)$ can be reduced by applying some straightforward operations

as follows:

- For $v \in V$, if there is no directed path from any source node to v or there is no directed path from v to any sink node, then node v and its incoming and outgoing edges can be removed without affecting the network capacity.
- Parallel edges $e_1, \dots, e_m$ with capacity $C_1, \dots, C_m$, respectively, between a pair of nodes can be combined into a single edge of capacity $\sum_{i=1}^m C_i$.
- The capacity of a n -hop line network, where each node is only able to communicate with its immediate downstream node, is determined by the bottleneck edge and thus it can be equivalently represented by a single-hop network.

Besides the aforementioned simple reductions, there are some scenarios where more sophisticated reduction techniques can be applied by carefully analyzing the network elements. For instance, the authors in [74] proved that the network shown in Figure 3.2(a) can be reduced to that in Figure 3.2(c), called *Y-network*, with $C_a = C_1 + C_2$, $C_b = C_3 + C_4$ and $C_c = C_5 + C_6$ if the following condition is satisfied:

$$\frac{C_2}{C_1} = \frac{C_4}{C_3} = \frac{C_6}{C_5} = \alpha, \quad (3.1)$$

where α is a constant. It was also shown in [74] that the network in Figure 3.2(b) is equivalent to the Y-network shown in Figure 3.2(c) with $C_a = C_1$, $C_b = C_2 + C_4$ and $C_c = C_5$ if

$$\frac{C_4}{C_2 + C_4} C_1 + \frac{C_2}{C_2 + C_4} C_5 \leq C_3. \quad (3.2)$$

Some generalizations of these results were also discussed in [74]. However, the frequency of the occurrence of these components in a randomly generated network is quite low [76]. Besides, even if such components exist, the complexity of searching and identifying them is quite high. In the following, we will present some reduction techniques that can be efficiently implemented and are applicable for randomly generated networks.

FIGURE 3.2: (a) Two Y-network with shared source and sink nodes. (b) Two source nodes communicating with a common sink. (c) A Y-network.

3.2 Main Results

3.2.1 Edge Elimination in the Network

While the network simplification techniques in the literature focus on *node* eliminations [74], we instead search for opportunities of *edge* elimination, which can be transformed to node reduction in the corresponding FDG.

Definition 3.1. Given a network $G = (V, E)$ and $e = (u, v) \in E$ with $u \notin S$, the reduced network after eliminating edge e , denoted as $\{G - e\}$, is modeled by a directed hypergraph $H = (V, E')$, where $E' = (E \setminus \{e \cup \text{In}(u)\}) \cup \{(w, \{u, v\}) : (w, u) \in E\}$ and the capacity for the hyperedge $(w, \{u, v\})$ is the same as the capacity of the edge (w, u) in the original network.

FIGURE 3.3: (a) The original network G . (b) The reduced network $\{G - e\}$.

The hypernetwork obtained by eliminating an edge e from the original network G is illustrated in Figure 3.3. Although the transformation from G to $\{G - e\}$

makes the network seem more complicated, it is in fact a reduction as reflected by their corresponding FDGs, denoted by $\bar{G}$ and $\{\bar{G} - U_e\}$ and shown in Figure 3.4(a) and Figure 3.4(b), respectively.

FIGURE 3.4: (a) The FDG of the original network. (b) The FDG of the reduced network.

Next, we give a set of sufficient conditions such that the elimination of edge e does not affect the network capacity.

Proposition 3.2. *The networks G and $\{G - e\}$ have the same capacity if the following conditions are satisfied:*

1. $\text{tail}(e) \cap S = \emptyset$, and
2. $C_e \geq \sum_{l: l \in \text{In}(\text{tail}(e))} C_l$.

Proof. Note that the reduced network $\{G - e\}$ differs from the original network G only on that the information available at $\text{tail}(e)$ is made directly available at $\text{head}(e)$, instead of being conveyed through edge e . This is equivalent to relaxing the capacity constraint of edge e of the original network. Therefore, any rate tuple that is achievable in the original network must also be achievable in the reduced network.

On the other hand, if a rate tuple is achievable in the reduced network, it is also achievable in the original network by constructing a code as follows. With both conditions given in Proposition 3.2, it follows that edge e is capable of forwarding all the received data from $\text{tail}(e)$ without any compression. Therefore, we can set the encoding function at edge e as a simple replication. Note that this construction does not violate any encoding constraint for the downstream edges of e since the

same information as in the reduced network is available to them for encoding. Therefore, all the variables in the original networks are the same as that in the reduced network and thus the same rate tuple is achievable. $\square$

Example 3.2. An example for applying Lemma 3.2 is shown in Figure 3.5. The 2-unicast network shown in Figure 3.5(a) is constructed in [77] with each edge having unit capacity. By applying Lemma 3.2 and subsequently removing those nodes without any outgoing edges, a reduced network is obtained, as represented by the hypergraph in Figure 3.5(b). Let Y_i represent the random variable associated with information at $s_i, i = 1, 2$ and U_j represent the random variable associated with the codeword sent on edge $e_j, j = 1, \dots, 12$, Figure 3.5(c) and Figure 3.5(d) depict their corresponding FDGs. Clearly, with the edge reduction by Lemma 3.2, an FDG of much smaller order is obtained.

FIGURE 3.5: (a) An example network. (b) The reduced network after edge elimination with Lemma 3.2. (c) The FDG of the original network. (d) The FDG of the reduced network.

With the correspondence between a network and its FDG, the elimination given

by Proposition 3.2 can be performed on the FDG directly to obtain the reduced FDG, which gives the following corollary:

Corollary 3.3. *The FDG $\bar{G}$ and $\{\bar{G} - U_e\}$ give the same capacity of their corresponding networks if the following conditions are satisfied:*

1. $\text{Up}(U_e) \cap Y_S = \emptyset$, and
2. $C_e \geq \sum_{l: U_l \in \text{Up}(U_e)} C_l$.

3.2.2 Edge Reduction in the FDG

In the preceding subsection, the reductions are based on edge eliminations in the original network, which are equivalent to node reductions in the corresponding FDG. In this subsection, we show that certain edges in the FDG are redundant, and the removal of which may create new opportunities for node reduction.

While the edges in a network represent communication channels, the edges in an FDG capture the functional dependence relationships of the random variables. For an FDG $\bar{G} = (\bar{V}, \bar{E})$, an edge $\bar{e} = (\bar{u}, \bar{v}) \in \bar{E}$ is redundant if the functional dependence relationship is implied by the subgraph $\bar{G} \setminus \bar{e} \triangleq (\bar{V}, \bar{E} \setminus \bar{e})$. From Section 2.4, it is known that for $A, B \subset \bar{V}$, if $A \rightarrow B$, we have $H(B|A) = 0$, i.e., the set of random variables in B does not provide any additional information once A is given. Therefore, if $\text{Up}(\bar{v}) \setminus \bar{u} \rightarrow \bar{u}$, the edge $(\bar{u}, \bar{v})$ can be removed without affecting the functional dependence relationships among the random variables. To determine whether $A \rightarrow B$, we remove all the nodes in A from the FDG and the outgoing edges from those nodes, and subsequently the nodes and the edges with no incoming edges and no tail nodes, respectively. During this procedure, if B is removed, we can state that $A \rightarrow B$ [38].

Proposition 3.4. *Given an FDG $\bar{G} = (\bar{V}, \bar{E})$ that corresponds to the network $G = (V, E)$, the edge $(U_i, U_j) \in \bar{E}$ can be removed without affecting the capacity of the corresponding network if $(\text{Up}(U_j) \setminus U_i) \rightarrow U_i$.*

Proof. Any rate tuple achievable in the reduced network (the network corresponding to $\bar{G} \setminus (U_i, U_j)$) can be achieved in the original network using the same network code by ignoring the information carried from edge i to j . On the other hand, any rate tuple that is achievable in the original network is also achievable in the

reduced network with the same network code, since the information available for constructing codeword on edge j is unchanged given $(\text{Up}(U_j) \setminus U_i) \rightarrow U_i$. $\square$

Remark 3.5. Proposition 3.4 is also applicable to those edges entering into or leaving from the nodes associated with source random variables in the FDG. The proof follows similarly by changing the information on the edge to the information generated by the source accordingly. In fact, the source random variables are indistinguishable from the edge random variables in an FDG.

Example 3.3. Consider the FDG shown in Figure 5.2(d). By applying Proposition 3.4, the redundant edges in the FDG are $(U_4, U_{12}), (U_7, U_{12})$ and (U_{12}, Y_1) . By removing those redundant edges and subsequently those nodes and edges with no outgoing edges and no head, respectively, we obtain a very simple FDG shown in Figure 3.6(a). The corresponding network of the reduced FDG that has the same capacity as the original network in Figure 5.2(a) is shown in Figure 3.6(b).

FIGURE 3.6: (a) The reduced FDG from Figure 5.2(d). (b) The corresponding reduced network.

3.2.3 An Iterative Algorithm for Reduced FDG

The proposed reduction techniques discussed in the preceding subsections are summarized in Algorithm 1. The inputs of Algorithm 1 include the corresponding FDG of the considered network and the capacity of all edges.

In Algorithm 1, the pseudo-codes from line 4 to line 8 reduce the FDG according to Corollary 3.3, and those from line 9 to line 14 remove the redundant edges based on Proposition 3.4. At line 15, the function $\text{prune}(\bar{G})$ deletes those nodes with no incoming or outgoing edges and subsequently those edges with no tail or head

Algorithm 1 ReducedFDG($\bar{G}(\bar{V}, \bar{E}), (C_e : U_e \in \bar{V})$)

```

1:  $flag = 1$ 
2: while  $flag = 1$  do
3:    $flag = 0$ 
4:   for  $U_e \in \bar{V}$  do
5:     if  $C_e \geq \sum_{l: U_l \in \text{Up}(U_e)} C_l$  then
6:        $\bar{G} = \{\bar{G} - U_e\}$ 
7:     end if
8:   end for
9:   for  $\bar{e} = (\bar{u}, \bar{v}) \in \bar{E}$  do
10:    if  $\text{Up}(\bar{v}) \setminus \bar{u} \rightarrow \bar{u}$  then
11:       $\bar{E} = \bar{E} \setminus \bar{e}$ 
12:       $flag = 1$ 
13:    end if
14:  end for
15:   $\bar{G} = \text{prune}(\bar{G})$ 
16: end while

```

node in a graph. Note that if any edge is removed from the FDG, Algorithm 1 goes back to check for new opportunities to further reduce the node of the FDG. The total number of iterations of this algorithm is $\alpha|\bar{V}| + \alpha|\bar{E}|$, where α is the number of iterations of the **while** loop. Note that α is upper bounded by $|\bar{E}|$ and our simulation results suggest that it is usually much smaller than $|\bar{E}|$.

3.2.4 Extensions of Proposed Reduction Techniques

In this section, we discuss some possible extensions of the proposed reduced techniques presented in the preceding sections.

3.2.4.1 Group of Nodes

A simple extension of Corollary 3.3 from single edge random variable to a group of edge random variables gives the following result.

Corollary 3.6. *Consider an FDG $\bar{G} = (\bar{V}, \bar{E})$ and $U_Z \subset (\bar{V} \setminus Y_S)$, $\bar{G}$ and $\{\bar{G} - U_Z\}$ are equivalent if the following conditions are satisfied:*

1. $\text{Up}(U_i) = \text{Up}(U_j), \forall i, j \in Z,$
2. $\text{Down}(U_i) = \text{Down}(U_j), \forall i, j \in Z,$ and

$$3. \text{Up}(U_Z) \cap Y_S = \emptyset \text{ and } \sum_{i \in Z} C_i \geq \sum_{l: U_l \in \text{Up}(U_Z)} C_l.$$

Corollary 3.6 follows from Corollary 3.3 by treating the set of random variables U_Z as a single “super random variable”. It can be easily incorporated in Algorithm 1 by grouping the nodes in the FDG before the elimination process.

Example 3.4. Figure 3.7(a) shows a component of a particular FDG. By treating nodes U_3 and U_4 as a group, i.e., $U_Z = \{U_3, U_4\}$, if $C_3 + C_4 \geq C_1 + C_2$, then U_Z can be eliminated according to Corollary 3.6, which yields a simpler structure as shown in Figure 3.7(b).

FIGURE 3.7: (a) A component of an FDG. (b) The reduced component of an FDG.

3.2.4.2 Constrained Source Capacity

In Section 3.2.1, we assume that all sources have unlimited capacity, i.e., the rate of information generated by the source is unbounded. Therefore, the elimination techniques introduced there are applicable only to those edges outgoing from non-source nodes. However, since the edges are capacity-limited, the information that can be transmitted by the source is actually also bounded. We extend our results to the edges outgoing from source nodes in this section. Without loss of generality, we assume that the capacity of the source is bounded by the max-flow min-cut value between the source and its corresponding sinks.

Definition 3.7 (Critical Cut). Consider a network $G(V, E)$ with a set of source nodes $S \subset V$ and a set of sink nodes $T \subset V$. A subset of edges $X \subseteq E$ is a cut for source $s \in S$ if the removal of edges in X disconnects s with any sink node that wants to decode it. Among all possible cuts for source s , those with minimum capacity are called *critical cuts* and the capacity of critical cuts is denoted by C_s ,

i.e.,

$$C_s = \min_X \sum_{e \in X} C_e. \quad (3.3)$$

In contrast to the edge capacities which are independent from network topologies, the source capacity may change after network simplification. However, as long as one of the critical cuts is unaffected after the simplification, the source capacity remains the same. For notational convenience, denote the collection of critical cuts of source s by $\mathcal{O}_s$. Then, Proposition 3.2 can be extended to include those edges that are outgoing from source nodes with finite source capacity, which gives the following result:

Proposition 3.8. *The networks G and $\{G - e\}$ have the same capacity if the following conditions are satisfied:*

1. $C_e \geq \sum_{l: l \in \text{In}(\text{tail}(e))} C_l + \sum_{s: s \in \text{In}(\text{tail}(e))} C_s$, and
2. $\mathcal{O}_s \setminus \{X : X \in \mathcal{O}_s, e \in X\} \neq \emptyset, \forall s \in \text{tail}(e)$.

With Proposition 3.8, further reductions can be performed, especially for small networks where a large portion of edges are outgoing from the source nodes. The resulting reductions in Proposition 3.8 can be incorporated in Algorithm 1 with $\mathcal{O}_s, s \in S$ as additional inputs.

3.3 Further FDG Reduction for Linear Network Coding

It has been proved in [33] that linear network coding is not sufficient to achieve the capacity for general networks. However, in many practical scenarios, it is preferred due to its simplicity. With linear network coding, the codeword sent on edge e is a *linear combination* of the input information available at $\text{tail}(e)$.

In the following, we present some scenarios where further reduction can be obtained when linear network coding capacity is considered.

Proposition 3.9. *Consider a network $G = (V, E)$ and an edge $e = (u, v) \in E$, the reduced network $\{G - e\}$ has the same linear network coding capacity as G if $\text{head}(e) \notin T$ and $C_e \geq \sum_{k: k \in \text{Out}(v)} C_k$.*

Proof. As shown in the proof of Proposition 3.2, if a rate tuple is achievable in the original network G , it is also achievable in the reduced network $\{G - e\}$ as the latter is less constrained.

On the other hand, if a rate tuple is achievable in the reduced network $\{G - e\}$ with *linear* network coding, the codeword sent on any edge in $\text{Out}(v)$ can be separated into two parts. One of them is a linear combination of the codewords sent on edges $\text{In}(u)$, denoted as f_k^u , and the other part is a linear combination of the codewords sent on edges $\text{In}(v) \setminus \text{In}(u)$. As $C_e \geq \sum_{k:k \in \text{Out}(v)} C_k$, all the necessary information $\{f_k^u : k \in \text{Out}(v)\}$ can be calculated at node u and conveyed to v through edge e without any compression in the original network. Therefore, the rate tuple is also achievable in the original network. $\square$

Proposition 3.9 can be applied to the corresponding FDG, which leads to further reduction as given by the following corollary.

Corollary 3.10. *The corresponding networks of the FDG $\bar{G}$ and $\{\bar{G} - U_e\}$ have the same linear network coding capacity if the following conditions are satisfied.*

1. $Y_S \cap \text{Down}(U_e) = \emptyset$, and
2. $C_e \geq \sum_{k:U_k \in \text{Down}(U_e)} C_k$.

Example 3.5. *Consider the network shown in Figure 3.8(a), which is obtained from [78] with slight modifications. All the edges in the network have unit capacity. With Algorithm 1, we obtain the reduced FDG shown in Figure 3.8(b). If only linear network coding capacity is considered, the FDG can be further reduced to the one shown in Figure 3.8(c) according to Corollary 3.10. U_{14} , U_{15} and U_{16} are reduced as they have only a single downstream node and thus the conditions in Corollary 3.10 are satisfied.*

The reductions presented in Proposition 3.9 or Corollary 3.10 is only applicable for linear network coding capacity. To show this, we present in Appendix A a counter example which demonstrates that the reduction is not valid for the more general non-linear network coding capacity.

FIGURE 3.8: (a) A 3-unicast network. (b) The reduced FDG from Algorithm 1. (c) The FDG with further reduction by applying Corollary 3.10.

3.4 Performance Evaluation

The proposed reduction techniques on the FDG can be used to simplify the analysis and design of network coding for a network. The efficiency of the proposed techniques is largely dependent on network topology. In the preceding sections, the efficiency is demonstrated through specific networks constructed in the literature. In practice, the network topology varies in different communication systems and can be modeled by different random graphs [79]. For instance, the generic wired networks and indoor wireless networks are modeled by the Érdős-Rényi (ER) graphs [80], whereas wireless sensor networks can be appropriately modeled by the random geometric (RG) graphs [81]. In this section, the proposed algorithm is evaluated on the random networks that are generated based on ER graphs¹.

A directed acyclic ER graph with parameter (n, p) is generated with the following steps:

- A set of n nodes are generated and randomly labeled, by which the probability

¹We have also performed simulations based on networks generated from RG graphs. Similar performance gains have been observed.

for two of them to be connected is equal to p . The direction of the arrow is defined by looking at the labeling of the nodes, i.e., if there is a link between two nodes, it is directed from the node with smaller index to the one with larger index.

- 2 pairs of source and sink nodes are added. The probability that a source node is connected to any of the n nodes equals to p . Similarly, with probability p , there is a link between any of the n nodes to a sink node.
- Those nodes that are not directly connected to any source node or any sink node are pruned since they do not help in forwarding the information.
- Each edge in the network is assumed to have unit capacity.

The efficiency of the reduction techniques proposed in Algorithm 1 and Section 3.3 are evaluated on the random networks with different sizes n and densities p . Consider a random network with the corresponding conventional FDG and the reduced FDG denoted by $\bar{G} = (\bar{V}, \bar{E})$ and $\bar{G}_R = (\bar{V}_R, \bar{E}_R)$, respectively. The size reduction of the FDG is given by $|\bar{V}| - |\bar{V}_R|$. For each set of parameters (n, p) , the size reduction is obtained by averaging over 100 random graph realizations and the results are shown in Figure 3.9. It is observed that substantial reduction is achieved for networks with low-to-moderate density. For instance, with $p = 0.3$, the size reductions achieved by Algorithm 1 are 164, 75 and 25 for $n = 20, 50$ and 100, respectively, which correspond to 31%, 18% and 10% of the original FDG size, respectively. As the computational complexity of the network capacity bounds usually grows exponentially with the size of the FDG, a small amount of reduction in the order of FDG leads to a significant complexity reduction, as will be shown in the following section. Another observation from Figure 3.9 is that the size reduction decreases with the network density. This is expected since an edge in a network is less likely to be able to convey all the information available to its tail node when the number of edges entering its tail node increases. It is also revealed by Figure 3.9 that most of the reduction can be obtained by applying Corollary 3.3 only. The additional reduction by iteratively applying Corollary 3.3 and Proposition 3.4 is substantial only for moderate-to-large networks with low density, but becomes marginal as the network density increases or the size of the network decreases.

In Figure 3.10, the line marked with green squares gives the reduction for preserving network capacity using Corollary 3.3 and Proposition 3.4, and the dash

FIGURE 3.9: Performance evaluation of the proposed reduction techniques on randomly generated networks.

line marked with yellow rounds depicts the reduction for preserving *linear* network coding capacity by further applying Corollary 3.10, both for networks with $n = 50$. It is observed that significant additional reduction can be achieved for linear network coding. For instance, with $p = 0.3$, additional 77 nodes of the FDG are reduced, which is 18% of the original FDG size.

3.5 Application of Reduced FDG

The proposed FDG reduction techniques have a wide range of applications, such as for various network capacity inner/outer bounds computations. It can also be used to simplify the network code construction by performing coding only over edges corresponding to the remaining nodes in the reduced FDG. In this section, the applications of the reduced FDG for LP bounds computation and algebraic formulation of scalar linear network coding will be discussed in detail.

FIGURE 3.10: Performance evaluation of the proposed reduction techniques for linear network coding.

3.5.1 Complexity Reduction for LP Bound Computation

Referring to Section 2.3, computing the LP bound for a networks is achieved by solving the linear programming optimization in (2.13) with different rate weights for the sources. For every set of chosen weights, the dimension of the optimization problem is $2^N - 1$, with $N = |E| + |S|$, and the number of constraints is $N + \binom{N}{2}2^{N-2} + 1 + 2|E| + |T|$, which includes $N + \binom{N}{2}2^{N-2}$ elemental information inequalities, one equality representing source independence, $|E|$ equalities for encoding requirements, $|E|$ inequalities capturing the edge capacity constraints, and $|T|$ equalities stating the decoding requirements.

The dimension of the optimization problem (2.13) grows exponentially with the order of its FDG, N . Therefore, any linear reduction of FDG size results in exponential reduction of the problem dimension. The complexity reduction in the constraints is more significant as it grows even faster with the size of FDG. Specifically, if the FDG size reduces from N_1 to N_2 , the percentage of reduction in terms of problem dimension and the computational complexity, denoted by $\rho_d\%$

and $\rho_c\%$, respectively, are given by

$$\rho_d\% = \frac{2^{N_1} - 2^{N_2}}{2^{N_1} - 1} \approx 1 - 2^{N_2 - N_1} = 1 - 2^{-\Delta_N}, \quad (3.4)$$

$$\rho_c\% \approx 1 - \frac{(2^{N_2} - 1)(3N_2 + \binom{N_2}{2}2^{N_2-2})}{(2^{N_1} - 1)(3N_1 + \binom{N_1}{2}2^{N_1-2})} \approx 1 - 2^{-2\Delta_N}, \quad (3.5)$$

where $\Delta_N \triangleq N_1 - N_2$.

Example 3.6. Consider the butterfly network shown in Figure 3.11(a), where all edges in the network have unit capacity. If the original FDG shown in Figure 3.11(b) is used, the number of variables involved is $N = 9$ and the dimension of the resulting linear programming is 511. The number of constraints is 4634. According to Corollary 3.3, U_6 and U_7 can be reduced as they have sufficient capacity to directly forward all the received information from U_5 . The resultant FDG is shown in Figure 3.11(c). Moreover, according to Proposition 3.8, the nodes U_1, U_2, U_3 and U_4 can be reduced as the entropy of both source variables Y_1 and Y_2 are bounded by the entropy of U_5 , which is further bounded by its unit capacity. Therefore, the reduced FDG contains only three nodes, as shown in Figure 3.11(d). With this reduced FDG, the number of variables N reduces to 3, thus the problem dimension and the constraint size reduce to 7 and 14×7 , respectively. The overall complexity reduction exceeds 99.99%. From Matlab simulation, the computational time for the LP bound is reduced from 108 seconds based on the original FDG (shown in Figure 3.11(b)) to 0.03 second based on the reduced FDG (shown in Figure 3.11(d)).

The complexity reduction of LP bounds computation has also been investigated in [82] by removing the redundant inequalities involved in (2.3)-(2.12). Note that, the reduction techniques introduced in this chapter and that in [82] are non-conflicting and can be applied together to benefit from both.

Example 3.7. Consider the butterfly network shown in Figure 3.11(a). By applying the complexity reduction techniques introduced in [82] on the original FDG, the problem dimension of LP bound computation is given by 39 and the number of constraints is given by 852. On the other hand, if the reduced FDG shown in Figure 3.11(d) is used, the dimension is further reduced to 4.

FIGURE 3.11: (a) The butterfly network. (b) The corresponding FDG. (c) The reduced FDG with reduction techniques proposed in Corollary 3.3. (d) The reduced FDG with extended reduction techniques proposed in Proposition 3.8.

3.5.2 Complexity Reduction in Algebraic Formulation of Linear Network Coding

Referring to Section 2.5.1, directed line graph is of great importance in constructing the algebraic formulation of the linear network coding. Consider a network $G = (V, E)$ and denote its corresponding directed line graph and FDG by $\tilde{G} = (\tilde{V}, \tilde{E})$ and $\bar{G} = (\bar{V}, \bar{E})$, respectively. It follows that $\tilde{G}$ is a subgraph of $\bar{G}$. Therefore, the reduction techniques proposed in this chapter, including the reductions for linear network coding capacity proposed in Section 3.3, can be used to simplify the algebraic formulation as well.

Let $\mathbf{A}$ be the matrix that maps the source to its outgoing edges, $\mathbf{F}$ be adjacency matrix of $\tilde{G}$ and $\mathbf{B}$ be the matrix that maps incoming edges of the sink nodes to decoded symbols. The transition matrix of the network is formulated as $\mathbf{M} = \mathbf{A}(\mathbf{I} - \mathbf{F})^{-1}\mathbf{B}^T$ [1]. The total number of variables involved in this formulation is $|\bar{E}|$, and the adjacency matrix $\mathbf{F}$ is of size $|\bar{V}| \times |\bar{V}|$. Let Δ_v and Δ_e denote the amount

of nodes and edges removed from the original FDG respectively. With reduced FDG, the variables involved in the formulation is reduced by Δ_e and the reduction in complexity for computing the transition matrix is proportional to Δ_v .

Example 3.8. Figure 3.12(a) shows a network that is solvable only over fields with characteristic 2 [33] and Figure 3.12(b) shows the corresponding reduced FDG. By labeling each edge in reduced FDG, we can formulate the scalar linear network coding as:

$$\mathbf{A} = \begin{bmatrix} \alpha_1 & \alpha_2 & 0 & 0 & 0 \\ 0 & \alpha_3 & \alpha_4 & 0 & 0 \\ 0 & 0 & \alpha_5 & 0 & \alpha_6 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & \gamma_6 & 0 & \gamma_5 \\ 0 & 0 & 0 & \gamma_3 & \gamma_4 \\ \gamma_1 & 0 & 0 & \gamma_2 & 0 \end{bmatrix} \quad \mathbf{F} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta_1 & \beta_2 \\ 0 & 0 & 0 & \beta_3 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{M} = \mathbf{A}(\mathbf{I} - \mathbf{F})^{-1}\mathbf{B}^T$$

$$= \begin{bmatrix} \alpha_2\beta_2\gamma_5 & \alpha_2\beta_1\gamma_3 + \alpha_2\beta_2\gamma_4 & \alpha_1\gamma_1 + \alpha_2\beta_1\gamma_2 \\ \alpha_4\gamma_6 + \alpha_3\beta_2\gamma_5 & (\alpha_3\beta_1 + \alpha_4\beta_3)\gamma_3 + \alpha_3\beta_2\gamma_4 & (\alpha_3\beta_1 + \alpha_4\beta_3)\gamma_2 \\ \alpha_5\gamma_6 + \alpha_6\gamma_5 & \alpha_5\beta_3\gamma_3 + \alpha_6\gamma_4 & \alpha_5\beta_3\gamma_2 \end{bmatrix} \quad (3.6)$$

The transition matrix is given in (3.6). By setting the diagonal element to be 1 and off-diagonal element to be 0, it can be proved that a solution exists only if all the variables are chosen from a field with characteristic 2. Note that the total number of variables used in this formulation is 15 and the size of $\mathbf{F}$ is only 5×5 . However, if the original line graph is used, the number of variables is 28 and the dimension of the adjacency matrix is 18×18 . Therefore, the total amount of reduction in problem dimension is 46% and in complexity is 92%.

On top of the proposed reduced FDG, the algebraic formulation of scalar linear network coding can be further simplified by using the source-to-sink path gain variables as introduced in [83].

FIGURE 3.12: A network that is solvable only over fields with characteristic 2

3.6 Chapter Summary

In this chapter, we have introduced some systematic graph processing techniques to reduce the size of the FDG. The first reduction technique is performed on the original networks, where “non-essential” edges are identified and eliminated. This is equivalent to node reduction in the corresponding FDG. Furthermore, we have shown that certain edges in the FDG may also be removed without affecting the functional dependence relationships of the random variables. With the reduced FDGs, various capacity bounds of the network, including outer bounds on linear coding capacity, can be computed much more efficiently without any loss in tightness. Simulation results have shown that with the proposed techniques, substantial reduction can be achieved for the FDG of the randomly generated networks, especially for networks with low-to-moderate density. Two applications of the reduced FDG have been discussed in this chapter, which are the computation of LP bound and algebraic formulation of scalar linear network coding. In both applications, the proposed techniques significantly reduce the problem size and complexity.

Chapter 4

Refined Functional Dependence Bounds for Network Coding Capacity

Explicit characterization of the capacity region for multi-session network coding is a long-standing open problem. Over the past few years, several computable outer bounds have been developed, [34, 36–39, 84], among which the linear programming (LP) [34, Chapter 21] bound is known to be the tightest. Although the graph processing techniques introduced in Chapter 3 can be used to simplify the LP bound computation, the complexity still increases double exponentially with the order of the reduced FDG [82], which hinders the LP bound to be used in many practical scenarios. On the other hand, most of the practical bounds, such as the cut-set bound [37], the network-sharing bound [36] and the functional dependence (FD) bound [38], are developed based on identifying the information bottlenecks in the network, and hence are usually referred to as “edge-cut bounds”. As pointed out in [37], one major limitation of such edge-cut bounds is that every source is treated “equally”, in the sense that the same weight is assigned for each source rate for bound computation. As a consequence, edge-cut bounds are usually quite loose, even for simple networks such as the double-unicast network [40].

Recently, refined edge-cut bounds have been introduced in [84], where different weights are assigned for the source rates. It is found in [84] that if different sink nodes have overlapping demands, the mutual information between the random variables associated with their incoming edges is lower bounded by the rate of their overlapping demands. Together with the information bottleneck, such lower

bounds lead to some *weighted* sum-rate bounds, which are tighter than the original edge-cut bounds. Furthermore, by bounding the mutual information between the edge random variables in the network, a set of generalized cut-set bounds have been proved in [85] for broadcast networks where all source messages collocate at a single source node. However, similar to that in [84], the bounds proved in [85] are still based on overlapped demands at different sink nodes. It was not clear how to extend the approach introduced in [84] and [85] for general networks where the sink nodes have disjoint demands. This motivates our work presented in this chapter.

In this chapter, we present a lower bound on the mutual information between any two subsets of source/edge random variables in a network. Such a bound is obtained based on the functional dependence relationships among the source/edge random variables, and hence it is applicable for any networks, regardless of whether the sink nodes have overlapping demands. Moreover, we will show that this lower bound on mutual information is the tightest one for networks with general network coding. Refined versions of edge-cut based bounds, e.g. [36–39], can be obtained using our proposed lower bound on the mutual information. A refined cut-set bound is demonstrated by an example.

Based on the developed lower bound on the mutual information, a systematic approach is presented to obtain the “refined” FD bounds. Note that since the proposed lower bound on the mutual information is a generalization of the results in [84], the refined FD bounds include the refined edge-cut bounds given in [84] as a subset, and hence is tighter. On the other hand, in [86], we have characterized a set of Shannon-type inequalities that renders weighted sum-rate bounds. It can be verified that the weighted bounds given in [86] are also a subset of the refined FD bounds presented in this chapter, and a network example where the refined FD bounds are *strictly* tighter than the bounds given in [86] is presented.

4.1 Related Works

A complete and comprehensive discussion of the aforementioned “edge-cut” type of bounds can be found in [87]. In this section, we will review some weighted bounds for network coding capacity, including the refined edge-cut bound presented in [84] and the weighted sum-rate FD bound introduced in [86].

4.1.1 Refined Edge-Cut Bound

In [84], refined edge-cut bounds are introduced for networks in which the receivers have overlapping demands. The main idea is illustrated with the following example.

Example 4.1. Consider the example network given in [84], shown in Figure 4.1, where two independent random variables, Y_1 and Y_2 , are generated by the source node s . Y_1 is demanded by sink nodes t_1 and t_2 and Y_2 is demanded by the sink node t_3 . Denote the capacity of edge e_i by C_i and assume that the information sent over edge i is represented by random variable U_i , where $i \in \{1, 2, 3, 4\}$.

FIGURE 4.1: A network with overlapping demands.

Since both t_1 and t_2 need to recover Y_1 , we have $H(Y_1|U_1) = 0$ and $H(Y_1|U_2) = 0$. Therefore, the mutual information between the edges random variables U_1 and U_2 is lower bounded as: $I(U_1; U_2) \geq H(Y_1) \geq R_1$. Moreover, since the removal of those two edges associated with random variables U_1 and U_2 disconnects the source from all receivers, (U_1, U_2) is an information bottleneck of the network and thus we have $R_1 + R_2 \leq H(U_1, U_2)$. The refined edge-cut bound is then obtained as follows.

$$\begin{aligned}
 R_1 + R_2 &\leq H(U_1, U_2) \\
 &= H(U_1) + H(U_2) - I(U_1; U_2) \\
 &\leq C_1 + C_2 - R_1 \\
 \Rightarrow 2R_1 + R_2 &\leq C_1 + C_2.
 \end{aligned} \tag{4.1}$$

Compared with the original edge-cut bound $R_1 + R_2 \leq 2$. The obtained bound in (4.1) is tighter by including different weights for these two source rates, hence it is called “refined” edge-cut bound in [84].

4.1.2 Weighted Sum-rate FD Bound

In [86], we characterized a class of Shannon-type inequalities that renders, together with the sub-additivity of entropy, a weighted sum-rate bound that is guaranteed to improve over the FD bound [38].

Consider a multi-session network $G = (V, E)$ with the set of source random variables Y_S and the set of edge random variables U_E . Let $\mathcal{M}$ be the collection of all maximal irreducible sets and $\mathcal{L} \subset \mathcal{M}$ be the collection of all maximal irreducible sets that contain edge random variables only. For $U_A \in \mathcal{L}$, $U_B \subset U_E$ and $W \subset S$, a weighted sum-rate bound that strictly improve over the FD bound can be obtained if the following three conditions are satisfied:

1. there exists a set of edge random variables $U_X \subset U_E$ such that $(Y_W, U_B) \subseteq \phi(U_X)$,
2. there exists a set of edge random variables $U_Z \subset U_E$ such that $U_B \subseteq \phi(Y_{W^c}, U_Z)$, and
3. $\sum_{e \in (A \setminus B)} C_e + \sum_{e \in X} C_e + \sum_{e \in Z} C_e < C_S + C_W$,

where C_S and C_W are sum-rate FD bound for the set of source S and W , i.e., the FD bounds of $\sum_{s \in W} R_s$ and $\sum_{s \in S} R_s$, respectively.

4.2 Main Result

For networks without any overlapping demands, the refined edge cut bound introduced in [84] is no longer applicable, as the simple lower bounds on the mutual information between the edge random variables as in Example 4.1 cannot be obtained. Although the weighted bound introduced in [86] may still apply, the procedure for finding this weighted bound can be quite tedious since we need to exhaustively search for all possible U_X and U_Z .

In this section, based on the functional dependence relationships in the FDG, we derive a lower bound on the mutual information between any two sets of source/edge random variables. The lower bound is applicable for general multi-session networks, regardless of the demands at the sink nodes. Based on such lower bounds, a new class of information-theoretic outer bounds, referred as *refined FD bounds* are derived for network coding capacity.

4.2.1 A Lower Bound on Mutual Information

Note that the FD bound given in Theorem 2.5 is established based on the sub-additivity property of entropy, which is equivalent to the non-negativity of the mutual information between the edge random variables. Therefore, the FD bound can be further improved if the mutual information can be lower bounded by a strictly positive value.

In order to establish the lower bound of the mutual information between the source/edge random variables in a network, we need to make use of the following properties of an FDG.

Lemma 4.1. *Let $\bar{G} = (\bar{V}, \bar{E})$ be an FDG, then $\forall A \subseteq \bar{V}$, $H(A) = H(\phi(A))$.*

Proof. Since $A \rightarrow \phi(A)$, according to Lemma 2.3, we have $H(\phi(A)|A) = 0$. Meanwhile, as $A \subseteq \phi(A)$, we have $0 \leq H(A|\phi(A)) \leq H(A|A) = 0$, i.e., $H(A|\phi(A)) = 0$. Therefore, we can conclude that $H(A) = H(\phi(A))$. $\square$

Lemma 4.2. *For a given FDG $\bar{G} = (\bar{V}, \bar{E})$ and $A, B \subseteq \bar{V}$*

$$I(A; B) \geq H(\phi(A) \cap \phi(B)). \quad (4.2)$$

Proof. For two sets of random variables $A, B \subseteq \bar{V}$

$$\begin{aligned} I(A; B) &= H(A) + H(B) - H(A, B) \\ &\stackrel{(a)}{=} H(\phi(A)) + H(\phi(B)) - H(A, B) \\ &\stackrel{(b)}{\geq} H(\phi(A) \cap \phi(B)) + H(\phi(A) \cup \phi(B)) - H(A, B) \\ &\stackrel{(c)}{=} H(\phi(A) \cap \phi(B)), \end{aligned}$$

where (a) follows from Lemma 4.1, (b) follows from the sub-modularity property of entropy, and (c) is true due to the facts that $(A, B) \rightarrow (\phi(A) \cup \phi(B))$ and $(A, B) \subseteq (\phi(A) \cup \phi(B))$. $\square$

Note that the mutual information lower bound in (4.2) generalizes the result presented in [84], where a refined bound can be obtained only when there are overlapping demands. Consider a network with sources denoted by S and sink nodes denoted by T . Let W_1 and W_2 denote the subsets of sources demanded by the sink nodes t_1 and t_2 , respectively, where $t_1, t_2 \in T$. Denote the set of random

variables associated with the incoming edges of t_1 and t_2 by U_{E_1} and U_{E_2} , respectively. According to [84], we have $I(U_{E_1}; U_{E_2}) \geq \sum_{s \in W_1 \cap W_2} R_s$. The same result can be obtained by applying Lemma 4.2 as

$$\begin{aligned} I(U_{E_1}; U_{E_2}) &\geq H(\phi(U_{E_1}) \cap \phi(U_{E_2})) \\ &\stackrel{(a)}{\geq} H(W_1 \cap W_2) \\ &\geq \sum_{s \in (W_1 \cap W_2)} R_s, \end{aligned}$$

where (a) follows from the facts that $U_{E_1} \rightarrow W_1$ and $U_{E_2} \rightarrow W_2$. Therefore, we have $W_1 \subset \phi(U_{E_1})$, $W_2 \subset \phi(U_{E_2})$ and $(W_1 \cap W_2) \subset (\phi(U_{E_1}) \cap \phi(U_{E_2}))$.

While the exact value of the mutual information between two sets of source/edge random variables in a network depends on the network code design, the lower bound given in Lemma 4.2 is valid for any network code design. Moreover, it is the tightest lower bound of this form for general network coding. In other words, the mutual information lower bound given in Lemma 4.2 cannot be further improved by inclusion of any other source/edge random variables on the righthand side of (4.2). Formally, we have the following result:

Lemma 4.3. *Consider a network $G = (V, E)$ with corresponding FDG $\bar{G} = (\bar{V}, \bar{E})$. Let $x \in \bar{V}$, $A, B \subset \bar{V}$ and $x \notin (\phi(A) \cap \phi(B))$, then there exists a network code design such that*

$$I(A; B) < H(\phi(A) \cap \phi(B), x). \quad (4.3)$$

Proof. Since $x \notin (\phi(A) \cap \phi(B))$, we know that either $x \notin \phi(A)$ or $x \notin \phi(B)$, or both. If $x \notin \phi(A)$, we sketch the proof by showing that there exists a network code such that $H(A) = 0$ while $H(x) > 0$.

Let $W = \{s : Y_s \in A\}$ be the set of sources and $E_A = \{E : U_E \in A\}$ be the set of edges whose corresponding random variables are in the set A . We can design a network code such that $\sum_{s \in W} R_s = 0$, i.e., no information is transmitted over the edges in E_A . For such network code design, we thus have $H(A) = 0$. Since $x \notin \phi(A)$, there exists at least one path from Y_{W^c} to x in $\bar{G}$ without going through A . Hence, we have $H(x) > 0$. With such a code design, we have $I(A; B) = 0$ as $H(A) = 0$, and $H(\phi(A) \cap \phi(B), x) \geq H(x) > 0$. Therefore, we have $I(A; B) < H(\phi(A) \cap \phi(B), x)$. Similar arguments can be made for the case of $x \notin \phi(B)$. $\square$

Example 4.2. Consider the network shown in Figure 4.1 and assume that each edge has unit capacity. It can be easily verified that the capacity region of the network is given by $\{R_1 \leq 1; 2R_1 + R_2 \leq 2\}$. From Lemma 4.2, we have

$$I(U_1; U_2) \geq H(\phi(U_1) \cap \phi(U_2)) = H(Y_1). \quad (4.4)$$

With a network code construction that achieves the rate pair $(R_1, R_2) = (0, 2)$, we must have $H(U_i) = 1, i \in \{1, \dots, 4\}$, $H(Y_2) = 2$ while $I(U_1; U_2) = 0$. Therefore, the bound given in (4.4) cannot be improved by inclusion of any other random variable in the network.

4.2.2 Refined FD Bound

In order to utilize the lower bound derived in Lemma 4.2 to obtain a refined FD bound, we need to apply the following results. To simplify the notation, let $C_X \triangleq \sum_{s: Y_s \in X} R_s + \sum_{e: U_e \in X} C_e$ for $X \subset \bar{V}$.

Proposition 4.4. Consider a network $G = (V, E)$ with the corresponding FDG denoted by $\bar{G} = (\bar{V}, \bar{E})$. For $A, B \subset \bar{V}$, let $C \subset \bar{V}$ be a minimal set of random variables such that $(A, B, C) \rightarrow \bar{V}$ and D be a minimal set of random variables such that $(\phi(A) \cap \phi(B), D) \rightarrow \bar{V}$, then we have

$$2 \sum_{s \in S} R_s \leq C_A + C_B + C_C + C_D \quad (4.5)$$

Proof. Since the set of random variables C is chosen such that $(A, B, C) \rightarrow \bar{V}$, we have

$$\sum_{s \in S} R_s \leq H(A, B, C) \quad (4.6)$$

$$\begin{aligned} &\leq H(C) + H(A, B) \\ &= H(C) + H(A) + H(B) - I(A; B) \\ &\stackrel{(a)}{\leq} H(C) + H(A) + H(B) - H(\phi(A) \cap \phi(B)), \end{aligned} \quad (4.7)$$

where (a) follows from Lemma 4.2. Furthermore, since D is a set of random variables

such that $(\phi(A) \cap \phi(B), D) \rightarrow \bar{V}$, then

$$\sum_{s \in S} R_s \leq H(D) + H(\phi(A) \cap \phi(B)). \quad (4.8)$$

By summing up (4.7) and (4.8), we can get rid of the term $H(\phi(A) \cap \phi(B))$ and have an additional bound on the sum-rate:

$$\begin{aligned} 2 \sum_{s \in S} R_s &\leq H(A) + H(B) + H(C) + H(D) \\ &\leq C_A + C_B + C_C + C_D \end{aligned} \quad (4.9)$$

□

Remark 4.5. Note that if $\{A, B\}$ contains a subset of source random variables, (4.5) corresponds to a weighted sum-rate bound; in contrast, when only edge random variables are included in $\{A, B\}$, it gives a sum-rate bound, which cannot be obtained by any FD bound.

Remark 4.6. For $X \subset \bar{V}$, let $W = \{s : Y_s \in X\}$. Based on (2.15), the FD bound on $\sum_{s \in W} R_s$, denoted by C_W , can be obtained. Let $C'_X \triangleq C_W + \sum_{e: U_e \in X} C_e$, which is a constant dependent on the network topology. Then, the refined FD bound given by (4.5) *strictly* improves over the original FD bounds given in (2.15) if

$$C'_A + C'_B + C'_C + C'_D < 2C_S. \quad (4.10)$$

Remark 4.7. In the above steps, if we choose C and D such that together with (A, B) and $\phi(A) \cap \phi(B)$, respectively, they form a *cut-set* of the network, instead of a *maximal set*, then we can obtain a class of refined cut-set bounds. This is demonstrated in Example 4.3. The lower bound on the mutual information given in Lemma 4.2 is useful to obtain the refined version of many known bounds that are derived based on “edge-cuts”.

Example 4.3 (Refined cut-set bound). *Consider a 2-unicast network shown in Figure 4.2, where each edge has unit capacity. It can be easily verified that the cut-set bound [88, Chapter 15] of this network is given by*

$$\begin{aligned} R_1 &\leq 2; \\ R_2 &\leq 2; \\ R_1 + R_2 &\leq 3. \end{aligned} \quad (4.11)$$

Denote the source random variable associated with source s_i by $Y_i, i \in \{1, 2\}$ and the edge random variable associated with edge e_i by $U_i, i \in \{1, 2, \dots, 14\}$. Let $A = \{U_5, U_8\}$ and $B = \{Y_1, U_4\}$, then we have $\phi(A) \cap \phi(B) = \{Y_1, U_1, U_2, U_8, U_{11}, U_{12}, U_{13}\}$. Since the edges $\{e_4, e_5, e_8\}$ is a cut-set of the network, we can choose $C = \emptyset$. Moreover, as $Y_1 \in (\phi(A) \cap \phi(B))$ and $\{e_8, e_{14}\}$ disconnect t_2 from both sources, an appropriate choice of D is $\{U_{14}\}$. From (4.5), we can obtain a refined cut-set bound for this network as

$$R_1 + 2R_2 \leq 4, \quad (4.12)$$

which strictly improves over the original cut-set bounds in (4.11).

FIGURE 4.2: A 2-unicast network with unit edge capacity.

Remark 4.8. In [86], we characterized a set of Shannon-type inequalities that render weighted bounds for networks. The weighted bounds derived in [86] is a subset of the refined FD bounds presented in this chapter. Specifically, if the conditions specified in Section 4.1.2 are satisfied, we can obtain the same weighted bound as that in [86] by choosing $A = U_X, B = \{Y_{W^c}, U_Z\}$ in Proposition 4.4. Moreover, there exists network instances such that the refined FD bounds are *strictly* tighter; such a network will be shown in Example 4.6.

4.3 An Algorithm for the Refined FD Bounds

In the preceding subsection, we have presented the procedures to find the refined FD bounds based on lower bounding the mutual information between two sets of random variables, A and B , of an FDG. However, the tightness of the resulting bounds are closely related to the choice of A and B . Exhaustively searching for all possible sets of random variables A and B in a network incurs high computational complexity, which is undesired. In this section, we provide some guidelines for selecting A and B and present an efficient method for determining the associated edge random variables C and D , based on which an algorithm is designed to find the refined FD bounds.

4.3.1 Guidelines for Determining A and B

In this subsection, we give some rules for the selection of A and B so that the searching space is reduced. The improper choices of A and B are excluded either by showing that they can be replaced by a strictly better alternative, or by proving that the refined FD bound generated based on them does not provide any improvement over the original FD bounds.

Rule 1: A and B must be irreducible.

Proof. If A is reducible, then we can find $A_d \subsetneq A$ such that $A_d \rightarrow A$, and thus $\phi(A_d) = \phi(A)$. With the same choice of B, C, D and replacing A by A_d , a strictly tighter bound can be found, which is given by $2 \sum_s R_s \leq C_{A_d} + C_B + C_C + C_D$. Similarly, we can show that B must also be irreducible. $\square$

Rule 2: A, B must not be maximal set, i.e., $A \not\rightarrow \bar{V}$ and $B \not\rightarrow \bar{V}$.

Proof. If A is a maximal set, from (2.15), we can conclude that $C'_A \geq C_S$. At the same time, since $(\phi(A) \cap \phi(B), D) \rightarrow \bar{V}$, we must have $(B, D) \rightarrow \bar{V}$ as $(B, D) \rightarrow (\phi(B), D) \rightarrow (\phi(A) \cap \phi(B), D)$. Therefore, we can conclude that $C'_B + C'_D \geq C_S$. Combining with the condition $C'_A \geq C_S$, it gives $C'_A + C'_B + C'_D \geq 2C_S$, which violates (4.10) and thus the refined FD bound will not improve over the original FD bound. Similar arguments apply for the choice of B . $\square$

Rule 3: $\phi(A) \neq A$ and $\phi(B) \neq B$.

Proof. Denote $(\phi(A) \cap \phi(B))$ by J . If $\phi(A) = A$, then we have $J = (A \cap \phi(B))$ and thus $J \subseteq A$ and $J \subseteq \phi(B)$. Therefore, we have $B \rightarrow \phi(B) \rightarrow J$.

Since $(A, B, C) \rightarrow \bar{V}$, which can be equivalently written as $(A \setminus J, J, B, C) \rightarrow \bar{V}$, we must also have $(A \setminus J, B, C) \rightarrow \bar{V}$ as $B \rightarrow J$. Therefore, $(A \setminus J, B, C)$ is a maximal set and we can conclude that

$$C'_{A \setminus J} + C'_B + C'_C \geq C_S. \quad (4.13)$$

Moreover, as $(J, D) \rightarrow \bar{V}$, we also have

$$C'_J + C'_D \geq C_S. \quad (4.14)$$

Summing up (4.13) and (4.14) gives

$$\begin{aligned} 2C_S &\geq C'_{A \setminus J} + C'_B + C'_C + C'_J + C'_D \\ &= C'_A + C'_B + C'_C + C'_D, \end{aligned}$$

which violates (4.10) and thus the refined FD bound will not improve over the original FD bound. Similarly, it can be proved that we should have $\phi(B) \neq B$. $\square$

For $X \subset \bar{V}$, denote as $\text{Up}(X)$ the set of immediate upstream nodes of X , i.e., $\text{Up}(X) \triangleq \{v' \in \bar{V} : (v', v) \in \bar{E}, \forall v \in X\}$. Based on the definition of FDG in Definition 2.2, we have $\text{Up}(X) \rightarrow X$.

According to Rule 3, $\phi(A) \neq A$, i.e., $\exists v \in (\phi(A) \setminus A)$, such that $\text{Up}(v) \subseteq A$. In other words, the set A should contain all the immediate upstream nodes of at least one node in the FDG. Similar arguments apply for the choice of the set B .

Rule 4: $\nexists A_d \subsetneq A$ such that $(\phi(A) \setminus \phi(A_d)) \subset A$ and $\nexists B_d \subsetneq B$ such that $(\phi(B) \setminus \phi(B_d)) \subset B$.

Proof. Rule 4 can be shown by contradiction. If there exists $A_d \subsetneq A$ such that $(\phi(A) \setminus \phi(A_d)) \subset A$, then we can obtain a better refined FD bound by replacing A with A_d in Proposition 4.4. Let $J \triangleq (\phi(A) \setminus \phi(A_d))$, i.e., $\phi(A) = (J, \phi(A_d))$. Therefore, $\phi(A) \cap \phi(B) = (\phi(A_d) \cap \phi(B)) \cup (J \cap \phi(B))$. Denote $J \cap \phi(B)$ by K , which is a subset of J and may be an empty set. We have $B \rightarrow \phi(B) \rightarrow K$.

With A, C, D replaced by $A_d, (C, J \setminus K)$ and (D, K) , respectively, it can be verified that $(A_d, B, C, J \setminus K) \rightarrow \bar{V}$ as $B \rightarrow K$ and $(J, A_d) \rightarrow \phi(A) \rightarrow A$. Moreover,

we have $(\phi(A_d) \cap \phi(B), D, K) \rightarrow \bar{V}$ as $(\phi(A) \cap \phi(B)) = (\phi(A_d) \cap \phi(B)) \cup K$. Therefore, the refined FD bound is given by

$$\begin{aligned} 2 \sum_{s \in W} R_s &\leq C_{A_d} + C_B + C_C + C_{J \setminus K} + C_D + C_K \\ &= C_{(A_d, J)} + C_B + C_C + C_D. \end{aligned} \quad (4.15)$$

From the definition of J , we know that $J \cap \phi(A_d) = \emptyset$. As $A_d \subseteq \phi(A_d)$, we must also have $J \cap A_d = \emptyset$. Moreover, since $A_d \subseteq A$ and $J \subseteq A$, it can be concluded that $(A_d, J) \subseteq A$. Therefore, the bound in (4.15) is at least as tight as the bound given in (4.5).

By symmetry, similar arguments can be made for the choice of set B . $\square$

Note that the main idea given by Rule 4 is that we should not include any node in A (or B) that does not help to enlarge $\phi(A)$ (or $\phi(B)$).

Based on Rule 3 and Rule 4, we may reduce the searching space of the sets A and B to the following collection without compromising the tightness of the resulting refined FD bounds:

$$\mathcal{O} \triangleq \{\text{Up}(X) : X \subset \bar{V}\}. \quad (4.16)$$

Moreover, Rule 1 and Rule 2 exclude those sets that are reducible or maximal. Let $\mathcal{A}$ denote the the collection of non-maximal irreducible sets of a network. Then, the searching space for sets A and B is given by

$$\mathcal{S} \triangleq \mathcal{O} \cap \mathcal{A}. \quad (4.17)$$

Since the dimensions of both $\mathcal{O}$ and $\mathcal{A}$ grow exponentially with the network size, it is desirable to further reduce the searching space for sets A and B in Proposition 4.4 by considering sub-optimal schemes. In (4.16), when $|X|$ increases, it is more likely that $\text{Up}(X)$ is a reducible set. Hence, it is reasonable for us to only consider the case when $|X| = 1$, in which the collection $\mathcal{O}$ reduces to

$$\mathcal{O}' = \{\text{Up}(v) : v \in \bar{V}\}. \quad (4.18)$$

The size of $\mathcal{O}'$, given by $|\bar{V}|$, grows linearly with the size of the FDG. Note that if the reduced FDG introduced in Chapter 3 is used, all elements in $\mathcal{O}'$ are irreducible,

i.e., $\mathcal{O} \subseteq (\mathcal{A} \cup \mathcal{M})$, as shown in Lemma 4.9. Hence, the searching space can be represented by

$$\mathcal{S}' = \mathcal{O} \setminus \mathcal{M}. \quad (4.19)$$

Lemma 4.9. *In a reduced FDG $\bar{G} = (\bar{V}, \bar{E})$, all the elements in (4.18) are irreducible.*

Proof. Suppose, on the contrary, that there exists a node $v \in \bar{V}$ such that $\text{Up}(v)$ is reducible, i.e., $\exists U_i \in \text{Up}(v)$ such that $(\text{Up}(v) \setminus U_i) \rightarrow U_i$. According to the edge reduction in the FDG introduced in Section 3.2.2 of Chapter 3, the edge from U_i to v will be removed and hence U_i is no longer an element in $\text{Up}(v)$, which contradicts with the assumption $U_i \in \text{Up}(v)$. Hence, in the reduced FDG, $\text{Up}(v)$ must be irreducible. $\square$

Example 4.4. *Consider the network shown in Figure 4.3(a) and its corresponding reduced FDG shown in Figure 4.3(b).*

FIGURE 4.3: (a) A network example. (b) The corresponding reduced FDG.

Note that $|\bar{V}| = 7$, but the edge random variable U_3 and U_4 have the same set of upstream nodes. Hence, there are 6 elements in $\mathcal{O}'$, which are given by

$$\{U_3, U_{12}\}, \{U_{10}, U_{12}, U_{14}\}, \{Y_2\}, \{Y_1, U_3\}, \{U_4, U_{10}\}, \{Y_1, U_3, U_4\}. \quad (4.20)$$

It can be verified that all sets in (4.20) are irreducible, which is consistent with Lemma 4.9. By removing the maximal irreducible sets $\{U_{10}, U_{12}, U_{14}\}$ and $\{Y_1, U_3, U_4\}$, the collection of proper choice of A and B is given by

$$\mathcal{S}' = \{Y_2, \{Y_1, U_3\}, \{U_4, U_{10}\}, \{U_3, U_{12}\}\}. \quad (4.21)$$

For the simple network considered in this example, the searching space given in (4.21) is actually the same as the optimal searching space defined in (4.17).

4.3.2 Determining the Capacities of C and D

For any given random variables A and B in a network, the corresponding sets C and D are defined as the minimum capacity of the maximal irreducible set in the subgraph of $\bar{G}$ induced by $\bar{V} \setminus (\phi(A) \cup \phi(B))$ and $\bar{V} \setminus (\phi(A) \cap \phi(B))$, respectively. Maximal irreducible sets of a graph can be found using Algorithm 2 in [38] and we will show that the maximal irreducible sets of a subgraph can be obtained directly from those of the original graph.

Lemma 4.10. *Consider a FDG $\bar{G} = (\bar{V}, \bar{E})$ and let the collection of all maximal irreducible sets of $\bar{G}$ denoted by $\mathcal{M}$. Then, the collection of all maximal irreducible sets of the subgraph induced by $\bar{G} \setminus U$, where $U \subset \bar{V}$, is given by all the unique elements in $\{X \setminus \phi(U) : X \in \mathcal{M}\}$.*

Proof. This lemma can be proved by contradiction. Assume that $P \subset (\bar{V} \setminus \phi(U))$ is a maximal irreducible set for the subgraph induced by $\bar{G} \setminus U$, but it is not an element of $\{X \setminus \phi(U) : X \in \mathcal{M}\}$. Then, there must exist a subset of $\phi(U)$, denoted by U_Z , such that $U_Z \cup P$ is a maximal irreducible set of $\bar{G}$, but is not within $\mathcal{M}$, which contradicts with the definition of $\mathcal{M}$. On the other hand, if there is a unique element Q in $\{X \setminus \phi(U) : X \in \mathcal{M}\}$ that is not a maximal irreducible set of the subgraph induced by $\bar{G} \setminus U$, i.e., $Q \not\rightarrow (\bar{V} \setminus U)$, we must have $(Q, \phi(U)) \not\rightarrow \bar{V}$. Clearly, this contradicts with the fact that all elements in $\mathcal{M}$ are maximal irreducible sets. $\square$

From Lemma 4.10, C and D can be determined directly from the collection of maximal irreducible sets of the original FDG $\bar{G} = (\bar{V}, \bar{E})$. Furthermore, based on the description in Section 4.2.2, the choices of C and D may be non-unique. Nevertheless, only the capacities of C and D are required for deriving the refined FD bound in (4.5).

Example 4.5. Consider the network shown in Figure 4.3(a) and its corresponding reduced FDG shown in Figure 4.3(b). All maximal irreducible sets of the FDG can be obtained using Algorithm 2 in [38], which are

$$\begin{aligned} &\{Y_1, U_3, U_4\}, \{Y_2, U_{10}\}, \{Y_2, U_{12}\}, \{U_3, U_4, U_{10}\}, \{U_3, U_4, U_{12}\}, \\ &\{U_3, U_{12}, U_{14}\}, \{U_{10}, U_{12}, U_{14}\}. \end{aligned} \quad (4.22)$$

The FD bound can be obtained for this network based on Theorem 2.5 as

$$\begin{aligned} R_1 &\leq 1 \\ R_2 &\leq 2 \\ R_1 + R_2 &\leq 3. \end{aligned} \quad (4.23)$$

For instance, if $A = Y_2$ and $B = \{Y_1, U_3\}$, it can be verified that $\phi(A) \cup \phi(B) = \bar{V}$, hence C must be an empty set. On the other hand, $\phi(A) \cap \phi(B) = U_3$ and hence D can be selected from

$$\{Y_1, U_4\}, \{Y_2, U_{10}\}, \{Y_2, U_{12}\}, \{U_4, U_{10}\}, \{U_4, U_{12}\}, \{U_{12}, U_{14}\}, \{U_{10}, U_{12}, U_{14}\}.$$

Since the source is assumed to be unbounded, the minimum capacity of D is 2, which is given by either $D = \{U_4, U_{10}\}$, $D = \{U_4, U_{12}\}$ or $D = \{U_{12}, U_{14}\}$.

4.3.3 A Summarized Algorithm

In this subsection, we present an algorithm that summarizes the procedures for producing the refined FD bounds for general multi-session networks with its reduced FDG as an input.

Remark 4.11. Algorithm 2 can be easily modified to consider the sub-optimal searching space given in (4.18). By symmetry, the choices of A and B can be interchanged without affecting the resulting refined FD bounds. Therefore, the total number of resulting refined FD bounds is upper bounded by $\binom{|S|}{2}$.

Example 4.6. Consider the network shown in Figure 4.3(a) and its corresponding reduced FDG shown in Figure 4.3(b). The collection of proper choice of A, B is given by (4.21). With Algorithm 2, there are 6 possible refined FD bounds for this network, which are listed in Table 4.1.

Algorithm 2 RefinedFDB($\bar{G} = (\bar{V}, \bar{E})$)

```

Initialize:  $\mathcal{S} = \emptyset$ 
 $\mathcal{A} = \{Z \subset \bar{V} : \nexists Z_d \subsetneq Z, \text{ such that } Z_d \rightarrow Z \ \& \ Z \not\rightarrow \bar{V}\}$ 
for  $X \subset \bar{V}$  do
  if  $\text{Up}(X) \in \mathcal{A}$  then
     $\mathcal{S} = \mathcal{S} \cup \text{Up}(X)$ 
  end if
end for
for  $A, B \in \mathcal{S}$  and  $A \neq B$  do
   $C = \arg \min_{C:(A,B,C) \rightarrow \bar{V}} C'_C$ 
   $D = \arg \min_{D:(\phi(A) \cap \phi(B), D) \rightarrow \bar{V}} C'_D$ 
  Output  $2 \sum_{s \in W} R_s \leq C_A + C_B + C_C + C_D$ 
end for

```

TABLE 4.1: A list of refined FD bounds

A	B	C	D	Refined FD bound
Y_2	(Y_1, U_3)	$\emptyset$	(U_4, U_{10})	$R_1 + R_2 \leq 3$
Y_2	(U_4, U_{10})	$\emptyset$	(U_3, U_{10})	$2R_1 + R_2 \leq 4$
Y_2	(U_3, U_{12})	$\emptyset$	(U_4, U_{10})	$2R_1 + R_2 \leq 4$
(Y_1, U_3)	(U_4, U_{10})	$\emptyset$	(U_3, U_4)	$R_1 + 2R_2 \leq 5$
(Y_1, U_3)	(U_3, U_{12})	U_4	U_4	$R_1 + 2R_2 \leq 5$
(U_4, U_{10})	(U_3, U_{12})	$\emptyset$	U_{14}	$R_1 + R_2 \leq 2.5$

Note that the last bound given in Table 4.1 strictly improves over the original FD bound given in (4.23). In fact, it can be shown that such improved bounds meet the capacity of this network, hence it is tight.

4.4 Application on Degraded Multicast Networks

The refined FD bounds generalize some graph theoretic bounds considered in the literature for networks with special topologies [84, 85, 89]. In this section, we apply the refined FD bounds on a class of 3-user degraded multicast networks shown in Fig 4.4, where three sources are available at the common source node s and the sink node t_i wants to decode a subset of the sources $\{Y_1, \dots, Y_i\}$, for $i = 1, 2, 3$. For any double-unicast network, a corresponding 3-user degraded multicast network which has the same solvability over any alphabet can be constructed. Since determining the capacity region for general double-unicast networks is proven to be NP-hard

[90], characterizing the capacity for general 3-user degraded multicast networks is also difficult.

FIGURE 4.4: 3-user degraded multicast network.

Let $\mathcal{C}_i$ denote a min-cut between s and t_i and $|\mathcal{C}_i| \triangleq \sum_{e \in \mathcal{C}_i} C_e$. The edge-cut bound of 3-user degraded multicast networks is given by

$$\begin{aligned} R_1 &\leq \min |\mathcal{C}_1| \\ R_1 + R_2 &\leq \min |\mathcal{C}_2| \\ R_1 + R_2 + R_3 &\leq \min |\mathcal{C}_3|. \end{aligned} \tag{4.24}$$

Let $A = \mathcal{C}_1$ and $B = \mathcal{C}_2$ in Proposition 4.4. We have

$$I(A; B) = I(\mathcal{C}_1; \mathcal{C}_2) \geq H(\phi(\mathcal{C}_1) \cap \phi(\mathcal{C}_2)) \tag{4.25}$$

$$\stackrel{(a)}{\geq} H(Y_1), \tag{4.26}$$

where (a) follows as $\mathcal{C}_1 \rightarrow Y_1$ and $\mathcal{C}_2 \rightarrow (Y_1, Y_2)$. Given A and B , the optimal choice of C and D can be obtained from Proposition 4.4 based on the network topology. Without knowing the exact topology of the network, a sub-optimal choice of C and D can be determined as $C = \mathcal{C}_3 \setminus (\mathcal{C}_1 \cup \mathcal{C}_2)$ and $D = (Y_2, Y_3)$. The resulting refined

FD bound is then given by

$$\begin{aligned} 2R_1 + R_2 + R_3 &\leq |\mathcal{C}_1| + |\mathcal{C}_2| + |\mathcal{C}_3 \setminus (\mathcal{C}_1 \cup \mathcal{C}_2)| \\ &= |\mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3| + |\mathcal{C}_1 \cap \mathcal{C}_2|. \end{aligned} \quad (4.27)$$

By symmetry, two other refined FD bounds can be obtained with $A = \mathcal{C}_1, B = \mathcal{C}_3$ and $A = \mathcal{C}_2, B = \mathcal{C}_3$, respectively, which yields

$$2R_1 + R_2 + R_3 \leq |\mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3| + |\mathcal{C}_1 \cap \mathcal{C}_3| \quad (4.28)$$

$$2R_1 + 2R_2 + R_3 \leq |\mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3| + |\mathcal{C}_2 \cap \mathcal{C}_3| \quad (4.29)$$

By considering all possible minimum cuts, the following refined FD bounds can be obtained

$$\begin{aligned} 2R_1 + R_2 + R_3 &\leq \min_{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3} \{|\mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3| + |\mathcal{C}_1 \cap \mathcal{C}_2|\} \\ 2R_1 + R_2 + R_3 &\leq \min_{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3} \{|\mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3| + |\mathcal{C}_1 \cap \mathcal{C}_3|\} \\ 2R_1 + 2R_2 + R_3 &\leq \min_{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3} \{|\mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3| + |\mathcal{C}_2 \cap \mathcal{C}_3|\} \end{aligned} \quad (4.30)$$

Together with the edge-cut bounds in (4.24), the refined FD bounds in (4.30) is tight for 3-user degraded multicast *combinational* networks shown in Figure 4.5 [85][91].

Specifically, consider a realization of the combinational network shown in Figure 4.5 with capacities of the edges $(e_1, e_2, e_3, e_{12}, e_{13}, e_{23}, e_{123})$ given by $(2, 2, 2, 1, 5, 5, 0)$. The edge-cut bounds are compared with the refined FD bounds in Figure 4.6. The rate tuple $(R_1, R_2, R_3) = (8, 0, 4)$ satisfies the original edge cut bounds in (4.24), but it is actually not achievable as it violates the refined FD bound in (4.27): $2R_1 + R_2 + R_3 \leq 18$.

4.5 Chapter Summary

In this chapter, we have derived a new class of information-theoretic outer bounds, refined FD bounds, for network coding capacity based on lower bounding the mutual information between two sets of source/edge random variables in a network. The lower bounds on the mutual information are obtained by exploring the functional

FIGURE 4.5: 3-user degraded multicast combinational network.

dependence relationships among the source/edge random variables, hence they are applicable for general multi-session networks. These new class of outer bounds improve over the original edge-cut bounds by including different rate weights. Moreover, an efficient algorithm has been designed for computing the refined FD bounds of network coding capacity. The tightness of the refined FD bounds has been demonstrated with a class of 3-user degraded multicast networks.

FIGURE 4.6: Compare the edge-cut bound with the refined FD bound for a 3-user degraded multicast combinational network.

Chapter 5

Linear Network Coding Schemes for Multi-Session Networks

While various outer bounds for network coding capacity have been proposed in the literature, as discussed in the preceding two chapters, only limited results on the inner bounds are available for general networks. This is mainly due to the fact that network code design is largely dependent on the network topology, and hence it is difficult to find a universal network code that works for general networks. Therefore, most of the existing works on network code constructions focus on specific classes of networks only, such as the single-session networks [29], single-source two-receiver networks [48, 49], and double-unicast networks [50, 67, 92]. In this chapter, we first present a linear network coding scheme for double-unicast networks, where the intermediate nodes perform RLNC and the source and sink nodes optimize their strategies to maximize the achievable rate region. Furthermore, a joint routing and RLNC scheme is proposed for degraded 2-message multicast networks and double-unicast networks based on the achievable regions obtained with RLNC. The proposed joint routing and RLNC scheme can be extended for general multi-session networks.

5.1 Achievable Region for Double-Unicast Networks with Linear Network Coding

Multiple-unicast networks, in which each message is generated by one source node and demanded by exactly one sink node [93], is a class of multi-session networks

of particular interest. It has been found that for any directed acyclic multi-session network, there exists a corresponding multiple-unicast network that has the same solvability over any alphabet [78]. As the simplest case of general multiple-unicast networks where network coding is relevant, double-unicast networks with two source-sink pairs have received significant research interests recently. For such networks, the achievability of the rate pair $(1, 1)$ has been established with various methods [92, 94, 95]. On the other hand, since the capacity region for double-unicast networks remains unknown, many works have been devoted to obtaining better outer and inner bounds for the capacity region [34, 40, 45, 50, 67, 96–98].

In [40], based on the idea of information dominance, a generalized network sharing outer bound for double-unicast networks was introduced. Later, such a bound was improved in [96] by including some additional weighted bounds. In [34], a linear programming (LP) bound was derived based on solving LP optimization problems, as described in Section 2.3. LP bound is the tightest computable outer bound known to date and it was conjectured to meet the capacity region for double-unicast networks [40]. Recently, this conjecture was proven to be incorrect in [90] by showing that the double-unicast problem is as hard as any general multiple-unicast problem.

As for the inner bounds of double-unicast networks, most achievable schemes are based on some suboptimal, yet simple to implement, linear network coding strategies. In [45], the design problem was formulated as an LP optimization problem by packing the butterfly structures in the network. In [50], an achievable region was obtained by the so-called “rate-exchange” method, where starting from the single-user rate for one of the users, a non-zero rate for the other user is achieved by sacrificing the single-user rate via interference nulling. In [67, 97, 98], achievable rate regions were obtained by applying RLNC at each intermediate node, together with some linear/nonlinear precessing at the source and sink nodes.

In this chapter, we consider the achievable rate region for double-unicast networks by assuming that *RLNC* is employed by the intermediate nodes, which, though suboptimal in general, is of practical interest due to its simplicity, distributiveness, and robustness to network dynamics. Under such an assumption, the design problem reduces to finding the optimal strategies at the source and sink nodes so that the achievable rate region is maximized. Different from that in single-session networks, the optimal strategy at each source node for double-unicast networks needs to achieve a good balance between maximizing the rate to its corresponding

sink node and minimizing the inter-session interference to the other. For a given set of network coding coefficients chosen by the intermediate nodes, such a scenario can be modeled as a deterministic interference channel (IC), whose capacity region has been obtained in [58]. For the particular class of *linear* deterministic IC of our interest, in which the outputs and the interferences are *linear* deterministic functions of the inputs, we show that the capacity region specified in [58] can be achieved by linear precoding and decoding at the source and sink nodes, respectively. The main techniques used in the proposed linear scheme are Han-Kobayashi rate splitting [99] and concatenated precoding composed of zero-forcing and random block-diagonal precoding matrices. Specifically, the information bearing symbols are first split into a common part which is decodable at both sink nodes, and a private part which is decodable at the designated sink node only. The resulting data symbols are then precoded by concatenated precoding matrices, which are carefully designed such that the interferences caused by the undesired data symbols are minimized at both sink nodes. For any given set of coding coefficients chosen by the intermediate nodes, the proposed linear scheme for the source and sink nodes thus gives the maximum achievable rate region for double-unicast network. We further derive a suboptimal but easy to compute achievable region that is independent of the network coding coefficients used at the intermediate nodes, which is instead specified by the min-cuts of the network. We analytically show that even this suboptimal rate region is strictly larger than the existing regions obtained in the literature [50, 67, 97, 98].

A similar achievable region is obtained in [98] by directly applying the capacity result of the deterministic IC in [58]. However, the achievable scheme given in [58] is based on the information-theoretic superposition coding, which is difficult to be implemented in practice. In contrast, by exploiting the *linear* structure of the deterministic IC, our proposed achievable scheme requires simple linear precoding and decoding at the source and sink nodes, respectively, which is easy to be implemented. Moreover, the rate region expressed in terms of the min-cuts of the network derived in this chapter is shown to be strictly larger than that obtained in [98].

5.1.1 System Model

We consider double-unicast networks with two source nodes $S = \{s_1, s_2\} \subset V$ and two sink nodes $T = \{t_1, t_2\} \subset V$, where s_1 and s_2 are intended to send independent messages $\mathbf{d}_1$ and $\mathbf{d}_2$ to t_1 and t_2 , respectively. Without loss of generality, we assume unit edge capacity, i.e., each edge is capable of carrying one symbol per time slot. By Merger's theorem, the minimum cut between the source subset $S_{N_1} \subseteq S$ and the sink subset $T_{N_2} \subseteq T$ equals to the number of edge disjoint paths from S_{N_1} to T_{N_2} , denoted as $k_{N_1-N_2}$, where $N_1, N_2 \subseteq \{1, 2\}$. It then follows that the number of outgoing edges from the source node s_i , denoted as m_i , is no smaller than k_{i-12} . Furthermore, if $m_i > k_{i-12}$, we may add an imaginary source s'_i connecting to s_i via k_{i-12} edges without affecting the network capacity. Therefore, without loss of generality, we may assume that $m_i = k_{i-12}$, i.e., there are k_{i-12} outgoing edges from source s_i . Similarly, the number of incoming edges to sink t_i , denoted as n_i , is assumed equal to k_{12-i} , $i \in \{1, 2\}$ [67]¹.

With RLNC performed at each intermediate node, the input-output relation between the source and sink nodes of double-unicast networks is given as [1]²

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{H}_{11}\mathbf{x}_1 + \mathbf{H}_{12}\mathbf{x}_2 \\ \mathbf{y}_2 &= \mathbf{H}_{21}\mathbf{x}_1 + \mathbf{H}_{22}\mathbf{x}_2, \end{aligned} \tag{5.1}$$

where $\mathbf{y}_i \in \mathbb{F}_q^{k_{12-i}}$ denotes the symbols received by sink t_i ; $\mathbf{x}_i \in \mathbb{F}_q^{k_{i-12}}$ consists of the symbols sent over the k_{i-12} outgoing edges of s_i and it is generated from the information bearing symbols $\mathbf{d}_i$; and $\mathbf{H}_{ji} \in \mathbb{F}_q^{k_{12-j} \times k_{i-12}}$ represents the transition matrix from s_i to t_j , which is determined by the network topology and the coding coefficients used at the intermediate nodes [1].

It is observed that the input-output relation given in (6.22) belongs to a special class of the deterministic IC considered in [58], in which the outputs $\mathbf{y}_1$ and $\mathbf{y}_2$ and the interferences $\mathbf{z}_1 = \mathbf{H}_{21}\mathbf{x}_1$ and $\mathbf{z}_2 = \mathbf{H}_{12}\mathbf{x}_2$ are *linear* deterministic functions of the inputs $\mathbf{x}_1$ and $\mathbf{x}_2$. By applying the result in [58], the capacity region of (6.22)

¹The techniques presented in this chapter can be extended to the cases with arbitrary m_i 's and n_i 's (see Remark 5.9 in Section 6.1.3).

²Throughout this chapter, the input and output symbols are represented by column vectors.

is given by the union of the set of all rate pairs (R_1, R_2) satisfying

$$R_1 \leq H(\mathbf{y}_1|\mathbf{z}_2) \quad (5.2)$$

$$R_2 \leq H(\mathbf{y}_2|\mathbf{z}_1) \quad (5.3)$$

$$R_1 + R_2 \leq H(\mathbf{y}_1) + H(\mathbf{y}_2|\mathbf{z}_1, \mathbf{z}_2) \quad (5.4)$$

$$R_1 + R_2 \leq H(\mathbf{y}_1|\mathbf{z}_1, \mathbf{z}_2) + H(\mathbf{y}_2) \quad (5.5)$$

$$R_1 + R_2 \leq H(\mathbf{y}_1|\mathbf{z}_1) + H(\mathbf{y}_2|\mathbf{z}_2) \quad (5.6)$$

$$2R_1 + R_2 \leq H(\mathbf{y}_1) + H(\mathbf{y}_1|\mathbf{z}_1, \mathbf{z}_2) + H(\mathbf{y}_2|\mathbf{z}_2) \quad (5.7)$$

$$R_1 + 2R_2 \leq H(\mathbf{y}_1|\mathbf{z}_1) + H(\mathbf{y}_2) + H(\mathbf{y}_2|\mathbf{z}_1, \mathbf{z}_2), \quad (5.8)$$

over all input distributions $\mathbf{x}_1 \in \mathbb{F}_q^{k_1-12}$ and $\mathbf{x}_2 \in \mathbb{F}_q^{k_2-12}$.

In the next subsection, we will show that for the *linear* deterministic IC given by (6.22), the capacity region specified in (5.2)-(5.8) can be achieved with simple *linear* precoding and decoding at the source and the sink nodes, which thus provides a rate region achievable by linear schemes for double-unicast networks.

5.1.2 A Linear Network Coding Scheme

Figure 5.1 provides a schematic overview for our proposed linear network coding scheme, which consists of rate-splitting, block-diagonal precoding, and invertible linear transformations at the source nodes, and invertible linear transformations and Gaussian eliminations at the sink nodes.

(a) Rate splitting and linear precoding at s_1 .

(b) Decoding process at t_1

FIGURE 5.1: The proposed achievability scheme with linear precoding and decoding at s_1 and t_1 , respectively.

Before presenting the detailed designs for each block in Figure 5.1, we first reformulate the capacity region specified in (5.2)-(5.8) in terms of the channel ranks by applying Lemma 5.1

Lemma 5.1. *Let $\mathbf{x}$ be a random vector of dimension $l \times 1$ and $\mathbf{A} \in \mathbb{F}_q^{p_1 \times l}$ and $\mathbf{B} \in \mathbb{F}_q^{p_2 \times l}$ are two given matrices. We then have $H(\mathbf{Ax}|\mathbf{Bx}) \leq \text{rank} \left(\begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \right) - \text{rank}(\mathbf{B})$, where the equality holds when the entries of $\mathbf{x}$ are independently and uniformly chosen from the finite field $\mathbb{F}_q$.*

Proof. Lemma 5.1 can be proved by applying similar techniques as that in Section III of [89]. For notational convenience, denote by r_A , r_B and r_{AB} the ranks of the matrices $\mathbf{A}$, $\mathbf{B}$ and $\begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix}$, respectively. Let $\mathbf{N}_1 \in \mathbb{F}_q^{l \times (l-r_{AB})}$ be a matrix whose columns form a basis for $\mathcal{N} \left(\begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \right)$. Then we can find a matrix $\mathbf{N}_2$ of size $l \times (r_{AB} - r_B)$ such that the columns of $\mathbf{N}_1$ and $\mathbf{N}_2$ form a basis for $\mathcal{N}(\mathbf{B})$. Moreover, let $\mathbf{N}_3 \in \mathbb{F}_q^{l \times r_B}$ be the basis of $\mathcal{R}(\mathbf{B}^T)$. Therefore, the columns in $\begin{bmatrix} \mathbf{N}_1 & \mathbf{N}_2 & \mathbf{N}_3 \end{bmatrix}$ span the l -dimensional space. As a consequence, any given $\mathbf{x} \in \mathbb{F}_q^{l \times 1}$ can be represented as $\mathbf{x} = \begin{bmatrix} \mathbf{N}_1 & \mathbf{N}_2 & \mathbf{N}_3 \end{bmatrix} \mathbf{x}'$. Thus,

$$\begin{aligned}
 & H(\mathbf{Ax}|\mathbf{Bx}) \\
 &= H \left(\mathbf{A}[\mathbf{N}_1 \quad \mathbf{N}_2 \quad \mathbf{N}_3] \mathbf{x}' \mid \mathbf{B}[\mathbf{N}_1 \quad \mathbf{N}_2 \quad \mathbf{N}_3] \mathbf{x}' \right) \\
 &= H(\mathbf{AN}_2 \mathbf{x}'_2 + \mathbf{AN}_3 \mathbf{x}'_3 \mid \mathbf{BN}_3 \mathbf{x}'_3) \\
 &\stackrel{(a)}{=} H(\mathbf{AN}_2 \mathbf{x}'_2 \mid \mathbf{BN}_3 \mathbf{x}'_3, \mathbf{x}'_3) \\
 &\leq H(\mathbf{AN}_2 \mathbf{x}'_2) \leq \text{rank}(\mathbf{AN}_2) \\
 &= \text{rank}(\mathbf{N}_2) = r_{AB} - r_B.
 \end{aligned}$$

where (a) follows since $\mathbf{x}'_3$ can be uniquely determined from $\mathbf{BN}_3 \mathbf{x}'_3$.

When all the entries of $\mathbf{x}$ are independently and uniformly distributed in $\mathbb{F}_q$, so are those in $\mathbf{x}'$. In this case, all the inequalities above become equalities.

This completes the proof of Lemma 5.1. $\square$

With Lemma 5.1, the region given in (5.2)-(5.8) for the linear deterministic IC

can be equivalently written as ³

$$R_1 \leq \text{rank}(\mathbf{H}_{11}) \quad (5.9)$$

$$R_2 \leq \text{rank}(\mathbf{H}_{22}) \quad (5.10)$$

$$R_1 + R_2 \leq \text{rank} \left(\begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \end{bmatrix} \right) + \text{rank} \left(\begin{bmatrix} \mathbf{H}_{12} \\ \mathbf{H}_{22} \end{bmatrix} \right) - \text{rank}(\mathbf{H}_{12}) \quad (5.11)$$

$$R_1 + R_2 \leq \text{rank} \left(\begin{bmatrix} \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix} \right) + \text{rank} \left(\begin{bmatrix} \mathbf{H}_{11} \\ \mathbf{H}_{21} \end{bmatrix} \right) - \text{rank}(\mathbf{H}_{21}) \quad (5.12)$$

$$R_1 + R_2 \leq \text{rank} \left(\begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{0} \end{bmatrix} \right) + \text{rank} \left(\begin{bmatrix} \mathbf{H}_{21} & \mathbf{H}_{22} \\ \mathbf{0} & \mathbf{H}_{12} \end{bmatrix} \right) - \text{rank}(\mathbf{H}_{21}) - \text{rank}(\mathbf{H}_{12}) \quad (5.13)$$

$$2R_1 + R_2 \leq \text{rank} \left(\begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \end{bmatrix} \right) + \text{rank} \left(\begin{bmatrix} \mathbf{H}_{11} \\ \mathbf{H}_{21} \end{bmatrix} \right) + \text{rank} \begin{bmatrix} \mathbf{H}_{21} & \mathbf{H}_{22} \\ \mathbf{0} & \mathbf{H}_{12} \end{bmatrix} \\ - \text{rank}(\mathbf{H}_{21}) - \text{rank}(\mathbf{H}_{12}) \quad (5.14)$$

$$R_1 + 2R_2 \leq \text{rank} \left(\begin{bmatrix} \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix} \right) + \text{rank} \left(\begin{bmatrix} \mathbf{H}_{12} \\ \mathbf{H}_{22} \end{bmatrix} \right) + \text{rank} \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{0} \end{bmatrix} \\ - \text{rank}(\mathbf{H}_{21}) - \text{rank}(\mathbf{H}_{12}). \quad (5.15)$$

Furthermore, with RLNC employed at each intermediate node over $\mathbb{F}_q$, we know from Lemma 2.7 that the following equalities hold with high probability as $q \rightarrow \infty$.

$$\begin{aligned} \text{rank}(\mathbf{H}_{11}) &= k_{1-1}; & \text{rank}(\mathbf{H}_{12}) &= k_{2-1}; \\ \text{rank}(\mathbf{H}_{21}) &= k_{1-2}; & \text{rank}(\mathbf{H}_{22}) &= k_{2-2}; \\ \text{rank} \left(\begin{bmatrix} \mathbf{H}_{11} \\ \mathbf{H}_{21} \end{bmatrix} \right) &= k_{1-12}; & \text{rank} \left(\begin{bmatrix} \mathbf{H}_{12} \\ \mathbf{H}_{22} \end{bmatrix} \right) &= k_{2-12}; \\ \text{rank} \left(\begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \end{bmatrix} \right) &= k_{12-1}; & \text{rank} \left(\begin{bmatrix} \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix} \right) &= k_{12-2}. \end{aligned} \quad (5.16)$$

The remaining matrix ranks on the right hand sides of (5.9)-(5.15) cannot be explicitly represented in terms of the min-cuts of the network. However, they can be reformulated into equivalent forms involving the min-cuts by introducing a decomposition of the interfering transition matrices $\mathbf{H}_{12}$ and $\mathbf{H}_{21}$. Taking the transition matrix $\mathbf{H}_{12}$ with size $k_{12-1} \times k_{2-12}$ and rank k_{2-1} as an example, we let

³As the analysis given in the following is based on the finite field of order q , the data rate in the following context is given in terms of symbols, which is equivalent to the conventional rate expressed in bits normalized by $\log(q)$.

- $\mathbf{V}_{10} \in \mathbb{F}_q^{k_{2-12} \times (k_{2-12} - k_{2-1})}$ be a matrix whose columns span the null space of $\mathbf{H}_{12}$, denoted as $\mathcal{N}(\mathbf{H}_{12})$,
- $\mathbf{V}_{11} \in \mathbb{F}_q^{k_{2-12} \times k_{2-1}}$ be a matrix such that $\mathcal{R}(\mathbf{V}_{11})$ and $\mathcal{R}(\mathbf{V}_{10})$ are complementary subspaces ([100], p.90).
- $\mathbf{U}_{10} \in \mathbb{F}_q^{k_{12-1} \times (k_{12-1} - k_{2-1})}$ be a matrix whose columns span the left null space of $\mathbf{H}_{12}$, denoted as $\mathcal{N}(\mathbf{H}_{12}^T)$, and
- $\mathbf{U}_{11} \in \mathbb{F}_q^{k_{12-1} \times k_{2-1}}$ be a matrix such that $\mathcal{R}(\mathbf{U}_{11})$ and $\mathcal{R}(\mathbf{U}_{10})$ are complementary subspaces,

Let $\mathbf{U}_1 = \begin{bmatrix} \mathbf{U}_{11} & \mathbf{U}_{10} \end{bmatrix}$ and $\mathbf{V}_1 = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{10} \end{bmatrix}$. Following from the definition of the complementary subspaces ([100], p.90), we have $\dim(\mathcal{R}(\mathbf{U}_{10}) + \mathcal{R}(\mathbf{U}_{11})) = k_{12-1}$ and $\dim(\mathcal{R}(\mathbf{V}_{10}) + \mathcal{R}(\mathbf{V}_{11})) = k_{2-12}$. $\mathbf{U}_1$ and $\mathbf{V}_1$ are full rank square matrices and hence are invertible. Let $\mathbf{\Lambda}_1 = \mathbf{U}_1^T \mathbf{H}_{12} \mathbf{V}_1$, we then have

$$\begin{aligned}
 \mathbf{\Lambda}_1 &= \begin{bmatrix} \mathbf{U}_{11}^T \\ \mathbf{U}_{10}^T \end{bmatrix} \mathbf{H}_{12} \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{10} \end{bmatrix} \\
 &= \begin{bmatrix} \mathbf{U}_{11}^T \mathbf{H}_{12} \mathbf{V}_{11} & \mathbf{U}_{11}^T \mathbf{H}_{12} \mathbf{V}_{10} \\ \mathbf{U}_{10}^T \mathbf{H}_{12} \mathbf{V}_{11} & \mathbf{U}_{10}^T \mathbf{H}_{12} \mathbf{V}_{10} \end{bmatrix} \\
 &= \begin{bmatrix} \mathbf{U}_{11}^T \mathbf{H}_{12} \mathbf{V}_{11} & \mathbf{0}_{k_{2-1} \times (k_{2-12} - k_{2-1})} \\ \mathbf{0}_{(k_{12-1} - k_{2-1}) \times k_{2-1}} & \mathbf{0}_{(k_{12-1} - k_{2-1}) \times (k_{2-12} - k_{2-1})} \end{bmatrix}.
 \end{aligned}$$

Therefore, $\mathbf{H}_{12}$ can be decomposed as

$$\mathbf{H}_{12} = (\mathbf{U}_1^T)^{-1} \mathbf{\Lambda}_1 \mathbf{V}_1^{-1}. \quad (5.17)$$

Lemma 5.2. [100] Let $\mathbf{A} \in \mathbb{F}_q^{p \times l}$ and $\mathbf{B} \in \mathbb{F}_q^{l \times k}$, then we have $\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{A}) - \dim(\mathcal{R}(\mathbf{A}^T) \cap \mathcal{N}(\mathbf{B}^T)) = \text{rank}(\mathbf{B}) - \dim(\mathcal{N}(\mathbf{A}) \cap \mathcal{R}(\mathbf{B}))$.

Proof. Please refer to Fact 2.10.14 on page 116 in [100]. $\square$

Lemma 5.3. The $k_{2-1} \times k_{2-1}$ square matrix $\mathbf{D}_{12} \triangleq \mathbf{U}_{11}^T \mathbf{H}_{12} \mathbf{V}_{11}$ is nonsingular.

Proof. To prove Lemma 5.3, we need to show that $\text{rank}(\mathbf{D}_{12}) = k_{2-1}$. First, accord-

ing to Lemma 5.2, $\text{rank}(\mathbf{U}_{11}^T \mathbf{H}_{12})$ can be computed as

$$\begin{aligned} \text{rank}(\mathbf{U}_{11}^T \mathbf{H}_{12}) &= \text{rank}(\mathbf{U}_{11}^T) - \dim(\mathcal{R}(\mathbf{U}_{11}) \cap \mathcal{N}(\mathbf{H}_{12}^T)) \\ &= \text{rank}(\mathbf{U}_{11}) - \dim(\mathcal{R}(\mathbf{U}_{11}) \cap \mathcal{R}(\mathbf{U}_{10})) \\ &\stackrel{(a)}{=} \text{rank}(\mathbf{U}_{11}) = k_{2-1}. \end{aligned}$$

where (a) follows since $\mathbf{U}_{10}$ and $\mathbf{U}_{11}$ are complementary subspaces. Therefore, we have $\dim(\mathcal{N}(\mathbf{U}_{11}^T \mathbf{H}_{12})) = k_{2-12} - \text{rank}(\mathbf{U}_{11}^T \mathbf{H}_{12}) = k_{2-12} - k_{2-1}$. According to Lemma 2.4.1 on page 94 of [100], we have $\mathcal{N}(\mathbf{H}_{12}) \subseteq \mathcal{N}(\mathbf{U}_{11}^T \mathbf{H}_{12})$. Since $\dim(\mathcal{N}(\mathbf{H}_{12})) = \dim(\mathcal{N}(\mathbf{U}_{11}^T \mathbf{H}_{12})) = k_{2-12} - k_{2-1}$, we must have $\mathcal{N}(\mathbf{H}_{12}) = \mathcal{N}(\mathbf{U}_{11}^T \mathbf{H}_{12})$. Hence, by applying Lemma 5.2, we have

$$\begin{aligned} \text{rank}(\mathbf{D}_{12}) &= \text{rank}(\mathbf{U}_{11}^T \mathbf{H}_{12} \mathbf{V}_{11}) \\ &= \text{rank}(\mathbf{V}_{11}) - \dim(\mathcal{N}(\mathbf{U}_{11}^T \mathbf{H}_{12}) \cap \mathcal{R}(\mathbf{V}_{11})) \\ &= k_{2-1} - \dim(\mathcal{N}(\mathbf{H}_{12}) \cap \mathcal{R}(\mathbf{V}_{11})) \\ &= k_{2-1} - \dim(\mathcal{R}(\mathbf{V}_{10}) \cap \mathcal{R}(\mathbf{V}_{11})) \\ &= k_{2-1}, \end{aligned}$$

where the last equality follows since $\mathcal{R}(\mathbf{V}_{10})$ and $\mathcal{R}(\mathbf{V}_{11})$ are complementary subspaces. This thus completes the proof of Lemma 5.3. $\square$

Similarly, $\mathbf{H}_{21}$ can be decomposed as

$$\begin{aligned} \mathbf{H}_{21} &= (\mathbf{U}_2^T)^{-1} \mathbf{\Lambda}_2 \mathbf{V}_2^{-1} \\ &= \begin{bmatrix} \mathbf{U}_{21}^T \\ \mathbf{U}_{20}^T \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{D}_{21} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{21} & \mathbf{V}_{20} \end{bmatrix}^{-1}. \end{aligned} \tag{5.18}$$

Lemma 5.4. *With the above decompositions of $\mathbf{H}_{12}$ and $\mathbf{H}_{21}$, we have*

$$\begin{aligned} \text{rank} \left(\begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{0} \end{bmatrix} \right) &= \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) + k_{1-2} + k_{2-1} \\ \text{rank} \left(\begin{bmatrix} \mathbf{H}_{21} & \mathbf{H}_{22} \\ \mathbf{0} & \mathbf{H}_{12} \end{bmatrix} \right) &= \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) + k_{1-2} + k_{2-1}. \end{aligned}$$

Proof. Since the matrices $\mathbf{U}_1, \mathbf{U}_2, \mathbf{V}_1$, and $\mathbf{V}_2$ are non-singular, the block-diagonal matrices $\begin{bmatrix} \mathbf{U}_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_2^T \end{bmatrix}$ and $\begin{bmatrix} \mathbf{V}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_1 \end{bmatrix}$ are non-singular as well. With the fact that

the matrix rank is unchanged after multiplying a non-singular matrix, we have

$$\begin{aligned}
\text{rank} \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{0} \end{bmatrix} &= \text{rank} \left(\begin{bmatrix} \mathbf{U}_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_2^T \end{bmatrix} \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_1 \end{bmatrix} \right) \\
&= \text{rank} \begin{bmatrix} \mathbf{U}_1^T \mathbf{H}_{11} \mathbf{V}_2 & \mathbf{\Lambda}_1 \\ \mathbf{\Lambda}_2 & \mathbf{0} \end{bmatrix} \\
&= \text{rank} \begin{bmatrix} \mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{21} & \mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{20} & \mathbf{D}_{12} & \mathbf{0} \\ \mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21} & \mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20} & \mathbf{0} & \mathbf{0} \\ \mathbf{D}_{21} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \\
&\stackrel{(a)}{=} \text{rank} \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{D}_{12} \\ \mathbf{0} & \mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20} & \mathbf{0} \\ \mathbf{D}_{21} & \mathbf{0} & \mathbf{0} \end{bmatrix} \\
&= \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) + k_{1-2} + k_{2-1},
\end{aligned}$$

where (a) can be obtained by elementary row and column operations with the fact that $\mathbf{D}_{12}$ and $\mathbf{D}_{21}$ are non-singular square matrices. With similar techniques, we can show that $\text{rank} \left(\begin{bmatrix} \mathbf{H}_{21} & \mathbf{H}_{22} \\ \mathbf{0} & \mathbf{H}_{12} \end{bmatrix} \right) = \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) + k_{1-2} + k_{2-1}$. $\square$

With (5.16) and Lemma 5.4, (5.9)-(5.15) can be equivalently expressed as

$$R_1 \leq k_{1-1} \tag{5.19}$$

$$R_2 \leq k_{2-2} \tag{5.20}$$

$$R_1 + R_2 \leq k_{12-1} + k_{2-12} - k_{2-1} \tag{5.21}$$

$$R_1 + R_2 \leq k_{12-2} + k_{1-12} - k_{1-2} \tag{5.22}$$

$$R_1 + R_2 \leq \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) + k_{1-2} + k_{2-1} \tag{5.23}$$

$$2R_1 + R_2 \leq k_{12-1} + k_{1-12} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) \tag{5.24}$$

$$R_1 + 2R_2 \leq k_{12-2} + k_{2-12} + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) \tag{5.25}$$

Given a rate pair (R_1, R_2) that satisfies the inequalities in (5.19)-(5.25), we will next show that it can be achieved by applying the linear scheme depicted in Figure 5.1. First, with $\mathbf{H}_{12}$ and $\mathbf{H}_{21}$ decomposed as in (5.17) and (5.18), respectively, and by absorbing $\mathbf{V}_j^{-1}$ into the input vector $\mathbf{x}_i$ and multiplying the output vector $\mathbf{y}_i$ with $\mathbf{U}_i^T$, $i, j \in \{1, 2\}, i \neq j$, the channel model given in (6.22) can be equivalently

written as

$$\begin{aligned} \mathbf{y}'_1 &= \mathbf{U}_1^T \mathbf{y}_1 = \mathbf{H}'_{11} \mathbf{x}'_1 + \mathbf{\Lambda}_1 \mathbf{x}'_2 \\ \mathbf{y}'_2 &= \mathbf{U}_2^T \mathbf{y}_2 = \mathbf{H}'_{22} \mathbf{x}'_2 + \mathbf{\Lambda}_2 \mathbf{x}'_1, \end{aligned} \quad (5.26)$$

where $\mathbf{x}'_1 = \mathbf{V}_2^{-1} \mathbf{x}_1$, $\mathbf{x}'_2 = \mathbf{V}_1^{-1} \mathbf{x}_2$, $\mathbf{H}'_{11} = \mathbf{U}_1^T \mathbf{H}_{11} \mathbf{V}_2$ and $\mathbf{H}'_{22} = \mathbf{U}_2^T \mathbf{H}_{22} \mathbf{V}_1$. The advantage of the equivalent channel model (5.26) is that it results in block-diagonal interfering channel matrices $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$, which are easier to deal with. Similar transformations have been used in [101] for deriving the optimal degrees of freedom (DoF) of the two-user wireless MIMO-IC. To determine the input signal vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, it is then sufficient to obtain $\mathbf{x}'_1$ and $\mathbf{x}'_2$ since they are related by the invertible linear transformations $\mathbf{x}_1 = \mathbf{V}_2 \mathbf{x}'_1$ and $\mathbf{x}_2 = \mathbf{V}_1 \mathbf{x}'_2$.

Motivated by the rate-splitting technique used in the celebrated Han-Kobayashi schemes in ICs [99], we decompose the R_1 symbols in $\mathbf{d}_1$ into two parts: the common information $\mathbf{d}_{1c} \in \mathbb{F}_q^{R_{1c}}$, which is decodable at both t_1 and t_2 , and the private information $\mathbf{d}_{1p} \in \mathbb{F}_q^{R_{1p}}$, which is decodable at t_1 only. We then have $\mathbf{d}_1 = \begin{bmatrix} \mathbf{d}_{1c} \\ \mathbf{d}_{1p} \end{bmatrix}$ and $R_1 = R_{1c} + R_{1p}$. To map $\mathbf{d}_1$ to the k_{1-12} -dimensional input vector $\mathbf{x}'_1$, a block-diagonal linear precoding is applied as

$$\mathbf{x}'_1 = \begin{bmatrix} \mathbf{E}_{1c} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{1p} \end{bmatrix} \mathbf{d}_1 = \begin{bmatrix} \mathbf{E}_{1c} \mathbf{d}_{1c} \\ \mathbf{E}_{1p} \mathbf{d}_{1p} \end{bmatrix}, \quad (5.27)$$

where $\mathbf{E}_{1c} \in \mathbb{F}_q^{k_{1-2} \times R_{1c}}$ and $\mathbf{E}_{1p} \in \mathbb{F}_q^{(k_{1-12}-k_{1-2}) \times R_{1p}}$ are randomly and independently generated. Note that with the above block-diagonal precoding, the common and private symbols $\mathbf{d}_{1c}$ and $\mathbf{d}_{1p}$ are constrained to the first k_{1-2} and the last $(k_{1-12} - k_{1-2})$ components of $\mathbf{x}'_1$, respectively. As such, the private symbols $\mathbf{d}_{1p}$ will not interfere the received signal vector $\mathbf{y}'_2$ at t_2 , and at the same time, the common symbols $\mathbf{d}_{1c}$ can be decoded at both t_1 and t_2 if the rate pair (R_1, R_2) satisfies the inequalities given in (5.19)-(5.25), as will become clear later.

Similarly, $\mathbf{x}'_2$ can be obtained as $\mathbf{x}'_2 = \begin{bmatrix} \mathbf{E}_{2c} \mathbf{d}_{2c} \\ \mathbf{E}_{2p} \mathbf{d}_{2p} \end{bmatrix}$, where $\mathbf{d}_{2c} \in \mathbb{F}_q^{R_{2c}}$, $\mathbf{d}_{2p} \in \mathbb{F}_q^{R_{2p}}$, $\mathbf{E}_{2c} \in \mathbb{F}_q^{k_{2-1} \times R_{2c}}$, and $\mathbf{E}_{2p} \in \mathbb{F}_q^{(k_{2-12}-k_{2-1}) \times R_{2p}}$. $\mathbf{E}_{2c}$ and $\mathbf{E}_{2p}$ are randomly and independently generated.

The output $\mathbf{y}'_1$ at t_1 given in (5.26) can then be written as

$$\begin{aligned}
 \mathbf{y}'_1 &= \mathbf{U}_1^T \mathbf{H}_{11} \mathbf{V}_2 \mathbf{x}'_1 + \mathbf{A}_1 \mathbf{x}'_2 \\
 &= \begin{bmatrix} \mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{21} & \mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{20} \\ \mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21} & \mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20} \end{bmatrix} \begin{bmatrix} \mathbf{E}_{1c} \mathbf{d}_{1c} \\ \mathbf{E}_{1p} \mathbf{d}_{1p} \end{bmatrix} + \begin{bmatrix} \mathbf{D}_{12} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{E}_{2c} \mathbf{d}_{2c} \\ \mathbf{E}_{2p} \mathbf{d}_{2p} \end{bmatrix} \\
 &= \begin{bmatrix} \mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{21} & \mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{20} & \mathbf{D}_{12} \\ \mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21} & \mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{E}_{1c} \mathbf{d}_{1c} \\ \mathbf{E}_{1p} \mathbf{d}_{1p} \\ \mathbf{E}_{2c} \mathbf{d}_{2c} \end{bmatrix}. \tag{5.28}
 \end{aligned}$$

Equation (5.28) clearly shows that the private symbols in vector $\mathbf{d}_{2p}$ transmitted by s_2 does not interfere with $\mathbf{y}'_1$. Furthermore, the desired symbols $\mathbf{d}_{1c}$ and $\mathbf{d}_{1p}$ can be decoded at t_1 if the linear system equations given in (5.28) are uniquely solvable.

Due to symmetry, at t_2 we have

$$\mathbf{y}'_2 = \begin{bmatrix} \mathbf{U}_{21}^T \mathbf{H}_{22} \mathbf{V}_{11} & \mathbf{U}_{21}^T \mathbf{H}_{22} \mathbf{V}_{10} & \mathbf{D}_{21} \\ \mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{11} & \mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{E}_{2c} \mathbf{d}_{2c} \\ \mathbf{E}_{2p} \mathbf{d}_{2p} \\ \mathbf{E}_{1c} \mathbf{d}_{1c} \end{bmatrix}. \tag{5.29}$$

To find a sufficient condition on R_{1p} , R_{1c} , R_{2p} , and R_{2c} such that the system of linear equations given by (5.28) and (5.29) are uniquely solvable, the following results will be used.

Lemma 5.5. ([100], p.100) *Given the linear relationship $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{A} \in \mathbb{F}_q^{p \times l}$, $\mathbf{x} \in \mathbb{F}_q^l$, and $\mathbf{y} \in \mathbb{F}_q^p$, $\mathbf{x}$ can be uniquely determined from $\mathbf{y}$ if and only if $\mathbf{A}$ is of full column rank, i.e., $\text{rank}(\mathbf{A}) = l$.*

For notational convenience, let

$$\begin{aligned}
 \mathbf{M}_1 &\triangleq \begin{bmatrix} \mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{21} & \mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{20} & \mathbf{D}_{12} \\ \mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21} & \mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{E}_{1c} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{1p} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{E}_{2c} \end{bmatrix} \\
 &= \begin{bmatrix} \left(\mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{21} \right) \mathbf{E}_{1c} & \left(\mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{20} \right) \mathbf{E}_{1p} & \left(\mathbf{D}_{12} \right) \mathbf{E}_{2c} \\ \left(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21} \right) \mathbf{E}_{1c} & \left(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20} \right) \mathbf{E}_{1p} & \left(\mathbf{0} \right) \mathbf{E}_{2c} \end{bmatrix}. \tag{5.30}
 \end{aligned}$$

Hence, (5.28) can be written as

$$\mathbf{y}'_1 = \mathbf{M}_1 \begin{bmatrix} \mathbf{d}_{1c} \\ \mathbf{d}_{1p} \\ \mathbf{d}_{2c} \end{bmatrix}.$$

According to Lemma 5.5, the sink node t_1 is able to decode $\mathbf{d}_{1c}$, $\mathbf{d}_{1p}$, and $\mathbf{d}_{2c}$ if $\mathbf{M}_1$ is of full column rank. Next, we will find a sufficient condition on the data rates R_{1c} , R_{1p} , and R_{2c} such that $\mathbf{M}_1$ has full column rank.

Lemma 5.6. *Let $\mathbf{A}_1 \in \mathbb{F}_q^{p \times l_1}$, $\mathbf{A}_2 \in \mathbb{F}_q^{p \times l_2}$ and $\mathbf{A}_3 \in \mathbb{F}_q^{p \times l_3}$ be given matrices, and $\mathbf{E}_1 \in \mathbb{F}_q^{l_1 \times k_1}$, $\mathbf{E}_2 \in \mathbb{F}_q^{l_2 \times k_2}$ and $\mathbf{E}_3 \in \mathbb{F}_q^{l_3 \times k_3}$ be random matrices whose entries are uniformly generated from $\mathbb{F}_q$. Then as $q \rightarrow \infty$, $\text{rank}([\mathbf{A}_1 \mathbf{E}_1 \quad \mathbf{A}_2 \mathbf{E}_2 \quad \mathbf{A}_3 \mathbf{E}_3]) = k_1 + k_2 + k_3$ holds with high probability if the following conditions are satisfied:*

- $k_1 \leq \text{rank}(\mathbf{A}_1)$
- $k_2 \leq \text{rank}(\mathbf{A}_2)$
- $k_3 \leq \text{rank}(\mathbf{A}_3)$
- $k_1 + k_2 \leq \text{rank}([\mathbf{A}_1 \quad \mathbf{A}_2])$
- $k_1 + k_3 \leq \text{rank}([\mathbf{A}_1 \quad \mathbf{A}_3])$
- $k_2 + k_3 \leq \text{rank}([\mathbf{A}_2 \quad \mathbf{A}_3])$
- $k_1 + k_2 + k_3 \leq \text{rank}([\mathbf{A}_1 \quad \mathbf{A}_2 \quad \mathbf{A}_3])$

Proof. Please refer to Appendix B. □

By applying Lemma 5.6 to (5.30), a sufficient condition for $\mathbf{M}_1$ to be of full

column rank, and hence (5.28) is uniquely solvable, is given by

$$R_{1c} \leq \text{rank} \left(\begin{bmatrix} \mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{21} \\ \mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21} \end{bmatrix} \right) = \text{rank}(\mathbf{U}_1^T \mathbf{H}_{11} \mathbf{V}_{21}) \stackrel{(a)}{=} \text{rank}(\mathbf{H}_{11} \mathbf{V}_{21}) \quad (5.31)$$

$$R_{1p} \leq \text{rank} \left(\begin{bmatrix} \mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{20} \\ \mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20} \end{bmatrix} \right) = \text{rank}(\mathbf{U}_1^T \mathbf{H}_{11} \mathbf{V}_{20}) \stackrel{(b)}{=} k_{1-12} - k_{1-2} \quad (5.32)$$

$$R_{2c} \leq \text{rank} \left(\begin{bmatrix} \mathbf{D}_{12} \\ \mathbf{0} \end{bmatrix} \right) = k_{2-1} \quad (5.33)$$

$$R_{1c} + R_{1p} \leq \text{rank} \left(\begin{bmatrix} \mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{21} & \mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{20} \\ \mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21} & \mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20} \end{bmatrix} \right) = \text{rank}(\mathbf{U}_1^T \mathbf{H}_{11} \mathbf{V}_2) = k_{1-1} \quad (5.34)$$

$$R_{1c} + R_{2c} \leq \text{rank} \left(\begin{bmatrix} \mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{21} & \mathbf{D}_{12} \\ \mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21} & \mathbf{0} \end{bmatrix} \right) \stackrel{(c)}{=} k_{2-1} + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21}) \quad (5.35)$$

$$R_{1p} + R_{2c} \leq \text{rank} \left(\begin{bmatrix} \mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{20} & \mathbf{D}_{12} \\ \mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20} & \mathbf{0} \end{bmatrix} \right) \stackrel{(d)}{=} k_{2-1} + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) \quad (5.36)$$

$$R_{1c} + R_{1p} + R_{2c} \leq \text{rank} \left(\begin{bmatrix} \mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{21} & \mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{20} & \mathbf{D}_{12} \\ \mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21} & \mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20} & \mathbf{0} \end{bmatrix} \right) \stackrel{(e)}{=} k_{12-1}, \quad (5.37)$$

where (a) holds since $\mathbf{U}_1$ is invertible, and (b) follows from Lemma 5.2 since

$$\begin{aligned} \text{rank}(\mathbf{U}_1^T \mathbf{H}_{11} \mathbf{V}_{20}) &= \text{rank}(\mathbf{H}_{11} \mathbf{V}_{20}) \\ &= \text{rank}(\mathbf{V}_{20}) - \dim(\mathcal{N}(\mathbf{H}_{11}) \cap \mathcal{R}(\mathbf{V}_{20})) \\ &= \text{rank}(\mathbf{V}_{20}) - \dim(\mathcal{N}(\mathbf{H}_{11}) \cap \mathcal{N}(\mathbf{H}_{21})) \\ &\stackrel{(f)}{=} \text{rank}(\mathbf{V}_{20}) \\ &= k_{1-12} - k_{1-2}, \end{aligned} \quad (5.38)$$

where (f) follows from (5.16), i.e., $\begin{bmatrix} \mathbf{H}_{11} \\ \mathbf{H}_{21} \end{bmatrix}$ is of full column rank. Moreover, (c) and (d) can be obtained by applying elementary column operations since $\mathbf{D}_{12}$ is nonsingular as given in Lemma 5.3; and (e) can be shown with elementary column

operations together with a similar proof as that for (b), i.e.,

$$\begin{aligned}
& \text{rank} \left(\begin{bmatrix} \mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{21} & \mathbf{U}_{11}^T \mathbf{H}_{11} \mathbf{V}_{20} & \mathbf{D}_{12} \\ \mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21} & \mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20} & \mathbf{0} \end{bmatrix} \right) \\
&= \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_2) + \text{rank}(\mathbf{D}_{12}) \\
&= \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11}) + k_{2-1} \\
&= \text{rank}(\mathbf{U}_{10}) - \dim(\mathcal{R}(\mathbf{U}_{10}) \cap \mathcal{N}(\mathbf{H}_{11}^T)) + k_{2-1} \\
&= k_{12-1} - \dim(\mathcal{N}(\mathbf{H}_{12}^T) \cap \mathcal{N}(\mathbf{H}_{11}^T)) \\
&= k_{12-1}
\end{aligned} \tag{5.39}$$

By symmetry, a sufficient condition for sink t_2 to successfully decode $\mathbf{d}_{2c}$, $\mathbf{d}_{2p}$ and $\mathbf{d}_{1c}$ is given by

$$R_{2c} \leq \text{rank}(\mathbf{H}_{22} \mathbf{V}_{11}) \tag{5.40}$$

$$R_{2p} \leq k_{2-12} - k_{2-1} \tag{5.41}$$

$$R_{1c} \leq k_{1-2} \tag{5.42}$$

$$R_{2c} + R_{2p} \leq k_{2-2} \tag{5.43}$$

$$R_{2c} + R_{1c} \leq k_{1-2} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{11}) \tag{5.44}$$

$$R_{2p} + R_{1c} \leq k_{1-2} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) \tag{5.45}$$

$$R_{2c} + R_{2p} + R_{1c} \leq k_{12-2} \tag{5.46}$$

By substituting with $R_1 = R_{1c} + R_{1p}$ and $R_2 = R_{2c} + R_{2p}$, the sufficient conditions on the data rate R_1 and R_2 to ensure full decodability at the respective destinations can be obtained by applying the standard Fourier-Motzkin elimination over (5.31)-(5.37) and (5.40)-(5.46). The detailed steps involving the Fourier-Motzkin elimination are presented Appendix C and the resulting achievable rate region is given by (5.19)-(5.25). Furthermore, given any valid rate pair (R_1, R_2) satisfying (5.19)-(5.25), there exists at least one valid common-private rate splitting $(R_{1c}, R_{1p}, R_{2c}, R_{2p})$ that can be found by solving $R_{1p} = R_1 - R_{1c}$ and $R_{2p} = R_2 - R_{2c}$ subject to the constraints specified in (5.31)-(5.37) and (5.40)-(5.46).

Remark 5.7. Since (5.9)-(5.15) is the capacity region for the linear deterministic IC given by (6.22), it gives the maximum achievable rate region for double-unicast networks with any given transition matrices $\{\mathbf{H}_{ji}\}$. In other words, for any given coding coefficients chosen by the intermediate nodes, there are no other source and

sink strategies, linear or non-linear, that can outperform the proposed linear scheme described above.

Remark 5.8. Different from the classical information theoretic coding scheme in [58] where the code length is assumed to be sufficiently large, the achievability scheme in Figure 5.1 operates symbol-wise. When the target achievable rates R_1 and R_2 are integers, block length 1 is sufficient; otherwise, time extension is needed, i.e., integer number of symbols are transmitted using the network over n time slots, corresponding to a block length of n . The resulting time-extended network can be viewed as a larger network, to which the scheme presented above can be readily applied. Notice that time extension is a commonly used technique in the literature to deal with fractional degrees of freedom (DoF)/rates [102]. Specifically, for a transmission over n time slots, the transition matrices from the source to the sink nodes are of the form $\mathbf{H}_{ij}^{(n)} = \mathbf{I}_n \otimes \mathbf{H}_{ij}$, $i, j \in \{1, 2\}$. Then, we have

$$\begin{aligned}
\text{rank}(\mathbf{H}_{ij}^{(n)}) &= n[\text{rank}(\mathbf{H}_{ij})], i, j \in \{1, 2\} \\
\text{rank} \left(\begin{bmatrix} \mathbf{H}_{1i}^{(n)} \\ \mathbf{H}_{2i}^{(n)} \end{bmatrix} \right) &= \text{rank} \left(\mathbf{I}_n \otimes \begin{bmatrix} \mathbf{H}_{1i} \\ \mathbf{H}_{2i} \end{bmatrix} \right) = n \left[\text{rank} \left(\begin{bmatrix} \mathbf{H}_{1i} \\ \mathbf{H}_{2i} \end{bmatrix} \right) \right], i \in \{1, 2\} \\
\text{rank} \left(\begin{bmatrix} \mathbf{H}_{i1}^{(n)} & \mathbf{H}_{i2}^{(n)} \end{bmatrix} \right) &= \text{rank} \left(\mathbf{I}_n \otimes \begin{bmatrix} \mathbf{H}_{i1} & \mathbf{H}_{i2} \end{bmatrix} \right) = n \left[\text{rank} \left(\begin{bmatrix} \mathbf{H}_{i1} & \mathbf{H}_{i2} \end{bmatrix} \right) \right], i \in \{1, 2\} \\
\text{rank} \left(\begin{bmatrix} \mathbf{H}_{11}^{(n)} & \mathbf{H}_{12}^{(n)} \\ \mathbf{H}_{21}^{(n)} & \mathbf{0} \end{bmatrix} \right) &= n \left[\text{rank} \left(\begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{0} \end{bmatrix} \right) \right] \\
\text{rank} \left(\begin{bmatrix} \mathbf{H}_{21}^{(n)} & \mathbf{H}_{22}^{(n)} \\ \mathbf{0} & \mathbf{H}_{12}^{(n)} \end{bmatrix} \right) &= n \left[\text{rank} \left(\begin{bmatrix} \mathbf{H}_{21} & \mathbf{H}_{22} \\ \mathbf{0} & \mathbf{H}_{12} \end{bmatrix} \right) \right]
\end{aligned} \tag{5.47}$$

By substituting (5.47) into (5.9)-(5.15), it can be obtained that if rate pair (R_1, R_2) is achievable in the original network, then (nR_1, nR_2) is achievable in the time-extended network with proposed precoding/decoing scheme.

Remark 5.9. While the proposed linear achievability scheme has been presented mainly based on the rate region specified in (5.19)-(5.25) under the assumption that RLNC over sufficiently large field is applied at the intermediate nodes, the same techniques can be applied to show the achievability of the rate region specified in the original form in (5.9)-(5.15) with any given set of transition matrices resulting from any (not necessarily random) linear network coding schemes. Specifically, for general transition matrices $\mathbf{H}_{12} \in \mathbb{F}_q^{n_1 \times m_2}$ and $\mathbf{H}_{21} \in \mathbb{F}_q^{n_2 \times m_1}$ with ranks r_{12} and r_{21} , respectively, the corresponding precoding matrices in Figure 5.1 can be similarly obtained as

- $\mathbf{V}_2 = \begin{bmatrix} \mathbf{V}_{21} & \mathbf{V}_{20} \end{bmatrix} \in \mathbb{F}_q^{m_1 \times m_1}$, where the columns of $\mathbf{V}_{20} \in \mathbb{F}_q^{m_1 \times (m_1 - r_{21})}$ span $\mathcal{N}(\mathbf{H}_{21})$ and $\mathcal{R}(\mathbf{V}_{21})$ and $\mathcal{R}(\mathbf{V}_{20})$ are complementary subspaces.
- $\mathbf{U}_1 = \begin{bmatrix} \mathbf{U}_{11} & \mathbf{U}_{10} \end{bmatrix} \in \mathbb{F}_q^{n_1 \times n_1}$, where the columns of $\mathbf{U}_{10} \in \mathbb{F}_q^{n_1 \times (n_1 - r_{12})}$ span $\mathcal{N}(\mathbf{H}_{12}^T)$ and $\mathcal{R}(\mathbf{U}_{11})$ and $\mathcal{R}(\mathbf{U}_{10})$ are complementary subspaces.
- $\mathbf{E}_{1c} \in \mathbb{F}_q^{r_{21} \times R_{1c}}$ and $\mathbf{E}_{1p} \in \mathbb{F}_q^{(m_1 - r_{21}) \times R_{1p}}$ are generated randomly.

The proposed linear achievability scheme presented above is illustrated with the following simple example.

Example 5.1. Consider the network shown in Figure 5.2(a) with unit edge capacity. The min-cut values of this network are respectively given by $k_{1-1} = 2$, $k_{2-2} = 3$, $k_{2-1} = k_{1-2} = 2$, $k_{1-12} = k_{12-1} = 2$ and $k_{2-12} = k_{12-2} = 3$. Assume that random network coding coefficients have been chosen in the field of size⁴ $q = 7$ with the coding coefficients at each intermediate node indicated in the figure, and the resulting transition matrices are given by:

$$\mathbf{H}_{11} = \begin{bmatrix} 2 & 0 \\ 2 & 3 \end{bmatrix}, \mathbf{H}_{12} = \begin{bmatrix} 2 & 1 & 0 \\ 2 & 1 & 1 \end{bmatrix}, \mathbf{H}_{21} = \begin{bmatrix} 1 & 0 \\ 2 & 3 \\ 2 & 3 \end{bmatrix}, \mathbf{H}_{22} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 2 & 1 & 1 \end{bmatrix}. \quad (5.48)$$

FIGURE 5.2: (a) An example of double-unicast networks. (b) The achievable rate regions.

⁴One commonly used field size for RLNC is 2^8 . Here, a small field size is used for illustration purposes.

Then the achievable region (after removing redundant inequalities) specified in (5.9)-(5.15) for this particular example is given by

$$\begin{aligned} R_1 &\leq 2 \\ R_1 + R_2 &\leq 3 \\ 2R_1 + R_2 &\leq 4, \end{aligned}$$

which is plotted in Figure 5.2(b). As a comparison, the achievable rate region with routing is also plotted, which is shown to be strictly smaller than that achieved with network coding.

By applying the above proposed scheme, we next give the specific designs that achieve the outmost rate pair $(R_1, R_2) = (1, 2)$. Let $\mathbf{d}_1 = d_{11}$ and $\mathbf{d}_2 = [d_{21} \ d_{22}]^T$ be the information-bearing symbols. Then, following (5.31)-(5.37) and (5.40)-(5.46), the private-common rate splitting is given by $\mathbf{d}_{1c} = d_{11}$, $\mathbf{d}_{1p} = \emptyset$, $\mathbf{d}_{2c} = d_{21}$ and $\mathbf{d}_{2p} = d_{22}$, where $\emptyset$ denotes an empty vector/matrix.

Based on $\mathbf{H}_{12}$ and $\mathbf{H}_{21}$ given in (5.48), the following component matrices can be obtained:

$$\begin{aligned} \mathbf{U}_{11} &= \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \mathbf{U}_{10} = \emptyset, \mathbf{V}_{11} = \begin{bmatrix} 2 & 2 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}, \mathbf{V}_{10} = \begin{bmatrix} 1 \\ 5 \\ 0 \end{bmatrix}, \\ \mathbf{U}_{21} &= \begin{bmatrix} 1 & 0 \\ 2 & 3 \\ 2 & 3 \end{bmatrix}, \mathbf{U}_{20} = \begin{bmatrix} 0 \\ 3 \\ 4 \end{bmatrix}, \mathbf{V}_{21} = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}, \mathbf{V}_{20} = \emptyset. \end{aligned}$$

Assume that the following matrices are randomly generated, which will be applied as the linear precoding matrices to map the data symbols to the transmit vectors:

$$\mathbf{E}_{1c} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}, \mathbf{E}_{1p} = \emptyset, \mathbf{E}_{2c} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \mathbf{E}_{2p} = \begin{bmatrix} 3 \end{bmatrix}.$$

The transmit vector by s_1 can then be obtained as

$$\mathbf{x}_1 = \mathbf{V}_2 \mathbf{x}'_1 = \mathbf{V}_{21} \mathbf{E}_{1c} \mathbf{d}_{1c} = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} d_{11} = \begin{bmatrix} 3d_{11} \\ 2d_{11} \end{bmatrix}$$

Similarly, we have

$$\mathbf{x}_2 = \mathbf{V}_1 \mathbf{x}'_2 = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{10} \end{bmatrix} \begin{bmatrix} \mathbf{E}_{2c} \mathbf{d}_{2c} \\ \mathbf{E}_{2p} \mathbf{d}_{2p} \end{bmatrix} = \begin{bmatrix} 2 & 2 & 1 \\ 1 & 1 & 5 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2d_{21} \\ 3d_{21} \\ 3d_{22} \end{bmatrix} = \begin{bmatrix} 3d_{21} + 3d_{22} \\ 5d_{21} + d_{22} \\ 3d_{21} \end{bmatrix}$$

Therefore, the received symbol vectors at t_1 and t_2 are respectively given by

$$\mathbf{y}_1 = \mathbf{H}_{11} \begin{bmatrix} 3d_{11} \\ 2d_{11} \end{bmatrix} + \mathbf{H}_{12} \begin{bmatrix} 3d_{21} + 3d_{22} \\ 5d_{21} + d_{22} \\ 3d_{21} \end{bmatrix} = \begin{bmatrix} 6d_{11} + 4d_{21} \\ 5d_{11} \end{bmatrix} \quad (5.49)$$

$$\mathbf{y}_2 = \mathbf{H}_{22} \begin{bmatrix} 3d_{21} + 3d_{22} \\ 5d_{21} + d_{22} \\ 3d_{21} \end{bmatrix} + \mathbf{H}_{21} \begin{bmatrix} 3d_{11} \\ 2d_{11} \end{bmatrix} = \begin{bmatrix} 3d_{21} + 3d_{22} + 3d_{11} \\ 4d_{21} + 5d_{11} \\ 5d_{11} \end{bmatrix}. \quad (5.50)$$

It can be verified that both t_1 and t_2 can recover their desired symbols by solving the system of linear equations given by (5.49) and (5.50), respectively.

5.1.3 An Achievable Rate Region in Terms of Min-cuts

The achievable rate region for double-unicast networks specified in (5.19)-(5.25) is given in terms of the min-cuts of the network, as well as the ranks of the matrices $\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}$ and $\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}$. In this section, we derive another rate region for (6.22) only in terms of the min-cuts of the network, which is sub-optimal but can be more easily computed.

Lemma 5.10. *If the coding coefficients at the intermediate nodes are randomly selected from a sufficiently large field, then we have the following inequalities:*

$$\text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) \geq (k_{12-1} + k_{1-12} - k_{1-1} - k_{1-2} - k_{2-1})^+, \quad (5.51)$$

$$\text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) \geq (k_{12-2} + k_{2-12} - k_{2-2} - k_{2-1} - k_{1-2})^+. \quad (5.52)$$

Proof. With Frobenius inequality, we have

$$\begin{aligned} \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) &\geq (\text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11}) + \text{rank}(\mathbf{H}_{11} \mathbf{V}_{20}) - \text{rank}(\mathbf{H}_{11}))^+ \\ &\stackrel{(a)}{=} (k_{12-1} - k_{2-1} + k_{1-12} - k_{1-2} - k_{1-1})^+, \end{aligned}$$

where (a) follows from Lemma 5.2.

Similarly, the inequality $\text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) \geq (k_{12-2} + k_{2-12} - k_{2-2} - k_{2-1} - k_{1-2})^+$ follows. $\square$

By substituting (5.51) and (5.52) into (5.19)-(5.25), the following achievable rate region only in terms of min-cuts of the network can be obtained:

$$R_1 \leq k_{1-1} \quad (5.53)$$

$$R_2 \leq k_{2-2} \quad (5.54)$$

$$R_1 + R_2 \leq k_{12-1} + k_{2-12} - k_{2-1} \quad (5.55)$$

$$R_1 + R_2 \leq k_{12-2} + k_{1-12} - k_{1-2} \quad (5.56)$$

$$R_1 + R_2 \leq (k_{12-2} + k_{2-12} - k_{2-2} - k_{1-2} - k_{2-1})^+ \quad (5.57)$$

$$+ (k_{12-1} + k_{1-12} - k_{1-1} - k_{1-2} - k_{2-1})^+ + k_{1-2} + k_{2-1} \quad (5.58)$$

$$2R_1 + R_2 \leq k_{12-1} + k_{1-12} + (k_{12-2} + k_{2-12} - k_{2-2} - k_{1-2} - k_{2-1})^+ \quad (5.59)$$

$$R_1 + 2R_2 \leq k_{12-2} + k_{2-12} + (k_{12-1} + k_{1-12} - k_{1-1} - k_{1-2} - k_{2-1})^+. \quad (5.60)$$

Example 5.2. While the achievable region (5.53)-(5.60) can be easily computed based on the min-cuts of the network, it is usually smaller than the original region specified in (5.9)-(5.15). This is illustrated by the rail network shown in Figure 5.3[92].

FIGURE 5.3: (a) The rail network. (b) The achievable regions.

5.1.3.1 Comparison with Existing Rate Regions

As discussed in Remark 5.7, the region specified in (5.9)-(5.15) gives the largest possible achievable rate region of double-unicast networks for any given set of network coding coefficients, which cannot be outperformed by any other schemes. In this subsection, we will show that even the rate region specified in (5.53)-(5.60) is strictly larger than those achievable regions obtained in the literature [50, 67, 97, 98].

With “rate-exchange” method, the authors in [50] have obtained an achievable region for double-unicast networks that is given by the convex hull of the following two regions:

$\mathcal{R}_1$	$\mathcal{R}_2$
$R_1 \leq k_{1-1}$	$R_1 + 2R_2 \leq k_{1-1}$
$2R_1 + R_2 \leq k_{2-2}$	$R_2 \leq k_{2-2}$

Proposition 5.11. *The achievable rate region specified in (5.53)-(5.60) for double-unicast networks is strictly larger than the convex hull of $\mathcal{R}_1$ and $\mathcal{R}_2$.*

Proof. Since the region specified in (5.53)-(5.60) is convex, it is sufficient to show that it includes both $\mathcal{R}_1$ and $\mathcal{R}_2$. First, we show that any rate pair (R'_1, R'_2) in $\mathcal{R}_1$ satisfies the inequalities (5.53)-(5.60). With $R'_1 \leq k_{1-1}$ and $2R'_1 + R'_2 \leq k_{2-2}$, we

have:

$$\begin{aligned}
R'_1 &\leq k_{1-1} \\
R'_2 &\leq k_{2-2} - 2R'_1 \\
&\leq k_{2-2} \\
R'_1 + R'_2 &\leq k_{2-2} - R'_1 \\
&\leq k_{2-2} \leq k_{2-12} \\
&\leq k_{2-12} + k_{12-1} - k_{2-1} \\
R'_1 + R'_2 &\leq k_{2-2} - R'_1 \\
&\leq k_{2-2} \leq k_{12-2} \\
&\leq k_{12-2} + k_{1-12} - k_{1-2} \\
R'_1 + R'_2 &\leq k_{2-12} \\
&\leq k_{2-12} + k_{12-2} - k_{2-2} \\
&\leq (k_{12-2} + k_{2-12} - k_{2-2} - k_{1-2} - k_{2-1})^+ + k_{1-2} + k_{2-1} \\
&\leq (k_{12-2} + k_{2-12} - k_{2-2} - k_{1-2} - k_{2-1})^+ + k_{1-2} + k_{2-1} \\
&\quad + (k_{12-1} + k_{1-12} - k_{1-1} - k_{1-2} - k_{2-1})^+ \\
2R'_1 + R'_2 &\leq k_{2-2} \leq k_{12-2} \\
&\leq k_{12-2} + (k_{2-12} - k_{2-2}) + (k_{12-1} - k_{2-1}) + (k_{1-12} - k_{1-2}) \\
&\leq k_{12-1} + k_{1-12} + (k_{12-2} + k_{2-12} - k_{2-2} - k_{1-2} - k_{2-1})^+ \\
R'_1 + 2R'_2 &\leq 2k_{2-2} \leq k_{2-12} + k_{12-2} \\
&\leq k_{12-2} + k_{2-12} + (k_{12-1} + k_{1-12} - k_{1-1} - k_{1-2} - k_{2-1})^+.
\end{aligned}$$

By symmetry, any rate pair (R'_1, R'_2) in $\mathcal{R}_2$ can also be shown to satisfy (5.53)-(5.60). Thus, the achievable rate region specified in (5.53)-(5.60) is strictly larger than the convex hull of $\mathcal{R}_1$ and $\mathcal{R}_2$. $\square$

Another achievable rate region for double-unicast networks has been obtained

in [98], which is given by

$$R_1 \leq k_{1-1} \quad (5.61)$$

$$R_2 \leq k_{2-2} \quad (5.62)$$

$$R_1 + R_2 \leq k_{12-1} + k_{2-12} - k_{2-1} \quad (5.63)$$

$$R_1 + R_2 \leq k_{12-2} + k_{1-12} - k_{1-2} \quad (5.64)$$

$$R_1 + R_2 \leq k_{1-12} + k_{2-12} - k_{1-2} - k_{2-1} \quad (5.65)$$

$$2R_1 + R_2 \leq k_{12-12} + k_{1-12} - k_{1-2} - k_{2-1} \quad (5.66)$$

$$R_1 + 2R_2 \leq k_{12-12} + k_{2-12} - k_{1-2} - k_{2-1}. \quad (5.67)$$

Proposition 5.12. *The achievable rate region specified in (5.53)-(5.60) for double-unicast networks is strictly larger than that given in (5.61)-(5.67).*

Proof. Since the two regions are specified with similar set of inequalities and (5.61)-(5.64) are identical to (5.53)-(5.56), we only need to show that the right hand sides (RHS) of (5.58)-(5.60) are larger than the RHS of (5.65)-(5.67) as follows:

$$\begin{aligned} & (k_{12-2} + k_{2-12} - k_{2-2} - k_{1-2} - k_{2-1})^+ + (k_{12-1} + k_{1-12} - k_{1-1} - k_{1-2} - k_{2-1})^+ \\ & + k_{1-2} + k_{2-1} \\ & \geq k_{12-2} + k_{2-12} - k_{2-2} + k_{12-1} + k_{1-12} - k_{1-1} - k_{1-2} - k_{2-1} \\ & \geq k_{1-12} + k_{2-12} - k_{1-2} - k_{2-1}, \end{aligned} \quad (5.68)$$

$$\begin{aligned} & k_{12-1} + k_{1-12} + (k_{12-2} + k_{2-12} - k_{2-2} - k_{1-2} - k_{2-1})^+ \\ & \geq (k_{12-1} + k_{12-2}) + k_{1-12} + (k_{2-12} - k_{2-2}) - k_{1-2} - k_{2-1} \\ & \geq k_{12-12} + k_{1-12} - k_{1-2} - k_{2-1}, \end{aligned} \quad (5.69)$$

$$\begin{aligned} & k_{12-2} + k_{2-12} + (k_{12-1} + k_{1-12} - k_{1-1} - k_{1-2} - k_{2-1})^+ \\ & \geq (k_{12-2} + k_{12-1}) + k_{2-12} + (k_{1-12} - k_{1-1}) - k_{1-2} - k_{2-1} \\ & \geq k_{12-12} + k_{2-12} - k_{1-2} - k_{2-1}. \end{aligned} \quad (5.70)$$

This completes the proof of proposition 5.12. $\square$

In [97], an achievable rate region for double-unicast networks has been obtained by separately treating the low-interference case with $k_{1-2} + k_{2-1} \leq \min(k_{12-1}, k_{12-2})$ and the high-interference case with $k_{1-2} + k_{2-1} \geq \min(k_{12-1}, k_{12-2})$. The obtained region can be expressed based on the following three sub-regions:

$\bar{\mathcal{R}}_1$	$\bar{\mathcal{R}}_2$	$\bar{\mathcal{R}}_3$
$R_1 \leq k_{12-1} - k_{2-1}$	$R_1 \leq k_{1-1}$	$R_1 \leq \min\{k_{12-1}, k_{12-2}\} - k_{2-1}$
$R_2 \leq k_{12-2} - k_{1-2}$	$R_2 \leq \min\{k_{12-1}, k_{12-2}\} - k_{1-2}$	$R_2 \leq k_{2-2}$
	$R_1 + R_2 \leq \text{rank} \left(\begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12}\mathbf{M}_2 \end{bmatrix} \right)$	$R_1 + R_2 \leq \text{rank} \left(\begin{bmatrix} \mathbf{H}_{21}\mathbf{M}_1 & \mathbf{H}_{22} \end{bmatrix} \right)$

where M_1 and M_2 are some mapping matrices of size $k_{1-12} \times R_1$ and $k_{2-12} \times R_2$, respectively. It was shown in [97] that for the low-interference case, the achievable rate region is given by the convex hull of $\bar{\mathcal{R}}_1$, $\bar{\mathcal{R}}_2$ and $\bar{\mathcal{R}}_3$. For the high-interference case, the rate pairs in the convex hull of $\bar{\mathcal{R}}_2$ and $\bar{\mathcal{R}}_3$ are achievable.

Proposition 5.13. *The achievable rate region specified in (5.53)-(5.60) with our proposed scheme for double-unicast networks is strictly larger than the convex hull of $\bar{\mathcal{R}}_1$, $\bar{\mathcal{R}}_2$ and $\bar{\mathcal{R}}_3$.*

Proof. Since the region specified in (5.53)-(5.60) is convex, it is sufficient to show that it includes all the three regions $\bar{\mathcal{R}}_1$, $\bar{\mathcal{R}}_2$ and $\bar{\mathcal{R}}_3$.

First, we show that any rate pair (R'_1, R'_2) in $\bar{\mathcal{R}}_1$ satisfies the inequalities (5.53)-(5.60) as

$$\begin{aligned}
R'_1 &\leq k_{12-1} - k_{2-1} \leq k_{1-1} \\
R'_2 &\leq k_{12-2} - k_{1-2} \leq k_{2-2} \\
R'_1 + R'_2 &\leq k_{12-1} + k_{12-2} - k_{2-1} - k_{1-2} \leq k_{12-1} + k_{2-2} - k_{2-1} \\
&\leq k_{12-1} + k_{2-12} - k_{2-1} \\
R'_1 + R'_2 &\leq k_{12-1} + k_{12-2} - k_{2-1} - k_{1-2} \leq k_{12-2} + k_{1-1} - k_{1-2} \\
&\leq k_{12-2} + k_{1-12} - k_{1-2} \\
R'_1 + R'_2 &\leq k_{12-1} + k_{12-2} - k_{2-1} - k_{1-2} + (k_{2-12} - k_{2-2}) + (k_{1-12} - k_{1-1}) \\
&\leq (k_{12-2} + k_{2-12} - k_{2-2} - k_{1-2} - k_{2-1})^+ \\
&\quad + (k_{12-1} + k_{1-12} - k_{1-1} - k_{1-2} - k_{2-1})^+ + k_{1-2} + k_{2-1} \\
2R'_1 + R'_2 &\leq k_{1-1} + k_{12-1} - k_{2-1} + k_{12-2} - k_{1-2} \\
&\leq k_{1-12} + k_{12-1} - k_{2-1} + k_{12-2} - k_{1-2} + (k_{2-12} - k_{2-2}) \\
&\leq k_{12-1} + k_{1-12} + (k_{12-2} + k_{2-12} - k_{2-2} - k_{1-2} - k_{2-1})^+ \\
R'_1 + 2R'_2 &\leq k_{2-2} + k_{12-1} - k_{2-1} + k_{12-2} - k_{1-2} \\
&\leq k_{12-2} + k_{2-12} + (k_{12-1} + k_{1-12} - k_{1-1} - k_{1-2} - k_{2-1})^+.
\end{aligned}$$

Similarly, it can be shown that any rate pair (R'_1, R'_2) in $\bar{\mathcal{R}}_2$ also satisfies the inequalities (5.53)-(5.60) by using the fact that

$$\text{rank} \left(\begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12}\mathbf{M}_2 \end{bmatrix} \right) \leq \text{rank} \left(\begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \end{bmatrix} \right) = k_{12-1}. \quad (5.71)$$

Finally, by symmetry, $\bar{\mathcal{R}}_3$ is also included in the rate region specified by (5.53)-(5.60).

This completes the proof of Proposition 5.13. $\square$

To provide a numerical example for the analytical rate comparisons given above, we consider the network shown in Figure 5.2(a). It can be verified that the rate pair $(R_1, R_2) = (1, 2)$ is not within the achievable regions obtained in [50, 67, 97, 98], but it is in the region given by (5.53)-(5.60).

5.2 Joint Routing and RLNC

Although the proposed precoding and decoding schemes in Figure 5.1 are optimal for the resulting linear deterministic IC given any set of RLNC applied at the intermediate nodes [58], there exists a gap between the rate region given in (5.9)-(5.15) and the capacity region of double-unicast networks, as RLNC is in general a suboptimal strategy for the intermediate nodes. Intuitively, with RLNC, the information is mixed at the intermediate nodes to the largest extent, which results in severe interference received at the sink node. On the other hand, with the conventional routing strategy, the path from the source to the corresponding receiver are pre-determined and the information are sent over these edge-disjoint paths. Hence, there is no interference across sessions.

In this section, we propose a joint routing and RLNC (J-R-RLNC) approach for multi-session networks where an achievable region with RLNC can be obtained. The proposed J-R-RLNC approach achieves higher rate region compared to the schemes with either pure routing or pure RLNC. With proposed J-R-RLNC, part of the information is routed to the receivers without network coding, to reduce the inter-session interference, and the rest is transmitted over the network via RLNC. The routing paths and the corresponding routing rates are optimized using Fourier-Motzkin elimination to maximize the overall achievable region. The main idea is illustrated with two special classes of the multi-session networks: the degraded

2-message multicast networks and the double-unicast networks, and it can be readily extended to any other multi-session networks.

5.2.1 J-R-RLNC for Degraded 2-message Multicast Networks

Consider the 3-user degraded 2-message multicast problem shown in Figure 5.4, where a common source sends two messages X_1 and X_2 to three sink nodes, denoted as t_1 , t_2 and t_3 , at rate R_1 and R_2 , respectively. The first two sink nodes, known as “public receivers”, want to recover message X_1 , while the third one would like to recover both messages. This model is motivated by the application of multicasting real-time video to two sets of receivers of different level of priority. This problem has been investigated in [103] where an outer bound of the capacity region has been presented.

FIGURE 5.4: 3-user degraded 2-message multicast network.

If all the intermediate nodes in the network shown in Figure 5.4 perform RLNC, the problem reduces to degraded broadcast network over a linear deterministic channel. Denote the transfer matrix from source s to t_i by $\mathbf{H}_i$, for $i = 1, 2, 3$. It has

be shown in [66] that the capacity of this linear deterministic channel is given by

$$\begin{aligned}
R_1 &\leq \min \{\text{rank}(\mathbf{H}_1), \text{rank}(\mathbf{H}_2)\} \\
R_1 + R_2 &\leq \text{rank}(\mathbf{H}_3) \\
2R_1 + R_2 &\leq \text{rank}(\mathbf{H}_1) + \text{rank}(\mathbf{H}_2) + \text{rank} \left(\begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \mathbf{H}_3 \end{bmatrix} \right) - \text{rank} \left(\begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{bmatrix} \right).
\end{aligned} \tag{5.72}$$

If the field size is large enough, we have

$$\begin{aligned}
\text{rank}(\mathbf{H}_j) &= k_j, j = 1, 2, 3, \\
\text{rank} \left(\begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{bmatrix} \right) &= k_{12}, \\
\text{rank} \left(\begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \mathbf{H}_3 \end{bmatrix} \right) &= k_{123},
\end{aligned}$$

where k_{12} is the min-cut values between the source s and sink nodes $\{t_1, t_2\}$ and k_{123} is the min-cut value between the source and all the three sink nodes. Therefore, with RLNC, we can obtain an achievable region for this network in terms of its min-cuts, which is given by

$$\begin{aligned}
R_1 &\leq \min \{k_1, k_2\} \\
R_1 + R_2 &\leq k_3 \\
2R_1 + R_2 &\leq k_1 + k_2 + k_{123} - k_{12}.
\end{aligned} \tag{5.73}$$

However, the achievable region given in (5.73) is not optimal in general. This is mainly because, with RLNC, the inter-session interference is maximized, which is undesired. Note that since X_2 is demanded only by the third receiver, it is viewed as interference at the other two receivers. Therefore, the achievable region may be enlarged if we can reduce the inter-session interference caused by X_2 to t_1 and t_2 , which can be achieved by sending part of X_2 to t_3 via pure routing.

Assume that there are totally l paths from s to t_3 , denoted as $P_1, \dots, P_l$, and the routing rate assigned for path P_i is denoted by α_i , for $i = 1, \dots, l$. Note that the

routing rate assignment is feasible if and only if it satisfies

$$\alpha_i \geq 0, i = 1, \dots, l \quad (5.74)$$

$$\sum_{\{i:e \in P_i\}} \alpha_i \leq C_e, \forall e \in E, \quad (5.75)$$

where the second constraint given in (5.75) follows from the capacity of edge e . After subtracting the amount of resource used for routing, we may re-label the capacity of each edge in G by

$$C'_e = C_e - \sum_{\{i:e \in P_i\}} \alpha_i, \forall e \in E. \quad (5.76)$$

Denote the resultant network by G' and the min-cut from the source s to receiver t_J in G' by k'_J , for $J \subset \{1, 2, 3\}$. The rate pair (R_1, R_2) is achievable in G' by applying random network coding with proper precoding if it satisfies

$$\begin{aligned} R_1 &\leq \min \{k'_1, k'_2\} \\ R_1 + R_2 &\leq k'_3 \\ 2R_1 + R_2 &\leq k'_1 + k'_2 + k'_{123} - k'_{12}. \end{aligned} \quad (5.77)$$

By including the message sent by routing, the overall achievable region is thus given by

$$\begin{aligned} R_1 &\leq \min \{k'_1, k'_2\} \\ R_1 + R_2 &\leq k'_3 + \sum_{i=1}^q \alpha_i \\ 2R_1 + R_2 &\leq k'_1 + k'_2 + k'_{123} - k'_{12} + \sum_{i=1}^q \alpha_i. \end{aligned} \quad (5.78)$$

With Fourier-Motzkin elimination, the achievable region given in (5.78) can be maximized subject to the constraints given in (5.74) and (5.75).

Example 5.3. Consider the 3-user degraded 2-message multicast network shown in Figure 5.5, where each edge in the network has unit capacity.

Since $k_1 = 1$, $k_2 = 2$ and $k_3 = k_{12} = k_{123} = 3$ for this network, the achievable region by random network coding is

$$\begin{aligned} R_1 &\leq 1 \\ 2R_1 + R_2 &\leq 3. \end{aligned} \quad (5.79)$$

FIGURE 5.5: (a) A 3-user degraded 2-message multicast network example. (b) The achievable regions obtained by RLNC and J-R-RLNC.

To obtain an achievable region by J-R-RLNC, we first list all paths from s to t_3 , which are

$$\begin{aligned}
 P_1 &= \{e_1, e_{10}\} \\
 P_2 &= \{e_1, e_4, e_{11}, e_{13}\} \\
 P_3 &= \{e_2, e_5, e_{11}, e_{13}\} \\
 P_4 &= \{e_2, e_6, e_{14}\} \\
 P_5 &= \{e_3, e_8, e_{14}\}
 \end{aligned}$$

Denote the routing rate along path P_i by α_i for $i = 1, \dots, 5$. The nontrivial constraints on α_i as given in (5.74) and (5.75) are

$$\begin{aligned}
 \alpha_i &\geq 0, i = 1, \dots, 5 \\
 \alpha_1 + \alpha_2 &\leq 1 \\
 \alpha_3 + \alpha_4 &\leq 1 \\
 \alpha_2 + \alpha_3 &\leq 1 \\
 \alpha_4 + \alpha_5 &\leq 1
 \end{aligned} \tag{5.80}$$

The new achievable region, as that given in (5.78), is

$$\begin{aligned} R_1 &\leq \min\{1 - \alpha_1 - \alpha_2, 2 - \alpha_2 - \alpha_3 - \alpha_5\} \\ R_1 + R_2 &\leq \min\{3 - \alpha_2, 3 - \alpha_4\} \\ 2R_1 + R_2 &\leq \min\{6 - \alpha_2 + \alpha_4, 7 - 2\alpha_2 + \alpha_4 - \alpha_5\} - \min\{3, 3 - \alpha_2 + \alpha_4\}. \end{aligned} \quad (5.81)$$

By performing Fourier-Motzkin elimination over (5.80) and (5.81), we can obtain the achievable region on the rate pair (R_1, R_2) , which is given by

$$\begin{aligned} R_1 &\leq 1 \\ 3R_1 + 2R_2 &\leq 6. \end{aligned} \quad (5.82)$$

Specifically, the rate pair $(R_1, R_2) = (1, \frac{3}{2})$ is achieved with $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_5 = 0$ and $\alpha_4 = 0.5$, i.e., we first route X_2 from s to t_3 via P_4 at the rate of 0.5 and then applying random network coding with proper precoding on the rest of the network to achieve $(R_1, R_2) = (1, 1)$. Note that the achievable region by proposed J-R-RLNC given in (6.21) actually matches the capacity region of this network and is thus optimal.

Lemma 5.14. *The new achievable region specified by (5.74), (5.75) and (5.78) is no smaller than that achieved by pure RLNC in (5.73) and that achieved by pure routing.*

Proof. Note that RLNC is a special setting of the proposed J-R-RLNC by choosing $\alpha_i = 0, i = 1, \dots, q$. Since the proposed scheme optimize the achievable region over all feasible α_i , it is guaranteed to include all the achievable rate pairs in (5.73).

Moreover, if a rate pair (R_1, R_2) is achieved by pure routing, there exists paths from s to t_3 of capacity at least R_2 . After subtracting those resources used for routing X_2 , the min-cut from s to t_i for $i = 1, 2, 3$ should be no smaller than R_1 since the routing capacity from s to t_{123} is upper bounded by $\min\{k_1, k_2, k_3\}$. Therefore, choosing the same routing paths for X_2 as that in this routing solution and performing RLNC over the rest of graph, it is guaranteed to achieve rate pair (R_1, R_2) . $\square$

It is observed that the number of variables involved in formulating the new achievable region equals to the number of possible paths from s to t_3 , which may be quite large. Without loss of optimality, we can simplify the formulation by only

considering the paths satisfying the following condition:

$$\begin{aligned}
& \exists 0 < \alpha \leq \min_{e \in P} C_e \text{ such that} \\
& \alpha + \min_{\mathcal{C}_1} \{|\mathcal{C}_1| - \alpha|\mathcal{C}_1 \cap P|\} + \min_{\mathcal{C}_2} \{|\mathcal{C}_2| - \alpha|\mathcal{C}_2 \cap P|\} \\
& - \min_{\mathcal{C}_{12}} \{|\mathcal{C}_{12}| - \alpha|\mathcal{C}_{12} \cap P|\} + \min_{\mathcal{C}_{123}} \{|\mathcal{C}_{123}| - \alpha|\mathcal{C}_{123} \cap P|\} \\
& > k_1 + k_2 + k_{123} - k_{12}.
\end{aligned} \tag{5.83}$$

Such a choice follows the intuition that the first two inequalities given in (5.73) matches the outer bound and thus cannot be improved. Therefore, the achievable region can be improved only if the third inequality in (5.73) can be relaxed.

Example 5.4. Consider the network shown in Figure 5.5. Although there are 5 possible paths from s to t_3 , the only path that satisfies the condition given in (5.83) is P_4 . Therefore, the previous procedures in obtaining the new achievable region can be significantly simplified by only considering routing along the path P_4 .

Unless the achievable region by RLNC in (5.73) is optimal, the new achievable region by proposed J-R-RLNC in (5.78) will *strictly* improve over it if there exists at least one path from s to t_3 in the network that satisfies the condition given in (5.83).

5.2.2 J-R-RLNC for Double-Unicast Networks

Consider a general double-unicast network $G = (V, E)$ where the capacity of edge $e \in E$ is denoted by C_e . With the proposed precoding scheme and RLNC performed at the intermediate nodes, the region specified in (5.53)-(5.60) can be achieved. To further enlarge this region by incorporating with routing, we first need to choose the proper routing paths and determine the associated routing rates. Assume that there are m paths from s_1 to t_1 , denoted as $P_1^1, \dots, P_m^1$, and n paths from s_2 to t_2 , denoted as $P_1^2, \dots, P_n^2$. Let the routing rate assigned to path P_i^1 and P_j^2 be α_i^1 and α_j^2 , respectively. We then have the following constraints:

$$\alpha_i^1 \geq 0, i = 1, \dots, m \tag{5.84}$$

$$\alpha_j^2 \geq 0, j = 1, \dots, n \tag{5.85}$$

$$\sum_{\{i: e \in P_i^1\}} \alpha_i^1 + \sum_{\{j: e \in P_j^2\}} \alpha_j^2 \leq C_e, \forall e \in E, \tag{5.86}$$

where (5.84) and (5.85) are non-negativity constraints of routing rates, and (5.86) says that the total amount of information routed along an edge should not exceed its capacity.

After subtracting the resource used for routing, we can obtain a new network G' with the capacity for link e reduced to:

$$C'_e = C_e - \sum_{\{i:e \in P_i^1\}} \alpha_i^1 - \sum_{\{j:e \in P_j^2\}} \alpha_j^2, \forall e \in E. \quad (5.87)$$

Denote the min-cut between the set of source s_I and the set of receivers t_J in G' by k'_{I-J} , where $I, J \subseteq \{1, 2\}$. We can obtain an achievable rate region on G' employing RLNC by substituting the min-cut values into (5.53)-(5.60), denoted as $\mathcal{R}'(R_1, R_2)$. Adding the rates achieved by routing for each session, the new achievable rate region is thus given by $\mathcal{R}'(R_1 + \sum_{i=1}^m \alpha_i^1, R_2 + \sum_{j=1}^n \alpha_j^2)$. Therefore, the routing rates $\{\alpha_i^1, \alpha_j^2\}$ can be solved by maximizing the new achievable region subject to the constraints given in (5.84)-(5.86). This linear optimization problem can be solved by using the standard Fourier-Motzkin elimination techniques.

Example 5.5. Consider the double-unicast network in Figure 5.6, where each edge has unit capacity. With RLNC performed at the intermediate nodes, the optimal achievable region is given by

$$\begin{aligned} R_1 &\leq 2 \\ R_1 + R_2 &\leq 3. \end{aligned} \quad (5.88)$$

By applying the proposed J-R-RLNC, the achievable region can be enlarged to

$$\begin{aligned} R_1 &\leq 2 \\ R_1 &\leq 3 \\ R_1 + R_2 &\leq 4. \end{aligned} \quad (5.89)$$

For instance, the rate pair $(R_1, R_2) = (2, 2)$ is achieved by routing X_1 through the path $P_1^1 = \{e_2, e_{12}\}$, routing X_2 through $P_1^2 = \{e_4, e_{10}\}$ each at unit rate and then applying random network coding over the rest of the network.

FIGURE 5.6: (a) An example of double-unicast networks. (b) The achievable region with RLNC and J-R-RLNC.

5.2.3 Extension to General Multi-session Networks

The proposed J-R-RLNC can be applied to any multi-session network where an achievable region with RLNC can be obtained. By routing part of those information that are not required by all receivers, the inter-session interference can be reduced and thus the achievable region may be enlarged compared to the one achieved by RLNC. The routing path and the corresponding rate can be determined based on the optimization result using simple Fourier-Motzkin elimination, similar to that introduced in previous sections for degraded two-message multicast network and double-unicast networks.

5.3 Chapter Summary

In this chapter, we have proposed a linear precoding and decoding scheme for the source and sink nodes of double-unicast networks by assuming that RLNC is performed at each intermediate node. The main idea of the proposed scheme is rate splitting and the concatenated precoding composed of zero-forcing and random block-diagonal precoding matrices. Since the proposed scheme achieves the capacity region of the resulting linear deterministic IC, it gives the maximum achievable rate region for double-unicast network for any given set of coding coefficients used

at the intermediate nodes. A suboptimal rate region in terms of the min-cuts of the network has also been obtained. Analytical comparisons have been provided to show the strict rate improvement over the existing schemes.

Since RLNC is in general a suboptimal strategy for the intermediate nodes, there exists a gap between the obtained achievable rate region and the capacity region of double-unicast networks. We have shown that by routing some information through properly chosen paths, the inter-session interference can be reduced and an even larger achievable rate region may be obtained. However, the complexity of the proposed optimization problem for determining the routing rates grows exponentially with the network size. One possible future work direction is to formulate the problem with edge variables, instead of path variables, so that the computational complexity can be reduced.

Chapter 6

Spatial-Temporal Network Coding for Erasure Networks

In the preceding chapters, we have focused on deterministic networks, i.e., all the communication links in the networks are lossless. In this chapter, the focus will be on erasure networks, i.e., each link is erased with certain probability. The study of erasure networks with network coding was pioneered in [56], where each link is an independent erasure channel. Later, the broadcast nature of wireless environment was accounted for the erasure networks in [53]. The capacity of multiple-source single-receiver erasure networks has been obtained in [53], which has a nice max-flow min-cut interpretation. Single-source single-receiver erasure networks with finite-field additive interference has been investigated in [104], which was later extended to multiple-source single-receiver erasure networks in [105]. In the first part of this chapter, we derive an analytical expression for the throughput of single-session multicast erasure networks with finite-length spatial-temporal network coding. In the second part, we propose a random linear network coding approach for multi-source single-sink erasure networks and show that the analytical results derived in the first part can be directly applied for multi-source single-sink networks with finite temporal network coding length.

6.1 Throughput Analysis of Finite-Length Spatial -Temporal Network Coding

For *deterministic* single-session multicast networks, the capacity is achievable by combining the packets received at each of the intermediate nodes coming from different edges, which is known as *spatial* network coding [106]. To achieve the capacity for general deterministic multi-session networks, *temporal* network coding, i.e., coding over the packets arriving at successive time slots is necessary in general [107]. For erasure networks considered in [53, 56, 104, 105], spatial-temporal network coding across infinite number of time slots is needed to achieve the network capacity.

In general, the throughput of spatial-temporal network coding for erasure networks increases monotonically with the temporal coding length M and it approaches the network capacity as M goes to infinity. In practical implementations, M is usually limited to a small-to-moderate number due to constraints such as tolerable communication delay, available buffer size, and/or coding/decoding complexity [108]. Practically, one would like to determine the minimum temporal coding length M to guarantee a target throughput. To this end, we derive the expected throughput of the erasure networks employed with spatial-temporal network coding as a function of M . When $M = 1$, the problem reduces to finding the network capacity with spatial network coding only, which has been shown to be equal to the statistical mean of the min-cut value of the network [109]. Evaluating this statistical mean is rather challenging, and its asymptotical behavior for layered network has been studied in [110]. In [111], the authors studied the statistical mean of the min-cut value of fully-connected networks, i.e., there is an edge between any pair of nodes.

Different from prior works [109–111] which dealt with $M = 1$, the results derived in this chapter are applicable to any temporal coding length M for arbitrary erasure network topologies with single-session communication. Since finding the exact throughput is intractable [110], we obtain a tight approximate expression in terms of the min-cut value k , number of minimum cuts N , and number of critical edges L of the network. When $M \rightarrow \infty$, the obtained result reduces to the capacity of the network. Numerical examples demonstrate a very good match between the analytical and simulation results for $M < \infty$. While k and L of a network can be determined efficiently, counting the number of minimum cuts N may be difficult for large networks. Hence, a lower bound for the throughput that is independent of N is also developed.

The derived results can be used as a benchmark for evaluating the efficiency of various practical network coding schemes [112–115]. In particular, the batched sparse code (BATS code) [57], which is designed based on joint fountain code and RLNC, is found to achieve the derived throughput for erasure networks.

6.1.1 Illustrative Example: Line Network

We first illustrate our main idea with a simple two-hop line network as shown in Figure 6.1, where a set of packets is to be transmitted from the source node s to the sink node t via the relay node r . Without loss of generality, we assume that both edges e_1 and e_2 are capable of transmitting one packet per channel use with packet loss probability p , with $0 < p < 1$. It is not difficult to see that the routing capacity of this network is given by $(1 - p)^2$ per channel use. With temporal network coding of sufficiently large coding length, the network capacity, which is $(1 - p)$, can be achieved. On the other hand, if the temporal coding length is limited to a finite value M , obtaining the expected throughput is non-trivial, which we will pursue next.

FIGURE 6.1: A two-hop line network with packet loss rate p .

During each coding block of M time slots, the number of packets that can be successfully delivered by an edge e_i , denoted by $\mathbf{X}_i$, is a random variable following the Binomial distribution $\mathcal{B}(M, 1 - p)$, i.e.,

$$\Pr(\mathbf{X}_i = x) = \binom{M}{x} (1 - p)^x p^{M-x}, \quad x = 0, 1, \dots, M. \quad (6.1)$$

With spatial-temporal network coding of length M , the relay node will generate M packets based on the $\mathbf{X}_1$ received packets. Among those M packets, $\mathbf{X}_2$ number of packets will be successfully received by t . Therefore, the total number of innovative packets at t is $\mathbf{Y} = \min(\mathbf{X}_1, \mathbf{X}_2)$. With ordered statistics, the expected throughput can be computed as

$$\mathbb{E}[\mathbf{Y}] = \sum_{x=0}^M x f(x) \left(f(x) + 2 \sum_{i=x+1}^M f(i) \right), \quad (6.2)$$

where $f(x) = \binom{M}{x}(1-p)^x p^{M-x}$, $x = 0, 1, \dots, M$.

For moderate-to-large M , $\mathcal{B}(M, 1-p)$ can be well approximated by a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, with $\mu = M(1-p)$ and $\sigma^2 = Mp(1-p)$. As a result, (6.2) can be evaluated as

$$\begin{aligned} \mathbb{E}[\mathbf{Y}] &\approx \int_0^\infty \frac{2x}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) Q\left(\frac{x-\mu}{\sigma}\right) dx \\ &= \mu - \frac{\sigma}{\sqrt{\pi}}, \end{aligned} \quad (6.3)$$

where $Q(\cdot)$ is the Gaussian Q function defined as $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{\tau^2}{2}} d\tau$. The detailed steps for deriving (6.3) can be found in Appendix D.

Therefore, the normalized throughput is given by

$$\rho = \frac{\mathbb{E}[\mathbf{Y}]}{M} \approx (1-p) - \sqrt{\frac{p(1-p)}{\pi M}}. \quad (6.4)$$

It is observed from (6.4) that as $M \rightarrow \infty$, ρ approaches $(1-p)$, which is the capacity of this two-hop line network.

In the remaining parts of this chapter, we will derive the normalized throughput for general erasure networks employing spatial-temporal network coding with finite temporal coding length.

6.1.2 Single-Session Erasure Networks

We consider a directed acyclic single-session network represented by a directed graph $G = (V, E)$, where V is the set of nodes and E is the set of edges. Let $s \in V$ and $t \in V$ denote the source and sink nodes, respectively. We assume that each edge is capable to convey one packet at each time slot with independent and memoryless packet loss at rate p . Let C denote a cut of the network, i.e., C contains a set of edges whose removal will disconnect the source node s and the sink node t , and let Ψ be the collection of all such cuts. The max-flow k of the network is then given by the cardinality of the minimum cuts, i.e., $k = \min_{C \in \Psi} |C|$. Let Ψ^* denote the collection of all *minimum* cuts, i.e., $\Psi^* \triangleq \{C : |C| = k\} \subseteq \Psi$. Since each edge is assumed to have unit capacity, the cardinality of the minimum cuts is the same as their capacity.

Consider an intermediate node $v \in V$ with A incoming edges and B outgoing

FIGURE 6.2: An intermediate node v with spatial-temporal network coding.

edges, as shown in Figure 6.2. With spatial-temporal network coding of temporal coding length M applied at node v , all the packets received from the A incoming edges during the M consecutive time slots are combined to generate MB packets, which are to be sent over the B outgoing edges during the next M time slots. Denote by $P_1^{(a)}, \dots, P_M^{(a)}$ the packets conveyed by the incoming edge e_a during one coding block of M time slots, where $a = 1, \dots, A$. Due to packet loss, some of these packets may not be received by node v . Denote the set of successfully received packets at node v during one coding block as $\Lambda_v \subseteq \{P_i^{(a)} : 1 \leq i \leq M, 1 \leq a \leq A\}$. The packet $Z_i^{(b)}$ sent by node v to the b th outgoing edge during the i th time slot is then a function of all the received packets at node v during this coding block, i.e., $Z_i^{(b)} = f_i^{(b)}(\Lambda_v)$, $i = 1, \dots, M, b = 1, \dots, B$, where $f_i^{(b)}(\cdot)$ is the encoding function. After all the MB packets have been sent, node v then clears the buffer containing those received packets Λ_v and proceeds to the next coding block.

In the special case of $M = 1$, the above scheme reduces to spatial network coding only, whose capacity is [109]

$$\rho_1 = \mathbb{E}[\text{mincut}_{s \rightarrow t}(G)] = \mathbb{E} \left[\min_{C \in \Psi} \sum_{i: e_i \in C} \mathbf{Z}_i \right], \quad (6.5)$$

where e_i denotes the i th edge of the network and $\mathbf{Z}_i$'s are independent and identically distributed (i.i.d.) Bernoulli random variables with $\Pr(\mathbf{Z}_i = 0) = p$ and $\Pr(\mathbf{Z}_i = 1) = 1 - p$.

In the other extreme case with $M \rightarrow \infty$, the network coding scheme is proven to achieve the capacity of the network [7], which is

$$\rho_\infty = k(1 - p), \quad (6.6)$$

where k is the min-cut between the source and sink node.

For the general scenario with $1 < M < \infty$, the spatial-temporal network

coding scheme described above can be equivalently treated as a spatial network coding scheme over a network with the capacity of each edge scaled up by M folds. Therefore, with erasure probability p at each edge, the number of packets that can be successfully delivered by edge $e_i \in E$ during each coding block of M time slots, denoted by $\mathbf{X}_i$, is a random variable that follows the Binomial distribution $\mathcal{B}(M, 1 - p)$. The throughput of the network with spatial-temporal network coding of length M can thus be written as [109]:

$$\rho_M = \frac{\mathbb{E}[\min_{C \in \Psi} \sum_{e_i \in C} \mathbf{X}_i]}{M}, \quad (6.7)$$

where $\mathbf{X}_i$'s are i.i.d. according to $\mathcal{B}(M, 1 - p)$.

Finding the exact value for ρ_M in (6.7) is challenging since it requires evaluating the statistical distribution for all cuts of the network. In fact, even finding the cuts has been shown to be NP-hard [116]. In the remaining part of this chapter, we will derive a tight approximation for (6.7) represented in terms of M and p .

6.1.3 Throughput Analysis

For notational convenience, we define $\mathbf{X}_C \triangleq \sum_{i: e_i \in C} \mathbf{X}_i$. ρ_M in (6.7) can then be alternatively written as

$$\begin{aligned} \rho_M &= \frac{1}{M} \mathbb{E} \left[\min_{C \in \Psi} (\mathbf{X}_C) \right] \\ &= \frac{1}{M} \sum_{C \in \Psi} \mathbb{E}[\mathbf{X}_C] \mathbf{1}(\mathbf{X}_C \leq \mathbf{X}_{C'}, \forall C' \in \Psi), \end{aligned} \quad (6.8)$$

where $\mathbf{1}(T)$ is an indicator function for event T , i.e., $\mathbf{1}(T) = 1$ if event T happens and it returns 0 otherwise.

The following lemma shows that to evaluate ρ_M in (6.8), it is sufficient to consider the collection of minimum cuts Ψ^* only for reasonably large M .

Lemma 6.1. *Let $C \in \Psi \setminus \Psi^*$ be a non-minimum cut and $C^* \in \Psi^*$ be a minimum cut, then $\Pr(\mathbf{X}_C \leq \mathbf{X}_{C^*})$ decreases exponentially fast with M .*

Proof. Since $C \notin \Psi^*$ and $C^* \in \Psi^*$, we have $|C| > k$ and $|C^*| = k$. Let $|C| = k + m$, $m \geq 1$ and $|C \cap C^*| = l$. As those edges in $C \cap C^*$ are common to both C and C^* , to compare $\mathbf{X}_C$ with $\mathbf{X}_{C^*}$, we only need to compare $\mathbf{X}_{C \setminus C^*}$ with $\mathbf{X}_{C^* \setminus C}$, which are

independent Binomial random variables. Let $\mu = M(1 - p)$ and $\sigma^2 = Mp(1 - p)$. When $(k - l)\mu$ is reasonably large, $\mathbf{X}_{C \setminus C^*}$ and $\mathbf{X}_{C^* \setminus C}$ can be approximated as Gaussian random variables distributed according to

$$\begin{aligned}\mathbf{X}_{C \setminus C^*} &\sim \mathcal{N}((k + m - l)\mu, (k + m - l)\sigma^2), \\ \mathbf{X}_{C^* \setminus C} &\sim \mathcal{N}((k - l)\mu, (k - l)\sigma^2).\end{aligned}\tag{6.9}$$

Moreover, $\mathbf{X}_{C \setminus C^*}$ and $\mathbf{X}_{C^* \setminus C}$ are independent variables as the erasure events happen independently over different links. Therefore, $\mathbf{X}_{C^* \setminus C} - \mathbf{X}_{C \setminus C^*}$ can be approximated as a Gaussian random variable distributed as $\mathcal{N}(-m\mu, (2k + m - 2l)\sigma^2)$ and we have

$$\begin{aligned}\Pr(\mathbf{X}_C \leq \mathbf{X}_{C^*}) &= \Pr(\mathbf{X}_{C \setminus C^*} \leq \mathbf{X}_{C^* \setminus C}) \\ &= \Pr(\mathbf{X}_{C^* \setminus C} - \mathbf{X}_{C \setminus C^*} \geq 0) \\ &= Q\left(\frac{m\mu}{\sigma\sqrt{2k + m - 2l}}\right) \\ &\stackrel{(a)}{\leq} \frac{1}{2} \exp\left(-\frac{m^2(1 - p)M}{(4k + 2m - 4l)p}\right),\end{aligned}$$

where (a) follows from the Chernoff bound of Q -function, i.e., $Q(x) \leq \frac{1}{2} \exp(-\frac{x^2}{2})$. $\square$

With Lemma 6.1, for reasonably large M , we have

$$\begin{aligned}\forall C \in \Psi \setminus \Psi^*, \mathbf{1}(\mathbf{X}_C \leq \mathbf{X}_{C'}, \forall C' \in \Psi) &\approx 0 \\ \forall C^* \in \Psi^*, \mathbf{1}(\mathbf{X}_{C^*} \leq \mathbf{X}_{C'}, \forall C' \in \Psi) &\approx \mathbf{1}(\mathbf{X}_{C^*} \leq \mathbf{X}_{C'}, \forall C' \in \Psi^*).\end{aligned}$$

Therefore, (6.8) can be approximated as

$$\begin{aligned}\rho_M &\approx \frac{1}{M} \sum_{C^* \in \Psi^*} \mathbb{E}[\mathbf{X}_{C^*}] \mathbf{1}(\mathbf{X}_{C^*} \leq \mathbf{X}_{C'} : C' \in \Psi^*) \\ &= \frac{\mathbb{E}[\min_{C^* \in \Psi^*} (\sum_{e_i \in C^*} \mathbf{X}_i)]}{M}.\end{aligned}\tag{6.10}$$

The expression given in (6.10) shows that the throughput of the network can be approximated by the statistical mean of all the *minimum cuts*. To evaluate ρ_M in (6.10), we denote the number of minimum cuts and the number of edges that belong to at least one of the minimum cuts by N and L , respectively, i.e., $N = |\Psi^*|$ and $L = |\{ \cup C^*, C^* \in \Psi^* \}|$. Note that these L edges are known as *critical edges* [117]

and the removal of any of them will decrease the network capacity. Further denote by $\{\mathbf{X}_{i_1}, \dots, \mathbf{X}_{i_L}\}$ the number of packets that are successfully delivered by those L critical edges during the M time slots, where $\mathbf{X}_{i_j} \sim \mathcal{B}(M, 1 - p), j = 1, \dots, L$. We use $\mathbf{Y}_j$ to denote the j th smallest random variable among $\{\mathbf{X}_{i_1}, \dots, \mathbf{X}_{i_L}\}$, e.g., $\mathbf{Y}_1 = \min\{\mathbf{X}_{i_1}, \dots, \mathbf{X}_{i_L}\}$, $\mathbf{Y}_L = \max\{\mathbf{X}_{i_1}, \dots, \mathbf{X}_{i_L}\}$. By approximating the Binomial distribution with Gaussian distribution, the statistic means of $\mathbf{Y}_j$ can be obtained as [118]

$$\mathbb{E}[\mathbf{Y}_j] \approx \mu + \sigma \Phi^{-1} \left(\frac{j - 0.375}{L + 0.25} \right), j = 1, \dots, L, \quad (6.11)$$

where $\mu = M(1 - p)$, $\sigma^2 = Mp(1 - p)$, and $\Phi(\cdot)$ is the cumulative distribution function (cdf) of the standard Gaussian distribution $\mathcal{N}(0, 1)$.

Since the variables in $\{\mathbf{X}_{i_1}, \dots, \mathbf{X}_{i_L}\}$ are i.i.d., we have

$$\Pr(\mathbf{Y}_j = \mathbf{X}_{i_l}) = 1/L, \forall l = 1, \dots, L.$$

Let I denote a set of randomly chosen k indices from $\{1, \dots, L\}$. The probability that the set $\{\mathbf{Y}_i, i \in I\}$ corresponds to a valid minimum cut of the network is

$$q = \frac{N}{\binom{L}{k}}. \quad (6.12)$$

Consider two sets of indices $I_1 \subset \{1, \dots, L\}$ and $I_2 \subset \{1, \dots, L\}$ with $|I_1| = |I_2| = k$. If both $\{\mathbf{Y}_i, i \in I_1\}$ and $\{\mathbf{Y}_j, j \in I_2\}$ correspond to valid minimum cuts of the network, the throughput of the network will be limited by the smaller one, i.e., $\min\{\sum_{i \in I_1} \mathbf{Y}_i, \sum_{j \in I_2} \mathbf{Y}_j\}$. Recall that $\{\mathbf{Y}_j\}$ are ordered random variables, i.e., $\mathbf{Y}_1 \leq \mathbf{Y}_2 \leq \dots \leq \mathbf{Y}_L$, and their statistical mean given in (6.11) is approximately linear in j . To simplify the analysis, we make the following approximations:

$$\sum_{i \in I_1} \mathbf{Y}_i < \sum_{j \in I_2} \mathbf{Y}_j, \text{ if } \sum_{i \in I_1} i < \sum_{j \in I_2} j \quad (6.13)$$

$$\sum_{i \in I_1} \mathbf{Y}_i = \sum_{j \in I_2} \mathbf{Y}_j, \text{ if } \sum_{i \in I_1} i = \sum_{j \in I_2} j \quad (6.14)$$

Example 6.1. Consider the butterfly network shown in Figure 2.1. The collection of the minimum cuts between s and t_1 is

$$\Psi^* = \{\{e_1, e_2\}, \{e_1, e_5\}, \{e_1, e_7\}, \{e_1, e_8\}, \{e_3, e_7\}, \{e_3, e_8\}\}.$$

Therefore, we have $k = 2$, $N = 6$, and $L = 6$. The numbers of successfully

FIGURE 6.3: The Butterfly network.

transmitted packets over the 6 edges appearing in Ψ^* are then denoted by the random variables $\{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_5, \mathbf{X}_7, \mathbf{X}_8\}$ and the j th smallest value is denoted by $\mathbf{Y}_j, j = 1, \dots, 6$. The probability that $\{\{\mathbf{Y}_i, \mathbf{Y}_j\}, 1 \leq i < j \leq 6\}$ corresponds to a valid minimum cut in Ψ^* can be calculated as $q = \frac{2}{5}$. All possible combinations of $\{\{\mathbf{Y}_i, \mathbf{Y}_j\}, 1 \leq i < j \leq 6\}$ are listed according to the sum of the indices $n = i + j$ in Table 6.1

TABLE 6.1: List of all possible combinations of $\{\{\mathbf{Y}_i, \mathbf{Y}_j\}, 1 \leq i < j \leq 6\}$

$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$
$\{\mathbf{Y}_1, \mathbf{Y}_2\}$	$\{\mathbf{Y}_1, \mathbf{Y}_3\}$	$\{\mathbf{Y}_1, \mathbf{Y}_4\}$ $\{\mathbf{Y}_2, \mathbf{Y}_3\}$	$\{\mathbf{Y}_1, \mathbf{Y}_5\}$ $\{\mathbf{Y}_2, \mathbf{Y}_4\}$	$\{\mathbf{Y}_1, \mathbf{Y}_6\}$ $\{\mathbf{Y}_2, \mathbf{Y}_5\}$ $\{\mathbf{Y}_3, \mathbf{Y}_4\}$
$n = 8$	$n = 9$	$n = 10$	$n = 11$	
$\{\mathbf{Y}_2, \mathbf{Y}_6\}$ $\{\mathbf{Y}_3, \mathbf{Y}_5\}$	$\{\mathbf{Y}_3, \mathbf{Y}_6\}$ $\{\mathbf{Y}_4, \mathbf{Y}_5\}$	$\{\mathbf{Y}_4, \mathbf{Y}_6\}$	$\{\mathbf{Y}_5, \mathbf{Y}_6\}$	-

We are now ready to evaluate the expression given in (6.10). The probability that the $\min_{C^* \in \Psi^*}(\cdot)$ operator in (6.10) returns $\{\mathbf{Y}_1, \mathbf{Y}_2\}$ is q . It will return $\{\mathbf{Y}_1, \mathbf{Y}_3\}$ if $\{\mathbf{Y}_1, \mathbf{Y}_2\}$ does not correspond to a valid minimum cut while $\{\mathbf{Y}_1, \mathbf{Y}_3\}$ does, which happens with probability $(1 - q)q$. Furthermore, based on the assumption given in (6.14), $\{\mathbf{Y}_1, \mathbf{Y}_4\}$ and $\{\mathbf{Y}_2, \mathbf{Y}_3\}$ will be returned with equal probability $\frac{(1-q)^2[1-(1-q)^2]}{2}$.

Let $\Pr(\{\mathbf{Y}_i, \mathbf{Y}_j\})$ be the probability that the $\min_{C^* \in \Psi^*}(\cdot)$ operator in (6.10) returns the combination $\{\mathbf{Y}_i, \mathbf{Y}_j\}$, and $\Pr(\mathbf{Y}_i)$ be the probability that $\mathbf{Y}_i$ is one of

the returned random variable. For instance, with the example given in Table 6.1, we have $\Pr(\mathbf{Y}_1) = \Pr(\{\mathbf{Y}_1, \mathbf{Y}_2\}) + \Pr(\{\mathbf{Y}_1, \mathbf{Y}_3\}) + \Pr(\{\mathbf{Y}_1, \mathbf{Y}_4\}) + \Pr(\{\mathbf{Y}_1, \mathbf{Y}_5\}) + \Pr(\{\mathbf{Y}_1, \mathbf{Y}_6\})$. Hence, the throughput of this butterfly network can be computed as

$$\begin{aligned} \rho_M &\approx \frac{\mathbb{E}[\min_{C^* \in \Psi^*}(\sum_{e_i \in C^*} \mathbf{X}_i)]}{M} \\ &= \frac{\sum_{i=1}^6 \Pr(\mathbf{Y}_i) \mathbb{E}[\mathbf{Y}_i]}{M}, \end{aligned} \quad (6.15)$$

where $\mathbb{E}[\mathbf{Y}_i]$ can be computed from (6.11).

Now, we are ready to extend the analysis for the butterfly network shown in Example 6.1 to general erasure networks. Consider a general network with L critical edges, N minimum cuts and the min-cut value k . Let $\Pr(\mathbf{Y}_I)$ be the probability that the $\min_{C^* \in \Psi^*}(\cdot)$ operator in (6.10) returns the set $\mathbf{Y}_I$ and $\Pr(n)$ be the probability that the sum of the indices of the returned variables equals to n , where $n \in \left\{ \frac{k(k+1)}{2}, \dots, \frac{k(2L-k+1)}{2} \right\}$, i.e., $\Pr(n) = \sum_{I: \sum_{i \in I} i = n} \Pr(\mathbf{Y}_I)$.

To evaluate $\Pr(n)$, let $P(n, k, L)$ be the number of ways to partition a positive integer n into k distinct parts with each part being a positive integer not exceeding L . Note that the problem of determining $P(n, k, L)$ has been extensively studied in [119] and they can be computed recursively as¹

$$\begin{aligned} P(n, 2, L) &= \begin{cases} \lfloor \frac{n-1}{2} \rfloor, & 3 \leq n \leq L \\ L - \lfloor \frac{n}{2} \rfloor, & L < n \leq 2L - 1, \\ 0, & \text{otherwise} \end{cases} \\ P(n, k, L) &= \sum_{i=2}^L Q(n-i, k-1, i-1), \forall k > 2. \end{aligned}$$

For notational convenience, let $F(n, k, L) \triangleq \sum_{i=\frac{k(k+1)}{2}}^n P(i, k, L)$ be the cumulative function of $P(n, k, L)$. Based on the approximations made in (6.13), the $\min(\cdot)$ operator in (6.10) will return a set of variables with indices summing up to n only if all these sets of variables whose indices summing up to a value smaller than n do not correspond to the valid minimum cuts. For example, let $n_1 = \frac{k(k+1)}{2}$, which is the smallest sum of k distinct integers. Since $P(n_1, k, L) = P(n_1 + 1, k, L) = 1$, we have $\Pr(n_1) = q$ and $\Pr(n_1 + 1) = (1 - q)q$. Furthermore, according to (6.14), all

¹The details for deriving the recursive formulas for $P(n, k, L)$ and $J(n, k, L, i)$ can be found in Appendix E.

the $P(n, k, L)$ sets of variables whose indices summing up to n are returned with equal probability. Therefore, $\Pr(n)$ can be approximated as

$$\Pr(n) \approx \frac{(1-q)^{F(n-1,k,L)}[1-(1-q)^{P(n,k,L)}]}{1-q^{\binom{L}{k}}}, \quad (6.16)$$

where the numerator gives the probability that all partitions of the integers smaller than n do not correspond to any valid minimum cut, while at least one of the partitions of n corresponds to a valid minimum cut. The denominator in (6.16) is a normalization factor such that $\sum_n \Pr(n) = 1$.

Let $J(n, k, L, i)$ be the number of partitions among the total $P(n, k, L)$ partitions of n that contain a part $i, i = 1, \dots, L$. The value of $J(n, k, L, i)$ can be computed recursively as

$$J(n, 2, L, i) = \begin{cases} 1, & i+1 \leq n \leq L+i, n \neq 2i \\ 0, & \text{otherwise} \end{cases},$$

$$J(n, k, L, i) = P(n-i, k-1, L) - J(n-i, k-1, L, i), \forall k > 2.$$

The probability that $\mathbf{Y}_i, i \in \{1, \dots, L\}$, is one of the variable returned by the $\min(\cdot)$ operator in (6.10), denoted as $\Pr(\mathbf{Y}_i)$, can then be computed from $\Pr(n)$ as

$$\Pr(\mathbf{Y}_i) = \sum_n \Pr(n) \frac{J(n, k, L, i)}{P(n, k, L)}. \quad (6.17)$$

Therefore, the expected throughput given in (6.10) can be evaluated as

$$\begin{aligned} \rho_M &= \frac{\sum_{i=1}^L \Pr(\mathbf{Y}_i) \mathbb{E}[\mathbf{Y}_i]}{M} \\ &\approx \sum_{i=1}^L \left[(1-p) + \sqrt{\frac{p(1-p)}{M}} \Phi^{-1} \left(\frac{i-0.375}{L+0.25} \right) \right] \Pr(\mathbf{Y}_i) \\ &= (1-p) \sum_{i=1}^L \Pr(\mathbf{Y}_i) + \underbrace{\sqrt{\frac{p(1-p)}{M}} \sum_{i=1}^L \Phi^{-1} \left(\frac{i-0.375}{L+0.25} \right) \Pr(\mathbf{Y}_i)}_{\alpha(k, N, L)} \\ &\stackrel{(b)}{=} k(1-p) + \alpha(k, N, L) \sqrt{\frac{p(1-p)}{M}}. \end{aligned} \quad (6.18)$$

where (b) is true because

$$\begin{aligned}
 \sum_{i=1}^L \Pr(\mathbf{Y}_i) &= \sum_{i=1}^L \sum_n \Pr(n) \frac{J(n, k, L, i)}{P(n, k, L)} \\
 &= \sum_n \frac{\Pr(n)}{P(n, k, L)} \sum_{i=1}^L J(n, k, L, i) \\
 &= \sum_n k \Pr(n) = k.
 \end{aligned}$$

Note that we have used the identity $\sum_j J(n, k, L, i) = kP(n, k, L)$, which is true because each partition of $P(n, k, L)$ contains k distinct parts, thus it is counted k times when we sum up $J(n, k, L, i)$ over all possible i . The network parameter $\alpha(k, N, L)$ in (6.18) is a *negative* number independent of M . Therefore, the obtained throughput given in (6.18) monotonically increases with M , as expected. Furthermore, when $M \rightarrow \infty$, $\rho_M \rightarrow \rho_\infty = k(1 - p)$, which is the capacity of the network as given in (6.6).

6.1.4 A Simple Lower Bound

Note that to evaluate the normalized throughput ρ_M in (6.18), the parameters k , L and N need to be determined. While k and L can be efficiently determined for arbitrary networks, counting the number of minimum cuts, i.e., determining N , has been shown to be #P-complete [116]. To reduce the computational complexity of the throughput evaluation for large networks, we derive a simple lower bound for ρ_M that is independent of N , as given by the following lemma.

Lemma 6.2. *Consider a directed acyclic network with the cardinality of minimum cuts being k and has L critical edges in total. Assume that each link has independent erasure p , the normalized throughput given in (6.10) is lowered bounded by*

$$\rho_M \geq \frac{\sum_{j=1}^k \mathbb{E}(\mathbf{Y}_j)}{M} = k(1 - p) + \sqrt{\frac{p(1 - p)}{M}} \sum_{j=1}^k \Phi^{-1} \left(\frac{j - 0.375}{L + 0.25} \right). \quad (6.19)$$

Proof. Since $\mathbf{Y}_j, j = 1, \dots, L$, is the j th smallest variable of $\{\mathbf{X}_{i_1}, \dots, \mathbf{X}_{i_L}\}$ and $|C^*| = k$, we have $\sum_{i: e_i \in C^*} \mathbf{X}_i \geq \sum_{j=1}^k \mathbf{Y}_j, \forall C^* \in \Psi^*$. Therefore, the normalized

throughput given in (6.10) is lower bounded as

$$\rho_M \geq \frac{\mathbb{E} \left[\sum_{j=1}^k \mathbf{Y}_j \right]}{M} = \frac{\sum_{j=1}^k \mathbb{E}(\mathbf{Y}_j)}{M}$$

□

6.1.5 Numerical Examples

In this section, numerical examples are provided to verify the analytical results in (6.18) and (6.19). We will first consider the simple butterfly network, followed by random networks generated based on Érdős-Rényi (ER) graphs.

6.1.5.1 Butterfly Network

Consider the communication between s and t_1 in the butterfly network shown in Figure 6.3(a), which has the parameters $(k, N, L) = (2, 6, 6)$. By substituting these values into (6.18), it can be obtained that $\alpha(k, N, L) = -1.3871$. Therefore, the normalized throughput as a function of M and p can be approximated by the following closed-form expression:

$$\rho_M \approx 2(1 - p) - 1.3871 \sqrt{\frac{p(1 - p)}{M}}. \quad (6.20)$$

To evaluate the accuracy of (6.20), we obtain the simulated throughput for different M and p values based on 10^4 independent realizations of $\{\mathbf{X}_i, i = 1, \dots, 9\}$, and the results are plotted in Figure 6.4(b). It is observed that the analytical results given in (6.20) match well with the simulation values, which thus validates the approximations made in Section 6.1.3.

It is also observed that when temporal network coding is applied, i.e., $M > 1$, substantial performance enhancement can be achieved, even with temporal coding length as small as $M = 2$. In fact, as observed in Figure 6.4(b), most of the performance gain due to temporal-coding can be achieved by coding with a small M . For example, it can be obtained from (6.20) that $M = 5, 20$ and 49 are sufficient to achieve 90% of the network capacity for $p = 0.1, 0.3$ and 0.5 , respectively.

FIGURE 6.4: Normalized throughput ρ_M versus network edge erasure probability p for various temporal coding length M .

6.1.5.2 Random Networks

Next, we verify the accuracy of our analytical results using random networks generated based on ER graphs. Figure 6.5 shows the simulation and analytical results versus p for different M values based on 10^4 ER graphs, each with 30 nodes and edge density 0.2. A good match between the simulation and analytical results are observed in Figure 6.5. Furthermore, it is observed that the lower bound obtained in (6.19) gives good estimations of the normalized throughput for large M , whereas it is loose for small M such as $M = 2$.

6.1.6 An Achievable Scheme

To further corroborate our analytical results given in the pervious sections, we will show that a practical network coding scheme called batched sparse codes (BATS codes) [57] achieves the throughput predicted by (6.18). A BATS code contains an outer code and an inner code in finite field, as shown in Figure 6.6. The outer code is an extension into the matrix form of the traditional fountain code [120]. Specifically, the source node first obtains a degree d by sampling a predefined degree distribution, and then randomly picks d distinct input packets to generate a batch of M fountain-coded packets. The batches are transmitted sequentially. The inner code of BATS employs RLNC at the intermediate nodes. Note that only packets

FIGURE 6.5: Normalized throughput ρ_M versus network edge erasure probability p for random networks.

within the same batch are coded together, which is equivalent to a spatial-temporal network code with temporal coding length M . Finally, the inner and outer codes are jointly decoded at the receiver with belief-propagation (BP) decoding [121].

FIGURE 6.6: The structure of BATS.

Similar as Fountain code, the performance of BATS code is largely dependent on the degree distribution. The degree distribution for BATS code has been optimized in [122] by assuming that the file size is sufficiently large, i.e., the total number of

packets to be transmitted is much larger than the batch size M . The result has been further extended for finite number of packets in [123]. Besides, by applying the precoding and inactivation decoding as that in Raptor Q codes [124], it is observed that BATS codes achieve the analytical throughput of erasure networks.

Example 6.2. *In this example, BATS code is applied on the butterfly network shown in Figure 6.3. A BATS code with batch size $M = 8$ is applied to send an image consisting of 2083 packets from s to both sink nodes t_1 and t_2 . The simulation results averaged over the two sink nodes are shown in Figure 6.7. It is observed that the throughput achieved by BATS code matches quite well with our analytical results in (6.20) for the butterfly network.*

FIGURE 6.7: Compare the achieved throughput of BATS code with the analytical throughput for $M = 8$ in the butterfly network.

6.1.6.1 Implementation Details

This subsection compares the experimental performance of BATS code with the derived theoretical throughput over a 3-node testbed based on IEEE802.11abg standard. The testbed consists of 3 laptops, which form a line network as shown in Figure 6.1. The source and the receiver are far apart, and hence they cannot hear each other. We use HP ProBook 430G1 for laptops which has inbuilt Wi-Fi. The

operating system used in the laptops is Ubuntu 13.10. The source and receivers are connected in Ad-Hoc mode. UDP is used as transport layer protocol.

A picture file of size 2.1MB is sent from the source to the receiver, via the relay node. The file is divided 2083 packets, each of 1000 bytes. The erasure probability is tuned by purposely dropping the packets at the relay and receiver nodes. The experimental results are shown in Figure 6.8, where the normalized throughput with BATS code is obtained by averaging over 10 runs. It is observed that the throughput achieved by BATS code matches perfectly with our analytical results for this two hop line network in (6.4).

FIGURE 6.8: Compare the achieved throughput of BATS code with the analytical throughput for various M in two-hop line network.

6.2 Multi-Source Single-Sink Erasure Networks

Multiple-source single-sink networks model the communication scenarios where all sources wish to transmit their respective information to the common sink node. It can also be used to model a special case of network computing in which all the information of disjoint sources is reproduced at the sink [125] so that some arbitrary functions can be computed, or a sensor/grid network where data from different sources are collected at a common node for some joint processing. For

deterministic multiple-source single-sink networks where all links are error-free, the capacity is given by the max-flow min-cut theorem, which is achieved by pure routing [126]. However, when the links are error-prone, network coding is necessary to achieve the network error/erasure correction capability [127]. In this section, we study the multi-source single-sink networks where each link is subject to erasure probability p . A RLNC approach with sufficiently large temporal coding length is proposed, which is proven to be capacity achieving. By relating the ranks of the transition matrices resulting from RLNC with the expected min-cuts of the network, the analytical throughput derived in the preceding section is shown to be applicable for multi-source single-sink erasure networks as well.

6.2.1 System Model

A multi-source single-sink erasure network can be modeled by a directed acyclic graph $G(V, E)$, as shown in Figure 6.9, where the source nodes $S = \{s_1, \dots, s_k\}$ want to send independent messages $\mathbf{d}_1, \dots, \mathbf{d}_k$ to the common sink node t . An edge e is assumed to have unit capacity subject to an erasure probability p , i.e., it is able to convey one packet per unit time with probability $(1 - p)$. Let $C(V') \subset E$ be a set of edges whose removal disconnects the set of nodes $V' \subsetneq V$ from the destination node t .

FIGURE 6.9: A multi-source single-sink network.

Theorem 6.3. [53] *Consider a multi-source single-sink erasure network $G(V, E)$ as shown in Figure 6.9. The capacity region of the network with erasure locations²*

²Erasure locations refers to the set of erased link at each time slot.

known at the destination is given by

$$\mathcal{R}(G) \triangleq \left\{ (R_i, s_i \in S) \mid 0 \leq \sum_{s_i \in V' \cap S} R_i \leq \min_{V' \subset (V \setminus d)} (1-p)|C(V')| \right\}. \quad (6.21)$$

(6.21) shows that the sum rate from source $S' \subset S$ to t should not exceed the capacity of any edge cut that disconnects S' and t .

6.2.2 A Random Linear Network Coding Approach

While the converse of Theorem 6.3 follows from the celebrated max-flow min-cut theorem, its proof of achievability in [53] is based on constructing the random code book with complex non-linear decoding. Specifically, with the erasure locations as side information, the sink node simulates the network with all possible input packets and compare them with the received information. Such decoding is not practical due to the large amount of possible input packets.

In this section, we propose a RLNC approach to achieve the capacity of multi-source single-sink networks given in Theorem 6.3. Specifically, with RLNC performed at all nodes, we show that the multi-source single-sink networks can be modeled by linear finite field deterministic multiple access channels where the sink node can decode all the desired messages by solving a set of linear equations.

Assume that there are m_i outgoing edges from source $s_i \in S$ and n incoming edges to the sink node t , i.e., $\text{Out}(s_i) = m_i, i = 1, \dots, k$ and $\text{In}(t) = n$. With RLNC performed at the intermediate nodes, the input-output relation between S and t of the multi-source single-sink erasure networks can be represented by

$$\mathbf{y} = \sum_{i=1}^k \mathbf{H}_i \mathbf{x}_i, \quad (6.22)$$

where $\mathbf{x}_i \in \mathbb{F}_q^{m_i}$ consists of symbols conveyed by the m_i outgoing edges of s_i ; $\mathbf{y} \in \mathbb{F}_q^n$ contains the symbols received by t ; and $\mathbf{H}_i \in \mathbb{F}_q^{n \times m_i}$ denotes the transition matrix from s_i to t , for $i = 1, \dots, k$, which is determined by the network topology, network coding coefficients and the erasure events.

Note that the mathematical relation given in (6.22) belongs to a class of multiple access channel (MAC) which has been extensively studied in wireless communications [128]. However, there are two main differences between the channel

model given in (6.22) and wireless MAC. First, the additive noise terms are absent in (6.22), which renders a finite field *deterministic* MAC. Second, in contrast to wireless channels where channel matrices are usually assumed to be generic and independent from each other, the transition matrices in (6.22) are correlated due to the network topology constraint.

From Theorem 3 in [41], when the finite field size q is sufficiently large, the ranks of the transition matrices equal to the min-cuts of the network. With channel extension over a sufficient number of time slots, the erasure network shown in Fig 6.9 can be viewed as a deterministic network where the normalized capacity of edge e is $(1 - p)$. Hence, it follows that the normalized expected ranks of the transition matrices are given by

$$\text{rank}(\mathbf{H}_i) = \min_{V': s_i \in V'} (1 - p)|C(V')|. \quad (6.23)$$

Further, denote by $\mathbf{H}_I, I \subset \{1, 2, \dots, k\}$, the concatenated matrix composing $\mathbf{H}_i, i \in I$. For instance, $\mathbf{H}_{\{1,2\}} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \end{bmatrix}$. Assume that there is an imaginary source node s'_I connecting to the set of sources $\{s_i, i \in I\}$ with edges of infinite capacity. Therefore the symbols can be forwarded from s'_I to $\{s_i, i \in I\}$ without coding. The resulting transition matrices from s'_I to t is then given by $\mathbf{H}_I$. Following similar arguments as for (6.23), we have

$$\text{rank}(\mathbf{H}_I) = \min_{V': \{s_i, i \in I\} \subset V'} (1 - p)|C(V')|, \forall I \subseteq \{1, 2, \dots, k\}. \quad (6.24)$$

Therefore, the capacity region of the multiple-access erasure networks given in (6.21) can be equivalently written as

$$\mathcal{R}(G) = \left\{ (R_i, s_i \in S) \mid 0 \leq \sum_{i \in I} R_i \leq \text{rank}(\mathbf{H}_I), \forall I \subset \{1, 2, \dots, k\} \right\}. \quad (6.25)$$

With RLNC applied at the source nodes, the information bearing symbols $\mathbf{d}_i$ are mapped to the source output $\mathbf{x}_i \in \mathbb{F}_q^{m_i}$ via a random matrix $\mathbf{E}_i \in \mathbb{F}_q^{m_i \times |\mathbf{d}_i|}$. Hence, t receives a linear combination of the information-bearing symbols as

$$\mathbf{y} = \sum_{i=1}^k \mathbf{H}_i \mathbf{E}_i \mathbf{d}_i, \quad (6.26)$$

which is shown in Figure 6.10.

FIGURE 6.10: The channel model of (6.26).

Let $\mathbf{G}_i = \mathbf{H}_i \mathbf{E}_i$ and denote by $\mathbf{G}_I$ a concatenated matrix composing of $\{\mathbf{G}_i, i \in I\}$. Clearly, t can solve all the desired messages $\{\mathbf{d}_i, i = 1, \dots, k\}$ from (6.26) if the transition matrix $\mathbf{G}_{\{1, \dots, k\}}$ is of full column rank, i.e., $\text{rank}(\mathbf{G}_{\{1, \dots, k\}}) = \sum_{i=1}^k |\mathbf{d}_i|$. The full rank condition is asymptotically satisfied, which is formally stated in the following proposition.

Proposition 6.4. *If $(|\mathbf{d}_i|, i = 1, 2, \dots, k)$ is within the capacity region given in (6.25), we have $\text{rank}(\mathbf{G}_{\{1, \dots, k\}}) = \sum_{i=1}^k |\mathbf{d}_i|$ with high probability when the finite field size $q \rightarrow \infty$.*

Proof. Proposition 6.4 can be proved by recursively applying Fact B.2 in Appendix B as follows.

For notational convenience, we define

$$\begin{aligned} \gamma_i &= \text{rank} \left(\begin{bmatrix} \mathbf{G}_{\{1, \dots, i\}} & \mathbf{H}_{\{i+1, \dots, k\}} \end{bmatrix} \right) \\ &= \text{rank} \left(\begin{bmatrix} \mathbf{H}_1 \mathbf{E}_1 & \cdots & \mathbf{H}_i \mathbf{E}_i & \mathbf{H}_{i+1} & \cdots & \mathbf{H}_k \end{bmatrix} \right). \end{aligned}$$

According to Fact B.2, as $q \rightarrow \infty$, the following result holds with high probability:

$$\begin{aligned} \gamma_1 &= \text{rank} \left(\begin{bmatrix} \mathbf{H}_1 \mathbf{E}_1 & \mathbf{H}_2 & \cdots & \mathbf{H}_k \end{bmatrix} \right) \\ &= \min \left\{ \text{rank}(\mathbf{H}_{\{1, \dots, k\}}), |\mathbf{d}_1| + \text{rank}(\mathbf{H}_{\{2, \dots, k\}}) \right\}. \end{aligned} \quad (6.27)$$

Similarly, γ_2 can be computed as

$$\begin{aligned}
\gamma_2 &= \text{rank} \left(\begin{bmatrix} \mathbf{H}_1 \mathbf{E}_1 & \mathbf{H}_2 \mathbf{E}_2 & \mathbf{H}_3 & \cdots & \mathbf{H}_k \end{bmatrix} \right) \\
&= \min \left\{ \gamma_1, |\mathbf{d}_2| + \text{rank} \left(\begin{bmatrix} \mathbf{H}_1 \mathbf{E}_1 & \mathbf{H}_3 & \cdots & \mathbf{H}_k \end{bmatrix} \right) \right\} \\
&= \min \left\{ \gamma_1, |\mathbf{d}_2| + \text{rank} (\mathbf{H}_{\{1,3,4,\dots,k\}}), |\mathbf{d}_1| + |\mathbf{d}_2| + \text{rank} (\mathbf{H}_{3,\dots,k}) \right\} \\
&\stackrel{(a)}{=} \min \left\{ \begin{aligned} &\text{rank}(\mathbf{H}_{\{1,\dots,k\}}), |\mathbf{d}_1| + \text{rank}(\mathbf{H}_{\{2,\dots,k\}}), |\mathbf{d}_2| + \text{rank} (\mathbf{H}_{\{1,3,4,\dots,k\}}), \\ &|\mathbf{d}_1| + |\mathbf{d}_2| + \text{rank} (\mathbf{H}_{\{3,\dots,k\}}) \end{aligned} \right\} \\
&= \min_{J \subseteq \{1,2\}} \left\{ \sum_{j \in J} |\mathbf{d}_j| + \text{rank} (\mathbf{H}_{\{1,\dots,k\} \setminus J}) \right\}, \tag{6.28}
\end{aligned}$$

where (a) follows by substituting γ_1 obtained in (6.27).

Repeating the above procedures renders

$$\begin{aligned}
\gamma_k &= \text{rank} (\mathbf{G}_{\{1,\dots,k\}}) \\
&= \min_{J \subseteq \{1,\dots,k\}} \left\{ \sum_{j \in J} |\mathbf{d}_j| + \text{rank} (\mathbf{H}_{\{1,\dots,k\} \setminus J}) \right\} \\
&= \sum_{i=1}^k |\mathbf{d}_i|, \tag{6.29}
\end{aligned}$$

where the last equality follows from the fact that $\{|\mathbf{d}_i|, i = 1, \dots, k\}$ is within the region specified in (6.25), hence $\sum_{i \in I} |\mathbf{d}_i| \leq \text{rank} (\mathbf{H}_I), \forall I \subseteq \{1, \dots, k\}$. $\square$

6.2.3 Throughput with Finite-Length Spatial-Temporal Network Coding

In the preceding subsection, we have shown that the capacity of multi-source single-sink erasure networks can be achieved by RLNC with sufficiently large temporal coding length. If the temporal coding length is limited to a finite value M , the achievable region presented in (6.25) deviate from the capacity region given in (6.21). Instead, the expected rank of the transition matrices after normalization is given by

$$\mathbb{E}[\text{rank}(\mathbf{H}_i)] = \frac{\mathbb{E}[\min_{C(s_i)} \sum_{e_j \in C(s_i)} \mathbf{X}_j]}{M}, \tag{6.30}$$

where $\mathbf{X}_j \sim \mathcal{B}(M, 1 - p)$.

Let k_I , N_I and L_I denote the min-cut value, the number of min-cuts and the number of critical edges between the set of source nodes $\{s_i, i \in I\}$ with the sink node t . A virtual super-source node s' connecting with $\{s_i, i \in I\}$ via links of infinite capacities can be added. Applying the results obtained in Section 6.1 for the communication between this virtual source node s' and the sink node t , a closed-form expression for $\mathbb{E}[\text{rank}(\mathbf{H}_I)]$ can be obtained as

$$\mathbb{E}[\text{rank}(\mathbf{H}_I)] \approx k_I(1-p) + \alpha(k_I, N_I, L_I) \sqrt{\frac{p(1-p)}{M}}, \quad (6.31)$$

where $\alpha(k_I, N_I, L_I)$ is defined in (6.18).

Therefore, with spatial-temporal network coding of length M , the expected throughput of the multi-source single-sink networks is specified by

$$\mathcal{R}_M(G) = \left\{ (R_i, s_i \in S) \mid 0 \leq \sum_{i \in I} R_i \leq k_I(1-p) + \alpha(k_I, N_I, L_I) \sqrt{\frac{p(1-p)}{M}}, \right. \\ \left. \forall I \subset \{1, 2, \dots, k\} \right\}. \quad (6.32)$$

Although the min-cuts between various source nodes and the sink node may have common edges and be correlated in statistics, evaluating the expectation of the min-cuts is the same as that for the point-to-point network considered in Section 6.1. Therefore, the results derived in (6.18) can be applied directly in (6.32).

Example 6.3. A simple 2-source single-sink erasure network is shown in Figure 6.11, where two source nodes s_1 and s_2 want to send independent information to the common sink node t via direct links or through the relay node r . Assume that each link is capable to convey one packet per time slot with packet loss rate $p = 0.2$. The capacity of this network is given by (6.21), which is

$$\begin{aligned} R_1 &\leq 2(1-p) = 1.6 \\ R_2 &\leq 2(1-p) = 1.6 \\ R_1 + R_2 &\leq 3(1-p) = 2.4. \end{aligned} \quad (6.33)$$

In Section 6.2.2, we have shown that RLNC over sufficiently large number of time slots is sufficient to achieve the capacity of this network. When the temporal coding length is limited to a finite value M , the throughput of this network can be

FIGURE 6.11: A 2-source single-sink erasure network.

evaluated from (6.32) with

$$\begin{aligned}
 (k_1, N_1, L_1) &= (2, 2, 3), \\
 (k_2, N_2, L_2) &= (2, 2, 3), \\
 (k_{\{1,2\}}, N_{\{1,2\}}, L_{\{1,2\}}) &= (3, 2, 3).
 \end{aligned}
 \tag{6.34}$$

The analytical throughput given in (6.32) is compared with the simulation results in Figure 6.12 for $M = 16$. It is observed that the analytical throughput matches well with the simulation result. A throughput degradation is observed for $M = 16$ compared with the network capacity, which is achieved when $M \rightarrow \infty$.

FIGURE 6.12: Compare the analytical throughput with the simulation results.

Example 6.4. Another 2-source single-sink erasure network constructed from the butterfly network is shown in Figure 6.13. We assume that each link in the network has unit capacity with erasure probability $p = 0.1$. The throughput of this network with temporal coding length $M = 8$ can be obtained from (6.32) and it is compared with the simulation result in Figure 6.14.

FIGURE 6.13: A 2-source single-sink erasure network.

6.3 Chapter Summary

In this chapter, the throughput of single-session erasure networks with finite-length spatial-temporal network coding has been analyzed. The derived analytical results are shown to match well with the simulation values. Our results show that given M and erasure probability p , the throughput is mainly determined by the min-cut value k , the number of minimum cuts N , and the number of critical edges L of the network. One practical application of the obtained results is to determine the minimum temporal coding length M needed to achieve a target throughput for a given erasure network. It is observed that BATS code achieves the analytical throughput for erasure networks with finite temporal coding length. Furthermore, we show that the capacity of multi-source single-sink erasure networks can be achieved by RLNC over sufficiently large number of time slots. By applying the analytical results derived in the first part of this chapter, the expected throughput of multi-source single-sink networks with finite-length spatial-temporal network coding can be obtained.

FIGURE 6.14: Compare the analytical throughput with the simulation results.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

In this thesis, we have studied the multi-session networks applied with network coding schemes. Specifically, network simplification techniques have been proposed to simplify the network coding capacity analysis, novel outer and inner network capacity bounds have been obtained, and the throughput of erasure networks with finite-length spatial-temporal network coding has been derived.

First, to simplify the analysis of network coding capacity, some systematic graph reduction techniques have been developed to reduce the order of the functional dependence graph (FDG) of a network. To this end, the “non-essential” edges in the network are identified and eliminated, which results in a hypernetwork whose corresponding FDG is of a smaller order than the original network. Besides, we show that certain edges in the FDG can be removed without affecting the functional dependence relationships of the random variables. Such edge removal creates new opportunities for reducing the order of the FDG. An iterative algorithm has been designed to achieve this by jointly applying the proposed reduction techniques. With the reduced FDGs, various capacity bounds of the network, such as the linear programming bound and functional dependence bound, can be computed much more efficiently without any loss in tightness. Moreover, we have shown that the FDG can be further reduced if only linear network coding capacity is of interest. Simulation results have shown that with the proposed techniques, substantial reduction can be obtained for FDG of the randomly generated networks, especially for those networks with low-to-moderate density. Specifically, on average, 65%, 30% and 18% of the

original FDG size are reduced for random networks with density 0.1, 0.2 and 0.3, respectively.

Based on the functional dependence relationship of the source/edge random variables in a network, we have derived a lower bound on the mutual information between any two subsets of source/edge random variables and proved that the proposed bound is the tightest bound for general network coding. Based on the proposed lower bound on mutual information, we have derived a new class of information-theoretic outer bounds for network coding capacity. The proposed outer bounds improve over the existing edge-cut bounds by including different rate weights. An algorithm has been designed to efficiently produce this new class of outer bounds for a network.

With random linear network coding (RLNC) performed at the intermediate nodes, we have proposed an optimal linear precoding and decoding scheme for the source and sink nodes of double-unicast networks. The proposed scheme utilizes Han-Kobayashi rate splitting and a concatenated precoding composed of zero-forcing and random block-diagonal precoding matrices. It has been shown to give the maximal achievable rate region for double-unicast networks for any given network coding coefficients used at the intermediate nodes. Furthermore, a suboptimal, but easy to compute, achievable region in terms of min-cuts of the network has also been derived, which is strictly larger than existing rate regions in the literature. Furthermore, since RLNC is generally a suboptimal strategy for the intermediate node in multi-session networks, we show that the achievable region can be further enlarged by applying a joint routing and RLNC scheme. With joint routing and RLNC, part of information is routed to the sink nodes without network coding to reduce the inter-session interference, and the rest is transmitted with RLNC.

Lastly, the throughput for erasure networks (networks with non-zero packet loss rate) with spatial-temporal network coding is derived as a function of the temporal coding length. The derived throughput formulas are shown to match well with the simulation results. One practical application of the obtained results is to determine the minimum temporal coding length needed to achieve a target throughput for a given erasure networks. The BATS code is found to be a spatial-temporal network code that achieves the derived throughput with finite temporal coding length. The analytical result is further extended for multi-source single-sink erasure networks with finite temporal network coding length. Furthermore, we show that RLNC with sufficiently large temporal coding length achieve the capacity of multi-source

single-sink erasure networks.

7.2 Future Work

As finding the network coding capacity for general multi-session networks is challenging, characterizing the outer/inner bounds is an effective way for studying network coding in multi-session networks. The refined FD bound presented in Chapter 4 may be extended to include more rate weights besides $\{0, 1, 2\}$ by considering more than two subsets of source/edge random variables together. However, in general, it is difficult to characterize a graph theoretic bound that outperforms the LP bound. Hence, an efficient algorithm for computing the LP bound needs to be designed. Since the variables and constraints for the LP bound have been minimized in general, further simplification is expected to be dependent on the network topology. With the reduction techniques introduced in Chapter 3, the FDG of a network can be reduced, but finding an efficient way to determine the minimum FDG order without compromising the tightness for the resulting bounds is still an open problem. It is promising to have some systematic method to develop the minimal FDG of a network, so that the complexity of LP bound computation can be further reduced.

If RLNC is applied at the intermediate nodes, multiple-unicast networks reduce to multi-user interference channels (ICs). As a result, many existing techniques on IC can be applied for the network code design. In Chapter 5 of this thesis, we have developed the optimal linear precoding and decoding schemes for double-unicast networks. However, extending the results to triple or higher order unicast networks is highly non-trivial. With a particular interference alignment (IA) scheme, called subspace alignment chain, we are able to design an optimal precoding scheme for 3-user rank-deficient multiple-input multiple-output (MIMO) IC [129]. To apply the scheme in [129] to 3-user unicast networks, we also need to consider the correlation between the channel matrices induced by the network topology.

Compared with deterministic networks, erasure networks actually benefit more from network coding since the temporal network coding is analogous to channel coding over packets. The derived throughput for erasure networks in Chapter 6 may be extended to a more general form by considering different erasure probabilities at different network edges, which is more realistic. The network model can be further extended to incorporate the broadcast nature of wireless communications. While the optimal throughput for general multi-session erasure networks is a long standing

open problem, it is promising to study the throughput for some special multi-session erasure networks with suboptimal network coding schemes, such as intra-session network coding.

Another possible work direction is to investigate the capacity region for a mobile network. Due to network dynamics, designing an efficient and robust routing algorithm is very difficult. In contrast, random linear network coding provides a simple and non-coherent solution to such networks, that is more robust to packet loss and link failures. It is interesting to analyze the throughput that can be achieved in a mobile network with random linear network coding. Furthermore, some scheduling algorithms may be jointly designed with random linear network coding to further improve its performance.

Author's Publications

Journal Papers

- [1] **X. Xu**, Y. Zeng, Y. L. Guan, and T. Ho “An Achievable Region for Double-Unicast Networks with Linear Network Coding,” *IEEE Transactions on Communications*, vol. 62, no. 10, pp. 3621 - 3630, Aug. 2014.
- [2] **X. Xu**, S. Thakor, and Y. L. Guan, “Reduced Functional Dependence Graphs,” *IET Networks*, vol. 3, no. 2, Jun. 2014.
- [3] **X. Xu**, Y. Zeng, and Y. L. Guan, “Performance Analysis of Finite-Length Spatial-Temporal Network Coding,” *IEEE Communications Letters*, vol. 18, no. 7, pp. 1163-1166, Jul. 2014.
- [4] Y. Zeng, **X. Xu**, Y. L. Guan, E. Gunawa, and C. Wang, “Degree of Freedom of 3-user Rank-Deficient MIMO Interference Channel,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 8, pp. 4179-4192, Aug. 2014.

Conference Papers

- [1] **X. Xu**, and Y. L. Guan, “GMD-based Erasure-and-Error Decoding of Lifted MRD Codes for Network Error Correction,” in *International Symposium on Network Coding (NetCod)*, (Beijing, China), Jul. 2011.
- [2] **X. Xu**, X. T. Xu, and Y. L. Guan, “Generalized Binned Linear Binary Codes as Dirty-Paper Codes for Digital Watermarking Application,” in *International Conference on Information, Communications, and Signal Processing (ICICS)*, (Singapore), Dec. 2011.
- [3] **X. Xu**, S. Thakor, and Y. L. Guan, “Reduced Functional Dependence Graph and its Applications,” *International Symposium on Network Coding (NetCod)*, (Boston, USA), Jun. 2012.

- [4] **X. Xu**, S. Thakor, and Y. L. Guan, "Weighted Sum-rate Functional Dependence Bound for Network Coding Capacity," in *International Symposium on Information Theory and its Applications (ISITA)* (Hawaii, USA), Oct., 2012.
- [5] **X. Xu**, Y. Zeng, and Y. L. Guan, "On the degrees of freedom of 3-user rank-deficient MIMO interference channel," in *Proc. IEEE 2013 International Conference on Communications (ICC'13)*, (Budapest, Hungary), Jun., 2013.
- [6] **X. Xu**, S. Thakor, Y. L. Guan, "On the Mutual Information between Random Variables in Networks", *Information Theory Workshop*, (Seville, Spain), Sep., 2013.
- [7] **X. Xu** and Y. L. Guan, "Joint Routing and Random Network Coding for Multi-session Networks", *IEEE International Conference on Networks*, (Singapore), Dec., 2013.

Bibliography

- [1] R. Koetter and M. Médard, “An algebraic approach to network coding,” *IEEE/ACM Trans. on Networking*, vol. 11, no. 5, pp. 2345–2364, 2003.
- [2] R. W. Yeung and Z. Zhang, “Distributed source coding for satellite communications,” *IEEE Trans. Inform. Theory*, vol. 45, no. 4, pp. 1111–1120, May 1999.
- [3] R. Ahlswede, N. Cai, S. Y. R. Li, and R. W. Yeung, “Network information flow,” *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.
- [4] A. Agarwal and M. Charikar, “On the advantage of network coding for improving network throughput,” in *Information Theory Workshop*, TX, USA, Oct. 2004, pp. 247 – 249.
- [5] C. Chekuri, C. Fragouli, and E. Soljanin, “On average throughput and alphabet size in network coding,” in *IEEE Inter. Symposium on Inform. Theory*, Adelaide, SA, Sep. 2005, pp. 1593 – 1597.
- [6] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, “XORs in the air: practical wireless network coding,” *IEEE/ACM Trans. On Networking*, vol. 16, no. 3, pp. 497–510, 2008.
- [7] D. S. Lun, M. Medard, and M. Effros, “On coding for reliable communication over packet networks,” in *Proc. 42nd Annual Allerton Conference on Commun., Control and Computing*, Sep. 2004.
- [8] M. Ghaderi, D. Towsley, and J. Kurose, “Reliability gain of network coding in lossy wireless networks,” in *IEEE Inter. Conf. on Computer Communications*, Phoenix, AZ, Apr. 2008.
- [9] R. Khalili, M. Ghaderi, J. Kurose, and D. Towsley, “On the performance of random linear network coding in relay networks,” in *IEEE Military Communications Conference*, San Diego, CA, Nov. 2008.

- [10] Y. Wu, P. A. Chou, and S. Y. Kung, "Minimum-energy multicast in mobile ad hoc networks using network coding," *IEEE Trans. Commun.*, vol. 53, no. 11, pp. 1906–1918, Nov. 2004.
- [11] C. Fragouli, J. Widmer, and J.-Y. L. Boudec, "A network coding approach to energy efficient broadcasting: from theory to practice," in *IEEE Inter. Conf. on Computer Communications*, Barcelona, Spain, Apr. 2006, pp. 1–11.
- [12] L. Lima, M. Medard, and J. Barros, "Random linear network coding: a free cipher?" in *IEEE Inter. Symp. On Inform. Theory*, Nice, Jun. 2007.
- [13] R. W. Yeung, S. Y. Li, N. Cai, and Z. Zhang, *Network Coding Theory*. Now Publisher, 2006.
- [14] C. Fragouli and E. Soljanin, *Network Coding Fundamentals*. Now Publisher, 2007.
- [15] ———, *Network Coding Applications*. Now Publisher, 2008.
- [16] T. Ho and D. S. Lun, *Network Coding: An Introduction*. Cambridge University Press, 2008.
- [17] 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.
- [18] L. Li, R. Ramjee, M. Buddhikot, and S. Miller, "Network coding-based broadcast in mobile ad-hoc networks," in *IEEE International Conference on Computer Communications (INFOCOM)*, Anchorage, AK, May 2007, pp. 1739 – 1747.
- [19] Y. E. Sagduyu and A. Ephremides, "On joint mac and network coding in wireless ad hoc networks," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3697 – 3713, Oct. 2007.
- [20] S. Omiwade, R. Zheng, and C. Hua, "Practical localized network coding in wireless mesh networks," in *IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, San Francisco, CA, Jun. 2008, pp. 332 – 340.

- [21] S. Katti, D. Katabi, H. Balakrishnan, and M. Medard, “Symbol-level network coding for wireless mesh networks,” in *Proceedings of the ACM SIGCOMM 2008*, SEATTLE, WA, Aug. 2008, pp. 401–412.
- [22] A. A. Hamra, C. Barakat, and T. Turletti, “Network coding for wireless mesh networks: a case study,” in *International Symposium on a World of Wireless, Mobile and Multimedia Networks*, Buffalo-Niagara Falls, NY, 2006, pp. 99 – 114.
- [23] H. Liu and Y. Gu, “TCP with hop-oriented network coding in multi-radio multi-channel wireless mesh networks,” *IET Networks*, vol. 1, no. 3, pp. 171 – 180, Sep. 2012.
- [24] C. Gkantsidis and P. Rodriguez, “Network coding for large scale content distribution,” in *INFOCOM*, Mar. 2005, pp. 2235 – 2245.
- [25] C. Gkantsidis, J. Miller, and P. Rodriguez, “Comprehensive view of a live network coding P2P system,” in *Association for Computing Machinery*, Oct. 2006.
- [26] ———, “Anatomy of a P2P content distribution system with network coding,” in *Proceedings of IPTPS*, Feb. 2006.
- [27] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, “Network coding for distributed storage systems,” *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4539–4551, Sep. 2010.
- [28] S.-Y. R. Li, W. Yeung, and N. Cai, “Linear network coding,” *IEEE Trans. Inform. Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.
- [29] S. Jaggi, P. Sanders, P. A. Chou, M. Effros, S. Egner, K. Jain, and L. M. G. M. Tolhuizen, “Polynomial time algorithms for multicast network code construction,” *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 1973 – 1982, May 2005.
- [30] X. Yan, R. W. Yeung, and Z. Zhang, “The capacity region for multi-source multi-sink network coding,” in *Proc. 2007 IEEE Int. Symp. Inform. Theory*, Nice, France, Jun. 2007, pp. 116–120.
- [31] T. H. Chan and A. Grant, “Dualities between entropy functions and network codes,” *IEEE Trans. Inform. Theory*, vol. 54, pp. 4470–4487, 2008.

- [32] F. Matus, “Infinitely many information inequalities,” in *Proc. 2007 IEEE Int. Symp. Inform. Theory*, Nice, France, Jun. 2007, pp. 41–44.
- [33] R. Dougherty, C. Freiling, , and K. Zeger, “Insufficiency of linear coding in network information flow,” *IEEE Trans. Inform. Theory*, vol. 51, no. 8, pp. 2745–2759, Aug. 2005.
- [34] R. W. Yeung, *Information Theory and Network Coding*. New York: Springer, 2008.
- [35] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.
- [36] X. Yan, J. Yang, and Z. Zhang, “An outer bound for multisource multisink network coding with minimum cost consideration,” *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2373–2385, Jun. 2006.
- [37] G. Kramer and S. Savari, “Edge-cut bounds on network coding rates,” *Journal of Netw. Syst. Manage.*, vol. 14, pp. 49–67, Mar. 2006.
- [38] S. Thakor, A. Grant, and T. Chan, “Network coding capacity: a functional dependence bound,” in *Proc. 2009 IEEE Int. Symp. Inform. Theory*, Seoul, Korea, Jun. 2009, pp. 263–267.
- [39] N. Harvey, R. Kleinberg, and A. Lehman, “On the capacity of information networks,” *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2345–2364, Jun. 2006.
- [40] S. U. Kamath, D. N. C. Tse, and V. Anantharam, “Generalized network sharing outer bound and the two-unicast problem,” in *Proc. of the IEEE International Symposium on Network Coding*, Beijing, China, Jul. 2011, pp. 1–6.
- [41] T. Ho, R. Koetter, M. Medard, M. Effros, J. Shi, and D. Karger, “A random linear network coding approach to multicast,” *IEEE Trans. Inform. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.
- [42] T. Ho and H. Viswanathan, “Dynamic algorithms for multicast with intra-session network coding,” *IEEE Trans. Inform. Theory*, vol. 55, no. 2, pp. 797 – 815, Feb. 2009.

- [43] N. Ratnakar, R. Koetter, and T. Ho, “Linear flow equations for network coding in the multiple unicast case,” in *Proc. DIMACS Working Group Network Coding*, Piscataway, NJ, Jan. 2005.
- [44] N. Ratnakar, D. Traskov, and R. Koetter, “Approaches to network coding for multiple unicasts,” in *Proc. 2006 International Zurich Seminar on Communications (IZS 2006)*, Zurich, Switzerland, Feb. 2006, pp. 70–73.
- [45] D. Traskov, N. Ratnakar, D. Lun, R. Koetter, and M. Medard, “Network coding for multiple unicasts: An approach based on linear optimization,” in *IEEE Inter. Symposium on Inform. Theory*, Seattle, WA, Jul. 2006, pp. 1758–1762.
- [46] C. C. Wang and N. B. Shroff, “Pairwise intersession network coding on directed networks,” *IEEE Trans. Inform. Theory*, vol. 56, no. 8, pp. 3879–3900, 2010.
- [47] M. Heindlmaier, D. S. Lun, D. Traskov, and M. Medard, “Wireless inter-session network coding - an approach using virtual multicasts,” in *IEEE International Conference on Communications (ICC)*, Kyoto, Japan, Jun. 2011, pp. 1–5.
- [48] E. Erez and M. Feder, “Capacity region and network codes for two receivers multicast with private and common data,” in *Proc. Workshop on Coding, Cryptography and Combinatorics*, Jun. 2003.
- [49] C. K. Ngai and R. Yeung, “Multisource network coding with two sinks,” in *2004 International Conference on Communications, Circuits and Systems*, 2004, pp. 34–37.
- [50] E. Erez and M. Feder, “Improving the multicommodity flow rate with network codes for two sources,” *IEEE J. Select. Areas Comm.*, vol. 27, no. 8, pp. 814–824, Jun. 2009.
- [51] A. Das, S. Vishwanath, S. Jafar, and A. Markopoulou, “Network coding for multiple unicasts: An interference alignment approach,” in *IEEE Inter. Symposium on Inform. Theory*, Austin, TX, Jun. 2010, pp. 1878–1882.
- [52] A. Ramakrishnan, A. Das, H. Maleki, and A. Markopoulou, “Network coding for three unicast sessions: Interference alignment approaches,” in *48th Annual*

- Allerton Conference on Commun., Control, and Computing*, Allerton, IL, Sep. 2010, pp. 1054 – 1061.
- [53] A. F. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros, “Capacity of wireless erasure networks,” *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 789–804, Mar. 2006.
- [54] R. Khalili and K. Salamatian, “On the capacity of erasure relay channel: Multi-relay case,” in *IEEE Inform. Theory Workshop*, Rotorua, New Zealand, Aug. 2005.
- [55] D. S. Lun, M. Medard, R. Koetter, and M. Effros, “Further results on coding for reliable communication over packet networks,” in *IEEE Int. Symp. Information Theory*, Adelaide, Australia, Sep. 2005, pp. 1848–1852.
- [56] D. Julian, “Erasure networks,” in *IEEE Inter. Symposium on Inform. Theory*, 2002.
- [57] S. Yang and R. W. Yeung, “Coding for a network coded fountain,” in *IEEE International Symposium on Information Theory*, St. Petersburg, Jul. 2011, pp. 2647–2651.
- [58] A. E. Gamal and M. H. M. Costa, “The capacity region of a class of deterministic interference channels,” *IEEE Trans. Inform. Theory*, vol. 28, no. 2, pp. 343–346, Mar. 1982.
- [59] S. Thakor, T. Chan, and A. Grant, “Bounds for network information flow with correlated sources,” in *Australia Communications Theory Workshop*, Melbourne, Australia, Jan. 2011, pp. 43–48.
- [60] —, “Characterising correlation via entropy functions,” in *Information Theory Workshop (ITW)*, Sevilla, Sep. 2013, pp. 1–2.
- [61] R. W. Yeung, “Multilevel diversity coding with distortion,” *IEEE Trans. Inform. Theory*, vol. 41, no. 2, pp. 412–422, Mar. 1995.
- [62] L. Song, R. W. Yeung, and N. Cai, “Zero-error network coding for acyclic networks,” *IEEE Trans. Inform. Theory*, vol. 49, no. 12, pp. 3129–3139, Dec. 2003.
- [63] G. Kramer, *Directed Information for Channels with Feedback*. Zurich: PhD thesis, Swiss Federal Institute of Technology, 1998.

- [64] O. Trullols-Cruces, J. M. Barcelo-Ordinas, and M. Fiore, “Exact decoding probability under random linear network coding,” *IEEE Communications Letters*, vol. 15, no. 1, pp. 67 – 69, Jan. 2011.
- [65] X. Guang and F. Fu, “The failure probability at sink node of random linear network coding,” in *IEEE International Conference on Information Theory and Information Security*, Beijing, China, Dec. 2010, pp. 876 – 879.
- [66] S. Saeedi, S. Diggavi, C. Fragouli, and V. Prabhakaran, “On degraded two message set broadcasting,” in *Information Theory Workshop*, Taormina, Oct. 2009, pp. 406–410.
- [67] S. Huang and A. Ramamoorthy, “An achievable region for the double unicast problem based on a minimum cut analysis,” *IEEE Trans. Communications*, vol. 61, no. 7, pp. 2890–2899, Jul. 2013.
- [68] S. Avestimehr, S. Diggavi, and D. Tse, “Wireless network information flow: A deterministic approach,” *IEEE Trans. Inform. Theory*, vol. 57, no. 4, pp. 1872–1905, Apr. 2011.
- [69] A. Newell, J. Dong, and C. Nita-Rotaru, “On the practicality of cryptographic defenses against pollution attacks in wireless network coding,” *ACM Computing Surveys*, vol. 45, no. 3, Jun. 2013.
- [70] M. Jafari, L. Keller, C. Fragouli, and K. Argyraki, “Compressed network coding vectors,” in *IEEE International Symposium on Information Theory*, Seoul, Dec. 2009, pp. 109 – 113.
- [71] S. Li and A. Ramamoorthy, “Improved compression of network coding vectors using erasure decoding and list decoding,” *IEEE Communications Letters*, vol. 14, no. 8, pp. 749 – 751, Aug. 2010.
- [72] R. Koetter, M. Effros, and M. Médard, “A theory of network equivalence - part i: Point-to-point channel,” *IEEE Trans. Inform. Theory*, vol. 57, no. 2, pp. 972–995, Feb. 2011.
- [73] —, “A theory of network equivalence - part ii: Multiterminal channels,” *ArXiv e-prints*, 2010.
- [74] T. Ho, M. Effros, and S. Jalali, “On equivalence between network topologies,” in *48th Annual Allerton Conference on Communication, Control and Computing*, Monticello, USA, Sep. 2010, pp. 391–398.

- [75] G. Gallo, G. Longo, S. Nguyen, and S. Pallottino, “Directed hypergraphs and applications,” *Discrete Applied Mathematics*, vol. 40, pp. 117–201, 1992.
- [76] C. W. Yu, “Computing subgraph probability of random geometric graphs with applications in quantitative analysis of ad hoc networks,” *IEEE J. Sel. A. Communi.*, vol. 27, no. 7, pp. 1056–1065, Aug 2009.
- [77] W. Song, K. Cai, R. Feng, and W. Rui, “Solving the two simple multicast network coding problem in time $o(|e|)$,” in *Proc. of IEEE 3rd International Conference on Communication Software and Networks*, Xi’an, China, May 2011, pp. 427–431.
- [78] R. Dougherty, C. Freiling, and K. Zeger, “Unachievability of network coding capacity,” *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2365–2372, Jun. 2006.
- [79] L. Lima, D. Ferreira, and J. Barros, “Topology matters in network coding,” *Telecommun. Syst.*, vol. 51, pp. 247–257, Mar. 2012.
- [80] M. E. J. Newman, “The structure and function of complex networks,” *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.
- [81] H. Kenniche, “Random geometric graphs as model of wireless sensor networks,” in *The 2nd International Conference on Computer and Automation Engineering (ICCAE)*, Singapore, Feb. 2010, pp. 103 – 107.
- [82] S. Thakor, A. Grant, and T. Chan, “On complexity reduction of the LP bound computation and related problems,” in *Proc. of the IEEE International Symposium on Network Coding*, Beijing, China, Jul. 2011, pp. 1–6.
- [83] A. T. Subramanian and A. Thangaraj, “Path gain algebraic formulation for the scalar linear network coding problem,” *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4520 – 4531, Sep. 2010.
- [84] G. Kramer, S. T. Yazdi, and S. A. Savari, “Refined edge-cut bounds for network coding,” in *Proc. of the Int. Zurich Seminar on Communications (IZS)*, Zurich, Switzerland, Mar. 2010.
- [85] A. Salimi, T. Liu, and S. Cui, “Generalized cut-set bounds for networks with colocated messages,” in *International Symposium on Network Coding (NetCod)*, Cambridge, MA, Jun. 2012.

- [86] X. Xu, S. Thakor, and Y. L. Guan, “Weighted sum-rate functional dependence bound for network coding capacity,” in *International Symposium on Information Theory and its Applications (ISITA)*, Hawaii, USA, Oct. 2012.
- [87] S. Thakor, “Characterization and computation of network coding capacity bounds,” 2010, available online at <http://www.itr.unisa.edu.au/research/thesis/st.pdf>.
- [88] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.
- [89] V. Prabhakaran, S. Diggavi, and D. Tse, “Broadcasting with degraded message sets: A deterministic approach,” in *Proc. of the 45th Annual Allerton Conf. on Commu., Control and Comput.*, 2007.
- [90] S. Kamath, D. Tse, and C. C. Wang, “Two-unicast is hard,” in *IEEE Inter. Symposium on Inform. Theory*, Honolulu, Hawaii, Jun. 2014.
- [91] C. K. Nagi and R. W. Yeung, “Network coding gain of combination networks,” in *Information Theory Workshop*, TX, USA, Oct. 2004, pp. 283 – 287.
- [92] C. C. Wang and N. B. Shroff, “Beyond the butterfly a graph-theoretic characterization of the feasibility of network coding with two simple unicast sessions,” in *IEEE Inter. Symposium on Inform. Theory.*, Nice, Jun. 2007, pp. 121–125.
- [93] Z. Li and B. Li, “Network coding: The case of multiple unicast sessions,” in *42th Annual Allerton Conference on Commu., Control, and Compu.*, Oct. 2004.
- [94] K. Cai, K. B. Letaief, P. Fan, and R. Feng, “On the solvability of 2-pair unicast networks - a cut-based characterization,” Jul. 2010, available online at <http://arxiv.org/abs/1007.0465>.
- [95] S. Shenvi and B. K. Dey, “A simple necessary and sufficient condition for the double unicast problem,” in *Proc. 2010 IEEE Int. Conference on Communications*, South Africa, May. 2010.
- [96] X. Xu, S. Thakor, and Y. L. Guan, “On the mutual information between random variables in networks,” in *IEEE Information Theory Workshop*, Seville, Spain, Sep. 2013.

- [97] S. Huang and A. Ramamoorthy, “An achievable region for the double unicast problem based on a minimum cut analysis,” in *Proc. Inform. Theory Workshop*, Paraty, Oct. 2011, pp. 120–124.
- [98] W. Zeng, V. Cadambe, and M. Medard, “An edge reduction lemma for linear network coding and an application to two-unicast networks,” in *50th Annual Allerton Conference on Commu., Control, and Compu.*, Oct. 2012.
- [99] T. Han and K. Kobayashi, “A new achievable region for the interference channel,” *IEEE Trans. Inform. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.
- [100] D. S. Bernstein, *Matrix Mathematics: Theory, Facts and Formulas*. Princeton University Press, 2009.
- [101] S. A. Jafar and M. J. Fakhreddin, “Degrees of freedom for the MIMO interference channel,” *IEEE Trans. Inform. Theory*, vol. 53, no. 7, pp. 2637–2642, Jul. 2007.
- [102] V. R. Cadambe and S. A. Jafar, “Interference alignment and degrees of freedom of the K -user interference channel,” *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.
- [103] S. Bidokhti and C. Fragouli, “Degraded two-message multicast over graphs,” in *Proc. 2011 IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, Jul. 2011, pp. 1703–1707.
- [104] B. Smith and S. Vishwanath, “Unicast transmission over multiple access erasure networks: capacity and duality,” in *IEEE Info. Theory Workshop*, Tahoe City, CA, Sep. 2007, pp. 331–336.
- [105] K. T. Gowda and S. Sumei, “Multicast capacity of multiple access erasure networks,” in *Inter. Symposium on Inform. Theory and Its Applications*, Auckland, Dec. 2008.
- [106] S. Y. R. Li, R. W. Yeung, and N. Cai, “Linear network coding,” *IEEE Trans. Inform. Theory*, vol. 49, no. 2, pp. 371–381, 2003.
- [107] M. Medard, M. Effros, T. Ho, and D. R. Karger, “On coding for non-multicast networks,” in *Proc. 41st Annu. Allerton Conf. Communication, Control and Computing*, Monticello, IL, Oct. 2013.

- [108] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *Allerton Conference on Communication, Control, and Computing*, Oct. 2003, invited paper. [Online]. Available: <http://research.microsoft.com/apps/pubs/default.aspx?id=78167>
- [109] M. Salmasi and S. J. Golestani, "Capacity of erasure networks in the presence of spatial network coding," in *IEEE Information Theory Workshop*, Tahoe City, CA, Sep. 2007, pp. 114 – 119.
- [110] F. Farnia and S. J. Golestani, "Asymptotic behavior of network capacity under spatial network coding," in *Wireless Communications and Networking Conference*, Shanghai, Apr. 2013, pp. 2434–2439.
- [111] H. Wang, P. Fan, and K. B. Letaief, "Maximum flow and network capacity of network coding for ad-hoc networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 12, pp. 4193–4198, Dec. 2007.
- [112] P. Maymounkov, N. J. A. Harvey, and D. S. Lun, "Methods for efficient network coding," in *Allerton Conference on Communication, Control, and Computing*, Sep. 2006.
- [113] D. Silva, W. Zeng, and F. R. Kschischang, "Sparse network coding with overlapping classes," in *International Symposium on Network Coding (NetCod)*, Lausanne, Jun. 2009, pp. 74–79.
- [114] A. Heidarzadeh and A. H. Banihashemi, "Overlapped chunked network coding," in *Information Theory Workshop (ITW)*, Cairo, Jan. 2010, pp. 1–5.
- [115] Y. Li, E. Soljanin, and P. Spasojevic, "Effects of the generation size and overlap on throughput and complexity in randomized linear network coding," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 1111 – 1123, Feb. 2011.
- [116] J. S. Provan and M. O. Ball, "The complexity of counting cuts and of computing the probability that a graph is connected," *SIAM Journal of Computing*, vol. 12, no. 4, pp. 777–778, 1983.
- [117] K. Cai and P. Y. Fan, "An algebraic approach to link failures based on network coding," *IEEE Trans. Inform. Theory*, vol. 53, no. 2, pp. 775–779, 2007.

- [118] J. P. Royston, “Expected normal order statistics (exact and approximation),” *Journal of the Royal Statistical Society (Applied Statistics)*, vol. 31, no. 2, pp. 161–165, 1982.
- [119] G. E. Andrews, *The theory of partitions*. Cambridge: Cambridge University Press, 1998.
- [120] M. Luby, “LT codes,” in *IEEE Symposium on Foundations of Computer Science*, 2002, pp. 271–280.
- [121] R. G. Gallager, *Low Density Parity Check Codes*. M.I.T. Press, 1963.
- [122] S. Yang and R. W. Yeung, “Batched sparse codes,” *IEEE Trans. Inform. Theory*, Jul. 2014.
- [123] T. C. Ng and S. Yang, “Finite-length analysis of bats codes,” in *International Symposium on Network Coding (NetCod)*, Calgary, AB, Jun. 2013, pp. 1–6.
- [124] A. Shokrollahi and M. Luby, *Raptor Codes*. Now Publishers Inc, 2011.
- [125] R. Appuswamy, M. Franceschetti, N. Karamchandani, and K. Zeger, “Network coding for computing: Cut-set bounds,” *IEEE Trans. Inform. Theory*, vol. 57, no. 2, pp. 1015–1030, Feb. 2011.
- [126] A. R. Lehman and E. Lehman, “Complexity classification of network information flow problems,” in *Proc. Symp. Discrete Algorithms*, New Orleans, LA, Dec. 2004, pp. 142–150.
- [127] T. K. Dikaliotis, T. Ho, S. Jaggi, S. Vyetrenko, H. Yao, M. Effros, J. Kliewer, and E. Erez, “Multiple-access network information-flow and correction codes,” *IEEE Trans. Inform. Theory*, vol. 57, no. 2, pp. 1067–1079, Feb. 2011.
- [128] A. E. Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2012.
- [129] Y. Zeng, X. Xu, Y. L. Guan, E. Gunawan, and C. Wang, “Degrees of freedom of 3-user rank-deficient MIMO interference channel,” *IEEE Trans. Wireless Communications*, 2014.
- [130] A. Orlitsky and J. R. Roche, “Coding for computing,” *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 903–917, Aug. 2001.

Appendix A

A Counter-Example of Corollary 3.10 for Nonlinear Network Coding

Consider the network shown in Fig. A.1, which is constructed from the network given in [33] by adding a few more edges and nodes, shown in dotted line and gray circles. The newly added node 49 wants to recover the message sent by the source node 2. All the edges in the network are of unit capacity. It is known that the original network without the newly added parts is solvable over an alphabet of cardinality 4, using the nonlinear network coding as labeled in the figure [33]. All messages, a, b and c , are represented by 2-bit binary words. The symbol $+$ indicates the addition in the ring $\mathbf{Z}_4$ of integers modulo 4, the symbol $\oplus$ indicates addition in the ring $\mathbf{Z}_2 \oplus \mathbf{Z}_2$ (i.e., bitwise XOR), and $t(x)$ indicates the result of exchanging the order of the bits in a 2-bit binary word.

Focus on the portion of the network shown in Fig. A.2(a), with the corresponding FDG shown in Fig. A.2(b), which can be reduced to Fig. A.2(c) if Corollary 3.10 holds for nonlinear network codes. However, we will show that this reduction is not valid by arguing that the network shown in Fig. A.2(c) is feasible, while the original FDG in Fig. A.2(b) is not feasible (The FDG is feasible if the message b can be recovered from U_8).

FIGURE A.1: An counter example showing that Corollary 3.10 cannot be extended to nonlinear network coding. All the edges are of unit capacity and the independent message a, b, c, d, e are available at the source nodes $1 - 3, 11 - 12$. The sink nodes $37 - 46, 49$ each demand one of the messages, as indicated.

It is obvious that the reduced FDG shown in Fig. A.2(c) is feasible as $b = a + b - [(a \oplus b \oplus c) \oplus (b \oplus c)]$. In the original FDG, U_1 and U_2 is not directly available for constructing U_8 . Instead, they must be compressed to U_5 first. Considering $X = (U_1, U_2) = (a \oplus b \oplus c, a + b)$, $Y = U_6 = (b \oplus c)$ and $Z = Y_b = b$, this network is equivalent to the classical problem “lossless source coding for computing with side information” shown in Fig. A.3, which was firstly studied in [130]. The construction of U_5 in Fig. A.2(b) is equivalent to the encoding process in Fig. A.3 and the construction of U_8 maps to the decoding process in Fig. A.3.

FIGURE A.2: (a) A portion of the network shown in Fig. A.1. (b) The conventional FDG. (c) The reduced FDG by Corollary 3.10.

FIGURE A.3: Coding for lossless computing with side information.

Denote the optimal coding rate for this problem by R^* . Since the variable b is a 2-bit binary word, the normalized rate is equal to $R^*/2$. With the method given in [130], we can find that $R^*/2 = 3/2 = 1.5$. Since the normalized optimal rate is beyond the edge capacity, we cannot construct suitable U_5 with $H(U_5) \leq 1$ to ensure perfect recovering of b at the sink. Therefore, the FDG shown in Fig. A.2(b) is not feasible.

Appendix B

Proof of Lemma 5.6

To prove Lemma 5.6, we need to first show the following results.

Fact B.1. *Let $\mathbf{G} \in \mathbb{F}_q^{l \times u}$ be a given matrix with $\text{rank}(\mathbf{G}) = r$, and $\mathbf{B} \in \mathbb{F}_q^{k \times l}$ be a random matrix whose entries are uniformly generated from $\mathbb{F}_q$. Then, we have $\text{rank}(\mathbf{BG}) = \min(k, r)$ with high probability as $q \rightarrow \infty$.*

Proof. Fact B.1 can be shown by separately considering the two cases with $k \leq r$ and $k > r$.

Case I: $k \leq r$. In this case, we can construct a submatrix $\mathbf{G}^s \in \mathbb{F}_q^{l \times k}$ from $\mathbf{G}$ by selecting its k independent columns. Then we have

$$\text{rank}(\mathbf{BG}) \geq \text{rank}(\mathbf{BG}^s) = k, \text{ as } q \rightarrow \infty, \quad (\text{B.1})$$

where the last equality can be shown by considering the determinant of the square matrix $\mathbf{BG}^s \in \mathbb{F}_q^{k \times k}$. Specifically, as $\mathbf{B}$ is randomly generated, the determinant of $\mathbf{BG}^s$ is a polynomial on the entries of $\mathbf{B}$. If this polynomial is not identically zero, then with high probability as $q \rightarrow \infty$, the determinant of $\mathbf{BG}^s$ is nonzero, and hence $\text{rank}(\mathbf{BG}^s) = k$. To show that the polynomial is not identically zero, it is sufficient to choose one particular matrix for $\mathbf{B}$, e.g., $\mathbf{B} = (\mathbf{G}^s)^T$, so that

$\text{rank}(\mathbf{B}\mathbf{G}^s) = \text{rank}((\mathbf{G}^s)^T \mathbf{G}^s) = k$. Therefore, the determinant of $\mathbf{B}\mathbf{G}^s$ is not identically zero and we have $\text{rank}(\mathbf{B}\mathbf{G}^s) = k$ with high probability as $q \rightarrow \infty$.

Besides, by considering the matrix dimensions of $\mathbf{B}\mathbf{G}$, we obviously have $\text{rank}(\mathbf{B}\mathbf{G}) \leq k$. Together with (B.1), we have the desired result: $\text{rank}(\mathbf{B}\mathbf{G}) = k = \min\{k, r\}$ for $k \leq r$.

Case 2: $k > r$. In this case, $\mathbf{B}$ can be decomposed as $\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix}$, where $\mathbf{B}_1 \in \mathbb{F}_q^{r \times l}$ and $\mathbf{B}_2 \in \mathbb{F}_q^{(k-r) \times l}$. Following similar arguments as those in Case 1, we can show that $\text{rank}(\mathbf{B}\mathbf{G}) \geq \text{rank}(\mathbf{B}_1 \mathbf{G}) = r$. Together with the fact that $\text{rank}(\mathbf{B}\mathbf{G}) \leq \text{rank}(\mathbf{G}) = r$, we have $\text{rank}(\mathbf{B}\mathbf{G}) = \min\{k, r\}$ for $k > r$ as well.

This completes the proof of Fact B.1. $\square$

Fact B.2. Let $\mathbf{G}_1 \in \mathbb{F}_q^{l \times u_1}$ and $\mathbf{G}_2 \in \mathbb{F}_q^{l \times u_2}$ be two given matrices with $\text{rank}(\mathbf{G}_1) = r_1$, $\text{rank}(\mathbf{G}_2) = r_2$ and $\text{rank}\left(\begin{bmatrix} \mathbf{G}_1 & \mathbf{G}_2 \end{bmatrix}\right) = r_{12}$. Let $\mathbf{B} \in \mathbb{F}_q^{u_1 \times k}$, $k \leq r_1$, be a random matrix whose entries are uniformly generated from $\mathbb{F}_q$. Then we have $\text{rank}\left(\begin{bmatrix} \mathbf{G}_1 \mathbf{B} & \mathbf{G}_2 \end{bmatrix}\right) = \min\{r_{12}, k + r_2\}$ with high probability as $q \rightarrow \infty$.

Proof. Fact B.2 can be shown by separately considering the two cases with $k \leq r_{12} - r_2$ and $k > r_{12} - r_2$.

Case 1: $k \leq r_{12} - r_2$. Since $\text{rank}\left(\begin{bmatrix} \mathbf{G}_1 \mathbf{B} & \mathbf{G}_2 \end{bmatrix}\right) \leq \text{rank}(\mathbf{G}_1 \mathbf{B}) + \text{rank}(\mathbf{G}_2) = k + r_2$, we only need to show that $\text{rank}\left(\begin{bmatrix} \mathbf{G}_1 \mathbf{B} & \mathbf{G}_2 \end{bmatrix}\right) \geq k + r_2$. To this end, let $\mathbf{G}_2^s$ be a submatrix of $\mathbf{G}_2$ consisting of its r_2 independent columns. Since $\text{rank}(\mathbf{G}_2) = r_2$, the columns in $\mathbf{G}_2^s$ span the same subspace as those in $\mathbf{G}_2$. Thus, we have $\text{rank}\left(\begin{bmatrix} \mathbf{G}_1 & \mathbf{G}_2^s \end{bmatrix}\right) = \text{rank}\left(\begin{bmatrix} \mathbf{G}_1 & \mathbf{G}_2 \end{bmatrix}\right) = r_{12}$. Let $\mathbf{U}$ be a random matrix uniformly generated from $\mathbb{F}_q^{(k+r_2) \times l}$. According to Fact B.1, we have

$$\text{rank}\left(\mathbf{U} \begin{bmatrix} \mathbf{G}_1 & \mathbf{G}_2^s \end{bmatrix}\right) = \min\{k + r_2, r_{12}\} = k + r_2. \quad (\text{B.2})$$

Let $\tilde{\mathbf{G}}_1 = \mathbf{U}\mathbf{G}_1$ and $\tilde{\mathbf{G}}_2^s = \mathbf{U}\mathbf{G}_2^s$. Then we have

$$\begin{aligned} \text{rank} \left(\begin{bmatrix} \mathbf{G}_1\mathbf{B} & \mathbf{G}_2 \end{bmatrix} \right) &\geq \text{rank} \left(\begin{bmatrix} \mathbf{G}_1\mathbf{B} & \mathbf{G}_2^s \end{bmatrix} \right) \\ &\geq \text{rank} \left(\mathbf{U} \begin{bmatrix} \mathbf{G}_1\mathbf{B} & \mathbf{G}_2^s \end{bmatrix} \right) \\ &= \text{rank} \left(\begin{bmatrix} \tilde{\mathbf{G}}_1\mathbf{B} & \tilde{\mathbf{G}}_2^s \end{bmatrix} \right) = k + r_2, \end{aligned} \quad (\text{B.3})$$

where the last equality can be shown by verifying that the determinant of the square matrix $\begin{bmatrix} \tilde{\mathbf{G}}_1\mathbf{B} & \tilde{\mathbf{G}}_2^s \end{bmatrix}$ as a polynomial of entries in $\mathbf{B}$ is not identically zero. To this end, we can choose one particular $\mathbf{B}$ such that $\text{rank} \left(\begin{bmatrix} \tilde{\mathbf{G}}_1\mathbf{B} & \tilde{\mathbf{G}}_2^s \end{bmatrix} \right) = k + r_2$. According to (B.2), there are at least k independent columns in $\tilde{\mathbf{G}}_1$ that are not in the subspace spanned by $\tilde{\mathbf{G}}_2^s$. Denote the indices of these columns by $i_1, i_2, \dots, i_k$. If $\mathbf{B}$ is chosen such that the j th column is a unit vector with non-zero value only at the i_j th entry, $j = 1, \dots, k$, then $\tilde{\mathbf{G}}_1\mathbf{B}$ is a submatrix of $\tilde{\mathbf{G}}_1$ with columns not in the subspace spanned by $\tilde{\mathbf{G}}_2^s$; hence $\begin{bmatrix} \tilde{\mathbf{G}}_1\mathbf{B} & \tilde{\mathbf{G}}_2^s \end{bmatrix}$ is a full-rank matrix. Therefore, we can conclude that the determinant of $\begin{bmatrix} \tilde{\mathbf{G}}_1\mathbf{B} & \tilde{\mathbf{G}}_2^s \end{bmatrix}$ is not an identically zero polynomial and $\text{rank} \left(\begin{bmatrix} \tilde{\mathbf{G}}_1\mathbf{B} & \tilde{\mathbf{G}}_2^s \end{bmatrix} \right) = k + r_2$ with high probability as $q \rightarrow \infty$.

Case 2: $k > r_{12} - r_2$. Since $\text{rank} \left(\begin{bmatrix} \mathbf{G}_1\mathbf{B} & \mathbf{G}_2 \end{bmatrix} \right) \leq \text{rank} \left(\begin{bmatrix} \mathbf{G}_1 & \mathbf{G}_2 \end{bmatrix} \right) = r_{12}$, we only need to show that $\text{rank} \left(\begin{bmatrix} \mathbf{G}_1\mathbf{B} & \mathbf{G}_2 \end{bmatrix} \right) \geq r_{12}$. Let $\mathbf{V} \in \mathbb{F}_q^{(r_{12} \times l)}$ be a random matrix with entries uniformly generated from $\mathbb{F}_q$, and let $\bar{\mathbf{G}}_1 = \mathbf{V}\mathbf{G}_1$ and $\bar{\mathbf{G}}_2 = \mathbf{V}\mathbf{G}_2$. According to Fact B.1, we have $\text{rank}(\bar{\mathbf{G}}_1) = r_1$, $\text{rank}(\bar{\mathbf{G}}_2) = r_2$, and $\text{rank} \left(\begin{bmatrix} \bar{\mathbf{G}}_1 & \bar{\mathbf{G}}_2 \end{bmatrix} \right) = r_{12}$. Therefore, we can construct a matrix $\mathbf{G}_2^{s1}$ by choosing $(r_{12} - r_1)$ independent columns from $\bar{\mathbf{G}}_2$ that are not in the subspace spanned by $\bar{\mathbf{G}}_1$. Furthermore, let $\mathbf{G}_2^{s2}$ be a submatrix of $\bar{\mathbf{G}}_2$ consisting of $(r_1 - k)$ columns that are not in the subspace spanned by $\mathbf{G}_2^{s1}$. Such a choice is feasible since $(r_{12} - r_1) + (r_1 - k) = r_{12} - k \leq r_2$. Then, it follows that $\text{rank} \left(\begin{bmatrix} \bar{\mathbf{G}}_1 & \bar{\mathbf{G}}_2^{s1} & \bar{\mathbf{G}}_2^{s2} \end{bmatrix} \right) = r_{12}$. Following similar arguments as those for case 1, we can show that $\begin{bmatrix} \tilde{\mathbf{G}}_1\mathbf{B} & \tilde{\mathbf{G}}_2^{s1} & \tilde{\mathbf{G}}_2^{s2} \end{bmatrix}$ is a full-rank

square matrix with high probability as $q \rightarrow \infty$. Hence, we have the desired result:

$$\text{rank} \left(\begin{bmatrix} \mathbf{G}_1 \mathbf{B} & \mathbf{G}_2 \end{bmatrix} \right) \geq \text{rank} \left(\begin{bmatrix} \bar{\mathbf{G}}_1 \mathbf{B} & \bar{\mathbf{G}}_2 \end{bmatrix} \right) \geq \text{rank} \left(\begin{bmatrix} \bar{\mathbf{G}}_1 \mathbf{B} & \bar{\mathbf{G}}_2^{s_1} & \bar{\mathbf{G}}_2^{s_2} \end{bmatrix} \right) = r_{12}. \quad (\text{B.4})$$

This completes the proof of Fact B.2. $\square$

Now, we can prove Lemma 5.6 by recursively applying Fact B.2 as follows:

$$\begin{aligned} \gamma_1 &\triangleq \text{rank} \left(\begin{bmatrix} \mathbf{A}_1 \mathbf{E}_1 & \mathbf{A}_2 & \mathbf{A}_3 \end{bmatrix} \right) \\ &= \min \left\{ \text{rank}([\mathbf{A}_1 \quad \mathbf{A}_2 \quad \mathbf{A}_3]), k_1 + \text{rank}([\mathbf{A}_2 \quad \mathbf{A}_3]) \right\}. \end{aligned} \quad (\text{B.5})$$

$$\begin{aligned} \gamma_2 &\triangleq \text{rank} \left(\begin{bmatrix} \mathbf{A}_1 \mathbf{E}_1 & \mathbf{A}_2 \mathbf{E}_2 & \mathbf{A}_3 \end{bmatrix} \right) \\ &= \min \left\{ \gamma_1, k_2 + \text{rank}([\mathbf{A}_1 \mathbf{E}_1 \quad \mathbf{A}_3]) \right\} \\ &= \min \left\{ \gamma_1, k_2 + \text{rank}([\mathbf{A}_1 \quad \mathbf{A}_3]), k_1 + k_2 + \text{rank}(\mathbf{A}_3) \right\} \\ &= \min \left\{ \text{rank}([\mathbf{A}_1 \quad \mathbf{A}_2 \quad \mathbf{A}_3]), k_1 + \text{rank}([\mathbf{A}_2 \quad \mathbf{A}_3]), k_2 + \text{rank}([\mathbf{A}_1 \quad \mathbf{A}_3]), k_1 + k_2 + \text{rank}(\mathbf{A}_3) \right\} \end{aligned} \quad (\text{B.6})$$

$$\begin{aligned} \gamma_3 &\triangleq \text{rank}([\mathbf{A}_1 \mathbf{E}_1 \quad \mathbf{A}_2 \mathbf{E}_2 \quad \mathbf{A}_3 \mathbf{E}_3]) \\ &= \min \left\{ \gamma_2, k_3 + \text{rank}([\mathbf{A}_1 \mathbf{E}_1 \quad \mathbf{A}_2 \mathbf{E}_2]) \right\} \\ &= \min \left\{ \gamma_2, k_3 + \text{rank}([\mathbf{A}_1 \mathbf{E}_1 \quad \mathbf{A}_2]), k_2 + k_3 + \text{rank}([\mathbf{A}_1 \mathbf{E}_1]) \right\} \\ &= \min \left\{ \gamma_2, k_3 + \text{rank}([\mathbf{A}_1 \quad \mathbf{A}_2]), k_1 + k_3 + \text{rank}(\mathbf{A}_2), k_2 + k_3 + \text{rank}(\mathbf{A}_1), k_1 + k_2 + k_3 \right\} \\ &= \min \left\{ \begin{array}{l} \text{rank}([\mathbf{A}_1 \quad \mathbf{A}_2 \quad \mathbf{A}_3]), k_1 + \text{rank}([\mathbf{A}_2 \quad \mathbf{A}_3]), k_2 + \text{rank}([\mathbf{A}_1 \quad \mathbf{A}_3]), k_3 + \text{rank}([\mathbf{A}_1 \quad \mathbf{A}_2]), \\ k_1 + k_2 + \text{rank}(\mathbf{A}_3), k_1 + k_3 + \text{rank}(\mathbf{A}_2), k_2 + k_3 + \text{rank}(\mathbf{A}_1), k_1 + k_2 + k_3 \end{array} \right\} \\ &= k_1 + k_2 + k_3, \end{aligned} \quad (\text{B.7})$$

where the last equality in (B.7) follows from the conditions specified in Lemma 5.6.

This completes the proof of Lemma 5.6.

Appendix C

Detailed Steps of the Fourier-Motzkin Elimination

This appendix provides the detailed steps for obtaining the achievable region of the rate pair (R_1, R_2) via the obtained results on $(R_{1c}, R_{1p}, R_{2c}, R_{2p})$ presented in Chapter 5. By combining (5.31)-(5.37) and (5.40)-(5.46), we know that the rate pair $(R_{1c}, R_{1p}, R_{2c}, R_{2p})$ is achievable if it satisfies following inequalities.

$$R_{1c} \leq \text{rank}(\mathbf{H}_{11} \mathbf{V}_{21}) \quad (\text{C.1})$$

$$R_{1p} \leq k_{1-12} - k_{1-2} \quad (\text{C.2})$$

$$R_{2c} \leq k_{2-1} \quad (\text{C.3})$$

$$R_{1c} + R_{1p} \leq k_{1-1} \quad (\text{C.4})$$

$$R_{1c} + R_{2c} \leq k_{2-1} + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21}) \quad (\text{C.5})$$

$$R_{1p} + R_{2c} \leq k_{2-1} + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) \quad (\text{C.6})$$

$$R_{1c} + R_{1p} + R_{2c} \leq k_{12-1} \quad (\text{C.7})$$

$$R_{2c} \leq \text{rank}(\mathbf{H}_{22} \mathbf{V}_{11}) \quad (\text{C.8})$$

$$R_{2p} \leq k_{2-12} - k_{2-1} \quad (\text{C.9})$$

$$R_{1c} \leq k_{1-2} \quad (\text{C.10})$$

$$R_{2c} + R_{2p} \leq k_{2-2} \quad (\text{C.11})$$

$$R_{2c} + R_{1c} \leq k_{1-2} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{11}) \quad (\text{C.12})$$

$$R_{2p} + R_{1c} \leq k_{1-2} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) \quad (\text{C.13})$$

$$R_{2c} + R_{2p} + R_{1c} \leq k_{12-2}. \quad (\text{C.14})$$

Moreover, we also have non-negative requirements on the rate, i.e.,

$$R_{1c} \geq 0 \quad (\text{C.15})$$

$$R_{1p} \geq 0 \quad (\text{C.16})$$

$$R_{2c} \geq 0 \quad (\text{C.17})$$

$$R_{2p} \geq 0 \quad (\text{C.18})$$

Note that (C.3) is implied by (C.8) as $\text{rank}(\mathbf{H}_{22} \mathbf{V}_{11}) \leq \text{rank}(\mathbf{V}_{11}) = k_{2-1}$. Similarly, (C.10) is implied by (C.1). Removing the redundant inequalities and substituting with $R_{1p} = R_1 - R_{1c}$, $R_{2p} = R_2 - R_{2c}$, we obtain

$$R_{1c} \leq \text{rank}(\mathbf{H}_{11} \mathbf{V}_{21}) \quad (\text{C.19})$$

$$R_1 - R_{1c} \leq k_{1-12} - k_{1-2} \quad (\text{C.20})$$

$$R_1 \leq k_{1-1} \quad (\text{C.21})$$

$$R_{1c} + R_{2c} \leq k_{2-1} + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21}) \quad (\text{C.22})$$

$$R_1 - R_{1c} + R_{2c} \leq k_{2-1} + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) \quad (\text{C.23})$$

$$R_1 + R_{2c} \leq k_{12-1} \quad (\text{C.24})$$

$$R_{2c} \leq \text{rank}(\mathbf{H}_{22} \mathbf{V}_{11}) \quad (\text{C.25})$$

$$R_2 - R_{2c} \leq k_{2-12} - k_{2-1} \quad (\text{C.26})$$

$$R_2 \leq k_{2-2} \quad (\text{C.27})$$

$$R_{2c} + R_{1c} \leq k_{1-2} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{11}) \quad (\text{C.28})$$

$$R_2 - R_{2c} + R_{1c} \leq k_{1-2} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) \quad (\text{C.29})$$

$$R_2 + R_{1c} \leq k_{12-2} \quad (\text{C.30})$$

$$R_{1c} \geq 0 \quad (\text{C.31})$$

$$R_{1c} \leq R_1 \quad (\text{C.32})$$

$$R_{2c} \geq 0 \quad (\text{C.33})$$

$$R_{2c} \leq R_2 \quad (\text{C.34})$$

To eliminate R_{1c} , we get six upper bounds and three lower bounds on R_{1c}

$$R_{1c} \leq R_1 \quad (\text{C.35})$$

$$R_{1c} \leq \text{rank}(\mathbf{H}_{11} \mathbf{V}_{21}) \quad (\text{C.36})$$

$$R_{1c} \leq k_{2-1} + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21}) - R_{2c} \quad (\text{C.37})$$

$$R_{1c} \leq k_{1-2} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{11}) - R_{2c} \quad (\text{C.38})$$

$$R_{1c} \leq k_{1-2} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) - R_2 + R_{2c} \quad (\text{C.39})$$

$$R_{1c} \leq k_{12-2} - R_2 \quad (\text{C.40})$$

$$R_{1c} \geq 0 \quad (\text{C.41})$$

$$R_{1c} \geq R_1 - k_{1-12} + k_{1-2} \quad (\text{C.42})$$

$$R_{1c} \geq R_1 + R_{2c} - k_{2-1} - \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) \quad (\text{C.43})$$

Comparing the upper and lower bounds of R_{1c} gives 18 inequalities. Removing those inequalities that do not involve any variables and copying seven inequalities in (C.19)-(C.30) that do not involve R_{1c} , we obtain a new system of inequalities in

(R_{2c}, R_1, R_2) give by

$$R_1 \geq 0 \quad (\text{C.44})$$

$$R_1 \leq k_{1-1} \quad (\text{C.45})$$

$$R_1 \leq \text{rank}(\mathbf{H}_{11} \mathbf{V}_{21}) + k_{1-12} - k_{1-2} \quad (\text{C.46})$$

$$R_2 \leq k_{12-2} \quad (\text{C.47})$$

$$R_2 \leq k_{2-2} \quad (\text{C.48})$$

$$R_1 + R_2 \leq k_{12-2} + k_{1-12} - k_{1-2} \quad (\text{C.49})$$

$$R_1 + R_2 \leq \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) + k_{2-1} + k_{1-2} \quad (\text{C.50})$$

$$R_{2c} \geq 0 \quad (\text{C.51})$$

$$R_{2c} \leq R_2 \quad (\text{C.52})$$

$$R_{2c} \leq \text{rank}(\mathbf{H}_{22} \mathbf{V}_{11}) \quad (\text{C.53})$$

$$R_{2c} \leq k_{2-1} + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21}) \quad (\text{C.54})$$

$$R_{2c} \leq k_{1-2} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{11}) \quad (\text{C.55})$$

$$R_{2c} \leq k_{2-1} + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) \quad (\text{C.56})$$

$$R_2 - R_{2c} \leq k_{1-2} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) \quad (\text{C.57})$$

$$R_2 - R_{2c} \leq k_{2-12} - k_{2-1} \quad (\text{C.58})$$

$$R_1 + R_{2c} \leq k_{12-1} \quad (\text{C.59})$$

$$R_1 + R_{2c} \leq k_{2-1} + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21}) + k_{1-12} - k_{1-2} \quad (\text{C.60})$$

$$R_1 + R_{2c} \leq \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{11}) + k_{1-12} \quad (\text{C.61})$$

$$R_1 + R_{2c} \leq \text{rank}(\mathbf{H}_{11} \mathbf{V}_{21}) + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) + k_{2-1} \quad (\text{C.62})$$

$$R_1 + 2R_{2c} \leq 2k_{2-1} + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21}) + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) \quad (\text{C.63})$$

$$R_1 + 2R_{2c} \leq k_{1-2} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{11}) + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) + k_{2-1} \quad (\text{C.64})$$

$$R_1 + R_2 - R_{2c} \leq \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) + k_{1-12} \quad (\text{C.65})$$

$$R_1 + R_2 + R_{2c} \leq k_{12-2} + k_{2-1} + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) \quad (\text{C.66})$$

To identify the redundant inequalities in (C.44)-(C.66), we establish some useful facts in Section C.1. Therefore, we have:

- (C.46) is implied by (C.45) as $\text{rank}(\mathbf{H}_{11}\mathbf{V}_{21}) \geq k_{1-1} + k_{1-2} - k_{1-12}$;
- (C.47) is implied by (C.48) as $k_{12-2} \geq k_{2-2}$;
- (C.54) and (C.56) are implied by (C.53) as $\text{rank}(\mathbf{H}_{22}\mathbf{V}_{11}) \leq \text{rank}(\mathbf{V}_{11}) = k_{2-1}$;
- (C.55) is implied by (C.53) as $k_{1-2} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{11}) \geq k_{1-2} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22}) + \text{rank}(\mathbf{H}_{22}\mathbf{V}_{11}) - \text{rank}(\mathbf{H}_{22}) = \text{rank}(\mathbf{H}_{22}\mathbf{V}_{11}) + k_{12-2} - k_{2-2} \geq \text{rank}(\mathbf{H}_{22}\mathbf{V}_{11})$;
- (C.57) is implied by (C.58) as $k_{1-2} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) \geq k_{12-2} + k_{2-12} - k_{2-2} - k_{2-1} \geq k_{2-12} - k_{2-1}$;
- (C.60) is implied by (C.59) as $k_{2-1} + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21}) + k_{1-12} - k_{1-2} \geq k_{12-1}$;
- (C.62) is also implied by (C.59) as $\text{rank}(\mathbf{H}_{11}\mathbf{V}_{21}) + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) + k_{2-1} \geq k_{12-1}$.

Now, we are ready to eliminate R_{2c} from the remaining non-redundant inequal-

ities. In total, 7 upper bounds and 3 lower bounds are obtained for R_{2c} .

$$R_{2c} \leq R_2 \quad (\text{C.67})$$

$$R_{2c} \leq \text{rank}(\mathbf{H}_{22} \mathbf{V}_{11}) \quad (\text{C.68})$$

$$R_{2c} \leq k_{12-1} - R_1 \quad (\text{C.69})$$

$$R_{2c} \leq \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{11}) + k_{1-12} - R_1 \quad (\text{C.70})$$

$$2R_{2c} \leq k_{1-2} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{11}) + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) + k_{2-1} - R_1 \quad (\text{C.71})$$

$$2R_{2c} \leq 2k_{2-1} + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21}) + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) - R_1 \quad (\text{C.72})$$

$$R_{2c} \leq k_{12-2} + k_{2-1} + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) - R_1 - R_2 \quad (\text{C.73})$$

$$R_{2c} \geq 0 \quad (\text{C.74})$$

$$R_{2c} \geq R_2 - k_{2-12} + k_{2-1} \quad (\text{C.75})$$

$$R_{2c} \geq R_1 + R_2 - \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) - k_{1-12} \quad (\text{C.76})$$

Those remaining inequalities in (C.44)-(C.66) that do not involve R_{2c} are

$$R_1 \geq 0 \quad (\text{C.77})$$

$$R_1 \leq k_{1-1} \quad (\text{C.78})$$

$$R_2 \leq k_{2-2} \quad (\text{C.79})$$

$$R_1 + R_2 \leq k_{12-2} + k_{1-12} - k_{1-2} \quad (\text{C.80})$$

$$R_1 + R_2 \leq \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) + k_{2-1} + k_{1-2} \quad (\text{C.81})$$

Comparing the lower and upper bounds of R_{2c} , we get 21 inequalities. To save space, we remove those inequalities that are trivial or simply implied by (C.77)-(C.81) and

the remaining inequalities are:

$$R_2 \geq 0 \quad (\text{C.82})$$

$$R_1 + R_2 \leq k_{12-1} + k_{2-12} - k_{2-1} \quad (\text{C.83})$$

$$R_1 + 2R_2 \leq k_{1-2} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{11}) + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) + 2k_{2-12} - k_{2-1} \quad (\text{C.84})$$

$$R_1 + 2R_2 \leq k_{12-2} + k_{2-12} + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) \quad (\text{C.85})$$

$$R_1 + 2R_2 \leq \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21}) + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) + 2k_{2-12} \quad (\text{C.86})$$

$$2R_1 + R_2 \leq k_{12-1} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) + k_{1-12} \quad (\text{C.87})$$

$$2R_1 + R_2 \leq \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{11}) + 2k_{1-12} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) \quad (\text{C.88})$$

$$\begin{aligned} 3R_1 + 2R_2 &\leq \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{11}) + k_{2-1} + k_{1-2} + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) \\ &\quad + 2\text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) + 2k_{1-12} \end{aligned} \quad (\text{C.89})$$

$$\begin{aligned} 3R_1 + 2R_2 &\leq 2k_{2-1} + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21}) + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) \\ &\quad + 2\text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) + 2k_{1-12} \end{aligned} \quad (\text{C.90})$$

Now, we are proceed to identifying the redundant inequalities.

- (C.84) is implied by (C.85) as $\text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{11}) \geq k_{12-2} - k_{2-12} + k_{2-1} - k_{1-2}$.
- (C.86) is implied by (C.79) and (C.83) as

$$\begin{aligned} \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21}) + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) + 2k_{2-12} &\geq \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11}) + 2k_{2-12} \\ &\geq k_{12-1} - k_{2-1} + 2k_{2-12} \geq k_{12-1} + k_{2-12} - k_{2-1} + k_{2-2} \end{aligned}$$

- (C.89) can be obtained by summing up (C.81) and (C.88).
- (C.90) is implied by (C.81) and (C.87).

- (C.88) is implied by (C.78) and (C.80) as

$$\begin{aligned} \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{11}) + 2k_{1-12} + \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) &\geq \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22}) + 2k_{1-12} \\ &= k_{12-2} - k_{1-2} + 2k_{1-12} \geq k_{12-2} - k_{1-2} + k_{1-12} + k_{1-1} \end{aligned}$$

After removing all the redundant inequalities, we obtain the non-trivial set of inequalities.

$$R_1 \geq 0 \tag{C.91}$$

$$R_2 \geq 0 \tag{C.92}$$

$$R_1 \leq k_{1-1} \tag{C.93}$$

$$R_2 \leq k_{2-2} \tag{C.94}$$

$$R_1 + R_2 \leq k_{12-2} + k_{1-12} - k_{1-2} \tag{C.95}$$

$$R_1 + R_2 \leq k_{12-1} + k_{2-12} - k_{2-1} \tag{C.96}$$

$$R_1 + R_2 \leq \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) + \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) + k_{1-2} + k_{2-1} \tag{C.97}$$

$$2R_1 + R_2 \leq \text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) + k_{1-12} + k_{12-1} \tag{C.98}$$

$$R_1 + 2R_2 \leq \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) + k_{2-12} + k_{12-2} \tag{C.99}$$

Considering that (R_1, R_2) are the achievable rate pairs for the double-unicast networks, (C.91) and (C.92) are trivial by causality of the system.

For any given rate pair (R_1, R_2) within the achievable region (C.93)-(C.99). The common-private rates tuple $(R_{1c}, R_{1p}, R_{2c}, R_{2p})$ can be found as

1. With the value of R_1 and R_2 , the range of R_{2c} can be explicitly calculated from the inequalities in (C.67)-(C.76) and R_{2c} can be set as any value within this range;
2. With the value of R_1, R_2 and R_{2c} , the range of R_{1c} can be explicitly calculated

from the inequalities in (C.35)-(C.43) and R_{1c} can be set as any value within this range;

3. Set $R_{1p} = R_1 - R_{1c}$ and $R_{2p} = R_2 - R_{2c}$.

C.1 List of Facts

List of facts that are useful for in Fourier-Motzkin Elimination

1. $\text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11}) = k_{12-1} - k_{2-1}$
2. $\text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22}) = k_{12-2} - k_{1-2}$
3. $\text{rank}(\mathbf{H}_{11} \mathbf{V}_{20}) = k_{1-12} - k_{1-2}$
4. $\text{rank}(\mathbf{H}_{22} \mathbf{V}_{10}) = k_{2-12} - k_{2-1}$
5. $\text{rank}(\mathbf{H}_{11} \mathbf{V}_{21}) \geq k_{1-1} + k_{1-2} - k_{1-12}$
6. $\text{rank}(\mathbf{H}_{22} \mathbf{V}_{11}) \geq k_{2-2} + k_{2-1} - k_{2-12}$
7. $\text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21}) \geq k_{12-1} - k_{1-12} + k_{1-2} - k_{2-1}$
8. $\text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{11}) \geq k_{12-2} - k_{2-12} + k_{2-1} - k_{1-2}$
9. $\text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) \geq k_{12-1} + k_{1-12} - k_{1-1} - k_{2-1} - k_{1-2}$
10. $\text{rank}(\mathbf{U}_{20}^T \mathbf{H}_{22} \mathbf{V}_{10}) \geq k_{12-2} + k_{2-12} - k_{2-2} - k_{2-1} - k_{1-2}$

Proof. Fact 1) can be proved as

$$\begin{aligned}
\text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11}) &= \text{rank}(\mathbf{H}_{11}) - \dim(\mathcal{N}(\mathbf{U}_{10}^T) \cap \mathcal{R}(\mathbf{H}_{11})) \\
&= k_{1-1} - \dim(\mathcal{R}(\mathbf{H}_{12}) \cap \mathcal{R}(\mathbf{H}_{11})) \\
&= k_{1-1} - (k_{1-1} + k_{2-1} - \text{rank}([\mathbf{H}_{11} \quad \mathbf{H}_{12}])) \\
&= \text{rank}([\mathbf{H}_{11} \quad \mathbf{H}_{12}]) - k_{2-1} \\
&\stackrel{(a)}{=} k_{12-1} - k_{2-1}
\end{aligned}$$

where (a) follows as $\begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \end{bmatrix}$ is full row rank. By symmetry, Fact 2) can be proved.

To prove Fact 3)

$$\begin{aligned}
\text{rank}(\mathbf{H}_{11} \mathbf{V}_{20}) &= \text{rank}(\mathbf{V}_{20}) - \dim(\mathcal{N}(\mathbf{H}_{11}) \cap \mathcal{R}(\mathbf{V}_{20})) \\
&= k_{1-12} - k_{1-2} - \dim(\mathcal{N}(\mathbf{H}_{11}) \cap \mathcal{N}(\mathbf{H}_{21})) \\
&\stackrel{(b)}{=} k_{1-12} - k_{1-2}
\end{aligned}$$

where (b) follows as $\begin{bmatrix} \mathbf{H}_{11} \\ \mathbf{H}_{21} \end{bmatrix}$ is full column rank. By symmetry, Fact 4) can be proved.

Fact 5) can be proved as $\text{rank}(\mathbf{H}_{11}) = \text{rank}\left(\mathbf{H}_{11} \begin{bmatrix} \mathbf{V}_{21} & \mathbf{V}_{20} \end{bmatrix}\right) \leq \text{rank}(\mathbf{H}_{11} \mathbf{V}_{21}) + \text{rank}(\mathbf{H}_{11} \mathbf{V}_{20})$. Since $\text{rank}(\mathbf{H}_{11} \mathbf{V}_{20}) = k_{1-12} - k_{1-2}$, $\text{rank}(\mathbf{H}_{11} \mathbf{V}_{21}) \geq k_{1-1} + k_{1-2} - k_{1-12}$. Fact 6) follows similarly.

Fact 7) can be proved using Frobenius inequality,

$$\begin{aligned}
\text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{21}) &\geq \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11}) + \text{rank}(\mathbf{H}_{11} \mathbf{V}_{21}) - \text{rank}(\mathbf{H}_{11}) \\
&\geq k_{12-1} - k_{2-1} + k_{1-1} + k_{1-2} - k_{1-12} - k_{1-1} \\
&= k_{12-1} - k_{1-12} + k_{1-2} - k_{2-1}
\end{aligned}$$

Fact 8) follows similarly.

Fact 9) is also proved using Frobenius inequality,

$$\begin{aligned}\text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11} \mathbf{V}_{20}) &\geq \text{rank}(\mathbf{U}_{10}^T \mathbf{H}_{11}) + \text{rank}(\mathbf{H}_{11} \mathbf{V}_{20}) - \text{rank}(\mathbf{H}_{11}) \\ &= k_{12-1} - k_{2-1} + k_{1-12} - k_{1-2} - k_{2-2}\end{aligned}$$

Fact 10) follows similarly.

□

Appendix D

Detailed Steps for Deriving (6.3)

In this appendix, we present the detailed steps for deriving (6.3).

$$\begin{aligned}\mathbb{E}[\mathbf{Y}] &= \int_0^\infty \frac{2x}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) Q\left(\frac{x-\mu}{\sigma}\right) dx \\ &\stackrel{(a)}{\approx} \int_{-\infty}^\infty \frac{2x}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) Q\left(\frac{x-\mu}{\sigma}\right) dx\end{aligned}\quad (\text{D.1})$$

where (a) follows the fact that when μ is sufficiently large, the negative part of the integration is negligible. However, including the negative part introduces certain symmetry and allows us to manipulate the integration easily.

For brevity, let $t = x - \mu$. Hence, (D.1) can be written as

$$\begin{aligned}\mathbb{E}[\mathbf{Y}] &= \int_{-\infty}^\infty \frac{2(t+\mu)}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right) Q\left(\frac{t}{\sigma}\right) dt \\ &= \int_{-\infty}^0 \frac{2(t+\mu)}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right) Q\left(\frac{t}{\sigma}\right) dt + \int_0^\infty \frac{2(t+\mu)}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right) Q\left(\frac{t}{\sigma}\right) dt \\ &= \int_0^\infty \frac{2(-t+\mu)}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right) Q\left(\frac{-t}{\sigma}\right) dt + \int_0^\infty \frac{2(t+\mu)}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right) Q\left(\frac{t}{\sigma}\right) dt \\ &= \int_0^\infty \frac{2(-t+\mu)}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right) \left[1 - Q\left(\frac{t}{\sigma}\right)\right] dt + \int_0^\infty \frac{2(t+\mu)}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right) Q\left(\frac{t}{\sigma}\right) dt\end{aligned}$$

$$\begin{aligned}
&= \int_0^\infty \frac{2(-t + \mu)}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right) dt + \int_0^\infty \frac{4t}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right) Q\left(\frac{t}{\sigma}\right) dt \\
&= -2 \int_0^\infty \frac{t}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right) dt + 2\mu \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right) dt \\
&\quad + \int_0^\infty \frac{4t}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right) Q\left(\frac{t}{\sigma}\right) dt \\
&= \mu - \sqrt{\frac{2\sigma^2}{\pi}} + \int_0^\infty \frac{4t}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right) Q\left(\frac{t}{\sigma}\right) dt
\end{aligned}$$

The second part of above equation can be calculated as

$$\begin{aligned}
&\int_0^\infty \frac{4t}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right) Q\left(\frac{t}{\sigma}\right) dt \\
&= \int_0^\infty \frac{4t}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right) \int_t^\infty \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{u^2}{2\sigma^2}\right) du dt \\
&= \int_{u=0}^\infty \int_{t=0}^u \frac{4t}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{u^2}{2\sigma^2}\right) dt du \\
&= \int_{u=0}^\infty \frac{4}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{u^2}{2\sigma^2}\right) \int_{t=0}^u \frac{t}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right) dt du \\
&= \int_{u=0}^\infty \frac{4}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{u^2}{2\sigma^2}\right) \sqrt{\frac{\sigma^2}{2\pi}} \left[1 - \exp\left(-\frac{u^2}{2\sigma^2}\right)\right] du \\
&= 4\sqrt{\frac{\sigma^2}{2\pi}} \left[\int_{u=0}^\infty \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{u^2}{2\sigma^2}\right) du - \int_{u=0}^\infty \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{u^2}{\sigma^2}\right) du \right] \\
&\stackrel{(b)}{=} 4\sqrt{\frac{\sigma^2}{2\pi}} \left[\frac{1}{2} - \frac{1}{2\sqrt{2}} \right] \\
&= \sqrt{\frac{2\sigma^2}{\pi}} \left[1 - \frac{1}{\sqrt{2}} \right],
\end{aligned}$$

where (b) follows by the symmetry of zero-mean Gaussian distribution, i.e., the integration from 0 to ∞ of zero-mean Gaussian distribution returns $\frac{1}{2}$.

Hence, we have

$$\begin{aligned}
\mathbb{E}[\mathbf{Y}] &= \mu - \sqrt{\frac{2\sigma^2}{\pi}} + \sqrt{\frac{2\sigma^2}{\pi}} \left[1 - \frac{1}{\sqrt{2}} \right] \\
&= \mu - \frac{\sigma}{\sqrt{\pi}}.
\end{aligned}$$

Appendix E

Number Partitions

In this appendix, we present a recursive method for computing the partitions of a positive integer that has been used in deriving the throughput of erasure networks in Chapter 6.

To compute $P(n, k, L)$, we start from the simplest case when $k = 2$, i.e., each integer is to be partitioned into two parts. For $3 \leq n \leq L$, it can be partitioned into two distinct part as $n = (i) + (n - i), i = 1, \dots, \lfloor \frac{n-1}{2} \rfloor$. Similarly, $P(n, 2, L)$ can be determined for $L < n \leq 2L - 1$ and thus we have

$$P(n, 2, L) = \begin{cases} \lfloor \frac{n-1}{2} \rfloor, & 3 \leq n \leq L \\ L - \lfloor \frac{n}{2} \rfloor, & L < n \leq 2L - 1 \\ 0, & \text{otherwise} \end{cases} \quad (\text{E.1})$$

For $k \geq 3$, we will compute $P(n, k, L)$ recursively. If the largest part of the partition is $i, 2 \leq i \leq L$, the number of partitions is equal to total number of ways to partition $n - i$ into $k - 1$ parts with the largest part not exceeding $i - 1$, i.e.,

$P(n - i, k - 1, i - 1)$. Therefore, we have

$$P(n, k, L) = \sum_{i=2}^L P(n - i, k - 1, i - 1).$$

Similar method applies for computing $J(n, k, L, i)$. If $k = 2$, we have

$$J(n, 2, L, i) = \begin{cases} 1, & i + 1 \leq n \leq L + i, n \neq 2i \\ 0, & \text{otherwise} \end{cases}.$$

For $k \geq 3$, we fixed one part to i and partition $n - i$ into $k - 1$ distinct parts. Note that due to the requirement of distinction among parts, those $k - 1$ parts should not contain i . Therefore, we have

$$J(n, k, L, i) = P(n - i, k - 1, L) - J(n - i, k - 1, L, i).$$