Network Coding and Distributed Compression over Large Networks: Some Basic Principles

Thesis by

Mayank Bakshi

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2012
(Defended August 16, 2011)

Dedicated to my family

Acknowledgements

The work contained in this thesis was partially supported by NSF Grant Nos CCF-0220039, CCR-0325324, and CNS-0905615, Darpa Grant No W911NF-07-I0029, and Caltech's Lee Center for Advanced Networking. In addition, I would like to express my gratitude towards each and every person with whom I had the fortune of interacting during the course of my doctoral studies. In there own way, everyone I met and came to know made my stay at Caltech memorable and helped shape my research. At the risk of omitting many names, I would like to especially acknowledge the contribution of some key individuals.

Without doubt, the greatest contribution in this regard was that of my advisor, Prof Michelle Effros. I have many fond memories of my interactions with her. Her principled approach towards everything, technical as well as non-technical, was often my guiding light. Her constant encouragement along with countless insightful discussions allowed me to develop deeper understanding of my research topics even when new ideas were not forthcoming. I also appreciate Prof Effros' patient handling of the inevitable differences between her and mine approaches to certain problems. This understanding afforded me just enough flexibility to truly enjoy research, while ensuring that things kept progressing.

Simultaneously, I had the privilege of working closely with Prof Tracey Ho. I would like to thank her for providing very valuable feedback throughout my doctoral research. On a similar note, I would also like to mention committee members Prof Shuki Bruck, Prof Mani Chandy, Prof Adam Wierman, Prof Babak Hassibi, and Prof Andreas Krause. I also had the pleasure of being part of many fruitful discussions with former group mates Sid, Wei-Hsin, and Asaf. These discussions kept the creative juices flowing and ultimately led to several new research ideas. Through Prof Effros, I came to know many other prominent researchers in my field. Of special mention are Prof Ralf Koetter, Prof Muriel Medard and Prof Alon Orlitsky. Interactions with them were crucial in shaping parts of the thesis.

As a part my life as a Caltech student, I met many remarkable people. I learnt a lot through my former group mates Chris, Sukwon, Derek, Shirin, Ted, Tao, Svitlana, and Ming. Our meetings always led to interesting conclusions, both in research and outside of it. The EE staff, especially Shirley, Linda, and Tanya were instrumental in ensuring that logistics were take care of. Greg and

Christina at Caltech Y helped ensure that time was spent well even outside of research. The health center staff, especially Divina, also earn a special mention for keeping me fighting fit for the rigors of graduate life.

I was extremely lucky to make several truly wonderful friends who helped me stay sane during this time. Sushree made sure I did not miss my family as I toiled towards my thesis completion. Uday and Prabha were there with me every step of the way since landing in Pasadena. Shankar and Shwetank were always ready with helpful advice whenever I needed. In no particular order, I would also like to acknowledge JK, Teja, Saha, Sowmya, Vijay, Arundhati, Samir, Mansi, Setu, Pinkesh, Zeeshan, Shwetank, Vikram, Krish, Sukhada, Arthur, Jean-Loup, and Ashish for memories that I will cherish for long. Last, but not the least, I would like to acknowledge my parent and my younger brother for constant support and nagging without which none of this would have been possible.

Abstract

The fields of Network Coding and Distributed Compression have focused primarily on finding the capacity for families of problems defined by either a broad class of networks topologies (e.g., directed, acyclic networks) under a narrow class of demands (e.g., multicast), or a specific network topology (e.g. three-node networks) under different types of demands (e.g. Slepian-Wolf, Ahlswede-Körner). Given the difficulty of the general problem, it is not surprising that the collection of networks that have been fully solved to date is still very small. This work investigates several new approaches to bounding the achievable rate region for general network source coding problems - reducing a network to an equivalent network or collection of networks, investigating the effect of feedback on achievable rates, and characterizing the role of side information.

We describe two approaches aimed at simplifying the capacity calculations in a large network. First, we prove the optimality of separation between network coding and channel coding for networks of point-to-point channels with a Byzantine adversary. Next, we give a strategy for calculation the capacity of an error-free network by decomposing that network into smaller networks. We show that this strategy is optimal for a large class of networks and give a bound for other cases.

To date, the role of feedback in network source coding has received very little attention. We present several examples of networks that demonstrate that feedback can increases the set of achievable rates in both lossy and lossless network source coding settings. We derive general upper and lower bounds on the rate regions for networks with limited feedback that demonstrate a fundamental tradeoff between the forward rate and the feedback rate. For zero error source coding with limited feedback and decoder side information, we derive the exact tradeoff between the forward rate and the feedback rate for several classes of sources. A surprising result is that even zero rate feedback can reduce the optimal forward rate by an arbitrary factor.

Side information can be used to reduce the rates required for reliable information. We precisely characterize the exact achievable region for multicast networks with side information at the sinks and find upper and lower bounds on the achievable rate region for other demand types.

Contents

Chapter 1

Introduction

1.1 The Setup

Consider a communication system shown in Figure 1.1. Nodes 1 through m have source messages $W^{(1)}$ through $W^{(m)}$ available to them and wish to reconstruct functions $f_1(W^{(1)}, \ldots, W^{(m)})$ of all the source messages. Each edge in the network corresponds to a point-to-point channel whose inputs and outputs are related statistically according to transition probability $p_{Y|X}$. In order to achieve the communication objective, each node is allowed to perform the following *coding* operation. At time t, each ndoe v collects symbols $Y_1^{(v)}, \ldots, Y_{t-1}^{(v)}$ received at that node prior to time t, computes a function $X_t^{(v)}$ of these symbols, and transmits $X_t^{(v)}$ on outgoing edges from v. We call a transmission scheme of this form a *network code* when the source messages are independent, and a *network source code* when the sources are dependent.

Two figures of merits of network codes (and network source codes) that we consider are its *error probability* and *rate*. These terms are precisely defined later. Loosely speaking, the error probability of a network code (or a network source code) is the probability that at least one of the sinks incorrectly reconstructs its desired demands. The rate of a code measures its transmission efficiency. For a network code, this is typically measured in terms of the number of bits of each source message transmitted per time instant, while for a network source code it is usually measured as the number of transmitted codewords symbols transmitted per source symbol. In both these scenarios, the key quantity of interest to us is the set of *achievable rates* $\mathcal{R}$, which is defined as the collection of rates for which there exist network codes with arbitrarily small error probability.

The network model of Figure 1.1 may be thought of as a natural generalization of the point to point model of Figure 1.2. Shannon [1] showed that it is possible to reliably communicate a message in the presence of noise as long as the *"amount of information"* contained in the message is less than the *"capacity"* of the noisy channel. This model lends itself naturally to a vast number of channel coding and source coding scenarios – the two main themes explored in [1]. However, the above point-to-point model does not fully capture many network-based communication scenarios; the internet,

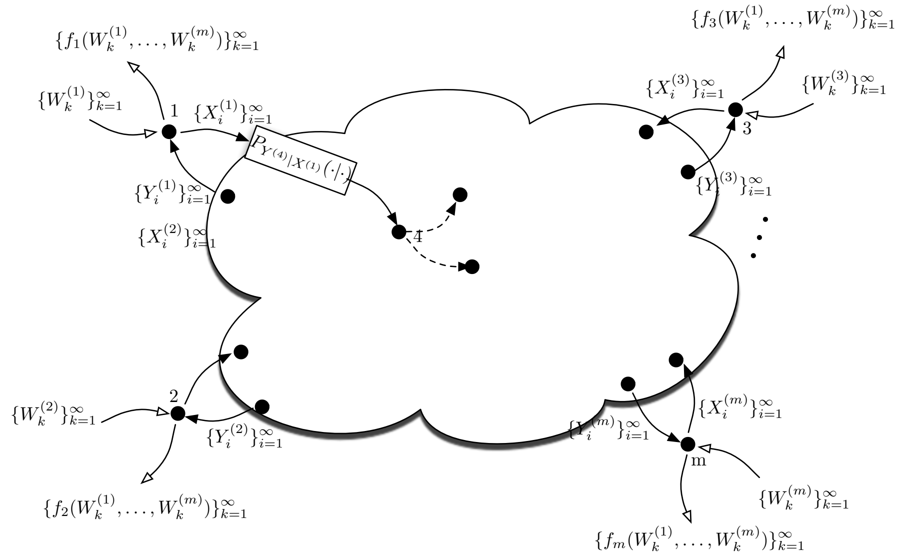


Figure 1.1: A general multi-terminal communication model

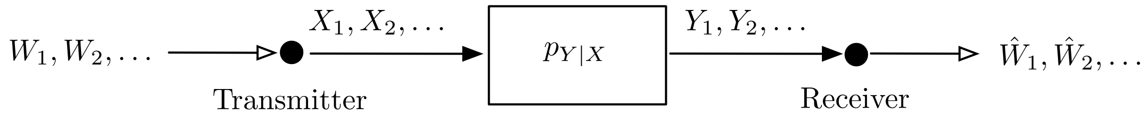


Figure 1.2: The point-to-point communication model

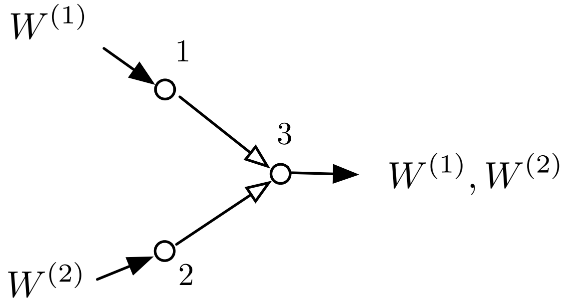


Figure 1.3: The Slepian-Wolf network

sensor networks, content distribution networks, and cellular networks are some examples of such systems. The following simple example makes this evident.

***Example* 1** (Slepian-Wolf [6])**.** *Consider the network shown in Figure 1.3. Node* 1 *and* 2 *observe source processes* $\{W_i^{(1)}\}_{i=1}^{\infty}$ *and* $\{W_i^{(1)}\}_{i=1}^{\infty}$*, respectively. For each* $i = 1, 2, \ldots,$*, the pair of random variables* $(W_i^{(1)}, W_i^{(2)})$ *is drawn independently from a known distribution* $p_{W^{(1)},W^{(2)}}(\cdot)$*. Node* 3 *wishes to reconstruct both the sources.*

By the Source Coding Theorem [1], the smallest rates achievable on the edges $(1,3)$ *and* $(2,3)$ *using the point-to-point approach are* $H(W^{(1)})$ *and* $H(W^{(2)})$*, respectively.*

On the other hand, Slepian and Wolf [6] show that all rates satisfying the following bounds allow recovering both $W^{(1)}$ *and* $W^{(2)}$ *at the decoder:*

$$\begin{aligned} R_1 &\geq H(W^{(1)}|W^{(2)}) \\ R_2 &\geq H(W^{(2)}|W^{(1)}) \\ R_1 + R_2 &\geq H(W^{(1)}, W^{(2)}). \end{aligned}$$

In particular, it follows that rates $H(W^{(1)})$ *and* $H(W^{(2)}|W^{(1)})$ *on edges* $(1,3)$ *and* $(2,3)$*, respectively, are achievable. Since* $H(W^{(2)}|W^{(1)}) < H(W^{(2)})$ *for a large class of interesting probability distributions, this illustrates that the point-to-point approach is suboptimal in general networked scenarios.*

The sub-optimality of the point-to-point approach is also well established for general multi-user channels (see [7, 8]) and transmission over networks with more than one source-destination pair (see [9]). These further highlighting the inadequacy of point-to-point approach when dealing with multi-terminal data compression scenarios. While these examples make a compelling case for accomplishing general multi-terminal communication goals through schemes that are optimized taking the entire network into account, in practice, such an approach is hindered by rapid increase in complexity of analysis with increase in network size. Indeed, even the set of rates achievable using

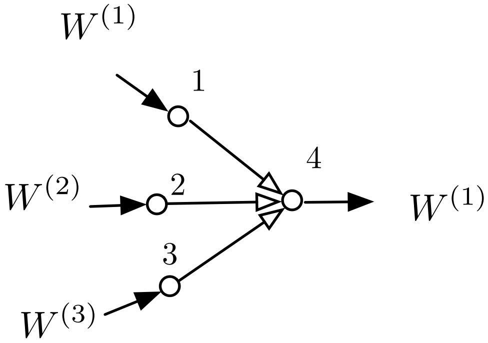


Figure 1.4: A network with two side information sources

communication schemes that are optimized for the entire network is known for only a few special classes of networks. The following seemingly simple network is an example of a network for which the entire set of achievable rates is unknown till date.

***Example* 2.** *Consider the network shown in Figure 1.4. Let* $W^{(1)}, W^{(2)}$*, and* $W^{(3)}$ *be sources observed at nodes* 1*,* 2*, and* 3*, respectively. Let* $P_{W^{(1)},W^{(2)},W^{(3)}}(\cdot)$ *be the joint distribution of these sources. Node* 4 *demands only* $W^{(1)}$*. Let edges* $(1,4),(2,4)$*, and* $(3,4)$ *be noiseless links. The set of achievable rates for this network is defined as the collection of rate triplets* (R_1, R_2, R_3) *such that for every* $\epsilon > 0$*, there exists a transmission scheme at rates* R_1*,* R_2*, and* R_3 *bits per unit time on links* $(1,4),(2,4)$*, and* $(3,4)$*, respectively, for which* $W^{(1)}$ *is reconstructed at node* 4 *with error probability at most* ϵ*.*

Traditionally, much of the research in analyzing general multi-terminal systems has focussed on solving networks such as the above by analyzing each such network separately. While considering networks one at a time has the advantage that resulting performance bounds and techniques are often specialized to the specific networks, it is perhaps also true that following this approach limits the class of networks that are well understood from an information-theoretic point of view. This motivates us to study properties that apply to general multi-terminal systems.

1.2 Our contribution

In this thesis, we attempt to gain insights into general networks by identifying several key information-theoretic principles. Specifically, we explore the following ideas that are applicable to multi-terminal systems in varying generality.

1.2.1 Network Reduction

In Chapters 2 and 3, we consider two approaches towards solving general network problems by reducing them to potentially simpler problems.

The approach considered in Chapter 2 shows that finding the adversarial network coding capacity of a network of independent point-to-point noisy channels is equivalent to finding the adversarial network coding capacity of another network which consists solely of error-free links of known capacity. This result extends previous works on this theme which deal with non-adversarial setups [10, 11]. This also shows that channels with same capacity are equivalent as far as their effect in network coding capacities are concerned. Combined with results of [10] and [11], this results enables us to focus our attention exclusively on networks of lossless links when the networks consists of point-to-point channels. In the rest of the thesis, we only consider networks of lossless capacitated links.

In the spirit of the above equivalence, one is tempted to believe that an analogous result should also hold true for sources themselves, i.e., if two networks that have the same topology have the same source coding rate regions if the corresponding sources in the two networks have the same joint entropies. In Chapter 2, we formally state this notion and show that while this equivalence holds for some collection of sources, it fails for general data compression problems.

Next, in Chapter 3, we explore the idea of decomposing a network into component networks and using the component networks to gain insights into for the original network. We focus specifically on Line Networks and show that for several classes of these networks, the set of achievable rates may be fully characterized using the decomposition approach. We also give counterexamples to show that this approach is suboptimal in general. Following the results from this chapter, the rest of the thesis focusses exclusively on networks of error-free links.

1.2.2 Feedback in Network Source Coding

In Chapter 4, we consider network source coding for error-free networks with feedback from sink nodes to source nodes. We present several examples of networks with unlimited feedback to demonstrate that the presence of feedback strictly increases the set of achievable rates in both lossy and lossless settings. We also observe that the presence of feedback can reduce the encoding and decoding complexity in some cases. Next, we derive general upper and lower bounds on the rate regions of networks with limited feedback demonstrating a fundamental tradeoff between the forward rates and the feedback rates. Finally we restrict our attention to zero-error source coding with limited feedback and decoder side information and derive the exact tradeoff between the forward rate and the feedback rate for several classes of sources. A surprising result is that even a zero-rate feedback can reduce the optimal forward rate by an arbitrary factor. Our results generalize prior works that focus largely on networks with two nodes [12, 13, 14, 15, 16, 17].

1.2.3 Side Information in Networks

Lastly, in Chapter 5, we focus on the role played by side information in network source coding problems. We first consider networks with multicast demands, i.e., all sinks want all sources, and characterize the exact rate region when each sink may have (possibly different) side information available to it. This result generalizes prior work on multicast networks without side information at the sinks [18] Next, we use this to derive a general achievable region for networks where side information may be present at a non-sink node in the network. Finally, we consider a general source-demand problem and use the results from the case with side information to derive an achievable rate region for this setup.

We present our findings in the next four chapters. Each chapter explore a different fundamental aspect of multi-terminal communication systems. In the interest of keeping the notation for each chapter consistent with prior literature on that topic, the notation for each chapter is independently set up at the beginning of the chapter.

Chapter 2

Network Equivalence

2.1 Introduction

One common approach for communicating in networks of noisy channels is to separate network coding and channel coding. In this approach, we operate each channel essentially losslessly with the help of a channel code. We then perform network coding on an essentially noise-free network. Indeed, in [19, 10], this approach is shown to be asymptotically optimal when the noise values on the distinct channels of the network are independent of each other. It is also known that when channels corresponding to different links are not independent, operating the channel code for each link independently may be strictly suboptimal (see Example 2 in [10]). In these cases, the dependence between the noise values on different links is exploited by first creating an appropriate dependence between the transmitted codewords on these channels and then jointly decoding them at the receiver.

In this chapter, we consider a network of independent point-to-point channels with the presence of a Byzantine adversary that observes all transmissions, messages, and channel noise values, and can corrupt some of the transmissions by replacing a constrained subset of the received channel outputs. The objective of the adversary is to maximize the probability of decoding error, and the capacity of the network is the set of vectors describing rates at which it is possible to reliably communicate across the network. It is tempting to believe that separation of network coding and channel coding is suboptimal in the case of our adversarial model due to the potential for statistical dependence between the "noise" observed on edges controlled by the adversary. We show, however, that the capacity of this network equals the adversarial capacity of another network in which each channel is replaced by a noise-free capacitated link of the same capacity. Thus, it is asymptotically optimal to operate the adversarial network code independently of the channel code in this framework. We do not assume any special structure on the topology of the network, e.g., we allow unequal link capacities and networks with cycles. We also allow arbitrary model of adversarial attack, e.g. edge-based or node-based attack. The result immediately extends previous adversarial network coding capacity results from noise-free networks (e.g. [20, 21, 22, 23, 24]) to that of networks of independent point-to-point channels.

The proof follows the strategy introduced in [19, 10]. In Section 2.3, we show that the adversarial capacity of a network is same as that of a *stacked network* comprised of many copies of the same network. In Section 2.4, we show that replacing one of the channels with a noiseless link of equal capacity does not alter the adversarial network coding capacity of the stacked network. We give a formal problem definition in Section 2.2.

The above result also shows that the notion of *Network Equivalence* introduced by [19] extends beyond networks of independent channels, and exact equivalence holds even when an adversary may introduce arbitrary errors. A key assumption in both our result and the results of [19] is that all sources are independent. Jalali et al. [11] have extended the result to networks with dependent sources. Unlike the previous setups which are limited to independent sources, in networks with dependent sources, the joint distribution of sources (and not just their respective entropies) is needed in the design the network code. As a result, the achievability of a given collection of demands over a given network may depend critically on the joint distribution of the sources present. We illustrate this in Section 2.5 by with the help of an example.

2.2 Preliminaries

2.2.1 Network model

We define a network $\mathcal{N}$ to be a pair $(\mathcal{G}, \mathcal{C})$. Here $\mathcal{G} = (\mathcal{V}, E)$ is a directed graph with vertices $\{1, \ldots, m\}$ and directed edges $E \subseteq \mathcal{V} \times \mathcal{V}$. Each edge $e \in E$ describes the input and output of a point-to-point channel $\mathcal{C}_e$. The full collection of channels is given by $\mathcal{C} = (\mathcal{C}_e : e \in E)$.

For each $e \in E$, channel $\mathcal{C}_e$ is given by a vector $(\mathcal{X}^{(e)}, \mathcal{Y}^{(e)}, \mathcal{Z}^{(e)}, P_e, \Upsilon_e)$, where $\mathcal{X}^{(e)}$, $\mathcal{Y}^{(e)}$, and $\mathcal{Z}^{(e)}$ are, respectively, the input, output, and noise alphabets of the channel, P_e is the probability distribution of the noise, and $\Upsilon_e : \mathcal{X}^{(e)} \times \mathcal{Z}^{(e)} \to \mathcal{Y}^{(e)}$ is the channel map that determines the channel output as a function of the channel input and noise. The noise distribution P_e and mapping Υ_e together induce a conditional probability distribution of the channel output given the channel input, here denoted by $p_e(\cdot|\cdot)$. Thus, the random variables $X^{(e)}$, $Y^{(e)}$, and $Z^{(e)}$ denoting the input, the output, and the noise value of the channel, are related as

$$Y^{(e)} = \Upsilon_e(X^{(e)}, Z^{(e)}),$$

with

$$p_e(y|x) = \int_{\{z:\Upsilon_e(x,z)=y\}} P_e(z)dz.$$

For each $t \in \mathbb{N}^+$ and $e \in E$, let $X_t^{(e)} \in \mathcal{X}^{(e)}$, $Y_t^{(e)} \in \mathcal{Y}^{(e)}$, and $Z_t^{(e)} \in \mathcal{Z}^{(e)}$, respectively, be the random variables denoting the transmitted, received, and noise values for edge e at time t. We assume that each transmission on edge e involves a delay of unit time and that the noise on all channels is independent and memoryless. Thus,

$$Y_{t+1}^{(e)} = \Upsilon_e(X_t^{(e)}, Z_t^{(e)}) \; \forall e \in E, \; t \in \mathbb{N}^+$$

and

$$P_E(Z_\tau^{(e)} : e \in E, \tau = 1, \ldots, t) = \prod_{e \in E} \prod_{\tau=1}^{t} P_e(Z_\tau^{(e)}).$$

Here, P_E denotes the joint distribution of the noise values.

For notational convenience, we adopt the following convention to represent collections of random vectors. For every collection of random variables $Q_1, Q_2, \ldots$ taking values from a set $\mathcal{Q}$, we denote the row vector $[Q_t, Q_{t+1}, \ldots, Q_{t+n-1}] \in \mathcal{Q}^n$ by $Q_{t:t+n-1}$. We specify column vectors by underlining them and the element of a given row from the column vector by parenthesis. Thus, $\underline{Q} \in \mathcal{Q}^N$ represents the column vector $[\underline{Q}(1), \underline{Q}(2), \ldots, \underline{Q}(N)]^T$ with $\underline{Q}(i) \in \mathcal{Q}$ for all i.

For each $v \in \mathcal{V}$ and $t \in \mathbb{N}^+$, let $X_t^{(v,*)} \triangleq (X_t^{(v,w)} : (v,w) \in E)$ and $X_t^{(*,v)} \triangleq (X_t^{(u,v)} : (u,v) \in E)$ denote the time-t random variables on edges outgoing from v and incoming to v, respectively; the alphabets for $X_t^{(v,*)}$ and $X_t^{(*,v)}$ are $\mathcal{X}^{(v,*)} = \prod_{u:(v,u) \in E} \mathcal{X}^{(v,u)}$ and let $\mathbf{X}_t \triangleq (X_t^{(e)} : e \in E)$ denote all transmitted random variables in the network at time t. Similarly, define $Y_t^{(v,*)}$, $Y_t^{(*,v)}$, $Z_t^{(v,*)}$, $Z_t^{(*,v)}$, $\mathbf{Y}_t$, and $\mathbf{Z}_t$ for each $v \in \mathcal{V}$ and $t \in \mathbb{N}^+$. Let $\mathcal{X}$, $\mathcal{X}^{(u,*)}$, and $\mathcal{X}^{(*,v)}$ denote the product sets $\prod_{e \in E} \mathcal{X}^{(e)}$, $\prod_{v:(u,v) \in E} \mathcal{X}^{(u,v)}$, and $\prod_{v:(u,v) \in E} \mathcal{X}^{(u,v)}$. Similarly define $\mathcal{Y}$, $\mathcal{Y}^{(u,*)}$, $\mathcal{Y}^{(*,v)}$, $\mathcal{Z}$, $\mathcal{Z}^{(u,*)}$, and $\mathcal{Z}^{(*,v)}$.

2.2.2 Network code

Let $\mathcal{M} = \{(u,V) : u \in \mathcal{V}, V \subseteq \mathcal{V} \setminus \{u\}\}$ denote the set of possible pairs of source nodes and sink sets. A network coding solution $\mathcal{S}(\mathcal{N})$ implemented over n time steps is defined by message alphabet $\mathcal{W} = \prod_{(u,V) \in \mathcal{M}} \mathcal{W}^{(u \to V)}$, the collection of encoder maps $\{f_t^{(u,v)} : (u,v) \in E, t \in \{1, \ldots, n\}\}$ with

$$f_t^{(u,v)} : \prod_{V \subseteq \mathcal{V}} \mathcal{W}^{(u \to V)} \times \prod_{v' : (v',u) \in E} (\mathcal{Y}^{(v',u)})^t \to \mathcal{X}^{(u,v)}$$

that determine the transmitted random variable $X_t^{(u,v)}$ as a function of the messages $(W^{(u \to V)} : V \subseteq \mathcal{V})$ and received vectors $Y_{1:t}^{(*,u)}$ at node u, and the decoder maps $\{g^{(u)} : u \in V\}$ with

$$g^{(u)} : \prod_{V \subseteq \mathcal{V} \setminus \{u\}} \mathcal{W}^{(u \to V)} \times \prod_{v:(v,u) \in E} (\mathcal{Y}^{(v,u)})^n \to \prod_{\substack{V \subseteq \mathcal{V} : u \in V \\ v \in \mathcal{V} \setminus \{u\}}} \mathcal{W}^{(v \to V)}$$

that determine the reconstructed messages $(\hat{W}^{(v\to V,u)} : (v,V) \in \mathcal{M}, v \in \mathcal{V}, u \in V)$ as a function of the messages $(W^{(u\to V')} : V' \subseteq \mathcal{V} \setminus \{u\})$ and received vectors $Y_{1:t}^{(*,u)}$ at node u for all $t = 1,\dots,n$ and $u \in \mathcal{V}$. Let $\mathbf{R} = (R(u,V) : (u,V) \in \mathcal{M}) \in \mathbb{R}^{|\mathcal{M}|}$. We say that a solution $\mathcal{S}(\mathcal{N})$ is a rate $\mathbf{R}$ solution if $|\mathcal{W}^{(u\to V)}| = 2^{nR(u,V)}$ for all $(u,V) \in \mathcal{M}$. Without loss of generality, we assume that all messages are either binary vectors or binary matrices of appropriate dimensions.

2.2.3 Adversarial model

We assume an omniscient Byzantine adversary that observes all messages $(W^{(u\to V)} : (u,V) \in \mathcal{M})$, noise values $\mathbf{Z}_{1:n}$, and the network code $\mathcal{S}$ in operation. Thus, the adversary can deduce all transmitted and received vectors, $\mathbf{X}_{1:n}$ and $\mathbf{Y}_{1:n}$. The adversary picks a subset σ from the set $\mathbf{\Sigma}$ of permissible attack-sets and replaces the vectors $(Y_t^{(e)} = \Upsilon(X_{t-1}^{(e)}, Z_{t-1}^{(e)}) : e \in \sigma, t = 1,\dots,n)$ of channel outputs on these edges with the vector $\mathbf{A}_{1:n} = (A_{1:n}^{(e)} : e \in \sigma)$ of his own choice. The set $\mathbf{\Sigma}$ is known to the designer of the network code, but the chosen attack set $\sigma \in \mathbf{\Sigma}$ is unknown.

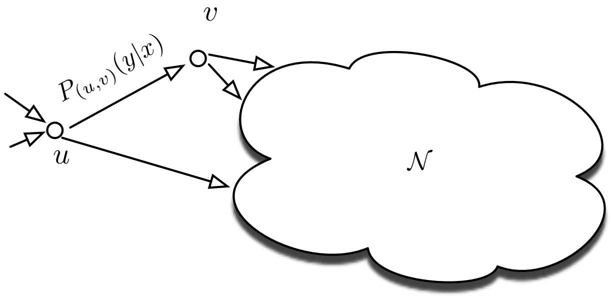
We say that there is a *decoding error* if $\hat{W}^{(v\to V,u)} \neq W^{(v\to V)}$ for some $v \in \mathcal{V}$, $V \subseteq \mathcal{V} \setminus \{v\}$, and $u \in V$. For a given solution $\mathcal{S}(\mathcal{N})$ that is implemented over n time steps, and for each $(u,V) \in \mathcal{M}$ and $v \in V$, $\hat{W}^{(u\to V,v)}$ is a deterministic function of the messages $W = (W^{(u\to V)} : (u,V) \in \mathcal{M})$, the noise values $\mathbf{Z}_{1:n}$, the attack-set σ, and the injected vector $\mathbf{A}_{1:n}$; let $G_{\mathcal{S}}^{(u\to V,v)} : \mathcal{W} \times \mathcal{Z}^n \times \mathbf{\Sigma} \times \mathcal{Y}^n \to \mathcal{W}$ denote this function. Since the adversary knows W and $\mathbf{Z}_{1:n}$, he can compute the decoded message for every possible choice of σ and $\mathbf{A}_{1:n}$. The adversary may then chose σ and $\mathbf{A}_{1:n}$ to minimize the rate of reliable communication. We define the set $\mathcal{E}(\mathcal{S}) \subseteq \mathcal{W} \times \mathcal{Z}^n$ as the collection of messages and noise values for which it is possible for the adversary to cause a decoding error for any of the messages, i.e.,

$$\begin{aligned}\mathcal{E}(\mathcal{S}) \triangleq \big\{&(w, \mathbf{z}_{1:n}) \in \mathcal{W} \times \mathcal{Z}^n : \\ &G_{\mathcal{S}}^{(u\to V,v)}(w, \mathbf{z}_{1:n}, \mathbf{a}_{1:n}, \sigma) \neq w^{(u,V)} \text{ for some} \\ &(u,V) \in \mathcal{M}, v \in V, \sigma \in \mathbf{\Sigma}, \text{ and } \mathbf{a}_{1:n} \in \prod_{e\in\sigma} (\mathcal{Y}^{(e)})^n \big\}.\end{aligned}$$

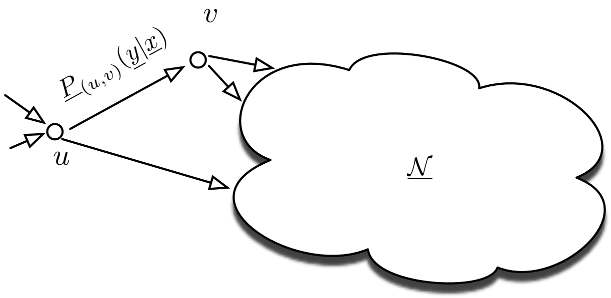
The *probability of error* for the solution $\mathcal{S}(\mathcal{N})$ is

$$\begin{aligned}P_{\mathcal{E}}(\mathcal{S}) &\triangleq \Pr_{W,\mathbf{Z}_{1:n}}((W, \mathbf{Z}_{1:n}) \in \mathcal{E}(\mathcal{S})) \\ &= \frac{1}{|\mathcal{W}|} \sum_{w\in W} \int_{\{\mathbf{z}_{1:n} : (w,\mathbf{z}_{1:n}) \in \mathcal{E}(\mathcal{S})\}} P_E(\mathbf{z}_{1:n}) d\mathbf{z}_{1:n}.\end{aligned}$$

We say that a solution $\mathcal{S}(\mathcal{N})$ that is implemented over n time steps is a $(\lambda, \mathbf{R})$-solution if $|\mathcal{W}^{(u\to V)}| = 2^{nR(u,V)}$ for every $(u,V) \in \mathcal{M}$ and $P_{\mathcal{E}}(\mathcal{S}) < \lambda$. The *capacity region* $\mathcal{R}(\mathcal{N})$ of a network $\mathcal{N}$ is the closure of the set of all rate vectors $\mathbf{R}$ for which a $(\lambda, \mathbf{R})$-solution exists for every $\lambda > 0$.



(a) The original network $\mathcal{N}$.



(b) The stacked network $\underline{\mathcal{N}}$.

Figure 2.1: The stacked network $\underline{\mathcal{N}}$ for a given network $\mathcal{N}$

2.3 Stacked network

Following the proof method employed in [10], we define the stacked network as follows (See Figure 2.3).

Let $\mathcal{N} = (\mathcal{G}, \mathcal{C})$ be a network with vertex set $\mathcal{V} = \{1, \ldots, m\}$ and edge set E. For each $e \in E$, let $\underline{P}_e$ be a probability distribution on $(\mathcal{Z}^{(e)})^N$ obtained by forming an N-fold product of P_e with itself, i.e., $\underline{P}_e(\underline{z}_e) = \prod_{i=1}^{N} P_e(\underline{z}_e(i))$ for all $\underline{z}_e \in (\mathcal{Z}^{(e))^N}$. Next, let $\underline{\Upsilon}_e : (\mathcal{X}^{(e)})^N \times (\mathcal{Z}^{(e)})^N \to (\mathcal{Y}^{(e)})^N$ represent a channel that maps pairs $(\underline{x}_e, \underline{z}_e) \in (\mathcal{X}^{(e)})^N \times \in (\mathcal{Z}^{(e)})^N$ to $\underline{\Upsilon}_e(\underline{x}_e, \underline{z}_e) = [\Upsilon_e(\underline{x}_e(1), \underline{z}_e(1)), \ldots, \Upsilon_e(\underline{x}_e(1), \underline{z}_e(N))]^T$. We define the N-fold stacked network $\underline{\mathcal{N}}$ derived from $\mathcal{N} = (\mathcal{G}, \mathcal{C})$ as a pair $(\underline{\mathcal{G}}, \underline{\mathcal{C}})$, where, $\underline{\mathcal{G}} \triangleq \mathcal{G}$ and $\underline{\mathcal{C}}_e \triangleq ((\mathcal{X}^{(e)})^N, (\mathcal{Y}^{(e)})^N, (\mathcal{Z}^{(e)})^N, \underline{P}_e, \underline{\Upsilon}_e)$ for all $e \in E$.

For the network $\underline{\mathcal{N}}$, we denote the messages corresponding to the pair $(u, V) \in \mathcal{M}$ by matrix $\underline{W}_{1:nR(u,V)}^{(u \to V)}$, and the transmitted, received, and noise values for the edge $(u, v) \in E$ by matrices $\underline{X}_{1:n}, \underline{Y}_{1:n}$, and $\underline{Z}_{1:n}$ respectively. Let $\underline{\mathcal{N}}(1), \underline{\mathcal{N}}(2), \ldots, \underline{\mathcal{N}}(N)$ be N copies of the network $\mathcal{N}$. For each $i = 1, \ldots, N$, associate vector $\underline{W}_{1:nR(u,V)}^{(u \to V)}(i)$ with the message corresponding to the pair $(u, V) \in \mathcal{M}$, and $\underline{X}_{1:n}(i), \underline{Y}_{1:n}(i)$, and $\underline{Z}_{1:n}(i)$, with the messages, and transmitted, received, and noise values, respectively, for the edge $(u, v) \in E$ in $\underline{\mathcal{N}}(i)$.

For ease of visualization, we think of $\underline{\mathcal{N}}$ as a stack with layers $\underline{\mathcal{N}}(1), \underline{\mathcal{N}}(2), \ldots, \underline{\mathcal{N}}(N)$ and infinite

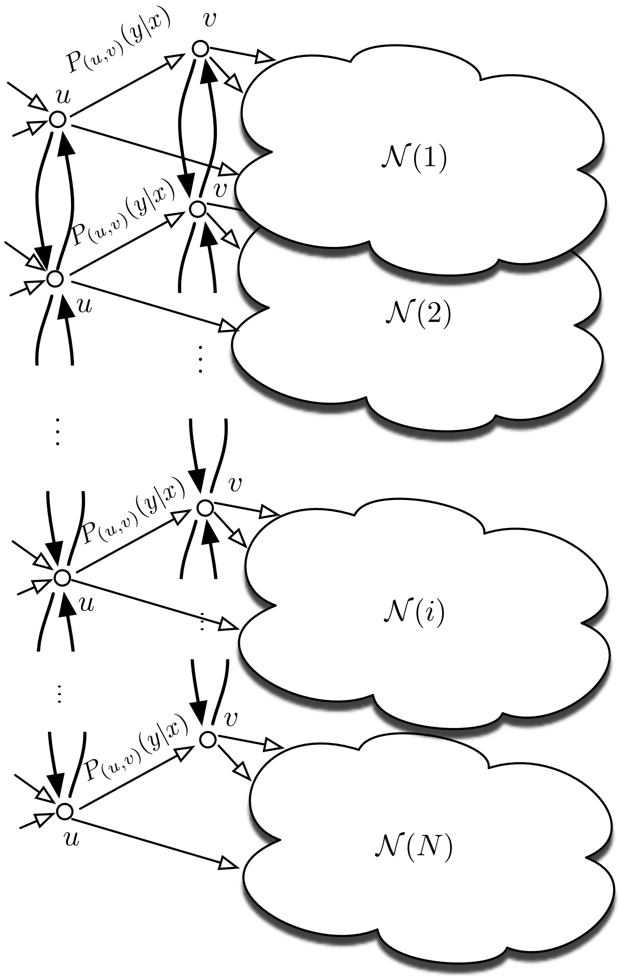


Figure 2.2: Visualizing $\underline{\mathcal{N}}$ as a stack with N layers

capacity bidirectional edges connecting all N copies of a giveb vertex $v \in \mathcal{V}$ to each other as shown in Figure 2.2. Thus, for each $v \in \mathcal{V}$ and $i = 1, \ldots, N$, the transmitted vector $\underline{X}_{1:n}^{(v,*)}(i)$ may be a function of all messages $(\underline{W}_{1:nR(v,U)}^{(v\to U)} : (v, U) \in \mathcal{M})$ and received vectors $\underline{Y}_{1:n}^{(*,v)}$.

The capacity region for the stacked network $\mathcal{R}(\underline{\mathcal{N}})$ is normalized by the number of layers N. In [10], it is shown that the capacity regions for $\mathcal{N}$ and $\underline{\mathcal{N}}$ are equal when none of the edges are corruptible by the adversary. Even though the presence of an adversary changes the network capacity, the arguments of lemma 1 of [10] extend readily to our setup. We state this in the following Lemma without proof.

***Lemma* 1.** *For any network* $\mathcal{N} = (\mathcal{G}, \mathcal{C})$, $\mathcal{R}(\mathcal{N}) = \mathcal{R}(\underline{\mathcal{N}})$.

Next, we show that there exists a sequence of solutions to the stacked network such that the error probability decays exponentially with the number of layers. In the non-adversarial case, the mutual independence of $(\underline{\mathbf{Z}}(i) : i = 1, \ldots, N)$ results in independent decoding errors for a solution that operates on each layer independently. Thus, applying a randomly generated error-correcting code to all messages $(\underline{W}^{(u\to V)}_{1:nR(u,V)} : (u,V) \in \mathcal{M})$ before they are processed by the network code ensures an exponential decay of error probability [25]. However, in the presence of an adversary, decoding errors may no longer be independent across the layers. We overcome this difficulty by first designing a solution to the stacked network for which the error probability is maximum when decoding errors are statistically independent across layers, and then showing that, under this condition, the error probability for this solution decays exponentially in the number of layers.

***Theorem* 1.** *Given any* $\mathbf{R} \in int(\mathcal{R}(\mathcal{N}))$*, there exists a* $(2^{-N\delta}, \mathbf{R})$*-solution* $\tilde{\mathcal{S}}(\underline{\mathcal{N}})$ *for* $\underline{\mathcal{N}}$ *for some* $\delta > 0$ *and for all* N *large enough.*

Proof. Let $\lambda > 0$ and let $\rho > H(2\lambda)$, where $H(\cdot)$ is the binary entropy function. Let $\mathcal{S}(\mathcal{N})$ be a $(\lambda, \mathbf{R})$-solution for $\mathcal{N}$ with $\mathcal{W} = \prod_{(u,V)\in\mathcal{M}} \mathcal{W}^{(u\to V)} = \{0,1\}^{nR(u,V)}$. We design the solution $\tilde{\mathcal{S}}(\underline{\mathcal{N}})$ as follows. For each $(u,V) \in \mathcal{M}$, let $\underline{w}^{(u\to V)}_{1:nR(u,V)}$ be a two-dimensional binary $(1-\rho)N \times nR(u,V)$ matrix. We first encode $\underline{w}^{(u\to V)}$ by using a different error-correcting code for each column. Next, we transmit each row of the resulting binary matrices $(\underline{\tilde{w}}^{(u\to V)}_{1:nR(u,V)} : (u,V) \in \mathcal{M})$ on a different layer using the solution $\mathcal{S}(\mathcal{N})$. Finally, we employ nearest-neighbor decoding at each node to reconstruct the messages.

Code Construction: Fix a pair $(u,V) \in \mathcal{M}$ and $k \in \{1, \ldots, nR(u,V)\}$. Consider a binary symmetric channel $\tilde{\mathcal{C}}$ with crossover probability 2λ. Let $\Psi^{(u\to V)}_k : \{0,1\}^{(1-\rho)N} \to \{0,1\}^N$ and $\Phi^{(u\to V)}_k : \{0,1\}^N \to \{0,1\}^{(1-\rho)N}$ be the encoder and decoder mappings for an error-correcting code for $\tilde{\mathcal{C}}$ of blocklength N and rate $(1-\rho)$ that is designed randomly as follows.

Select $\tilde{\mathcal{W}}^{(u\to V)}_k \subseteq \{0,1\}^N$ of size $2^{(1-\rho)N}$ by independently picking each element of $\tilde{\mathcal{W}}^{(u\to V)}_k$ from $\{0,1\}^N$ using a uniform distribution. The encoder $\Psi^{(u\to V)}_k$ maps each message $\underline{b} \in \{0,1\}^{(1-\rho)N}$ to a unique codeword $\underline{\tilde{b}} \in \tilde{\mathcal{W}}^{(u\to V)}_k$. The decoder $\Phi^{(u\to V)}_k$ maps each received vector $\hat{\underline{\tilde{b}}} \in \{0,1\}^N$ to the reconstruction $\underline{\hat{b}} \in \{0,1\}^{(1-\rho)N}$ that corresponds to the nearest valid codeword. Let $d_H(\cdot,\cdot)$ denote the Hamming distance between two binary vectors. In other words,

$$\Phi^{(u\to V)}_k(\hat{\underline{\tilde{b}}}) = \operatorname*{argmin}_{\underline{\hat{b}}\in\{0,1\}^{nR(u,V)}} d_H\big(\Psi^{(u\to V)}_k(\underline{\hat{b}}), \hat{\underline{\tilde{b}}}\big).$$

The encoding operation is shown in Figures 2.3.

We construct the solution $\tilde{\mathcal{S}}(\underline{\mathcal{N}})$ as follows. Let the message alphabet be $\underline{\mathcal{W}} = \prod_{(u,V)\in\mathcal{M}} \underline{\mathcal{W}}^{(u\to V)}$, where $\underline{\mathcal{W}}^{(u\to V)} \triangleq \{0,1\}^{nR(u,V)\times(1-\rho)N}$ is the set of $nR(u,V) \times (1-\rho)N$ binary matrices. Let $\underline{w}^{(u\to V)}_{1:nR(u,V)} \in \{0,1\}^{nR(u,V)\times(1-\rho)N}$ be the message intended for the connection $(u,V) \in \mathcal{M}$. The solution $\tilde{\mathcal{S}}(\underline{\mathcal{N}})$ performs the following sequence of operations:

1. For each pair of vertices $(u,V) \in \mathcal{M}$ and message $\underline{w}^{(u\to V)}_{1:nR(u,V)} \in \underline{\mathcal{W}}^{(u\to V)}$, let $\underline{\tilde{w}}^{(u\to V)}_{1:nR(u,V)} \in$

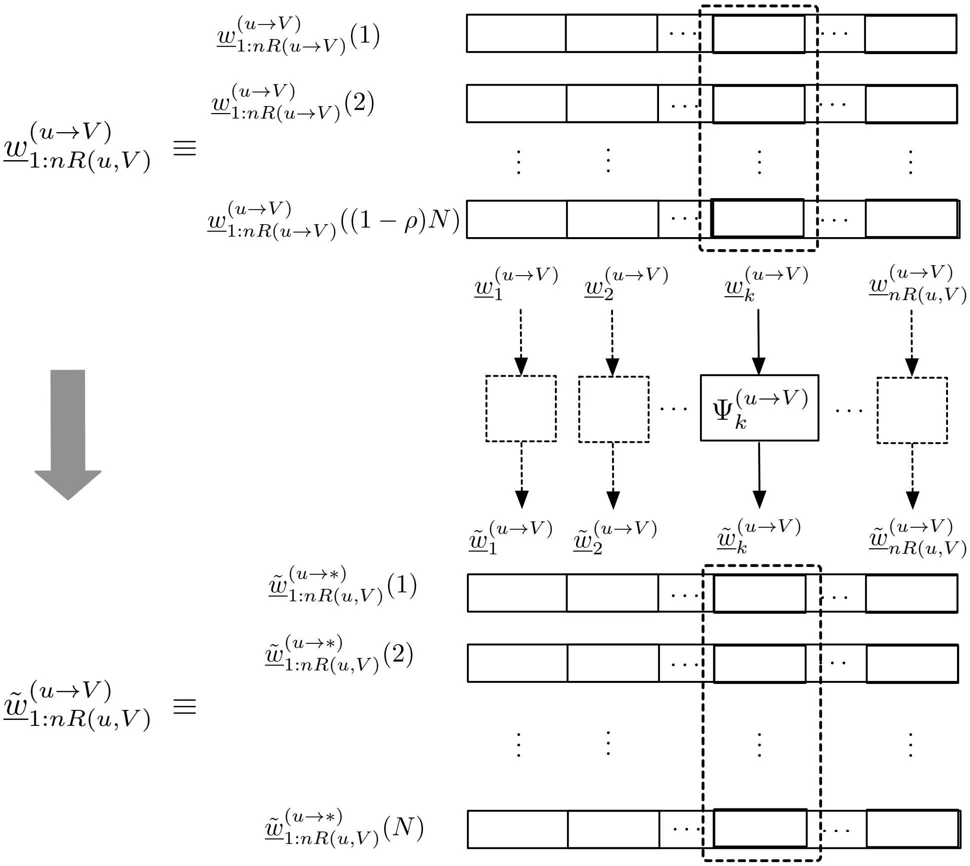


Figure 2.3: The encoding operation for each message

$\{0,1,\}^{nR(u,V)\times N}$ with

$$\tilde{\underline{w}}_k^{(u\to V)} = \Psi_k^{(u\to V)}\big(\underline{w}_k^{(u\to V)}\big)$$

for each $k = 1, 2, \ldots, nR(u,V)$. Thus, for each k, the code Ψ_k acts on the k-th column of the binary matrix $\underline{w}_{1:nR(u,V)}$. This is illustrated in Figure 2.3.

2. For each $i = 1, \ldots, N$, communicate messages $(\tilde{\underline{w}}^{(u\to V)}_{1:nR(u,V)}(i) : (u,V) \in \mathcal{M})$ using the solution $\mathcal{S}(\mathcal{N})$ on $\underline{\mathcal{N}}(i)$. Let $(\hat{\tilde{\underline{w}}}^{(u\to V)}_{1:nR(u,V)}(i) : (u,V) \in \mathcal{M})$ be the reconstructed messages after operating $\mathcal{S}(\mathcal{N})$ on $\underline{\mathcal{N}}(i)$. Figure 2.4(a) shows the layer-wise operation of $\mathcal{S}(\mathcal{N})$ on an edge (u,v) at time step $t < n$ and Figure 2.4(b) shows the layer-wise reconstruction at node u at the end of n time steps.

3. Finally, as shown in Figure 2.5, for every $(u,V) \in \mathcal{M}$, each vertex $v \in V$ outputs a reconstruction $\hat{\underline{w}}^{(u\to V,v)}_{1:nR(u,V)} \in \{0,1\}^{nR(u,V)\times(1-\rho)N}$ with

$$\hat{\underline{w}}_k^{(u\to V,v)} = \Phi_k^{(u\to V)}\big(\hat{\tilde{\underline{w}}}_k^{(u\to V,v)}\big)\ ,\ k = 1, \ldots, nR(u,V).$$

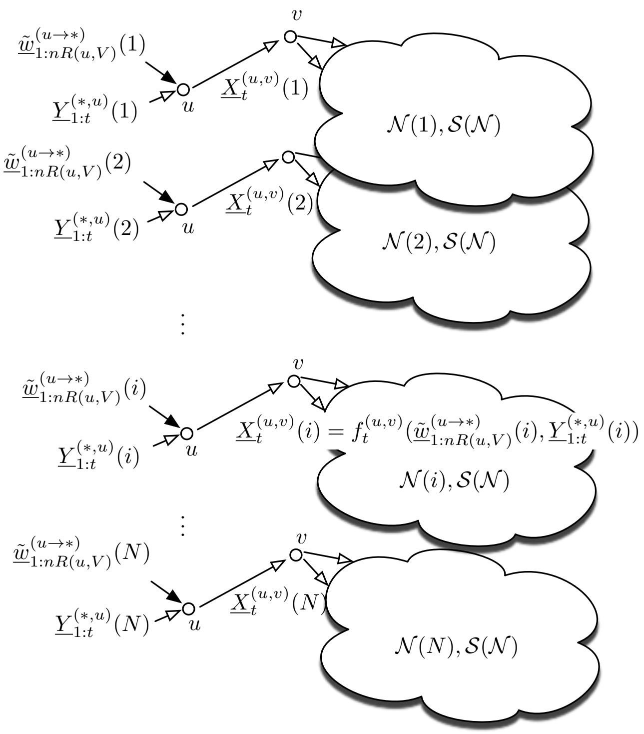
<u>*Analysis of error probability:*</u> Let $(u,V) \in \mathcal{M}$, $v \in V$ and $k \in \{1, \ldots, nR(u,V)\}$. Since $\tilde{\mathcal{C}}$ is symmetrical and the input is uniformly distributed, the decoder $\Phi_k^{(u\to V)}$ maps each received vector to the maximum likelihood estimate of the input given the received vector. By previous results on error exponents (See [25]), we know that such a code achieves an error probability of $2^{-N\tilde{\delta}}$ for some $\tilde{\delta} = \tilde{\delta}(\rho, \lambda)$, since the rate $1-\rho$ is less than the capacity $1 - H(2\lambda)$ of the channel $\tilde{\mathcal{C}}$. Denote the message and the received vector for the code $(\Psi_k^{(u\to V)}, \Phi_k^{(u\to V)})$ by random variables $\underline{B}$ and $\hat{\tilde{\underline{B}}}$, respectively. $\underline{B}$ is uniformly distributed on $\{0,1\}^{(1-\rho)N)}$ and $\hat{\tilde{\underline{B}}}$ is statistically related to $\Psi_k^{(u\to V)}(\underline{B})$ via the channel $\tilde{\mathcal{C}}$. Let $p_{\tilde{\mathcal{C}}}$ denote the joint distribution of the message $\underline{B}$ and the reconstruction $\Phi_k^{(u\to V)}(\hat{\tilde{\underline{B}}}$. Therefore,

$$p_{\tilde{\mathcal{C}}}(\underline{B} \neq \Phi_k^{(u\to V)}(\hat{\tilde{\underline{B}}})) < 2^{-N\tilde{\delta}}. \tag{2.1}$$

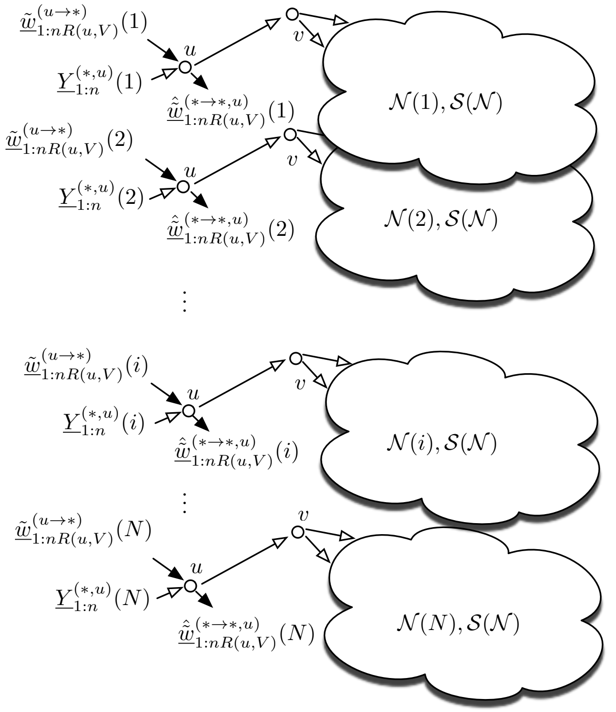
We now show that the block error probability for the code $(\Psi_k^{(u\to V)}, \Phi_k^{(u\to V)})$ over the channel $\tilde{\mathcal{C}}$ is an upper bound for the block error probability $\Pr\big(\underline{W}_k^{(u\to V)} \neq \hat{\underline{W}}_k^{(u\to V,v)}\big)$. Note that by the choice of the solution $\mathcal{S}(\mathcal{N})$, $\Pr(\underline{W}_k^{(u\to V)}(i) \neq \hat{\underline{W}}_k^{(u\to V,v)}(i)) < \lambda$ for each $i = 1, \ldots, N$. Thus,

$$\max_{b\in\{0,1\}} \Pr\big(\underline{W}_k^{(u\to V)}(i) \neq \hat{\underline{W}}_k^{(u\to V)}(i) | \underline{W}_k^{(u\to V)}(i) = b\big) < 2\lambda$$

for each $i = 1, \ldots, N$. Let $w_H(\cdot)$ denote the number of 1's in a binary vector. For every $\tilde{\underline{b}} \in \mathcal{W}_k^{(u\to V)}$, let $\pi_k^{(u\to V)}(\tilde{\underline{b}}) \subseteq \{0,1\}^N$ denote the set of all minimal weight error patterns that are decoded incorrectly by $\Phi_k^{(u\to V)}$, i.e., $\pi_k^{(u\to V)}(\tilde{\underline{b}})$ is the set of all vectors $\underline{e} \in \{0,1\}^N$ satisfying the following:

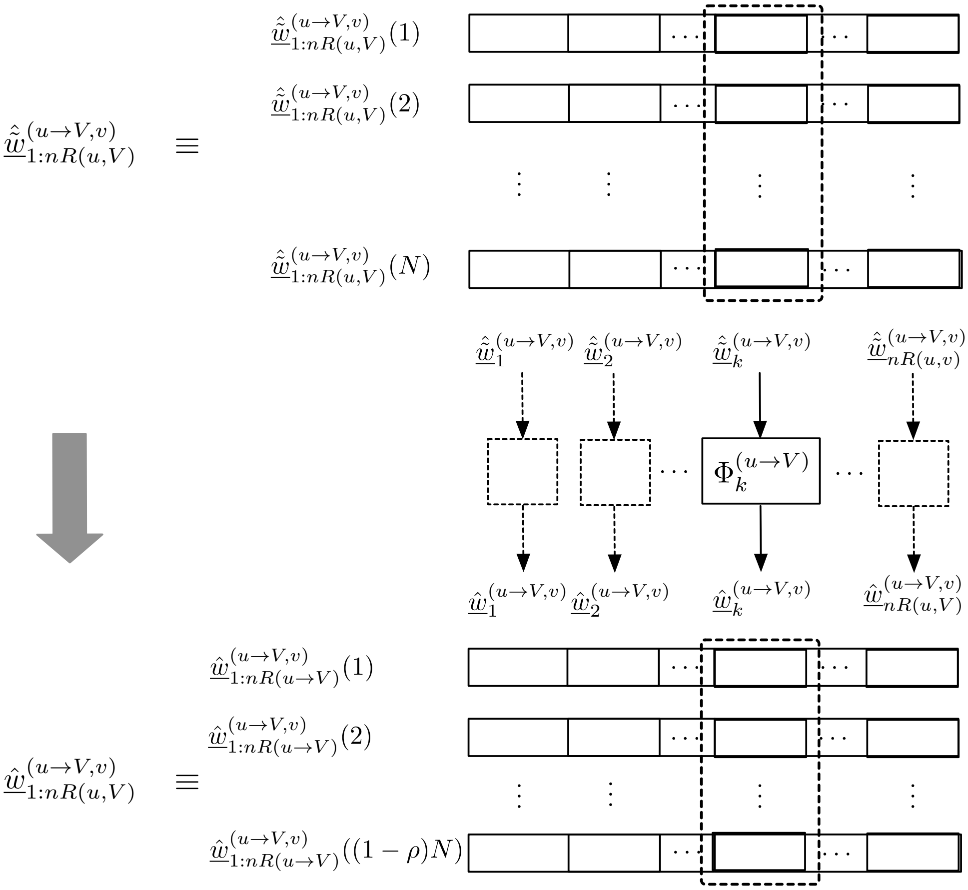


(a) The time-t transmitted vector $\underline{X}_t^{(u,v)}(i)$ on an edge (u, v) not attacked by the adversary in the i-th layer is a function only of the received vectors and input messages at the vertex u in the i-th layer until time-step t.



(b) The decoded messages for the i-th layer solution $\mathcal{S}(\mathcal{N})$ depend only on the received vectors and input messages seen at the vertex u in the i-layer over the entire blocklength n.

Figure 2.4: The layer-by-layer operation of the stacked solution



Functions $\Phi^{(u,V)}_{k}$ finally map the layer-by-layer reconstructions $\underline{\hat{\hat{w}}}^{(u\to V,v)})_{1:nR(u,V)}$ after performing the stacked solution to decoded messages $\underline{\hat{w}}^{(u\to V,v)})_{1:nR(u,V)}$ at each node v.

Figure 2.5: The decoder operation for each message

1. $\Phi_k^{(u\to V)}(\underline{\tilde{b}}) \neq \Phi_k^{(u\to V)}(\underline{\tilde{b}} \oplus \underline{e})$
2. $\Phi_k^{(u\to V)}(\underline{\tilde{b}}) = \Phi_k^{(u\to V)}(\underline{\tilde{b}} \oplus \underline{\tilde{e}})$ for all $\underline{\tilde{e}}$ such that $\{i : \tilde{e}(i) = 0\} \supseteq \{i : e(i) = 0\}$.

Let $\underline{c} \in \{0,1\}^N$. Then the probability that $\underline{W}_k^{(u,V)}$ is decoded incorrectly for some adversarial strategy on network $\mathcal{N}$ when the transmitted vector $\underline{\tilde{W}}_k^{(u,V)}$ is $\underline{c}$ is bounded from above by the probability of error when $\underline{\tilde{W}}_k^{(u,V)}$ is transmitted over the channel $\tilde{C}$ as follows:

$$\begin{aligned}
&\Pr(\underline{W}_k^{(u,V))} \neq \underline{\hat{W}}_k^{(u\to V,v)} | \underline{\tilde{W}}_k^{(u,V)} = \underline{c}) \\
&= \Pr(\Phi_k^{(u\to V)}(\underline{\tilde{W}}_k^{(u\to V)}) \neq \Phi_k^{(u\to V)}(\underline{\hat{\tilde{W}}}_k^{(u\to V,v)}) | \underline{\tilde{W}}_k^{(u\to V)} = \underline{c}) \\
&= \sum_{\underline{e}\in\pi_k^{(u\to V)}(\underline{c})} \Pr\big(\cap_{i:\underline{e}(i)=1} \{\underline{\hat{\tilde{W}}}_k^{u\to V,v}(i) = \underline{c}(i) \oplus \underline{e}(i)\} \big| \\
&\qquad\qquad\qquad\qquad\qquad\qquad \underline{\tilde{W}}_k^{(u\to V)} = \underline{c}\}\big) \\
&\leq \sum_{\underline{e}\in\pi_k^{(u\to V)}(\underline{c})} (2\lambda)^{w_H(\underline{e})} \qquad (2.2) \\
&= p_{\tilde{\mathcal{C}}}\big(\Phi_k^{(u\to V)}(\underline{c}) \neq \Phi_k^{(u\to V)}(\underline{\hat{\tilde{W}}}_k^{(u\to V,v)}) | \underline{\tilde{W}}_k^{(u,V)} = \underline{c}\big). \qquad (2.3)
\end{aligned}$$

Therefore,

$$\Pr(\underline{W}_k^{(u,V))} \neq \underline{\hat{W}}_k^{(u\to V,v)} | \underline{\tilde{W}}_k^{(u,V)} = \underline{c}) < 2^{-N\tilde{\delta}}.$$

The bound in (2.2) is a consequence of the fact that for each $i = 1, \ldots, N$ and $k = 1, \ldots, nR(u,V)$, whether or not the adversary's actions can cause the event $\{\underline{\tilde{W}}_k^{(u\to V)}(i) \neq \underline{\hat{\tilde{W}}}_k^{(u\to V)}(i)\}$ depends only on the noise values $\underline{\mathbf{Z}}_{1:n}(i)$ and the messages $\underline{\tilde{W}}(i)$, and can occur with probability at most 2λ under all possible adversarial actions. Step (2.3) follows from the fact that the transition probability for the channel $\tilde{\mathcal{C}}$ is 2λ. Finally, applying a union bound over all values of k and (u,V), we obtain

$$P_{\mathcal{E}}(\tilde{\mathcal{S}}) \leq n|\mathcal{M}|2^{-N\tilde{\delta}} \max_{(u,V)\in\mathcal{M}} R(u,V) \leq 2^{-N\delta}$$

for every $\delta < \tilde{\delta}$ for large enough N. □

2.4 Network equivalence

In this section, we prove that finding the capacity region of the network $\mathcal{N}$ is equivalent to finding the capacity of a network $\hat{\mathcal{N}}_R$ where one of the links $\hat{e} = (1,2)$ is replaced by a noiseless links of capacity $R = C(\hat{e})$. Koetter et al. [10] prove this by first showing that for every $R < C(\hat{e})$, the capacity region of the $\hat{\mathcal{N}}_R$ is a subset of that of $\mathcal{N}$. Next, they show that for every $R > C(\hat{e})$, the capacity region of $\mathcal{N}$ is a subset of that of $\hat{\mathcal{N}}_R$. We follow a similar proof outline for our case. The proof of the first part follows the arguments of [10] exactly. We state it without proof in the following theorem.

***Theorem* 2.** *For the networks $\mathcal{N}$ and $\hat{\mathcal{N}}_R$ defined above, $\mathcal{R}(\hat{\mathcal{N}}_R) \subseteq \mathcal{R}(\mathcal{N})$ if $R < C$.*

Next, to prove that $\mathcal{R}(\hat{\mathcal{N}}_R) \supseteq \mathcal{R}(\mathcal{N})$ if $R > C$, Koetter et al.argue that a noisy channel may be emulated on the lossless link $\hat{e}$ in the stacked network $\underline{\mathcal{N}}_R$ by using a randomly generated source code that operates across the layers. Their proof relies on typicality of the vector $\underline{X}_t^{(u,v)}$, which follows from the statistical independence of random variables corresponding to different layers. This assumption is not true in our case because the adversary may introduce dependence between different layers. To accommodate this possibility, we modify the proof of [10] to use a universal source code that first determines the type of the received sequence and then emulates the channel using a source code designed specifically for the observed type. We assume here that the alphabet $\mathcal{X}^{(\hat{e})}$ is finite.

***Theorem* 3.** *For the networks $\mathcal{N}$ and $\hat{\mathcal{N}}_R$ defined above, $\mathcal{R}(\hat{\mathcal{N}}_R) \supseteq \mathcal{R}(\mathcal{N})$ if $R > C$.*

Proof. First, we design a sequence of universal source codes $\{(\alpha_{N,t}, \beta_{N,t})\}_{t=1,2,\ldots}$, each at rate $\hat{R} \in (C(\hat{e}), R)$, that operate by describing the type of the input followed by the index of the codeword picked from a codebook designed for that type. Next, we modify a given $(2^{-N\delta}, \mathbf{R})$-solution $\mathcal{S}(\underline{\mathcal{N}})$ for $\underline{\mathcal{N}}$ to a solution $\hat{\mathcal{S}}(\underline{\mathcal{N}}_R)$ for $\underline{\mathcal{N}}_R$ by emulating the channel on the link $\hat{e}$ by the source codes $\{(\alpha_{N,t}, \beta_{N,t})\}_{t=1,2,\ldots}$. Finally, we conclude that the error probability for the solution $\hat{\mathcal{S}}(\underline{\mathcal{N}}_R)$ vanishes as N grows without bound.

Construction of $\hat{\mathcal{S}}(\underline{\mathcal{N}}_R)$: Let $\underline{\mathcal{S}}(\underline{\mathcal{N}})$ be a $(2^{-N\delta}, \mathbf{R})$-solution to $\underline{\mathcal{N}}$ implemented over n time steps and obtained from the solution $\mathcal{S}(\mathcal{N})$ by following the construction in Theorem 1. We modify $\underline{\mathcal{S}}(\underline{\mathcal{N}})$ to obtain a solution $\hat{\mathcal{S}}(\underline{\mathcal{N}}_R)$ by first designing a sequence of universal source codes $\{\alpha_{N,t}, \beta_{N,t}\}_{t=1,2,\ldots}$ that operate at a rate $\hat{R} \in (C(\hat{e}), R)$ and next appending the above source code to $\underline{\mathcal{S}}(\underline{\mathcal{N}})$ to emulate the channel conditional probability $p_{\hat{e}}$ on the the channel $\hat{\mathcal{C}}_{\hat{e}}$ in network $\hat{\underline{\mathcal{N}}}_R$.

Design of $\{\alpha_{N,t}, \beta_{N,t}\}_{t=1,2,\ldots}$: Let $\mathcal{P}_N$ be the set of all types of N length sequences from $\mathcal{X}^{(\hat{e})}$. For a vector $\underline{x} \in (\mathcal{X}^{(\hat{e})})^N$, let $\hat{Q}(\underline{x}) \in \mathcal{P}_N$ denote the empirical distribution of $\underline{x}$. Let $Q \in \mathcal{P}_N$. Let (X_Q, Y_Q) be random variables jointly distributed on $\mathcal{X}^(\hat{e}) \times \mathcal{Y}^(\hat{e})$ such that $\Pr(X_Q = x) = Q(x)$ and the conditional distribution of Y_Q given X_Q is $p_{\hat{e}}$. For each $t = 1, \ldots, n$, select $\mathcal{B}_t^Q \subseteq (\mathcal{Y}^(\hat{e}))^N$ by choosing $2^{N\hat{R}}$ elements uniformly at random from $A_\epsilon^{(N)}(Y_Q)$.

For $t = 1, \ldots, n$, let the encoder

$$\alpha_{N,t} : (\mathcal{X}^{(\hat{e})})^N \to \{0,1\}^{|\mathcal{X}^{(\hat{e})}| \log_2 (N+1)} \times \{1, \ldots, 2^{N\hat{R}}\}$$

consist of the pair of maps to $(\alpha_{\mathcal{P}}, \alpha_{\mathcal{B}})$, where, for each $\underline{x} \in (\mathcal{X}^{(\hat{e})})^N$, $\alpha_{\mathcal{P}}(\underline{x})$ is the binary description of $\hat{Q}(\underline{x})$, and $\alpha_{\mathcal{B}}(\underline{x})$ is the index of some vector $\underline{\hat{y}} \in \mathcal{B}_t^{\hat{Q}(\underline{x})}$ such that $(\underline{x}, \underline{\hat{y}}) \in A_\epsilon^{(N)}(X_{\hat{Q}(\underline{x})}, Y_{\hat{Q}(\underline{x})})$. If no such vector $\underline{\hat{y}}$ exists, then $\alpha_{\mathcal{B}}(\underline{x})$ is set to be 1. Next, let the decoder

$$\beta_{N,t} : \{0,1\}^{|\mathcal{X}^{(\hat{e})}|\log_2(N+1)} \times \{1,\ldots,2^{N\hat{R}}\} \to (\mathcal{Y}^{(\hat{e})})^N$$

map pairs $(x_{\mathcal{P}}, x_{\mathcal{B}})$ to the vector with index $x_{\mathcal{B}}$ in $\mathcal{B}_t^Q$, where Q is the type described by $x_{\mathcal{P}}$.

Appending $\{\alpha_{N,t}, \beta_{N,t}\}_{t=1,2,\ldots}$ to $\underline{\mathcal{S}}(\mathcal{N})$: The solution $\hat{\mathcal{S}}(\hat{\mathcal{N}}_R)$ is identical to $\underline{\mathcal{S}}(\mathcal{N})$ except for the maps at nodes 1 and 2. For $(u,v) \in \mathcal{V} \times \mathcal{V}$, let $f^{(u,v)} : \underline{\mathcal{W}}^{(u\to *)} \times (\mathcal{Y}^{(*,u)})^{nN} \to (\mathcal{X}^{(u,*)})^{nN}$ denote the encoder that determines the codeword on the edge (u,v) and let $g^{(u)} : \underline{\mathcal{W}}^{(u\to *)} \times (\mathcal{Y}^{(*,u)})^{nN} \to \prod_{V\subseteq\mathcal{V}\setminus\{u\}} \underline{\mathcal{W}}^{(*\to\{u\}\cup V)}$ denote the decoder for the messages meant for node u in the solution $\underline{\mathcal{S}}(\underline{\mathcal{N}})$.

Let $\underline{\hat{X}}_{1:n}^{(e)} = \underline{X}_{1:n}^{(e)}$ and $\underline{\hat{Y}}_{1:n}^{(e)} = \underline{Y}_{1:n}^{(e)}$ for all $e \neq \hat{e}$. Let $\underline{\hat{X}}_{1:n}^{(\hat{e})} = f^{(1,2)}(\underline{W}^{1\to *}, \underline{\hat{Y}}_{1:n}^{(*,1)}))$ and $\underline{\hat{Y}}_t^{(\hat{e})} = \beta_{N,t-1}(\alpha_{N,t-1}(\underline{\hat{X}}_{t-1}^{(\hat{e})}))$. Let $\hat{\mathcal{S}}(\hat{\mathcal{N}}_R)$ be a solution with encoder and decoder mappings $(\hat{f}^{(u,v)} : (u,v) \in E)$ and $(\hat{g}^{(u)} : u \in \mathcal{V})$, where,

$$\hat{f}^{(u,v)}(\underline{W}^{(u\to *)}, \underline{\hat{Y}}_{1:n}^{(*,u)}) \triangleq \begin{cases} f^{(u,v)}(\underline{W}^{(u\to *)}, \underline{\hat{Y}}_{1:n}^{(*,u)}) \\ \quad \text{if } (u,v) \neq \hat{e} \text{ and } u \neq 2 \\ \alpha(f^{(u,v)}(\underline{W}^{(u\to *)}, \underline{\hat{Y}}_{1:n}^{(*,u)})) \\ \quad \text{if } (u,v) = \hat{e} \\ f^{(u,v)}(\underline{W}^{(u\to *)}, \underline{\hat{Y}}_{1:n}^{(*,u)}) \\ \quad \text{if } u = 2 \end{cases}$$

and

$$\hat{g}^{(u)}(\underline{W}^{(u\to *)}, \underline{Y}_{1:n}^{(*,u)}) \triangleq \begin{cases} g^{(u)}(\underline{W}^{(u\to *)}, \underline{Y}_{1:n}^{(*,u)}) \\ \quad \text{if } u \neq 2 \\ g^{(u)}(\underline{W}^{(u\to *)}, \underline{\hat{Y}}_{1:n}^{(*,u)}) \\ \quad \text{if } u = 2. \end{cases}$$

Analysis of error probability: Let $t = 1,\ldots,n$. Let $\mathbf{Z}_{1:n}^{[\hat{e}]} = (\underline{Z}_{1:n}^{(e)} : e \in E \setminus \{\hat{e}\})$ be the noise values on all edges except the edge $\hat{e}$ and let $\underline{W} = (\underline{W}^{(u\to V)} : (u,V) \in \mathcal{M})$ denote the collection of transmitted messages. Note that for $t = 1,\ldots,n$, $\underline{\hat{X}}_{1:t}^{(\hat{e})}$ is a deterministic function of $\underline{W}$, $\mathbf{Z}_{1:t-1}^{[\hat{e}]}$, and $\underline{\hat{Y}}_{1:t-1}^{(\hat{e})}$ while $\underline{\hat{Y}}_{1:t}^{(\hat{e})}$ is a random variable due to the random design of $\hat{\mathcal{S}}$. Let $\hat{p}_t$ denote the conditional probability distribution of $\underline{\hat{Y}}_t$ given $\underline{\hat{X}}_{t-1}$ under a random choice of $\hat{\mathcal{S}}$ as described above. By Lemma 11 in [10],

$$\hat{p}_t(\underline{\hat{y}}_t|\underline{\hat{x}}_{t-1}) \leq \prod_{i=1}^{N} p_{\hat{e}}(\underline{\hat{y}}_t(i)|\underline{\hat{x}}_{t-1}(i)) \cdot 2^{Na(\epsilon,N,t)}.$$

for every $(\underline{\hat{x}}_t, \underline{\hat{y}}_{t-1})$ such that $\underline{\hat{x}}_t \in (\mathcal{X}^(\hat{e}))^N$ and $(\underline{\hat{x}}_{t-1}, \underline{\hat{y}}_t) \in A_\epsilon^{(N)}(X_{\hat{Q}(\underline{x}_{t-1})}, Y_{\hat{Q}(\underline{x}_{t-1})})$. Further, since $R > C(\hat{e})$, by standard random coding arguments (e.g., proof of Rate Distortion Theorem, [26]), for large enough N,

$$p_t\big(\{\underline{\hat{y}}_t : (\underline{\hat{y}}_t, \underline{\hat{x}}_{t-1}) \notin A_\epsilon^{(N)}(X_{\hat{Q}(\underline{x}_{t-1})}, Y_{\hat{Q}(\underline{x}_{t-1})}\}|\underline{\hat{x}}_{t-1})\big) \leq \epsilon$$

Next, note that messages $\underline{w}$, noise values $\underline{\mathbf{z}}^{[\hat{e}]} = (\underline{z}^{(e)} : e \in E \setminus \hat{e})$, transmitted vector $\underline{x}_{1:n}^{(\hat{e})}$, and received vector $\underline{y}_{1:n}^{(\hat{e})}$ result in a decoding error under the solution $\underline{\mathcal{S}}$ if there exists $\underline{\mathbf{z}}^{(\hat{e})}$ such that $(\underline{w}, \underline{\mathbf{z}}^{[\hat{e}]}, \underline{\mathbf{z}}^{(\hat{e})}) \in \mathcal{E}(\underline{\mathcal{S}}(\underline{\mathcal{N}}))$, and $\underline{y}_t^{(\hat{e})} = \underline{\Upsilon}_{\hat{e}}(\underline{x}_{t-1}^{(\hat{e})}, \underline{z}_{t-1}^{(\hat{e})})$. Let $\hat{\mathcal{E}}(\underline{\mathcal{S}}, \underline{w}, \underline{\mathbf{z}}^{[\hat{e}]}) = \{\underline{\mathbf{z}}^{(\hat{e})} : (\underline{w}, \underline{\mathbf{z}}^{[\hat{e}]}, \underline{\mathbf{z}}^{(\hat{e})}) \in \mathcal{E}(\underline{\mathcal{S}}(\underline{\mathcal{N}}))\}$. The expected probability of a decoding error over the choice of $\hat{\mathcal{S}}(\underline{\hat{\mathcal{N}}}_R)$ for given values of $\underline{\mathbf{z}}_{1:n}^{[\hat{e}]}$ and $\underline{w}$ is given by

$$\begin{aligned}
&\int_{\hat{\mathcal{E}}(\underline{\mathcal{S}}, \underline{w}, \underline{\mathbf{z}}^{[\hat{e}]})} \left\{ \prod_{t=1}^{n} \hat{p}_t(\underline{\Upsilon}_{\hat{e}}(z_{t-1}^{(\hat{e})}) | \underline{\hat{x}}_{t-1}^{(\hat{e})}) \right\} d\underline{\mathbf{z}}_{1:n}^{(\hat{e})} \\
&\leq \int_{\hat{\mathcal{E}}(\underline{\mathcal{S}}, \underline{w}, \underline{\mathbf{z}}^{[\hat{e}]})} \left\{ \prod_{t=1}^{n} \prod_{i=1}^{N} \right. \\
&\qquad \left. p_{\hat{e}}(\Upsilon_{\hat{e}}(z_{t-1}^{(\hat{e})}(i), \hat{x}_{t-1}^{(\hat{e})}(i)) | \hat{x}_{t-1}^{(\hat{e})}(i)) 2^{Na(\epsilon,N,t)} \right\} d\underline{\mathbf{z}}_{1:n}^{(\hat{e})} + \epsilon.
\end{aligned}$$

Taking expectation over $\underline{W}$ and $\underline{\mathbf{Z}}_{1:n}^{[\hat{e}]}$,

$$\begin{aligned}
\mathbf{E}_{\hat{\mathcal{S}}, \underline{W}, \mathbf{Z}_{1:n}}\big[P_{\mathcal{E}}(\hat{\mathcal{S}})\big] &= \sum_{\underline{w}} \Big[2^{-nN \sum_{(u,V) \in \mathcal{M}} R(u,V)} \\
&\quad \int_{\underline{\mathbf{z}} : (\underline{\mathbf{z}}, \underline{w}) \in \mathcal{E}(\underline{\mathcal{S}})} \left\{ \prod_{t=1}^{n} \prod_{i=1}^{N} p_{\hat{e}}\big(\Upsilon(z_{t-1}^{(\hat{e})}(i), \hat{x}_{t-1}^{(\hat{e})}(i)) | \hat{x}_{t-1}^{(\hat{e})}(i)\big) \right. \\
&\qquad\qquad \left. 2^{Na(\epsilon,N,t)} P_E(\underline{\mathbf{z}}_{1:n}^{[\hat{e}]}) \right\} d\underline{\mathbf{Z}}_{1:n} \Big] + \epsilon \\
&= P_{\mathcal{E}}(\underline{\mathcal{S}}) \cdot 2^{nNc(\epsilon,N)} + \epsilon.
\end{aligned}$$

Since we assumed that $\underline{\mathcal{S}}$ is a $(2^{-N\delta}, \mathbf{R})$ solution, we get

$$\mathbf{E}_{\hat{\mathcal{S}}, \underline{W}, \mathbf{Z}_{1:n}}\big[P_{\mathcal{E}}(\hat{\mathcal{S}})\big] \leq 2^{-N\delta} \cdot 2^{nNa(\epsilon,N,t)} + \epsilon.$$

Finally, for a fixed value of n, let $\epsilon < 1/n$, and choose N large enough to conclude that $\mathbf{R} \in \mathcal{R}(\underline{\hat{\mathcal{N}}}_R)$. □

2.5 Equivalence for network source coding

In point-to-point source coding (Figure 1.2), all rates greater than the entropy of the source are sufficient to successfully describe the source losslessly. As a result, from the point of view of rate calculation, all sources with the same entropy are equivalent.

Note that the number of typical sequences of length n for a memoryless source W is roughly equal to $2^{nH(W)}$. Further, by the Asymptotic Equipartition Property (henceforth, AEP) [27], the observed sequence lies in the set of typical sequence with probability approaching one asymptotically. Thus, from a coding point of view, the notion of equivalence is further reinforced by noting that a code designed for a source $W^{(1)}$ can be applied to another source $W^{(2)}$ of the same entropy by first mapping typical sequences for $W^{(1)}$ to typical sequences for $W^{(2)}$ and then employing the code designed for $W^{(1)}$. This property is especially useful in systems with multi-step coding as it enables the system designer to design subsequent encoders independently of the exact source distribution. For example, in designing the channel code, it suffices to restrict the attention to sources that are uniformly distributed binary vectors of length equal to the binary entropy of the source.

This motivates us to examine whether a similar property holds for network source coding problems as well. Here, we restrict our attention to networks of error-free links. For a network $\mathcal{N}$ defined over the graph $\mathcal{G} = (\mathcal{V}, E)$ with sources $(W^{(v)} : v \in \mathcal{V})$ distributed according to a joint distribution $P_{\mathcal{V}}(\cdot)$, we define the achievable rate region $\mathcal{R}(\mathcal{N}) \subseteq \mathbf{R}^{|E|}$ as the set of rate vectors such that for each $\mathbf{R} = (R_e : e \in E) \in \mathcal{R}(\mathcal{N})$ and for every $\lambda > 0$, there exists a network source code that meets the desired demands with error probability at most λ. For a precise definition, the reader is referred to Chapters 4. To formalize notion of equivalence for this setup, let $\mathcal{N}_1$ and $\mathcal{N}_2$ be two networks defined on isomorphic graphs $\mathcal{G}_1 = (\mathcal{V}_1, E_1)$ and $\mathcal{G}_2 = (\mathcal{V}_2, E_2)$, respectively. Let $\mathcal{I} : \mathcal{V}_1 \to \mathcal{V}_2$ be the one-to-one map from $\mathcal{V}_1$ to $\mathcal{V}_2$ associated with the isomorphism of $\mathcal{G}_1$ and $\mathcal{G}_2$. For each $v \in \mathcal{V}_1 \cup \mathcal{V}_2$, let $W^{(v)}$ be a random variable observed at node v such that the collections of random variables $(W^{(v)} : v \in \mathcal{V}_1)$ and $(W^{(v)} : v \in \mathcal{V}_2)$ are distributed according to probability mass functions $P_{\mathcal{V}_1}$ and $P_{\mathcal{V}_2}$, respectively.

***Definition* 1** (Networks with Isomorphic Demands)**.** *We say that* $\mathcal{N}_1$ *and* $\mathcal{N}_2$ *are* networks with isomorphic demands *if the following are true:*

1. $\mathcal{N}_1$ *and* $\mathcal{N}_2$ *are defined on isomorphic graphs.*
2. *For each pair of vertices* $(u, v) \in \mathcal{V}_1$*, the source* $W^{(u)}$ *is losslessly present as a lossless demand at vertex* v *in* $\mathcal{N}_1$ *if and only if* $W^{\mathcal{I}(u)}$ *is present as a lossless demand at vertex* $\mathcal{I}(v)$ *in* $\mathcal{N}_2$*.*

For each $i \in \{1, 2\}$, $\mathcal{H}(\mathcal{N}_i)$ denote the entropic vector for sources in $\mathcal{N}_i$, i.e.,

$$\mathcal{H}(\mathcal{N}_i) = (H(W^{(v)} : v \in V) : V \subseteq \mathcal{V}) \text{ for } i = 1, 2.$$

Thus, the entropic vector for a given network consists of the collection of all possible joint entropies corresponding to the sources associated with the network. Note that the entropic vector captures the AEP for multiple sources in much the same way as the entropy does for a single source. For example, the number of jointly typical sequences of length n corresponding to sources $(W^{(v)} : v \in V)$

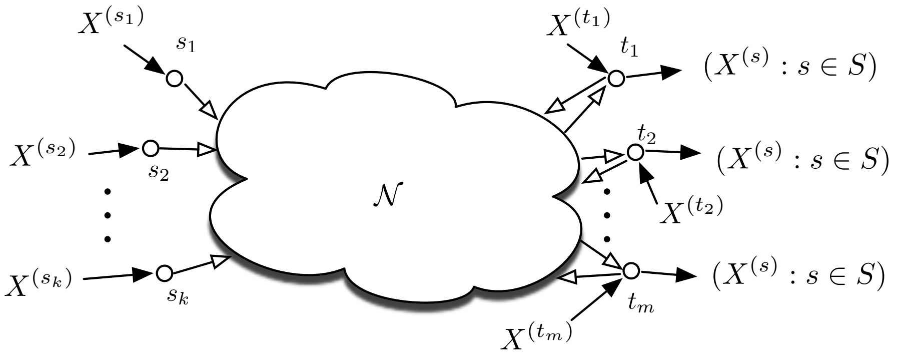


Figure 2.6: A multicast network with side information at the sinks

is approximately equal to $2^{nH(W^{(v)}:v\in V)}$ for every $V \subseteq \mathcal{V}$. Similarly, the probability of observing jointly typical sequences approaches one as the blocklength n grows without bound. Thus, it is tempting to believe that the following proposition holds:

***Proposition* 1.** *Let $\mathcal{N}_1$ and $\mathcal{N}_2$ be networks with isomorphic demands, and let $\mathcal{H}(\mathcal{N}_1) = \mathcal{H}(\mathcal{N}_2)$. Then,*

$$\mathcal{R}(\mathcal{N}_1) = \mathcal{R}(\mathcal{N}_2).$$

In the rest of the section, we examine the correctness of the above proposition under various conditions.

2.5.1 Multicast Demands with Side Information at Sinks

Let $\mathcal{N}_1$ and $\mathcal{N}_2$ be networks networks with isomorphic demands. Let $T_1 \subseteq \mathcal{V}_1$ and $T_2 \triangleq \{\mathcal{I}(v) : v \in T_1\}$ be the set of sink nodes in $\mathcal{N}_1$ and $\mathcal{N}_2$, respectively.

***Definition* 2** (Multicast demands with side information at the sinks)**.** *We say that $\mathcal{N}_1$ and $\mathcal{N}_2$ have multicast demands with side information at sinks if all sources in the collection $(W^{(v)} : v \notin T_1)$ (resp. $(W^{(v)} : v \notin T_2)$) is demanded at all vertices in T_1 (resp. T_2). See Figure 2.6 for an example of a multicast network.*

Note that the above definition allows side information at the sink nodes. The rate region for this setup is characterized in Chapter 5. As a corollary of this characterization, Proposition 1 follows. We state this as the following lemma.

***Lemma* 2.** *Let $\mathcal{N}_1$ and $\mathcal{N}_2$ be multicast networks with isomorphic demands and let $\mathcal{H}(\mathcal{N}_1) = \mathcal{H}(\mathcal{N}_2)$. Then, $\mathcal{R}(\mathcal{N}_1) = \mathcal{R}(\mathcal{N}_2)$.*

Proof. The proof follows directly from the fact that that rate regions for both $\mathcal{N}_1$ and $\mathcal{N}_2$ are fully characterized in terms of their respective source entropies. □

2.5.2 Networks with independent sources

Let $\mathcal{N}_1$ and $\mathcal{N}_2$ be networks with isomorphic demands and independent sources satisfying the conditions of Proposition 1. By Corollary 2.3.3 of [28], it follows that the above proposition holds.

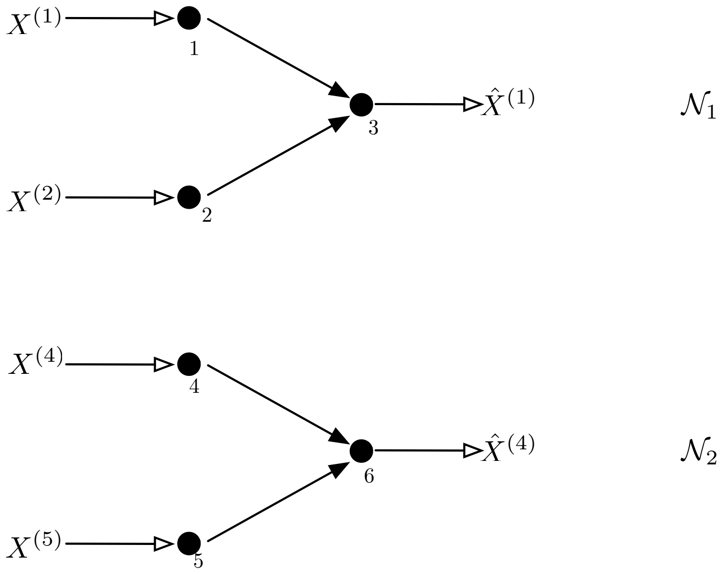


Figure 2.7: Two isomorphic networks with coded side information

2.5.3 An example where equivalence does not apply

Next, we show that 1 does not hold for all source distributions and for all demand scenarios. In particular, we consider two instances of the Ahlswede-Körner network of [29] and show that the rate regions for these are different even though the entropic vectors are the same.

***Example* 3.** *Consider the coded-side information networks $\mathcal{N}_1$ and $\mathcal{N}_2$ shown in Figure 2.7.*

Let $W^{(1)}$, $W^{(2)}$, $W^{(4)}$, and $W^{(5)}$ be random variables defined on alphabets $\mathcal{W}^{(1)}$, $\mathcal{W}^{(2)}$, $\mathcal{W}^{(4)}$, and $\mathcal{W}^{(5)}$ respectively. Let each of these alphabets be equal to $\{w_1, w_2, w_3, w_4\}$. Let $W^{(1)}$ and $W^{(4)}$ be distributed uniformly over their respective source alphabets. Let $W^{(2)}$ be jointly distributed with $W^{(1)}$ according to the conditional probability distribution

$$p_{W^{(2)}|W^{(1)}}(w_j|w_i) = \begin{cases} 1/2 & \text{if } i,j \in \{1,2\} \\ 1/2 & \text{if } i,j \in \{3,4\} \\ 0 & \text{otherwise.} \end{cases}$$

Let $W^{(5)}$ be jointly distributed with $W^{(4)}$ according to the conditional probability distribution

$$p_{W^{(5)}|W^{(4)}}(w_j|w_i) = \begin{cases} 1/2 & \text{if } j = i \\ 1/2 & \text{if } j = i+1 (mod\ 4) \\ 0 & \text{otherwise.} \end{cases}$$

Consider network $\mathcal{N}_1$ consisting of the sources $W^{(1)}$ and $W^{(2)}$ observed at vertices 1 and 2, respectively, and lossless demand $W^{(1)}$ at vertex 3. Similarly, network $\mathcal{N}_2$ consists of the sources $W^{(4)}$ and $W^{(5)}$ observed at vertices 4 and 5, respectively, and demand $W^{(4)}$ at vertex 6

By [29], $\mathcal{R}(\mathcal{N}_1)$ is the set of rate pairs $(R_{(1,3)}, R_{(2,3)})$ such that

$$R_{(1,3)} \geq H(W^{(1)}|U) \tag{2.4}$$

$$R_{(2,3)} \geq I(W^{(2)}; U) \tag{2.5}$$

for some random variable U forming the Markov chain $W^{(1)} \to W^{(2)} \to U$. Similarly, $\mathcal{R}(\mathcal{N}_2)$ is the set of rate pairs $(R_{(4,6)}, R_{(5,6)})$ such that

$$R_{(4,6)} \geq H(W^{(4)}|V) \tag{2.6}$$

$$R_{(5,6)} \geq I(W^{(5)}; V) \tag{2.7}$$

for some random variable V forming the Markov chain $W^{(4)} \to W^{(5)} \to V$.

Claim: *There exist random variables $(W^{(1)}, W^{(2)})$ and $(W^{(4)}, W^{(5)})$ satisfying:*

$$\begin{aligned} & H(W^{(1)}) = H(W^{(4)}), \\ & H(W^{(2)}) = H(W^{(5)}), \\ \text{and} \qquad & H(W^{(1)}, W^{(2)}) = H(W^{(4)}, W^{(5)}), \end{aligned}$$

such that $\mathcal{R}(\mathcal{N}_1) \neq \mathcal{R}(\mathcal{N}_2)$.

Proof. Let $W^{(1)}$ and $W^{(4)}$ be uniformly distributed on $\{1, 2, 3, 4\}$. Let $W^{(2)}$ and $W^{(5)}$ be random variables taking values in $\{1, 2, 3, 4\}$ and related to $W^{(1)}$ and $W^{(4)}$ through the transition probabilities $P_{W^{(2)}|W^{(1)}}(\cdot|\cdot)$ and $P_{W^{(5)}|W^{(4)}}(\cdot|\cdot)$ shown in Figures 2.8 and 2.9 . It is easy to verify that $H(W^{(1)}) = H(W^{(4)}) = H(W^{(2)}) = H(W^{(5)}) = 2$ and $H(W^{(1)}, W^{(2)}) = H(W^{(4)}, W^{(5)}) = 3$.

By (2.4) and (2.5), for every $(R_{(4,6)}, R_{(5,6)}) \in \mathcal{R}(\mathcal{N}_2)$,

$$R_{(4,6)} + R_{(5,6)} \geq H(W^{(4)}|V) + I(W^{(5)}|V) \tag{2.8}$$

for some V s.t. $W^{(4)} \to W^{(5)} \to V$ is a Markov chain. The first term on the right-hand side of the

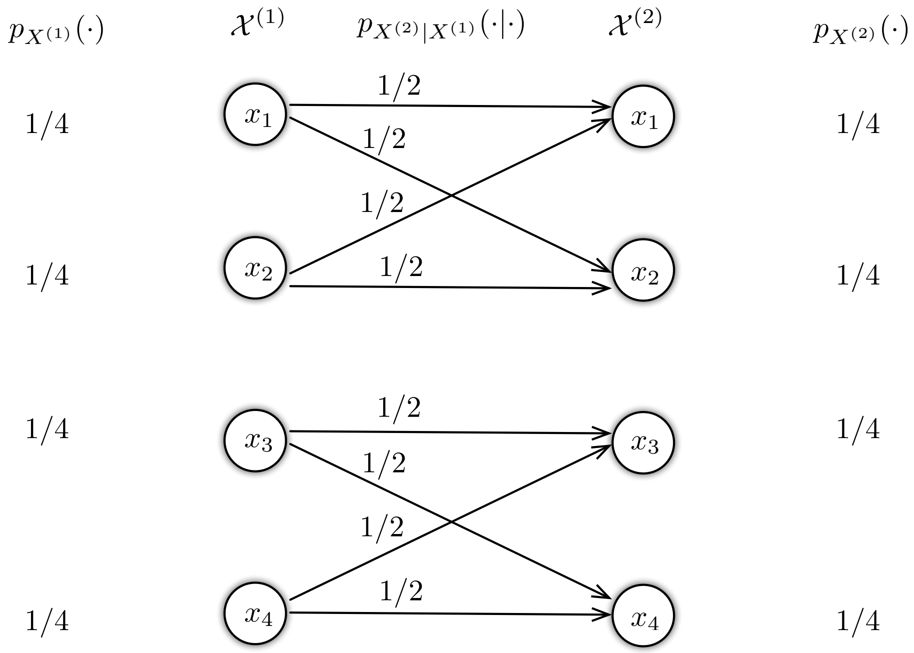


Figure 2.8: Transition probabilities for $W^{(2)}$ given $W^{(1)}$ for network $\mathcal{N}_1$

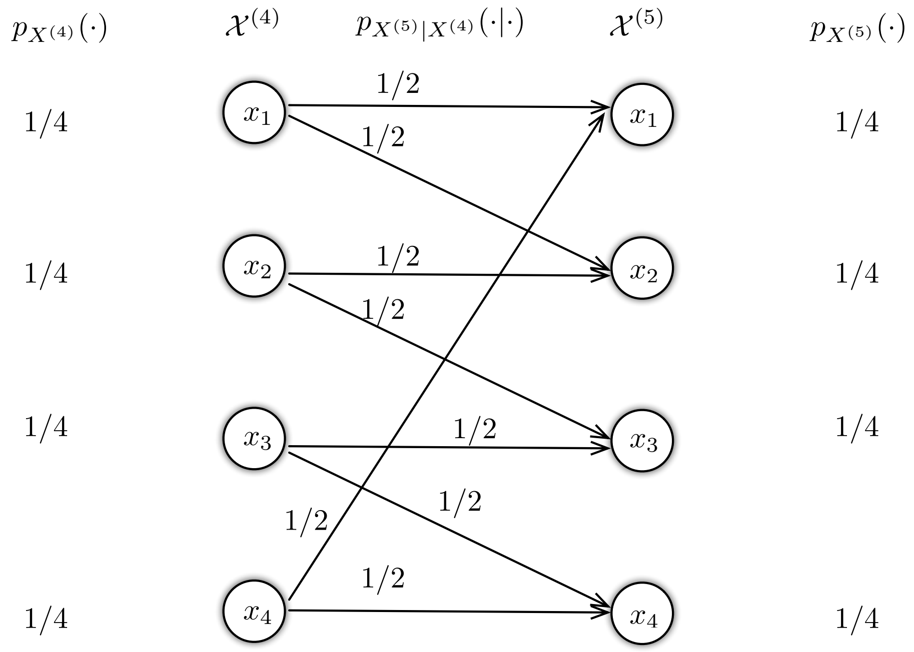


Figure 2.9: Transition probabilities for $W^{(5)}$ given $W^{(4)}$ for network $\mathcal{N}_2$

above equation can be lower bounded as follows:

$$\begin{aligned}
H(W^{(4)}|V) &= \sum_{i=1}^{4}\sum_{j=1}^{4} P_V(i)P_{W^{(4)}|V}(j|i)\log\frac{1}{P_{W^{(4)}|V}(j|i)}\\
&= \sum_{i=1}^{4}\sum_{j=1}^{4} P_V(i)\Big(\frac{1}{2}P_{W^{(5)}|V}(j|i)+\frac{1}{2}P_{W^{(5)}|V}(j(\text{mod } 4)+1|i)\Big)\\
&\qquad\qquad \log\frac{1}{\Big(\frac{1}{2}P_{W^{(5)}|V}(j|i)+\frac{1}{2}P_{W^{(5)}|V}(j(\text{mod } 4)+1|i)\Big)}\\
&\geq 1+\frac{1}{2}\sum_{i=1}^{4}\sum_{j=1}^{4} P_V(i)P_{W^{(5)}|V}(j|i)\\
&\qquad \log\frac{1}{\Big(P_{W^{(5)}|V}(j|i)+\frac{1}{2}P_{W^{(5)}|V}(j(\text{mod } 4)+1|i)\Big)\Big(P_{W^{(5)}|V}(j(\text{mod } 4)+3|i)\Big)}.
\end{aligned}$$

Now, consider rate points $(R_{(4,6)}, R_{(5,6)})$ for which $R_{(4,6)} \leq 1$. Note that when $H(W^{(4)}|V) = 1$, the above implies that $P_{W^{(5)}|V}(j|i) = 0$ or 1. Thus, $I(W^{(5)}; V) = 2$. Therefore, if $R_{(4,6)} \leq 1$,

$$R_{(4,36} + R_{(5,6)} \geq 3. \tag{2.9}$$

Next, note that $(1,1) \in \mathcal{R}(\mathcal{N}_1)$. This can be seen by choosing a binary valued U that is a deterministic function of $W^{(2)}$ as follows:

$$U = \begin{cases} 0 & \text{if } W^{(2)} \in \{1,2\} \\ 1 & \text{if } W^{(2)} \in \{3,4\}. \end{cases}$$

Thus,

$$H(W^{(1)}|U) = H(U) = 1,$$

and

$$I(W^{(2)}; U) = H(U) - H(U|W^{(2)}) = 1.$$

Clearly, $(1,1) \notin \mathcal{R}(\mathcal{N}_2)$, as it violates (2.9). Thus, $\mathcal{R}(\mathcal{N}_1) \neq \mathcal{R}(\mathcal{N}_2)$.

Chapter 3

The Decomposition Approach

3.1 Introduction

It is fair to say that the field of network coding has focused primarily on finding solutions for families of problems defined by a broad class of networks (e.g., networks representable by directed, acyclic graphs) and a narrow class of demands (e.g., multicast or multiple unicast demands). In this chapter, we investigate a family of network coding problems defined by a completely general demand structure and a narrow family of networks. Precisely, we give the complete solution to the problem of network coding with independent sources and arbitrary demands on a directed line network. We then generalize that solution to accommodate special cases of dependent sources.

Given independent sources and a special class of dependent sources, we fully characterize the capacity region of line networks for all possible demand structures (e.g., multiple unicast, mixtures of unicasts and multicasts, etc.). Our achievability bound is derived by first decomposing a line network into components that have exactly one demand and then adding the component rate regions to get rates for the parent network. Theorem 4 summarizes those results.

***Theorem* 4.** *Given an n-node line network $\mathcal{N}$ (shown in Figure 3.1) with memoryless sources $X_1, X_2, \ldots, X_n$ and demands $Y_1, Y_2, \ldots, Y_n$ satisfying $H(Y_i|X_1, \ldots, X_i) = 0$ for all $i \in \{1, \ldots, n\}$, the rate vector $(R_1, \ldots, R_{n-1})$ is achievable if and only if, for $1 \leq i \leq n-1$,*

$$R_i \geq \sum_{j=i+1}^{n} H(Y_j|X_{i+1}, \ldots, X_j, Y_{i+1}, \ldots, Y_{j-1}), \tag{3.1}$$

provided one of the following conditions holds:

A. Sources $X_1, X_2, \ldots, X_n$ are independent and for each $i \in \{1, 2, \ldots, n\}$, receiver i demands a subset of the sources $X_1, \ldots, X_i$.

B. Sources $X_1, \ldots, X_n$ have arbitrary dependencies and for each $i \in \{1, 2, \ldots, n\}$, either $Y_i =$constant, or $Y_i = (X_1, X_2, \ldots, X_n)$.

C. For each $i \in \{1, 2, \ldots, n\}$*, source* X_i *is any subset of independent sources* $W_1, \ldots, W_k$ *and demand* Y_i *is any subset of those* $W_1, \ldots, W_k$ *that appear in* $X_1, \ldots, X_i$*.*

For general dependent sources, we give an achievability result and provide examples where the result is and is not tight.

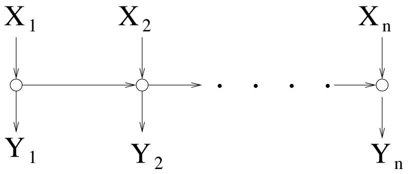


Figure 3.1: An n-node line network with sources $X_1, \ldots, X_n$ and demands $Y_1, \ldots, Y_n$

Lemmas 5, 6, and 7 of Sections 3.3.2, 3.3.3, and 3.3.4 give formal statements and proofs of this result under conditions A, B, and C, respectively. Case B is the multicast result of [18] generalized to many sources and specialized to line networks. We include a new proof of this result for this special case as it provides an important example in developing our approach.

Central to our discussion is a formal network decomposition described in Section 3.2. The decomposition breaks an arbitrary line network into a family of component line networks. Each component network preserves the original node demands at exactly one node and assumes that all demands at prior nodes in the line network have already been met. (See Fig. 3.2 for an illustration. Formal definitions follow in Section 3.2.) Sequentially applying the component network solutions in the parent network to meet first the first node's demands, then the second node's demands (assuming that the first node's demands are met), and so on, achieves a rate equal to the sum of the rates on the component networks. The given solution always yields an achievability result. The proofs of Lemmas 5, 6, and 7 additionally demonstrate that the given achievability result is tight under each of conditions A, B, and C.

Theorem 5 shows that the achievability result given by our additive solution is tight for an extremely broad class of sources and demands in 3-node line networks. In particular, this result allows arbitrary dependencies between the sources and also allows demands that can be functions of those sources (rather than simply the sources themselves).

The form of our solution lends insight into the types of coding sufficient to achieve optimal performance in the given families of problems. Primary among these are entropy codes including Slepian-Wolf codes for examples with dependent courses. These codes can be implemented, for example, using linear encoders and typical set or minimum entropy decoders [30]. The other feature illustrated by our decomposition is the need to retrieve information from the nearest preceding node where it is achievable (which may be a sink rather than a source), thereby avoiding sending multiple copies of the same information over any link (as can happen in pure routing solutions).

Unfortunately, the given decomposition fails to capture all of the information known to prior nodes in some cases, and thus the achievability result given by the additive construction is not tight in general. Theorem 6 gives a 3-node network where the bound is provably loose. The failure of additivity in this case arises from the fact that for the given functional source coding problem, the component network decomposition fails to capture other information that intermediate nodes can learn beyond their explicit demands. The same problem can also be replicated in a 4-node network where all sources (including the one previously described as a function) are available in the network.

3.2 Preliminaries

An *n-node line network* $\mathcal{N}$ is a directed graph (V, E) with $V = \{1, 2, \ldots, n\}$ and $E = \{(1, 2), (2, 3), \ldots, (n-1, n)\}$, as shown in Figure 3.1. Node i observes the source $X_i \in \mathcal{X}_i$ and requires the demand $Y_i \in \mathcal{Y}_i$. The random process $\{(X_1(i), X_2(i), \ldots, X_n(i), Y_1(i), Y_2(i), \ldots, Y_n(i))\}_{i=1}^{\infty}$ is drawn i.i.d. according to the probability mass function $p(X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_n)$. For $S \subseteq \{1, 2, \ldots, n\}$, X_S denotes the vector $(X_i : i \in S)$. A rate allocation for the n-node line network $\mathcal{N}$ is a vector $(R_i : i \in \{1, 2, \ldots, n-1\})$, where R_i is the rate on link $(i, i+1)$. We assume that there are no errors on any of the links. Line networks have been studied earlier in the context of reliable communication (e.g., [31]).

A *simple line network* is a line network with exactly one demand; thus Y_i is a constant at all but one node in the network. We next define the *component networks* $\mathcal{N}_1, \mathcal{N}_2, \ldots, \mathcal{N}_n$ for an n-node line network $\mathcal{N}$ with sources $X_1, X_2, \ldots, X_n$ and demands $Y_1, Y_2, \ldots, Y_n$. Figure 3.2 illustrates these definitions. For each $i \in \{1, 2, \ldots, n\}$, component $\mathcal{N}_i$ is an i-node simple line network. For each $j \in \{1, 2, \ldots, i-1\}$, the source and demand at node j of network $\mathcal{N}_j$ are $X_j^{(i)} = (X_j, Y_j)$ and $Y_j^{(i)} = c$, respectively; the source and demand at node i are $X_i^{(i)} = X_i$ and $Y_i^{(i)} = Y_i$, respectively.

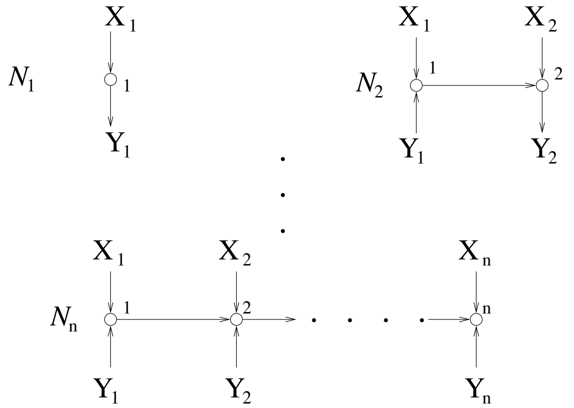


Figure 3.2: Component networks

3.3 Results

3.3.1 Cutset bounds and line networks

In this section, we first find a simple characterization of cutset bounds for line networks. Next, we show that whenever cutset bounds are tight for a 3-node line network, the achievable region for the network can be decomposed into a sum of two rate regions.

***Lemma* 3.** *For an n-node line network $\mathcal{N}$ with sources $(X_i : i = 1, \ldots, n)$ and demands $(Y_i : i = 1, \ldots, n)$, the cutset bounds are equivalent to the following set of inequalities:*

$$R_i \geq \max_{j \geq i+1} H(Y_{i+1}, \ldots, Y_j | X_{i+1}, \ldots, X_j) \; \forall 1 \leq i \leq n-1. \tag{3.2}$$

Proof. First consider any rate vector $(R_i : 1 \leq i \leq n-1)$ which satisfies all of the cutset bounds. Then, the inequalities in Equation (3.2) are satisfied, since they are a subset of the cutset bounds. Now, consider any rate vector $(R_i^* : 1 \leq i \leq n-1)$ that satisfies Equation (3.2). Let T be a cut, where each cut is expressed as a union of intervals, i.e., $T = \cup_{k=1}^{l} T(k)$ with $T(k) = \{m(k), \ldots, m(k) + l(k) - 1\} \subseteq \{1 \ldots, n\}$ and $m(k) > m(k-1) + l(k-1)$. Then,

$$\begin{aligned} \sum_{k=1}^{l} R^*_{m(k)-1} &\geq \sum_{k=1}^{l} \max_{j \geq 0} H(Y_{m(k)}, \ldots, Y_{m(k)+j} | X_{m(k)}, \ldots, X_{m(k)+j}) \\ &\geq \sum_{k=1}^{l} H(Y_{T(k)} | X_{T(k)}) \\ &\geq H(Y_T | X_T). \end{aligned} \tag{3.3}$$

Since the choice of the set T above was arbitrary, the set of inequalities of the type obtained in Equation (3.3) is exactly the set of cutset bounds. Thus, $(R_j^* : 1 \leq j \leq n-1)$ satisfies the cutset bounds.

□

***Lemma* 4.** *Let $\mathcal{N}$ be a 3-node line network for which the cutset bounds are tight on each of the component networks. Then, the achievable rate region for $\mathcal{N}$ is the set $\mathcal{R} = \{(R_1, R_2) : R_i > R_i^{(m)}\}$, where,*

$$\begin{aligned} R_1^{(m)} &= H(Y_2|X_2) + H(Y_3|X_2, Y_2, X_3), \\ R_2^{(m)} &= H(Y_3|X_3). \end{aligned}$$

Proof. Converse: Let C_1 and C_2 be any m-dimensional codes for the link $(1, 2)$ and $(2, 3)$ of the network $\mathcal{N}$ operating at rate (R_1, R_2) and satisfying the fidelity requirement

$$(1/m)H(Y_i(1), \ldots, Y_i(m)|X_i(1), \ldots, X_i(m), C_{i-1}) \leq \epsilon.$$

Then,

$$\begin{aligned} mR_2 &\geq H(C_2) \\ &\geq mH(Y_3|X_3), \end{aligned}$$

and

$$\begin{aligned} mR_1 &\geq H(C_1) \\ &\geq H(Y_2(1), \ldots, Y_2(m), C_2|X_2(1), \ldots, X_2(m)) \\ &= mH(Y_2|X_2) + H(C_2|X_2(1), \ldots, X_2(m), Y_2(1), \ldots, Y_2(m)). \end{aligned}$$

Now,

$$\begin{aligned} &H(C_2|X_2(1), \ldots, X_2(m), Y_2(1), \ldots, Y_2(m)) \\ &\quad\geq\quad I(C_2; Y_3(1), \ldots, Y_3(m)|(X_2, Y_2, X_3)(1), \ldots, (X_2, Y_2, X_3)(m)) \\ &\quad\geq\quad mH(Y_3|X_2, Y_2, X_3) - H(Y_3(1), \ldots, Y_3(m)|C_2, X_3(1), \ldots, X_3(m)) \\ &\quad\geq\quad mH(Y_3 X_2, Y_2, X_3) - m\epsilon. \end{aligned}$$

Thus, $R_1 \geq H(Y_2|X_2) + H(Y_3|X_2, Y_2, X_3) - \epsilon$ and $R_2 \geq H(Y_3|X_3$, implying that for codes which achieve arbitrarily high fidelity ($\epsilon \to 0$), $(R_1, R_2) \in \mathcal{R}$.

Achievability: Now, let $(R_1, R_2) \in \mathcal{R}$ s.t. $R_i \geq R_i^{(m)} + \epsilon$. Since the cutset bound is tight on the components $\mathcal{N}_2$ and $\mathcal{N}_3$, for sufficiently large m, there exist m-dimensional codes $C_1^{(2)}, C_1^{(3)}$, and $C_2^{(3)}$ for the links $(1, 2)$ in $\mathcal{N}_2$, $(1, 2)$ in $\mathcal{N}_3$, and $(2, 3)$ in $\mathcal{N}_3$, respectively, s.t.,

$$\frac{1}{m}H(C_1^{(2)}) \quad\leq\quad H(Y_2|X_2) + \epsilon/3, \tag{3.4}$$

$$\frac{1}{m}H(C_1^{(3)}) \quad\leq\quad H(Y_3|X_2, Y_2, X_3) + \epsilon/3, \tag{3.5}$$

$$\frac{1}{m}H(C_2^{(3)}) \quad\leq\quad H(Y_3|X_3) + \epsilon/3, \tag{3.6}$$

and $\frac{1}{m}H(Y_i(1), \ldots, Y_i(m)|X_i(1), \ldots, X_i(m), C_{i-1}^i) \leq \epsilon/3$ for $i = 2, 3$. Let $C_1 = C_1^{(2)}C_1^{(3)}$ and $C_2 = C_2^{(3)}$. Then, (C_1, C_2) is a code for $\mathcal{N}$ s.t. for $i = 2, 3$, $\frac{1}{m}H(C_i) < R_i$ for $i = 1, 2$, $\frac{1}{m}H(Y_i(1), \ldots, Y_i(m)|X_i(1), \ldots, X_i(m), C_{i-1}) < \epsilon$ and $H(C_2|C_1) < \epsilon$. Thus, (C_1, C_2) achieve the rate (R_1, R_2). □

3.3.2 Independent sources, arbitrary demands

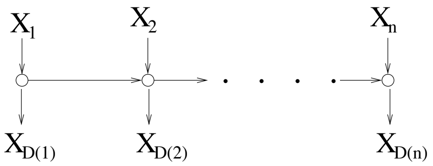


Figure 3.3: Line network with independent sources

***Lemma* 5.** *Let $\mathcal{N}$ be an n-node line network (shown in Figure 3.3) with $X_1, X_2, \ldots, X_n$ independently distributed and for $i \in \{1, 2, \ldots, n\}$, $Y_i = X_{D(i)}$ for some $D(i) \subseteq \{1, 2, \ldots, i\}$. Then, the rate region is fully characterized by the cutset bound. In particular, the rate allocation $(R_i : i \in \{1, 2, \ldots, n-1\})$ given by:*

$$R_i \geq H(X_{\cup_{j \geq k+1} D(i) \setminus \{k+1, \ldots, n\}})$$

is achievable.

Proof. First, let $(R_i : 1 \leq i < n)$ be any rate allocation for the network . Let $Q(i) \triangleq \cup_{j \geq i} D(j)$ and $Q(i,k) \triangleq \cup_{i \leq j \leq k} D(j)$. The cutset bound for the set $\{i, i+1, \ldots, k\}$ is

$$\begin{aligned} R_{i-1} &\geq H(X_{Q(i,k)} | X_i, X_{i+1}, \ldots, X_k) \\ &= H(X_{Q(i,k) \setminus \{i, i+1, \ldots, k\}}) \\ &= \sum_{j \in Q(i,k) \setminus \{i, i+1, \ldots, k\}} H(X_j). \end{aligned} \tag{3.7}$$

Now, since $D(j) \subseteq \{1, 2, \ldots, j\}$, it follows that $Q(i,k) \subseteq \{1, 2, \ldots, k\}$. The following argument shows that $Q(i, k+1) \setminus \{i, i+1, \ldots, k+1\} \supseteq Q(i,k) \setminus \{i, i+1, \ldots, k\}$:

$$\begin{aligned} & Q(i, k+1) \setminus \{i, i+1, \ldots, k+1\} \\ & \quad = (Q(i,k) \cup D(k+1)) \setminus \{i, i+1, \ldots, k+1\} \\ & \quad = (Q(i,k) \setminus \{i, i+1, \ldots, k+1\}) \\ & \qquad \cup (D(k+1) \setminus \{i, i+1, \ldots, k+1\}) \\ & \quad = (Q(i,k) \setminus \{i, i+1, \ldots, k\}) \\ & \qquad \cup (D(k+1) \setminus \{i, i+1, \ldots, k+1\}) \\ & \quad \supseteq Q(i,k) \setminus \{i, i+1, \ldots, k\}. \end{aligned} \tag{3.8}$$

It follows that the bound on R_{i-1} in Equation (3.7) is greatest when $k = n$. Hence,

$$R_{i-1} \quad \geq \quad \sum_{j \in Q(i) \setminus \{i, i+1, \ldots, n\}} H(X_j).$$

Next, we will demonstrate that this lower bound is achievable by coding for the component networks separately. To this end, consider the network $\mathcal{N}_j$. Since the demand at the end of the network is $X_{D(j)}$, calculating the demand across the various links starting from the very last link, we obtain the following rate allocation to be feasible:

$$\begin{aligned} R^j_{j-1} &= H(X_{D(j)\setminus\{j\}}) \\ R^j_{j-2} &= H(X_{(D(j)\setminus\{j\})\setminus(D(j-1)\cup\{j-1\})}). \\ &= H(X_{D(j)\setminus(\{j-1,j\}\cup Q(j-1,j-1))}). \\ &\vdots \\ \text{and} \qquad R^j_1 &= H(X_{D(j)\setminus(\{2,\ldots,j\}\cup Q(2,j-1))}). \end{aligned}$$

Now, adding the rates across the component networks for the link $(i, i+1)$,

$$\begin{aligned} \sum_{j=i+1}^{n} R^j_i &= \sum_{j=i+1}^{n} H(X_{D(j)\setminus(\{i+1,\ldots,j\}\cup Q(i+1,j-1))}) \\ &= \sum_{j=i+1}^{n} H(X_{D(j)\setminus(\{i+1,\ldots,n\}\cup Q(i+1,j-1))}) \\ &= \sum_{j=i+1}^{n} H(X_{(D(j)\setminus Q(i+1,j-1))\setminus\{i+1,\ldots,n\}}) \\ &= H(X_{(\cup_{j=i+1}^{n} D(j))\setminus\{i+1,\ldots,n\}}) \\ &= H(X_{Q(k+1)\setminus\{k+1,\ldots,n\}}) \\ &\geq R_i. \end{aligned}$$

Finally, Theorem 4-A follows from the observation that

$$H(X_{D(j)\setminus(\{i+1,\ldots,j\}\cup Q(i+1,j-1))}) = H(Y_j | X_{i+1}, \ldots, X_j, Y_{i+1}, \ldots, Y_{j-1}).$$

□

3.3.3 Dependent sources, multicast

***Lemma* 6.** *Let $\mathcal{N}$ be an n-node line network with arbitrarily dependent sources $X_1, X_2, \ldots, X_n$. For $i \in \{1, 2, \ldots, n\}$, let the demand Y_i be either a constant or the vector $(X_1, X_2, \ldots, X_n)$, such* *that each demand is feasible for the network $\mathcal{N}$. Under these conditions, Theorem 4 holds.*

Proof. Define $M = \{m_1, m_2, \ldots, m_k\}$ as the set of vertices where the demands are non-null, i.e., $M \triangleq \{i \in \{1, 2, \ldots, n\} : Y_i = (X_1, X_2, \ldots, X_n)\}$. Without loss of generality, assume that $m_i < m_j$ whenever $i < j$. For each $i \in \{1, 2, \ldots, n\}$, let $d(i)$ denote the position of the first non-null demand after i, i.e., $d(i) \triangleq \min\{m : m \in M \text{ and } m > i\}$. Consider an achievable rate allocation $(R_i : i \in \{1, 2, \ldots, n-1\})$ for the line network $\mathcal{N}$. For any $i \in \{1, 2, \ldots, n-1\}$, the cutset bound on R_i is the tightest possible if we choose the set of vertices for the cutset to be the set $\{i+1, \ldots, d(i)\}$. This is true because adding extra vertices to this set adds additional sources to it without increasing the set of demands. Therefore, $(R_i : i \in \{1, 2, \ldots, n-1\})$ satisfies the following inequality for all $i \in \{1, 2, \ldots, n-1\}$:

$$\begin{aligned} R_i &\geq H(X_1, \ldots, X_n | X_{i+1}, \ldots, X_{d(i)}) \\ &\overset{(a)}{=} H(X_1, \ldots, X_{d(i)} | X_{i+1}, \ldots, X_{d(i)}) \\ &= H(X_1, \ldots, X_i | X_{i+1}, \ldots, X_{d(i)}) \end{aligned} \tag{3.9}$$

where, (a) is a consequence of the fact that $H(X_1, X_2, \ldots, X_n | X_1, X_2, \ldots, X_{d(i)}) = 0$ as the demand is feasible for the network, and (b) follows from the fact that $H(A, B|A) = H(B|A)$ for any random variables A and B. Next, we show that for the component networks $\{\mathcal{N}_j\}_{j \in \{1,2,\ldots,n\}}$, there exist rate allocations $\{(R_i^j : j \in \{1, \ldots, j-1\})\}_{j=1}^n$ which come arbitrarily close to satisfying the above bounds with equality. Observe that it suffices to restrict our attention to the networks $\{N_j\}_{j \in M}$. Fix $\epsilon > 0$ and consider any $m_i \in M$. There are two cases:

<u>$i = 1$</u> Using Slepian-Wolf encoding for multiple sources ([32]), it suffices to encode X_j at a rate $H(X_j | X_{j+1}, \ldots, X_{m_1}) + \epsilon/n$. Summing the rates required for all the sources that need to use the link $(r, r+1)$ gives the rate allocation $(R_r^{m_1} : r \in \{1, \ldots, m_1 - 1)$ to be achievable for the network $\mathcal{N}_{m_1}$, where R_r is given by

$$\begin{aligned} R_r^{m_1} &= \sum_{j=1}^{r} H(X_j | X_{j+1}, \ldots, X_{m_1}) + \epsilon/n \\ &\leq H(X_1, X_2, \ldots, X_r | X_{r+1}, \ldots, X_{m_1}) + \epsilon \\ &= H(X_1, X_2, \ldots, X_r | X_{r+1}, \ldots, X_{d(r)}) + \epsilon. \end{aligned} \tag{3.10}$$

<u>$i > 1$</u> In this case, note that $X_1, X_2, \ldots, X_n$ are available at the node m_{i-1}. Hence, the rate required over the link $(r, r+1)$ is zero for all $r < m_{i-1}$. Now, using Slepian Wolf encoding to encode the sources present at the remaining nodes, the rate allocation $(R_r^{m_i} : r \in \{m_{i-1}, \ldots, m_i - 1)$ is

achievable for the network $\mathcal{N}_{m_i}$, with R_r being given by

$$\begin{aligned} R_r^{m_i} &= H(X_1, X_2, \ldots, X_n | X_{m_{i-1}+1}, \ldots, X_{m_i}) \\ &\quad + \sum_{j=m_{i-1}+1}^{r} H(X_j | X_{j+1}, \ldots, X_{m_i}) + \epsilon/n \\ &\leq H(X_1, X_2, \ldots, X_r | X_{r+1}, \ldots, X_{m_i}) + \epsilon \\ &= H(X_1, X_2, \ldots, X_r | X_{r+1}, \ldots, X_{d(r)}) + \epsilon. \end{aligned} \tag{3.11}$$

Finally, adding the rates over all component networks for the link $(i, i+1)$, we have

$$\begin{aligned} \sum_{j=i+1}^{n} R_i^j &= R_i^{d(i)} \\ &\leq H(X_1, X_2, \ldots, X_i | X_{i+1}, \ldots, X_{d(i)}) + \epsilon \\ &\leq R_i + \epsilon. \end{aligned}$$

This shows that the cutset bound in Equation (3.9) is tight and is achievable by the approach based on component networks. Further, each R_i^j is of the form $H(Y_j | X_{i+1}, \ldots, X_j, Y_{i+1}, \ldots, Y_{j-1})$. □

3.3.4 A special class of dependent sources with dependent demands

In this section, we consider sources and demands which are dependent in the following way. We assume the existence of underlying independent sources $W_1, W_2, \ldots, W_k$ such that the sources are $X_i = W_{S(i)}$ and the demands are $Y_i = W_{D(i)}$ for $i = 1, 2, \ldots, n$ for $\{S(i)\}_{i=1}^n$ and $\{D(i)\}_{i=1}^n$ subsets of $\{1, 2, \ldots, k\}$. In order for the demands to be feasible for the network, we require $D(i) \subseteq \cup_{j=1}^{i} S(j)$ for each $i = 1, 2, \ldots, n$. Lemma 7 characterizes the rate region for line networks with the above kind of sources and demands. We need the following notation in order to state the lemma. For $j \in \{1, 2, \ldots, n-1\}$, define $d_i(j)$ and $s_i(j)$ as the first occurrence after the vertex j of W_i in a demand and source, respectively.

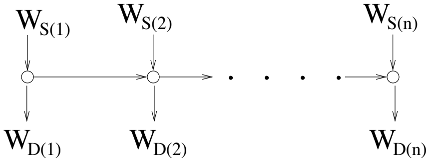


Figure 3.4: Line network with dependent sources

***Lemma* 7.** *Let $\mathcal{N}$ be an n-node line network (shown in Figure 3.4) with sources $\{W_{S(i)}\}_{i=1}^n$ and demands $\{W_{D(i)}\}_{i=1}^n$ as defined above. Then, $(R_j : j \in \{1, 2, \ldots, n-1\})$ is an achievable rate* *allocation if and only if*

$$R_j \geq \sum_{\substack{i=1 \\ i:s_i(j)>d_i(j)}}^{k} H(W_i)$$

for all $j \in \{1, 2, \ldots, n-1\}$.

Proof. We proceed by first decomposing the network $\mathcal{N}$ into k different networks $\{\mathcal{N}^i\}_{i=1}^k$, each corresponding to a different W_i out $W_1, W_2, \ldots, W_k$.

For each $i \in \{1, 2, \ldots, k\}$ and $A \subseteq \{1, 2, \ldots, k\}$, let $\widetilde{X}_A^i = W_i$ if $i \in A$ and constant otherwise. Let $\mathcal{N}_i$ be an n-node line networks $\mathcal{N}_i$ with sources $\widetilde{X}^i_{S(1)}, \widetilde{X}^i_{S(2)}, \ldots, \widetilde{X}^i_{S(n)}$ and demands $\widetilde{X}^i_{D(1)}, \widetilde{X}^i_{D(2)}, \ldots, \widetilde{X}^i_{D(n)}$. Note that each $\mathcal{N}^i$ is a line network in which both the sources and demands are either W_i or constant. By result of Section 3.3.3, it follows that cutset bound is tight for such networks and the rate allocation $(R_{j,i} : j \in \{1, 2, \ldots, n\})$, defined by

$$R_{j,i} = \begin{cases} H(W_i) & \text{if } s_i(j) > d_i(j) \\ 0 & \text{otherwise,} \end{cases}$$

is achievable. Thus, for the parent network $\mathcal{N}$, the rate allocation $(R_j : j \in \{1, \ldots, n\})$ is achievable, where $R_j = \sum_{i=1}^k R_{j,i}$. Further, as all the sources are independent, this approach is optimal. Therefore, the rate region for the network $\mathcal{N}$ is given by all $(R_j : j \in \{1, \ldots, n\})$ such that

$$R_j \geq \sum_{\substack{i=1 \\ i:s_i(j)>d_i(j)}}^{k} H(W_i).$$

Next, we show that the same can be obtained by decomposing $\mathcal{N}$ into simple networks $\mathcal{N}_1, \ldots, \mathcal{N}_n$. To this end, we first decompose the network $\mathcal{N}^i$ into simple networks $\mathcal{N}_1^i, \ldots, \mathcal{N}_n^i$, noting that the minimum rate $R_{j,i}^l$ required for the link $(j, j+1)$ in $\mathcal{N}_l^i$ is 0 if there is a demand or a source present in one of the nodes $j+1, \ldots, l$ and $H(W_i)$ otherwise. Hence, $R_{j,i}^l = H(\widetilde{X}^i_{D(l)} | \widetilde{X}^i_{S(j+1)}, \ldots, \widetilde{X}^i_{S(l)}, \widetilde{X}^i_{D(j+1)}, \ldots, \widetilde{X}^i_{D(l-1)})$. This is an optimal decomposition, since

$$R_{j,i} = \sum_{l=j+1}^{n} R_{j,i}^l.$$

Adding the rates for the l-th components of $\mathcal{N}^i$ for $i = 1, 2, \ldots, k$, the vector $(R_j^l : j \in \{1, 2, \ldots, l-1\})$ defined by

$$\begin{aligned} R_j^l &= \sum_{i=1}^{k} R_{j,i}^l \\ &= \sum_{i=1}^{k} H(\widetilde{X}^i_{D(l)} | \widetilde{X}^i_{S(j+1)}, \ldots, \widetilde{X}^i_{S(l)}, \widetilde{X}^i_{D(j+1)}, \ldots, \widetilde{X}^i_{D(l-1)}) \end{aligned}$$

$$= \quad H(X_{D(l)}|X_{S(j+1)},\ldots,X_{S(l)},X_{D(j+1)},\ldots,X_{D(l-1)}) \tag{3.12}$$

is an achievable rate allocation for $\mathcal{N}_l$. Finally, note that

$$\begin{aligned} \sum_{l=j+1}^{n} R_j^l &= \sum_{l=j+1}^{n}\sum_{i=1}^{k} R_{j,i}^l \\ &= \sum_{i=1}^{k}\sum_{l=j+1}^{n} R_{j,i}^l \\ &= \sum_{i=1}^{k} R_{j,i} \\ &= R_j. \end{aligned}$$

This shows that the rate allocation for $\mathcal{N}$ can also be obtained by summing linkwise the rate allocation for the component networks. Equation (3.12) shows that the sum is, infact, of the form claimed in Theorem 4-C.

3.3.5 3-node line networks with dependent sources

In this section, we restrict our attention to 3-node line networks of the kind shown in Figure 3.5. Sources X_1 X_2 and X_3 in the network $\mathcal{N}$ are arbitrarily dependent taking values in finite alphabets $\mathcal{X}_1$, $\mathcal{X}_2$ and $\mathcal{X}_3$ respectively. The demands at the node 2 and 3 are of the general form $Y_2 = f(X_1)$ and $Y_3 = g(X_1, X_2, X_3)$ respectively, for some $f : \mathcal{X}_1 \to \mathcal{Y}_1$ and $g : \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3 \to \mathcal{Y}_2$. The following result shows the tightness of the decomposition based approach in this case.

***Theorem* 5.** *Given $\epsilon > 0$, for every rate vector (R_1, R_2) achievable for $\mathcal{N}$, there exist achievable rate allocations $R_1^{(2)}$ and $(R_1^{(3)}, R_2^{(3)})$ for the component networks $\mathcal{N}_2$ and $\mathcal{N}_3$ (defined as in Section 3.2) such that:*

$$\begin{aligned} R_1^{(2)} + R_1^{(3)} &< R_1 + \epsilon \\ \textit{and } R_2^{(3)} &= R_2 + \epsilon. \end{aligned}$$

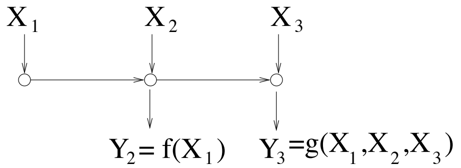


Figure 3.5: The three node line network

Proof. Let (R_1, R_2) be an achievable rate allocation for the network $\mathcal{N}$. Then, for i large enough, there exist codes (a_i, b_i) for the first and second link, respectively such that $(1/i)H(a_i(X_1(1,i))) < R_1 + \epsilon$, $(1/i)H(b_i(a_i(X_1(1,i)), X_2(1,i))) < R_2 + \epsilon$ and $Pr((\widehat{Y}_2(1,i), \widehat{Y}_3(1,i)) \neq (Y_2(1,i), Y_3(1,i)) < \epsilon$.

For the rest of this section, we use $B_i(k)$, $F_i(k)$ $X_{1,i}(k)$ and $X_{2,i}(k)$ to denote $b_i(a_i(X_1(i(k-1)+1, ik)), X_2(i(k-1)+1, ik))$, $(f(X_1(i(k-1)+1), \ldots, f(X_1(ik)))$, $X_1(i(k-1)+1, ik)$ and $X_2(i(k-1)+1, ik)$ respectively. Allowing a probability of error ϵ, the problem of coding for the network $\mathcal{N}$ can be reformulated as a functional source coding problem for the network $\mathcal{N}_i$ with sources $X_{1,i}$ and $X_{2,i}$ and the demand (F_i, B_i) as shown in Figure 3.6. Standard results on functional coding ([33],[34], [35]) can now be applied. Specifically, by evaluating the functional rate distortion function in [34], [35] at zero distortion, the minimum rate at which $X_{1,i}$ can be coded is given by

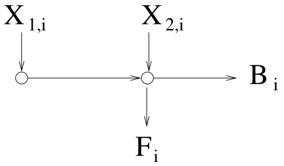


Figure 3.6: An equivalent functional coding problem

$$R^* = \inf_{\widehat{X}_{1,i}} I(X_{1,i}; \widehat{X}_{1,i} | X_{2,i})$$

where, the infimum is over the set $\mathcal{P}$ consisting of all $\widehat{X}_{1,i}$ for which $\widehat{X}_{1,i} \rightarrow X_{1,i} \rightarrow X_{2,i}$ forms a markov chain and $H(F_i, B_i | \widehat{X}_{1,i}, X_{2,i}) = 0$.

We show that the above rate can be split into two parts - the rate required to to encode $X_{1,i}$ so as to reconstruct F_i with $X_{2,i}$ as the side information, and the rate required to be able to reconstruct B_i with $X_{2,i}$ and F_i as the side information. To this end, let $X_{F,B} \in \mathcal{P}$. Then, the following hold:

$$\begin{aligned}
& I(X_{F,B}; X_{1,i} | X_{2,i}) \\
&= I(X_{F,B}, F_i; X_{1,i} | X_{2,i}) - I(F_i; X_{1,i} | X_{2,i}, X_{F,B}) \\
&= I(X_{F,B}, F_i; X_{1,i} | X_{2,i}) \\
&= H(X_{F,B}, F_i | X_{2,i}) - H(X_{F,B}, F_i | X_{2,i}, X_{1,i}) \\
&= H(X_{F,B}, F_i | X_{2,i}) - H(X_{F,B}, | X_{2,i}, X_{1,i}, F_i) \\
&= H(F_i | X_{2,i}) + H(X_{F,B} | F_i, X_{2,i}) \\
&\quad - H(X_{F,B}, | X_{2,i}, X_{1,i}, F_i) \\
&= H(F_i | X_{2,i}) + I(X_{F,B}; X_{1,i} | F_i, X_{2,i}). \qquad (3.13)
\end{aligned}$$

Since F_i is a function of $X_{1,i}$, $\frac{1}{i}H(F_i | X_{2,i})$ is an achievable rate for the network $\mathcal{N}_1$. Further, since $X_{F,B} \in \mathcal{P}$, it follows that $X_{F,B} \rightarrow X_{1,i} \rightarrow (F_i, X_{2,i})$ is a markov chain. Combining it with the fact that, $H(B_i|)X_{F,B}, X_{1,i}, F_i) = 0$, we note that $I(X_{F,B}; X_{1,i}|F_i, X_{2,i})$ is a sufficient rate for functional source coding with regards to the function B_i given $X_{2,i}$ and F_i as the side information. Therefore, $(\frac{1}{i}I(X_{F,B}; X_{1,i}|F_i, X_{2,i}), \frac{1}{i}H(B_i)$ is an achievable rate for the network $\mathcal{N}_3$, hence proving the theorem. □

3.3.6 Networks where additivity does not hold

***Theorem* 6.** *There exists a* 3*-node network* $\mathcal{N}$*, such that the sum of best possible rate allocations for its component networks is strictly greater than the best possible rate allocation for* $\mathcal{N}$

Proof. Consider the 3-node line network $\mathcal{N}$ shown in Figure 3.7

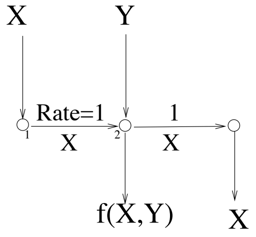


Figure 3.7: A set of achievable rates for $\mathcal{N}$

Let X be distributed uniformly on $\{0, 1\}$ and Y be independently distributed uniformly on $\{0, 1, 2, 3\}$. Let

$$f(x,y) \triangleq \begin{cases} 0 & (x,y) \in \{(0,0),(0,2),(1,1),(1,2)\} \\ 1 & (x,y) \in \{(0,1),(0,3),(1,0),(1,3)\} \end{cases}$$

Figure 3.7 shows an achievable rate allocation for the network, which can be achieved by the transmitting X over both the links. We will show that the best possible rates obtained by adding the rates over the component networks is strictly greater than the above rate allocation. To this end, consider the first component shown in Figure 3.8. Using known results for coding for computing [33], we can evaluate the best rate required over the link $(1, 2)$ to be 1.

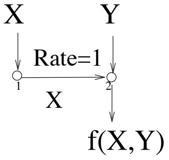


Figure 3.8: A tight rate allocation for the first component $\mathcal{N}_1$

Next, let us look at the second component. To prove our claim, it suffices to show that the rate required over the $(1,2)$ link in $\mathcal{N}_2$ is non-zero. This holds true because $H(X|Y, f(X,Y)) > 0$ and cutset bound requires a rate of atleast $H(X|Y, f(X,Y)) > 0$ across the link $(1,2)$. □

Theorem 7. *For any $n \geq 3$, there exists a $2n$-node networks $\mathcal{N}$, such that for all achievable rates $\{R_i^j\}_{i=1}^{j-1}$ for the networks $\mathcal{N}_j$, there exists an achievable rate allocation $\{R_i\}_{i=1}^{n-1}$ for $\mathcal{N}$ such that*

$$\sum_{j=i+1}^{2n} R_i^j > R_i + \Omega(n).$$

Proof. Let X and Y be independent sources uniformly distributed over $\{0,1\}$ and $\{0,\ldots,2n\}$ respectively. Define $f : \{0,1\} \times \{0,\ldots,2n\} \to \{0,\ldots,2n\}$ as

$$f(x,y) = \begin{cases} y & \text{if } y \in \{1,\ldots,2n-1\} \\ x & \text{if } y = 2n \end{cases}$$

Consider the $2n$-node line network $\mathcal{N}$ with sources $X_1,\ldots,X_{2n}$ and demands $Y_1,\ldots,Y_{2n}$ as follows. $X_1 = X$, $X_2 = X_3 = \ldots, X_{2n-1} = Y$, $X_{2n} = Y_1 =$ constant, and $Y_i = f(X, Y \oplus (i-1)$ for $i = 2,\ldots,2n$.

Using results from functional source coding, the rate required on each of the links of the i-th component is atleast 1. Further, this rate is achieved by sending X on all the links. Thus,

$$\sum_{j=i+1}^{2n} R_i^j = 2n - i.$$

On the other hand, for the network $\mathcal{N}$, a rate of 1 is sufficient (by sending X over all the links). Therefore,

$$\sum_{j=i+1}^{2n} R_i^j - R_i \geq 2n - i - 1$$

. Next, note that the LHS can be atmost $\mathcal{O}(n)$ for a network with $2n$ components. Hence,

$$\sum_{j=i+1}^{2n} R_i^j - R_i = \Omega(n).$$

□

Chapter 4

Feedback In Network Source Coding

4.1 Introduction

The networks studied in the source coding literature are typically directed, acyclic graphs. Just as it is well known that feedback cannot increase the capacity of the canonical point-to-point channel [36], it is also evident that feedback cannot increase the rate region of the canonical point-to-point lossless and lossy source coding problems. We here examine the role of feedback in network source coding, demonstrating that feedback can increase the rate region for network source coding in some networks where the rate region is well understood, and that feedback can increase the known set of achievable rates in one example network where the rate region remains unsolved. While we focus here on examples where feedback increases the achievable rate region, it is important to note that feedback does not increase the rate region for all networks or even all network topologies where feedback has the potential to increase the channel capacity. For example, in the Slepian-Wolf system shown [6] in Figure 4.1, the presence of feedback from receiver node 3 to source nodes 1 and 2 does not increase the min-cut and therefore does not enable operation at lower rates on the forward links, whereas, it is well known that feedback can increase the capacity region of the multiple access channel [37].

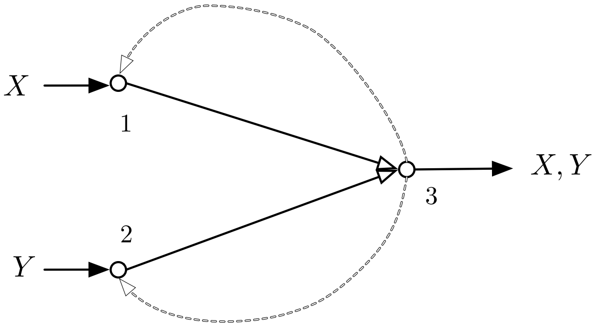


Figure 4.1: Slepian-Wolf network with feedback

Following the typical approach from channel coding, we first assume unbounded capacity on the feedback links and then consider the rate region for the forward links only. While this approach is chosen for its simplicity, the resulting insights may be directly applicable in networks where the cost of operating the feedback link is negligible compared to the cost of the forward links. For example, in sensor networks, where the central receiver node usually has much more power available than the remote sensors, the cost of sending information from the central processor back to the sensors may be far less than the cost of forward links. If transmitting information from the central processor to the sensors decreases the rate required on the forward links, then an overall system benefit might be realized.

In Section 4.3, we show through several examples that feedback can enable operation at rate points that are not achievable otherwise. Examples 4 and 6 illustrate that by knowing all the information that is available at the receiver node, a transmitter node can design codes that require less rate. Example 5 demonstrates that even with independent sources in a multi-source multi-sink network, feedback from the sinks to the sources can act as a virtual link between the sources and effectively, reduce the network to a collection of one-source one-sink networks.

In Section 4.4, we examine a multiterminal lossy source coding problem with two encoders. While the rate region without feedback remains unsolved, we show that feedback enables lower rates than the best achievable rates known to date. The result of Example 6 is a special case of this network that demonstrates the tightness of our bound at the extreme points, showing that feedback strictly enlarges the rate region for this network.

The above examples show that feedback has the potential to reduce the rates required on forward links, in general. Unfortunately, fully characterizing the rate tradeoffs for networks with bidirectional links is a difficult problem. The difficulty arises in part from the fact that a potentially unbounded number of rounds of transmissions may be required to achieve optimal rates [17]. Traditional information theoretic techniques do not readily extend to these situations.

To develop insights into the tradeoff between forward and feedback rates, we consider a simple network. The setup that we consider is shown in Figure 4.3. The source process X is observed at $\mathcal{E}_x$ and demanded at $\mathcal{E}_y$. Terminal $\mathcal{E}_y$ observes source Y jointly distributed with X. We wish to characterize the set of rates $(R_{(1,2)}, R_{(2,1)})$ required to enable $\mathcal{E}_y$ to reconstruct X with precisely zero probability of error.

A relaxed version of this problem is the asymptotically lossless setting, where the process X is demanded with a vanishing error probability as the block length increases without bound. For this setting, it is easy to see that even unlimited rate on the feedback link does not change the rate required. In particular, for the system shown in Figure 4.2, Slepian and Wolf [6] have shown that the minimum rate required on the forward link equals the cutset bound. Since the addition of the feedback link does not alter the cutset bound, it follows that the rate required cannot be reduced

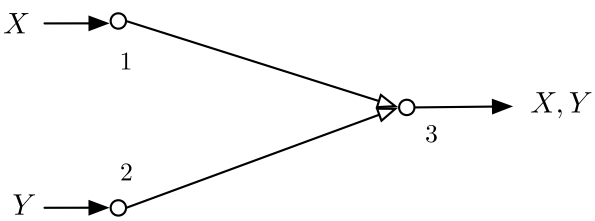


Figure 4.2: Slepian-Wolf network without feedback

any further.

In the zero-error setting, the cutset bound is not achievable for the system shown in Figure **??** (c.f. [38],[39],[40],[41]) since $\mathcal{E}_x$ has to distinguish between all possible pairs of source and side information sequences, not just those that are typical. We show that a two-way communication between $\mathcal{E}_x$ and $\mathcal{E}_y$ allows them to decide whether or not their observed sequences are typical and enables a tradeoff between the rate on the forward and the backward links. Of special interest are two extreme cases that are discussed in Examples 7 and 8. In the first example, asymptotically zero rate on the feedback link enables $\mathcal{E}_x$ to operate at rates arbitrarily close to the cutset bound, while in the second, the sum rate required on the two links is bounded from below by $H(X)$.

It should be noted that the study of feedback in source coding is not entirely new [16, 17, 42]. For the zero-error problem with feedback, prior works in the computer science literature have examined this problem from a communication complexity perspective (see for example, [12, 13, 14, 15]). In this work, we make the observation that insights from both communication complexity theory and asymptotically lossless source coding are useful in fully understanding problems of feedback in source coding. Further, we show that feedback is useful in the zero error setting, even for networks where it does not help in the asymptotically zero error setting.

We begin by describing our setup and introducing necessary notation in Section 4.2.

4.2 Preliminaries

1. *Network model:*

   We define a network as a directed graph along with a set of sources and demands. We specify a network as a collection $\mathcal{N} = (V, E, X, Y, T)$, where $V = \{1, 2, \ldots, |V|\}$ denotes the set of vertices, E denotes the set of directed edges, X and Y denote the sources observed at vertices 1 and 2 respectively, and T denotes the set of sink nodes. The demand at nodes in the set T is specified by the problem under consideration.

   An edge $(u, v) \in E$ denotes a lossless link with vertices $u \in V$ as the transmitter and $v \in V$ as the receiver. Let $F = \cup_{t \in T}\{(t, 1), (t, 2)\}$. We refer to the set $E \cap F$ as the set of *feedback*

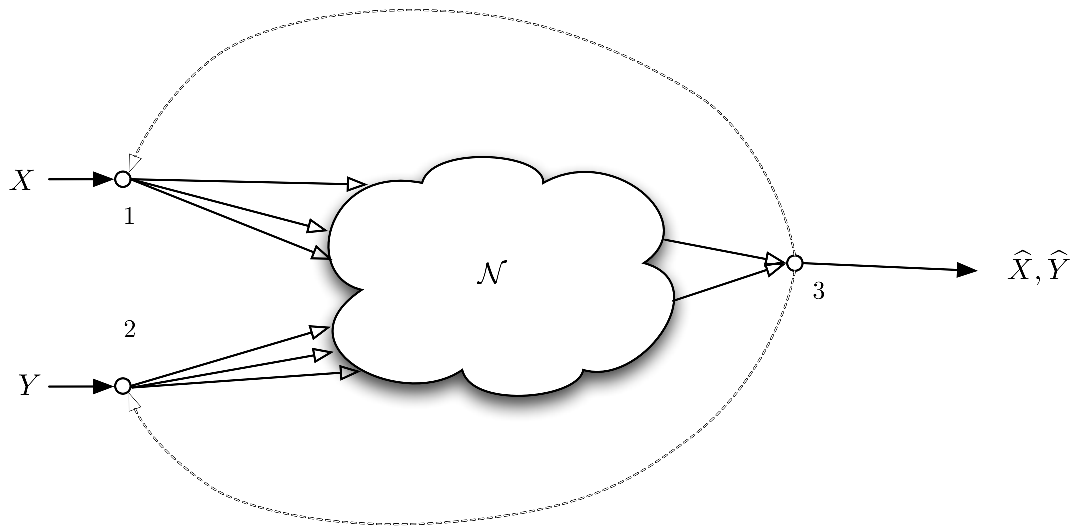


Figure 4.3: A network with feedback

links, and the set $E \setminus F$ as the set of *forward links*. We assume that the subgraph $(V, E \setminus F)$ formed by the forward links is acyclic.

2. *Source model:*

   The sequence $\mathbf{X}$ is observed at node 1, and $\mathbf{Y}$ is observed at node 2. Let P_{XY} be a probability mass function on a finite alphabet $\mathcal{X} \times \mathcal{Y}$. Denote by p_X (resp. p_Y), the marginal of P_{XY} on $\mathcal{X}$ (resp. $\mathcal{Y}$). For each $n \in \mathbb{N}$, the collection of random variable $(\mathbf{X}, \mathbf{Y}) \triangleq (X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ is drawn i.i.d. from the distribution P_{XY}.

3. *Demand model:*

   Each sink $t \in T$ wishes to reproduce either one or both of the sequences $\mathbf{X}$ and $\mathbf{Y}$. Let $\hat{\mathbf{X}}_t \in \hat{\mathcal{X}}$ and $\hat{\mathbf{Y}}_t \in \hat{\mathcal{Y}}$ denote the reconstructions of $\mathbf{X}$ and $\mathbf{Y}$ respectively at sink t. Here, $\hat{\mathcal{X}}$ and $\hat{\mathcal{Y}}$ are the reproduction alphabets corresponding to $\mathbf{X}$ and $\mathbf{Y}$ respectively and are assumed to be same for all the sinks. The objective of the transmission is to reconstruct $\mathbf{X}$ and $\mathbf{Y}$ subject to a suitable decoding criteria defined by the specifics of the problem. We consider four different decoding criteria in this paper, the details of which are presented in Section 6

4. *Network Source Codes:*

   We consider transmission schemes that are defined by *block network source codes*, i.e., the source messages are divided into blocks of fixed size (henceforth referred to as the *dimension of the code*) , and each block of source messages is encoded and transmitted independently of other blocks using a network source code. We limit our attention to codes for which all transmitted codewords are binary strings. While this may be a restrictive assumption form the point of view of code design, it does not alter the set of rates achievable asymptotically.

We define a *round* as a sequence of transmissions on forward links $E \setminus F$ followed by transmissions on the feedback links $E \cup F$. When the network is cycle-free, all transmission schemes may be thought to be spread over a single round. However, in a cyclic network, transmission schemes may involve several rounds before the messages are decoded at the sink. In view of this property, we allow block network codes to be implemented over multiple rounds. Note that feedback received by the source nodes in the final round of transmission does not alter the decoded output at the sink. Hence, we restrict our attention to schemes in which the final round involves transmission only on the forward links.

Let $\tau(k, E) = \{1, 2, \ldots, k\} \times E \setminus \{k\} \times F$ for all $k \in \mathbb{N}^+$. Formally, we define an n-dimensional, k-round network source code code (f, g) as a collections of functions $f = \{f_e^{(i)} : (i, e) \in \tau(k, E)\}$ and $g = (g_t : t \in T)$, with

$$\begin{aligned} & f : \mathcal{X}^n \times \mathcal{Y}^n \to (\{0,1\}^*)^{|\tau(k,E)|} \text{ and} \\ \text{and} \qquad & g : (\{0,1\}^*)^{|\tau(k,E)|} \to (\hat{\mathcal{X}}^n \times \hat{\mathcal{Y}}^n)^{|T|}, \end{aligned}$$

that satisfies the following properties:

(a) $H(f_{(1,v)}^{(i)}(\mathbf{X}, \mathbf{Y}) | \mathbf{X}, (f_{(t,1)}^{(j)}(\mathbf{X}, \mathbf{Y}) : j < i)) = 0$ for each $i = 1, 2, \ldots, k$ and for all v such that $(1, v) \in E$.

(b) $H(f_{(2,v)}^{(i)}(\mathbf{X}, \mathbf{Y}) | \mathbf{Y}, (f_{(t,2)}^{(j)}(\mathbf{X}, \mathbf{Y}) : j < i)) = 0$ for each $i = 1, 2, \ldots, k$ and for all v such that $(2, v) \in E$.

(c) $H(f_{(v,w)}^{(i)}(\mathbf{X}, \mathbf{Y}) | (f_{(u,v)}^{(j)}(\mathbf{X}, \mathbf{Y}) : j \leq i, (u, v) \in E)) = 0$ for each $i = 1, 2, \ldots, k$ and for all $(v, w) \in E$ such that $v \notin \{1, 2\}$.

(d) $H(g_t(f(\mathbf{X}, \mathbf{Y})) | (f_{(v,t)}^{(i)}(\mathbf{X}, \mathbf{Y}) : 1 \leq i \leq k, (v, t) \in E)) = 0$ for each $t \in T$.

Here, $f_e^{(i)}(\mathbf{x}, \mathbf{y})$ denotes the codeword transmitted on the edge e in the i-th round of transmission and $g_t(\mathbf{x}, \mathbf{y})$ denotes the reconstruction at a node $t \in T$ when $\mathbf{x}$ and $\mathbf{y}$ are observed at node 1 and 2 respectively. When the network has only one sink node t, we drop the subscript and refer to the reconstruction as $g(\mathbf{x}, \mathbf{y})$. Properties 4a and 4b are consequences of the i-th round codeword on edges outgoing from a source node being a function of only the observede source sequence and the feedback received prior to the i-th round, while property 4c are a consequences of i-th round codewords on edges outgoing from an intermediate node being a function only of the codewords received in the i-th round and earlier. Property 4d stipulates that the decoded message at the sink is a function of all received codewords at the sink at the end of k rounds.

5. <u>*Rate of a code:*</u>

Let $\{0,1\}^* = \cup_{i=1}^{\infty} \{0,1\}^i$ denote the set of all finite-length binary strings. Let $l : \{0,1\}^* \to \mathbb{N}$ be the length function on $\{0,1\}^*$, i.e. if b is a string of n bits, then $l(b) = n$. For a n-dimensional, k-round code (f, g), we define the *rate* as the vector $\mathbf{R} = (R_e : e \in E)$, where

$$R_e \triangleq \frac{1}{n}\sum_{i=1}^{k} \mathbf{E}[l(f_e^{(i)}(\mathbf{X}, \mathbf{Y}))].$$

We say that code (f, g) of rate $\mathbf{R}$ is a *fixed-rate code* if, for some vector $(l_e^{(i)} : (i, e) \in \tau(k, E)$,

$$\begin{aligned} & l(f_e^{(i)}(\mathbf{x}, \mathbf{y})) = l_e^{(i)} \text{ for all } (\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n \text{ and } (i, e) \in \tau(k, E), \\ & \text{and } R_e = \frac{1}{n} \sum_{i:(i,e)\in\tau(k,E)} l_e^{(i)} \text{ for all } e \in E. \end{aligned}$$

6. <u>*Achievable Rate Region:*</u>

 The achievable rate region denotes the set of rate vectors for which there exist sequences of codes satisfying a desired decoding criterion. In this paper, we consider the following decoding criteria, each of which corresponds to a different achievable rate region. Let $\mathcal{N} = (V, E, X, Y, T)$ be a given network. Let $\mathcal{N}_F = (V, E \setminus F, X, Y, T)$ be the network obtained by deleting the feedback edges $E \cap F$ from $\mathcal{N}$.

 (a) <u>*Lossless Source Coding with Coded Side Information:*</u>

 Consider the network shown in figure 4.4. Each sink node $t \in T$ demands only sequence $\mathbf{X}$ with an error probability that vanishes asymptotically with the dimension of the code. Thus, we say that a rate vector $\mathbf{R} \in \mathbb{R}^E$ is *losslessly achievable with coded side information* if there exist a sequence of codes $\{(f^n, g^n)\}_{n=1}^{\infty}$ such that

 i. (f^n, g^n) is an n-dimensional k-round code of rate $\mathbf{R}$ for some $k \geq 1$.

 ii. $\lim_{n\to\infty} \Pr(g_t^n(f^n(\mathbf{X}, \mathbf{Y})) \neq \mathbf{X}) = 0$ for each $t \in T$.

 (b) <u>*Lossless Multicast:*</u>

 Consider the network shown in figure 4.5. The sink t demands the pair $(\mathbf{X}, \mathbf{Y})$ with an error probability that vanishes asymptotically with the dimension of the code. Thus, we say that a rate vector $\mathbf{R} \in \mathbb{R}^E$ is *losslessly achievable for multicast* if there exist a sequence of codes $\{(f^n, g^n)\}_{n=1}^{\infty}$ such that

 i. (f^n, g^n) is an n-dimensional k-round code of rate $\mathbf{R}$ for some $k \geq 1$.

 ii. $\lim_{n\to\infty} \Pr(g_t^n(f^n(\mathbf{X}, \mathbf{Y})) \neq (\mathbf{X}, \mathbf{Y})) = 0$ for each $t \in T$.

 (c) <u>*Zero-error source coding:*</u>

 Consider the network shown in Figure 4.7. The sink t demands $\mathbf{X}$ with exactly zero error probability. We say that a rate vector $\mathbf{R} \in \mathbb{R}^E$ is *zero-error achievable* if for some $n, k \in \mathbb{N}$, there exists a n-dimensional k-round code (f, g) of rate $\mathbf{R}$ such that $g(f(\mathbf{x}, \mathbf{y})) = \mathbf{x}$ for all $(\mathbf{x},\mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$.

(d) Lossy Multicast:

Consider the network shown in figure 4.8. The sink t demands the pair $(\mathbf{X},\mathbf{Y})$ within a distortion constraint on the reproduction. Let $d_X : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}^+$ and $d_Y : \mathcal{Y} \times \hat{\mathcal{Y}} \to \mathbb{R}^+$ be given distortion measures and let $(D_X, D_Y) \in \mathbb{R}^+ \times \mathbb{R}^+$ be specified distortion thresholds. Reusing the above notation, for each $n \geq 1$, let $d_X : \mathcal{X}^n \times \hat{\mathcal{X}}^n \to \mathbb{R}^+$ and $d_Y : \mathcal{Y}^n \times \hat{\mathcal{Y}}^n \to \mathbb{R}^+$ denote be distortion measures for n-dimensional sequences such that for all $(\mathbf{x},\mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ and $\mathbf{z} = (\hat{\mathbf{x}},\hat{\mathbf{y}}) \in \hat{\mathcal{X}}^n \times \hat{\mathcal{Y}}^n$, $d_X(\mathbf{x},\mathbf{z}) = \frac{1}{n}\sum_{i=1}^n d_X(x_i,\hat{x}_i)$ and $d_Y(\mathbf{y},\mathbf{z}) = \frac{1}{n}\sum_{i=1}^n d_Y(y_i,\hat{y}_i)$. We say that a rate vector $\mathbf{R} \in \mathbb{R}^E$ is *achievable with distortion threshold* (D_X, D_Y) if there exist a sequence of codes $\{(f^n, g^n)\}_{n=1}^\infty$ such that

i. (f^n, g^n) is an n-dimensional k-round code of rate $\mathbf{R}$ for some $k \geq 1$.

ii. $\mathbf{E}[d_X(\mathbf{X}, g^n(f^n(\mathbf{X},\mathbf{Y})))] \leq D_X$.

iii. $\mathbf{E}[d_Y(\mathbf{Y}, g^n(f^n(\mathbf{X},\mathbf{Y})))] \leq D_Y$.

For each of the above decoding criteria, the achievable rate region with feedback is defined as the set of rate vectors $\mathbf{R} \in \mathbb{R}^E$ that are achievable on $\mathcal{N}$ under the given decoding criterion and is usually denoted by $\mathcal{R}(\mathcal{N})$. The achievable rate region without feedback $\mathcal{R}_F(\mathcal{N})$is defined as the set of rate vectors $\mathbf{R} \in \mathbb{R}^{E \setminus F}$ that are achievable on $\mathcal{N}_F$ under the given decoding criterion. Thus, $\mathcal{R}_F(\mathcal{N}) = \mathcal{R}(\mathcal{N}_F)$. For the case of Lossy Multicast, when the distortion thresholds (D_X, D_Y) are not clear from the context, the abeove rate regions are denoted by $\mathcal{R}(\mathcal{N}, D_X, D_Y)$ and $\mathcal{R}_F(\mathcal{N}, D_X, D_Y)$.

7. Other notation:

Let $A_\epsilon^{(n)}(X)$ denote the set of ϵ-strongly typical sequences in $\mathcal{X}^n$. Similarly, define $A_\epsilon^{(n)}(Y)$ and $A_\epsilon^{(n)}(X,Y)$. Denote the set of types of n-length sequences drawn from $\mathcal{X}^n$ by $\mathcal{P}_n(\mathcal{X})$ and the type class of a probability mass function $Q \in \mathcal{P}_n(\mathcal{X})$ by $T^n(Q)$ (see [26, Section 12.1] for details). For a given $\mathcal{Z}$-valued random variable Z, we denote the entropy-coded binary description of Z by $e(Z)$. Thus, $e : \mathcal{Z} \to \{0,1\}^*$ satisfies

$$H(Z) \leq \mathbf{E}[l(e(Z))] \leq H(Z) + 1.$$

.

4.3 The role of feedback in source coding networks

To understand if feedback helps at all, we first permit the rate on feedback links to be unbounded and examine the achievable rates on the forward links with and without feedback . Given a network

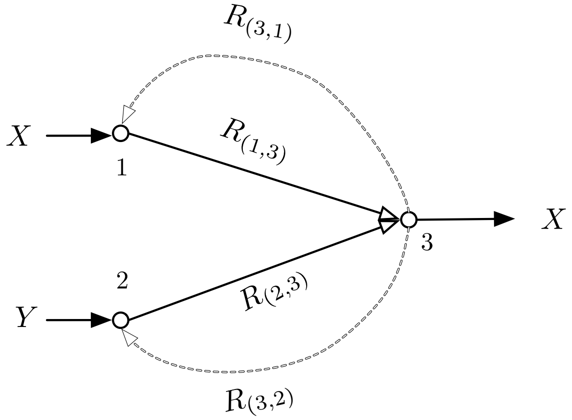


Figure 4.4: Lossless source coding problem with coded side information and feedback

$\mathcal{N} = (V, E, X, Y, T)$ with rate region $\mathcal{R}(\mathcal{N})$, we define the rate region with unbounded feedback as the set

$$\begin{aligned}\mathcal{R}_\infty(\mathcal{N}) \triangleq \{&(R_e : e \in E \setminus F) : \\ &(R_e : e \in E) \in \mathcal{R}(\mathcal{N}) \text{ for some } (R_e : e \in F) \in \mathbb{R}^F\}.\end{aligned}$$

Thus, $\mathcal{R}_\infty(\mathcal{N})$ is the set of rates achievable on the forward links that are achievable for some collection of rates on the reverse link. We give three examples of networks where $\mathcal{R}_\infty(\mathcal{N}) \supsetneq \mathcal{R}_F(\mathcal{N})$.The first two examples demonstrate that feedback can expand the rate region even for lossless coding. This is in contrast to the point-to-point case, where feedback cannot increase the rate region. The first example is the coded side information network, which has been studied previously without feedback in [29].

4.3.1 Source coding with coded side information

Consider the network $\mathcal{N}$ shown in Figure 4.4 with vertex set $V = \{1, 2, 3\}$ and edge set $E = \{(1, 3), (2, 3), (3, 1), (3, 2)\}$. The encoders 1 and 2 observe sources X and Y respectively, and the decoder at node 3 wishes to reconstruct X losslessly with Y as the coded side information. Without feedback, the rate region, $\mathcal{R}_F(\mathcal{N})$ is the collection of all rate pairs $\mathbf{R} = (R_{13}, R_{23})$ that satisfy the following inequalities for some random variable U forming a Markov chain $U \to Y \to X$[29]:

$$R_{13} \geq H(X|U) \tag{4.1}$$

$$R_{23} \geq I(Y;U). \tag{4.2}$$

We first derive the achievable region $\mathcal{R}(\mathcal{N})$ in the following theorem:

***Theorem* 8.** *For the network $\mathcal{N}$ shown in Figure 4.4, a rate pair $\mathbf{R} = (R_{13}, R_{23})$ is losslessly achievable with coded side information if and only if*

$$R_{13} \geq H(X|Y) \tag{4.3}$$

$$R_{13} + R_{23} \geq H(X). \tag{4.4}$$

Proof. The necessity of (4.3) and (4.4) follows from simple cutset arguments. The achievability of these rates can be shown by a two-step Slepian-Wolf code.

Fix $\epsilon > 0$. Let $\mathbf{R} = (R_{13}, R_{23})$ satisfy (4.3) and (4.4).

Let $\tilde{\mathcal{N}} = (\{1,2\}, \{(1,2)\}, X, Y, \{2\})$ be a network in which node 1 observes X and node 2 wishes to reconstruct it with asymptotically vanishing error probability with the help of side information Y. Let $(\tilde{\alpha}, \tilde{\beta})$ be an n-dimensional Slepian-Wolf code of rate R_{13} that operates with an error probability $p_{\tilde{\alpha}}$.

Next, let $\mathcal{N}' = (V, E \setminus F, X, X, \{3\})$ be a network in which both nodes 1 and 2 observe the source X and node 3 demands X losslessly with asymptotically vanishing error probability. Let (α, β) be an n-dimensional Slepian-Wolf code [6] of rate (R_{13}, R_{23}) that operates with an error probability p_α with

$$\tilde{\alpha}_{(1,3)}(\mathbf{x}, \mathbf{y}) = \alpha_{(1,3)}(\mathbf{x}) \tag{4.5}$$

for all $(\mathbf{x}, \mathbf{y}) \in \mathcal{X} \times \mathcal{Y}$. Note that the constraint imposed by Eq (4.5) does not result in a loss of asymptotic optimality as the code for each node is generated by randomly binning all source sequences and is independent of the code code for the other node. Thus, the codebook for each node depends only on the encoding rate and the probability distribution of the source observed at that node.

Consider the 2-round n-dimensional code (f, g) for $\mathcal{N}$ defined by functions $(f_e : e \in E)$ and g that satisfy

$$\begin{aligned}
f_{(1,3)}(\mathbf{x}, \mathbf{y}) &= (\tilde{\alpha}_{(1,3)}(\mathbf{x}, \mathbf{y}), 0), \\
f_{(2,3)}(\mathbf{x}, \mathbf{y}) &= (0, \tilde{\alpha}_{(2,3)}(\tilde{\beta}(\alpha_{(1,3)}(\mathbf{x}), \mathbf{y})) \\
f_{(3,1)}(\mathbf{x}, \mathbf{y}) &= 0, \text{ and} \\
f_{(3,2)}(\mathbf{x}, \mathbf{y}) &= \alpha_{(1,3)}(\mathbf{x}) \\
g(f(\mathbf{x}, \mathbf{y})) &= \beta(\alpha_{(1,3)}(\mathbf{x}), \alpha_{(2,3)}(\hat{\mathbf{x}}_2))
\end{aligned}$$

for all $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$. The above code implements the following sequence of transmissions:

- Node 1 transmits $\alpha_{(1,3)}(\mathbf{x})$ to node 3 at a rate R_{13}, which is then made available to node 2 via feedback link $(3, 2)$.

- Node 2 performs a Slepian-Wolf decoding to create a reconstruction $\hat{\mathbf{x}}_2 = \tilde{\beta}(\alpha_{(1,3)}(\mathbf{x}), \mathbf{y})$, and transmits the codeword $\alpha_{(2,3)}(\hat{\mathbf{x}}_2)$ to node 3 using rate R_{23}.
- Node 3 finally outputs the reconstruction $\hat{\mathbf{x}} = \beta(\alpha_{(1,3)}(\mathbf{x}), \alpha_{(2,3)}(\hat{\mathbf{x}}_2))$.

Using the above strategy, a decoding error occurs node if either node 2 is unable to correctly reconstruct $\mathbf{x}$ using $(\tilde{\alpha}_{(1,3)}(\mathbf{x}, \mathbf{y}), \tilde{\alpha}_{(2,3)}(\mathbf{x}, \mathbf{y}))$, or node 3 is unable to do so using $(\alpha_{(1,3)}(\mathbf{x}), \alpha_{(2,3)}(\hat{\mathbf{x}}))$. Thus,

$$\begin{aligned}
&\Pr(\mathbf{X} \neq g(f(\mathbf{X}, \mathbf{Y}))) \\
&\quad = \Pr(\mathbf{X} \neq \beta(\alpha_{(1,3)}(\mathbf{X}), \alpha_{(2,3)}(\hat{\mathbf{X}}))) \\
&\quad = \Pr(\mathbf{X} \neq \hat{\mathbf{X}}) \Pr(\beta(\alpha_{(1,3)}(\mathbf{X}), \alpha_{(2,3)}(\hat{\mathbf{X}})) \neq \mathbf{X} | \mathbf{X} \neq \hat{\mathbf{X}}) \\
&\qquad + \Pr(\mathbf{X} = \hat{\mathbf{X}}) \Pr(\beta(\alpha_{(1,3)}(\mathbf{X}), \alpha_{(2,3)}(\mathbf{X})) \neq \mathbf{X}) \\
&\quad \leq \Pr(\mathbf{X} \neq \hat{\mathbf{X}}) + \Pr(\beta(\alpha_{(1,3)}(\mathbf{X}), \alpha_{(2,3)}(\mathbf{X})) \neq \mathbf{X}) \\
&\quad = p_\alpha + p_{\tilde{\alpha}}.
\end{aligned}$$

Next, we choose n large enough such

$$\max\{p_\alpha, p_{\tilde{\alpha}}\} < \epsilon/2.$$

This is possible since (4.3) and (4.4) imply that the rates for codes (α, β) and $(\tilde{\alpha}, \tilde{\beta})$ are sufficient for existence of Slepain-Wolf codes [6] on networks $\mathcal{N}'$ and $\tilde{\mathcal{N}}$, respectively. Thus,

$$\begin{aligned}
\Pr(\mathbf{X} \neq g(f(\mathbf{X}, \mathbf{Y}))) &\leq \epsilon/2 + \epsilon/2 \\
&= \epsilon.
\end{aligned}$$

Since ϵ is arbitrary, this shows that $\mathbf{R} \in \mathcal{R}_\infty(\mathcal{N})$. □

The rate region derived above suggests that the addition of feedback links enables greater cooperation between the encoders at node 1 and node 2. The following example demonstrates this rigorously by proving that the region $\mathcal{R}_\infty(\mathcal{N})$ may include points that are not contained in $\mathcal{R}_F(\mathcal{N})$. Further, we also show that the ratio of the rates required on a given forward link with feedback to the rate required on the same link without feedback may be arbitrarily small when we fix the rate on the other link.

***Example* 4.** *Let $N \in \mathbb{N}^+$, $N \geq 2$. Let $\mathcal{X} = \mathcal{Y} = \{1, 2, \ldots, 2N\}$. Consider the coded side information network Let X be distributed uniformly over $\mathcal{X}$, and let the conditional distribution of Y given $X = x$*

be given as

$$p_{Y|X}(Y=y|X=x) = \begin{cases} 1/N & y = x+i \ (mod\ N) \\ & \quad \text{for some } i \in \{1,2,\dots,N\}, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $H(X) = H(Y) = \log_2 2N$ *and* $H(X|Y) = H(Y|X) = \log_2 N$.

Let $R_\infty \triangleq \min\{R_{(2,3)} : (H(X|Y), R_{(2,3)}) \in \mathbf{R}_\infty(\mathcal{N})\}$ *be the minimum rate required from node* 2 *in* $\mathcal{N}$ *when node* 1 *transmits at rate* $H(X|Y)$. *Let* $R_F \triangleq \min\{R_{(2,3)} : (H(X|Y), R_{(2,3)}) \in \mathbf{R}_F(\mathcal{N})\}$ *be the corresponding rate in* $\mathcal{N}_F$.

By Theorem 8, $R_\infty = I(X;Y) = 1$. *On the other hand, for the network* $\mathcal{N}_F$, *achieving rate* $H(X|Y)$ *on the link* $(1,2)$ *requires at least rate* $H(Y) = \log_2 2N$ *for the above probability mass functions [43]. Thus,* $R_F/R_\infty = \log_2 2N$.

This shows that feedback can strictly reduce the forward rates required. Moreover, the gain R_F/R_∞ *depends on the distribution of* (X,Y) *and may be arbitrarily large.*

Next, we apply the result of Theorem 8 to derive the achievable rate region for arbitrarily sized networks with one sink and one side information source. For a network $\mathcal{N} = (V,E,X,Y,T)$, a rate vector $\mathbf{R} \in \mathbb{R}^{|E \setminus F|}$, and subsets s and t of V, let the *min-cut* from s to t under the rate allocation $\mathbf{R}$ be defined as

$$M(\mathbf{R},s,t) \triangleq \min_{C: s \subseteq C, t \subseteq V\setminus C} \sum_{(u,v): u\in C, v \in V\setminus C} R_{(u,v)}.$$

***Theorem* 9.** *Let* (V,E) *be a directed graph. Let* $t \in V$ *and* $\{(t,1),(t,2)\} \subseteq E$. *Consider the network* $\mathcal{N} = (V,E,X,Y,\{t\})$ *in which source* X *is observed at node* 1, *source* Y *is observed at node* 2, *and* X *is demanded at node* t. *Then* $\mathbf{R} \in \mathcal{R}_\infty(\mathcal{N})$ *iff*

$$M(\mathbf{R},\{1,2\},\{t\}) \geq H(X), \tag{4.6}$$

and

$$M(\mathbf{R},\{1\},\{t\}) \geq H(X|Y), \tag{4.7}$$

Proof. As in Theorem 8, the necessity of the above conditions follows from cut-set arguments. Equation (4.6) is obtained by considering the sum of the rates on all outgoing edges from cuts containing both X and Y, while Equation (4.7) is obtained by applying the argument to all cuts that contain the source X, but not necessarily the source Y.

To prove the achievability of the above region, let $\mathbf{R} = (R_e : e \in E \setminus F)$ satisfy equations (4.6) and (4.7). We claim that there exist $\tilde{\mathbf{R}}, \hat{\mathbf{R}} \in \mathbb{R}_+^{E\setminus F}$ such that

$$\mathbf{R} = \tilde{\mathbf{R}} + \hat{\mathbf{R}}, \tag{4.8}$$

$$M(\tilde{\mathbf{R}},\{1\},\{t\}) \geq H(X|Y), \tag{4.9}$$

$$\text{and} \qquad M(\hat{\mathbf{R}},\{2\},\{t\})+M(\tilde{\mathbf{R}},\{1\},\{t\}) \geq H(X,Y). \tag{4.10}$$

To this end, let G' be a graph with vertex set $V' \triangleq V \cup \{v_s\}$ and edge set $E' \triangleq E \cup \{(s,1),(s,2)\}$. Let $\mathbf{R}' \in \mathbf{R}^{|E' \setminus F|}$ be defined as follows:

$$R'_e = \begin{cases} R_e & \forall\ e \in E \\ \sum_{e' \in E} R_{e'} & \text{if } e \in \{(s,1),(s,2)\}. \end{cases}$$

It follows that

$$M(\mathbf{R}',\{s\},\{t\}) = M(\mathbf{R},\{1,2\},\{t\})$$

and

$$M(\mathbf{R}',\{1\},\{t\}) = M(\mathbf{R},\{1\},\{t\}).$$

By the min-cut max-flow theorem [44], there exists a flow $(f_e : e \in E')$ over the graph G' such that

1. $0 \leq f_e \leq R'_e$ for all $e \in E'$,
2. $\sum_{v:(s,v)\in E'} f_{(s,v)} = \sum_{v:(v,t)\in E'} f_{(v,t)} = M(\mathbf{R}',\{s\},\{t\})$, and
3. $\sum_{u:(u,v)\in E'} f_{(u,v)} = \sum_{u':(v,u')\in E'} f_{(v,u')}$ for all $v \in V \setminus \{1,t\}$.

Further, note that

$$\sum_{v:(s,v)\in E'} f_{(s,v)} = \sum_{v:(1,v)\in E\setminus(1,2)} f_{(1,v)} + \sum_{v:(2,v)\in E\setminus(2,1)} f_{(2,v)}.$$

To this end, let $\tilde{\mathbf{R}} \in \mathbb{R}_+^{|E\setminus F|}$ such that

1. $M(\tilde{\mathbf{R}},\{1\},\{t\}) = M(\mathbf{R},\{1\},\{t\})$.
2. For all $\tilde{\mathbf{R}}' \in \{\mathbf{R}' : R'_e \leq \tilde{R}_e \forall e \in E \setminus F\}$, $M(\tilde{\mathbf{R}}',\{1\},\{t\}) < M(\tilde{\mathbf{R}},\{1\},\{t\})$.

Next, let $\hat{\mathbf{R}} \in \mathbb{R}^{|E\setminus F|}$ be the vector $(R_e - \tilde{R}_e : e \in E \setminus F)$. Since

$$M(\tilde{\mathbf{R}},\{1\},\{t\}) + M(\hat{\mathbf{R}},\{1\},\{t\}) \leq M(\tilde{\mathbf{R}}+\hat{\mathbf{R}},\{1\},\{t\}),$$

it follows that $M(\hat{\mathbf{R}},\{1\},\{t\}) = 0$. Let $C_1 \subseteq V$ be the cut achieving the min-cut rate $M(\hat{\mathbf{R}},\{1\},\{t\})$ for the rate vector $\hat{\mathbf{R}}$, i.e., $1 \in C_1$ and

$$M(\hat{\mathbf{R}},\{1\},\{t\}) = \sum_{(u,v):u\in C_1, v\in V\setminus C_1} R_{(u,v)}.$$

Likewise, let $C_2 \subseteq V$ be the cut containing vertex 2 that achieves the min-cut rate $M(\hat{\mathbf{R}}, \{2\}, \{t\})$ for the rate vector $\hat{\mathbf{R}}$. Then, the min-cut rate from $\{1, 2\}$ to $\{t\}$ is upper bounded as follows:

$$\begin{aligned} M(\hat{\mathbf{R}}, \{1,2\}, \{t\}) &\leq \sum_{(u,v):u \in C_1 \cup C_2, v \in V \setminus (C_1 \cup C_2)} R_{(u,v)} \\ &\leq \sum_{(u,v):u \in C_1, v \in V \setminus C_1} R_{(u,v)} + \sum_{(u,v):u \in C_2, v \in V \setminus C_2} R_{(u,v)} \\ &= M(\hat{\mathbf{R}}, \{1\}, \{t\}) + M(\hat{\mathbf{R}}, \{2\}, \{t\}) \\ &= M(\hat{\mathbf{R}}, \{2\}, \{t\}). \end{aligned}$$

Further, the following is always true:

$$M(\hat{\mathbf{R}}, \{1,2\}, \{t\}) \geq M(\hat{\mathbf{R}}, \{2\}, \{t\}).$$

Therefore,

$$M(\hat{\mathbf{R}}, \{1,2\}, \{t\}) = M(\hat{\mathbf{R}}, \{2\}, \{t\}).$$

□

4.3.1.1 Networks with multicast demands

Next, we consider networks with multicast demands, which have been studied in the context of network coding in [9, 18]. Once again, we demonstrate that feedback enables rates which are not achievable otherwise.

Consider the network shown in Figure 4.5. The rate region for this network, both with and without feedback, is given by the cut-set bounds, as shown in [18]. In particular, with feedback, all rate points $\mathbf{R} = (R_e : e \in E)$ satisfying the following inequalities are achievable:

$$\sum_{e \in \Gamma_o(C)} R_e \geq H(X_1, X_2, \ldots, X_k) \tag{4.11}$$

for all $C \subseteq V$ s.t. $S \subseteq C$ and $T \nsubseteq C$.

We show that the above region may be strictly larger than the region obtained without feedback with the help of the following example

***Example* 5.** *Consider the network $\mathcal{N}$ shown in Figure 4.6. We show that $\mathcal{R}_\infty(\mathcal{N}) \supsetneq \mathcal{R}_F(\mathcal{N})$ in the following.*

Let $\mathbf{R} = (R_e : e \in E)$ satisfy (4.11). Let $n \in \mathbb{N}$. Consider a 2-round n-dimensional code (f, g) of rate $\mathbf{R}$, where

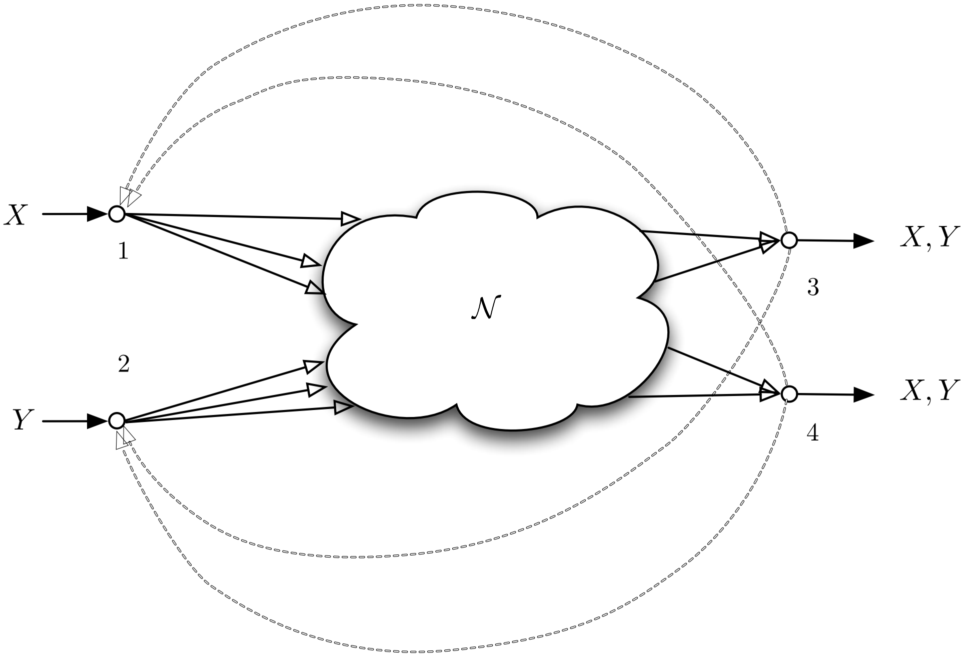


Figure 4.5: Multicast with feedback

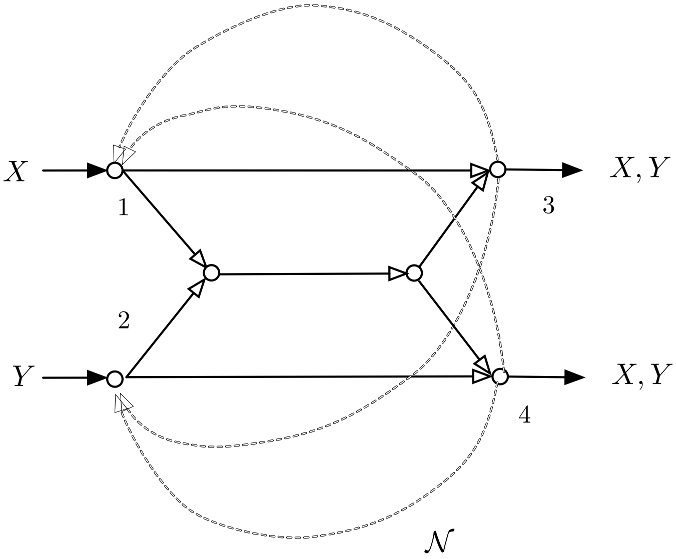


Figure 4.6: The Butterfly Network with feedback

$$\begin{aligned} f_{(1,5)}(\mathbf{x},\mathbf{y}) &= (\mathbf{x}, \alpha_Y(\mathbf{x},\mathbf{y})), \\ f_{(2,6)}(\mathbf{x},\mathbf{y}) &= (\mathbf{y}, \alpha_X(\mathbf{x},\mathbf{y})), \\ f_{(5,2)}(\mathbf{x},\mathbf{y}) &= \mathbf{x}, \\ \textit{and } f_{(6,1)}(\mathbf{x},\mathbf{y}) &= \mathbf{y} \end{aligned}$$

for all $(\mathbf{x},\mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$*. that shows that* $\mathbf{R}$ *is achievable:*

- *Transmit source* X *to sink* 5 *at rates* $H(X)$ *on the link* $(1,5)$*, and* Y *to sink* 6 *at rate* $H(Y)$ *on* $(2,6)$*, respectively.*
- *After feedback, both* X *and* Y *are available at both the source nodes. Perform a Slepian-Wolf encoding to transmit* Y *to sink* 5 *and* X *to sink* 6 *at rates* $H(Y|X)$ *and* $H(X|Y)$ *respectively.*

For a sufficiently large block lengths, both the sources can be reconstructed exactly at nodes 1 *and* 2 *with probability approaching* 1*.*

On the other hand, not all rate points satisfying Equation (4.11) *are achievable without feedback. For example, the rate vector* $\mathbf{R} = (R_e : e \in E)$*, where* $R_{15} = R_{26} = H(X,Y), R_e = 0 \ \forall \ e \notin \{(1,5),(2,6)\}$*, is not achievable without the feedback links.*

The next example is a lossy source coding problem where feedback can increase the rate region [45]. We use this result in order to prove Theorem 11.

***Example* 6** (Rate-Distortion coding with Side Information)**.** *Consider the network shown in Figure 4.7. Source* X *is observed at node* 1 *and source* Y *is present at node* 2 *as side information.The decoder demands a lossy reconstruction* $\hat{X}$ *of* X *subject to a distortion criterion* $Ed(X,\hat{X}) \leq D$*. Without feedback, the minimum rate achievable is described by the Wyner-Ziv region [45]:*

$$R_{12} = \min_{\substack{(U,g):Ed(X,g(U,Y)) \leq D \\ U \to X \to Y}} I(X;U|Y). \tag{4.12}$$

When feedback is present in the network, both the encoder and the decoder have knowledge of Y*. In this case, the minimum achievable rate is given by the conditional rate-distortion function* $R_{X|Y}(D_X)$*, given by [45]:*

$$R_{X|Y}(D_X) = \min_{U:Ed(X,U) \leq D} I(X;U|Y). \tag{4.13}$$

For some choices of sources X *and* Y *and distortion measure* d*, the expression in Eq* (4.12) *is strictly greater than* $R_{X|Y}(D_X)$ *[45]. Thus, feedback increases the rate region for this network.*

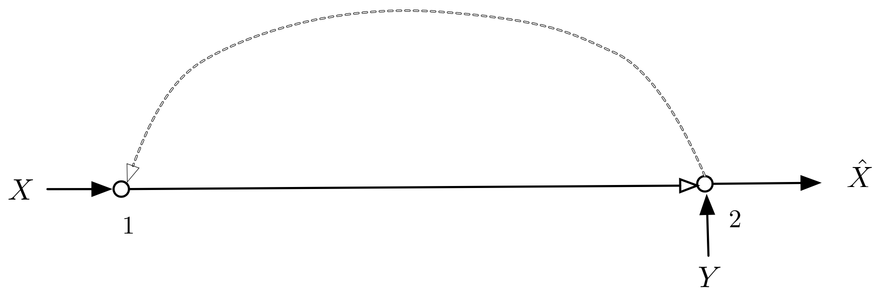


Figure 4.7: Source coding with side information

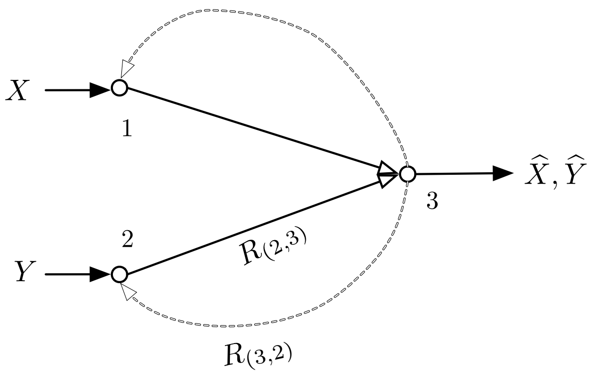


Figure 4.8: A multi-terminal source coding network

4.4 Achievable rates for multi-terminal lossy source coding with feedback

In this section, we examine the network $\mathcal{N}$ shown in Figure 4.8. Sources X and Y are sources present at nodes 1 and 2 respectively. The receiver (node 3) wishes to reconstruct both sources subject to the fidelity criteria:

$$\begin{aligned} & Ed_X(X,\hat{X}) \leq D_X \\ \text{and} \qquad & Ed_Y(Y,\hat{Y}) \leq D_Y \end{aligned}$$

where, d_X and d_Y are finite valued distortion measures, and D_X and D_Y are the respective distortion thresholds. We derive an achievable region $\mathcal{R}_{\text{in,fb}}$ with feedback, and show that $\mathcal{R}_{\text{in,fb}}$ is a strict superset of $\mathcal{R}_{\text{in}}$, the best known achievable region without feedback. This is proved by evaluating both $\mathcal{R}_{\text{in,fb}}$ and $\mathcal{R}_{\text{in}}$ for the network considered in example 6, which is a special case. .

Let $\mathcal{D}(D_X, D_Y)$ denote the set of pairs of random variables $(U,V) \in \mathcal{U} \times \mathcal{V}$ for which there exist functions $f : \mathcal{U} \times \mathcal{V} \to \hat{\mathcal{X}}$ and $g : \mathcal{U} \times \mathcal{V} \to \hat{\mathcal{Y}}$ such that. $Ed_X(X, f(U,V)) \leq D_X$ and $Ed_Y(X, g(U,V)) \leq D_Y$. Define the set $\mathcal{R}_X$ to be the set of all rate pairs (R_{13}, R_{23}) that satisfy the conditions

$$R_{(1,3)} > I(X;U|V), \tag{4.14}$$

$$\text{and} \qquad R_{(2,3)} > I(Y;V), \tag{4.15}$$

for some pair of random variables $(U,V) \in \mathcal{D}(D_X, D_Y)$ for which $X \to Y \to V$ and $U \to (X,V) \to Y$ form Markov chains. In a symmetric fashion, define the set $\mathcal{R}_Y$ to be the set of all rate pairs (R_{13}, R_{23}) that satisfy the conditions

$$R_{(1,3)} > I(X;U), \tag{4.16}$$

$$\text{and} \qquad R_{(2,3)} > I(Y;V|U), \tag{4.17}$$

for some pair of random variables $(U,V) \in \mathcal{D}(D_X, D_Y)$ for which $Y \to X \to U$ and $V \to (Y,U) \to X$ form Markov chains. Both $\mathcal{R}_{(1,2)}$ and $\mathcal{R}_{(2,1)}$ are non-empty since choosing $(U,V) = (X,Y)$ satisfies all the Markov chain conditions. Finally, let $\mathcal{R}_{\text{in,fb}}$ be the convex hull of $\mathcal{R}_X \cup \mathcal{R}_Y$, and again, let $\mathcal{R}_\infty(\mathcal{N})$ denote the set of achievable rates with feedback for the network shown in Figure 4.8. The following theorem relates $\mathcal{R}_\infty(\mathcal{N})$ and $\mathcal{R}_{\text{in,fb}}$.

***Theorem* 10.** $\mathcal{R}_\infty(\mathcal{N}) \supseteq \mathcal{R}_{\text{in,fb}}$.

The proof of this result relies on the notion of strong joint typicality [46], which is reviewed here briefly. Let $N_{abcd}(\mathbf{x},\mathbf{y},\mathbf{u},\mathbf{v})$ denote the number of occurrences of the quadruplet (a,b,c,d) in the sequence $(\mathbf{x},\mathbf{y},\mathbf{u},\mathbf{v})$. Define the typical set:

$$\begin{aligned} A_\epsilon^{(n)}(X,Y,U,V) \triangleq \{(\mathbf{x},\mathbf{y},\mathbf{u},\mathbf{v}) : \\ \left| \frac{1}{n} N_{abcd}(\mathbf{x},\mathbf{y},\mathbf{u},\mathbf{v}) - p(a,b,c,d) \right| < \frac{\epsilon}{|\mathcal{X}||\mathcal{Y}||\mathcal{U}||\mathcal{V}|} \\ \forall\ (a,b,c,d) \in \mathcal{X}\times\mathcal{Y}\times\mathcal{U}\times\mathcal{V}. \end{aligned} \tag{4.18}$$

Similarly, for each subset W of $\{X,Y,U,V\}$, define $A_\epsilon^{(n)}(W)$ to be the typical sets corresponding to n-length sequences drawn from the distribution of W. The above definition implies that if a collection of sequences is jointly typical with respect to their joint distribution, then any subset of the collection is also jointly typical with respect to the joint distribution of that subset; for example, if $(\mathbf{x},\mathbf{y},\mathbf{u},\mathbf{v}) \in A_\epsilon^{(n)}(X,Y,U,V)$, then $(\mathbf{x},\mathbf{y},\mathbf{u}) \in A_\epsilon^{(n)}(X,Y,U)$. Therefore, whenever the set of underlying random variables is clear from the context, we denote the corresponding typical set by the simplified notation $A_\epsilon^{(n)}$. Another useful property of this notion of typicality that it implies distortion typicality; namely, if $(U,V) \in \mathcal{D}(D_X,D_Y)$ and $(\mathbf{x},\mathbf{u},\mathbf{v}) \in A_\epsilon^{(n)}$, then $\frac{1}{n}\sum_{i=1}^n d_X(x_i, f(u_i,v_i)) < D_X + d_{\max}\cdot\epsilon$.

Proof of Theorem 10: By the symmetry in the definition of $\mathcal{R}_{\text{in,fb}}$ and the convexity of $\mathcal{R}_\infty(\mathcal{N})$, it suffices to show that $\mathcal{R}_X \subseteq \mathcal{R}_\infty(\mathcal{N})$. Let $\mathbf{R} = (R_{13}, R_{23}) \in \mathcal{R}_X$. By definition, there exists a pair $(U,V) \in \mathcal{D}(D_X,D_Y)$ for which $X \to Y \to V$ and $U \to (X,V) \to Y$ form Markov chains and the inequalities (4.14) and (4.15) are satisfied.

Fix an integer n and an $\epsilon > 0$. Choose R'_{13} such that $I(X,V;U) < R'_{13} < R_{13} + I(U;V)$. The reason for this choice will become clear later.To show that the rate pair (R_{13},R_{23}) is achievable, consider the following encoding and decoding strategy over a block of length n.

Codebook generation: At encoder 1, first generate $2^{nR'_{13}}$ n-length sequences sequences $\mathbf{U}(1), \mathbf{U}(2), \ldots,$ $\mathbf{U}(2^{nR'_{13}})$ by drawing each $\mathbf{U}(j)$ i.i.d. according to the probability rule

$$\Pr(\mathbf{U}(j) = \mathbf{u}) = \Pi_{i=1}^n p(u_i)$$

for every $\mathbf{u} \in \mathcal{U}^n$. Uniformly bin these $2^{nR'_{13}}$ sequences into $2^{nR_{13}}$ bins. We use $B_n(j)$ to describe the index of the bin into which $\mathbf{U}(j)$ falls. At the second encoder, generate $2^{nR_{23}}$ sequences $\mathbf{V}(1), \mathbf{V}(2), \ldots, \mathbf{V}(2^{nR_{23}})$ drawn i.i.d. with the probability rule

$$\Pr(\mathbf{V}(j) = \mathbf{v}) = \Pi_{i=1}^n p(v_i).$$

Both these codebooks are assumed known to both encoders and the decoder.

Encoding: Let $\alpha_n^{(1)}(\mathbf{Y}) = k$ if $(\mathbf{Y}, V^n(k)) \in A_\epsilon^{(n)}$. Otherwise, let $\alpha_n^{(1)}(\mathbf{Y}) = 1$. Transmit $\alpha_n^{(1)}(\mathbf{Y})$ to node 3, and also to node 1 via the feedback link. Let $\alpha_n^{(2)}(\mathbf{Y},\mathbf{X}) = B_n(j)$ if

$(\mathbf{X}, V^n(\alpha_n^{(1)}(\mathbf{Y})), U^n(j)) \in A_\epsilon^{(n)}$. Otherwise, let $\alpha_n^{(2)}(\mathbf{Y}, \mathbf{X}) = 1$.

Decoding: The decoder first decodes $\alpha_n^{(1)}(\mathbf{Y})$ to the sequence $\hat{V}^n = V^n(\alpha_n^{(1)}(\mathbf{Y}))$. Next, it looks for a sequence $\hat{U}^n$ in the bin $\alpha_n^{(2)}(\mathbf{Y}, \mathbf{X})$ s.t. $(\hat{U}^n, \hat{V}^n) \in A_\epsilon^{(n)}$. Finally, it produces the reconstructions $\hat{X}^n = f(\hat{U}_1, \hat{V}_1), \ldots, f(\hat{U}_n, \hat{V}_n)$ and $\hat{Y}^n = g(\hat{U}_1, \hat{V}_1), \ldots, g(\hat{U}_n, \hat{V}_n)$.

Since ϵ can be made arbitrarily small, it is clear that the above code can operate at rates as close to $\mathbf{R}$ as desired. Further, since d_X and d_Y are finite distortion measures, in order to show that the expected distortion of this code can be made arbitrarily close to (D_X, D_Y), it suffices to show that $\Pr(\frac{1}{n} d_X(\mathbf{X}, f(\hat{U}^n, \hat{V}^n)) > D_X + \delta)$ and $\Pr(\frac{1}{n} d_Y(\mathbf{Y}, g(\hat{U}^n, \hat{V}^n)) > D_X + \delta)$ can be made arbitrarily small for each $\delta > 0$. Thus, it is enough to prove that $\Pr(\{(\mathbf{X}, \mathbf{Y}, \hat{U}^n, \hat{V}^n) \notin A_\epsilon^{(n)}\}$ can be made arbitrarily small for each $\epsilon > 0$. Note that

$$\{(\mathbf{X}, \mathbf{Y}, \hat{U}^n, \hat{V}^n) \notin A_\epsilon^{(n)}\} \subseteq E_1 \cup E_2 \cup E_3 \cup E_4,$$

where, the events E_1, E_2, E_3, and E_4 are defined as follows:

- $E_1 = \{(\mathbf{X}, \mathbf{Y}) \notin A_\epsilon^{(n)}\}$. By AEP, the probability of this event can be made arbitrarily small by choosing n large enough.

- $E_2 = E_1^c \cap \{(\mathbf{X}, \mathbf{Y}, \hat{V}^n \notin A_\epsilon^{(n)}\}$. By noting that $X \to Y \to V$ is a Markov chain, and using the Markov lemma [47], the probability of this event can be made to asymptotically vanish with n as long as $R_{23} > I(Y; V)$ (see the proof of the rate distortion theorem in [46, 26] for further details on this argument).

- $E_3 = (E_1 \cup E_2)^c \cap \{(\mathbf{X}, \mathbf{Y}, \hat{V}^n, U^n(j)) \notin A_\epsilon^{(n)} \ \forall\ j = 1, 2, \ldots, 2^{nR'_{13}}\}$. By following a similar reasoning as above, as long as $R'_{13} > I(X, V; U)$, the probability of this event can be made arbitrarily small.

- $E_4 = (E_1 \cup E_2 \cup E_3)^c \cap \{(u^n, \hat{V}^n) \in A_\epsilon^{(n)}$ for some $u^n \neq U^n(\alpha_n^{(2)}(\mathbf{Y}, \mathbf{X}))$ s.t. u^n is in the bin $B_n(\alpha_n^{(2)}(\mathbf{Y}, \mathbf{X}))\}$. The probability of this event can be made arbitrarily small too by choosing a large enough n, the number of elements in each bin is less than $2^{nI(U;V)}$ with probability approaching 1 as n grows without bound.

Thus, for any rate $\mathbf{R} = (R_{13}, R_{23}) \in \mathcal{R}_{(1,2)}$, there exists a sequence of 2-round $((2^{nR_{13}}, 2^{nR_{23}}), n)$ codes for this network. By a similar reasoning, $\mathcal{R}_{(2,1)}$ is achievable. By the convexity of $\mathcal{R}_\infty(\mathcal{N})$, $\mathcal{R}_{\text{in,fb}}$ is achievable. Hence, $\mathcal{R}_{\text{in,fb}} \subseteq \mathcal{R}_\infty(\mathcal{N})$. □

Let $\mathcal{R}_F(\mathcal{N})$ denote the set of achievable rate pairs for the network in Figure 4.8 without the feedback links. Example 6 demonstrates that $\mathcal{R}_F(\mathcal{N}) \subsetneq \mathcal{R}_\infty(\mathcal{N})$. It should be pointed out that finding a single letter characterization of $\mathcal{R}_F(\mathcal{N})$ is not known. Berger and Tung proposed an inner bound [47, 48] $\mathcal{R}_{\text{in}} \subseteq \mathcal{R}_F(\mathcal{N})$, which was shown to be optimal for Gaussian sources [49]. For other classes of sources, the question of tightness of this bound is still open. The inner bound is defined as follows:

***Definition* 3** (Berger-Tung inner bound)**.** *[47, 48] The Berger-Tung inner bound $\mathcal{R}_{\text{in}}$ is defined to be the set of all rate pairs $(R_{(1,3)}, R_{(2,3)})$ that satisfy the conditions*

$$R_{(1,3)} > I(X;U|V), \tag{4.19}$$

$$R_{(2,3)} > I(Y;V|U), \tag{4.20}$$

$$\textit{and} \quad R_{(1,3)} + R_{(2,3)} > I(X,Y;U,V), \tag{4.21}$$

for some random variables U and V taking values in alphabets $\mathcal{U}$ and $\mathcal{V}$ respectively, and satisfying the following properties:

1. *$U \to X \to Y \to V$ forms a Markov chain, and*
2. *$(U,V) \in \mathcal{D}(D_X, D_Y)$.*

Our next result relates $\mathcal{R}_{\text{in,fb}}$ to the Berger-Tung inner bound.

***Theorem* 11.** *For every source pair (X,Y), $\mathcal{R}_{\text{in,fb}} \supseteq \mathcal{R}_{\text{in}}$. Further, there exists a source pair such that $\mathcal{R}_{\text{in,fb}} \supsetneq \mathcal{R}_{\text{in}}$.*

Proof. In order to prove that $\mathcal{R}_{\text{in,fb}} \supseteq \mathcal{R}_{\text{in}}$, first note that $\mathcal{R}_{\text{in}}$ can be viewed as the convex hull of $\mathcal{R}_{X,\text{nf}} \cup \mathcal{R}_{Y,\text{nf}}$, where $\mathcal{R}_{X,\text{nf}}$ (and in a similar manner, $\mathcal{R}_{Y,\text{nf}}$) is defined as the set of all rate pairs $\mathbf{R} = (R_{13}, R_{23}) \in \mathcal{R}_{\text{in}}$ satisfying

$$R_{(1,3)} \geq I(X;U), \tag{4.22}$$

$$R_{(2,3)} \geq I(Y;V|U). \tag{4.23}$$

To prove that $\mathcal{R}_{\text{in}} = conv(\mathcal{R}_{X,\text{nf}} \cup \mathcal{R}_{Y,\text{nf}})$, note that for each $\mathbf{R} \in \mathcal{R}_{\text{in}}$ and $\lambda \in [0,1]$,

$$\begin{aligned} R_{13} + R_{23} &> I(X,Y;U,V) \\ &= (1-\lambda)I(X,Y;U,V) + \lambda I(X,Y;U,V) \\ &= (1-\lambda)(I(X;U) + I(Y;U|X) \\ &\quad +I(Y;V|U) + I(X;V|Y,U)) \\ &\quad +\lambda(I(X;V|Y) + I(Y;V) \\ &\quad +I(Y;U|V,X) + I(X;U|V)). \end{aligned} \tag{4.24}$$

It follows that $\mathbf{R}$ can be written as a convex combination of points from $\mathcal{R}_{X,\text{nf}}$ and $\mathcal{R}_{Y,\text{nf}}$. Therefore, it is sufficient to prove that $\mathcal{R}_{X,\text{nf}} \subseteq \mathcal{R}_{(1,2)}$. This is easy to see because the Markov condition

$U \to X \to Y \to V$ that is satisfied by every element in $\mathcal{R}_{X,\text{nf}}$ implies the Markov conditions $U \to X \to Y$ and $X \to (Y,U) \to V$. Hence, $\mathcal{R}_{X,\text{nf}} \subseteq \mathcal{R}_{(1,2)}$, and therefore, $\mathcal{R}_{\text{in}} \subseteq \mathcal{R}_{\text{in,fb}}$.

Next, we show that there exist sources X and Y for which $\mathcal{R}_{\text{in,fb}} \supsetneq \mathcal{R}_{\text{in}}$. As in Example 6, let $\mathcal{X} = \mathcal{Y} = \hat{\mathcal{X}} = \hat{\mathcal{Y}}$ such that $p(X=0) = p(X=1) = 1/2$ and $p(Y=x|X=x) = p$ for some $p \in (0, 1/2)$. Let d_X and d_Y be hamming distortion measures on on $\mathcal{X} \times \hat{\mathcal{X}}$ and $\mathcal{Y} \times \hat{\mathcal{Y}}$ respectively, i.e. $d_i(z, \hat{z}) = 1$ if $z \neq \hat{z}$ and 0 is $z = \hat{z}$. Let $0 < D_X < 1/2$ and $D_Y = 0$. Fix the rate for the second encoder to be $R_{23} = H(Y) = 1$, which is achievable by choosing $V = Y$. Then, $\min\{R_{13} : (R_{13}, R_{23}) \in \mathcal{R}_{\text{in,fb}}\} = \min_{U:(U,Y)\in\mathcal{D}(D_X,0)} I(X;U|V) = R_{X|Y}(D_X)$, where $R_{X|Y}(\cdot)$ is the conditional rate distortion function of X when Y is known. Thus, the rate pair $(R_{X|Y}(D_X), H(Y))$lies in $\mathcal{R}_{\text{in,fb}}$.

On the other hand, $\min\{R_{13} : (R_{13}, R_{23}) \in \mathcal{R}_{\text{in}}\} = \min_{U:(U,Y)\in\mathcal{D}(D_X,0), T\to X\to\ Y} I(X;U|Y)$. By the result of [45], this is strictly greater than $R_{X|Y}(D_X)$. Hence, $\mathcal{R}_{\text{in,fb}} \supsetneq \mathcal{R}_{\text{in}}$. □

4.5 Finite feedback

In this section, we consider the case when the rate on the feedback link is finite.

As noted in [17], in cyclic networks, unbounded rounds of transmission may be required to achieve all points in the rate region. As a result, it may be necessary to characterize the rate region as an optimization involving possibly infinite number of terms, which may not always lead to a computable single-letter characterization.

Notwithstanding the above difficulty, we first find inner and outer bounds to the rate region by characterizing key properties of codes with feedback. Next, we consider zero error source coding with feedback and provide an achievable region with a matching converse for certain classes of probability distributions.

4.5.1 Upper and lower bounds

Consider the network $\mathcal{N}$ in Figure 4.4. We find upper and lower bounds on the rates required as follows.

4.5.1.1 Upper bound

Consider the network $\mathcal{N}_U$ shown in figure 4.9 with the vertex set $V_U = \{1, 2, 3, c\}$, edge set $E_U = \{(1,3), (1,c), (c,2), (c,3), (2,3)\}$, sources X and Y at nodes 1 and 2 respectively, and desired reproduction X at node 3. We first argue that every valid code for the network $\mathcal{N}_U$ can be modified to obtain a valid code for the network $\mathcal{N}$. This leads to a sufficiency condition for achievability of a given rate vector for the network $\mathcal{N}$ in terms of the achievable rate region for network $\mathcal{N}_U$

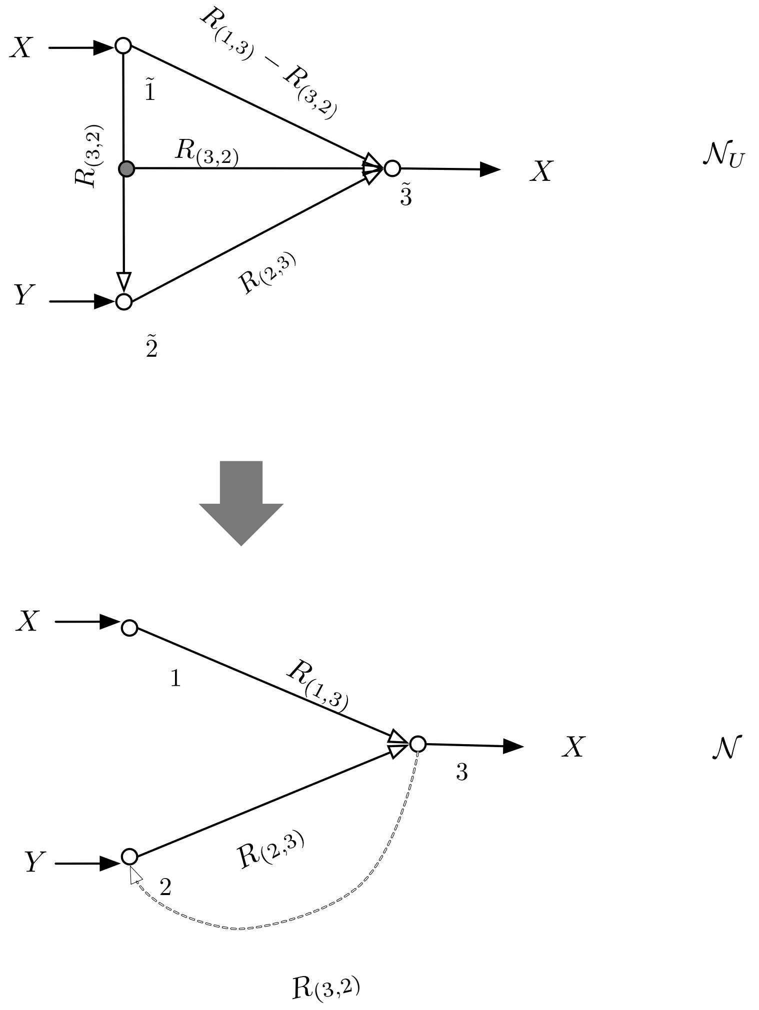


Figure 4.9: An upper-bounding network for the coded side information problem with feedback

***Theorem* 12.** $\mathbf{R} \in \mathcal{R}(\mathcal{N})$ *if there exists* $\tilde{\mathbf{R}} \in \mathcal{R}(\mathcal{N}_U)$ *such that*

(a) $\tilde{R}_{(c,v)} = \tilde{R}_{(1,c)}$ *for* $v \in \{2,3\}$.

(b) $\min_{C \subseteq V \setminus \{1\}} \sum_{u \in C \cup \{1\}} \sum_{v \notin C \cup \{1\} : (u,v) \in E} R_{(u,v)} = \tilde{R}_{(1,2)} + \tilde{R}_{(1,c)}$.

(c) $\min_{C \subseteq V \setminus \{2\}} \sum_{u \in C \cup \{2\}} \sum_{v \notin C \cup \{2\} : (u,v) \in E} R_{(u,v)} = \tilde{R}_{(2,3)}$.

(d) $\min_{C \subseteq V \setminus \{1,2\}} \sum_{u \in C \cup \{1,2\}} \sum_{v \notin C \cup \{1,2\} : (u,v) \in E} R_{(u,v)} = \tilde{R}_{(1,2)} + \tilde{R}_{(1,c)} + \tilde{R}_{(2,3)}$.

(e) $R_{(3,2)} \geq \tilde{R}_{(c,2)}$.

Proof. Fix $\epsilon > 0$. Let $\tilde{\mathbf{R}} \in \mathcal{N}_U$ such that $\tilde{R}_{(c,v)} = \tilde{R}_{(1,c)}$ for $v \in \{2,3\}$. Chose n large enough such that there exists a $((2^{n\tilde{R}_e})_{e \in E_i}, n)$ code $((\tilde{f}_e : e \in E_U), \tilde{g})$ for $\mathcal{N}_U$ such that $\Pr(\tilde{g}(\mathbf{X},\mathbf{Y}) \neq \mathbf{X}) \leq \epsilon$ and $f_{(1,c)}(\mathbf{x},\mathbf{y}) = f_{(c,2)}(\mathbf{x},\mathbf{y}) = f_{(c,3)}(\mathbf{x},\mathbf{y})$ for all $(\mathbf{x},\mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$. Note that such a solution exists because the rates on links $(1,c)$, $(c,2)$, and $(c,3)$ are equal and $H(f_{(c,2)}(\mathbf{X},\mathbf{Y})|f_{(1,c)}(\mathbf{X},\mathbf{Y})) = H(f_{(c,2)}(\mathbf{X},\mathbf{Y})|f_{(1,c)}(\mathbf{X})) = 0$.

Consider $\mathbf{R} \in \mathbb{R}^E$ that satisfies the conditions (b)-(e) with respect to the above rate vector $\tilde{\mathbf{R}}$. We design a 2-round code for $\mathcal{N}$ that implements the following steps:

1. Transmit $(\tilde{f}_{(1,c)}(\mathbf{x},\mathbf{y}), \tilde{f}_{(1,3)}(\mathbf{x},\mathbf{y}))$ from node 1 to node 3 over the edges $E \setminus F$ using single-source unicast.
2. Transmit $\tilde{f}_{(1,c)}(\mathbf{x},\mathbf{y})$ from node 3 to node 2 over the edge $(3,2)$.
3. Transmit $\tilde{f}_{(2,3)}(\mathbf{x},\mathbf{y})$ from node 2 to node 3 over the edges $E \setminus F$ using single-source unicast.

To be precise, let $\mathcal{N}^{(1)}$, $\mathcal{N}^{(c)}$ and $\mathcal{N}^{(2)}$ be networks with edge sets $E \setminus F$ and vertex sets V. Networks $\mathcal{N}^{(1)}$ and $\mathcal{N}^{(c)}$ have respectively $f_{(1,3)}(\mathbf{X},\mathbf{Y}) \in \{1,2,\ldots,2^{n\tilde{R}_{(1,3)}}\}$ and $f_{(1,c)}(\mathbf{X},\mathbf{Y}) \in \{1,2,\ldots,2^{n\tilde{R}_{(1,c)}}\}$ as sources at at node 1 and demand at node 3, while network $\mathcal{N}^{(2)}$ has $f_{(2,3)}(\mathbf{X},\mathbf{Y}) \in \{1,2,\ldots,2^{n\tilde{R}_{(2,3)}}\}$ as source at node 2 and demand node 3. Let $\mathbf{R}^{(1)} \in \mathbb{R}^{E \setminus F}$ and $\mathbf{R}^{(2)} \in \mathbb{R}^{E \setminus F}$ be rate vectors that satisfy conditions (b) and (c) respectively and $R_e^{(1)} + R_e^{(2)} = R_e$ for all $e \in E \setminus F$. The existence of such rate vectors is guaranteed since $\mathbf{R}$ satisfies (b)-(d) and the right hand side of condition (d) is simply the sum of the right hand sides of conditions (b) and (c). Let $((f_e^{(1)} : e \in E \setminus F), g^{(1)})$ be a $((2^{nR_e^{(1)}})_{e \in E \setminus F}, 1)$ random code for $\mathcal{N}^{(1)}$ that is obtained by random binning at each node, i.e., for each $(v,w) \in E \setminus F$, the map $f_{(v,w)}^{(1)} : \prod_{u:(u,v) \in E \setminus F} \{1,2,\ldots,2^{nR_{(u,v)}^{(1)}}\} \to \{1,2,\ldots,2^{nR_{(v,w)}^{(1)}}\}$ is designed by independently assigning each input codeword vector to a codeword that is chosen uniformly at random from one of the $2^{nR_{(v,w)}}$ possible values. Similarly, design a $((2^{nR_e^{(2)}})_{e \in E \setminus F}, 1)$ random code $((f_e^{(2)} : e \in E \setminus F), g^{(2)})$ for the network $\mathcal{N}^{(2)}$.

Consider a 2-round $((2^{nR_e})_{e \in E}, n)$ code $((f_e : e \in E), g)$ for $\mathcal{N}$, where the encoder and decoder functions are defined as follows:

$$f_e(\mathbf{x},\mathbf{y}) = \begin{cases} (f_e^{(1)}(\mathbf{x},\mathbf{y}), f_e^{(2)}(\mathbf{x},\mathbf{y})) & e \in E \setminus F \\ (\tilde{f}_{(1,c)}(\mathbf{x},\mathbf{y}), 0) & e = (3,2) \\ (0,0) & e = (3,1) \end{cases}$$

and

$$g(f_E(\mathbf{x},\mathbf{Y})) = \tilde{g}(g^{(1)}(f^{(1)}(\mathbf{x},\mathbf{y}), g^{(2)}(f^{(2)}(\mathbf{x},\mathbf{y}))).$$

For the code defined above, the error probability satisfies the following set of inequalities:

$$\begin{aligned} \Pr(\mathbf{X} \neq \hat{\mathbf{X}}) &= \Pr(\mathbf{X} \neq g(\mathbf{X},\mathbf{Y})) \\ &\leq \Pr(\mathbf{X} \neq \tilde{g}(\mathbf{X},\mathbf{Y})) + \Pr(\tilde{g}(\mathbf{X},\mathbf{Y}) \neq g(\mathbf{X},\mathbf{Y})) \\ &\leq 2\epsilon \end{aligned}$$

□

4.5.1.2 Lower bound

Similar to the result in previous section, we find a lower bound to the rates required for the network of Figure 4.4 by comparing it to the network shown in Figure 4.10.

4.5.2 Zero-error source coding

Next, we consider network $\mathcal{N} = (\{1,2\}, \{(1,2),(2,1)\}, X, Y, \{2\})$ shown in Figure 4.7. Source X is observed at node 1 and demanded with zero error at node 2 that also has source Y as side information.

We first derive an achievable region which is stated in the following theorem. Let $H_Z(X|Y)$ denote the minimum rate required for the zero error source coding of X when Y is known to the decoder and there is no feedback. Clearly, $H(X|Y) \leq H_Z(X|Y) \leq H(X)$. Further, it is known that each bound is achieved with equality for some choices of P_{XY}. However, a general characterization of $H_Z(X|Y)$ is known only asymptotically, i.e., through a possibly infinite sequence of minimizations over increasingly large alphabets [38],[39],[40],[41]. Theorem 13 derives an achievable subset of $\mathcal{R}(\mathcal{N})$ as a function of $H_Z(X|Y)$. Using this result, any upper bound on $H_Z(X|Y)$ gives an achievable region for the feedback network.

***Theorem* 13.** $\mathcal{R}(\mathcal{N}) \supseteq \{(R_{(1,2)}, R_{(2,1)}) : R_{(1,2)} \geq H(X|Y),\ R_{(1,2)} + R_{(2,1)} \geq H_Z(X|Y)\}$.

In order to prove the above result, we first show a weaker version of Theorem 13 in the following lemma.

***Lemma* 8.** $\mathcal{R}(\mathcal{N}) \supseteq \{(R_{(1,2)}, R_{(2,1)}) : R_{(1,2)} \geq H(X|Y),\ R_{(1,2)} + R_{(2,1)} \geq H(X)\}$

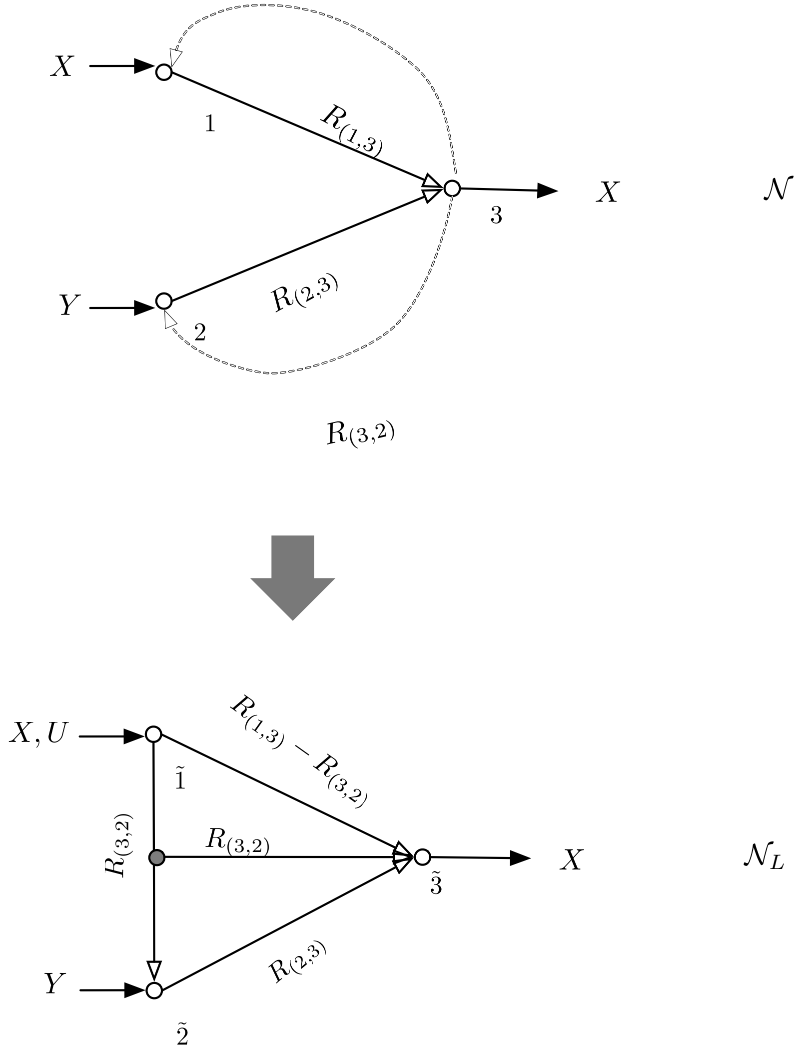


Figure 4.10: A lower-bounding network for the coded side information problem with feedback

Proof. Let $R > H(X|Y)$. Consider the following code construction.

Fix a block length $n \geq 1$. Partition $A_\epsilon^{(n)}(X)$ into 2^{nR} bins $\{\mathcal{B}_i : i = 1, 2, \ldots, 2^{nR}\}$ by assigning each $\mathbf{x} \in A_\epsilon^{(n)}(X)$ a bin chosen uniformly at random. Define $B : A_\epsilon^{(n)}(X) \to \{1, 2, \ldots, 2^{nR}\}$ to be the mapping from sequences in $\mathcal{X}^n$ to the corresponding bin number. For $i = 1, 2, \ldots, 2^{nR}$, denote by $I^i : \mathcal{B}_i \to \{1, 2, \ldots, |\mathcal{B}_i|\}$ a numbering of sequences in the i-th bin.

Suppose that $\mathbf{x}$ and $\mathbf{y}$ are the n-length sequences observed by node 1 and 2 respectively. Consider the n-dimensional, 2-round variable rate code (f, g) defined by the following protocol:

1. Node 1 sends $f(\mathbf{x}, \mathbf{y})$, where

$$f_1(\mathbf{x}, \mathbf{y}) = \begin{cases} 0\ e(\mathbf{x}) & \text{if } \mathbf{x} \notin A_\epsilon^{(n)}(X) \\ 1\ e(B(\mathbf{x})) & \text{otherwise.} \end{cases}$$

2. If $f_1(\mathbf{x}, \mathbf{y}) = 0e(\mathbf{x})$, then the procedure stops. Else,

   a. $\mathcal{E}_y$ sends

$$g_1(\mathbf{x}, \mathbf{y}) = \begin{cases} 0 & \text{if } \mathbf{y} \notin A_\epsilon^{(n)}(Y) \text{ or} \\ & \quad (\tilde{\mathbf{x}}, \mathbf{y}) \notin A_\epsilon^{(n)}(X, Y) \\ & \qquad \forall\ \tilde{\mathbf{x}} \in \mathcal{B}_{B(\mathbf{x})} \\ 1\ e(I^{B(\mathbf{x})}(\hat{\mathbf{x}})) & \text{otherwise} \end{cases}$$

   where,

$$\hat{x} = \arg \min_{\tilde{\mathbf{x}} \in \mathcal{B}_{B(\mathbf{x})} : (\tilde{\mathbf{x}}, \mathbf{y}) \in A_\epsilon^{(n)}(X,Y)} I^{B(\mathbf{x})}(\tilde{\mathbf{x}}).$$

   b. $\mathcal{E}_x$ sends

$$f_2(\mathbf{x}, \mathbf{y}) = \begin{cases} 0\ e(\mathbf{x}) & \text{if } g_1(\mathbf{x}, \mathbf{y}) = 0 \text{ or} \\ & \quad g_1(\mathbf{x}, \mathbf{y}) \neq 0 \text{ and} \\ & \qquad I^{B(\mathbf{x})}(\hat{\mathbf{x}}) \neq I^{B(\mathbf{x})}(\mathbf{x}) \\ 1 & \text{otherwise.} \end{cases}$$

Since the mapping from $\mathbf{x}$ to the pair $(B(\mathbf{x}), I^{B(\mathbf{x})}(\mathbf{x}))$ is one-to-one, the above protocol ends with $\mathcal{E}_y$ decoding $\mathbf{x}$ correctly for each $\mathbf{x} \in \mathcal{X}^n$. Let P_n denote the probability that the sequence of transmissions is

$$(f_1(\mathbf{x}, \mathbf{y}), g_1(\mathbf{x}, \mathbf{y}), f_2(\mathbf{x}, \mathbf{y})) = (1\ e(B(\mathbf{x})),\ 1\ e(I^{B(\mathbf{x})}(\mathbf{x})), 1)$$

Let $R_{(1,2)}^n$ and $R_{(2,1)}^n$ denote the expected rates on the forward link and the backward link respectively. These can be bounded from above as

$$\begin{aligned} R^n_{(1,2)} &\leq P_n(R+2/n)+(1-P_n)(H(X)+3/n) \\ \text{and } R^n_{(2,1)} &\leq P_n((1/n)\mathbf{E}[\log|\mathcal{B}_{B(\mathbf{X})}|+1])+(1-P_n)(1/n) \\ &= P_n((H(X)-R+\epsilon)+(1-P_n)(1/n). \end{aligned}$$

Following previous results on random binning (c.f.[6]), it is easily seen that P_n approaches 1 as n grows without bound. Thus,

$$\begin{aligned} \limsup_{n\to\infty} R^n_{(1,2)} &\leq R \\ \text{and } \limsup_{n\to\infty} R^n_{(2,1)} &\leq H(X)-R+\epsilon. \end{aligned}$$

Since ϵ is arbitrary, this proves the desired result. □

Theorem 13 can now be proved as follows.

Proof of Theorem 13: Let $R > H_Z(X|Y)$. It follows from the operational definition of $H_Z(X|Y)$ that for some block length n, there exists a function $c : \mathcal{X}^n \to \mathcal{C}$ satisfying the following properties:

1. Let $\mathbf{x}, \mathbf{x}' \in \mathcal{X}^n$ such that there exists $\mathbf{y} \in \mathcal{Y}^n$ for which $p_{XY}(\mathbf{x},\mathbf{y}) > 0$ and $p_{XY}(\mathbf{x}',\mathbf{y}) > 0$. Then, $c(\mathbf{x}) \neq c(\mathbf{x}')$.
2. $H(c(\mathbf{X})) = nR$.

Observe that knowing $c(\mathbf{x})$ is sufficient for $\mathcal{E}_y$ to decode $\mathbf{x}$ with zero error. Let

$$\underline{\mathcal{N}}_c \triangleq (\{1,2\},\{(1,2),(2,1)\},c(\mathbf{X}),\mathbf{Y},\{2\})$$

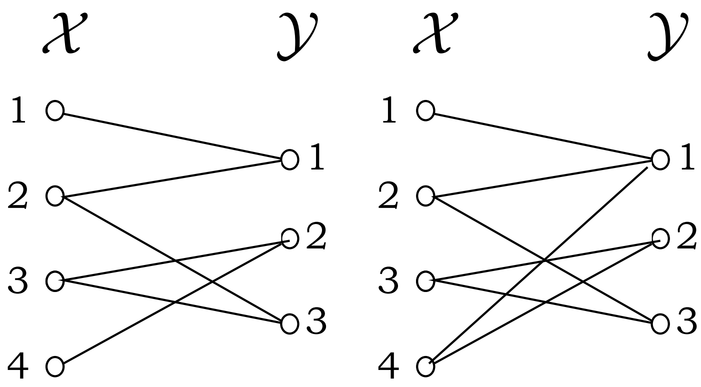
be a network derived from $\mathcal{N}$ by replacing the sources X and Y by $c(\mathbf{X})$ and $\mathbf{Y}$, respectively, and a $c(\mathbf{X})$ demanded with zero error at node 2. It follows that $\mathcal{R}(\mathcal{N}) \supseteq \frac{1}{n}\mathcal{R}(\underline{\mathcal{N}}_c)$. Further, $H(c(\mathbf{X})|\mathbf{Y})) > nH(X|Y)$ from Slepian and Wolf's result [6].

Using Lemma 8, $\mathcal{R}(\underline{\mathcal{N}}_c) \supseteq \{(R_{(1,2)}, R_{(2,1)}) : R_{(1,2)} \geq H(c(\mathbf{X})|\mathbf{Y}), R_{(1,2)} + R_{(2,1)} \geq H(c(\mathbf{X}))\}$. It follows that $\mathcal{R}(\mathcal{N}) \supseteq \{(R_{(1,2)}, R_{(2,1)}) : R_{(1,2)} \geq H(X|Y), R_{(1,2)} + R_{(2,1)} \geq H_Z(X|Y)\}$. □

The above result shows the achievability of the region

$$\{(R_{(1,2)}, R_{(2,1)}) : R_{(1,2)} \geq H(X|Y), R_{(1,2)} + R_{(2,1)} \geq H_Z(X|Y)\}$$

for all probability distributions P_{XY}. When P_{XY} satisfies the property that $p_{XY}(x,y) > 0$ for all pairs $(x,y) \in \mathcal{X} \times \mathcal{Y}$, function $H_Z(X|Y)$ is equal to $H(X)$. For this class of distributions, the above region equals $\mathcal{R}(\mathcal{N})$. We state this as Theorem 14.



(a) A cycle-free p.m.f. (b) A a p.m.f. with a cycle

Figure 4.11: Example of $G(P_{XY})$ for two different distributions

***Theorem* 14.** *Let* $X \in \mathcal{X}$ *and* $Y \in \mathcal{Y}$ *be random variables such that* $p_{XY}(x, y) > 0$ *for all* $(x, y) \in \mathcal{X} \times \mathcal{Y}$*. Then,* $\mathcal{R}(\mathcal{N}) = \{(R_{(1,2)}, R_{(2,1)}) : R_{(1,2)} \geq H(X|Y), R_{(1,2)} + R_{(2,1)} \geq H(X)\}$.

Proof. The achievability of the given rates follows from Theorem 13. Note that $H_Z(X|Y)$ is equal to $H(X)$ as for any block length n, all pairs $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ may be observed with non-zero probability and thus, without feedback, the encoder has to distinguish between all possible observations of $\mathbf{x} \in \mathcal{X}^n$ in order to achieve precisely zero error.

The converse follows from the lower bound on the sum $R_{(1,2)} + R_{(2,1)}$ derived in Corollary 1 of [15]. In particular, it is shown that for distributions that satisfy the conditions of theorem, every rate pair $(R_{(1,2)}, R_{(2,1)}) \in \mathcal{R}(\mathcal{N})$ satisfies

$$R_{(1,2)} + R_{(2,1)} \geq H(X).$$

From the converse to the Slepian-Wolf result, it also follows that

$$R_{(1,2)} \geq H(X|Y).$$

This completes the proof of theorem. □

For a probability mass function P_{XY} on $\mathcal{X} \times \mathcal{Y}$, define $G(P_{XY})$, the connectivity graph of X and Y, as a graph with vertices $\mathcal{X} \cup \mathcal{Y}$ and edges $\{(x, y) : p_{XY}(x, y) > 0\}$. We say that P_{XY} is cycle-free if $G(P_{XY})$ has no cycles. See Figure 4.11 for an example of such a probability mass function. We next show that the rate $(R_{(1,2)}, R_{(2,1)}) = (H(X|Y), 0)$ is in the zero error region $\mathcal{R}(X, Y)$ when P_{XY} is cycle-free.

***Theorem* 15.** *Let X and Y be random variables drawn from a joint distribution P_{XY} that is cycle-free. Then,* $\mathcal{R}(\mathcal{N}) = \{(R_{(1,2)}, R_{(2,1)}) : R_{(1,2)} \geq H(X|Y), R_{(2,1)} \geq 0\}$.

Proof. The converse follows immediately from the Slepian-Wolf problem since the rate required on the forward link for zero-error coding is no less than the rate required for the asymptotically lossless Slepian-Wolf code. We now show the achievability of the claimed rates.

As in the proof of Theorem 13, let $R > H(X|Y)$ and partition $A_\epsilon^{(n)}(X)$ into 2^{nR} bins $\{\mathcal{B}_i : i = 1, 2, \ldots, 2^{nR}\}$ by assigning each $\mathbf{x} \in A_\epsilon^{(n)}(X)$ a bin chosen uniformly at random. Let $B : A_\epsilon^{(n)}(X) \to \{1, 2, \ldots, 2^{nR}\}$ denote the corresponding mapping from sequences in $A_\epsilon^{(n)}(X)$ to bin numbers. Let $\mathbf{x}$ and $\mathbf{y}$ be n-length sequences observed at $\mathcal{E}_x$ and $\mathcal{E}_y$ respectively. Number all the types $P \in \mathcal{P}_n(\mathcal{X})$ and denote the index of type of $\mathbf{x}$ by $\hat{T}^n(\mathbf{x})$. Consider the following protocol.

1. $\mathcal{E}_x$ sends $f_1(\mathbf{x}, \mathbf{y}) = e(\hat{T}^n(\mathbf{x}))$.

2. $\mathcal{E}_y$ sends
$$g_1(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{if } (\mathbf{x}, \mathbf{y}) \in A_\epsilon^{(n)}(X, Y) \\ 0 & \text{otherwise.} \end{cases}$$

3. $\mathcal{E}_x$ sends
$$f_2(\mathbf{x}, \mathbf{y}) = \begin{cases} e(B(\mathbf{x})) & \text{if } g_1(\mathbf{x}, \mathbf{y}) = 1 \\ e(\mathbf{x}) & \text{otherwise} \end{cases}$$

4. If there is a unique $\mathbf{x}' \in \mathcal{B}_{B(\mathbf{x})}$ such that $(\mathbf{x}', \mathbf{y}) \in A_\epsilon^{(n)}(X, Y)$, or if $g_1(\mathbf{x}, \mathbf{y}) = 0$, transmission stops. Otherwise, $\mathcal{E}_y$ sends $g_2(\mathbf{x}, \mathbf{y}) = 0$.

5. $\mathcal{E}_x$ sends $f_3(\mathbf{x}, \mathbf{y}) = e(\mathbf{x})$

Lemma 9 shows that given individual type classes of $\mathbf{x}$ and $\mathbf{y}$, the joint type class is uniquely determined. Therefore, the above protocol always outputs the correct value $\mathbf{x}$. Finally using the same argument as in Theorem 13, as long as $R > H(X|Y)$, the expected rate on the forward link for the above code approaches R as n grows without bound, while the rate on the backward link approaches 0.

□

***Lemma* 9.** *Let $T^n(Q_X) \subseteq \mathcal{X}^n$, $T^n(Q_Y) \subseteq \mathcal{Y}^n$, and $T^n(Q_{XY}) \subseteq \mathcal{X}^n \times \mathcal{Y}^n$ be type classes that are consistent with each other, i.e., the marginal of Q_{XY} on $\mathcal{X}$ (resp. $\mathcal{Y}$) is Q_X (resp. Q_Y). Further, assume that Q_{XY} is cycle-free.*

Under the above conditions, if $\mathbf{x} \in T^n(Q_X)$, $\mathbf{y} \in T^n(Q_Y)$, and $Q_{X,Y}(\mathbf{x}, \mathbf{y}) > 0$, then $(\mathbf{x}, \mathbf{y}) \in T^n(Q_{XY})$.

Proof. Let $N(a, \mathbf{x})$ denote the number of occurrences of a symbol $a \in \mathcal{X}$ in the sequence $\mathbf{x}$. Likewise, define $N((a, b), (\mathbf{x}, \mathbf{y}))$ to be the number of simultaneous occurrences of the pair (a, b) in the sequence $(\mathbf{x}, \mathbf{y})$. To prove the lemma, we apply induction on the size of $\mathcal{X} \times \mathcal{Y}$. The smallest non-trivial case corresponds to $|\mathcal{X} \cup \mathcal{Y}| = 3$, i.e., either $|\mathcal{X}| = 2$ and $|\mathcal{Y}| = 1$ or $|\mathcal{X}| = 1$ and $|\mathcal{Y}| = 2$. In the first case, $N((a, b), (\mathbf{x}, \mathbf{y})) = N(a, \mathbf{x})$ for all $(a, b) \in \mathcal{X} \times \mathcal{Y}$. Thus, $\mathbf{x} \in T^n(Q_X)$ implies that $(\mathbf{x}, \mathbf{y}) \in T^n(Q_{XY})$. A similar argument holds for the second case.

Assume that the lemma is true whenever $|\mathcal{X} \cup \mathcal{Y}| < K$. Suppose now that for Q_{XY}, $|\mathcal{X} \cup \mathcal{Y}| = K$. Notice that if Q_{XY} has no cycles, then the connectivity graph $G(Q_{XY})$ on $\mathcal{X} \cup \mathcal{Y}$ has at least one vertex with exactly one edge connected to it. To see this, pick any vertex v_1 in $\mathcal{X} \cup \mathcal{Y}$ and construct a sequence of vertices $v_1, v_2, \ldots$ such that (v_i, v_{i+1}) are pairs of connected vertices and $v_i \neq v_{i+2}$ for each $i \geq 1$. Since $\mathcal{X} \cup \mathcal{Y}$ is a finite set, either $v_j = v_1$ for some $j > 1$ or the sequence terminates at a vertex v_k which has exactly one edge connected to it. Since Q_{XY} is cycle-free, it follows that the second condition must be true. Further, the vertex v_k also satisfies the property that the transition probability from v to its neighbour is 1 under Q_{XY}.

Fix any $\mathbf{x} \in T^n(Q_X)$ and $\mathbf{y} \in T^n(Q_Y)$ such that $Q_{XY}(\mathbf{x}, \mathbf{y}) > 0$. Let v be a vertex in $G(Q_{XY})$ that has exactly one connected edge. The following argument shows that $(\mathbf{x}, \mathbf{y}) \in T^n(Q_{XY})$. Since the argument is symmetrical in X and Y, without loss of generality, assume that $v \in \mathcal{X}$ and $w \in \mathcal{Y}$ be the vertex connected to v in $G(Q_{XY})$. Since $\mathbf{x} \in T^n(Q_X)$, $N(v, \mathbf{x}) = nQ_X(v)$ and therefore, $N((v, w), (\mathbf{x}, \mathbf{y})) = nQ_X(v) = nQ_{X,Y}(v, w)$.

Now, let $\mathcal{X}' = \mathcal{X} \setminus \{v\}, \mathcal{Y}' = \mathcal{Y}$, and $Q'_Y = Q_Y$. Define probability mass functions Q'_X on $\mathcal{X}'$ and Q'_{XY} on $\mathcal{X}' \times \mathcal{Y}'$ as follows:

$$\begin{aligned} Q'_X(x) &\triangleq Q_X(x)/(1 - Q_X(v)) \ \forall\ x \in \mathcal{X}' \text{ and} \\ Q'_{XY}(x, y) &\triangleq Q_{XY}(x, y)/(1 - Q_{XY}(v, w)) \\ &\qquad\qquad \forall\ (x, y) \in \mathcal{X}' \times \mathcal{Y}'. \end{aligned}$$

Let $(\mathbf{x}', \mathbf{y}')$ be the subsequence of $(\mathbf{x}, \mathbf{y})$ of length $n' = n - N(v, \mathbf{x})$ obtained by deleting the indices that correspond to occurrences of (v, w) in $(\mathbf{x}, \mathbf{y})$. It can be verified that $\mathbf{x}' \in T_X^{(n')}(Q'_X)$ and $\mathbf{y}' \in T_Y^{(n')}(Q'_Y)$. Since, Q'_{XY} is cycle-free and $|\mathcal{X}' \cup \mathcal{Y}'| = K - 1$, by the induction hypothesis, $(\mathbf{x}', \mathbf{y}') \in T^{(n')}(Q'_{XY})$. Hence, $\forall (x, y) \in \mathcal{X} \times \mathcal{Y} \setminus \{(v, w)\}$,

$$\begin{aligned} N((x, y), (\mathbf{x}, \mathbf{y})) &= N((x, y), (\mathbf{x}', \mathbf{y}')) \\ &= n' Q'_{XY}(x, y) \\ &= (n - nQ_{XY}(v, w)) Q_{XY}(x, y) \times \\ &\qquad\qquad 1/(1 - Q_{XY}(v, w)) \end{aligned}$$

$$= \quad nQ_{XY}(x,y).$$

This shows that $(\mathbf{x},\mathbf{y}) \in T^n(Q_{XY})$. $\square$

4.5.3 Discussion

We have shown that for every pair (X,Y) such that the rate $H(X|Y)$ on the forward link is not achievable without feedback, the addition of the feedback link enables us to lower the forward transmission rate. In particular, for certain classes of sources, Theorem 15 shows that even asymptotically zero feedback is useful. The following example illustrates this.

***Example* 7** (Binary erasure channel)**.** *Let random variable X be distributed uniformly on $\mathcal{X} = \{0,1\}$ and let Y be distributed on $\mathcal{Y} = \{0,E,1\}$ with the transition probability*

$$p_{Y|X}(y|x) = \begin{cases} 1-p & \textit{if } y = x \\ p & \textit{if } y = e. \end{cases}$$

From prior results (c.f. [38]), it follows that without feedback, the minimum rate for zero error coding of X is $H(X) = 1$. On the other hand, Theorem 15 shows that even with asymptotically zero feedback, a rate of $H(X|Y) = p$ is achievable on the forward link.

An interesting contrast to the above example is provided by the following example.

***Example* 8** (Binary symmetric channel)**.** *Let random variable X be distributed uniformly on $\mathcal{X} = \{0,1\}$ and let Y be distributed on $\mathcal{Y} = \{0,1\}$ with the following transition probability*

$$p_{Y|X}(y|x) = \begin{cases} 1-p & \textit{if } y = x \\ p & \textit{if } y \neq x. \end{cases}$$

The minimum rate possible without feedback for this example is the same as that in Example 7. However, the presence of asymptotically zero feedback does not reduce the minimum rate required on the forward link. From Theorem 14, using non-zero rate on the feedback link may lower rates on the forward link. In particular, $\mathcal{R}(\mathcal{N})$ is given by $\{(R_{(1,2)}, R_{(2,1)}) : R_{(1,2)} \geq H(p), R_{(1,2)} + R_{(2,1)} \geq 1\}$.

Finally, note that the cycle free condition of Theorem 15 is not necessary for asymptotically achieving a forward rate of $H(X|Y)$ with asymptotically zero feedback rate. This is shown in the following example.

***Example* 9** (Binary Erasure Channel with Two Erasures)**.** *Let X be distributed uniformly on $\mathcal{X} =$*

$\{0,1\}$ and let Y distributed on $\mathcal{Y} = \{0, E_1, E_2, 1\}$ with the transition probability

$$p_{Y|X}(y|x) = \begin{cases} 1-p & \text{if } y = x \\ p/2 & \text{if } y = e_1 \text{ or } e_2. \end{cases}$$

Lemma 10 shows that this example is equivalent to Example 7. Thus, a rate $H(X|Y) = p$ on the forward link can be achieved with asymptotically zero rate on the feedback link even though P_{XY} is not cycle free.

***Lemma* 10.** *Let $f : \mathcal{Y} \to \mathcal{Z}$ be such that $H(X|Y) = H(X|f(Y))$. Then, $\mathcal{R}(\mathcal{N}) = \mathcal{R}(X, f(Y))$.*

Proof. Clearly, $\mathcal{R}(\mathcal{N}) \supseteq \mathcal{R}(X, f(Y))$ since $\mathcal{E}_y$ can compute $f(Y)$ and hence, operate at all rate points in $\mathcal{R}(X, f(Y))$. To see the reverse inclusion, define a $\mathcal{Y}$-valued random variable Y' satisfying the Markov chain $Y - f(Y) - Y'$ and let $p_{Y'|f(Y)}(y|z) = p_{Y|f(Y)}(y|z)$ for all $(y, z) \in \mathcal{Y} \times \mathcal{Z}$. It follows that $H(X|Y) = H(X|Y')$, $p_Y = p_{Y'}$, and therefore, $p_{X|Y}(x|y) = p_{X|Y'}(x|y)$. Hence, the joint distribution of X and Y is same as that of X and Y', which implies that $\mathcal{R}(\mathcal{N}) = \mathcal{R}(X, Y')$. Finally, note that given $f(Y)$, $\mathcal{E}_y$ can generate Y' randomly. Therefore, $\mathcal{R}(X, f(Y)) \supseteq \mathcal{R}(X, Y') = \mathcal{R}(\mathcal{N})$.

Chapter 5

Side Information in Networks

5.1 Introduction

In this chapter, we investigate the network source coding rate region for networks with multiple sources and multicast demands in the presence of side information. This work generalizes earlier results on multicast rate regions without side information. For the case when side information is present only at the terminal nodes, we show that the rate region is precisely characterized by cut-set bounds and that random linear coding suffices to achieve the optimal performance. When side information is present at a non-terminal node, we present an achievable region. Finally, we apply these results to obtain an inner bound on the rate region for networks with general source-demand structures.

The rate region for a multicast network was characterized for independent sources by Ahlswede et al. in [9], wherein the achievability of the cut-set bounds was shown. Ho et al. [18] proved that cut-set bounds are tight for multicast codes on dependent sources and that linear codes suffice to achieve the optimal performance.

We consider several generalizations of this simple model that incorporate side information random variables that are jointly distributed with the source random variables. Theorem 16 treats a generalization of [18] where each terminal sink node has access to a distinct side information random variable (see Figure 5.2), showing that the cut-set region is again achievable and that linear codes suffice to achieve the optimal performance. Theorem 17 generalizes Theorem 16 to the case when some sink nodes may not be terminal nodes. Lemma 11 gives an alternative proof of this achievability result that does not rely on linear codes; this approach is useful in the derivations that follow. Theorem 19 generalizes Lemma 11 to allow side information at one non-sink node. Theorem 20 applies these results to find an achievable region for networks with general source-demand structures.

5.2 Preliminaries

5.2.1 Network model

We define a network $\mathcal{N}$ as a directed graph with a set of sources and a set of demands. The networks that we consider are defined on directed, acyclic graphs of the form $\mathcal{G} = (V, E)$, where V is the vertex set and E is the edge set.

For each $v \in V$, we use $\Gamma_i(v) \subseteq E$ and $\Gamma_o(v) \subseteq E$ to denote the incoming and outgoing edges, respectively, for node v. With some abuse of notation, we use $\Gamma_i(A) = \cup_{v \in A} \Gamma_i(v)$ and $\Gamma_o(A) = \cup_{v \in A} \Gamma_o(v)$, respectively, to represent the set of edges coming into and emerging from a set of vertices $A \subseteq V$. By a *cut*, we mean a subset of the vertex set V. For any cut $C \subseteq V$ and vertex $v \in V$, we define notation $\mathcal{I}_C(v)$ as

$$\mathcal{I}_C(v) = \begin{cases} 1 & \text{if } v \in C \\ 0 & \text{otherwise.} \end{cases}$$

5.2.2 Sources and sinks

We denote by $S \subseteq V$ and $T \subseteq V$, respectively, the sets of source and sink nodes. We assume that each source node s has exactly one output edge denoted by $e_s \in E$, and each sink node has exactly one input edge. These assumptions do not imply any loss of generality as a network with more than one output edge from a source node or more than one input edge to a sink node may be modified by adding intermediate nodes to the graph. Figure 5.1 shows this operation.

Each node $v \in V$ observes a random process $X^{(v)} \in \mathcal{X}^{(s)}$ and demands to reconstruct random processes observed at all nodes in subset S_v of V. The random process $\{(X_i^{(v)} : v \in V)\}_{i=1}^{\infty}$ is drawn i.i.d. from known probability mass function $P(\cdot)$.

We say that a node $s \in V$ is a *source node* if $X^{(s)}$ is present as a demand at another node in the network. We call a node $t \in V$ a sink node if t demands a non-null subset of the sources. Let S and T be the set of all source and sink nodes, respectively, and let $S_t \subset S$ be the set of sources that are demanded at a sink node $t \in T$. We call the collections $(X^{(s)} : s \in S)$ and $(X^{(s)} : s \notin S)$, as the set of *sources*, and *side information* for the network $\mathcal{N}$.

5.2.3 Demand models

We consider three different demand models in this chapter — multicast with side information at sinks, multicast with side information at a non-sink node, and general demands.

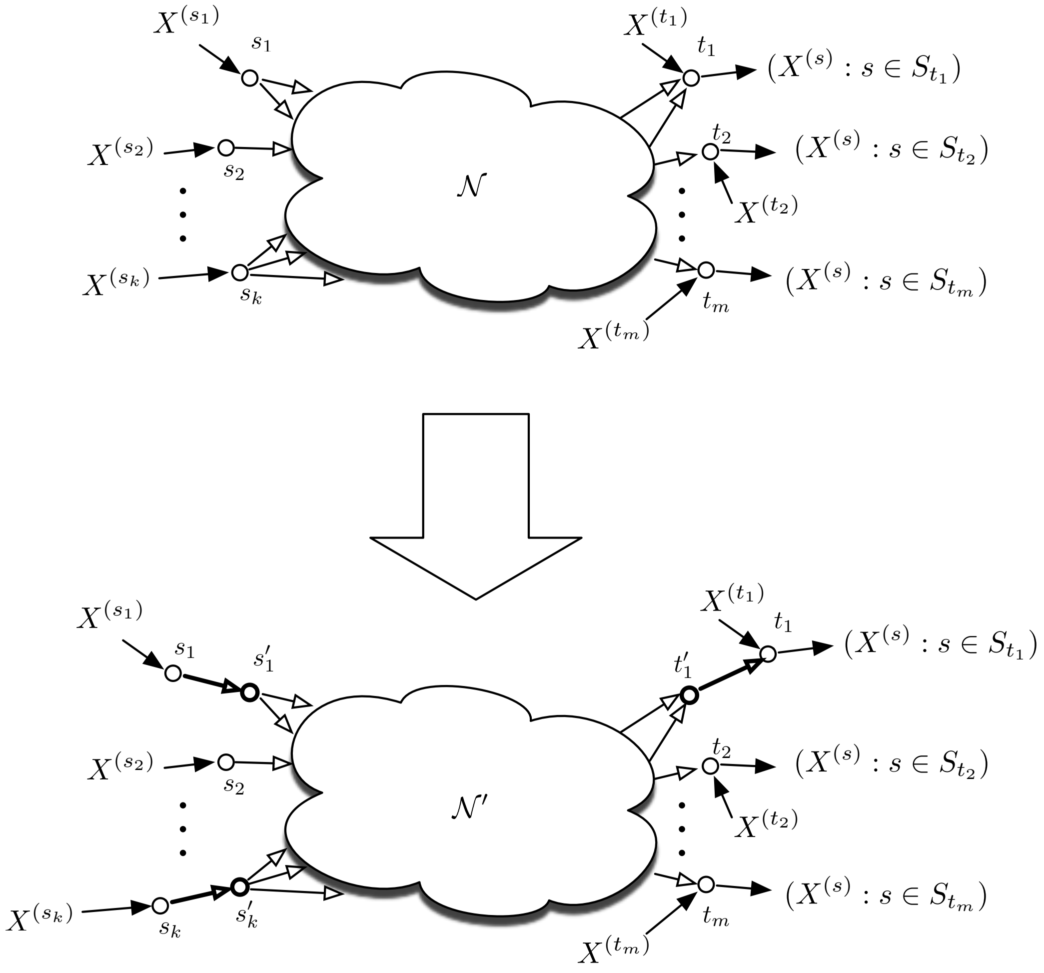


Figure 5.1: Modifying a network with more than one outgoing edge at a source node or more than one incoming edge at a sink node to a network with exactly one outgoing edge at each source node and exactly one incoming edge at each sink node

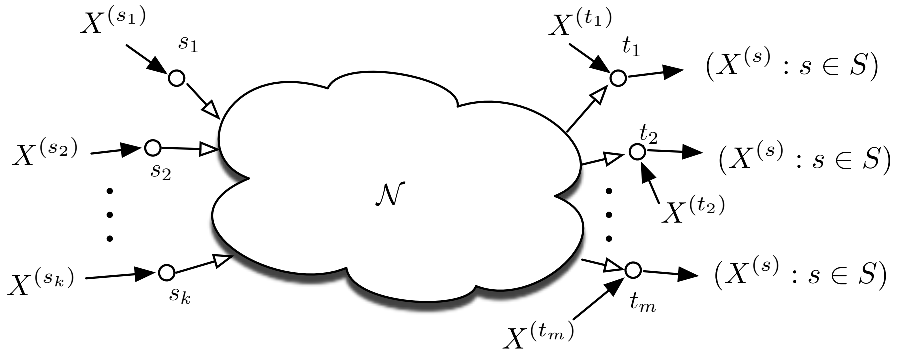


Figure 5.2: A multicast network with side information at the sink

5.2.3.1 Multicast with side information at the sinks

We say that the network $\mathcal{N}$ has *multicast demands with side information at the sinks* (Figure 5.2) if $S_t = S$ for all $t \in T$ and $X^{(v)} = 0$ for all $v \notin S \cup T$. Thus, each sink demands all the sources $(X^{(s)} : s \in T)$; the side information may differ from one sink to the next. We consider this demand model in Section 5.3.

5.2.3.2 Multicast with side information at a non-sink node

We say that the network $\mathcal{N}$ has *multicast demands with side information at a non-sink node* (Figure 5.8) if there exists a node $z \in V \setminus (S \cup T)$ such that $S_t = S$ for all $t \in T$ and $X^{(v)} = 0$ for all $v \notin S \cup T \cup \{z\}$. Thus, each sink demands all the sources $(X^{(s)} : s \in T)$; the side information may differ from one sink to the next. In addition, $X^{(z)}$ is present as a side information at a non-sink node in the network. This demand model is considered in Section 5.4

5.2.3.3 General demands

Finally, we consider a general demand model in Section 5.5. Here, we do not place any restriction on the set of sources demanded at each sink (Figure 5.10). Thus, each sink node $t \in T$ may demand a different subset $S_t \subseteq S$ with $S = \cup_{t \in T} S_t$. We assume that $X^{(v)} = 0$ for all $v \notin S$.

5.2.4 Network source codes

For any collection of rates $R_e \geq 0$, $e \in E$, an $(n, (2^{nR_e})_{e \in E})$ network source code (F_n, G_n) comprises of encoder maps $F_n = (f_n^e : e \in E)$ and decoder maps $G_n = (g_n^{(t)} : t \in T)$ with

$$\begin{aligned} f_n^{(v,v')} &: \mathcal{X}_{1:n}^{(v)} \to \{1, \ldots, 2^{nR_{(v,v')}}\} \quad \forall v \in S \\ f_n^{(v,v')} &: \prod_{e \in \Gamma_i(v)} \{1, \ldots, 2^{nR_e}\} \to \{1, \ldots, 2^{nR_{(v,v')}}\} \end{aligned}$$

$$\forall (v,v') \in \Gamma_o(V \setminus (S \cup T))$$

$$g_n^{(t)} \quad : \quad \prod_{e \in \Gamma_i(t)} \{1, \ldots, 2^{nR_e}\} \times \mathcal{X}_{1:n}^{(t)} \to \prod_{k=1}^{K} \mathcal{X}^{(k)} \quad \forall t \in T.$$

During transmission, the above maps are appropriately sequenced to ensure that at each node, the maps corresponding to the incoming edges are applied (and their outputs received) prior to applying the maps corresponding to the outgoing edges. We say that $\{(F_n, G_n)\}_{n=1}^{\infty}$ is a *valid* sequence of codes if the probability of an error at the receivers vanishes as n increases without bound; more precisely,

$$\lim_{n\to\infty} \Pr(g_n^t((f_e : e \in \Gamma_i(t)), X_{1:n}^{(t)}) \neq (X_{1:n}^{(s)} : s \in S_t)) \to 0$$

for each $t \in T$. The set of *achievable* rate vectors $\mathcal{R}(\mathcal{N})$ for the network $\mathcal{N}$ is all rates $\mathbf{R}$ for which valid sequences of $(n, (2^{nR_e})_{e\in E})$ codes exist.

5.3 Multicast with side information at the sinks

The proofs of previous multicast results without side information use random binning for code design on independent sources [9] and random linear coding for code design on (both independent and) dependent sources [18]. Random linear code design is a form of random binning that adds additional structure to the random bin choices. This extra structure is extremely useful from the perspective of implementation. Theorem 16 generalizes the proof of [18] to allow decoder side information.

To make the discussion precise, a $(n, (2^{nR_e})_{e\in E})$ *linear code* (F_n, G_n) is a set of mappings $(f_n^e : e \in E)$ and $(g_n^{(t)} : t \in T)$ such that for any $(v, v') \in E$,

$$f_n^{(v,v')} = \begin{cases} \sum_{e\in\Gamma_i(v)} a_e f_n^e & v \notin S, (v,v') \in E \\ b_n^{(v)}(X_v^n) & v \in S. \end{cases}$$

Here, $a^e \in \mathbb{F}_2^{nR_{\text{in},e} \times nR_e}$ defines a linear map with $R_{\text{in},e} = \sum_{e'\in\Gamma_i(v)} R_{e'}$ for $e = (v,v')$, and $b_n^{(s)} : (\mathcal{X}^{(s)})^n \to \mathbb{F}_2^{nR_e}$ is an arbitrary (possibly non-linear) function. The decoder mappings $(g_n^{(t)} : t \in T)$ are suitably chosen (often non-linear) functions. We say that a $(n, (2^{nR_e})_{e\in E})$ linear code (F_n, G_n) is a *random linear code* if the coefficients $\{a_e\}$ and the maps $\{b_n^{(v)}\}$ are chosen uniformly at random. Let $\mathcal{R}_L(\mathcal{N})$ denote the set of rate vectors that are achievable through linear coding, and let $\mathcal{R}_C(\mathcal{N})$ be the set of rate vectors that satisfy the cut-set bound for the given network [26, 18], i.e., $\mathbf{R} = (R_e : e \in E) \in \mathcal{R}_C(\mathcal{N})$ if and only if for any $C \subseteq V$,

$$\sum_{e\in\Gamma_o(C)} R_e \geq H((X^{(s)} : s \in C \cap S)|(X^{(s)} : s \in C^c \cap S), (X^{(t)} : t \in C^c \cap T)). \tag{5.1}$$

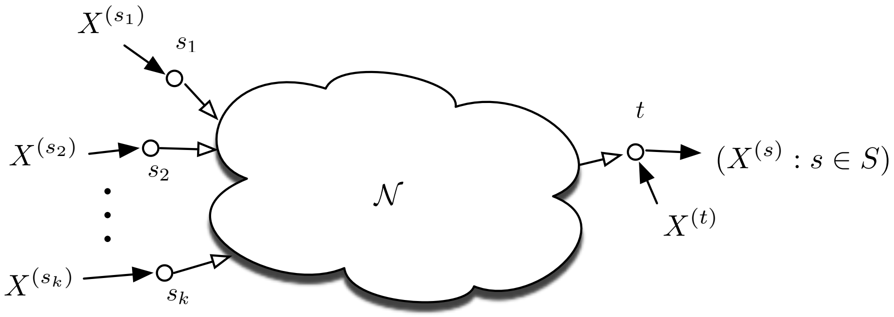


Figure 5.3: The network $\mathcal{N}_t$

The following theorem characterizes the rate region for multicast networks in the presence of side information at the sinks when each sink is a terminal node, i.e., there are no outgoing edges from any sink nodes. We also show that linear codes are sufficient achieving this region.

***Theorem* 16.** *Let $\mathcal{N}$ be a network in which each sink is a terminal node. Then, $\mathcal{R}(\mathcal{N}) = \mathcal{R}_C(\mathcal{N}) = \mathcal{R}_L(\mathcal{N})$.*

Proof. All rates achievable through random linear coding lie in $\mathcal{R}_C(\mathcal{N})$. Thus, $\mathcal{R}_L(\mathcal{N}) \subseteq \mathcal{R}(\mathcal{N}) \subseteq \mathcal{R}_C(\mathcal{N})$. In the following, we show that $\mathcal{R}_C(\mathcal{N}) \subseteq \mathcal{R}_L(\mathcal{N})$. Define $\mathcal{N}_t$ to be the network obtained from $\mathcal{N}$ by deleting all the sink nodes except the node t. The remaining network has sources $(X^{(s)} : s \in S)$ and side information $X^{(t)}$ at the only sink t. (See Fig. 5.3.) Since the side information is available at the sink, network $\mathcal{N}_t$ is equivalent to a pure multicast problem with sources $(X^{(s)} : s \in S), X^{(t)}$. Thus, by Theorem 6 of [18], random linear codes achieve any rate that satisfies:

$$\sum_{e \in \Gamma_o(C)} R_e \geq H((X^{(s)} : s \in C \cap S) | (X^{(s)} : s \in C^c \cap S), X^{(t)}) \tag{5.2}$$

for all $C \subseteq V$. Denote by $\mathcal{R}_C(\mathcal{N}_t)$ the set of rate vectors that satisfy (5.2).

Now, let $\mathbf{R} \in \cap_{t \in T} \mathcal{R}_C(\mathcal{N}_t)$, and consider a code $\{(F_n, G_n)\}$ obtained by assigning random coefficients to a linear code at rate $\mathbf{R}$. For this code,

$$\begin{aligned} & \Pr(g_n^{(t)}(f_n^e : e \in \Gamma_i(t), X^{(t)}) \neq (X^{(s)} : s \in S) \text{ for some } t \in T) \\ \leq \; & \sum_{t \in T} \Pr(g_n^{(t)}(f_n^e : e \in \Gamma_i(t), X^{(t)}) \neq (X^{(s)} : s \in S)) \\ < \; & M\epsilon \end{aligned}$$

for sufficiently large n since rate $\mathbf{R}$ is achievable for each $\mathcal{N}_t$ and the random linear encoding operation depends only on the input rates and output rates at each node. Therefore, $\mathbf{R}$ is achievable for the network $\mathcal{N}$.

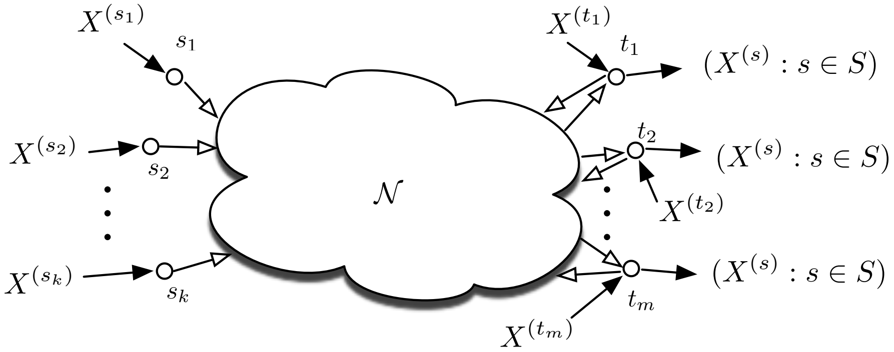


Figure 5.4: A multicast network where sinks are not terminal nodes

Next, we show that $\mathcal{R}_C(\mathcal{N}) = \cap_{t\in T}\mathcal{R}_C(\mathcal{N}_t)$. Since all $\mathbf{R} \in \cap_{t\in T}\mathcal{R}_C(\mathcal{N}_t)$ are achievable, $\cap_{t\in T}\mathcal{R}_C(\mathcal{N}_t) \subseteq \mathcal{R}_C(\mathcal{N})$. To see the reverse inclusion, let $\mathbf{R} \in \mathcal{R}_C(\mathcal{N})$. For any $C \subseteq V$, s.t. $T \nsubseteq C$ let $C_t = C \cup (T \setminus \{t\})$, and let $\tilde{C}_t = C \setminus (T \setminus \{t\})$. Then, $\Gamma_o(C_t) = \Gamma_o(\tilde{C}_t)$ and the cut-set inequality corresponding to the cut C_t in the network $\mathcal{N}$ is

$$\sum_{e\in\Gamma_o(C_t)} R_e \geq H((X^{(s)} : s \in C_t \cap S)|(X^{(s)} : s \in {C_t}^c), (X^{(t')} : t' \in {C_t}^c \cap T)), \tag{5.3}$$

which is the same as the cut-set inequality corresponding to the cut $\tilde{C}_t$ in $\mathcal{N}_t$.

□

In the proof of the above theorem, a key assumption is that the side information available at one sink is not useful for other sinks. This assumption is justified when we assume that each sink is a terminal node. It is easy to see that the above theorem also extends to the case when a sink may be a non-terminal node, but none of the sinks are downstream from another sink, i.e., there is no directed path from one sink to another. However, the argument does not extend directly to the case when a sink node may be downstream from another. The reason for this is that there may exist coding schemes that transmit the encoded version of the side information at a sink t_1 to a downstream t_2 and use it to reduce the rate required on other links.

In the following theorem, we argue that all rate vectors satisfying cut-set bounds may be achieved even when one sink node may be downstream from another. Note that the achievability of all rates vector in the cut-set region does not follow from the previous proof as achievable schemes for networks in which sink nodes are terminal nodes do not involve transmission on outgoing edges from any sink node.

***Theorem* 17.** *Let $\mathcal{N}$ be any acyclic multicast network with side information at the sinks. Then,* $\mathcal{R}(\mathcal{N}) = \mathcal{R}_C(\mathcal{N})$.

Proof. We first define the notion of *decoding order* for the network $\mathcal{N}$.

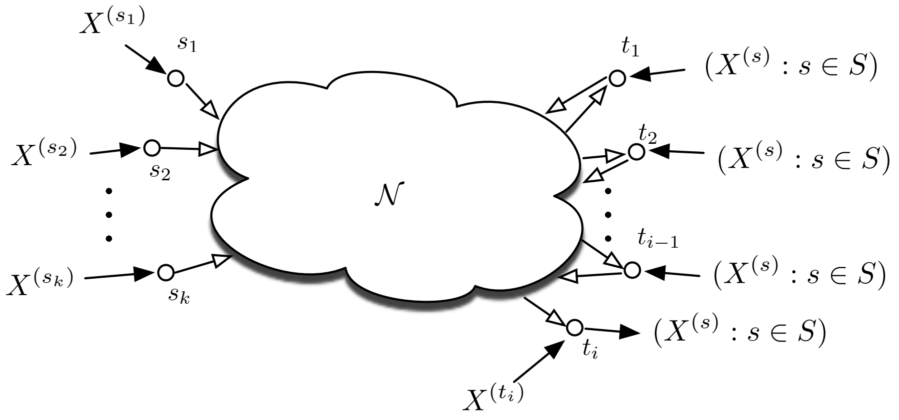


Figure 5.5: The network $\tilde{\mathcal{N}}_i$

Let (F, G) be a code for $\mathcal{N}$. Let $\prec_E$ be an ordering on the set of edges satisfying the property that if the codeword $f^{(e_2)}$ is applied after $f^{(e_1)}$, then $e_1 \prec_E e_2$. Define the ordering $\prec_V$ on the set of vertices of $\mathcal{N}$ as follows. For every pair $u, v \in V$, let $u \prec_V v$ if and only if $e_1 \prec_E e_2$ for every $e_1 \in \Gamma_i(u)$ and $e_2 \in \Gamma_i(v)$. Note that whenever a node u is downstream from a node v, the sequence of transmissions may be arranged such that $v \prec_{(V)} u$.

We define the *decoding order* for the code (F, G) as the ordering on the set of sinks T under the order $\prec_V$. For ease of notation, we use the symbol $\prec_{(F)}$ to denote the restriction of $\prec_V$ on T. Thus, the decoding order for a code (F, G) is a $|T|$-tuple $\tau(t_1, t_2, \ldots, t_{|T|})$, where $t_j \in T$ for all $j = 1, 2, \ldots, |T|$ and $t_i \prec_{(F)} t_j$ for all $i < j$. For a given $\epsilon > 0$, we say that the decoding order $\tau \in T^{|T|}$ is ϵ feasible at rate $\mathbf{R}$ if there exists a code (F, G) of rate $\mathbf{R}$ such that

1. Probability of error under code (F, G) is less than ϵ.
2. The decoding order for the code (F, G) is τ.

Fix $\epsilon > 0$. Let $\mathbf{R} \in \mathcal{R}(\mathcal{N})$ and let $(t_1, t_2, \ldots, t_{|T|})$ be an ϵ-feasible decoding order. For $i = 1, 2, \ldots, |T|$, define $\tilde{\mathcal{N}}_i$ as shown in Figure 5.5. The network $\tilde{\mathcal{N}}_i$ has vertex set $\tilde{V}_i = V \setminus \{t_j : i < j \leq |T|\}$ and edge set $\tilde{E}_i = E \setminus (\cup_{i=j+1}^{|T|}(\Gamma_i(j) \cup \Gamma_o(j)))$. Thus, $\tilde{\mathcal{N}}_i$ is the network obtained by deleting the vertices $t_{i+1}, t_{i+2}, \ldots, t_{|T|}$ and their connected edges from the network $\mathcal{N}$. Let the sources for network $\tilde{\mathcal{N}}_i$ be as follows.

$$\tilde{X}^{(v)} = \begin{cases} X^{(v)} & v \in S \\ (X^{(u)} : u \in S), & v \in T \cap E_i \setminus \{t_i\} \\ (X^{(u)} : u \in S), X^{(v)} & v = t_i \\ 0 & \text{otherwise} \end{cases}$$

The demand in network $\tilde{\mathcal{N}}_i$ is $(X^{(v)} : v \in S)$ at node t_i and 0 everywhere else.

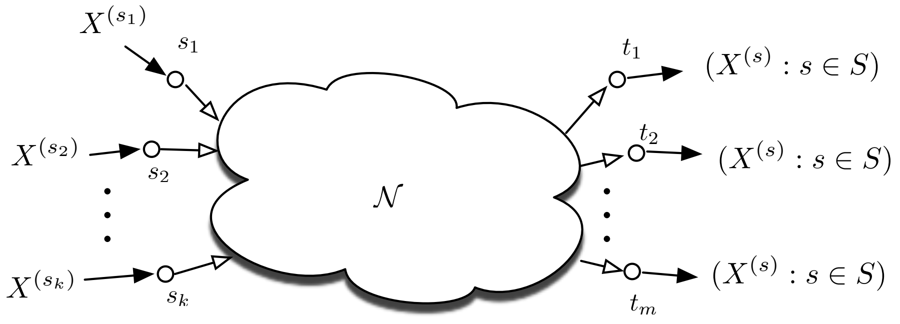


Figure 5.6: A multicast network without side information

First, we note that cut-set bounds are achievable for the network $\tilde{\mathcal{N}}_i$. This is provable by performing random linear network coding described as follows. Fix a blocklength n. Let $\mathbf{R} \in \mathcal{R}_C(\tilde{\mathcal{N}}_i)$. For each edge $e = (u, v) \in E$ such that $u \notin T$, let the map f_n^e be a randomly chosen linear map that maps the incoming codewords and the source $X_{1:n}^{(u)}$ at node u to a value in $\mathbf{F}_2^{nR_e}$. For edges $e = (u, v)$ such that $u \in T \cap V_i$, f_n^e is a randomly chosen linear map that accepts the incoming codewords at u and the vector $(X^{(w)} : w \in S)$ as input and maps it to a value in $\mathbf{F}_2^{nR_e}$.

Since the only node that observes random variables other than those belonging to the collection $(X^{(v)} : v \in S)$ is node t_i, the above network is a multicast network with side information at the sink t_i. By Theorem 16, all rate vectors in the cut-set region $\mathcal{R}_C(\tilde{\mathcal{N}}_i)$ are achievable using random linear codes.

Further, note that $\cup_{\tau=(t_1,t_2,\ldots,t_{|T|})} \cap_{i=1}^{|T|} \mathcal{R}_C(\tilde{\mathcal{N}}_i) = \mathcal{R}_C(\mathcal{N})$. This follows from arguments similar to Theorem 16. Therefore, all points in the rate region $\mathcal{R}_C(\mathcal{N})$ are achievable via random linear network coding as the restriction of a random linear network code designed for $\mathcal{N}$ to the network $\tilde{\mathcal{N}}_i$ achieves an equal or lower probability than the code for $\mathcal{N}$.

This shows that $\mathcal{R}(\mathcal{N}) = \mathcal{R}_C(\mathcal{N})$. $\square$

5.3.1 Achievability via random binning

While random linear coding is a low-complexity method for achieving random binning, the proof for a direct random binning argument requires the codewords corresponding to different inputs to be independent of each other. Thus, the above result does not imply the achievability of the rate region through random binning. In the following theorem, we generalize [9] (See Figure 5.6) first to dependent sources and then to allow side information, thereby giving an alternative proof to Theorem 16. This proof is a critical component in our proof of Theorem 19.

A $(n, (2^{nR_e})_{e\in E})$ code (F_n, G_n) is generated by *random binning* if for each $s \in S$ and each $x_{1:n} \in \mathcal{X}_{1:n}^{(s)}$, $f_n^{e_s}(x_{1:n})$ is chosen uniformly at random from alphabet $\{1, \ldots, 2^{nR_{e_s}}\}$; and for each

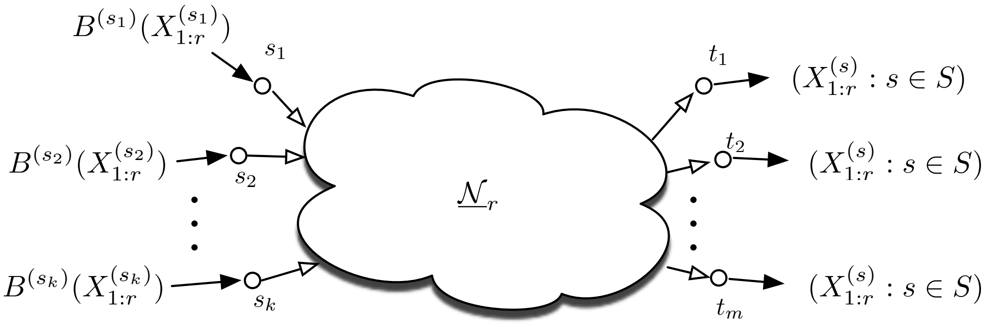


Figure 5.7: The network $\underline{\mathcal{N}}_r$ used in Lemma 11

$(v,v') \in \Gamma_o(V \setminus (S \cup T))$ and each $i \in \prod_{e \in \Gamma_i(v)}$, $f_n^{(v,v')}(i)$ is chosen uniformly at random from alphabet $\{1, \ldots, 2^{nR_{v,v'}}\}$. We use $\mathcal{R}_B(\mathcal{N})$ to denote the set of rate vectors that are achievable by for a given network using random binning.

***Theorem* 18.** *Let $\mathcal{N}$ be any multicast network with arbitrary dependence between its sources $(X^{(s)} : s \in S)$ and side information $(X^{(t)} : t \in T)$. Then, $\mathcal{R}_B(\mathcal{N}) = \mathcal{R}_C(\mathcal{N})$.*

Proof. The proof of the above result follows closely the proof of Theorem 16. As in the earlier proof, we consider the auxiliary network $\mathcal{N}_t$ for each $t \in T$, which contains exactly one sink and one side information source. Next,we select a random-binning-based code for $\mathcal{N}$ operating at a rate $\mathbf{R}$ that satisfies the cut-set inequalities for each $\mathcal{N}_t$ (and hence for $\mathcal{N}$). Finally, we note that code design using random-binning depends relies only on the rates on the links. Therefore, by Lemma 11, there exists a sequence of rate $\mathbf{R}$ codes based on random binning that achieve asymptotically vanishing error probability on $\mathcal{N}$. □

In the following lemma, we prove that random binning is sufficient to achieve all points in the rate region for multicast with dependent sources. This provides an alternative proof to the achievability region for multicast with dependent sources proved in [18].

***Lemma* 11.** *Let $\mathcal{N}$ be any multicast network with arbitrary dependence between its sources $(X^{(s)} : s \in S)$. Then, $\mathcal{R}_B(\mathcal{N}) = \mathcal{R}_C(\mathcal{N})$.*

Proof. Since all rates in $\mathcal{R}_B(\mathcal{N})$ are achievable, $\mathcal{R}_B(\mathcal{N}) \subseteq \mathcal{R}_C(\mathcal{N})$.

To prove $\mathcal{R}_B(\mathcal{N}) \supseteq \mathcal{R}_C(\mathcal{N})$, we first show that $\mathcal{R}_B(\mathcal{N})$ is a convex set. Let $\mathbf{R}, \hat{\mathbf{R}} \in \mathcal{R}_B(\mathcal{N})$. Let $\{F_n, G_n\}_{n=1}^{\infty}$ and $\{\hat{F}_n, \hat{G}_n\}_{n=1}^{\infty}$ be two valid sequences of codes that achieve the rates $\mathbf{R}$ and $\hat{\mathbf{R}}$ respectively. Define a $(n, (2^{\lceil n\lambda \rceil R_e + \lfloor n(1-\lambda) \rfloor \hat{R}_e})_{e \in E})$ code $(\tilde{F}_n, \tilde{G}_n)$ by time-sharing F_n and $\hat{F}_n$ and appropriately defining the decoder functions $\tilde{G}_n$ as follows:

$$\tilde{f}_n^e = f^e_{\lceil \lambda n \rceil} * \hat{f}^e_{\lfloor (1-\lambda) n \rfloor} \quad \forall e \in E$$

$$\tilde{g}_n^{(t)} = g_{\lceil \lambda n \rceil}^{(t)} * \hat{g}_{\lfloor (1-\lambda)n \rfloor}^{(t)} \quad \forall t \in T.$$

For the code thus formed,

$$\begin{aligned} & \Pr(\tilde{g}_n^{(t)}((\tilde{f}_n^e : e \in \Gamma_i(t)), X_{1:n}^{(t)}) \neq (X_n^{(s)} : s \in S)) \\ \leq \; & \Pr(g_{\lceil \lambda n \rceil}^{(t)}((f_{\lceil \lambda n \rceil}^e : e \in \Gamma_i(t)), X_{\lceil \lambda n \rceil}^{(t)} \neq (X_{\lceil \lambda n \rceil}^{(s)} : s \in S)) \\ & + \Pr(\hat{g}_{\lfloor (1-\lambda)n \rfloor}^{(t)}((f_{\lfloor (1-\lambda)n \rfloor}^e : e \in \Gamma_i(t)), (X^{(t)})_{\lceil \lambda n \rceil + 1:n}) \neq (X_{\lceil \lambda n \rceil + 1:n}^{(s)} : s \in S)). \end{aligned}$$

Thus,

$$\begin{aligned} & \lim_{n \to \infty} \Pr(\tilde{g}_n^{(t)}((\tilde{f}_n^e : e \in \Gamma_i(t)), X_{1:n}^{(t)}) \neq (X_n^{(s)} : s \in S)) \\ \leq \; & \lim_{n \to \infty} \Pr(g_{\lceil \lambda n \rceil}^{(t)}((f_{\lceil \lambda n \rceil}^e : e \in \Gamma_i(t)), X_{\lceil \lambda n \rceil}^{(t)} \neq (X_{\lceil \lambda n \rceil}^{(s)} : s \in S)) \\ & + \lim_{n \to \infty} \Pr(\hat{g}_{\lfloor (1-\lambda)n \rfloor}^{(t)}((f_{\lfloor (1-\lambda)n \rfloor}^e : e \in \Gamma_i(t)), (X^{(t)})_{\lceil \lambda n \rceil + 1:n}) \neq (X_{\lceil \lambda n \rceil + 1:n}^{(s)} : s \in S)) \\ = \; & 0. \end{aligned}$$

Thus, $\{(\tilde{F}_n, \tilde{G}_n)\}$ is a valid sequence of codes. Since the f_j^e's and $\hat{f}_j^e$'s are chosen independently and each of them is uniformly random mapping, it follows that so are the $\tilde{f}_n^e$'s, and hence, they have the same distribution as a code that would have been formed by random binning. Further, the rate vector corresponding to the code thus constructed approaches $\lambda \mathbf{R}_1 + (1-\lambda)\mathbf{R}_2$ asymptotically. Thus, $\lambda \mathbf{R}_1 + (1-\lambda)\mathbf{R}_2 \in \mathcal{R}_B(\mathcal{N})$. Therefore, $\mathcal{R}_B(\mathcal{N})$ is a convex set.

Now, let $\mathbf{R}$ be a boundary point of $\mathcal{R}_C(\mathcal{N})$, i.e., no component of $\mathbf{R}$ can be lowered without increasing at least one other component. We claim that $\sum_{e \in \Gamma_i(S)} R_e = H(X^{(s)} : s \in S))$.

To see this, let us assume otherwise. By the achievability of $\mathcal{R}_C(\mathcal{N})$ proved in [18], there exists a sequence of valid $(n, (2^{nR_e})_{e \in E})$ codes, say $\{(F_n, G_n)\}$.

Let $r > 1$. Let $\mathbf{R}' \in \mathbb{R}^S$ such that

$$\sum_{s' \in S'} R'_s \geq H(X^{(s')} : s' \in S' | X^{(s)} : s \in S \setminus S')$$

for all $S' \subseteq S$. Perform random binning on the inputs to obtain the network $\underline{\mathcal{N}}_r$ which has the same set of vertices and edges as $\mathcal{N}$, while the r-dimensional sources $(X_{1:r}^{(s)} : s \in S)$ are replaced by sources $(B^{(s)} : s \in S)$, where $B^{(s)} : (\mathcal{X}^{(s)})^r \to (\mathcal{X}^{(s)})^r$ is the output of a uniform random binning operation at rate R'_s. We assume that the functions $(B^{(s)} : s \in S)$ satisfy:

1. $x_{1:r} \in (B^{(s)})^{-1}(\{x_{1:r}\}) \; \forall \; x_{1:r} \in B^{(s)}((\mathcal{X}^{(s)})^r)$, and
2. $|\{B^{(s)}(x_{1:r}) : x_{1:r} \in (\mathcal{X}^{(s)})^r\}| = 2^{nR'_s}$.

The first condition can be ensured by appropriately relabeling the value of $B^{(s)}(\cdot)$ in each bin. Since the code sequence $\{(F_n, G_n)\}_{n=1}^{\infty}$ is valid for a multicast with $(X^{(s)} : s \in S)$ as sources, it follows that it is a valid sequence of codes for the multicast with sources $(B^{(s)}(X_{1:r}^{(s)}) : s \in S)$ as well. For the sequence of codes $\{(\tilde{F}_n, \tilde{G}_n)\}_{n=1}^{\infty}$, the encoding functions $(\tilde{f}_n^e : e \in E)$ are defined as

$$\tilde{f}_n^e = \begin{cases} B^{(s)} \circ f_n^e & e = e_s \text{ for some } s \in S \\ f_n^e & \text{otherwise} \end{cases}$$

and the decoding functions $(\tilde{g}_t : t \in T)$ are suitably defined. By the Slepian-Wolf theorem [6], $(B^{(s)}(X_{1:r}^{(s)}) : s \in S)$ is sufficient to reconstruct $(X_{1:r}^{(s)} : s \in S)$ with error probability vanishing asymptotically as r grows without bound, as long as $\sum_{s' \in S'} R'_s \geq H(X^{(s')} : s' \in S' | X^{(s)} : s \in S \backslash S')$. This shows that there is an achievable rate $\tilde{\mathbf{R}}$ s.t. $\tilde{R}_e \leq R_e \ \forall \ e \in E$ and $\sum_{s \in S} \tilde{R}_{e_s} = H(X^{(s)} : s \in S)$, which contradicts the assumption that $\mathbf{R}$ is a tight rate. Thus, $\sum_{s \in S} R_{e_s} = H(X^{(s)} : s \in S)$.

Define the network $\underline{\mathcal{N}}_r$ with the same set of vertices and edges as $\mathcal{N}$ and the sources $(X^{(s)} : s \in S)$ replaced by $(B^{(s)}(X_{1:r}^{(s)}) : s \in S)$, where each $B^{(s)}(\cdot)$ is a random binning operation at rate $R_s + \epsilon$. Let $\tilde{\mathcal{N}}_r$ be the network obtained by replacing $(B^{(s)}(X_{1:r}^{(s)}) : s \in S)$ in $\underline{\mathcal{N}}_r$ by $(\tilde{B}^{(s)} : s \in S)$, where $\tilde{B}^{(s)}$'s are independent of each other, but have the same first-order marginal distribution as $B^{(s)}(X_{1:r}^{(s)})$'s. By the proof used in the achievability result in [9] for the case of multicast with independent sources, the error probability for random binning codes on $\tilde{\mathcal{N}}_r$ approaches zero asymptotically. Further, since $\mathbf{R}$ is a tight point, $\sum_{s \in S} H(\tilde{B}^{(s)}) = \sum_{s \in S} H(B^{(s)}(X_{1:r}^{(s)})) < H(B^{(s)}(X_{1:r}^{(s)}, : s \in S) + r|S|\epsilon$. Since ϵ is arbitrary, it follows that by using the same code on each link as $\tilde{\mathcal{N}}_r$, we see that random binning achieves the rate vector $r\mathbf{R}$ in $\underline{\mathcal{N}}_r$, and hence, $\mathbf{R}$ in $\mathcal{N}$. To establish that the error probability for a code formed by random binning approaches 0, consider any sequence of codes $\{(\tilde{F}^{(n)}, \tilde{G}^{(n)})\}$ that is valid for the network $\tilde{\mathcal{N}}_r$. By using the same codes on the network $\hat{\mathcal{N}}_r$, the error probability satisfies the following:

$$\begin{aligned} & \Pr(\tilde{g}_n^{(t)}(\tilde{f}^e : e \in \Gamma_i(t)) \neq (B^{(s)}(X_{1:r}^{(s)} : s \in S))) \\ & = \sum_{\substack{y_{1:n} \in (\prod_{s \in S} (B^{(s)}((\mathcal{X}^{(s)})^r))^n \\ y_{1:n} : \tilde{g}_n^{(t)}(\tilde{f}^e : e \in \Gamma_i(t)) \neq y_{1:n}}} \mathrm{P}_{(B^{(s)}(X_{1:r}^{(s)}) : s \in S)}(y_{1:n}) \\ & \leq \sum_{\substack{y_{1:n} \in (\prod_{s \in S} B^{(s)}(\mathcal{X}s^r))^n \\ y_{1:n} : \tilde{g}_n^{(t)}(\tilde{f}^e : e \in \Gamma_i(t)) \neq y_{1:n}}} \mathrm{P}_{(\tilde{B}^{(s)} : s \in S)}(y_{1:n}) + d_V(P_{(B^{(s)}(X_{1:r}^{(s)}) : s \in S)}, P_{(\tilde{B}^{(s)} : s \in S)}), \end{aligned}$$

where $d_V(p, q)$ denotes the variational distance between the distributions p and q. The sequence of inequalities is furthered by the use of Pinsker's inequality as follows:

$$\Pr(\tilde{g}_n^{(t)}(\tilde{f}_n^e : e \in \Gamma_i(t)) \neq (B^{(s)}(X_{1:r}^{(s)} : s \in S)^n)$$

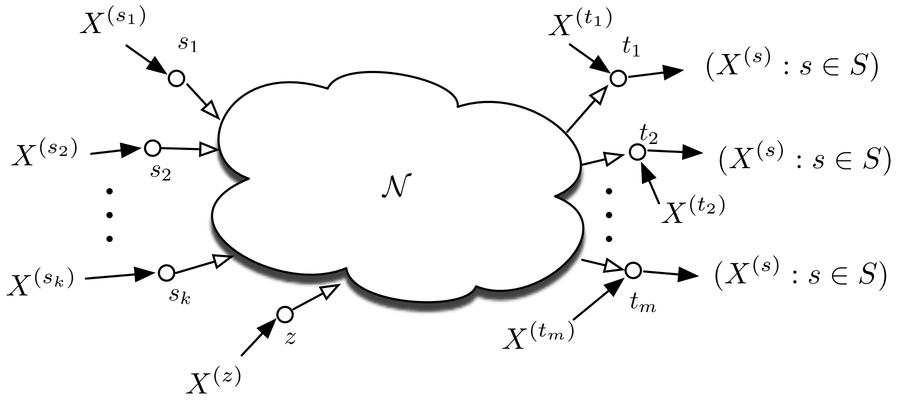


Figure 5.8: Side information at a non-sink node

$$\leq \sum_{\substack{y_{1:n} \in (\prod_{s\in S} B^{(s)}((\mathcal{X}^{(s)})^r))^n \\ y_{1:n} : \tilde{g}_n^{(t)}(\tilde{f}_n^e : e \in \Gamma_i(t)) \neq y_{1:n}}} \mathrm{P}_{(\tilde{B}^{(s)}:s\in S)}(y_{1:n}) \sqrt{2nD(P_{(B^{(s)}(X_{1:r}^{(s)}):s\in S)} || P_{(\tilde{B}^{(s)}:s\in S)})}$$

$$\leq \sum_{\substack{y_{1:n} \in \prod_{s\in S} B^{(s)}((\mathcal{X}^{(s)})^r))^n \\ y_{1:n} : \tilde{g}_n^{(t)}(\tilde{f}_n^e : e \in \Gamma_i(t)) \neq y_{1:n}}} \mathrm{P}_{(\tilde{B}^{(s)}:s\in S)}(y_{1:n}) + \sqrt{2n\epsilon}.`$$

Since the choice of ϵ is independent of n, we can make the second term vanish by choosing $\epsilon = 1/n^2$. The first term vanishes because the code $(\tilde{F}, \tilde{G})$ is a valid code for the network $\tilde{N}_r$. Thus, $\mathbf{R} \in \mathcal{R}_B(\mathcal{N})$. $\square$

5.4 Multicast with side information at a non-sink node

We now consider the case where the side information may be present at a node other than the sink nodes. (See Figure 5.8.) The key idea here is to encode the side information source into separate codewords for each subset of the sinks. We then multicast each codeword to its corresponding sinks, treating the codewords received earlier at each sink as side information. We then use Theorem 16 to find the rate required for this transmission scheme.

Again, denote the sources by $(X^{(s)} : s \in S)$ and assume that each sink node $t \in T$ has access to side information $X^{(t)}$ and wishes to reconstruct all sources. Let $z \in V \setminus (S \cup T)$ observe the side information $X^{(z)}$ and let $X^{(v)} = 0$ for all $v \notin S \cup T \cup \{z\}$. Let $\mathcal{T} = \{\tau : \tau \subseteq T\}$ and $\mathcal{P}_{\mathcal{T}} = \{\sigma : \sigma \text{ is an ordering of } \mathcal{T}\}$. The following theorem gives an achievable region for this network.

***Theorem* 19.** *Let $\{U^{(\tau)} : \tau \subseteq T\}$ be a set of random variables satisfying the Markov chains $U^{(\tau)} - X^{(z)} - (X^{(v)} : v \in S \cup T)$. Let $\sigma \in \mathcal{P}_{\mathcal{T}}$. Then, a rate vector $\mathbf{R}$ is achievable if*

$$\sum_{e \in \Gamma_o(C)} R_e^{\tau} \tag{5.4}$$

$$
\begin{aligned}
&\geq \mathcal{I}_C(z) \sum_{\tau : C^c \cap \tau \neq \phi} \Big[H\Big(U^{(\tau)} | X^{T \cap C^c}, \big(U^{(\tau')} : \tau' \in \cup_{t \in C^c \cap T} \mathcal{T}_\sigma(t,\tau)\big)\Big) \\
&-H(U^{(\tau)}|X^{(z)})\Big] + H(X^k : s_k \in C | (X^k : s_k \in C^c), (U^{(\tau)} : C^c \cap \tau \neq \phi), X^{T \cap C^c}) \qquad (5.5)
\end{aligned}
$$

for all $C \subseteq V$.

Proof. We take an approach similar to the one used in the coded side-information problem of [29]. The auxiliary random variables $U^{(\tau)}$ relate to the information present in Z that is useful to all $t \in \tau$. Consequently, we allow a different $U^{(\tau)}$ for each $\tau \subseteq T$.

For each $\tau \subseteq T$, generate $2^{n\tilde{R}_\tau}$ length-n codewords $u_{1:n}^{(\tau)}(1), \ldots, u_{1:n}^{(\tau)}(2^{n\tilde{R}_\tau})$ such that for each i, $u_1^{(\tau)}(i), \ldots, u_n^{(\tau)}(i)$ is drawn i.i.d. according to the distribution of $U^{(\tau)}$. Define the encoder mapping $h^{(\tau)} : (\mathcal{X}^{(z)})^n \to \{1, \ldots, 2^{n\tilde{R}_\tau}\}$, where $h(x_{1:n})$ is chosen to be an index i for which $(x_{1:n}, u^{(\tau)}(i)) \in A_\epsilon^{(n)}(X^{(z)}, U^{(\tau)})$. Following the proof in [29], the existence of such an index occurs with probability approaching 1. Denote the index corresponding to the random variable $U^{(\tau)}$ by $I^{(\tau)}$. Then consider the following transmission scheme.

1. Fix $\sigma \in \mathcal{P}_\mathcal{T}$.
2. For $i = 1, \ldots, 2^{|T|} - 1$,

   multicast $I^{(\sigma(i))}$ to the vertices in $\sigma(i)$ using random binning at a rate allocation $\mathbf{R}^{\sigma(i)}$.

For $t \in T$, $\sigma \in \mathcal{P}_\mathcal{T}$, and $\tau \subseteq T$, let $\mathcal{T}(t) \triangleq \{\tau \subseteq T : t \in \tau\}$, and

$$\mathcal{T}_\sigma(t,\tau) \triangleq \mathcal{T}(t) \cap \{\tau' \subseteq T : \sigma^{-1}(\tau') < \sigma^{-1}(\tau)\}.$$

In this notation, the sets $\{I^{(\tau)} : \tau \in \mathcal{T}(t)\}$ and $\{I^{(\tau')} : \tau' \in \mathcal{T}_\sigma(t,\tau)\}$ are, respectively, all the generated indices that are intended for the sink t, and all those indices that are transmitted to the sink t earlier than the index $I^{(\tau)}$.

Then, by using the binning argument of [45], and by applying the observation made in Lemma 11, in order to achieve asymptotically vanishing error probability for decoding $I^{(\tau)}$ at the sink t, it suffices to perform random binning at each intermediate node while ensuring that the rate R^τ satisfies:

$$\sum_{e \in \Gamma_o(C)} R_e^\tau \geq \mathcal{I}_C(z)[I(U^{(\tau)}; X^{(z)}) - I(U^{(\tau)}; X^{(t)}, (U^{(\tau')} : \tau' \in \mathcal{T}_\sigma(t,\tau)))]$$

for each $C \subseteq V$. Replacing the right side of the above bound by its maximum value across all sinks $t_m \in C^c$, we obtain that any $\mathbf{R}^\tau$ satisfying

$$\sum_{e \in \Gamma_o(C)} R_e^\tau \geq \mathcal{I}_C(z)[I(U^{(\tau)}; X^{(z)}) - \min_{t \in T \cap C^c} I(U^{(\tau)}; X^{(t)}, (U^{(\tau')} : \tau' \in \mathcal{T}_\sigma(t,\tau)))]$$

for each $C \subseteq V$, is sufficient to make the error probability vanish asymptotically. Simplifying the term on the right side of the inequality,

$$\begin{aligned}
&\sum_{e\in\Gamma_o(C)} R_e^\tau \\
&\geq \mathcal{I}_C(z)[I(U^{(\tau)};Z) - H(U^{(\tau)}) + \max_{t\in C^c\cap T} H(U^{(\tau)}|X^{(t)},(U^{(\tau')}:\tau'\in\mathcal{T}_\sigma(t,\tau)))] \\
&= \mathcal{I}_C(z)[\max_{t\in C^c\cap T} H(U^{(\tau)}|X^{(t)},(U^{(\tau')}:\tau'\in\mathcal{T}_\sigma(t,\tau))) - H(U^{(\tau)}|X^{(z)})].
\end{aligned}$$

Now, by using an argument similar to the one that lets us obtain the region given by (5.3) from the one given by (5.2) in Theorem 16, the above region is the same as the set of rate allocations $\mathbf{R}^\tau$ that satisfy

$$\sum_{e\in\Gamma_o(C)} R_e^\tau \geq \mathcal{I}_C(z)[H(U^{(\tau)}|(X^{(v)}:v\in T\cap C^c),(U^{(\tau')}:\tau'\in\cup_{t\in C^c\cap T}\mathcal{T}_\sigma(t,\tau))) - H(U^{(\tau)}|X^{(z)})].$$

For all $\tau'\in T_\sigma(t,\tau)$, $T_\sigma(t,\tau)\supseteq\{\tau'\}\cup T_\sigma(t,\tau')$, with equality if and only if $\sigma^{-1}(\tau')=\max_{\tau''\in T_\sigma(t,\tau)}\sigma^{-1}(\tau'')$. Therefore, adding the rates over all $\tau\subseteq T$, and the rate required to multicast $(X^{(s)}:s\in S)$ with $X^{(t)}$ and $(U^{(\tau)}:t\in\tau)$ present as side information at the sink node t, we obtain the achievability result given in ((5.4)).

□

***Remark* 1.** *When there is exactly one sink node t_1 and one source node s_1, the region described in Theorem 19 reduces to the set of vectors $\mathbf{R}$ that satisfy*

$$\sum_{e\in\Gamma_o(C)} R_e \geq \mathcal{I}_C(s_1)H(X^{(s_1)}|U) + \mathcal{I}_C(z)I(X^{(z)};U).$$

By the converse to the source coding theorem with coded side information [29], it follows that the above region is tight when no directed path from s_1 to t_1 has an edge in common with a directed path from z to t_1. Further discussion on the rate region for networks with one source appears in [50].

***Remark* 2.** *When a directed path from a source node to a sink node has one or more edges in common with a directed path from the side information to that sink, the above characterization may not include all points in the achievable rate region. We illustrate this with the help of network $\mathcal{N}$ shown in Figure 5.9. Let $X^{(1)}$ and $X^{(2)}$ be binary sources with*

$$P(X^{(1)}=x_1,X^{(2)}=x_2) = \begin{cases} 1/3 & \text{if } x_1=x_2 \\ 1/6 & \text{if } x_1\neq x_2. \end{cases}$$

Using the characterization derived in Theorem 19, a rate vector $\mathbf{R}$ *is achievable if*

$$\begin{aligned} R_{(1,3)} &\geq H(X^{(1)}|U), && (5.6)\\ R_{(2,3)} &\geq I(X^{(2)};U), && (5.7)\\ & && (5.8) \end{aligned}$$

and

$$R_{(3,4)} \geq H(X^{(1)}|U) + I(X^{(2)};U), \tag{5.9}$$

for some random variable U *satisfying the Markov chain* $X^{(1)} \to X^{(2)} \to U$*. In particular, consider the rate point* $\tilde{\mathbf{R}}$ *satisfying the above bounds that minimizes* $\tilde{R}_{(1,3)}$*. By [43], this is achieved when* $U = X^{(2)}$*. Thus,*

$$\begin{aligned} \tilde{R}_{(1,3)} &= \log_2 3 - 2/3,\\ \tilde{R}_{(2,3)} &= 1, \end{aligned}$$

and,

$$\tilde{R}_{(3,4)} = \log_2 3 + 1/3.$$

On the other hand, note that if $X^{(1)}$ *is decodable at node* 4*, it is also decodable at node* 3*, since node* 4 *has no additional information. Thus, for every code, it suffices to decode* $X^{(1)}$ *at node* 3 *and forward the reconstruction to node* 4*. Hence, for all rate points* $\mathbf{R} \in \mathcal{R}(\mathcal{N})$ *and* $\epsilon > 0$*, there exists* $\hat{\mathbf{R}} \in \mathcal{R}(\mathcal{N})$ *such that*

$$\hat{R}_e = \begin{cases} R_e & \textit{if } e \neq (3,4) \\ H(X^{(1)}) + \epsilon & \textit{if } e = (3,4). \end{cases}$$

Specifically, with reference the rate point $\tilde{\mathbf{R}}$*, the rate point* $\hat{\mathbf{R}}$ *is achievable, where*

$$\begin{aligned} \hat{R}_{(1,3)} &= \log_2 3 - 2/3,\\ \hat{R}_{(2,3)} &= 1, \end{aligned}$$

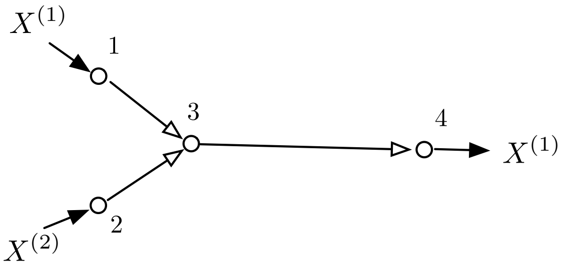


Figure 5.9: An example that shows that the conditions of Theorem 19 are not necessary

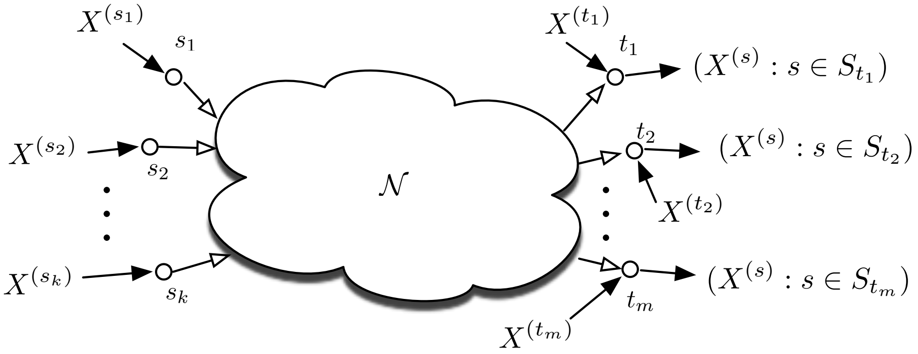


Figure 5.10: A general source-demand network

and,

$$\hat{R}_{(3,4)} = 4/3.$$

Since $\hat{R}_{(3,4)} < \log_2 3 + 1/3$, $\hat{\mathbf{R}}$ *does not satisfy Equations* (5.6)–(5.9)*. This shows that the conditions of Theorem 19 are not necessary for achievability of a rate vector.*

5.5 An inner bound on the rate region with general demand structures

In this section, we use the result of the previous section to find an inner bound on the rate region for general demands.

We again denote the sources by $(X^{(s)} : s \in S)$. The demands are denoted by $(X^{(t)} : t \in T)$, where each $X^{(t)}$ is $(X^{(s)} : s \in S_t)$ for some $S_t \subseteq S$. For $s \in S$, let $T_s = \{t : s \in S_t\}$. Let $\mathcal{P}_S = \{\sigma : \sigma \text{ is a permutation on } S\}$. For $\sigma \in \mathcal{P}_S$, let $\mathcal{R}_\sigma$ denote the set of rate vectors $\mathbf{R}$ satisfying

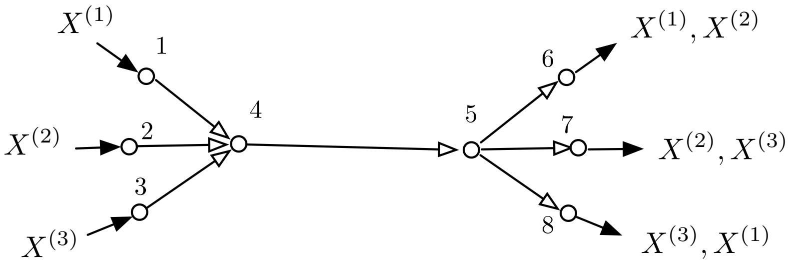


Figure 5.11: The network for Example 10

the following for all $C \subseteq V$:

$$\sum_{e\in\Gamma_o(C)} R_e \geq \sum_{k=1}^{|S|} \max_{t\in T\cap C^c} H((X^{(v)} : v \in \{\sigma(k)\} \cap S_t)|(X^{(s)} : s \in \{\sigma(1),\dots,\sigma(k-1)\} \cap S_t)), \quad (5.10)$$

and let $\mathcal{R}_i = conv(\cup_{\sigma\in\mathcal{P}_K}\mathcal{R}_\sigma)$. The following theorem asserts the achievability of $\mathcal{R}_i$.

***Theorem* 20.** *Let $\mathcal{R}$ denote the set of achievable rates for a network $\mathcal{N}$. Then,*

$$\mathcal{R}(\mathcal{N}) \supseteq \mathcal{R}_i.$$

Proof. The demands $(X^{(t)} : t \in T)$ can be met by treating them as a sequence of multicast sessions, where the demands met in previous multicast sessions are treated as side information for the current session. Thus, for any ordering $\sigma \in \mathcal{P}_S$ of S, the k-th multicast session has multicast demands $X^{\sigma(k)}$ at the sink nodes t for which $\sigma(k) \in S_t$. By Theorem 16, the rate vector $\mathbf{R}^{(\sigma,k)}$ is sufficient to meet the demands for the k-th multicast session, if the following condition is satisfied:

$$\sum_{e\in\Gamma_o(C)} R_e^{(\sigma,k)} \quad \geq \quad \max_{t\in T\cap C^c} H(X^{\{\sigma(k)\}\cap S_t}|(X^{(s)} : s \in \{\sigma(1),\dots,\sigma(k-1)\} \cap S_t)). \qquad (5.11)$$

Adding the rates required for each of the multicast sessions gives the achievability of the region $\mathcal{R}_\sigma$ given in (5.10) for each σ. By the convexity of the rate region, $\mathcal{R}_i$ is achievable. □

While the closed-form expression for the above rate-region may be difficult to analyze, it is easily computable algorithmically. It should be noted, however, that the above rate region is not tight in general. In the following example, $\mathcal{R}_i$ has no tight rate points.

***Example* 10.** *Consider the network shown in Fig 5.11. Let X^1 take values uniformly in $\{0,1,2\}$, $p_{X^{2+1}|X^2}(y|x) = 1/2$ for $y \in \{x,(x+1)(mod\ 3)\}$ and $X^3 = (1 - X^2 - X^1)(mod\ 3)$. Then, for all possible $\sigma \in \mathcal{P}_3$, any $\mathbf{R} \in \mathcal{R}_\sigma$ satisfies:*

$$R_e \quad \geq \quad \log_2 3 + 1 + 1$$

$$= \quad \log_2 3 + 2.$$

On the other hand, for any $\mathbf{R} \in \mathcal{R}_\sigma$*, the vector* (X^1, X^2, X^3) *is decodable at the node* a*. Thus, a rate of* $H(X^1, X^2, X^3) = \log_2 3 + 1$ *is sufficient on the link* e*, thereby proving the sub-optimality of* $\mathbf{R}$.

5.6 Discussion

In this chapter, we generalize earlier multicast rate region bounds to allow side information at the decoders. We also generalize to networks with side information at one intermediate node in addition to the side information at the sinks. The generalization takes an approach similar to that used in the coded side information problem. The given bounds are interesting both on their own and for their applicability in proving other interesting bounds. For example, we can bound the rate region for a network with multiple multicasts by considering each multicast in turn and treating information received from “earlier” multicasts as side information available to the corresponding sinks.

Bibliography

[1] Claude E. Shannon. A mathematical theory of communication. *Communication, Bell System Technical Journal*, 27:379–423, 1948.

[2] Robert G Gallager. *Low-density parity-check codes.* The MIT Press, 1963.

[3] E Arikan. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *Information Theory*, 2009.

[4] D.A. Huffman. A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, 40(9):1098 –1101, sept. 1952.

[5] J Ziv and Lempel. A universal algorithm for sequential data compression. *Information Theory, IEEE Transactions on*, 23(3):337–343, 1977.

[6] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, IT-19:471–480, 1973.

[7] T Cover. Broadcast channels. *Information Theory*, 1972.

[8] T Cover. Capacity theorems for the relay channel. *Information Theory*, 1979.

[9] R. Ahlswede, N. Cai, S. Y. R. Li, and R. Yeung. Network information flow. *IEEE Transactions on Information Theory*, IT-46(4):1204–1216, July 2000.

[10] R. Koetter, M. Effros, and M. Médard. A theory of network equivalenceâĂŤpart i: Point-to-point channels. *IEEE Transactions on Information Theory*, 57(2), February 2011.

[11] S. Jalali and M. Effros. On the separation of lossy source-network coding and channel coding in wireline networks. In *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pages 500 –504, june 2010.

[12] A Orlitsky and A El Gamal. Communication Complexity. In Y Abu Mustafa, editor, *Complexity in Information Theory*, pages 16–61. Springer-Verlag, 1986.

[13] A Orlitsky and A El Gamal. Average and randomized communication complexity. *IEEE Transactions on Information Theory*, 1990.

[14] A Orlitsky. Two messages are almost optimal for conveying information. *Annual ACM Symposium on Principles of Distributed Computing*, 1990.

[15] A. Orlitsky. Average-case interactive communication. *Information Theory, IEEE Transactions on*, 38(5):1534 –1547, sep 1992.

[16] En hui Yang, Da-Ke He, T. Uyematsu, and R.W. Yeung. Universal multiterminal source coding algorithms with asymptotically zero feedback: Fixed database case. *Information Theory, IEEE Transactions on*, 54(12):5575 –5590, dec. 2008.

[17] Nan Ma and P Ishwar. Two-terminal distributed source coding with alternating messages for function computation. *IEEE International Symposium on Information Theory, 2008*, pages 51 – 55, Jul 2008.

[18] T. Ho, R. Koetter, M. Médard, M. Effros, J. Shi, and D. Karger. A random linear network coding approach to multicast. *IEEE Transactions on Information Theory*, 52(10):4413–4430, October 2006.

[19] R. Koetter, M. Effros, and M. Médard. On the theory of network equivalence. In *IEEE Inform. Theory Workshop (ITW)*, 2009.

[20] S. Kim, T. Ho, M. Effros, and S. Avestimehr. New results on network error correction: Capacities and upper bounds.

[21] S. Kim, T. Ho, M. Effros, and S. Avestimehr. Network error correction with unequal link capacities. In *Proceedings of the Allerton Conference on Communication, Control, and Computing*, Monticello, IL, September 2009.

[22] O. Kosut, L. Tong, , and D. Tse. Nonlinear network coding is necessary to combat general byzantine attacks. In *Proceedings of the Allerton Conference on Communication, Control, and Computing*, Monticello, IL, September 2009.

[23] O. Kosut, L. Tong, , and D. Tse. Polytope codes against adversaries in networks. In *Proceedings of the International Symposium on Information Theory*, Austin, TX, 2010.

[24] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, and M. Medard. Resilient network coding in the presence of byzantine adversaries. In *Proceedings of INFOCOM*, pages 616–624, September 2007.

[25] R. G. Gallager. *Information Theory and Reliable Communication.* New York: John Wiley, 1968.

[26] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley, 1991.

[27] T M Cover and J A Thomas. *Elements of information theory.* Segunda Ediçao. John Wiley & Sons, 2006.

[28] WeiHsin Gu. On Achievable Rate Regions for Source Coding over Networks. *Thesis*, pages 1–154, May 2009.

[29] R. Ahlswede and J. Körner. Source coding with side information and a converse for degraded broadcast channels. *IEEE Transactions on Information Theory*, IT-21(6):629–637, November 1975.

[30] Imre Csiszar. Linear codes for sources and source networks: Error exponents, universal coding. *IEEE Transactions on Information Theory*, IT-28:585–592, July 1982.

[31] P. Pakzad, C. Fragouli, and A. Shokrollahi. Coding schemes for line networks. In *Proceedings of the IEEE International Symposium on Information Theory*, Adelaide, Australia, September 2005.

[32] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, IT-19:471–480, 1973.

[33] A. Orlitsky and J. R. Roche. Coding for computing. In *IEEE Symposium on Foundations of Computer Science*, pages 502–511, 1995.

[34] H. Yamamoto. Wyner-ziv theory for a general function of correlated sources. *IEEE Transactions on Information Theory*, IT-28(5):803–807, September 1982.

[35] H. Feng, M. Effros, and S. A. Savari. Functional source coding for networks with receiver side information. In *Proceedings of the Allerton Conference on Communication, Control, and Computing*, Monticello, IL, September 2004.

[36] C. E. Shannon. The zero-error capacity of a noisy channel. *IRE Transactions on Information Theory*, IT-2:8–19, 1956.

[37] N.T. Gaarder and J. K. Wolf. The capacity region of a multiple-access discrete memoryless channel can increase with feedback. *IEEE Transactions on Information Theory*, IT-21:100–102, jan 1975.

[38] N Alon and A Orlitsky. Source coding and graph entropies. *IEEE Transactions on Information Theory*, 42(5):1329 – 1339, Sep 1996.

[39] H Witsenhausen. The zero-error side information problem and chromatic numbers (corresp.). *IEEE Transactions on Information Theory*, 22(5):592 – 593, Jan 1976.

[40] P Koulgi, E Tuncel, S Regunathan, and K Rose. On zero-error source coding with decoder side information. *IEEE Transactions on Information Theory*, 49(1):99 – 111, Jan 2003.

[41] E Tuncel. Kraft inequality and zero-error source coding with decoder side information. *IEEE Transactions on Information Theory*, 53(12):4810 – 4816, Dec 2007.

[42] M Bakshi and M Effros. On feedback in network source coding. *IEEE International Symposium on Information Theory, 2009*, pages 1348 – 1352, Jun 2009.

[43] D. Marco and M. Effros. A partial solution for lossless source coding with coded side information. In *Proceedings of the Information Theory Workshop*, Punta del Este, Uruguay, 2006. IEEE.

[44] L.R. Ford and D.R. Fulkerson. Maximal flow through a network. *Canadian Journal of Mathematics*, 8:399âĂŞ404, 1956.

[45] A. D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Transactions on Information Theory*, IT-22(1):1–10, January 1976.

[46] I. Csiszár and J. Köner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. London, U.K.:Academic, 1981.

[47] T. Berger. Multiterminal source coding. in: The information theory approach to communications. In *CISM Courses and Lecture Notes No. 229*, pages 171–231. Springer-Verlag, 1977.

[48] S.-Y. Tung. Multiterminal source coding. Technical report, Ph.D. dissertation, School of Electrical Engineering, Cornell University, Ithaca, NY, 1978.

[49] A. Wagner, S. Tavildar, and P. Viswanath. Rate region of the quadratic gaussian two-encoder source-coding problem. *IEEE Transactions on Information Theory*, IT-54:1938–1961, May 2008.

[50] A Cohen, S Avestimehr, and M Effros. On networks with side information. *IEEE International Symposium on Information Theory, 2009*, Jun 2009.