FUNDAMENTAL LIMITS OF DELAY AND SECURITY IN DEVICE-TO-DEVICE COMMUNICATION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Ebad Ahmed

May 2013

FUNDAMENTAL LIMITS OF DELAY AND SECURITY IN DEVICE-TO-DEVICE COMMUNICATION

Ebad Ahmed, Ph.D.

Cornell University 2013

Distributed communication poses novel challenges for efficient operation of such networks and requires design considerations that are fundamentally different from those of classical point-to-point communication systems. This thesis studies two such design issues, (1) delay management and (2) security, and attempts to understand the information-theoretic limits of distributed communication with regard to these issues.

First, the tradeoff between delay and partial reconstruction in peer-to-peer (P2P) networks is studied, *i.e.,* the number of messages a peer must obtain to reconstruct a given fraction of the data. Using a binary erasure version of the multiple descriptions (MD) problem to model the P2P network, the thesis presents coding schemes based on systematic MDS (maximum distance separable) codes and random binning strategies that achieve a Pareto optimal delay-reconstruction tradeoff. The erasure MD setup is then used to propose a layered coding framework for MD, which is then applied to vector Gaussian MD and shown to be optimal for symmetric scalar Gaussian MD with two levels of receivers and no excess rate at the central receiver.

Second, delay-reconstruction tradeoffs are studied for a more decentralized network in which peers are allowed to encode and generate their own messages based on their current partial knowledge of the file, and a coding scheme based on erasure compression and Slepian-Wolf binning is presented. The coding scheme is shown to provide a Pareto optimal delay-reconstruction tradeoff for the case of symmetric peers (*i.e.,* each peer generates packets of the same rate). In the process of characterizing the aforementioned tradeoff, an improved outer bound on the rate region of the general multi-terminal source coding problem from information theory is also established. It is further shown that in the case of asymmetric peers, the aforementioned coding scheme is not optimal.

Third, lossy compression is studied from the viewpoint of security. An adversarial lossy source coding problem is considered in which a source is encoded into n packets, any t of which may be altered in an arbitrary way by Byzantine adversaries. The decoder receives the n packets and, without knowing which packets were altered, seeks to reconstruct the original source to meet a distortion constraint. A layered architecture for this problem is examined, which separates lossy compression from coding for adversarial errors. This architecture is shown to be optimal for binary sources with Hamming distortion and Gaussian sources with quadratic distortion, yet suboptimal in general.

Finally, an adversarial 3-encoder lossless source coding problem with multiple sources is considered in which the number of packets corrupted by adversaries is unknown to the honest entities in the network. It is shown that this problem is equivalent to an instance of the symmetric MLD (multi-level diversity) coding problem with three sources and three encoders, in which there are no adversaries but the decoder may receive only a subset of the three messages and then reconstructs a subset of the three sources.

BIOGRAPHICAL SKETCH

Ebad Ahmed was born in Karachi, Pakistan in 1984. After completing high school in Pakistan, he came to MIT for his undergraduate work. He received the Bachelor of Science degree in Electrical Engineering and Computer Science in 2006 and the Master of Engineering degree in Electrical Engineering and Computer Science in 2007, both from MIT, and then started his dcotoral work at Cornell University in August 2007 under the supervision of Professor Aaron Wagner. In the summer of 2011, he was at RIM Corp. in Rolling Meadows, Illinois, working on coding techniques for telecommunication standards. His research interests lie broadly in the area of information theory and coding theory with applications to multi-user networks. some of the areas he has investigated are lossy data compression techniques for content distribution, including distributed source coding for peer-to-peer networks, network coding, stability in networks, and network security.

ACKNOWLEDGEMENTS

In his *Epistulae Morales*, Letter XVI, Seneca, the Roman philosopher, says, "*Quidquid bene dictum est ab ullo meum est*. [Anything that is well said by anyone is mine.]" Seneca says this because in the next sentence, he quotes Epicurus, and in the previous sentence, he tells Lucilius, the procurator of Sicily to whom the letter is addressed: "*Non est quod mireris animum meum; adhuc de alieno liberalis sum*. [You need not wonder at any genius of mine; for as yet I am lavish only with other men's property.]" I share Seneca's sentiment as I write this dissertation, for although the document bears my name as the sole author, I certainly cannot bear sole credit for this work.

I thank my advisor Aaron Wagner: without him this thesis would definitely not have been possible. His contribution, indeed, extends far beyond these 141 pages. He taught me how to think, approach and analyze theoretical problems with thoroughness and rigor. These are skills that will remain with me all my life. He made research engaging and fun, and his advice and insights on life beyond graduate school have benefited me immensely.

I thank my thesis committee members Kevin Tang, Salman Avestimehr, and Stephen Wicker for their invaluable feedback and suggestions on my research. Kevin Tang was not only interested in my work but also genuinely helpful when I was trying to decide my next step after graduate school, and I sincerely thank him for that. Salman Avestimehr provided extremely useful insights on the link between my research and network coding. His class on advanced topics in network information theory not only broadened my understanding of the subject but also exposed me to the latest ideas in communication networks and peer-to-peer systems. Stephen Wicker not only provided feedback on my research but also taught me coding theory, which informs the bulk of Chapter 2, and wireless communications. His lectures were fun and I particularly enjoyed how questions on our mid-terms ranged from coding theory and mobile communications to volcanoes in Iceland and Dante's *Inferno*.

I thank Lang Tong for his support and encouragement throughout the five years that I spent here. He has been most helpful since my very first day in Ithaca, which, incidentally, was during a visit weekend in Spring 2007 before I started my Ph.D., and he is one of the reasons I chose Cornell for graduate school. His work on Byzantine networks is in fact what got me interested in the ongoing research in this area, and Chapters 4 and 5 are a culmination of that interest. I thank Scott Coldren and Daniel Richter: without them life would be a lot more complicated for ECE faculty and students. I thank my group members and colleagues in ECE: Benjamin Kelly, Saifur Rahman, Amine Laourine, Yücel Altuğ, Ilan Shomorony, Alireza Vahid, Nithin Michael, Meng Wang, Enrique Mallada, Ania Kacewicz, Sergio Bermudez, and Oliver Kosut. I am indebted to Oliver in particular for his insights on Byzantine networks and his proof for Theorem 16 in Section 4.4.

I thank my professors and colleagues from undergrad who helped me develop my interest and expertise in the field of communication networks: Muriel Médard, Asuman Özdağlar, Atilla Eryılmaz, and Desmond Lun. I thank Eoin Buckley for being an excellent supervisor and a constant source of support during my internship at RIM.

I thank Dr. Shawkat Toorawa, Aunty Parvine, Maryam, and Asiya for opening their house to us and making Mews Hall a home away from home, and for the familial feel they imparted to Cornell. My Cornell experience would most definitely not have been the same without them. Dr. T, specifically for teaching me all the stuff that I otherwise might never have learned, Aunty Parvine specifically for her amazing culinary skills, and Maryam and Asiya for all the fun times at Mews. I thank other friends at Cornell for the friendship and good memories: Ahmed Jaber, Anas Abognah, Andrew Amstutz, John Robbins, Hadi Fathallah, Muhammad Adnan, Amandeep Singh, Ajeet Kumar, Swarnavo Sarkar, Hamza Mahmood, Wasif Syed, Gökhan Arıkan, Shaan Qamar, Abdurrahman Gümüş, Mohamed Elhawary, Wan Lutfi Wan Johari, Ian Hossein, Adil Gangat, Mahmud Burton, Aaqib Habib, Zubair Azad, Jasmin Sahbaz, Rong Ma, Rhani Abd Elrahman, Mohamed Ismail, Mahmoud Zeid, Mahmoud Hassan, Mahmoud Sadek, Saad Ahsan, Imad Kariapper, Mohammad Radiyat, Humale Khan, Muhammad Sameed, Farhan Quasem, Fathi Abdelsalam, Haroon Ismail, Annas bin Adil, and many others. I thank in particular Erkan Özdoğan for teaching me Turkish, and Ülkü Kozluca, Muharrem Es, H. Yunus Taş, Kenan Dağcı, Tuncay Güloğlu, and Adem Birson for giving me the opportunity to practice and improve. I also thank David Owen for hosting us at his Near Eastern Studies discussion table and Michela Baraldi for hosting us at *Cene Italiane A Cornell* every week.

Finally, I would like to thank my parents and my family for all that they have done for me, for always being there for me, and for always encouraging and supporting me through thick and thin.

This work was supported by the National Science Foundation under Grant CCF-0642925 (CAREER), by the Air Force Office of Scientific Research under Grant FA9550-08-1-0060, and by the Army Research Office under Contract W911NF-10-1-0419.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Modern-day networks are constantly and rapidly growing in size. Not only an appreciably larger amount of data is being transferred over vast geographic distances, but users are also expecting improved QoS, leading to more stringent requirements on delay, error rates, and dropped packets. The sheer size of the networks and the amount of traffic has led to a push in building distributed architectures to increase efficiency and reduce cost. Distributed systems provide a number of advantages over centralized systems; they are, for instance, more scalable as the number of users grows and require only partial knowledge of the network. However, while centralized point-to-point communication has been relatively well-studied and its fundamental limits well-understood, we still lack an understanding of many fundamental problems in decentralized communication. For instance, how do we best allocate network resources in distributed/cloud storage systems? How do sensors in a distributed sensor network communicate efficiently while meeting power constraints? What are the communication requirements to meet performance and QoS guarantees in decentralized networks and how are they different from those in centralized networks? What security and privacy issues can arise in distributed systems and how do they affect communication? In this thesis, we focus on two such design issues, delay and security, and attempt to understand the information-theoretic limits of communication with regard to these issues.

1.1 Delay-reconstruction Tradeoffs in Content-sharing Networks

Typically, content is distributed from servers to clients via transmission of packets over a network. For the purposes of sharing content, e.g., a file, participants can act as both server and client by both uploading and downloading packets to and from other participants, as is the case in gossip communication or peer-to-peer architectures (e.g., [1, 2, 3]). One metric that is widely used to measure the performance of file sharing systems is the average amount of time a user must spend in the network, *i.e.*, the average time taken to download the whole file. There are two schools of thought about how to build such a system; one is to divide the file into a number of pieces which are then circulated among participants without any coding. Participants therefore acquire a partial copy of the file as soon as they download their first packet. Such a strategy is susceptible to the coupon collector problem; the initial few packets can be acquired rapidly, but it takes much longer to collect the final few packets [4] which significantly increases the overall download time per participant. The delay performance of BitTorrent, a prominent P2P architecture based on this school of thought, has been thoroughly analyzed [5]-[8].

A competing school of thought is to first encode file pieces using random linear network coding [9] or rateless fountain codes. P2P protocols based on fountain codes have been considered in [11, 18], and random linear network coding has been employed in P2P technology in [12]-[15]. The advent of fountain codes [16]-[19] has been one of the most important recent advances in coding for packet networks. Fountain codes operate by generating a virtually infinite number of encoded packets such that the original source can be reconstructed from any sufficiently large subset of these packets. Fountain codes are capacity achieving and universal for the class of binary erasure channels, and they have the additional advantage of being *rateless* in the sense that the number of packets to be produced at the encoder can be decided in real time. Fountain codes also obviate the need for feedback from the receiver to the transmitter, and they have low encoding and decoding complexity.

Fountain codes can eliminate the coupon collector problem [10, Sec. 2], but they suffer from poor intermediate performance. In the extreme case, it is not possible to reconstruct any portion of the original source until all of it can be reconstructed. In contrast, for feedback-based retransmission schemes for the erasure channel, each received packet reveals some of the original source. A user-perceived delay is therefore introduced with fountain codes; if, for instance, users are downloading a movie, then they must wait for all of the movie to be downloaded before they can begin watching it.

A fundamental question that arises is whether it is possible to mitigate this user-perceived delay via partial reconstruction of the source without increasing the overall transmission time. In particular, assuming that the code remains capacity achieving over erasure channels, how much of the source can be reconstructed from a given number of received encoded packets? We distinguish between two types of partial reconstruction: “in-order” reconstruction refers to sequential reconstruction in which earlier parts of the source are reconstructed before the latter parts. While many network applications require in-order reconstruction, “out-of-order” reconstruction may be sufficient for others. It is sufficient for files that are not organized linearly, for example, such as an unsorted database. Videos with out-of-order reconstruction could be played by interpolating over the missing portions. This playback would be at a lower quality, but it could still be useful, say for determining whether the downloaded file is the desired one.

Methods for improving the intermediate performance of fountain codes, assuming out-of-order reconstruction, have been investigated in [20]-[22] in a coding-theoretic setting. The source is encoded into a large number of packets such that any k suffice for reconstructing the source. Intermediate performance is then characterized by the fraction of the original source string that can be reconstructed when m encoded packets are received, where $0 < m < k$. An upper bound on this fraction is provided in [20], and lower bounds, based on the designing of suitable output degree distributions for various values of m, that perform close to the upper bound are provided in [20]-[22]. This enhanced intermediate performance comes at the cost of an increased overall transmission time, however, i.e., the codes are no longer capacity achieving. Moreover, as mentioned in [22], designing degree distributions to boost intermediate performance for a particular value of m exacerbates intermediate performance for other values of m.

In this work, we take a more information-theoretic approach and address the issue of optimal partial reconstruction *without* increasing the overall transmission time. We model the source as a bit string that is encoded into n packets. We impose the constraint that the receiver be able to reconstruct a fraction $1 - D$ of the source from any k packets, and we require that the sum rate of these packets equal the minimum rate for which this is possible. We then ask what fraction of the source block can be reconstructed from m packets, where $m < k$, allowing for out-of-order reconstruction. For this setup, we provide a coding scheme based on MDS codes that yields significant partial reconstruction while meeting the aforementioned constraints, and is provably Pareto optimal in m for $1 \leq m \leq k$ and any n and k, and absolutely optimal for certain values of m, n and k.

Our source coding problem can be viewed as a binary erasure version of the multiple descriptions (MD) problem [23]-[32]. Multiple description coding is a technique in which a source is encoded into several messages that are sent to the decoder, only a subset of which are assumed to reach their destination. The decoder uses them to reproduce the source, with the fidelity of the reproduction depending on which packets are received. The problem considered in this work amounts to an MD problem with distortion measured using the *erasure distortion measure* [33, p. 338]: the decoder's reproduction of the source may contain erasures but not errors, and the fraction of erasures in the reproduction is defined to be its "distortion" with respect to the original. In the terminology of multiple descriptions, our rate constraint is called a "no excess rate" condition [25].

It is worth noting that the erasure version of the MD problem has some unique virtues. The erasure distortion measure is universal in that it can be reasonably employed for a wide array of digital data sources. This sidesteps the difficult question of how to measure distortion for complicated, real-world data sources such as video. The binary erasure MD problem with no excess rate and no distortion for every k out of n messages is particularly relevant to peer-to-peer networks, since it can be used to study the tradeoff between the performance of fountain codes and a competing technology: BitTorrent [3]. For large n and small k, our MD problem mimics rateless fountain codes, since out of a large number of descriptions, only a relatively small number must be received (collected) in order to reconstruct the source with a specified distortion. For $k = n$, the MD problem resembles BitTorrent, where all of the relevant packets must be received to allow for complete reconstruction of the source. BitTorrent provides good intermediate performance but suffers from the "coupon collector" problem: the initial packets can be acquired quickly at the receiver, but it takes much longer to obtain the last few packets. By varying n and k in the binary erasure MD model, the middle ground between fountain codes and BitTorrent can be explored. Our results suggest that choosing n to be an integer multiple of k would provide some of the advantages of both technologies.

The erasure MD problem could also serve as a starting point for the design of practical codes for network rate distortion. In the theoretical development of modern channel codes such as LDPC, many of the code designs and performance characterizations were first established for the erasure channel [34]. Finally, the erasure MD problem yields results that are more positive in nature than those of other MD instances. In particular, for many sources, the no excess rate assumption necessarily yields poor intermediate performance (e.g., [24]): if a coding scheme is near-optimal for k receptions, it often yields high distortion for $m < k$ receptions. For the binary erasure MD problem, however, we shall see that it is possible to obtain good intermediate performance under no excess rate.

1.1.1 Results

We shall henceforth use the terms packets, messages, and descriptions interchangeably. We focus on binary erasure MD with no excess rate for every k out of n descriptions, *i.e.*, any subset consisting of k messages, has a total rate of $R(D_k)$, where D_k is the distortion constraint when k messages are received. We consider symmetric descriptions, *i.e.*, the rates of the n descriptions are the same and the distortion constraint depends only on the number of messages received. In fact, no excess rate implies symmetric descriptions for $k < n$: if every k out of n descriptions have sum rate $R(D_k)$, then each rate must be $R(D_k)/k$. We examine two distortion criteria; a *worst-case* distortion criterion, which measures the reconstruction fidelity by the maximum of the per-letter distortion over all source sequences, and an *average-case* distortion criterion, which measures the reconstruction fidelity by the average of the per-letter distortion over all source sequences. The average-case criterion is the standard criterion used in the literature. The worst-case criterion is less commonly used but it has the advantage of being universal in the sense that it is insensitive to the source distribution, which in practice is often unknown. Our main contributions are:

1. proposing, for all n and k, coding schemes for both worst-case and average-case distortion criteria and characterizing their achievable distortion region when $m \leq k$ descriptions are received at the decoder. The scheme for worst-case distortion is a zero-error coding scheme based on MDS (*maximum distance separable*) codes. The scheme for average-case distortion is based on random binning and can be viewed as a concatenation of $(n, 1)$ and (n, k) source-channel erasure codes [29].
2. providing, for both worst-case and average-case distortion criteria, a tight

lower bound on the distortion when a single message is received at the decoder. For worst-case distortion, the lower bound holds for all n and k. Moreover, we show that the MDS coding scheme is Pareto optimal in the achievable distortions $D_1, \dots, D_k$ for all n and k, and, for certain ranges of n and k, is also absolutely optimal when more than one message is received at the decoder. For average-case distortion, our lower bound holds, modulo a closure operation, for all n and k satisfying $\left(1 - \frac{1}{n}\right)^k \leq \frac{1}{2}$. In addition, for $n > 3$ and $k = 2$, we provide a lower bound on the optimal single-message distortion that differs by exactly $1/n$ from the distortion achieved by the random binning scheme. Our results for the special case in which there is no distortion for k messages (*i.e.*, any k messages allow the decoder to construct the original source sequence completely) have appeared in [35] (average-case distortion) and [36] (worst-case distortion).

3. proposing a coding scheme, based on the binary erasure MD coding schemes, for vector Gaussian MD and showing that it is optimal for scalar Gaussian MD with two levels of receivers and no excess rate for the central receiver. The scheme involves quantizing the vector Gaussian source according to a given quadratic distortion constraint and then transmitting the quantized version over the n channels according to the aforementioned binary erasure coding schemes. This demonstrates how the binary erasure coding schemes can be used as part of a more general, layered coding scheme for multiple descriptions with a generic source distribution and arbitrary distortion metric.

1.1.2 In-Order Reconstruction

Several prior works have considered the problem with in-order reconstruction. Albanese *et al.* [38] propose a coding method that involves assigning a priority level to messages and encoding them into packets. A message can be decoded from any subset of packets; however, the priority level of a message determines the minimum number of packets required to reconstruct that message. This amounts to in-order reconstruction, because messages with higher priority must be reconstructed before messages with lower priority. The in-order reconstruction problem can also be viewed as an instance of symmetric multilevel diversity coding (MLD) [39]. Comparing these results with those in this work shows that guaranteeing in-order reconstruction requires significantly higher rates. Walsh *et al.* [40] study the rate-delay tradeoff for in-order reconstruction in multi-path networks where time-ordered source packets arrive out of order at the destination. The channel between the transmitter and receiver is therefore different from the packet erasure channel considered here, since any packet sent by the transmitter eventually arrives at the receiver, albeit not in the order it was transmitted in. The authors introduce *delay mitigating codes* with the aim of minimizing delay at the receiver when the source bits are reconstructed in order from encoded packets arriving out of order.

1.2 Decentralized Encoding

We next focus on delay-reconstruction tradeoffs in P2P networks with decentralized encoding, *i.e.*, peers generate coded packets based on their own partial copies of the file. Within this context, we address the question posed in

Section 1.1: if we assume optimal decentralized encoding and that the packets might be received in any order, then how much of the file can be reconstructed from a given number of received packets?

Since the centralized version of the problem was addressed by posing it as a multiple description problem, it is natural to study the decentralized version by posing it as an instance of multiple descriptions with distributed encoding, which in the literature is actually called the *robust CEO problem* [47]. In the CEO problem, n encoders observe independently corrupted versions of a source and then transmit encoded messages, based on their partial knowledge of the source, to a decoder that attempts to reconstruct the source from the n messages to meet a distortion constraint. There is no communication among the encoders, as shown in Figure 3.1. In the *robust* variant of the problem, the encoders behave as in the CEO problem, but instead of using all n messages to reconstruct the source, the decoder must reconstruct it from *any* subset of the n messages subject to different distortion constraints for each subset.

We employ a particular instance of the robust CEO problem that we call the *binary erasure robust CEO problem.* In this instance, the source to be communicated is binary and *i.i.d.* uniform. The encoders observe this binary source passed through independent binary erasure channels. Thus, some of each encoder's file is missing, but none of it is incorrect. Moreover, when the decoder reconstructs the file, it is not permitted to introduce errors, although it is allowed to output an erasure for any source bit about which it is uncertain. The "distortion" is the fraction of erasures in its reconstruction. In turn, the decoder could then create new coded packets from its reconstruction and distribute them to other peers. Although we focus on the case in which the source is binary, we expect that most of the analysis will carry over to uniform sources over larger alphabets, which could be used to model audio samples, transform coefficients, video frames, or BitTorrent pieces.

The binary erasure robust CEO problem lends itself to a natural coding scheme in which individual encoders (peers) perform vector quantization of their observed partial source sequences using erasure test channels, followed by Slepian-Wolf binning. This is a particularization of the general scheme proposed in [47]. We first consider the case of symmetric peers and we show, using very different techniques from those used in the centralized case [36], that this coding scheme achieves a delay-reconstruction tradeoff that is Pareto optimal over a range of received messages. The same problem for Gaussian sources and quadratic distortion measure has been considered in [48] and an achievable information-theoretic rate region has been derived. Optimality results for the symmetric case of the Gaussian problem have been presented in [49].

In the process of proving our result, we also establish a new outer bound for the general multi-terminal source coding problem that improves upon the outer bound of Wagner and Anantharam [44]. We further show that if we relax the symmetry assumptions about the encoders, then the coding scheme is no longer optimal, even for a simple setup with two encoders.

1.3 Lossy Source Coding with Byzantine Adversaries

While the rapid growth of modern-day communication networks makes them increasingly useful, it also makes them increasingly difficult to protect against attacks. This is especially true of those networks, such as peer-to-peer systems, in which the nodes are controlled by different entities. In the case of peer-to-peer networks, malicious users could sabotage the file-sharing process by intentionally transmitting a corrupted version of the file. Similar problems can potentially arise in ad-hoc networks and distributed storage systems.

There has been considerable work on how to protect transmitted information against malicious users within the context of channel- and network-coding, and a number of significant results are available. Yeung and Cai [53] show that if z unit-capacity edges in an acyclic multicast network are subject to random or adversarial errors, then the network capacity is $C - 2z$, where C is the network capacity when all edges are error-free. Thus if an adversary controls z edges, it effectively removes $2z$ edges from the original adversary-free network (see also [54]-[59] and the references therein). This is reminiscent of the Singleton bound, and we refer to it as the "factor-of-2" rule. The factor-of-2 rule was also shown to hold for lossless source coding: it is well known that if a source X is to be losslessly communicated via n packets, then the sum rate of those packets must be at least $H(X)$. Kosut and Tong [60] have shown that if t of the n packets can be altered in arbitrary ways by adversaries, then every $n-2t$ packets must have sum rate at least $H(X)$. Thus t traitors effectively remove $2t$ packets from the original adversary-free problem, *i.e.,* the factor-of-2 rule obtains.

In the context of peer-to-peer systems, often the ultimate goal is to communicate a file approximately rather than reliably. Codes and fundamental limits for this problem are less well understood (but see [61]-[62]). One natural approach to this problem is to perform separate compression and adversarial error-protection. That is, one combines rate-distortion-optimal lossy compression with network codes that are optimal for the adversarial model at hand.

We show that this approach is optimal in some cases but suboptimal in general, even for networks with one sender, one receiver, and no intermediate nodes. Specifically, we consider the problem in which a source is compressed to form n packets, any t of which can be altered in an arbitrary way. The decoder receives the n packets and, without knowing which packets were altered, must estimate the source to meet a given distortion constraint. We show that separate compression and adversarial error correction achieve rate-distortion performance governed by the factor-of-2 rule, and that this is optimal for binary sources with the Hamming distortion measure and Gaussian sources with the mean square error distortion measure. These two optimality results hinge on a combinatorial result of Kleitman [66] on the maximum size of subsets of Hamming space with a given diameter, and the Brunn-Minkowski inequality, respectively.

We then show by means of a counterexample, involving a binary source with erasure distortion, that separation is not optimal in general. We consider a 3-encoder problem with one traitor such that one encoder has rate $R < 1$, while the other two have rate 1 and can therefore transmit the source sequence exactly. We determine the optimal distortion for this problem as a function of R and show that separation cannot achieve it. We note that while source-channel separation has long been known known to fail in many scenarios (e.g., [63, 64, 65]), the reason that it fails here seems to be fundamentally different from the standard examples.

1.4 Multi-level Lossless Source Coding with Byzantine Adversaries

In Section 1.3, we looked at instances of adversarial lossy source coding with the assumption that the number of adversaries is fixed and known to the honest entities in the network. In real-world networks, this assumption does not hold; in a P2P system where peers are constantly entering and leaving the network, it is impossible for a user to know how many malicious peers are present in the system at any given time. In such a situation, a peer that is receiving packets from other peers must be able to detect which of the received packets is corrupted without knowing a priori how many are corrupted, discard the ones that are corrupted, and use only the uncorrupted ones to reconstruct the file.

In order to study such networks, the assumption that the number of adversaries in the system is fixed and known to the honest entities must be relaxed. Natural questions that then arise are: what strategies can the decoder use to detect corrupted packets without knowing a priori how many are corrupted? How much compression can the encoders perform? How does the performance of the decoder vary as the number of adversaries changes? What performance guarantees can be provided when a large number of adversaries is present in the network as compared to when a smaller number is present? In this regard, the ″adversarial multi-level diversity problem″ is a useful model to study. In the adversarial multi-level diversity problem, a number of independent information sources are compressed by encoders and transmitted to the decoder. Some of the encoders are adversaries, and the number and identity of these adversaries is not known to the honest encoders and the decoder. Based on the messages received from the encoders, the decoder attempts to detect potentially corrupted messages and then completely reproduce as many of the sources as it can. For instance, if there are n sources, it attempts to reconstruct all n sources if it does not detect a corrupted message, $n-1$ sources if it detects one corrupted message, and so on. The goal is to guarantee the reconstruction of a certain number of sources when a certain number of packets are corrupted.

We look at a simple example with two sources X and Y and three encoders. The three encoders are either all honest, or at most one of them is a traitor. In the latter case, the identity of the traitor (and that one is in fact present in the system) is unknown to the honest encoders and the decoder. The decoder receives all three messages and losslessly reconstructs both X and Y if it does not detect a corrupted message. If however, it does detect a corrupted message, it reconstruct only X and outputs a flag sequence instead of Y indicating that it has detected a corrupted message.

We show that the rate region of this problem is the same as the rate region of the 3-encoder symmetric multi-level diversity (MLD) problem [68]. In the 3-encoder symmetric MLD problem, three independent sources X, Y, and Z are encoded and transmitted by three encoders. There are no traitors, but there is packet loss in the network, and the decoder may not receive all three messages. Depending on how many messages it receives, the decoder losslessly reconstructs a subset of the sources. If it receives one message, then it lossless reconstructs X alone. If it receives two messages, then it losslessly reconstructs X and Y, and if it receives all three messages, then it losslessly reconstructs all three sources. The rate region for the general n-encoder MLD problem was established in [68], and it was shown that a coding strategy based on superposition coding is optimal: X, Y, and Z can be encoded separately at each encoder, and the resulting codewords then concatenated to form the final message.

The connection between the adversarial multi-level diversity problem and the MLD problem stems from the Singleton bound, which provides the link between adversarial packets and packet erasures. As stated in Section 1.3, a factor-of-2 rule obtains when adversarial error correction is required: t traitors effectively remove $2t$ honest packets from the system (*i.e.,* t corrupted messages have the same effect as $2t$ packet erasures). When adversarial error detection is required, however, a "factor-of-1" rule is in effect: t traitors effectively remove t packets from the system (*i.e.,* t corrupted messages have the same effect as t packet erasures)[1]. In the adversarial multi-level diversity problem, the decoder is performing adversarial error correction for X and adversarial error detection for Y. Therefore, if the decoder receives three messages, one of which is corrupted, then the decoder should be able to losslessly reconstruct X from every message (factor-of-2 rule), and losslessly reconstruct Y from every pair of messages (factor-of-1 rule). This is equivalent to the symmetric 3-encoder MLD problem in which the decoder must losslessly reconstruct X from any one message, Y from every pair of messages, and Z from all three messages. Since there are only two sources in the adversarial multi-level diversity problem, we can take Z to be a zero-entropy source, *i.e.,* a deterministic sequence, which is known to the encoders and the decoder and does not need to be transmitted. We show in Chapter 5 that the two problems have the same rate region, and hence superposition coding is optimal for adversarial multi-level diversity coding.

[1] The distinction between the factor-of-1 rule and factor-of-2 rule has an analog in coding theory: a binary code with minimum Hamming distance k can detect up to $k-1$ errors, but can correct up to $\lfloor \frac{k-1}{2} \rfloor$ errors.

1.5 Organization of the Thesis

Chapter 2: We study delay-reconstruction tradeoffs in P2P networks. We formulate the n-channel binary erasure MD problem in Section 2.1. Sections 2.2 and 2.3 are devoted to our results for worst-case distortion and average-case distortion, respectively. In Section 2.4, we describe the layered architecture for MD and present our results for vector Gaussian multiple descriptions.

Chapter 3: We study delay-reconstruction tradeoffs in P2P networks with decentralized encoding. In Section 3.1, we formulate the binary erasure robust CEO problem more precisely and describe our coding scheme. In Section 3.2 we consider the symmetric version of the binary erasure robust CEO problem and show that the above coding scheme provides a Pareto optimal delay-reconstruction tradeoff. In Section 3.3, we consider an asymmetric, two encoder version of the problem and show that the coding scheme is not optimal.

Chapter 4: We formulate the lossy source coding problem with Byzantine adversaries. In Section 4.2, we present the separation-based coding scheme for general sources and arbitrary distortion measures and show that it achieves the factor-of-2 rule. In Sections 4.3 and 4.4, respectively, we prove that our scheme is optimal for uniform binary sources with Hamming distortion and Gaussian sources with squared error distortion. In Section 4.5, we show that the factor-of-2 rule is pessimistic for binary sources and erasure distortion.

Chapter 5: We study lossless source coding with multiple sources and an unknown number of adversaries and show its equivalence to the symmetric MLD problem.

CHAPTER 2

ERASURE MULTIPLE DESCRIPTIONS

In this chapter, we study delay-reconstruction tradeoffs in P2P networks. We formulate the n-channel binary erasure MD problem, and state and prove our results for worst-case distortion and average-case distortion, respectively. We subsequently propose a layered architecture for MD and present our results for vector Gaussian multiple descriptions.

2.1 Notation and Problem Formulation

We use uppercase letters to denote random variables, bold letters to represent vectors and script letters to denote their ranges. Realizations of random variables are denoted by lowercase letters, and realizations of random vectors are denoted by bold lowercase letters. A superscript appearing with a vector (*e.g.*, $\mathbf{X}^l$) indicates the length of the vector. Matrices are also represented in boldface. Let $\{X_t\}_{t=1}^{\infty}$ be a memoryless uniform binary source, with the random variables X_t taking values in the alphabet $\mathcal{X} = \{+,-\}$. Let $\hat{\mathcal{X}}$ be the reconstruction space $\{+,-,0\}$, where 0 denotes the erasure symbol, with an associated distortion measure $d : \mathcal{X} \times \hat{\mathcal{X}} \to \{0,1,\infty\}$ such that

$$d(x,\hat{x}) = \begin{cases} 0 & \text{if } \hat{x} = x \\ 1 & \text{if } \hat{x} = 0 \\ \infty & \text{otherwise.} \end{cases}$$

The above per-letter measure is known as the erasure distortion measure [33, p. 338]. An *encoder* is a function $f_i^{(l)} : \mathcal{X}^l \to \{1,\ldots,M_i^{(l)}\}$. A *decoder* is a function $g_{\mathcal{K}}^{(l)} : \prod_{k\in\mathcal{K}}\{1,\ldots,M_k^{(l)}\} \to \hat{\mathcal{X}}^l$, where $\mathcal{K}$ is the set of descriptions received.

Let $\mathcal{N} = \{1, \ldots, n\}$. The n-channel multiple descriptions problem, illustrated in Figure 2.1, can be formulated as follows. There are n encoders. Encoder $f_i^{(l)},\ i \in \mathcal{N}$, encodes and transmits a description of a length-l source sequence $\mathbf{X}^l$ over channel i. The receiver either receives this description without errors or it does not receive it at all. Excluding the case where none of the descriptions is received, the receiver may receive $2^n - 1$ different combinations of the n descriptions. Thus it can be represented by the $2^n - 1$ decoding functions $g_{\mathcal{K}}^{(l)},\ \mathcal{K} \subseteq \mathcal{N}$, $\mathcal{K} \neq \emptyset$. Based on the set of descriptions received, the receiver employs the corresponding decoding function to output a reconstruction of the original source string subject to a distortion constraint. We consider symmetric descriptions, *i.e.*, each description has the same rate and the distortion constraint depends only on the number of descriptions received.

We measure the fidelity of the reconstruction using two distortion criteria: a *worst-case* distortion criterion, under which distortion is measured by taking the *maximum* of the per-letter distortion over all source sequences, and an *average-case* distortion criterion, under which distortion is measured by taking the average of the per-letter distortion over all source sequences. We define *achievability* for the two criteria as follows. Let $\hat{\mathbf{X}}_{\mathcal{K}}^l = g_{\mathcal{K}}^{(l)}(\{f_k^{(l)}(\mathbf{X}^l) : k \in \mathcal{K}\})$ be the reconstruction sequence corresponding to the source sequence $\mathbf{X}^l$.

Definition 1 (Worst-case distortion). *The rate-distortion vector $(R, D_1, \ldots, D_n)$ is* achievable *if for some l there exist encoders $f_i^{(l)}$, $i \in \mathcal{N}$ and decoders $g_{\mathcal{K}}^{(l)}$, $\mathcal{K} \subseteq \mathcal{N}$, $\mathcal{K} \neq \emptyset$, such that*

$$R \geq \frac{1}{l} \log M_i^{(l)} \text{ for all } i \text{, and}$$

$$D_k \geq \max_{\mathcal{K}:|\mathcal{K}|=k} \max_{\mathbf{x}^l \in \mathcal{X}^l} \left[\frac{1}{l} \sum_{t=1}^{l} d(x_t, \hat{x}_{\mathcal{K},t}) \right].$$

We use $\mathcal{RD}_{worst}$ to denote the set of achievable rate-distortion vectors.

Definition 2 (Average-case distortion). *The rate-distortion vector* $(R, D_1, \ldots, D_n)$ *is* achievable *if for some* l *there exist encoders* $f_i^{(l)}$, $i \in \mathcal{N}$ *and decoders* $g_{\mathcal{K}}^{(l)}$, $\mathcal{K} \subseteq \mathcal{N}$, $\mathcal{K} \neq \emptyset$, *such that*[1]

$$R \geq \frac{1}{l} \log M_i^{(l)} \text{ for all } i \text{, and}$$

$$D_k \geq \max_{\mathcal{K}:|\mathcal{K}|=k} \boldsymbol{E} \left[\frac{1}{l} \sum_{t=1}^{l} d(X_t, \hat{X}_{\mathcal{K},t}) \right].$$

We use $\mathcal{RD}_{avg}$ to denote the set of achievable rate-distortion vectors and $\overline{\mathcal{RD}}_{avg}$ to denote its closure. We describe our results for worst-case distortion in the next section and for average-case distortion in Section 2.3. For both distortion criteria, we consider the case where there is *no excess rate* for every k out of n descriptions, *i.e.*, $kR = R(D_k) = 1 - D_k$, where $R(\cdot)$ is the Shannon rate-distortion function. Thus $R = (1 - D_k)/k$. We shall henceforth use $R_k(D_k)$ to denote $(1 - D_k)/k$. Our goal is to characterize the achievable distortions $D_1, \ldots, D_n$ for both distortion criteria.

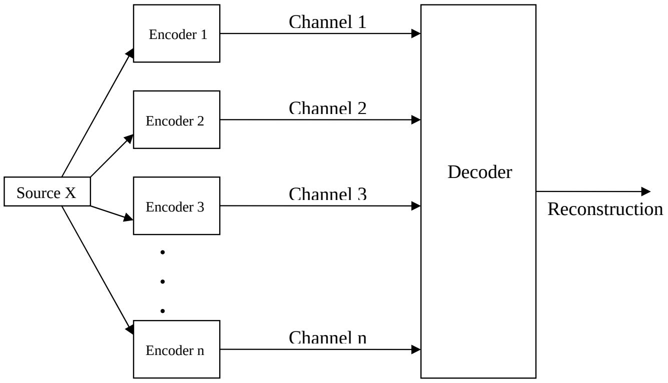

Figure 2.1: The n-channel multiple descriptions problem

It should be pointed out that the $k = n$ case is particularly simple. Let $D_i,\ i \in \mathcal{N}$ be the distortion constraint when the receiver receives i messages. No

[1] All logarithms and exponentiations have base 2 unless explicitly stated.

excess rate for n descriptions dictates that the sum-rate of the n messages is exactly $(1 - D_n)$, which in turn implies that the rate of each message is $(1 - D_n)/n$. The problem then reduces to characterizing the optimal $D_1, \ldots, D_n$. Consider a coding scheme that takes a source string of length l and erases the last lD_n bits. The remaining $l(1 - D_n)$ bits are divided into n disjoint parts, each consisting of $l(1 - D_n)/n$ bits. Encoder i transmits the $l(1 - D_n)/n$ bits in the i^{th} part to the decoder over the i^{th} channel, with erasures in place of the remaining $l - l(1 - D_n)/n$ bits. Thus upon reception of any k descriptions, the decoder can reconstruct $kl(1 - D_n)/n$ bits of the original source string. Clearly, this scheme achieves $D_k = 1 - k(1 - D_n)/n$ under both the worst-case and average-case distortion criteria. Moreover, for any code that achieves the rate-distortion vector $(1 - D_n/n, D_1, \ldots, D_n)$, every description has rate $(1 - D_n)/n$ and therefore the point-to-point rate distortion function dictates that any set of k message can reveal no more than a fraction $k(1 - D_n)/n$ bits of the original source string. Thus

$$\max_{\mathcal{K}:\mathcal{K}=k} \max_{\mathbf{x}^l \in \mathcal{X}^l} \left[\frac{1}{l} \sum_{t=1}^{l} d(x_t, \hat{x}_{\mathcal{K},t}) \right] \geq 1 - \frac{k(1 - D_n)}{n},$$

and

$$\max_{\mathcal{K}:\mathcal{K}=k} \mathbf{E} \left[\frac{1}{l} \sum_{t=1}^{l} d(X_t, \hat{X}_{\mathcal{K},t}) \right] \geq 1 - \frac{k(1 - D_n)}{n}.$$

Thus the aforementioned coding scheme achieves the optimal $D_1, \ldots, D_n$ under both the worst-case and average-case distortion criteria.

We use the insight obtained from the $k = n$ case to construct codes for the more complicated case in which $k < n$. No excess rate for a particular set of k descriptions requires that the information transmitted over those k channels be independent. Since we impose no excess rate for every size-k subset of descriptions, the information transmitted over any k channels must be mutually independent. The coding scheme for $k = n$ ensures that this condition is met by dividing an erased version of the source string into n disjoint (and therefore independent) parts and transmitting them uncoded over the n channels. This strategy of sending independent uncoded bits works as long as the bits transmitted over each channel are disjoint. In particular, if $R_k(D_k) = (1 - D_k)/k \leq 1/n$ (equivalently, $D_k \geq 1 - k/n$), the source string can always be divided into n disjoint, equal parts, each containing a fraction $R_k(D_k)$ of the total number of bits. If $D_k < 1 - k/n$, however, then $R_k(D_k) > 1/n$ and it is not possible to divide the source string into n disjoint parts each containing a fraction $R_k(D_k)$ of the total number of bits, since each part must then contain more than $1/n$ of the total number of bits. Transmitting uncoded bits, therefore, will be optimal for a rate up to $1/n$ only; in order to achieve a rate larger than $1/n$, additional information about the source must be transmitted along with each description, and this information must be mutually independent for every set of k descriptions. Random binning schemes can be designed in order to convey independent information about the source to the decoder such that any k messages reveal the source string to a specified distortion. Such schemes, however, suffer from the "cliff effect"; nothing can be reconstructed from fewer than k messages, and once k messages have been received, additional messages provide no reduction in distortion at all.

By using a hybrid of these two approaches, *i.e.*, transmission of uncoded bits and random binning, we can achieve an incremental reduction in distortion with each additional message while still satisfying the necessary independence conditions. With fewer than k messages, the decoder can partially reconstruct the source string using the uncoded bits alone. With k or more messages, the decoder can use the random binning component to decode the source string to a specified distortion D_k, and can then use the uncoded bits in the messages to further reduce distortion. The resulting distortion therefore decreases linearly with the number of messages received, with a sudden downward jump at k when additional information about the source can be decoded from the binning. Figure 2.2 depicts how the achievable distortion varies with the number of descriptions received at the decoder when $D_k = 0$. We provide outer bounds on the distortion region which show that such a hybrid scheme is optimal in a number of scenarios.

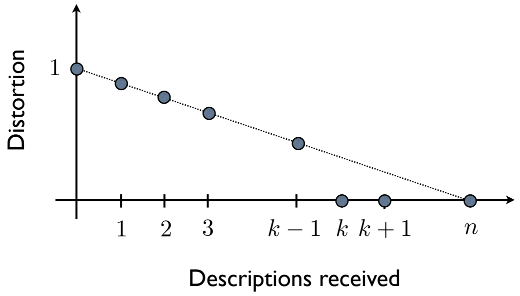

Figure 2.2: The achievable distortion region for $D_k = 0$. The achievable distortion decreases linearly with the number of descriptions received up to $k-1$ descriptions, and drops abruptly to zero upon reception of k or more descriptions.

The threshold $D_k = 1 - k/n$ plays an important role in our coding scheme. If $D_k \geq 1 - k/n$, then transmission of independent uncoded bits over the n channels as described above is sufficient. If $D_k < 1 - k/n$, then in addition to sending uncoded bits, we also send coded information. For the worst-case distortion measure, we describe this scheme in detail in Section 2.2.1, using MDS codes to realize the coding. Achievability for average-case distortion follows from the achievability result for worst-case; however, an alternative proof is included in Appendix A.8 that does not rely on MDS arguments by using random binning instead. The optimality results for the two distortion criteria are different, with our results for worst-case distortion being the stronger of the two.

2.2 The Worst-case Distortion Criterion

We begin by presenting a zero-error coding scheme based on systematic MDS codes that works for finite blocklengths. The scheme consists of two parts—uncoded bits and an MDS-code component. With each message, the encoder sends uncoded bits along with an encoded version of the source, using an (n, k) systematic MDS code for the encoding. The decoder outputs the bits revealed by the systematic part of the MDS code as the source reconstruction if less than k descriptions are received. If k or more descriptions are received, the decoder uses the uncoded bits and the bits revealed by the systematic part of the MDS code to decode the encoded erased version by applying an MDS decoding algorithm. The following subsection discusses the achievable distortion region of the MDS coding scheme.

2.2.1 An Achievability Result

Definition 3. *Given n, $k \leq n$, and $D_k \in [0, 1]$, define*

$$\begin{aligned}\tilde{\boldsymbol{R}} = (&R_k(D_k), 1 - R_k(D_k), \ldots, 1 - (k-1)R_k(D_k), D_k,\\ &D_k - R_k(D_k), \ldots, D_k - (n-k)R_k(D_k)), \text{ and}\\ \hat{\boldsymbol{R}} = \Big(&R_k(D_k), 1 - \frac{1}{n}, \ldots, 1 - \frac{k-1}{n}, D_k, \left(\frac{n-k-1}{n-k}\right) D_k,\\ &\left(\frac{n-k-2}{n-k}\right) D_k, \ldots, \left(\frac{1}{n-k}\right) D_k, 0\Big).\end{aligned}$$

Theorem 1. *Let D_k be a rational number in the interval $[0, 1]$. For any n and $k \leq n$, if $D_k \geq 1 - \frac{k}{n}$, then $\tilde{\boldsymbol{R}} \in \mathcal{RD}_{worst}$. If $D_k < 1 - \frac{k}{n}$, then $\hat{\boldsymbol{R}} \in \mathcal{RD}_{worst}$.*

Proof. **Case I**: $D_k \geq 1 - \frac{k}{n},\ D_k$ rational

Since D_k is rational, there exists a positive integer l' such that $l'R_k(D_k)$ is a positive integer. Choose a blocklength $l = \alpha n l'$, where α is any positive integer. Observe a length-l source sequence $\mathbf{X}^l$, and divide $\mathbf{X}^l$ into n disjoint parts such that each part contains $l/n = \alpha l'$ bits. (The division is the same regardless of the source realization.) Label the parts $\mathbf{X}_i,\ i \in \mathcal{N}$. Choose $lR_k(D_k)$ bits from each of the n parts (since $D_k \geq 1 - \frac{k}{n}$, $lR_k(D_k) \leq \frac{l}{n}$ and therefore $lR_k(D_k)$ bits can be chosen from each part). Denote by $\mathbf{Y}_i$ the set of $lR_k(D_k)$ bits chosen from $\mathbf{X}_i$. Transmit $\mathbf{Y}_i$ uncoded over the i^{th} channel.

The decoding is trivial. If m descriptions, say $(\mathbf{Y}_1, \ldots, \mathbf{Y}_m)$, are received, output $\hat{\mathbf{X}}^l_{\mathbf{m}}$ as the reconstruction of $\mathbf{X}^l$, where $\hat{\mathbf{X}}^l_{\mathbf{m}}$ is such that the $mlR_k(D_k)$ bits corresponding to $(\mathbf{Y}_1, \ldots, \mathbf{Y}_m)$ are non-erased and the other $(l - mlR_k(D_k))$ bits are erasures. Since the reconstruction sequence has $l - mlR_k(D_k)$ erasures regardless of the source sequence, the worst-case distortion D_m is $(l - mlR_k(D_k))/l = 1 - mR_k(D_k)$. When k descriptions are received, the worst-case distortion is $1 - kR_k(D_k) = D_k$. Thus $\tilde{\mathbf{R}} \in \mathcal{RD}_{worst}$.

Case II: $D_k < 1 - \frac{k}{n},\ D_k$ rational

For this case, we present an achievability scheme based on MDS (*maximum distance separable*) codes[2]. Let m be the smallest integer such that $2^m \geq n$ and $\frac{mnk(n-k)}{n(1-D_k)-k}$ is an integer (such an m exists because D_k is rational). Define $q = 2^m$, and construct a q-ary MDS code of length $q-1$ and dimension k. By repeatedly puncturing this $(q-1, k)$ MDS code, we obtain a punctured MDS code of size (n, k) [42, p. 190]. The punctured coordinates are revealed to the decoder. Let $\mathbf{G}_1$ be the generator matrix of the punctured (n, k) MDS code, and assume without

[2] A (n, k) MDS code is a linear code that satisfies the Singleton bound, *i.e.*, the Hamming distance between any two codewords is at least $n - k + 1$. Reed-Solomon codes, for instance, are MDS codes.

loss of generality that $\mathbf{G}_1$ is systematic, *i.e.*, $\mathbf{G}_1$ is of the form $[\mathbf{I}_k|\mathbf{A}]$, where $\mathbf{I}_k$ is the $k \times k$ identity matrix and $\mathbf{A}$ is a $k \times n - k$ matrix over the finite field GF(q). Construct matrices $\mathbf{G}_2, \ldots, \mathbf{G}_n$ by shifting the columns of $\mathbf{G}_1$ to the right, *i.e.*, $\mathbf{G}_i$ is the matrix formed by shifting the columns of $\mathbf{G}_1$ by $i-1$ places, with the last $i-1$ columns of $\mathbf{G}_1$ wrapping around. In particular, if $\mathbf{G}_1 = [\mathbf{I}_k|a_1 \ldots a_n]$, where $a_1, \ldots, a_n$ are the columns of $\mathbf{A}$, then $\mathbf{G}_i = [a_{n-i+2} \ldots a_n|\mathbf{I}_k|a_1 \ldots a_{n-i+1}]$.

Encoding: The encoding procedure is illustrated in Figure 2.3. Let $\mathbf{X}^l$ be the observed source string, of length $l = \frac{mnk(n-k)}{n(1-D_k)-k}$ bits. Divide $\mathbf{X}^l$ into n disjoint parts, each of length $\frac{mk(n-k)}{n(1-D_k)-k}$ bits. (The division is done the same way regardless of the source realization.) Let $\mathbf{X}_i,\ i \in \mathcal{N}$ denote the last $\frac{lD_k}{n-k}$ bits of the i^{th} part. Construct an erased version $\mathbf{X_e}^l$ by replacing the last $\frac{lD_k}{n-k}$ bits in each of the n parts by erasures. Thus $\mathbf{X}_e^l$ has $l(1-\frac{nD_k}{n-k}) = mnk$ bits. Each of the n parts of $\mathbf{X}_e^l$ has mk bits and can therefore be treated as a concatenation of k binary strings of length m, such that each of these binary strings is the binary representation of an element in GF(q). Thus each of the n parts of $\mathbf{X}_e^l$ can be mapped to a vector of length k in GF(q). Label these vectors $\mathbf{Z}_j,\ j \in \mathcal{N}$. Let $\mathbf{Y}_j = \mathbf{Z}_j\mathbf{G}_j,\ j \in \mathcal{N}$. Thus the $\mathbf{Y}_j$ are length-n vectors in GF(q). Let $Y_{ji} = \mathbf{Z}_j g_{ji}$ denote the i^{th} element of $\mathbf{Y}_j$ (here g_{ji} is the i^{th} column of $\mathbf{G}_j$). Transmit ($\mathbf{X}_i, Y_{ji} : j \in \mathcal{N}$) over the i^{th} channel.

Decoding: Suppose $c < k$ descriptions are received at the decoder. Let $\mathcal{M} \subset \mathcal{N}$ denote the set of indices of the received descriptions. Assume without loss of generality that $i \in \mathcal{M}$. Thus the decoder receives $\mathbf{X}_i$ and $Y_{ji} = \mathbf{Z}_j g_{ji} : j \in \mathcal{N}$. Thus $\frac{lD_k}{n-k}$ bits are revealed to the decoder via $\mathbf{X}_i$. Now for a fixed i, exactly k of the $\mathbf{G}_j,\ j \in \mathcal{N}$, (in particular, $\mathbf{G}_{i-k+1}, \ldots, \mathbf{G}_i$) will have their i^{th} column in the systematic part. Thus one symbol from k of the $\mathbf{Z}_j,\ j \in \mathcal{N}$, can be decoded. By mapping these decoded symbols to their binary representations, the

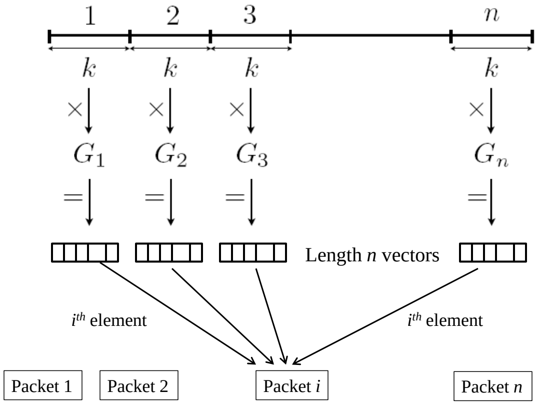

Figure 2.3: The MDS encoding procedure.

decoder can obtain a partial reconstruction of $\mathbf{X}^l$. Let $\hat{\mathbf{X}}_i$ represent the reconstructed source bits due to the i^{th} description. Output $(\hat{\mathbf{X}}_i : i \in \mathcal{M})$ as the reconstruction of $\mathbf{X}^l$. If $m > k$ descriptions are received, then any k descriptions reveal k symbols from each of the $\mathbf{Y}_j,\ j \in \mathcal{N}$. Also, since the punctured coordinates are known to the decoder, it can construct a longer codeword from every partially received codeword by adding erasures in place of the punctured coordinates. The longer codewords can be treated as codewords from the original $(q-1,k)$ MDS code. The original MDS code can subsequently be decoded by applying an erasure decoding algorithm [42, Ch. 9] and all the $\mathbf{Z}_j$ vectors can be recovered. Mapping the $\mathbf{Z}_j$ vectors to their binary representations reveals the erased version $\mathbf{X}_e^l$ of the original source string $\mathbf{X}^l$. Output $\{(\mathbf{X}_1,\ldots,\mathbf{X}_m)\} \cup \{\mathbf{X}_e^l \backslash (\mathbf{X}_1,\ldots,\mathbf{X}_m)\}$ as the reconstruction of $\mathbf{X}^l$.

Analysis: We now argue that the above scheme achieves the rate-distortion vector $(R_k(D_k), 1-\frac{1}{n}, 1-\frac{2}{n}, \ldots, 1-\frac{k-1}{n}, D_k, (\frac{n-k-1}{n-k})D_k, (\frac{n-k-2}{n-k})D_k, \ldots, (\frac{1}{n-k})D_k, 0)$. For any source string $\mathbf{X}^l$, every description (say the i^{th} description) consists of $(\mathbf{X}_i, Y_{ji} : j \in \mathcal{N})$. $\mathbf{X}_i$ consists of $lD_k/(n-k)$ bits. Now since Y_{ji} is an element of GF(q), it can be represented by m bits. Thus $(Y_{ji} : j \in \mathcal{N})$ is a length-n vector in GF(q), and can be represented by mn bits. Every description therefore consists of $mn + lD_k/(n-k)$ bits. Since the source string consists of $l = mnk(n-k)/(n(1-D_k)-k)$ source symbols, every description has rate

$$\frac{mn + lD_k/(n-k)}{l} = \frac{1-D_k}{k} = R_k(D_k).$$

Moreover, every description received at the decoder reveals $lD_k/(n-k)$ bits via $\mathbf{X}_i$, and exactly one symbol from k of the $\mathbf{Z}_j,\ j \in \mathcal{N}$. Each of these k symbols is an element of GF(q) and can be represented by m bits. Thus every description reveals $lD_k/(n-k) + mk$ bits to the decoder. (We note that the bits revealed by any two descriptions are disjoint. The uncoded bits X_a and X_b are disjoint by definition for any two descriptions a and b. Now suppose descriptions a and b revealed the same symbol from some $\mathbf{Z}_j$. Then $Y_{ja} = \mathbf{Z}_j g_{ja} = \mathbf{Z}_j g_{jb} = Y_{jb}$, which implies $a = b$.) Thus if $c < k$ descriptions are received, the decoder can reconstruct $c(lD_k/(n-k) + mk)$ bits of the original source sequence. Thus

$$\begin{aligned}
D_c &= 1 - \frac{c(\frac{lD_k}{n-k} + mk)}{l} \\
&= 1 - \frac{cD_k}{n-k} - \frac{cn(1-D_k) - ck}{n(n-k)} \\
&= 1 - \frac{c}{n}.
\end{aligned}$$

If $c \geq k$ descriptions are received, say descriptions $1, \ldots, m$, then $(\mathbf{X}_1, \ldots, \mathbf{X}_m)$ reveal $clD_k/(n-k)$ bits. Moreover, the erased version of the source sequence, $\mathbf{X}_e^l$, can be reconstructed by applying the MDS erasure decoding algorithm. The bits revealed by $(\mathbf{X}_1, \ldots, \mathbf{X}_m)$ are disjoint from the bits revealed by $\mathbf{X}_e^l$. The total number of bits revealed, therefore, is $clD_k/(n-k) + mnk$. Thus

$$D_c = 1 - \frac{c\frac{lD_k}{n-k} + mnk}{l}$$

$$
\begin{aligned}
&= 1 - \frac{cD_k}{n-k} - \frac{n(1-D_k)-k}{n-k} \\
&= \left(\frac{n-c}{n-k}\right) D_k.
\end{aligned}
$$

Thus $\hat{\mathbf{R}} \in \mathcal{RD}_{worst}$. □

2.2.2 Optimality Results

In this section we present optimality results for the MDS coding scheme described in the previous subsection. We first establish some preliminary results in Appendix A.1 which will be used in the proofs of the following theorems. The optimality results presented here are stronger than those for average-case distortion (Section 2.3.2) and yield a more complete characterization of the achievable distortion region. Since we are dealing with worst-case distortion constraints, the following results hold for any source distribution.

Theorem 2. *For any n and k, if $D_k \geq 1 - \frac{k}{n}$ and rational*[3]*, then $\forall\ (R_k(D_k), D_1, \ldots, D_k, \ldots, D_n) \in \mathcal{RD}_{worst}$, $D_m \geq 1 - mR_k(D_k)$ for all $m \in \mathcal{N}$.*

Proof. Let $D_k \geq 1 - \frac{k}{n}$. If a code achieves a certain distortion under worst-case distortion, then it will achieve that distortion under average-case distortion as well. The result therefore follows from the first part of Theorem 7. □

Definition 4. *Let $\mathbf{X}^l$ be a vector taking values in $\mathcal{X}^l$. An erased version of $\mathbf{X}^l$ is a vector $\tilde{\mathbf{X}}^l(X)$ (where $\tilde{\mathbf{X}}^l(\cdot)$ is a function of the X string), taking values in $\hat{\mathcal{X}}^l$, such that $\nexists\ t \in \{1, \ldots, l\}$ such that $\tilde{X}_t(X) = +$ and $X_t = -$ or $\tilde{X}_t(X) = -$ and $X_t = +$.*

[3] For this theorem and subsequent theorems in this subsection, we consider rational values for D_k since any code over a finite blocklength can yield rational distortions only.

The following lemma is integral to the proofs of our optimality results for worst-case distortion. Intuitively, the lemma says that for any code that encodes length-l source sequences into n pairwise independent messages, there exists a source sequence for which each of the l bits can be revealed by at most one of the n messages.

Lemma 1. *Let $\tilde{\mathbf{X}}_1^l(X), \tilde{\mathbf{X}}_2^l(X), \ldots, \tilde{\mathbf{X}}_n^l(X)$ be erased versions of the source string $\mathbf{X}^l \in \mathcal{X}^l$. Suppose $\mathbf{X}^l$ is* i.i.d. *uniform over $\mathcal{X}^l$. If for all $t \in \{1, \ldots, l\}$, $I(\tilde{X}_{it}(X); \tilde{X}_{jt}(X)) = 0 \; \forall \; i, j \in \mathcal{N}$, $i \neq j$, then*

$$\max_{\mathbf{x}^l \in \mathcal{X}^l} \sum_{i=1}^{n} \left[\frac{1}{l} \sum_{t=1}^{l} d(x_t, \tilde{X}_{it}(x)) \right] \geq n - 1.$$

Proof. See Appendix A.9. □

The following theorem proves that the MDS coding scheme is optimal for all n and k when a single-message is received at the decoder.

Theorem 3. *For any n and k, if $D_k < 1 - \frac{k}{n}$ and rational, then $\forall \; (R_k(D_k), D_1, \ldots, D_k, \ldots, D_n) \in \mathcal{RD}_{worst}$, $D_1 \geq 1 - \frac{1}{n}$.*

Proof. See Appendix A.2. □

The following theorem shows that the MDS coding scheme is Pareto optimal in the distortions $D_1, \ldots, D_{k-1}$.

Theorem 4. *For any n and k, $\hat{\boldsymbol{R}}$ is* Pareto optimal *in $D_1, \ldots, D_{k-1}$,* i.e.*, there does not exist $(R', D'_1, \ldots, D'_n) \in \mathcal{RD}_{worst}$ such that either $R' < R_k(D_k)$, or $R' \leq R_k(D_k)$, $D'_i \leq 1 - \frac{i}{n}$ for all $1 \leq i \leq k-1$ and $D'_j < 1 - \frac{j}{n}$ for at least one j, $1 \leq j \leq k-1$.*

Proof. See Appendix A.3. □

The following theorem shows that for certain values of m, n and k, the MDS coding scheme is optimal when m messages are received.

Theorem 5. *For any n and k, if $m \leq \frac{k}{2}$ and $m|n$ (m divides n), then $\forall\ (R_k(D_k), D_1, \ldots, D_k, \ldots, D_n) \in \mathcal{RD}_{worst}$, $D_m \geq 1 - \frac{m}{n}$.*

Proof. See Appendix A.4. □

It is worth noting that that our converse bounds for $D_k < 1 - \frac{k}{n}$ are sharper than the cooperative or cut-set bound, which is given by $D_m \geq 1 - mR_k(D_k)$.

2.3 The Average-case Distortion Criterion

2.3.1 An Achievability Result

Theorem 6. *Let $D_k \in [0, 1]$. For any n and $k \leq n$, if $D_k \geq 1 - \frac{k}{n}$, then $\tilde{\boldsymbol{R}} \in \overline{\mathcal{RD}}_{avg}$. If $D_k < 1 - \frac{k}{n}$, then $\hat{\boldsymbol{R}} \in \overline{\mathcal{RD}}_{avg}$.*

Proof. Theorem 6 is implied by Theorem 1. However, an alternate, more conventional proof based on random binning arguments, which also proves Theorem 6 for the closure region $\overline{\mathcal{RD}}_{avg}$, is included in Appendix A.8. □

2.3.2 Optimality Results

We now present optimality results for average-case distortion. These optimality results deal primarily with single-message optimality, *i.e.*, when only one message is received at the decoder, and are weaker than the optimality results proved earlier for worst-case distortion. Moreover, the optimality results pertain to the achievable region $\mathcal{RD}_{avg}$ itself rather than its closure $\overline{\mathcal{RD}}_{avg}$. In other words, there exists a "closure gap" between the inner bound in Theorem 6 and the outer bounds presented below. It should be evident from the proofs of the optimality results in the previous section that for converse proofs, only the pairwise independence condition between the component variables $\hat{X}_{it}$ and $\hat{X}_{jt}$ is important, and this condition follows from independence at the block level. The difficulty is that when we attempt to prove an outer bound for the closure, no excess rate imposes a weaker independence condition on the transmitted messages; messages need not be completely mutually independent but rather nearly mutually independent (*i.e.*, for any k messages $f_1, \dots, f_k$, no excess rate yields $I_k(f_1; \dots; f_k) \leq \epsilon n$ for some $\epsilon > 0$, rather than $I_k(f_1; \dots; f_k) = 0$).

A similar situation for the simpler case of two-channel MD with no excess rate for two descriptions was addressed by Ahlswede in [25], where he used "wringing techniques" to prove a tight outer bound without a closure gap. The wringing technique is a way to infer near independence at the component level given near independence at the block level. By conditioning on suitable random variables, the wringing technique ensures, given two random vectors that are nearly pairwise independent, that they are also nearly pairwise independent in each component. More precisely, if $I(\mathbf{X}_1^l; \mathbf{X}_2^l) \leq \epsilon l$ for some $\epsilon > 0$, then for any $\delta > 0$ there exist $t_1, \dots, t_m \in \{1, \dots, l\}$ (where $m \leq \epsilon l / \delta$) such that for all $t \in \{1, \dots, l\},\ I(X_{1t}; X_{2t} | X_{1t_1} X_{2t_2}, \dots, X_{1t_m} X_{2t_m}) < \delta$.

It seems natural to employ the wringing technique to remove the closure gap in the optimality results presented here. However, there is one important difference between our MD problem and the two-description problem considered in [25] which renders the wringing technique ineffective in our case. In the two-description case with no excess rate for two descriptions, there is only one set of descriptions (*i.e.*, the set containing both descriptions) for which no excess rate is imposed, resulting in a single pairwise independence condition. In our n-description case with no excess rate for every k descriptions, there are $\binom{n}{k}$ sets of k descriptions for which there is no excess rate, and thus there are $\binom{n}{k}$ independence conditions, one for each of the $\binom{n}{k}$ sets. If one applies existing wringing techniques here, then one would obtain a set of conditioning variables for each of the $\binom{n}{k}$ constraints. If these sets of variables happened to be the same for all of the constraints, then we could conclude component-wise independence in all $\binom{n}{k}$ cases, but there is no guarantee that this will happen. Developing wringing techniques for this setup would be useful future work.

The following theorem shows that when only one message is received at the decoder, our coding scheme is optimal, modulo a closure operation, for all n and k satisfying $\left(1-\frac{1}{n}\right)^k \leq \frac{1}{2}$. Recall that, given D_k, we use $R_k(D_k)$ to denote $(1-D_k)/k$.

Definition 5. *For any fixed D_k, define*

$$D_1^* = \inf\{D_1 : (R_k(D_k), D_1, \ldots, D_k, \ldots, D_n) \in \mathcal{RD}_{avg}\}.$$

Theorem 7. *For any n and $k \leq n$, if $D_k \geq 1 - \frac{k}{n}$, then for any $(R_k(D_k), D_1, \ldots, D_k, \ldots, D_n) \in \mathcal{RD}_{avg}$, $D_m \geq 1 - mR_k(D_k)$ for all $m \in \mathcal{N}$. If $D_k < 1 - \frac{k}{n}$, D_k is rational*[4]*, and $\left(1-\frac{1}{n}\right)^k \leq \frac{1}{2}$, then $D_1^* \geq 1 - \frac{1}{n}$.*

Proof. See Appendix A.5. □

[4] For this theorem and subsequent theorems in this subsection, we consider rational values for D_k since any code over a finite blocklength can yield only rational distortions.

We note that $\left(1 - \frac{1}{n}\right)^k \leq \frac{1}{2}$ implies $k \geq \frac{1}{\log(n/n-1)} := \lambda(n)$. Since $\lambda(n)/n \to 1/\log e$ as $n \to \infty$, the second part of Theorem 7 provides a lower bound on D_1^* for a large range of k when n is large.

The following theorem proves single-message optimality for the coding scheme when $n = 4$ and $k = 2$. This case is not included in Theorem 7.

Theorem 8. *Let $D_k < 1 - \frac{k}{n}$ and rational. If $n = 4$ and $k = 2$, then $D_1^* \geq 1 - \frac{1}{n}$.*

Proof. See Appendix A.6. □

Theorem 7 handles the regime in which k is large. We now study the other extreme, *i.e.*, when k is small. In particular, we look at the $k = 2$ case. The following theorem provides a lower bound on the optimal single-message distortion for $n > 3$ and $k = 2$. This lower bound differs from the distortion achieved by our coding scheme by exactly $1/n$, and thus becomes progressively tighter as n increases.

Theorem 9. *Let $D_k < 1 - \frac{k}{n}$ and rational. If $k = 2$, then for $n > 3$, $D_1^* \geq 1 - \frac{2}{n}$.*

Proof. See Appendix A.7. □

We conjecture that the lower bound in Theorem 9 is not tight and that our scheme is in fact optimal. Evidence for this is provided by Theorem 8.

2.4 A General Multiple Descriptions Architecture

The scheme described above provides a substrate that can be used to construct no-excess-rate multiple descriptions codes for a general source using only a point-to-point rate-distortion code for that source. We illustrate this idea for a Gaussian source, where the resulting scheme is optimal in a certain sense. The extension to arbitrary sources should be clear from the proof. Suppose that $(\mathbf{X}_t)_{t=1}^{\infty}$ is a memoryless Gaussian process, where $\mathbf{X}_t$ is a vector of length N and has a marginal distribution $\mathcal{N}(0, \mathbf{K}_x)$. The distortion for a source-reconstruction pair $(\mathbf{X}^l, \hat{\mathbf{X}}^l)$ is measured as $\mathbf{E}\left[\frac{1}{l}\sum_{t=1}^{l}(\mathbf{X}_t - \hat{\mathbf{X}}_t)(\mathbf{X}_t - \hat{\mathbf{X}}_t)^T\right]$. We compare distortions in the positive definite sense, *i.e.*, $\mathbf{D}_A \succcurlyeq \mathbf{D}_B$ iff $\mathbf{D}_A - \mathbf{D}_B \succcurlyeq \mathbf{0}$.

Definition 6. *The rate-distortion vector $(R, \boldsymbol{D}_1, \ldots, \boldsymbol{D}_n)$ is* achievable *if for some l there exist encoders $f_i^{(l)} : \mathbb{R}^{N\times l} \to \{1, \ldots, M_i^{(l)}\}$, $i \in \mathcal{N}$ and decoders $g_{\mathcal{K}}^{(l)} : \prod_{k\in\mathcal{K}}\{1, \ldots, M_k^{(l)}\} \to \mathbb{R}^{N\times l}$, $\mathcal{K} \subseteq \mathcal{N}$, $\mathcal{K} \neq \emptyset$, such that*

$$R \geq \frac{1}{l}\log M_i^{(l)} \; \forall\, i, \text{ and}$$

$$\boldsymbol{D}_k \succcurlyeq \boldsymbol{E}\left[\frac{1}{l}\sum_{t=1}^{l}(\boldsymbol{X}_t - \hat{\boldsymbol{X}}_{\mathcal{K},t})(\boldsymbol{X}_t - \hat{\boldsymbol{X}}_{\mathcal{K},t})^T\right] \; \forall\, \mathcal{K} \subseteq \mathcal{N}, |\mathcal{K}| = k,$$

where $\hat{\boldsymbol{X}}_{\mathcal{K}}^l = \boldsymbol{E}[\boldsymbol{X}^l | f_i^{(l)}(\boldsymbol{X}^l), i \in \mathcal{K}]$.

We use $\mathcal{RD}_{gauss}$ to denote the set of achievable rate-distortion vectors and $\overline{\mathcal{RD}}_{gauss}$ to denote its closure. We consider symmetric descriptions, *i.e.*, each description has the same rate R_g and the distortion constraint depends only on the number of descriptions received. We consider the case where there is *no excess rate* for every k out of n descriptions, *i.e.*, $kR_g = R(\mathbf{D}_k)$, where $R(\cdot)$ is the Shannon rate-distortion function and

$$R(\mathbf{D}_k) = \min_{\tilde{\mathbf{D}}} \frac{1}{2}\log\frac{|\mathbf{K}_x|}{|\tilde{\mathbf{D}}|}$$
$$\text{s.t. } \tilde{\mathbf{D}} \preccurlyeq \mathbf{D}_k \text{ and}$$
$$\tilde{\mathbf{D}} \preccurlyeq \mathbf{K}_x.$$

Thus $R_g = \frac{1}{k}R(\mathbf{D}_k)$ bits/symbol.

Definition 7.

$$\mathbf{R}_G = \Big(R_g, \frac{\boldsymbol{D}_k + (n-1)\boldsymbol{K}_x}{n}, \frac{2\boldsymbol{D}_k + (n-2)\boldsymbol{K}_x}{n}, \ldots, \\ \frac{(k-1)\boldsymbol{D}_k + (n-k+1)\boldsymbol{K}_x}{n}, \boldsymbol{D}_k, \ldots, \boldsymbol{D}_k \Big).$$

Theorem 10. $\mathbf{R}_G \in \overline{\mathcal{RD}}_{gauss}$.

Proof. It suffices to show that for any $\epsilon > 0$, the rate-distortion vector

$$\mathbf{R}_{G+\epsilon} = \left(R_g + \epsilon, \frac{\mathbf{D}_k + \epsilon\mathbf{I} + (n-1)\mathbf{K}_x}{n}, \ldots, \\ \frac{(k-1)(\mathbf{D}_k + \epsilon\mathbf{I}) + (n-k+1)\mathbf{K}_x}{n}, \\ \mathbf{D}_k + \epsilon\mathbf{I}, \ldots, \mathbf{D}_k + \epsilon\mathbf{I} \right)$$

is achievable. For any $\epsilon > 0$, we know from rate-distortion theory that there exist integers l and l', with $l' \leq l(R(\mathbf{D}_k) + \epsilon)$, such that any source sequence $\mathbf{X}^l$ of l symbols can be compressed to a sequence $\mathbf{Y}^{l'}$ consisting of l' bits and then reproduced from $\mathbf{Y}^{l'}$ with distortion $\preccurlyeq \mathbf{D}_k + \epsilon\mathbf{I}$. Fix ϵ and choose a blocklength nl. Using the aforementioned rate-distortion code, we can compress the length-nl source sequence (consisting of n blocks, each of length l) into a binary sequence $\mathbf{Y}^{nl'}$ taking values in $\mathcal{X}$. Now $\mathbf{Y}^{nl'}$ can be treated as n blocks of length l' each, and can be transmitted to the decoder over the n channels using the MDS-coding based scheme proposed in Section 2.2.1. Thus every description contains l' uncoded bits (*i.e.*, one of the n blocks) of $\mathbf{Y}^{nl'}$. In particular, the decoder should be able to completely reconstruct $\mathbf{Y}^{nl'}$ upon reception of any k descriptions, *i.e.*, there is no distortion for every k out of n descriptions (this corresponds to a special case of Theorem 1 with $D_k = 0$). Thus every set of k descriptions must reveal nl' bits, and therefore the rate of a single description is $\tilde{R} = nl'/knl = l'/kl$ bits per symbol of $\mathbf{X}^l$. Moreover, since every description contains l' uncoded bits, the decoder can reconstruct ml' bits (*i.e.*, m blocks) of $\mathbf{Y}^{nl'}$ upon reception of any $m < k$ descriptions.

We now argue that $\mathbf{R}_{G+\epsilon}$ is achievable. The rate of every description is $\tilde{R} = l'/kl \leq (R(\mathbf{D}_k) + \epsilon)/k \leq R_g + \epsilon$. Moreover, any $m < k$ descriptions reveal ml' bits (m blocks) of $\mathbf{Y}^{nl'}$ completely, and reveal nothing about the other $n - m$ blocks. Thus the decoder can reconstruct a fraction m/n of $\mathbf{X}^{nl}$ (*i.e.*, m out of the n blocks of $\mathbf{X}^{nl}$) from the m blocks of $\mathbf{Y}^{nl'}$ revealed to it with distortion at most $\mathbf{D}_k + \epsilon\mathbf{I}$, and must reconstruct the remaining fraction without any information (incurring distortion $\mathbf{K}_x$). If we take the time average over all blocks, we can see that the decoder can reconstruct $\mathbf{X}^{nl}$ with distortion at most $\frac{m(\mathbf{D}_k+\epsilon\mathbf{I})+(n-m)\mathbf{K}_x}{n}$. When k or more descriptions are received, the decoder is able to reconstruct $\mathbf{Y}^{nl'}$ completely and can reconstruct $\mathbf{X}^{nl}$ with distortion at most $\preccurlyeq \mathbf{D}_k + \epsilon\mathbf{I}$. □

Next, we show that, for the special case of symmetric scalar Gaussian multiple descriptions with two levels of receivers (where one receiver reconstructs the source from any k out of n descriptions with distortion $\mathbf{D}_k$ and the second receiver reconstruct the source from all n description with distortion $\mathbf{D}_n$), and no excess rate for the second receiver, the aforementioned scheme achieves the optimal $\mathbf{D}_k$. It has been shown by Wang and Viswanath [43, Theorem 1] that given distortion constraints $\mathbf{D}_k$ and $\mathbf{D}_n$, the symmetric multiple description rate for an *i.i.d.* vector Gaussian source with mean $\mathbf{0}$ and covariance $\mathbf{K}_x$ is

$$\hat{R} = \sup_{\mathbf{K}_z \succ \mathbf{0}} \quad \frac{1}{2} \log \left(\frac{|\mathbf{K}_x|^{\frac{1}{n}} |\mathbf{K}_x + \mathbf{K}_z|^{\frac{n-k}{kn}} |\mathbf{D}_n + \mathbf{K}_z|^{\frac{1}{n}}}{|\mathbf{D}_n|^{\frac{1}{n}} |\mathbf{D}_k + \mathbf{K}_z|^{\frac{1}{k}}} \right).$$

Thus the sum rate of the n descriptions is

$$n\hat{R} = \sup_{\mathbf{K}_z \succ \mathbf{0}} \quad \frac{1}{2} \log \left(\frac{|\mathbf{K}_x| |\mathbf{K}_x + \mathbf{K}_z|^{\frac{n-k}{k}} |\mathbf{D}_n + \mathbf{K}_z|}{|\mathbf{D}_n| |\mathbf{D}_k + \mathbf{K}_z|^{\frac{n}{k}}} \right). \tag{2.1}$$

Theorem 11. *For scalar Gaussian multiple descriptions (*i.i.d. $\mathcal{N}(0,\sigma_x^2)$ *Gaussian source) with two levels of receivers (distortion constraints D_k and D_n, respectively) and no excess rate for the second receiver, $D_k \geq \frac{k}{n}D_n + \frac{n-k}{n}\sigma_x^2$.*

Proof. Assume WLOG that $\sigma_x^2 = 1$. Reducing (2.1) to the scalar case and using the no excess rate condition gives

$$\frac{1}{2}\log\left(\frac{1}{D_n}\right) = \sup_{\lambda>0} \quad \frac{1}{2}\log\left(\frac{1}{D_n}\cdot\frac{(1+\lambda)^{\frac{n-k}{k}}(D_n+\lambda)}{(D_k+\lambda)^{\frac{n}{k}}}\right),$$

which implies

$$0 = \sup_{\lambda>0} \quad \frac{1}{2}\log\left(\frac{(1+\lambda)^{\frac{n-k}{k}}(D_n+\lambda)}{(D_k+\lambda)^{\frac{n}{k}}}\right).$$

Define $f(\lambda) = \frac{(1+\lambda)^{\frac{n}{k}-1}(D_n+\lambda)}{(D_k+\lambda)^{\frac{n}{k}}}$. Then

$$\begin{aligned}
0 &= \sup_{\lambda>0} \log_e f(\lambda) \\
&= \sup_{\lambda>0} \left(\frac{n}{k}-1\right)\log_e(1+\lambda) + \log_e(D_n+\lambda) - \frac{n}{k}\log_e(D_k+\lambda) \\
&= \sup_{\lambda>0} \log_e \frac{D_n+\lambda}{1+\lambda} + \frac{n}{k}\log_e\frac{1+\lambda}{D_k+\lambda} \\
&= \sup_{\lambda>0} \log_e\left(1+\frac{D_n-1}{1+\lambda}\right) + \frac{n}{k}\log_e\left(1+\frac{1-D_k}{D_k+\lambda}\right).
\end{aligned}$$

Define

$$g(\lambda) = \frac{(\frac{D_n-1}{1+\lambda})^2}{2(1-|\frac{D_n-1}{1+\lambda}|)^2} + \frac{(\frac{1-D_k}{D_k+\lambda})^2}{2(1-|\frac{1-D_k}{D_k+\lambda}|)^2}.$$

Using the fact that

$$\log_e(1+x) \geq x - \frac{x^2}{2(1-|x|)^2} \text{ for } |x| < 1$$

we obtain

$$0 \geq \sup_{\lambda>0}\left(\frac{D_n-1}{1+\lambda} + \frac{n}{k}\left(\frac{1-D_k}{D_k+\lambda}\right) - g(\lambda)\right)$$

$$
\begin{aligned}
\frac{1-D_n}{1+\lambda} &\geq \frac{n}{k}\left(\frac{1-D_k}{D_k+\lambda}\right) - g(\lambda) \\
\frac{D_k+\lambda}{1+\lambda} &\geq \frac{n}{k}\left(\frac{1-D_k}{1-D_n}\right) - \frac{D_k+\lambda}{1-D_n}g(\lambda).
\end{aligned}
$$

Now let $\lambda \to \infty$. Then $\frac{D_k+\lambda}{1-D_n}g(\lambda) \to 0$ and $\frac{D_k+\lambda}{1+\lambda} \to 1$. We thus have

$$
\begin{aligned}
1 &\geq \frac{n}{k}\left(\frac{1-D_k}{1-D_n}\right) \\
\Rightarrow D_k &\geq \frac{k}{n}D_n + \frac{n-k}{n}.
\end{aligned}
$$

□

Acknowledgement

The author would like to thank Chao Tian for suggesting the use of systematic MDS codes for the achievability scheme in Section 2.2.1.

CHAPTER 3

OPTIMAL DELAY-RECONSTRUCTION TRADEOFFS IN PEER-TO-PEER NETWORKS

In this chapter, we study delay-reconstruction tradeoffs in P2P networks with decentralized encoding. We formulate the binary erasure robust CEO problem more precisely and describe our coding scheme. We then consider the symmetric version of the binary erasure robust CEO problem and show that our coding scheme provides a Pareto optimal delay-reconstruction tradeoff. We also consider an asymmetric, two encoder version of the problem and show that the coding scheme for the symmetric case is not optimal for this example.

3.1 Problem Formulation and Coding Scheme

We begin with the formulation of the binary erasure robust CEO problem, depicted in Figure 3.1. Let $\mathcal{N} = \{1, \ldots, n\}$ and $\mathcal{X} = \{+, -\}$. Let X be a uniform binary random variable taking values in $\mathcal{X}$. We assume that this source is *i.i.d.* over time, and we denote a length-l sequence of X by X^l. Define $Y_i = N_i \cdot X,\ i \in \mathcal{N}$, where $N_1, \ldots, N_n$ are independent Bernoulli random variables with $0 < \Pr(N_i = 0) = p_i < 1$. Thus each Y_i is the output of passing X through a binary erasure channel (Figure 3.2) with erasure probability p_i, and takes values in $\hat{\mathcal{X}} = \{+, -, 0\}$, where 0 denotes the erasure symbol. There are n encoders, each of which is a function $f_i : \hat{\mathcal{X}}^l \to \left\{1, \ldots, M_i^{(l)}\right\},\ i \in \mathcal{N}$. Encoder $f_i,\ i \in \mathcal{N}$, observes Y_i^l and transmits an encoded version of it over channel i. The decoder either receives this description without error or does not receive it at all. Excluding the case in which none of the messages is received, the receiver may receive $2^n - 1$ different combinations of messages. Thus it can be represented by $2^n - 1$ decoding functions $g_{\mathcal{K}}$, $\mathcal{K} \subseteq \mathcal{N}$, $\mathcal{K} \neq \emptyset$ of the form $g_{\mathcal{K}} : \prod_{k \in \mathcal{K}} \left\{1, \ldots, M_k^{(l)}\right\} \to \hat{\mathcal{X}}^l$. Based on the set of received messages $\mathcal{K}$, the receiver employs the corresponding decoding function to output a reconstruction $\hat{X}_{\mathcal{K}}^l$ of the original source string X^l.

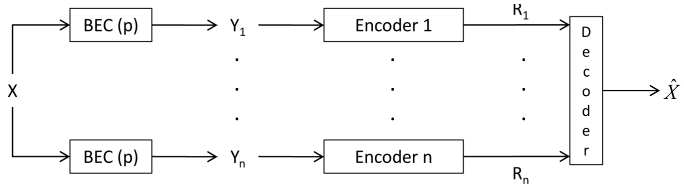

Figure 3.1: The binary erasure robust CEO problem

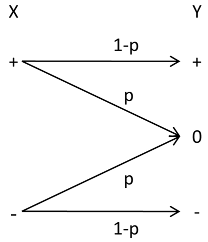

Figure 3.2: A binary erasure channel (BEC) with erasure probability p

We measure the fidelity of the reconstruction using a family of distortion measures, $\{d^\lambda\}_{\lambda>0}$, where

$$d^\lambda(x, \hat{x}) = \begin{cases} 0 & \text{if } \hat{x} = x \\ 1 & \text{if } \hat{x} = 0 \\ \lambda & \text{otherwise.} \end{cases}$$

We are particularly interested in the large-λ limit, wherein erasures incur unit cost while errors are penalized highly. In this regime, d^λ approximates the erasure distortion measure [33, p. 338].

In general, one could impose a distortion constraint for every subset of received messages. This generality is not needed here, however, so we will only measure the distortion as a function of the *number* of received messages.

Definition 8. $(R_1, R_2, \ldots, R_n, D_1, D_2, \ldots, D_n)$ *is an* achievable *rate-distortion vector if there exists a block length l for which there exist encoders $f_i,\ i \in \mathcal{N}$, and decoders $g_{\mathcal{K}}$, $\mathcal{K} \subseteq \mathcal{N}$, $\mathcal{K} \neq \emptyset$ such that*

$$\begin{aligned} R_i &\geq \frac{1}{l} \log M_i^{(l)} \\ D_k &\geq E\left[\frac{1}{l}\sum_{t=1}^{l} d^{\lambda}(X_t, \hat{X}_{\mathcal{K}t})\right] \text{ for all } \mathcal{K} \text{ s.t. } |\mathcal{K}| = k. \end{aligned} \tag{3.1}$$

Let $\mathcal{RD}_{CEO}(\lambda)$ denote the set of achievable rate-distortion vectors. Define

$$\mathcal{RD}_{CEO} = \bigcap_{\lambda \geq 1} \mathcal{RD}_{CEO}(\lambda).$$

It is worth noting that

$$D_k \geq \max_{\mathcal{K}, |\mathcal{K}|=k} \prod_{i \in \mathcal{K}} p_i,$$

since when all of the corresponding Y_i are erased for a given subset of messages, the decoder gets no information about X whatsoever and is forced to output erasures instead. We use $\overline{\mathcal{RD}}_{CEO}$ to denote the closure of $\mathcal{RD}_{CEO}$. In a P2P context, the encoders represent peers in the network that have access to partial copies Y_i of the received file X. Peers generate encoded packets in a decentralized fashion, without communicating with other peers, based on their own partial knowledge of the file. The erasure distortion measure measures how much of the file is reconstructed from these encoded messages.

A natural achievability scheme for this setup is vector quantization using erasure test channels followed by Slepian-Wolf binning at each encoder. Since this is a particularization of a scheme in [47], we provide only a high-level description and refer the reader to [47] for a detailed treatment. For a fixed block-length l, Encoder i, $i \in \mathcal{N}$ first performs vector quantization of the possible Y_i^l sequences using an erasure test channel (Figure 3.3).

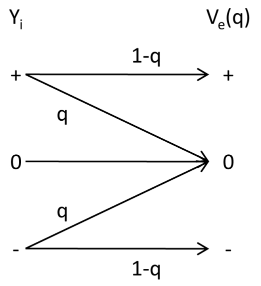

Figure 3.3: Erasure test channels

Specifically, Encoder i chooses a parameter q_i for the erasure test channel and generates codewords *i.i.d.* according to the output distribution of this channel when the input is Y_i. These codewords are then divided randomly amongst 2^{nR_i} bins. Then given the Y_i^l sequence, the encoder searches for a codeword with which it is typical, and transmits the index of the bin containing this codeword. The decoder receives the bin indices transmitted by a subset of the encoders and, if possible, uses typicality considerations to identify the codewords within the bins that were selected by those encoders. In particular, the decoder searches for codewords that are typical with respect to the output distributions of the encoders' test channels. These codewords will collectively reveal some of the source bits X^l but not others, and the decoder creates a reconstruction $\hat{X}^l$ of the file that specifies the known bits while leaving the remaining ones erased.

The aforementioned scheme exhibits a fundamental tradeoff between intermediate performance (*i.e.*, the fraction of the file that can be reconstructed when only a subset of the messages is received) and the overall efficiency of the file transfer (*i.e.*, the fraction of the file that can be reconstructed when all n messages are received). Although the scheme is valid for the case where p_i and R_i are different for different encoders, and we have stated it in its most general form, important insight can be gained into the above tradeoff if we consider the special case in which the encoders are symmetric. We therefore consider the scenario in which all of the Y_i have the same erasure probability, $p_i = p$, the rates are identical, $R_i = R$, and all of the encoders use the same test channel parameter, *i.e.*, $q_i = q$ for all i.

Let us first understand the performance of the scheme in the symmetric case. Consider the portion of the string X^l that the decoder will be able to reconstruct as a function of the number of messages received. For the first few messages, the decoder will be unable to recover the codewords chosen by the encoders. As such, it will be unable to reproduce any of the bits of X^l, and accordingly its reconstruction will be entirely erasures. After sufficiently many messages, say k, have been received, the decoder will be able to determine all k codewords from the bin indices and thereby determine some of the source bits. More precisely, the decoder will have access to k codewords, each of which is a copy of the source string with a fraction $p + (1 - p)q$ of the bits erased. Since the erasures in different codewords are independent, the fraction of erasures in the reconstruction will be

$$D_k = (p + (1 - p)q)^k$$

which by our choice of distortion measure is also the distortion. If additional messages are then received, their associated codewords can be determined through typicality considerations. These additional codewords will allow the decoder to reproduce even more of the bits of the source. In fact, the fraction of erasures in the reconstruction will be

$$D_m = (p + (1 - p)q)^m$$

where m is the number of received messages. In particular, we have $D_m = D_k^{m/k}$.

The relation $D_m = D_k^{m/k}$ captures the tradeoff between the fraction of the file that can be reconstructed from $m \leq n$ messages (intermediate performance) and the overall efficiency of the file transfer (*i.e.*, the fraction of the file that can be reconstructed from all n messages, which in this case is $1 - D_n = 1 - D_k^{n/k}$). Notice that the above scheme enables us to operate between two extremes in P2P technology which exhibit the same aforementioned tradeoff; *(i)* peers share packets without coding (*e.g.*, BitTorrent), and *(ii)*, peers share fully encoded packets (*e.g.*, network coding). Letting $k = 1$ allows us to recover the "no coding" case; there is no binning, and every message reveals a partial source string to the decoder. Letting $k = n$ allows us to recover the "coding" case; every quantized codeword is binned, and the decoder can only recover the codewords when all n messages have been received.

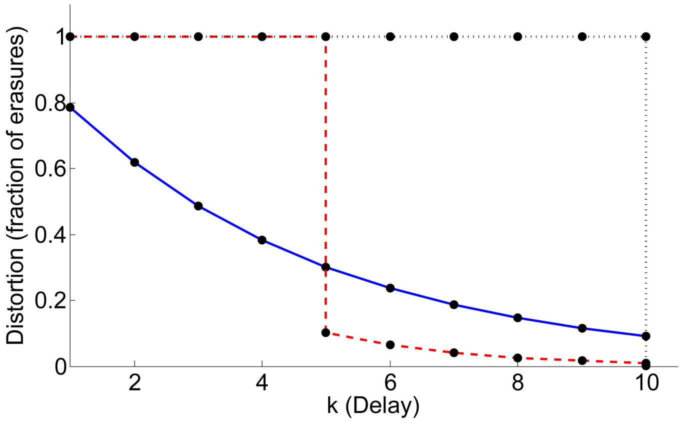

Figure 3.4: Performance of the achievability scheme for $n = 10$, $p = 0.1$, and encoder rate $R = 0.25$. The solid curve corresponds to $k = 1$ (no coding), the dotted curve to $k = 10$ (coding), and the dashed curve to $k = 5$.

By varying k, therefore, we can interpolate between the "coding" and "no coding" extremes. Figure 3.4 illustrates the performance of the scheme for $n = 10$ and $p = 0.1$. The solid curve corresponds to $k = 1$ (no coding), the dotted curve to $k = 10$ (coding), and the dashed curve to $k = 5$. An encoder rate $R = 0.25$ was used for all three cases. Notice that the "no coding" case yields good intermediate performance; 20% of the file can be reconstructed from a single message, and the distortion falls to 0.8. The overall efficiency, though, is not good; about 15% of the file cannot be reconstructed even when all of the messages have been received. The "coding" case performs contrary to the "no coding" case: nothing at all can be reconstructed with fewer than n messages ($D_m = 1$ for $m < n$), but once n messages have been received, everything can be reconstructed (the distortion is almost 0). The $k = 5$ curve, however, illustrates how the aforementioned scheme allows partial reconstruction of the source with fewer than n messages as opposed to the "coding case" (the decoder can output a partial reconstruction as long as k messages have been received), and also achieve a better overall efficiency with n messages than the "no coding" case (in fact, with $k = 5$, almost all of the file can be reconstructed from n messages).

The ability to partially reconstruct the source can prove vital in the context of distribution of content, *e.g.*, video files, in P2P networks. In such a scenario, our coding strategy can be implemented on the level of video frames rather than bits, treating the entire video file as a coding block. In this case, users with a partial reconstruction of the video file can watch the whole video by interpolating over the missing frames. This would lead to lower buffering delay (in Figure 3.4, for instance, the delay is halved for $k = 5$ as compared to the "coding" case) and might at the same time yield adequate playback quality, depending on the purposes of the user. As more messages are received, users would be able to reconstruct a higher quality video. Partial reconstruction also provides other advantages in this context; peers with partially reconstructed files can transmit uncoded bits to peers that are still waiting to receive enough messages to start decoding. This would lead to smaller user-perceived delays than with network coding, without compromising the overall efficiency of the download. Moreover, if users accidentally downloaded the wrong file, they would be able to stop the download after viewing the partial file.

In the next section, we prove optimality results for the delay-reconstruction tradeoff exhibited by the aforementioned coding scheme with symmetric encoders. In order to achieve distortion D_k with k messages and fixed encoder rate R, we must have $kR \geq R(D_k)$, where $R(\cdot)$ is the Shannon rate-distortion function for the robust CEO problem. In practice, having kR strictly greater than $R(D_k)$ is wasteful, since the additional rate can be used to convey useful information about the source and lower the distortion below D_k. We therefore focus on the case when $kR = R(D_k)$, which implies that the encoder rate is just sufficient to achieve distortion D_k with any k messages. This scenario is referred to as *no excess rate* in information theory.

3.2 Pareto Optimality of the Scheme in the Symmetric Case

We now show that for symmetric encoders, given k and D_k, the tradeoff $D_m = D_k^{m/k}$ between the distortion and the number of received messages is Pareto optimal. In particular, we will show that any scheme that achieves distortion D_k for k messages must have $D_m \geq D_k^{m/k}$. It is known from the results in [44] that the minimum per-encoder rate required to achieve a given distortion D_k when any k messages are received is

$$R = \frac{1 - D_k}{k} + g(D_k^{1/k}) \tag{3.2}$$

where[1] $g(\cdot)$ is given by

$$g(x) = \begin{cases} h(x) - (1-p)h(\frac{x-p}{1-p}) & p \leq x \leq 1 \\ 0 & x > 1. \end{cases}$$

[1] All logarithms and exponentiations in [44] have base e whereas we use base 2 here. Therefore the corresponding expression in [44] is $R = \frac{1}{k}(1 - D_k)\log 2 + g(D_k^{\frac{1}{k}})$.

By choosing the erasure test channel parameter q accordingly, the scheme described above can achieve equality in (3.2). We next show that with this choice of q, the scheme is Pareto optimal with respect to $(D_k, D_{k+1}, \ldots, D_n)$: any scheme (with the same rate R) that achieves a strictly lower D_m for some $k \leq m \leq n$ must achieve a strictly larger D_m for some $k \leq m \leq n$.

Theorem 12. *If* $(R, \ldots, R, D_1, \ldots, D_n) \in \overline{\mathcal{RD}}_{CEO}$, *and*

$$D_k = \inf \Big\{ D : (R, \ldots, R, \underbrace{1, \ldots, 1}_{k-1}, D, 1, \ldots, 1) \in \overline{\mathcal{RD}}_{CEO} \Big\},$$

i.e., R is as given by (3.2)*, then $D_m \geq (D_k)^{\frac{m}{k}}$ for all $m \geq k$.*

Note that this result makes no optimality claims about the performance of the scheme when fewer than k messages are received. Under this scheme, the decoder will be unable to recover the transmitted codewords in this regime, so it will be forced to declare an erasure for every bit in its reconstruction. It would be interesting to determine if the performance in this regime can be improved, perhaps by using the ideas in [37].

In order to prove this theorem, we first establish a new outer bound for a general problem in distributed rate-distortion.

3.2.1 Outer Bound on the Rate Region of the Multi-terminal Source Coding Problem

Consider the general problem in which we have an arbitrary number of discrete memoryless sources $Y_1, \ldots, Y_n$, with Y_i taking values in the set $\mathcal{Y}_i$, encoders $f_i,\ i \in \mathcal{N}$, a hidden source Y_0 that is not directly observed by any encoder or the decoder, and a side information source Y_{n+1}, taking values in the set $\mathcal{Y}_{n+1}$, which is observed by the decoder but not by any encoder. In particular, $\{Y_{0,t}, Y_{1,t}, \ldots, Y_{n,t}, Y_{n+1,t}\}_{t=1}^{\infty}$ is a vector-valued, finite-alphabet and memoryless source. Although we consider finite-alphabet sources here, the outer bound is extensible to continuous or countable alphabets, *e.g.*, Gaussian sources, by using the approach in [44]. Encoder f_i observes a length-l sequence of Y_i and transmits a message to the decoder based on the mapping

$$f_i^{(l)} : \mathcal{Y}_i^l \to \left\{1, \ldots, M_i^{(l)}\right\}.$$

The decoder seeks to reconstruct the sources, or functions of the sources, from subsets of messages $f_{\mathcal{K}} = \{f_k^{(l)},\ k \in \mathcal{K}\}$, where $\mathcal{K} \subset \mathcal{N}, \mathcal{K} \neq \emptyset$. Since we allow the reconstruction of functions of the sources instead of, or in addition to, the sources themselves, we represent the reconstructed sequences by $V_1^l, \ldots, V_J^l$ (with $V_{j,t},\ t \in \{1, \ldots, l\}, j = 1, \ldots, J$, taking values in the set $\mathcal{V}_j$). Given a subset of messages $\mathcal{K} \subset \mathcal{N}, \mathcal{K} \neq \emptyset$ and $j \in \{1, \ldots, J\}$, the decoder thus uses the mappings

$$\left(g_{\mathcal{K}}^j\right)^{(l)} : \mathcal{Y}_{n+1}^l \times \prod_{k \in \mathcal{K}} \left\{1, \ldots, M_k^{(l)}\right\} \to \mathcal{V}_j^l.$$

We have J distortion measures

$$d_j : \prod_{i=0}^{n+1} \mathcal{Y}_i \times \mathcal{V}_j \to \mathbb{R}^+,$$

one for each constraint.

For every $j \in \{1, \ldots, J\}$, we impose a common distortion constraint for all size-k subset of messages used to reconstruct V_j^l. More precisely, for every $j \in \{1, \ldots, J\}$, all $\binom{n}{k}$ subsets of messages of size k, when used to reconstruct V_j^l, must satisfy a single distortion constraint. Thus there are nJ distortion constraints in total. Let $\mathbf{Y}_{\mathcal{K}}$ denote $(Y_k)_{k \in \mathcal{K}}$, and Y_{i^c} denote $Y_{\{i\}^c}$. Moreover, $Y_{i,a:b}$ denotes $\{Y_{i,a}, Y_{i,a+1}, \ldots, Y_{i,b}\}$.

Definition 9. *The rate-distortion vector* $(\mathbf{R}, \mathbf{D}) =$

$$(R_1, \ldots, R_n, D_{1,1}, D_{2,1}, \ldots, D_{n,1}, D_{1,2}, \ldots, D_{n,2}, \ldots, \\ D_{1,J}, \ldots, D_{n,J})$$

is achievable if for some l *there exist encoders* $f_i^{(l)}$, $i \in \mathcal{N}$, *and decoders* $(g_{\mathcal{K}}^j)^l$, $\mathcal{K} \subset \mathcal{N}, \mathcal{K} \neq \emptyset, j = 1, \ldots, J$, *such that*

$$\begin{aligned} R_i &\geq \frac{1}{l} \log M_i^{(l)}, \text{ and} \\ D_{k,j} &\geq \max_{\mathcal{K}:|\mathcal{K}|=k} \mathbf{E}\left[\frac{1}{l}\sum_{t=1}^{l} d_j(Y_{0,t}, \mathbf{Y}_{\mathcal{K},t}, Y_{n+1,t}, V_{j,t})\right]. \end{aligned} \tag{3.3}$$

As in [44], we use $\mathcal{RD}_\star$ to denote the set of achievable rate-distortion vectors and $\overline{\mathcal{RD}_\star}$ to denote its closure. We use the following definitions from [44].

Definition 10. *Let* $Y_0, Y_1, \ldots, Y_{n+1}$ *be generic random variables with the distribution of the source at a single time. Let* Γ_o *denote the set of finite-alphabet random variables* $\gamma = (U_1, \ldots, U_n, V_1, \ldots, V_j, W, T)$ *satisfying*

(i) (W, T) *is independent of* $(Y_0, \mathbf{Y}_{\mathcal{N}}, Y_{n+1})$,

(ii) $U_i \leftrightarrow (Y_i, W, T) \leftrightarrow (Y_0, \mathbf{Y}_{i^c}, Y_{n+1}, \mathbf{U}_{i^c})$, *shorthand for "*$U_i$, (Y_i, W, T) *and* $(Y_0, \mathbf{Y}_{i^c}, Y_{n+1}, \mathbf{U}_{i^c})$ *form a Markov chain in this order", for all* $i \in \mathcal{N}$, *and*

(iii) $(Y_0, \mathbf{Y}_{\mathcal{N}}, W) \leftrightarrow (\mathbf{U}_{\mathcal{N}}, Y_{n+1}, T) \leftrightarrow (V_1, \ldots, V_j)$.

Definition 11. *Let* ψ *denote the set of finite-alphabet random variables* Z *with the property that* $Y_1, \ldots, Y_n$ *are conditionally independent given* (Z, Y_{n+1}).

There are many ways of coupling a given $Z \in \psi$ and $\gamma \in \Gamma_o$ to the source. We shall only consider the *Markov coupling* for which $Z \leftrightarrow (Y_0, \mathbf{Y}_{\mathcal{N}}, Y_{n+1}) \leftrightarrow \gamma$. We now state our outer bound.

Definition 12. *Let* $\mathcal{RD}_o(Z,\gamma) = \Bigg\{(\mathbf{R},\mathbf{D}):$

$$\sum_{i\in\mathcal{K}} R_i \geq \max\Big(I(Z;\mathbf{U}_{\mathcal{K}}|Y_{n+1},T), I(Z;\mathbf{U}_{\mathcal{K}}|\mathbf{U}_{\mathcal{K}^c},Y_{n+1},T)\Big)$$

$$+\sum_{i\in\mathcal{K}} I(Y_i;U_i|Z,Y_{n+1},W,T)\ \forall\mathcal{K}\subseteq\mathcal{N}, \textit{and}$$

$$D_{k,j} \geq \max_{\mathcal{K}:|\mathcal{K}|=k} \mathbf{E}[d_j(Y_0,\mathbf{Y}_{\mathcal{K}},Y_{n+1},V_j)],\ j=1,\ldots,J\Bigg\}.$$

Then define

$$\mathcal{RD}_o = \bigcap_{Z\in\psi}\bigcup_{\gamma\in\Gamma_o}\mathcal{RD}_o(Z,\gamma).$$

Theorem 13. $\mathcal{RD}_o$ *is an outer bound on the rate-distortion region for the general problem,* i.e., $\mathcal{RD}_\star \subseteq \mathcal{RD}_o$*.*

Proof. See Appendix B.1. □

The new bound is more general than the bound in [44]. Even if we apply it to the setup of [44], however, the new bound offers an improvement. Specifically, whereas the bound in [44] lower bounds the sum rate of a subset $\mathcal{K}$ of messages by $I(Z;\mathbf{U}_{\mathcal{K}}|\mathbf{U}_{\mathcal{K}^c},Y_{n+1},T)$, the new bound improves upon it by taking the maximum of $I(Z;\mathbf{U}_{\mathcal{K}}|\mathbf{U}_{\mathcal{K}^c},Y_{n+1},T)$ and $I(Z;\mathbf{U}_{\mathcal{K}}|Y_{n+1},T)$. This improvement is useful for establishing the main result. Notice that if we have the Markov chain $U_i \leftrightarrow (Y_i,T) \leftrightarrow (\mathbf{U}_{i^c})$, then $I(Z;\mathbf{U}_{\mathcal{K}}|\mathbf{U}_{\mathcal{K}^c},Y_{n+1},T) \leq I(Z;\mathbf{U}_{\mathcal{K}}|Y_{n+1},T)$. Since, in our setup, a weaker Markov chain condition (Definition 10, (ii)) is being imposed, the above inequality might not hold here. However, as we show in the proof of Theorem 12, using $I(Z;\mathbf{U}_{\mathcal{K}}|Y_{n+1},T)$ instead of $I(Z;\mathbf{U}_{\mathcal{K}}|\mathbf{U}_{\mathcal{K}^c},Y_{n+1},T)$ yields a tight lower bound, which suggests that the outer bound in [44] could be loose for our setup.

3.2.2 Proof of Theorem 12

We begin with the following lemma.

Lemma 2. *Suppose $p^m \leq D$ and that $(\mathbf{U}, X, \hat{X}_{\mathcal{K}}, \mathbf{Y}, W, T)$ for all $\mathcal{K}, |\mathcal{K}| = m$ is such that*

(i) $(X, \mathbf{Y}, \mathbf{U}_{\mathcal{K}^c}, W) \leftrightarrow (\mathbf{U}_{\mathcal{K}}, T) \leftrightarrow \hat{X}_{\mathcal{K}}$,

(ii) $U_i \leftrightarrow (Y_i, W, T) \leftrightarrow (X, \mathbf{Y}_{i^c}, \mathbf{U}_{i^c})$ *for all* $i \in \mathcal{N}$, *and*

(iii) $\frac{1}{m}\sum_{i\in\mathcal{K}} I(Y_i; U_i|X, W, T) \leq g(D^{1/m})$.

Let $\tilde{D} = \max_{\mathcal{K}:\mathcal{K}=m} E[d^\lambda(X, \hat{X}_{\mathcal{K}})]$. For $\delta \in (0, 1/2]$, if

$$\lambda \geq \max\left[4\left(\frac{32m}{\delta p(1-p)}\right)^{2m}, \left(\frac{\tilde{D}}{\delta}\right)^2\right],$$

then $\tilde{D} \geq D - \xi(\tilde{D}, \delta)$ for some continuous $\xi \geq 0$ satisfying $\xi(\tilde{D}, 0) = 0$.

Proof. See Appendix B.2. □

Proof of Theorem 12. It suffices to prove Theorem 12 for a single subset of messages of size $m \geq k$. Fix $\delta \in (0, 1/2]$, and suppose λ satisfies

$$\lambda \geq \max\left[4\left(\frac{32m}{\delta p(1-p)}\right)^{2m}, \left(\frac{D_k}{\delta}\right)^2\right].$$

It follows from Theorem 13 by taking $Z = X$ in the definition of $\mathcal{RD}_o(Z, \gamma)$ (Definition 12) and from the monotonicity of $\mathcal{R}_o(\mathbf{D}, \lambda)$ with respect to λ that there exist $R \in \mathbb{R}^+$ and $\gamma \in \Gamma_o$ such that, for all subsets $\mathcal{K}$ of size k,

$$\begin{aligned} D_k + \delta &\geq E[d^\lambda(X, \hat{X}_{\mathcal{K}})], \text{ and} \\ kR + \delta &\geq k\mathcal{R}_o(\mathbf{D}, \lambda) + \delta \qquad (3.4) \\ &\geq I(X; \mathbf{U}_{\mathcal{K}}|T) + \sum_{i\in\mathcal{K}} I(Y_i; U_i|X, W, T). \end{aligned}$$

From (3.2) and (3.4), it follows that

$$\frac{I(X;\mathbf{U}_{\mathcal{K}}|T)}{k}+\frac{1}{k}\sum_{i\in\mathcal{K}} I(Y_i;U_i|X,W,T) \\ \leq \frac{(1-D_k)}{k}+g(D_k^{\frac{1}{k}})+\frac{\delta}{k}. \tag{3.5}$$

Now by the data processing inequality, $I(X;\mathbf{U}_{\mathcal{K}}|T)=I(X;\mathbf{U}_{\mathcal{K}},T)\geq I(X;\hat{X}_{\mathcal{K}})$.

Let $\varepsilon = 1(X\cdot\hat{X}_{\mathcal{K}}=-1)$. We then have

$$\begin{aligned} I(X;\mathbf{U}_{\mathcal{K}}|T) &\geq H(X)-H(X|\hat{X}_{\mathcal{K}}) \\ &= 1-H(X,\varepsilon|\hat{X}_{\mathcal{K}}) \\ &= 1-H(\varepsilon|\hat{X}_{\mathcal{K}})-H(X|\varepsilon,\hat{X}_{\mathcal{K}}) \\ &\geq 1-h(D_k/\lambda)-\Pr(\hat{X}_{\mathcal{K}}=0) \\ &\geq (1-D_k)-h(\delta). \end{aligned}$$

Using this and (3.5), we can upper bound $\frac{1}{k}\sum_{i\in\mathcal{K}} I(Y_i;U_i|X,W,T)$ as

$$\frac{1}{k}\sum_{i\in\mathcal{K}} I(Y_i;U_i|X,W,T)\leq g(D_k^{\frac{1}{k}})+\frac{h(\delta)}{k}+\frac{\delta}{k}. \tag{3.6}$$

We will now show

$$\frac{1}{m}\sum_{i=1}^{m} I(Y_i;U_i|X,W,T)\leq g(D_k^{\frac{1}{k}})+\frac{h(\delta)}{k}+\frac{\delta}{k},\ m\geq k. \tag{3.7}$$

Suppose that the U_i are ordered according to the mutual informations $I(Y_i;U_i|X,W,T)$, *i.e.*, we have an ordered list of messages $U_1,\ldots,U_m$ in which, for all $i,j\in\{1,\ldots,m\},U_i$ and U_j are such that $I(Y_i;U_i|X,W,T)\leq I(Y_j;U_j|X,W,T)$ when $i\leq j$. The last k elements of this list, $U_{m-k+1},\ldots,U_m$, must satisfy (3.6), i.e.,

$$\frac{1}{k}\sum_{i=m-k+1}^{m} I(Y_i;U_i|Y_0,W,T)\leq g(D_k^{\frac{1}{k}})+\frac{h(\delta)}{k}+\frac{\delta}{k}. \tag{3.8}$$

All other elements in the list yield equal or strictly smaller mutual informations. Therefore, if we average over a larger subset of messages, the average will never increase. We thus have

$$\frac{1}{m}\sum_{i=1}^{m} I(Y_i;U_i|X,W,T) \leq \frac{1}{k}\sum_{i=m-k+1}^{m} I(Y_i;U_i|X,W,T).$$

Using this and (3.8), we obtain (3.7). Define

$$(D_k - \zeta(D_k,\delta))^{\frac{1}{k}} = g^{-1}\left(g(D_k^{\frac{1}{k}}) + \frac{h(\delta)}{k} + \frac{\delta}{k}\right)$$

for some continuous $\zeta \geq 0$ satisfying $\zeta(D_k, 0) = 0$. We then have

$$\frac{1}{m}\sum_{i=1}^{m} I(Y_i;U_i|X,W,T) \leq g((D_k - \zeta(D_k,\delta))^{\frac{1}{k}}). \tag{3.9}$$

From (3.9), we obtain, by using Lemma 2,

$$D_m \geq (D_k - \zeta(D_k,\delta))^{\frac{m}{k}} - \xi(D_m,\delta)$$

for some continuous $\xi \geq 0$ satisfying $\xi(D_m, 0) = 0$. The proof is completed by letting $\lambda \to \infty$ and then $\delta \to 0$. □

3.3 Suboptimality in the Asymmetric Case

In the previous section, we considered symmetric peers and showed that the coding scheme described in Section 3.1 provides a Pareto optimal delay-reconstruction tradeoff. If we consider asymmetric encoder observations, *i.e.*, the binary erasure probabilities p_i of the channels from X to Y_i are not identical, then it becomes natural that encoders encode at different rates, since some encoders (with smaller p_i) will have a better knowledge of the source.

We now consider a very simple asymmetric case with two encoders and show that the achievable scheme is no longer optimal; more precisely, the choice of an erasure test channel is no longer optimal. Encoder 1 observes the binary source X directly (*i.e.*, $p_1 = 0$), while Encoder 2 observes an erased version Y of the source with $p_2 = p > 0$. Both encoders transmit messages to a decoder, which then attempts to reconstruct X upon reception of both messages. This setup is referred to as a one-helper problem (Figure 3.5), and the two encoders, Encoders 1 and 2, are referred to as the main encoder and the helper, respectively. The goal is to characterize the tradeoff between the rate of the main encoder, R_1, the rate of the helper, R_2, and the resulting distortion.

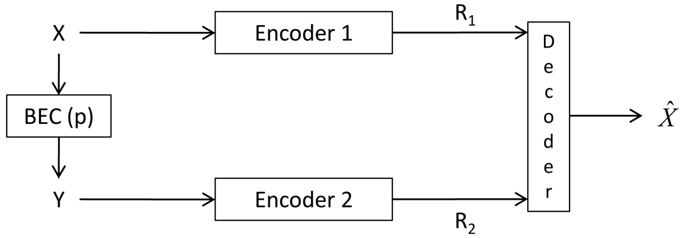

Figure 3.5: The erasure one-helper problem

Before showing that erasure test channels are suboptimal for this problem, it is worth mentioning why this suboptimality is unexpected. Existing results in distributed rate-distortion theory suggest a connection between binary erasure problems and their quadratic Gaussian counterparts. For instance, for the Wyner-Ziv problem, both instances have no rate loss [51], and this is shown using erasure and Gaussian test channels, respectively. Similarly, the only two instances of the CEO problem for which conclusive results are available at all rates are the erasure [44] and Gaussian [50] ones, and again the optimal schemes use erasure and Gaussian test channels, respectively. For the quadratic Gaussian version of the one-helper problem [52], Gaussian test channels are known to achieve the entire rate region. This suggests that erasure test channels might be optimal for the erasure version, yet we shall see that they are not in general, even if the decoder's goal is to reproduce X losslessly.

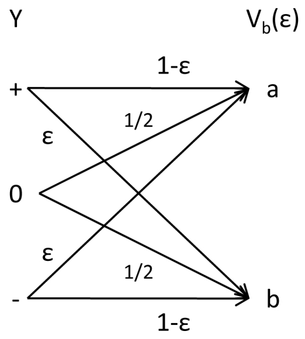

Figure 3.6: A family of symmetric test channels

Specifically, we show that for some rate constraints R_2 on the helper, the alternate family of test channels $V_b(\epsilon)$ depicted in Figure 3.6 meet the helper's rate constraint while allowing the primary encoder to use less rate. The optimal test channel for the lossless one-helper problem, given a rate constraint R_2 on the helper, is given by the optimal solution to the following optimization problem [33]:

$$\begin{aligned} &\min_{p(v|y)} H(X|V) \\ &\text{s.t. } I(Y;V) \leq R_2 \\ &X \leftrightarrow Y \leftrightarrow V. \end{aligned} \tag{3.10}$$

If we restrict the minimization to the family of channels $V_b(\epsilon)$ and the class of erasure channels $V_e(q)$, then it suffices to show that given a rate constraint R_2 on the helper, the optimal $H(X|V_b)$ is smaller than the optimal $H(X|V_e)$. Figure 3.7 depicts the optimal $H(X|V_b)$ and $H(X|V_e)$ against R_2. Notice that for low values of R_2, $H(X|V_b)$ is lower than $H(X|V_e)$, signifying that erasure test channels are the worse of the two families of channels.

The superiority of the $V_b(\epsilon)$ test channel can be understood as follows. Define a Bernoulli random variable E such that $E = 1$ when Y is erased and $E = 0$ when Y is not erased. Since E is a function of Y, we have

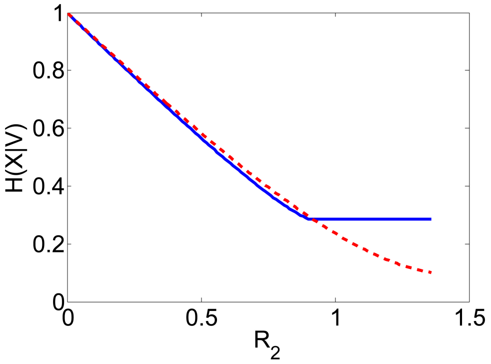

Figure 3.7: Plot of $H(X|V_b)$ (solid) and $H(X|V_e)$ (dashed) against R_2 for $p = 0.1$. For low values of R_2, $H(X|V_b)$ is smaller than $H(X|V_e)$.

$I(Y;V_b(\epsilon)) = I(Y,E;V_b(\epsilon)) = I(E;V_b(\epsilon)) + I(Y;V_b(\epsilon)|E)$. Likewise, $I(Y;V_e(q)) = I(E;V_e(q)) + I(Y;V_e(q)|E)$. Now $I(E;V_b(\epsilon)) = 0$, *i.e.*, $V_b(\epsilon)$ communicates no information about whether Y is erased. In contrast, $I(E;V_e(q)) > 0$, *i.e.*, erasure test channels expend positive rate transmitting information about the location of erasures in Y^l. This information is not pertinent to the problem of reconstructing X, and is therefore wasteful. Of course, when $\epsilon > 0$, X can never be determined with certainty from the output of the $V_b(\epsilon)$ channel. If the goal is to reproduce X^l from the helper's codeword, then the $V_b(\epsilon)$ would be a poor choice. Here, however, the helper's objective is simply to minimize $H(X|V)$.

Thus the erasure test channel is suboptimal, although Figure 3.7 shows that the benefit of using the alternate test channel $V_b(\epsilon)$ is small. Indeed, numerical solution to (3.10) for various problem instances suggest that erasure test channels are very nearly optimal and are therefore sufficient in practice. Showing this rigorously is an interesting problem for future work.

CHAPTER 4

LOSSY SOURCE CODING WITH BYANTINE ADVERSARIES

In this chapter, we study source coding in the presence of Byzantine adversaries. We formulate the lossy source coding problem with Byzantine adversaries. We present the separation-based coding scheme for general sources and arbitrary distortion measures and show that it achieves the factor-of-2 rule. We prove that our scheme is optimal for uniform binary sources with Hamming distortion and Gaussian sources with squared error distortion, and then show that the factor-of-2 rule is pessimistic for binary sources and erasure distortion, and provide a joint source-channel coding based scheme that is optimal for a three-encoder instance of this problem.

4.1 Problem Formulation

Let $\{X_t\}_{t=1}^{\infty}$ be an *i.i.d.* source, with the random variables X_t taking values in the (possibly infinite) alphabet $\mathcal{X}$. There are n encoders, t of which are traitors, that observe X^l and transmit a message to a decoder, which then attempts to reconstruct X^l from the received messages up to a specified distortion. The traitors' goal is to maximize the expected distortion in the decoder's reconstruction, and they choose their messages in order to fulfill this goal, with full knowledge of X^l, the other $n-1$ messages, and the decoder's decoding strategy. The number of traitors, t, is known to all the encoders and the decoder. However, their location among the n encoders (*i.e.*, which of the n encoders are traitors) is unknown to the honest encoders and the decoder. Moreover, the traitors can observe X^l and then decide which encoders to take over. The traitors' location among the n encoders and their actions can therefore be different for different source sequences.

Let $\hat{\mathcal{X}}$ denote the reconstruction space, with an associated distortion measure $d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}$. Let $\mathcal{N} = \{1, \ldots, n\}$. A *code* $(f_1, \ldots, f_n, g)$ is a collection of encoders $f_i : \mathcal{X}^l \to \{1, \ldots, M_i^{(l)}\}$, $i \in \mathcal{N}$, and a decoder $g : \prod_{i=1}^n \{1, \ldots, M_i^{(l)}\} \to \hat{\mathcal{X}}^l$. A rate-distortion vector $(R_1, \ldots, R_n, D)$ is said to be *achievable* if for all sufficiently large l, there exist encoders f_i and a decoder g such that

$$R_i \geq \frac{1}{l} \log M_i^{(l)} \text{ for all } i, \text{ and}$$
$$D \geq \mathbf{E}\left[\max_{\substack{H \subset \mathcal{N} \\ |H| = n-t}} \max_{C_{H^c}} \frac{1}{l} \sum_{t=1}^{l} d(X_t, g(C_1, \ldots, C_n)) \right], \text{ where } C_i = f_i(X^l) \ \forall i \in H.$$

Let $\mathcal{RD}$ denote the set of achievable rate-distortion vectors, and let $\overline{\mathcal{RD}}$ denote its closure. Moreover, let $R(\cdot)$ denote Shannon's rate-distortion function.

Definition 13. $\mathcal{RD}^* = \{(R_1, \ldots, R_n, D) : \forall \mathcal{S} \subset \mathcal{N},\ |\mathcal{S}| = n - 2t, \sum_{i \in \mathcal{S}} R_i \geq R(D)\}$.

Note that $\mathcal{RD}^*$ is the factor-of-2 region. The following theorem, proved in the next section, shows that $\mathcal{RD}^*$ is achievable.

Theorem 14. *Suppose there exists a reconstruction sequence $\hat{X}_0^l \in \hat{\mathcal{X}}^l$ such that $d(x^l, \hat{X}_0^l)$ is finite for all $x^l \in \mathcal{X}^l$. Then $\mathcal{RD}^* \subset \overline{\mathcal{RD}}$.*

4.2 A Separation-based Achievability Scheme

The achievability scheme we present in order to prove Theorem 19 consists of two stages: rate-distortion quantization and adversarial error correction. Our coding scheme separates the lossy source coding part of the problem from the adversarial error correction part. The lossy source coding part is taken care of in the first stage. The second stage deals with adversarial error correction, treating the quantized sequences generated in the first stage as a message to be transmitted over a channel with adversarial errors. The first stage corresponds to source coding (rate-distortion quantization) and the second stage corresponds to channel coding for transmitting the quantized sequences from the first step over the non-stochastic, packetized, adversarial channel depicted in Figure 4.1, where the original message W is transmitted to the decoder in the form of n packets, t of which are corrupted by traitors. Source-channel separation dictates that reliable communication can occur as long as $R(D) < C$, where C is the capacity of the channel. In Section 4.6 we show that the capacity of the channel shown in Figure 4.1 is in fact $\min_{S,S=|n-2t|} \sum_{i \in S} R_i$. With source-channel separation, reliable communication can occur as long as $R(D) \leq \min_{S,S=|n-2t|} \sum_{i \in S'}$ which is the statement of Theorem 19.

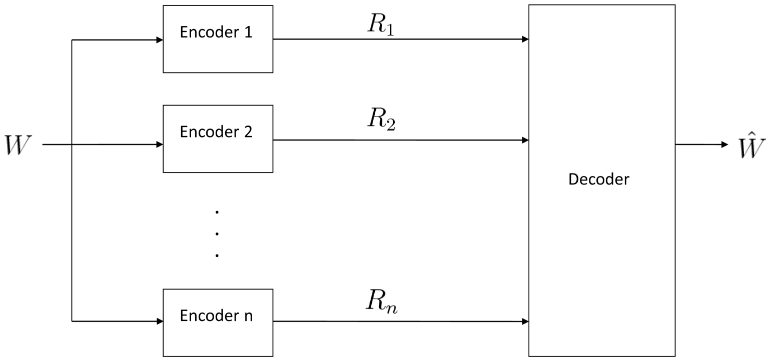

Figure 4.1: A non-stochastic, packetized adversarial channel where the original message is transmitted as n packets, t of which are corrupted by traitors.

Proof of Theorem 19. Choose $\epsilon > 0$, $\delta > 0$, and $0 \leq \alpha < (n-2t)\epsilon$. Given the source distribution $p(x)$, fix $p(\hat{x}|x)$ such that $I(X;\hat{X}) = R(D)$. Compute $p(\hat{x}) = \sum_x p(x)p(\hat{x}|x)$.

Rate-distortion quantization: Fix a blocklength l, and generate a codebook

$\mathcal{C}$ consisting of $2^{(lR(D)+\alpha)}$ sequences $\hat{X}^l$ drawn randomly and *i.i.d.* from the marginal distribution $p(\hat{x})$. Index the sequences in $\mathcal{C}$ by $w \in \{1, \ldots, 2^{l(R(D)+\alpha)}\}$.

Random binning: For all $i \in \mathcal{N}$, Encoder i bins the $2^{(lR(D)+\alpha)}$ sequences in $\mathcal{C}$ uniformly and independently into $2^{l(R_i+\epsilon)}$ bins.

Encoding: Observe a length-l source sequence X^l and find a w such that $(X^l, \hat{X}^l(w))$ are distortion typical [33, p. 319]. If there is more than one such w, pick w to be the least one. If there is no such w, set $w = 1$. Let $b_i = f_i(\hat{X}^l(w))$ be the bin index of $\hat{X}^l(w)$ at Encoder i. Encoder i transmits b_i to the decoder.

Decoding: For each set of $n - t$ messages, the decoder attempts to generate a reconstruction of X^l. In particular, for $H \subset \mathcal{N}$, $|H| = n - t$, the decoder searches the bins indexed by b_i, $i \in H$, for a sequence $\hat{X}^l_H$ such that $f_i(\hat{X}^l_H) = b_i$ for all $i \in H$. If there is exactly one such sequence $\hat{X}^l_H$ in the bins indexed by b_i, $i \in H$, set $\hat{X}^l_H$ to be the reconstruction for the set H. If there is no such sequence, or there is more than one such sequence, set $\hat{X}^l_H = \emptyset$.

Consider now the $\binom{n}{t}$ sequences $\hat{X}^l_H$ the decoder generates for $H \subset \mathcal{N}$, $|H| = n - t$. If there exists exactly one sequence $\hat{X}^l$ such that $\hat{X}^l_H = \hat{X}^l$ for all $\hat{X}^l_H \neq \emptyset$, output $\hat{X}^l$ as the reconstruction of X^l. If $\hat{X}^l_H = \emptyset$ for all H, or if $\hat{X}^l_{H_1} \neq \hat{X}^l_{H_2}$ for some $H_1, H_2 \subset \mathcal{N}$, output $\hat{X}^l_0$ as the reconstruction.

Error analysis: There is at least one set $\mathcal{H}$ of $n - t$ encoders that are all honest. By virtue of the encoding strategy, there is guaranteed to be at least one sequence common to all the bins indexed by b_i, $i \in \mathcal{H}$. If there is only one such sequence (and this would be the true quantized sequence $\hat{X}^l$), the decoder would output this sequence as the reconstruction for $\mathcal{H}$. If, however, there is more than one sequence common to all the bins, then the decoder would set

$\hat{X}^l_{\mathcal{H}} = \emptyset$. Define the error event

$$F_H : \hat{X}^l_H \neq \emptyset \text{ and } \hat{X}^l_H \neq \hat{X}^l, \forall H \in \mathcal{N}, |H| = n - t.$$

Define $E = \{\hat{X}^l_{\mathcal{H}} = \emptyset\} \cup (\bigcup_H F_H)$. Observe now that since there are t traitors, any set H of size $n - t$ has at least $n - 2t$ honest encoders. Denote the honest encoders in H by S_H. Note that the encoders in S_H will send bin indices corresponding to the true quantized sequence $\hat{X}^l$. Denote by E_{S_H} the event that the bins reported by S_H contain more than one common sequence. Then $E^c_{S_H}$ is the event that the bins reported by S_H contain exactly one common sequence (which would be $\hat{X}^l$). If $E^c_{S_H}$ occurs, then no matter what the traitors $S^c_H \cap H$ do, the decoder will output either $\hat{X}^l$ or $\emptyset$ for H (this is because if the traitors choose to send the bin indices for $\hat{X}^l$, then the decoder would find $\hat{X}^l$ in all the bins in H, and therefore output $\hat{X}^l$; if however, the traitors choose to send bin indices for a different sequence, then the decoder would not find that sequence in at least one of the bins reported by the honest encoders in S_H, and will therefore output $\emptyset$). Thus it is evident that F_H will occur only if E_{S_H} occurs. Hence $\Pr(F_H) \leq \Pr(E_{S_H})$.

Now let $f_i^{-1}(f_i(X^l))$ denote the preimage of the message that the i^{th} encoder sends for X^l. Define $E'_S = |\bigcap_{i \in S} f_i^{-1}(f_i(X^l))| \neq 1$. Thus E'_S is the event that the bins corresponding to the messages sent by the encoders in S do not contain exactly one common sequence. Notice that $E_{S_H} \subset \bigcup_{S, |S| = n-2t} E'_S$ for all H, and therefore $E \subset \bigcup_{S,|S|=n-2t} E'_S$.

Suppose now that for any set S of $n - 2t$ encoders, $\sum_{i \in S} R_i > R(D)$. We bound $\Pr(E)$ as follows. By the union bound,

$$\Pr(E) \leq \sum_{\substack{S: \\ |S| = n-2t}} \Pr(E'_S)$$

$$
\begin{aligned}
&= \sum_{\substack{S:\\ |S|=n-2t}} \Pr(\exists \tilde{X}^l \in \mathcal{C}, \tilde{X}^l \neq \hat{X}^l : f_i(\tilde{X}^l) = f_i(\hat{X}^l) \; \forall i \in S) \\
&= \sum_{C} p(C) \sum_{\substack{S:\\ |S|=n-2t}} \Pr(\exists \tilde{X}^l \in \mathcal{C}, \tilde{X}^l \neq \hat{X}^l : f_i(\tilde{X}^l) = f_i(\hat{X}^l) \; \forall i \in S \big| \mathcal{C} = C) \\
&= \sum_{C} p(C) \sum_{x^l} p(x^l) \sum_{\substack{S:\\ |S|=n-2t}} \Pr(\exists \tilde{X}^l \in \mathcal{C}, \tilde{X}^l \neq \hat{x}^l : f_i(\tilde{X}^l) = f_i(\hat{x}^l) \; \forall i \in S \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \Big| X^l = x^l, \mathcal{C} = C) \\
&\leq \sum_{C} p(C) \sum_{x^l} p(x^l) \sum_{\substack{S:\\ |S|=n-2t}} \sum_{\substack{\tilde{x}^l \in C\\ \tilde{x}^l \neq \hat{x}^l}} \Pr(f_i(\tilde{x}^l) = f_i(\hat{x}^l) \; \forall i \in S \big| X^l = x^l, \mathcal{C} = C) \\
&\leq \sum_{C} p(C) \sum_{x^l} p(x^l) \sum_{\substack{S:\\ |S|=n-2t}} |C| 2^{-l \sum_{i \in S}(R_i + \epsilon)} \\
&= \sum_{C} p(C) \sum_{x^l} p(x^l) \sum_{\substack{S:\\ |S|=n-2t}} 2^{l(R(D)+\alpha)} 2^{-l \sum_{i \in S}(R_i + \epsilon)} \\
&\leq \sum_{C} p(C) \sum_{x^l} p(x^l) \sum_{\substack{S:\\ |S|=n-2t}} 2^{-l((n-2t)\epsilon - \alpha)} \\
&= \sum_{C} p(C) \sum_{x^l} p(x^l) \binom{n}{2t} 2^{-l((n-2t)\epsilon - \alpha)} \\
&= \binom{n}{2t} 2^{-l((n-2t)\epsilon - \alpha)},
\end{aligned}
$$

where the last inequality follows because $\sum_{i \in S} R_i > R(D)$. Notice now that if E^c occurs, then the decoder outputs the true quantized sequence $\hat{X}^l$ for $\mathcal{H}$, and for every H, $H \neq \mathcal{H}$, the decoder either outputs $\hat{X}^l$ or $\emptyset$. Thus the decoder outputs $\hat{X}^l$ as the reconstruction of X^l. If, however, E occurs, then the decoder reconstructs the wrong quantized sequence. Let l be sufficiently large so that, by the rate-distortion theorem, the distortion when E^c occurs is less than $D + \delta$

when averaged over X^l and $\mathcal{C}$. We have

$$\begin{aligned}\mathbf{E}_{f,g}\mathbf{E}_{\mathbf{X}}\left[\max_{\substack{H\subset\mathcal{N}\\|H|=n-t}}\max_{C_{H^c}}\frac{1}{l}\sum_{t=1}^{l}d(X_t,\hat{X}_t)\right] &\le (D+\delta)(1-\Pr(E))+d_{max}\Pr(E)\\ &\le D+\delta+d_{max}\binom{n}{2t}2^{-l((n-2t)\epsilon-\alpha)}.\end{aligned}$$

The right hand side can be made smaller than $D+\epsilon$ by letting $l\to\infty$ and then $\alpha,\delta\to 0$. Thus there exists a code that achieves $(R_1+\epsilon,\dots,R_n+\epsilon,D+\epsilon)$. □

4.3 Converse for Uniform Binary Sources with Hamming Distortion

In this section, we prove that the achievability scheme in Section 4.2 is optimal for a uniform binary source with Hamming distortion. For Hamming distortion, given the binary source alphabet $\mathcal{X}=\{+,-\}$, the reconstruction space $\hat{\mathcal{X}}=\mathcal{X}=\{+,-\}$. The Hamming distortion measure $d:\mathcal{X}\times\mathcal{X}\to\{0,1\}$ is given by

$$d(x,\hat{x})=\begin{cases}0 & \text{if } \hat{x}=x\\ 1 & \text{otherwise.}\end{cases}$$

Theorem 15. *For a uniform binary source and Hamming distortion measure, $\mathcal{RD}^*=\overline{\mathcal{RD}}$.*

Given a target distortion D for some rate vector in $\mathcal{RD}^*$, the Shannon rate-distortion function for a uniform binary source with Hamming distortion is given by $R(D)=1-h(D)$, where $h(\cdot)$ is the binary entropy function. Therefore, in order to prove Theorem 15, we need to show that for any subset of encoders $\mathcal{S}$ of size $n-2t$, $D\ge h^{-1}(1-\sum_{i\in\mathcal{S}}R_i)$. Before proving Theorem 15, however, we shall state and prove a lemma which provides an upper bound on the number of binary strings of length ℓ such that any two strings differ in at most $2\ell\delta$ places, where $\delta < 1/2$. In proving this lemma, we shall make use of a result of Kleitman [66], earlier conjectured by Paul Erdős: the number of binary strings of length ℓ such that any two strings differ in at most $2k$ places is at most $\sum_{i=0}^{k} \binom{\ell}{i}$.

Lemma 3. *For any set $\mathcal{S}$ of binary strings of length ℓ, where $|\mathcal{S}| \geq 2$, there exists a pair of strings in $\mathcal{S}$ that differ in at least $\left(2\ell h^{-1}\left(\frac{1}{\ell}\log(|\mathcal{S}|-1)\right) - 1\right)$ places, where h^{-1} is the inverse of the binary entropy function.*

Proof. Let $\delta = h^{-1}\left(\frac{1}{\ell}\log(|\mathcal{S}|-1)\right)$. Thus $\delta < 1/2$ and $|\mathcal{S}| - 1 = 2^{\ell h(\delta)}$. We have

$$\begin{aligned}
|\mathcal{S}| &> |\mathcal{S}| - 1 \\
&= (|\mathcal{S}| - 1)(\delta + 1 - \delta)^{\ell} \\
&= (|\mathcal{S}| - 1)\sum_{i=0}^{\ell}\binom{\ell}{i}\delta^{i}(1-\delta)^{\ell-i} \\
&\geq (|\mathcal{S}| - 1)\sum_{i=0}^{\lfloor \ell\delta \rfloor}\binom{\ell}{i}\delta^{i}(1-\delta)^{\ell-i} \\
&= (|\mathcal{S}| - 1)\sum_{i=0}^{\lfloor \ell\delta \rfloor}\binom{\ell}{i}(1-\delta)^{\ell}\left(\frac{\delta}{1-\delta}\right)^{i} \\
&\geq (|\mathcal{S}| - 1)\sum_{i=0}^{\lfloor \ell\delta \rfloor}\binom{\ell}{i}(1-\delta)^{\ell}\left(\frac{\delta}{1-\delta}\right)^{\ell\delta} \\
&= (|\mathcal{S}| - 1)\sum_{i=0}^{\lfloor \ell\delta \rfloor}\binom{\ell}{i}\delta^{\ell\delta}(1-\delta)^{\ell(1-\delta)} \\
&= (|\mathcal{S}| - 1)\sum_{i=0}^{\lfloor \ell\delta \rfloor}\binom{\ell}{i}2^{-\ell h(\delta)} \\
&= \sum_{i=0}^{\lfloor \ell\delta \rfloor}\binom{\ell}{i}.
\end{aligned}$$

By the aforementioned result in [66], $\sum_{i=0}^{\lfloor \ell\delta \rfloor} \binom{\ell}{i}$ is the maximum number of binary strings such that any two strings differ in at most $2\lfloor \ell\delta \rfloor$ places. Since $|\mathcal{S}| > \sum_{i=0}^{\lfloor \ell\delta \rfloor} \binom{\ell}{i}$, there must exist a pair of strings in $\mathcal{S}$ that differ in at least $2\lfloor \ell\delta \rfloor + 1 \geq 2(\ell\delta - 1) + 1 = 2\ell\delta - 1$ places. □

We are now in a position to prove Theorem 15. Let $(f_1, \ldots, f_n, g)$ be a code that achieves the rate-distortion vector $(R_1, \ldots, R_n, D)$, and let $\mathcal{S} = \{1, \ldots, n - 2t\}$. Define, for any given source sequence x^l, the set $\mathcal{M}(x^l) = \{\tilde{x}^l \in \mathcal{X}^l : f_i(\tilde{x}^l) = f_i(x^l)\ \forall i \in \mathcal{S}\}$. Let $M_{c_\mathcal{S}}$, $c_\mathcal{S} \in \{1, \ldots, 2^{l \sum_{i\in\mathcal{S}} R_i}\}$ be the values taken by the set $\mathcal{M}(X^l)$. Thus $M_{c_\mathcal{S}}$ is the pre-image of the set of codewords $c_\mathcal{S}$. Since there are $2^{l \sum_{i\in\mathcal{S}} R_i}$ sets of codewords covering 2^l sequences, we have

$$\frac{1}{2^{l \sum_{i\in\mathcal{S}} R_i}} \sum_{c_\mathcal{S}=1}^{2^{l \sum_{i\in\mathcal{S}} R_i}} |M_{c_\mathcal{S}}| = \frac{2^l}{2^{l \sum_{i\in\mathcal{S}} R_i}} = 2^{l(1-\sum_{i\in\mathcal{S}} R_i)}.$$

Suppose that the set $\mathcal{S}$ contains honest encoders only, and the traitors constitute either the set of encoders $\mathcal{T}_1 = \{n - 2t + 1, \ldots, n - t\}$ or the set $\mathcal{T}_2 = \{n - t + 1, \ldots, n\}$. Suppose further that $(x')^l$ is the observed source sequence, and the encoders in $\mathcal{S}$ send codewords $c'_\mathcal{S}$ to the decoder. Thus $(x')^l \in M_{c'_\mathcal{S}}$. Since there are $|M_{c'_\mathcal{S}}|$ sequences in $M_{c'_\mathcal{S}}$, Lemma 3 tells us that there exists a sequence $(x'')^l \in M_{c'_\mathcal{S}}$ such that $d((x')^l, (x'')^l) \geq 2l\delta - 2$ bits, where $\delta = h^{-1}(\frac{1}{l}\log|M_{c'_\mathcal{S}}| - 1)$. Suppose $\mathcal{T}_1$ is the honest set, and the encoders in $\mathcal{T}_1$ send the codewords corresponding to $(x')^l$. Then the set $\mathcal{T}_2$ of traitors could send codewords corresponding to the fake sequence $(x'')^l$. Thus the decoder would receive the set of messages $(c'_\mathcal{S}, c_{\mathcal{T}_1}(x'), c_{\mathcal{T}_2}(x''))$. Note, however, that the same set of messages would be received by the decoder if $(x'')^l$ were the true source sequence and $\mathcal{T}_1$ rather than $\mathcal{T}_2$ were the traitorous set, and the traitors decided to report $(x')^l$ to the decoder. In either case, the decoder must output the same reconstruction, say $\bar{x}^l$, since it cannot distinguish between the two cases. We thus have the following sequence of inequalities:

$$
\begin{aligned}
&\sum_{x^l \in M_{c'_{\mathcal{S}}}} \max_{i=1,2} \max_{C_{\mathcal{T}_i}} \sum_{t=1}^{l} d(x_t, \hat{x}_t) \\
&\geq \max_{C_{\mathcal{T}_2}} \sum_{t=1}^{l} d(x'_t, \hat{x'}_t) + \max_{C_{\mathcal{T}_1}} \sum_{t=1}^{l} d(x''_t, \hat{x}''_t) + \sum_{\substack{x^l \in M_{c'_{\mathcal{S}}} \\ x^l \neq (x')^l, (x'')^l}} \max_{i=1,2} \max_{C_{\mathcal{T}_i}} \sum_{t=1}^{l} d(X_t, \hat{X}_t) \\
&\geq \sum_{t=1}^{l} d(x'_t, \bar{x}_t) + \sum_{t=1}^{l} d(x''_t, \bar{x}_t) + \sum_{\substack{x^l \in M_{c'_{\mathcal{S}}} \\ x^l \neq (x')^l, (x'')^l}} \max_{i=1,2} \max_{C_{\mathcal{T}_i}} \frac{1}{l} \sum_{t=1}^{l} d(X_t, \hat{X}_t) \\
&\geq \sum_{t=1}^{l} d(x'_t, x''_t) + \sum_{\substack{x^l \in M_{c'_{\mathcal{S}}} \\ x^l \neq (x')^l, (x'')^l}} \max_{i=1,2} \max_{C_{\mathcal{T}_i}} \frac{1}{l} \sum_{t=1}^{l} d(X_t, \hat{X}_t) \\
&\geq 2lh^{-1} \left(\frac{1}{l} \log |M_{c'_{\mathcal{S}}}| - 1 \right) - 2 + \sum_{\substack{x^l \in M_{c'_{\mathcal{S}}} \\ x^l \neq (x')^l, (x'')^l}} \max_{i=1,2} \max_{C_{\mathcal{T}_i}} \frac{1}{l} \sum_{t=1}^{l} d(X_t, \hat{X}_t),
\end{aligned}
$$

where the penultimate inequality follows from the triangle inequality. We can now remove $(x')^l$ and $(x'')^l$ from $M_{c'_{\mathcal{S}}}$ and apply Lemma 1 to the remaining $|M_{c'_{\mathcal{S}}}| - 2$ sequences. We can do this iteratively, stopping when 3 or fewer sequences remain. This yields the lower bound

$$
\begin{aligned}
\sum_{x^l \in M_{c'_{\mathcal{S}}}} \max_{i=1,2} \max_{C_{\mathcal{T}_i}} \sum_{t=1}^{l} d(x_t, \hat{x}_t) &\geq \sum_{k=0}^{\left\lfloor \frac{|M_{c'_{\mathcal{S}}}|}{2} \right\rfloor - 1} \left(2lh^{-1} \left(\frac{1}{l} \log(|M_{c'_{\mathcal{S}}}| - 1 - 2k) \right) - 2 \right) \\
&\geq \sum_{j=0}^{|M_{c'_{\mathcal{S}}}| - 2} \left(lh^{-1} \left(\frac{1}{l} \log(|M_{c'_{\mathcal{S}}}| - 1 - j) \right) - 1 \right).
\end{aligned}
$$

Let $N = \sum_{c_{\mathcal{S}}=1}^{2^{l \sum_{i \in \mathcal{S}} R_i}} (|M_{c_{\mathcal{S}}}| - 1) = 2^l - 2^{l \sum_{i \in \mathcal{S}} R_i}$. Now

$$
\mathbf{E}_X \left[\max_{H} \max_{C_{H^c}} \frac{1}{l} \sum_{t=1}^{l} d(X_t, \hat{X}_t) \right]
$$

$$
\begin{aligned}
&\geq \mathbf{E}_X \left[\max_{i=1,2} \max_{C_{\mathcal{T}_i}} \frac{1}{l} \sum_{t=1}^{l} d(X_t, \hat{X}_t) \right] \\
&= \sum_{c_{\mathcal{S}}=1}^{2^{l \sum_{i \in \mathcal{S}} R_i}} \sum_{x^l \in M_{c_{\mathcal{S}}}} \max_{i=1,2} \max_{C_{\mathcal{T}_i}} \frac{1}{l} \sum_{t=1}^{l} d(x_t, \hat{x}_t) p(x^l) \\
&\geq \sum_{c_{\mathcal{S}}=1}^{2^{l \sum_{i \in \mathcal{S}} R_i}} \sum_{j=0}^{|M_{c_{\mathcal{S}}}|-2} \frac{1}{l} \cdot \left(l h^{-1} \left(\frac{1}{l} \log(|M_{c_{\mathcal{S}}}| - 1 - j) \right) - 1 \right) \cdot 2^{-l} \cdot \frac{N}{N} \\
&\overset{(a)}{\geq} h^{-1} \left(\frac{1}{lN} \sum_{c_{\mathcal{S}}=1}^{2^{l \sum_{i \in \mathcal{S}} R_i}} \sum_{j=0}^{|M_{c_{\mathcal{S}}}|-2} \log(|M_{c_{\mathcal{S}}}| - 1 - j) \right) 2^{-l} N - \sum_{c_{\mathcal{S}}=1}^{2^{l \sum_{i \in \mathcal{S}} R_i}} \sum_{j=0}^{|M_{c_{\mathcal{S}}}|-2} \frac{1}{l} \cdot 2^{-l} \\
&\overset{(b)}{\geq} h^{-1} \left(\frac{1}{l} \sum_{c_{\mathcal{S}}=1}^{2^{l \sum_{i \in \mathcal{S}} R_i}} \left(\frac{(|M_{c_{\mathcal{S}}}| - 1) \ln(|M_{c_{\mathcal{S}}}| - 1) - (|M_{c_{\mathcal{S}}}| - 1)}{\ln 2} \right) \cdot \frac{1}{N} \right) 2^{-l} N \\
&\qquad - \frac{1}{l} 2^{-l} N \\
&= h^{-1} \left(\frac{1}{lN} \left(\sum_{c_{\mathcal{S}}=1}^{2^{l \sum_{i \in \mathcal{S}} R_i}} (|M_{c_{\mathcal{S}}}| - 1) \log(|M_{c_{\mathcal{S}}}| - 1) - \sum_{c_{\mathcal{S}}=1}^{2^{l \sum_{i \in \mathcal{S}} R_i}} \frac{|M_{c_{\mathcal{S}}}| - 1}{\ln 2} \right) \right) 2^{-l} N \\
&\qquad - \frac{1}{l} 2^{-l} N \\
&\geq h^{-1} \left(\frac{1}{lN} \sum_{c_{\mathcal{S}}=1}^{2^{l \sum_{i \in \mathcal{S}} R_i}} (|M_{c_{\mathcal{S}}}| - 1) \log(|M_{c_{\mathcal{S}}}| - 1) - \frac{1}{l \ln 2} \right) 2^{-l} N - \frac{1}{l} \\
&\overset{(c)}{\geq} h^{-1} \Bigg(\frac{1}{lN} \cdot 2^{l \sum_{i \in \mathcal{S}} R_i} \left(\frac{1}{2^{l \sum_{i \in \mathcal{S}} R_i}} \sum_{c_{\mathcal{S}}=1}^{2^{l \sum_{i \in \mathcal{S}} R_i}} (|M_{c_{\mathcal{S}}}| - 1) \right) \\
&\qquad \cdot \log \left(\frac{1}{2^{l \sum_{i \in \mathcal{S}} R_i}} \sum_{c_{\mathcal{S}}=1}^{2^{l \sum_{i \in \mathcal{S}} R_i}} (|M_{c_{\mathcal{S}}}| - 1) \right) - \frac{1}{l \ln 2} \Bigg) 2^{-l} N - \frac{1}{l} \\
&= h^{-1} \left(\frac{1}{lN} 2^{l \sum_{i \in \mathcal{S}} R_i} \frac{N}{2^{l \sum_{i \in \mathcal{S}} R_i}} \log \left(\frac{N}{2^{l \sum_{i \in \mathcal{S}} R_i}} \right) - \frac{1}{l \ln 2} \right) 2^{-l} N - \frac{1}{l} \\
&= h^{-1} \left(\frac{1}{l} \log(2^{l(1 - \sum_{i \in \mathcal{S}} R_i)} - 1) - \frac{1}{l \ln 2} \right) (1 - 2^{-l(1 - \sum_{i \in \mathcal{S}} R_i)}) - \frac{1}{l},
\end{aligned}
$$

where (a) follows from the convexity of $h^{-1}(x)$ in x, (b) follows from the fact that $\sum_{i=1}^{m} \ln x \geq m \ln m - m$ and because $h^{-1}(x)$ is nondecreasing in x, and (c) follows from the convexity of $x \log x$ in x and because $h^{-1}(x)$ is nondecreasing in x. Letting $l \to \infty$ completes the proof.

4.4 Converse for Gaussian Sources with Squared Error Distortion

In this section, we prove that the achievability scheme in Section 4.2 is optimal for a Gaussian source with squared error distortion. The squared error distortion measure $d : \mathbf{R} \times \mathbf{R} \to \mathbf{R}^+$ is given by $d(x, \hat{x}) = (x - \hat{x})^2$.

Theorem 16. *For a Gaussian source and squared error distortion measure, if there exists a reconstruction symbol $\hat{X}$ such that $\mathbf{E}[d(X, \hat{X})]$ is finite, then $\mathcal{RD}^* = \overline{\mathcal{RD}}$.*

Given a target distortion D, the Shannon rate-distortion function for a Gaussian source with squared error distortion is given by $R(D) = \frac{1}{2}\log \sigma^2/D$, where σ^2 is the variance of the source. Therefore, in order to prove Theorem 16, we need to show that for any subset of encoders $\mathcal{S}$ of size $n - 2t$, $\sum_{i \in \mathcal{S}} R_i \geq \frac{1}{2}\log \sigma^2/D$. Let $(f_1, \dots, f_n, g)$ be a code that achieves the rate-distortion vector $(R_1, \dots, R_n, D)$, and let C_i be the codeword transmitted by the i^{th} encoder. For any $\mathcal{S} \in \mathcal{N}$, $|\mathcal{S}| = n - 2t$, we have

$$\begin{aligned}
\sum_{i \in \mathcal{S}} R_i &\geq \frac{1}{l}\sum_{i \in \mathcal{S}} H(C_i) \\
&\geq \frac{1}{l} H(C_{\mathcal{S}}) \\
&\geq \frac{1}{l} I(X^l; C_{\mathcal{S}}) \\
&= \frac{1}{l} h(X^l) - \frac{1}{l} h(X^l | C_{\mathcal{S}}) \\
&= \frac{1}{2}\log 2\pi e \sigma^2 - \frac{1}{l} h(X^l | C_{\mathcal{S}}).
\end{aligned}$$

Thus, in order to prove Theorem 16, it suffices to show that $\frac{1}{l}h(X^l|C_{\mathcal{S}}) \leq \frac{1}{2}\log 2\pi eD$. Let H_1 and H_2 be two sets in $\mathcal{N}$ such that $|H_1| = |H_2| = n - t$ and $H_1 \cap H_2 = \mathcal{S}$. Define

$$Q_D = \{x^l : \max_{i=1,2} \max_{C_{H_i^c}} \frac{1}{l} \sum_{t=1}^{l} d(X_t, \hat{X}_t) \leq D\}.$$

For every codeword $c_{\mathcal{S}}$, let $Q_{D|c_{\mathcal{S}}} = \{x^l \in Q_D : f_{\mathcal{S}} = c_{\mathcal{S}}\}$. Then

$$\begin{aligned}
\Pr(X^l \in Q_{D'}) &= \sum_{c_{\mathcal{S}}} \Pr(C_{\mathcal{S}} = c_{\mathcal{S}}) \Pr(X^l \in Q_{D'}|C_{\mathcal{S}} = c_{\mathcal{S}}) \\
&\leq \sum_{c_{\mathcal{S}}} \Pr(C_{\mathcal{S}} = c_{\mathcal{S}}) \Pr(X^l \in Q_{D'|c_{\mathcal{S}}}|C_{\mathcal{S}} = c_{\mathcal{S}}) \\
&= \Pr(X^l \in Q_{D'|C_{\mathcal{S}}}). \qquad (4.1)
\end{aligned}$$

Since the code achieves distortion D, we have

$$\begin{aligned}
D &\geq \mathbf{E}\left[\max_{\substack{H \subset \mathcal{N} \\ |H|=n-t}} \max_{C_{H^c}} \frac{1}{l} \sum_{t=1}^{l} d(X_t, \hat{X}_t)\right] \\
&\geq \mathbf{E}\left[\max_{i=1,2} \max_{C_{H_i^c}} \frac{1}{l} \sum_{t=1}^{l} d(X_t, \hat{X}_t)\right] \\
&= \int_0^\infty \Pr\left[\max_{i=1,2} \max_{C_{H_i^c}} \frac{1}{l} \sum_{t=1}^{l} d(X_t, \hat{X}_t) > D'\right] dD' \\
&\geq \int_0^\infty \Pr\left[X^l \notin Q_{D'}\right] dD' \\
&\geq \int_0^\infty \Pr\left[X^l \notin Q_{D'|C_{\mathcal{S}}}\right] dD', \qquad (4.2)
\end{aligned}$$

where the last inequality follows from (4.1).

Let $\tilde{D} = \inf\{D' : X^l \in Q_{D'|C_{\mathcal{S}}}\}$. Since the set $Q_{D'|c_{\mathcal{S}}}$ is non-decreasing in D', the event $\{X^l \notin Q_{D'|C_{\mathcal{S}}}\}$ is identical to the event $\{\tilde{D} > D'\}$. Hence from (4.2),

$$D \geq \int_0^\infty \Pr(\tilde{D} > D')dD' = \mathbf{E}(\tilde{D}).$$

Fix $\Delta > 0$ and let $\tilde{D}_\Delta = \Delta\lceil \frac{\tilde{D}}{\Delta} \rceil$ be a quantized version of $\tilde{D}$. Observe that since $\tilde{D}_\Delta \leq \tilde{D} + \Delta$,

$$\mathbf{E}(\tilde{D}_\Delta) \leq \mathbf{E}(\tilde{D} + \Delta) \leq D + \Delta. \tag{4.3}$$

We then have

$$\begin{aligned}\frac{1}{l}h(X^l|C_\mathcal{S}) &= \frac{1}{l}h(X^l|C_\mathcal{S}, \tilde{D}_\Delta) + \frac{1}{l}I(X^l;\tilde{D}_\Delta|C_\mathcal{S}) \\ &\leq \frac{1}{l}h(X^l|C_\mathcal{S}, \tilde{D}_\Delta) + \frac{1}{l}H(\tilde{D}_\Delta).\end{aligned} \tag{4.4}$$

Consider the first term in (4.4). Note that $\tilde{D} \leq \tilde{D}_\Delta$, so $X^l \in Q_{\tilde{D}_\Delta|C_\mathcal{S}}$. Therefore, by the uniform bound on entropy,

$$h(X^l|C_\mathcal{S}, \tilde{D}_\Delta) \leq \mathbf{E}[\log \mathrm{Vol}(Q_{\tilde{D}_\Delta}|C_\mathcal{S})]. \tag{4.5}$$

Now consider the second term in (4.4). Since $\tilde{D}_\Delta$ is quantized, it can be shown using a maximum entropy distribution result that

$$H(\tilde{D}_\Delta) \leq \frac{\mathbf{E}(\tilde{D}_\Delta)}{\Delta} h\left(\frac{\Delta}{\mathbf{E}(\tilde{D}_\Delta)}\right), \tag{4.6}$$

where $h(q) = -q\log q - (1-q)\log(1-q)$ is the binary entropy function. The right hand side of (4.6) is increasing in $\mathbf{E}(\tilde{D}_\Delta)$, so using (4.3) gives

$$H(\tilde{D}_\Delta) \leq \frac{D+\Delta}{\Delta} h\left(\frac{\Delta}{D+\Delta}\right). \tag{4.7}$$

Now consider two sequences x^l, $x'^l \in Q_{D'|c_\mathcal{S}}$. Suppose the decoder receives the set of codewords $(c_\mathcal{S}, c_{H_1\backslash H_2} = f_{H_1\backslash H_2}(x^l), c_{H_2\backslash H_1} = f_{H_2\backslash H_1}(x'^l))$. First observe that this set of messages could have been produced if $X^l = x^l$ and H_1 were the set of honest encoders. Then the nodes in $H_2\backslash H_1$, which are all traitors, could send $c_{H_2\backslash H_1}$. Since $x^l \in Q_{D'|c_\mathcal{S}}$, the estimate $\hat{x}^l$ must by definition satisfy $\frac{1}{l}d(x^l, \hat{x}^l) \leq D'$. However, the same set of messages could have been produced if $X^l = x'^l$ and H_2 were the set of honest encoders, and the traitors $H_1\backslash H_2$ decide to send $c_{H_1 \setminus H_2}$. Since the decoder produces just one estimate for a given set of received codewords, the very same estimate $\hat{x}^l$, by the same reasoning, must satisfy $\frac{1}{l} d(x'^l, \hat{x}^l) \le D'$. Hence we have

$$\frac{1}{l}\sum_{t=1}^{l}(x(t)-\hat{x}(t))^2 \le D'$$
$$\frac{1}{l}\sum_{t=1}^{l}(x'(t)-\hat{x}(t))^2 \le D',$$

which can be rewritten as

$$||x-\hat{x}||_2 \le \sqrt{lD'}$$
$$||x'-\hat{x}||_2 \le \sqrt{lD'}.$$

Therefore, by the triangle inequality, for any $x^l, x'^l \in Q_{D'|c_{\mathcal{S}}}$, $||x - x'||_2 \le 2\sqrt{lD'}$. Thus $Q_{D'|c_{\mathcal{S}}}$ has diameter at most $2\sqrt{lD'}$. The following lemma from [67] upper bounds the volume of subsets of $\mathbf{R}^l$ as a function of their diameter.

Lemma 4. *The volume of any subset of $\mathbf{R}^l$ is no more than that of the l-ball with the same diameter.*

Lemma 4 tells us that the volume of $Q_{D'|c_{\mathcal{S}}}$ is no more than the volume of an l-ball with radius $\sqrt{lD'}$. The latter can be shown to be less than $(2\pi e D')^{\frac{l}{2}}$. Combining this with (4.5) and (4.7) gives

$$\begin{aligned}\frac{1}{l}h(X^l|C_{\mathcal{S}}) &\le \frac{1}{l}\mathbf{E}[\log(2\pi e\tilde{D}_\Delta)^{\frac{l}{2}}] + \frac{1}{l}\frac{D+\Delta}{\Delta}h\left(\frac{\Delta}{D+\Delta}\right)\\ &\le \frac{1}{2}\log(2\pi e\mathbf{E}[\tilde{D}_\Delta]) + \frac{1}{l}\frac{D+\Delta}{\Delta}h\left(\frac{\Delta}{D+\Delta}\right)\\ &\le \frac{1}{2}\log(2\pi e(D+\Delta)) + \frac{1}{l}\frac{D+\Delta}{\Delta}h\left(\frac{\Delta}{D+\Delta}\right),\end{aligned}$$

where the penultimate inequality follows from the concavity of $\log x$ in x and the last inequality follows from (4.3). Letting $l \to \infty$ and then $\Delta \to 0$ completes the proof.

4.5 Uniform Binary Sources with Erasure Distortion

In this section, we show that for uniform binary sources with erasure distortion, the factor-of-2 rule is pessimistic, and there exists a coding scheme which can achieve points outside the rate region proposed in Section 4.1. We will consider a special case of the 3-channel Byzantine multiple descriptions problem in which one of the channels transmits at rate R and the other two transmit at rate 1. One of the three encoders is a traitor. We shall henceforth refer to this special case as the $R - 1 - 1$ problem. Assume without loss of generality that Encoder 1 transmits at rate R and Encoders 2 and 3 transmit at rate 1. Thus Encoders 2 and 3 send the complete source sequence X^l to the decoder, since their respective channels are not rate-constrained. Given the source alphabet $\mathcal{X} = \{+, -\}$, define the reconstruction space $\hat{\mathcal{X}} = \{+, -, 0\}$, where 0 denotes the erasure symbol. The erasure distortion measure is given by

$$d(x, \hat{x}) = \begin{cases} 0 & \text{if } \hat{x} = x \\ 1 & \text{if } \hat{x} = 0 \\ \infty & \text{otherwise.} \end{cases} \tag{4.8}$$

Let $\mathcal{RD}_{eras}$ be the set of achievable rate-distortion pairs as defined in Section 4.1, and let $\overline{\mathcal{RD}}_{eras}$ denote its closure.

Theorem 17. *$\overline{\mathcal{RD}}_{eras} = \{(R, D) : D \geq 2h^{-1}(1 - R)\}$, where $h^{-1}(\cdot)$ is the inverse of the binary entropy function h.*

Proof. (*Achievability*) Note that $2h^{-1}(1 - R) < 1 - R$, which is the distortion-rate function for a uniform binary source with erasure distortion. We will show that for any D and any $R > 1 - h(D/2)$ (equivalently $D > 2h^{-1}(1 - R)$), the rate-distortion pair (R, D) is achievable. In particular, we will show that for any $\epsilon > 0$, there exists a code with rate less than $R + \epsilon$ and distortion less than $D + \epsilon$. Define $\tilde{D} = D/2$, and let $R > 1 - h(\tilde{D})$. Let $p(\tilde{x}|x)$ denote a binary symmetric channel (BSC) with crossover probability $\tilde{D}$. We construct an encoder similar to a rate-distortion encoder for a binary symmetric source (BSS) with Hamming distortion.

Random codebook generation: Compute $p(\tilde{x}) = \sum_x p(x)p(\tilde{x}|x)$. Fix a block-length l, and generate $2^{lR} + 1$ sequences $\tilde{X}^l$ drawn randomly and *i.i.d* from the marginal distribution $p(\tilde{x})$. Assign each codeword an index $w \in \{0, 1, \ldots, 2^{lR}\}$. The codebook is revealed to the encoders and the decoder.

Encoding: Choose $\delta > 0$. Encoder 1 observes a length-l source sequence X^l, and encodes X^l by w, $w \neq 0$, if X^l and $\tilde{X}^l(w)$ are jointly typical, *i.e.*, the Hamming distance between X^l and $\tilde{X}^l(w)$ is less than $l(\tilde{D} + \delta)$. If there is more than one such w, the smallest is used. If there is no such $w \in \{1, \ldots, 2^{lR}\}$, Encoder 1 sends $w = 0$. Since $lR + 1$ bits are required to describe the $2^{lR} + 1$ indices, the rate of this code is $R + 1/l$. Encoders 2 and 3 send the whole sequence X^l.

Decoding: If Encoders 2 and 3 send the same source sequence X^l, then X^l is the true source sequence, since at least one of Encoders 2 and 3 is honest. Output $\hat{X}^l = X^l$ as the reconstruction. If Encoders 2 and 3 send different source sequences, then one of them is the traitor, and Encoder 1 is honest. In this case, if Encoder 1 sent $w = 0$, output the all erasure string. Otherwise, if one of the sequences sent by Encoders 2 and 3 is not jointly typical with the index sent by Encoder 1, then that encoder is the traitor. Output the sequence sent by the other encoder as the reconstruction. If, however, Encoders' 2 and 3 sequences are both jointly typical with the index sent by Encoder 1, output a reconstruction $\hat{X}^l$ such that $\hat{X}^l$ has the same value for the bits for which Encoders' 2 and 3 sequences agree, and has erasures for the bits for which Encoders' 2 and 3 sequences differ.

Distortion analysis: Let us consider the possible traitor locations and actions for each source sequence. If the traitor chooses to take over Encoder 1, then Encoders 2 and 3 will send the true source sequence and the decoder will be able to decode correctly regardless of the source sequence. Suppose the traitor takes over one of Encoder 2 or Encoder 3. Assume without loss of generality that the traitor takes over Encoder 3. Then Encoder 2 will send the true source sequence, and Encoder 1 will send an index that is jointly typical with the true source sequence. If the traitor sends the true source sequence, the decoder will be able to decode correctly, so suppose the traitor chooses to send a spurious sequence. If the fake sequence sent by the traitor is not jointly typical with the index sent by Encoder 1, the decoder will output the sequence sent by Encoder 2, which is the true source sequence. Suppose now that the traitor sends a source sequence which is jointly typical with the index sent by Encoder 1, but different from the true source sequence. Then the decoder will output a partially erased reconstruction based on the bits that are common between the true sequence and the fake sequence. Thus, for any source sequence for which Encoder 1 transmits an index $w \neq 0$, the only strategy for the traitor which will yield non-zero distortion is to send a source sequence which is jointly typical with the index sent by Encoder 1, but different from the true source sequence. It therefore makes sense for the traitor to pursue this strategy for every source sequence. In this case, let X_2^l and X_3^l be the true and fake sequences respectively. Since both X_2^l and X_3^l are jointly typical with $\tilde{X}^l(w)$, where w is the index sent by Encoder 1, the Hamming distance between $\tilde{X}^l(w)$ and X_2^l (and $\tilde{X}^l(w)$ and X_3^l) is at most $l(\tilde{D} + \delta)$. By the triangle inequality, therefore, the Hamming distance between the true sequence X_2^l and the fake sequence X_3^l is at most $2l(\tilde{D}+\delta)$.

Let $P_w = \{x^l \in \mathcal{X}^l :$ the encoder transmits the index $w\}$. We thus have

$$\begin{aligned}
&\mathbf{E}_X\left[\max_{\alpha\in\{1,2,3\}}\max_{C_\alpha}\frac{1}{l}\sum_{t=1}^{l} d(X_t,\hat{X}_t)\right]\\
&=\sum_{x^l\in\mathcal{X}^l}\left[\max_{\alpha\in\{1,2,3\}}\max_{C_\alpha}\frac{1}{l}\sum_{t=1}^{l} d(x_t,\hat{x}_t)\right]p(x^l)\\
&=\sum_{w=0}^{2^{lR}}\sum_{x^l\in P_w}\left[\max_{\alpha\in\{1,2,3\}}\max_{C_\alpha}\frac{1}{l}\sum_{t=1}^{l} d(x_t,\hat{x}_t)\right]p(x^l)\\
&=\sum_{x^l\in P_0}\left[\max_{\alpha\in\{1,2,3\}}\max_{C_\alpha}\frac{1}{l}\sum_{t=1}^{l} d(x_t,\hat{x}_t)\right]p(x^l)\\
&\qquad\qquad+\sum_{w=1}^{2^{lR}}\sum_{x^l\in P_w}\left[\max_{\alpha\in\{1,2,3\}}\max_{C_\alpha}\frac{1}{l}\sum_{t=1}^{l} d(x_t,\hat{x}_t)\right]p(x^l)\\
&\le\sum_{x^l\in P_0}1\cdot p(x^l)+\sum_{w=1}^{2^{lR}}\sum_{x^l\in P_w}2(\tilde{D}+\delta)p(x^l),
\end{aligned}$$

since, when $w=0$, the decoder outputs the all-erasure string (which yields distortion 1), and when $w\neq 0$, the traitor sends a fake sequence X_3^l which differs in at most $2l(\tilde{D}+\delta)$ bits from the true source sequence, as described earlier. Thus the distortion when $w\neq 0$ is at most $2(\tilde{D}+\delta)$. Recall that the set P_0 (the set of sequences for which the encoder transmits $w=0$) is the set of sequences for which no typical codeword can be found. Let P_e be the total probability of these sequences. The total probability of the rest of the sequences can be bounded by 1. We thus have

$$\begin{aligned}
\mathbf{E}_X\left[\max_{\alpha\in\{1,2,3\}}\max_{C_\alpha}\frac{1}{l}\sum_{t=1}^{l} d(X_t,\hat{X}_t)\right] &\le\sum_{x^l\in P_0}1\cdot p(x^l)+\sum_{w=1}^{2^{lR}}\sum_{x^l\in P_w}2(\tilde{D}+\delta)p(x^l)\\
&\le 1\cdot P_e+2(\tilde{D}+\delta)\cdot 1\\
&=2\tilde{D}+2\delta+P_e.
\end{aligned}$$

Since we are using a rate-distortion code of rate $R>R(\tilde{D})=1-h(\tilde{D})$ for a BSS with Hamming distortion, P_e, averaged over a random choice of codebooks, can be made arbitrarily small as $l \to \infty$. Therefore, there exists a sufficiently large blocklength l such that $1/l < \epsilon$ and $2\delta + P_e < \epsilon$. Thus there exists a code with rate $R + 1/l < R + \epsilon$ and average distortion less than $2\tilde{D} + \epsilon = D + \epsilon$. □

We will now prove the converse to Theorem 17.

Proof. (*Converse to Theorem 17*) Let (f, g) be a code that achieves the rate-distortion pair (R, D). For this code, define, for any given source sequence x^l, the set $\mathcal{M}(x^l) = \{\tilde{x}^l \in \mathcal{X}^l : f(\tilde{x}^l) = f(x^l)\}$. Let M_i, $i \in \{1, \ldots, 2^{lR}\}$ be the values taken by the set $\mathcal{M}(X^l)$. Thus M_i is the pre-image of the i^{th} codeword. Since there are 2^{lR} codewords covering 2^l sequences, we have

$$\frac{1}{2^{lR}} \sum_{i=1}^{2^{lR}} |M_i| = \frac{2^l}{2^{lR}} = 2^{l(1-R)}.$$

Suppose Encoder 3 is the traitor, and suppose that unless the pre-image of the codeword sent by Encoder 1 contains a single source sequence (in which case Encoder 3 sends that source sequence), Encoder 3 always sends a fake source sequence $\tilde{X}^l$ which is in the pre-image of the codeword sent by Encoder 1 but is different from the sequence X^l sent by Encoder 2 (which is the true source sequence). The true and fake sequences will differ in at least one bit. The situation, from the point of view of the decoder, is identical to the situation where $\tilde{X}^l$ is the true source sequence, X^l is the fake source sequence, and Encoder 2, rather than Encoder 3, is the traitor. Since the distortion is maximized over all traitor locations and actions, the decoder cannot output either $+$ or $-$ for any bit in which X^l and $\tilde{X}^l$ differ, since outputting either would result in infinite distortion under one of the two aforementioned scenarios. The decoder, therefore, must output an erasure for any bit in which X^l and $\tilde{X}^l$ differ. Given a sequence $x^l \in M_i$, let $D_{M_i}(x^l)$ be the maximum Hamming distance of x^l from any sequence in M_i. In the extreme case that $|M_i| = 2^l$, *i.e.*, all sequences map to the same codeword, the traitor will be able to find a source sequence differing in l bits for every sequence. Thus $D_{M_i}(x^l) = l$ for all x^l, and the decoder would be forced to output the all erasure string for every source sequence, resulting in a distortion of 1.

Now suppose $|M_i| < 2^l$ for all i. Since there are $|M_i|$ source sequences in M_i, Lemma 3 tells us that the traitor will be able to find a sequence $x_w^l \in M_i$ such that $D_{M_i}(x_w^l) \geq 2l\delta - 2$ bits, where $\delta = h^{-1}(\frac{1}{l}\log|M_i| - 1)$. We can remove x_w^l from M_i and apply Lemma 1 to the remaining $|M_i| - 1$ sequences. Doing this iteratively yields the lower bound

$$\sum_{x^l \in M_i} D_{M_i}(x^l) \geq \sum_{j=0}^{|M_i|-2} \left(2lh^{-1}\left(\frac{1}{l}\log(|M_i| - 1 - j)\right) - 2\right).$$

Let $N = \sum_{i=1}^{2^{lR}}(|M_i| - 1) = 2^l - 2^{lR}$. Now

$$\begin{aligned}
&\mathbf{E}_X\left[\max_{\alpha\in\{1,2,3\}} \max_{C_\alpha} \frac{1}{l}\sum_{t=1}^{l} d(X_t, \hat{X}_t)\right] \\
&\geq \mathbf{E}_X\left[\max_{C_3} \frac{1}{l}\sum_{t=1}^{l} d(X_t, \hat{X}_t)\right] \\
&\geq \sum_{i=1}^{2^{lR}} \sum_{x^l\in M_i} \frac{1}{l} D_{M_i}(x^l)p(x^l) \\
&\geq \sum_{i=1}^{2^{lR}} \sum_{j=0}^{|M_i|-2} \frac{1}{l}\cdot\left(2lh^{-1}\left(\frac{1}{l}\log(|M_i| - 1 - j)\right) - 2\right)\cdot 2^{-l}\cdot\frac{N}{N} \\
&\overset{(a)}{\geq} 2h^{-1}\left(\frac{1}{l}\sum_{i=1}^{2^{lR}} \sum_{j=0}^{|M_i|-2} \log(|M_i| - 1 - j)\cdot\frac{1}{N}\right) 2^{-l}N - \sum_{i=1}^{2^{lR}} \sum_{j=0}^{|M_i|-2} \frac{2}{l}\cdot 2^{-l} \\
&\overset{(b)}{\geq} 2h^{-1}\left(\frac{1}{l}\sum_{i=1}^{2^{lR}}\left(\frac{(|M_i| - 1)\ln(|M_i| - 1) - (|M_i| - 1)}{\ln 2}\right)\cdot\frac{1}{N}\right) 2^{-l}N
\end{aligned}$$

$$- \frac{2}{l} \cdot 2^{-l} \sum_{i=1}^{2^{lR}} (|M_i| - 1)$$

$$= 2h^{-1}\left(\frac{1}{l}\left(\sum_{i=1}^{2^{lR}}(|M_i|-1)\log(|M_i|-1) - \sum_{i=1}^{2^{lR}}\frac{|M_i|-1}{\ln 2}\right)\cdot\frac{1}{N}\right)2^{-l}N$$

$$- \frac{2}{l} \cdot 2^{-l}N$$

$$\geq 2h^{-1}\left(\frac{1}{lN}\sum_{i=1}^{2^{lR}}(|M_i|-1)\log(|M_i|-1) - \frac{1}{l\ln 2}\right)2^{-l}N - \frac{2}{l}$$

$$\overset{(c)}{\geq} 2h^{-1}\left(\frac{2^{lR}}{lN}\left(\frac{1}{2^{lR}}\sum_{i=1}^{2^{lR}}(|M_i|-1)\right)\log\left(\frac{1}{2^{lR}}\sum_{i=1}^{2^{lR}}(|M_i|-1)\right) - \frac{1}{l\ln 2}\right)2^{-l}N$$

$$- \frac{2}{l}$$

$$= 2h^{-1}\left(\frac{1}{lN}\cdot 2^{lR}\cdot\frac{N}{2^{lR}}\log\frac{N}{2^{lR}} - \frac{1}{l\ln 2}\right)2^{-l}N - \frac{2}{l}$$

$$= 2h^{-1}\left(\frac{1}{l}\log(2^{l(1-R)}-1) - \frac{1}{l\ln 2}\right)(1-2^{-l(1-R)}) - \frac{2}{l},$$

where (a) follows from the convexity of $h^{-1}(x)$ in x, (b) follows from the fact that $\sum_{i=1}^{m}\ln x \geq m\ln m - m$ and because $h^{-1}(x)$ is nondecreasing in x, and (c) follows from the convexity of $x\log x$ in x and because $h^{-1}(x)$ is nondecreasing in x. Letting $l \to \infty$ completes the proof. □

4.6 Channel Coding Theorem

In this section we make precise the fact that separation breaks for erasure distortion. We begin by proving a capacity result for the channel depicted in Figure 4.1. A set of messages, indexed by $\{1, \ldots, 2^{lR}\}$ is to be transmitted over the channel. Encoder i encodes the message using the encoding function f_i :

$\{1,\ldots,2^{lR}\} \to \{1,\ldots,2^{lR_i}\}$. If $i \in H$, then Encoder i's codeword is $C_i = f_i(W)$, where W is the message to be transmitted. If $i \in H^c$, then Encoder i can choose C_i arbitrarily, with full knowledge of W and the other $n-1$ codewords. The decoders employs the decoding function $g : \prod_{i=1}^{n}\{1,\ldots,2^{lR_i}\} \to \{1,\ldots,2^{lR}\}$ produces an estimate $\hat{W} = g(C_1,\ldots,C_n)$ of the original message W.

For the message w, define the indicator function

$$1_{f,g}(w,H,C^{H^c}) = \begin{cases} 0 & \text{if } \hat{w} = w \\ 1 & \text{if } \hat{w} \neq w \end{cases}$$

given that the set of honest encoders is H and the traitors transmit the codewords C^{H^c}. For the code (f,g), we define the average probability of error as

$$P_e(f,g) = \frac{1}{2^{lR}} \sum_{w=1}^{2^{lR}} \max_{\substack{H \subset \mathcal{N} \\ H = |n-t|}} \max_{C^{H^c}} 1_{f,g}(w,H,C^{H^c}).$$

Definition 14. *A rate R is said to be achievable if there exists a sequence of codes (f,g), indexed by the blocklength l, such that $P_e(f,g) \to 0$ as $l \to \infty$.*

Theorem 18. *All rates R such that $R < \min_{\substack{S \subset \mathcal{N} \\ S = |n-2t|}} \sum_{i \in S} R_i$ are achievable. If $R > \min_{\substack{S \subset \mathcal{N} \\ S = |n-2t|}} \sum_{i \in S}$, then $P_e(f,g) \to 1$ as $l \to \infty$.*

The proof of achievability for Theorem 18 is very similar to the proof of Theorem 19 and is omitted. We prove the converse below. It is worth noting that Theorem 18 admits a strong converse.

Converse to Theorem 18. Suppose $R > \sum_{i=1}^{n-2t} R_i$ and let $\mathcal{T}_1 = \{n-2t+1,\ldots,n-t\}$ and $\mathcal{T}_2 = \{n-t+1,\ldots,n\}$. Fix a code $(f_1,\ldots,f_n,g)$ and consider the first $n-2t$ encoding functions $f_1,\ldots,f_{n-2t}$. Define $M(w)$ to be the set of source sequences such that for all $w' \in M(w)$, $f_i(w') = f_i(w)$, $i = 1,\ldots,n-2t$. Note that given a random message W, the range of the random variable $M(W)$ partitions the original set of messages $\{1, \ldots, 2^{lR}\}$. Since the sum rate of the first $n - 2t$ encoders is $\sum_{i=1}^{n-2t} R_i$, the number of partitions is $2^{l \sum_{i=1}^{n-2t} R_i}$. We can therefore label the partitions as $M_1, M_2, \ldots, M_{2^{l \sum_{i=1}^{n-2t} R_i}}$. For each partition M_i, we have the following two cases:

1. there exists $\tilde{w} \in M_i$ such that if the encoders in $\mathcal{T}_1$ transmit $f_i(\tilde{w})$ for all $i \in \mathcal{T}_1$, then the decoder outputs $\hat{W} = \tilde{w}$ regardless of the messages transmitted by encoders in $\mathcal{T}_2$. We refer to $\tilde{w}$ as a "leader" message.
2. for all $w \in M_i$, there exists a set of messages C_i, $i \in \mathcal{T}_2$, such that if the encoders in $\mathcal{T}_1$ transmit $f_i(\tilde{w})$ for all $i \in \mathcal{T}_1$ and the encoders in $\mathcal{T}_2$ transmit C_i, $i \in \mathcal{T}_2$, then the decoder outputs $\hat{W} \neq w$.

We now argue that for any message in M_i that is not a leader message, the traitors can cause the decoder to make an error by outputting a different message. If the first case holds, then the traitors simply have to take over the encoders in $\mathcal{T}_1$ and transmit $f_i(\tilde{w})$, $i \in \mathcal{T}_1$ where $\tilde{w}$ is the leader message. If Case 2 holds, then the traitors simply have to take over $\mathcal{T}_2$ and transmit the messages that would result in an error at the decoder. We can also argue that if there is more than one leader message in M_i, then the traitors can cause an error for every leader. More precisely, if $\tilde{w}$ and w' are two leader messages, then the traitors can cause an error for $\tilde{w}$ by taking over $\mathcal{T}_1$ and transmitting $f_i(w')$, $i \in \mathcal{T}_1$. If there is only one leader in M_i, then, by the above argument, the traitors can cause errors for all messages other than the leader. Therefore, at most one message for every partition can be decoded correctly. Since there are $2^{l \sum_{i=1}^{n-2t} R_i}$ partitions, at most $2^{l \sum_{i=1}^{n-2t} R_i}$ can be correctly decoded. We can therefore compute the probability of error as follows:

$$\begin{aligned}
P_e(f,g) &= \frac{1}{2^{lR}} \sum_{w=1}^{2^{lR}} \max_{\substack{H \subset \mathcal{N} \\ H=|n-t|}} \max_{C^{H^c}} 1_{f,g}(w, H, C^{H^c}) \\
&= \frac{1}{2^{lR}} \sum_{j=1}^{2^{l \sum_{i=1}^{n-2t} R_i}} \sum_{w \in M_j} \max_{\substack{H \subset \mathcal{N} \\ H=|n-t|}} \max_{C^{H^c}} 1_{f,g}(w, H, C^{H^c}) \\
&\geq \frac{1}{2^{lR}} (2^{lR} - 2^{l \sum_{i=1}^{n-2t} R_i}) \\
&= 1 - 2^{-l(R - \sum_{i=1}^{n-2t} R_i)},
\end{aligned}$$

which goes to 1 as $l \to \infty$. □

Notice now that according to Theorem 18, the capacity of the corresponding channel in the $R-1-1$ problem is R. For source-channel separation, the condition $R(D) \leq R$ must hold. But the scheme proposed in Section 4.5 achieves works for $R(D) > R$, which implies that the channel is being operated above capacity. Even though operating the channel above capacity is useless from the point of view of reliable communication, it appears to be beneficial from the point of view of rate-distortion. Note that in the end the decoder has to choose from two messages only, one of which it knows is the correct message. This is not sufficient for reliable communication, since the decoder cannot unequivocally determine which of the two messages is correct. However, the reduction to two messages, one of which is correct yields benefits from the point of view of rate-distortion since the two messages are constrained to be within a certain distortion-typical set.

It is instructive to consider the $R-1-1$ problem for the Hamming distortion case. Since separation is optimal in the Hamming case, we have $R(D) \leq R$ (cf. Theorem 19 and Theorem 15). Consider now the $R-R-R$ problem, *i.e.,* all three encoders have rate R. Theorems 19 and 15 tell us that we must again have

$R(D) < R$. Thus in the Hamming case, the $R-1-1$ and $R-R-R$ problems have the same rate region. This signifies that in the $R-1-1$ Hamming problem, the extra rate available to the encoder from Encoders 2 and 3 is useless; the same distortion could be achieved if Encoders 2 and 3 transmitted at the lower rate R. This is because it is optimal for the decoder to output the quantized sequence transmitted by Encoder 1 (which is the centroid of the corresponding Hamming ball) even if Encoders 2 and 3 send the complete source sequence as in the $R-1-1$ problem. The adversarial channel reveals a lot of information to the decoder through the source sequence that the traitor chooses to transmit, since the decoder eventually receives two source sequences, one of which is the true source sequence. However, this additional information is not useful at all since all the decoder needs to know to make an optimal decision is Encoder 1's quantized sequence.

The erasure distortion measure, however, is more stringent than the Hamming distortion measure, since it does not allow the decoder to make errors in its reconstruction. For this reason, the decoder needs to be absolutely certain about any non-erased bit it outputs in its reconstruction and output erasures for any bit about which it is not certain. This is not the case with Hamming distortion, since the decoder can always guess for any bit about which it is uncertain. In order to to achieve the same distortion, the erasure distortion measure requires the decoder to have more information than the Hamming distortion measure. It turns out that the information revealed by the adversarial channel, which is useless in the Hamming case, accounts for the additional information required in the erasure case. This allows Encoder 1 to transmit at a rate lower than the erasure rate-distortion function by performing Hamming quantization instead of erasure quantization, with the remaining information being supplied to the decoder by the adversarial channel.

Acknowledgements

The author would like to thank Sriram Vishwanath for helpful discussions regarding the formulation of the adversarial lossy source coding problem and Oliver Kosut for supplying the proof in Section 4.4.

CHAPTER 5
MULTI-LEVEL LOSSLESS SOURCE CODING WITH BYZANTINE ADVERSARIES

In this chapter, we study lossless source coding with multiple sources and an unknown number of adversaries. We consider a three-encoder version of the problem with two sources and one potential adversary, and show its equivalence to the three encoder symmetric MLD problem, proving in the process that superposition coding is optimal for adversarial multi-level diversity coding.

5.1 Problem Formulation

Let $\{X_t, Y_t\}_{t=1}^{\infty}$ be an *i.i.d.* source, with $\{X_t\}_{t=1}^{\infty}$ independent of $\{Y_t\}_{t=1}^{\infty}$, where the random variables X_t and Y_t take values in the finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively. There are three encoders, each of which observes the pair of length-ℓ sequences (X^ℓ, Y^ℓ) and transmits a message to a decoder. The three encoders are either all honest, or there is at most one traitor among them. The goal of the traitor, should one be present, is to sabotage the communication by not allowing the decoder to losslessly reconstruct the observed sequences, and it chooses its message in order to fulfill this goal, with full knowledge of (X^ℓ, Y^ℓ), the other two messages, and the decoder's decoding strategy. It is unknown to the honest encoders and the decoders a priori whether a traitor is present, and if there is one, its location among the three encoders. Moreover, if present, the traitor can observe (X^ℓ, Y^ℓ) and then decide which of the three encoders to take over. The traitor's location among the encoders and its actions can therefore be different for different realizations of the sources. The decoder receives the three messages from the encoders and attempts to detect the presence of a corrupted message.

If it does not detect a corrupted message, it outputs a lossless reconstruction of both X^ℓ and Y^ℓ. If it detects a corrupted message, then it losslessly reconstructs only X^ℓ and, along with its reconstruction for X^ℓ, outputs either Y^ℓ (if it can be correctly decoded) or a "flag" sequence indicating that it has detected a corrupted message.

Let $\tilde{\mathcal{Y}^\ell}$ denote the set $\{\mathcal{Y}^\ell \cup \emptyset\}$, where $\emptyset$ is a special "flag" sequence. A *code* (f_1, f_2, f_3, g) is a collection of encoders $f_i : \mathcal{X}^\ell \times \mathcal{Y}^\ell \to \{1, \dots, M_i^{(\ell)}\}$, $i \in \{1, 2, 3\}$, and a decoder $g : \prod_{i=1}^3 \{1, \dots, M_i^{(l)}\} \to \mathcal{X}^\ell \times \tilde{\mathcal{Y}^\ell}$. Denote by $\hat{X}^\ell$ and $\hat{Y}^\ell$ the decoder's reconstructions of X^ℓ and Y^ℓ. Let $\alpha \in \{1, 2, 3\}$ denote the location of the traitor if one is present, and C_α the message it transmits to the decoder. For a fixed code, fixed α and C_α, and a fixed source realization (x^ℓ, y^ℓ), define the indicator function $1_e(\mathbf{f}, g, \alpha, C_\alpha, x^\ell, y^\ell)$

$$= \begin{cases} 1 & \text{if } C_\alpha = f_\alpha(x^\ell, y^\ell) \text{ and } (\hat{x}^\ell, \hat{y}^\ell) \neq (x^\ell, y^\ell) \\ 1 & \text{if } C_\alpha \neq f_\alpha(x^\ell, y^\ell) \text{ and } (\hat{x}^\ell, \hat{y}^\ell) \neq (x^\ell, y^\ell) \text{ and } (\hat{x}^\ell, \hat{y}^\ell) \neq (x^\ell, \emptyset) \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathbf{f} = (f_1, f_2, f_3)$. The probability of error P_e is defined as

$$P_e \triangleq \mathbf{E}[\max_{\alpha} \max_{C_\alpha} 1_e(\mathbf{f}, g, \alpha, C_\alpha, X^\ell, Y^\ell)].$$

Definition 15. *A rate vector (R_1, R_2, R_3) is said to be* achievable *if for any $\epsilon > 0$ and sufficiently large ℓ, there exist encoders f_1, f_2, and f_3 and a decoder g such that*

$$R_i + \epsilon \geq \frac{1}{\ell} \log M_i^{(\ell)} \text{ for } i = 1, 2, 3, \text{ and} \tag{5.1}$$

$$\epsilon \geq P_e. \tag{5.2}$$

Let $\mathcal{R}$ denote the set of achievable rate-distortion vectors.

Definition 16. $\mathcal{R}^* = \{(R_1, R_2, R_3)$: for $i = 1, 2, 3$,

$$R_i = r_i^1 + r_i^2,$$

where $r_i^1, r_i^2 \geq 0$ and

$$\begin{aligned} r_i^1 &\geq H(X) \quad \text{for } 1 \leq i \leq 3 \\ r_i^2 + r_j^2 &\geq H(Y) \quad \text{for } 1 \leq i < j \leq 3\}. \end{aligned}$$

Theorem 19. $\mathcal{R} = \mathcal{R}^*$.

In proving Theorem 19, we shall use the rate region of the three-encoder symmetrical multi-level diversity (MLD) coding problem, which is a special case of the n-encoder MLD problem as defined in [68]. We use the notation in [68]. Suppose $\mathbf{v}$ is a vector in $\{0,1\}^3$. We say a vector $\mathbf{u} \geq \mathbf{v}$ if $u_i \geq v_i$ for $1 \leq i \leq 3$. Define

$$\Omega_3^\alpha = \{\mathbf{v} \in \{0,1\}^3 : |\mathbf{v}| = \alpha\}$$

where $|\mathbf{v}|$ denotes the Hamming weight of $\mathbf{v}$, and $\Omega_3 = \cup_{\alpha=1}^3 \Omega_3^\alpha$. Denote by $G_\mathbf{v}$ the set $\{i : v_i = 1\}$ and let (S_1, S_2, S_3) be 3 sources, taking values in alphabets $\mathcal{S}_i$ such that $S_1 = X$ and $S_2 = Y$, while $S_3 = Z$ is a zero-entropy source, *i.e.*, a deterministic sequence. Let

$$d_\alpha : \prod_{i=1}^{\alpha} \mathcal{S}_i \times \prod_{i=1}^{\alpha} \mathcal{S}_i \to \{0,1\}$$

denote the Hamming distortion measure. For a given blocklength ℓ, we define encoders

$$F_i : \prod_{i=1}^{3} \mathcal{S}_i^\ell \to \{1, \ldots, M_i^\ell\}, \quad i \in \{1,2,3\},$$

and decoders

$$T_\mathbf{v} : \prod_{i \in G_\mathbf{v}} \{1, \ldots, M_i^\ell\} \to \prod_{j=1}^{|\mathbf{v}|} S_j^\ell \quad \mathbf{v} \in \Omega_3$$

and

$$\Delta_{\mathbf{v}} = \ell^{-1}\mathbf{E}\left[\sum_{t=1}^{\ell} d_{|\mathbf{v}|}((S_{1,t},\ldots,S_{|\mathbf{v}|,t}),(\hat{S}_{1,t}(\mathbf{v}),\ldots,\hat{S}_{|\mathbf{v}|,t}(\mathbf{v})))\right] \quad \mathbf{v}\in\Omega_3$$

where $(\hat{S}_{1,t}(\mathbf{v}),\ldots,\hat{S}_{|\mathbf{v}|,t}(\mathbf{v})) = T_{\mathbf{v}}(F_i(S_1^\ell,S_2^\ell,S_3^\ell), i \in G_{\mathbf{v}}$. The rate vector (R_1,R_2,R_3) is said to be *achievable* if for every $\epsilon > 0$ and sufficiently large ℓ, there exist encoders F_i and decoders $T_{\mathbf{v}}$ as defined above such that

$$R_i + \epsilon \geq \frac{1}{\ell}\log M_i^\ell \qquad \text{for all } i \text{ and} \tag{5.3}$$

$$\epsilon \geq \Delta_{\mathbf{v}} \qquad \text{for all } \mathbf{v}\in\Omega_3. \tag{5.4}$$

We call $(F_i, T_{\mathbf{v}})$ an *MLD code*. Let $\mathcal{R}_{MLD}$ denote the set of achievable rate vectors. This definition of the MLD region is based on a symbol-error (*i.e.,* Hamming distortion) definition of error probability. We can similarly define a block-error probability version of the MLD rate region. Define

$$P_{\mathbf{v}} = \Pr((S_1^\ell,\ldots,S_{|\mathbf{v}|}^\ell) \neq (\hat{S}_1^\ell(\mathbf{v}),\ldots,\hat{S}_{|\mathbf{v}|}^\ell(\mathbf{v}))).$$

Definition 17. *The rate vector (R_1,R_2,R_3) is said to be achievable for the block-error MLD problem if for every $\epsilon > 0$ and sufficiently large ℓ, there exist encoders F_i and decoders $T_{\mathbf{v}}$ as defined above such that*

$$R_i + \epsilon \geq \frac{1}{\ell}\log M_i^\ell \qquad \textit{for all } i \textit{ and} \tag{5.5}$$

$$\epsilon \geq P_{\mathbf{v}} \qquad \textit{for all } \mathbf{v}\in\Omega_3. \tag{5.6}$$

Let $\mathcal{R}_{MLD,blk}$ denote the set of achievable rate vectors. We show in Appendix C.1 that $\mathcal{R}_{MLD} = \mathcal{R}_{MLD,blk}$. The superposition rate region, stated in Definition 18, is optimal for the MLD problem, as shown in [68]. Note that since Z is a zero-entropy source, this region is equivalent to $\mathcal{R}^*$.

Definition 18. $\mathcal{R}_{sup}\{(R_1, R_2, R_3)$:

$$R_i = \sum_{\alpha=1}^{3} r_i^\alpha \quad \textit{for } 1 \leq i \leq 3$$

for some $r_i^\alpha \geq 0$*, satisfying, for* $1 \leq \alpha \leq 3$*,*

$$\begin{aligned} r_i^1 &\geq H(X) \quad \textit{for } 1 \leq i \leq 3 \\ r_i^2 + r_j^2 &\geq H(Y) \quad \textit{for } 1 \leq i < j \leq 3 \\ r_1^3 + r_2^3 + r_3^3 &\geq H(Z)\}. \end{aligned}$$

In order to prove Theorem 19, we shall show that $\mathcal{R} = \mathcal{R}_{MLD,blk}(= \mathcal{R}_{MLD})$. This, together with the result in [68] which states that $\mathcal{R}_{MLD} = \mathcal{R}_{sup}(= \mathcal{R}^*)$, will then imply that $\mathcal{R} = \mathcal{R}^*$.

Theorem 20 (Theorem 1 in [68])**.** $\mathcal{R}_{MLD} = \mathcal{R}_{sup}$.

Theorem 21. $\mathcal{R} = \mathcal{R}_{MLD,blk}$.

Before proving Theorem 21, we will prove a few lemmas that will be integral to the proof. Consider an MLD code $(F_1, F_2, F_3, T_\mathbf{v}, \mathbf{v} \in \Omega_3)$ under the block-error probability definition (Definition 17). For each $\mathbf{v} \in \Omega_3$, define the set[1]

$$\mathcal{E}_{\mathbf{F},T_\mathbf{v}} := \{(x^\ell, y^\ell) : \hat{X}_\mathbf{v}^\ell \neq x^\ell \text{ if } \mathbf{v} \in \Omega_3^1 \text{ and } (\hat{X}_\mathbf{v}^\ell, \hat{Y}_\mathbf{v}^\ell) \neq (x^\ell, y^\ell) \text{ if } \mathbf{v} \in \Omega_3^2 \cup \Omega_3^3\},$$

where $\hat{X}_\mathbf{v}^\ell$ and $\hat{Y}_\mathbf{v}^\ell$ are the reconstructions output by decoder $T_\mathbf{v}$ for the source sequence (x^ℓ, y^ℓ).

Now consider a code (f_1, f_2, f_3, g) for the problem at hand. Define the set

$$\mathcal{E}_{\mathbf{f},g} := \{(x^\ell, y^\ell) : \max_\alpha \max_{C_\alpha} 1_e(\mathbf{f}, g, \alpha, C_\alpha, x^\ell, y^\ell) = 1\}.$$

[1] The 3-encoder MLD problem has three sources X, Y, and Z. However, since we assume Z is a zero-entropy source, *i.e.*, a deterministic sequence, we shall only consider sources X and Y.

Lemma 5. *Given an MLD code $(f_1, f_2, f_3, T_{\mathbf{v}}, \mathbf{v} \in \Omega_3)$, there exists a code (f_1, f_2, f_3, g) for the problem at hand, with the same encoders as the MLD code, such that $\mathcal{E}_{\mathbf{f},g} \subset \bigcup_{\mathbf{v}\in\Omega_3} \mathcal{E}_{\mathbf{f},T_{\mathbf{v}}}$.*

Proof. Consider an MLD code $(f_1, f_2, f_3, T_{\mathbf{v}}, \mathbf{v} \in \Omega_3)$. We will construct a code (f_1, f_2, f_3, g) for the problem at hand with the same encoders as the MLD encoders, and a decoder such that $\mathcal{E}_{\mathbf{f},g} \subset \bigcup_{\mathbf{v}\in\Omega_3} \mathcal{E}_{\mathbf{f},T_{\mathbf{v}}}$. The decoder receives all three messages, one of which could potentially be corrupted. The decoder reconstructs X^ℓ as follows. For each of the three singleton decoders $T_{\mathbf{v}}$, $\mathbf{v} \in \Omega_3^1$, the decoder reconstructs X^ℓ, thus obtaining three reconstructions $\hat{X}_1^\ell$, $\hat{X}_2^\ell$, and $\hat{X}_3^\ell$. It then outputs its own reconstruction of X^ℓ by choosing randomly from the sequences with maximal multiplicity among $\hat{X}_1^\ell$, $\hat{X}_2^\ell$, and $\hat{X}_3^\ell$: if at least two of the three reconstruction are equal, the decoder outputs the common sequence as $\hat{X}^\ell$. Otherwise, it outputs a randomly chosen sequence from $\hat{X}_1^\ell$, $\hat{X}_2^\ell$, and $\hat{X}_3^\ell$ as its reconstruction.

The decoding process for Y^ℓ is similar. For every pair of messages, the decoder produces reconstructions $\hat{Y}_{12}^\ell$, $\hat{Y}_{13}^\ell$, and $\hat{Y}_{23}^\ell$ using the MLD decoders $T_{\mathbf{v}}$ for $\mathbf{v} \in \Omega_3^2$. If all of the $\hat{Y}_{12}^\ell, \hat{Y}_{13}^\ell$, and $\hat{Y}_{23}^\ell$ sequences are identical, the decoder outputs the common sequence as its reconstruction for Y^ℓ. Otherwise, it outputs $\hat{Y}^\ell = \emptyset$.

Consider now a sequence $(x^\ell, y^\ell) \in \mathcal{E}_{\mathbf{f},g}$, and suppose the traitor takes over Encoder α, $\alpha \in \{1, 2, 3\}$ and transmits message C_α. Then either

1. $C_\alpha = f_\alpha(x^\ell, y^\ell)$ and $(\hat{x}^\ell, \hat{y}^\ell) \neq (x^\ell, y^\ell)$, or
2. $C_\alpha \neq f_\alpha(x^\ell, y^\ell)$ and $(\hat{x}^\ell, \hat{y}^\ell) \neq (x^\ell, y^\ell)$ and $(\hat{x}^\ell, \hat{y}^\ell) \neq (x^\ell, \emptyset)$.

Consider the first case. An error would occur for a source sequence (x^ℓ, y^ℓ)

(*i.e.*, $(x^\ell, y^\ell) \in \mathcal{E}_{\mathbf{f},g}$) if

(i) all three reconstructions $\hat{X}_1^\ell$, $\hat{X}_2^\ell$, and $\hat{X}_3^\ell$ are mutually different, or

(ii) at least two of $\hat{X}_1^\ell$, $\hat{X}_2^\ell$, and $\hat{X}_3^\ell$ are identical, but not equal to x^ℓ.

(iii) at least one of $\hat{Y}_{12}^\ell, \hat{Y}_{13}^\ell$, and $\hat{Y}_{23}^\ell$ is not equal to y^ℓ.

In the first two cases, it is evident that at least one of $\hat{X}_1^\ell$, $\hat{X}_2^\ell$, and $\hat{X}_3^\ell$ is not equal to x^ℓ, which implies that the sequence $(x^\ell, y^\ell) \in \mathcal{E}_{\mathbf{f},T_{\mathbf{v}}}$ for that encoder-decoder pair. Likewise, in the third case, since at least one of $\hat{Y}_{12}^\ell, \hat{Y}_{13}^\ell$, and $\hat{Y}_{23}^\ell$ is not equal to y^ℓ, $(x^\ell, y^\ell) \in \mathcal{E}_{\mathbf{f},T_{\mathbf{v}}}$ for that pair of encoders and the corresponding weight-2 decoder. Thus $(x^\ell, y^\ell) \in \bigcup_{\mathbf{v}\in\Omega_3} \mathcal{E}_{\mathbf{f},T_{\mathbf{v}}}$.

Consider the second case. An error would occur for a source sequence (x^ℓ, y^ℓ) (*i.e.*, $(x^\ell, y^\ell) \in \mathcal{E}_{\mathbf{f},g}$) if

(i) all three reconstructions $\hat{X}_1^\ell$, $\hat{X}_2^\ell$, and $\hat{X}_3^\ell$ are mutually different, or

(ii) at least two of $\hat{X}_1^\ell$, $\hat{X}_2^\ell$, and $\hat{X}_3^\ell$ are identical, but not equal to x^ℓ.

(iii) all of $\hat{Y}_{12}^\ell, \hat{Y}_{13}^\ell$, and $\hat{Y}_{23}^\ell$ are identical, but not equal to y^ℓ.

In the first two cases, it is evident that at least one of $\hat{X}_1^\ell$, $\hat{X}_2^\ell$, and $\hat{X}_3^\ell$ is not equal to x^ℓ. In the first case, at least two of the messages are uncorrupted, and hence if all three reconstructions are different, then the reconstructions from the two uncorrupted messages must also be different, which implies that the sequence $(x^\ell, y^\ell) \in \mathcal{E}_{\mathbf{f},T_{\mathbf{v}}}$ for one of the honest encoder and the corresponding decoder. Likewise, in the second case, at least one of the two identical reconstructions must be from an honest encoder, which implies that the sequence $(x^\ell, y^\ell) \in \mathcal{E}_{\mathbf{f},T_{\mathbf{v}}}$ for that encoder and decoder. Therefore, the sequence $(x^\ell, y^\ell) \in \mathcal{E}_{\mathbf{f},T_{\mathbf{v}}}$ for some

$\mathbf{v} \in \Omega_3^1$, and thus $(x^\ell, y^\ell) \in \bigcup_{\mathbf{v}\in\Omega_3} \mathcal{E}_{\mathbf{f},T_\mathbf{v}}$. Similarly, in the third case, one pair of encoders is honest, and the reconstruction from this honest pair of encoders is not equal to y^ℓ, which implies that $(x^\ell, y^\ell) \in \mathcal{E}_{\mathbf{f},T_\mathbf{v}}$ for those two encoders and the corresponding weight-2 decoder. Therefore, $(x^\ell, y^\ell) \in \bigcup_{\mathbf{v}\in\Omega_3} \mathcal{E}_{\mathbf{f},T_\mathbf{v}}$. □

Consider a code (f_1, f_2, f_3, g) for the problem at hand. For a given source realization (x^ℓ, y^ℓ), define $\mathcal{M}_{f_i}(x^\ell, y^\ell)$ to be the set of source sequences such that for all $(\tilde{x}^\ell, \tilde{y}^\ell) \in \mathcal{M}_{f_i}(x^\ell, y^\ell)$, $f_i(\tilde{x}^\ell, \tilde{y}^\ell) = f_i(x^\ell, y^\ell)$, for $i \in \{1, 2, 3\}$. Moreover, suppose $\mathcal{M}_{f_i}(X^\ell, Y^\ell)$ takes set values $\mathcal{M}^1_{f_i}, \mathcal{M}^2_{f_i}$, and so on. Then, for a fixed i, the sets $\mathcal{M}^j_{f_i}$ are the pre-images of the messages transmitted by Encoder f_i, and form a partition over the set of source sequences. For a fixed $\mathcal{M}^j_{f_i}$, define $\mathcal{G}_{\mathcal{M}^j_{f_i}} := \{(x^\ell, y^\ell) \in \mathcal{M}_{f_i,j} : (x^\ell, y^\ell) \notin \mathcal{E}_{\mathbf{f},g}\}$.

Lemma 6. *For $i \in \{1, 2, 3\}$, and for any j, if $(x^\ell, y^\ell) \in \mathcal{G}_{\mathcal{M}^j_{f_i}}$ and $(\tilde{x}^\ell, \tilde{y}^\ell) \in \mathcal{G}_{\mathcal{M}^j_{f_i}}$, then $x^\ell = \tilde{x}^\ell$.*

Proof. Assume WLOG that $i = 1$, and suppose there exists j such that the set $\mathcal{G}_{\mathcal{M}^j_{f_1}}$ contains two source sequences (x^ℓ, y^ℓ) and $(\tilde{x}^\ell, \tilde{y}^\ell)$ such that $x^\ell \neq \tilde{x}^\ell$. The decoder receives all three messages, one of which may be potentially corrupted. Suppose (x^ℓ, y^ℓ) is the true source sequence, and that the message transmitted by Encoder f_3 is corrupted by a traitor. Encoders f_1 and f_2, being honest, transmit $f_1(x^\ell, y^\ell)$ and $f_2(x^\ell, y^\ell)$, respectively. The traitor transmits $f_3(\tilde{x}^\ell, \tilde{y}^\ell)$. Since $(x^\ell, y^\ell) \in \mathcal{G}_{\mathcal{M}^j_{f_1}}$ (and therefore $(x^\ell, y^\ell) \notin \mathcal{E}_{\mathbf{f},g}$ by definition), the decoder outputs $\hat{X}^\ell = x^\ell$ regardless of Encoder f_3's message.

Now suppose $(\tilde{x}^\ell, \tilde{y}^\ell)$ is the true source sequence. Then Encoder f_1 transmits $f_1(\tilde{x}^\ell, \tilde{y}^\ell)$. Then the traitor can take over Encoder f_2 and transmit $f_2(x^\ell, y^\ell)$, which would mean, as described above, that the decoder would output $\hat{X}^\ell =$

$x^\ell \neq \tilde{x}^\ell$ regardless of Encoder f_3's message, leading to an error. Thus $(\tilde{x}^\ell, \tilde{y}^\ell) \in \mathcal{E}_{\mathbf{f},g}$ and therefore $(\tilde{x}^\ell, \tilde{y}^\ell) \notin \mathcal{G}_{\mathcal{M}^j_{f_i}}$, which is a contradiction. Hence $\tilde{x}^\ell$ must be equal to x^ℓ. □

Consider again a code (f_1, f_2, f_3, g) for the problem at hand. We can similarly define sets that are intersections of the pre-images of two encoder messages. More precisely, for a given source realization (x^ℓ, y^ℓ), define $\mathcal{M}_{f_i,f_j}(x^\ell, y^\ell)$ to be the set of source sequences such that for all $(\tilde{x}^\ell, \tilde{y}^\ell) \in \mathcal{M}_{f_i,f_j}(x^\ell, y^\ell)$, $f_i(\tilde{x}^\ell, \tilde{y}^\ell) = f_i(x^\ell, y^\ell)$ and $f_j(\tilde{x}^\ell, \tilde{y}^\ell) = f_j(x^\ell, y^\ell)$. Moreover, suppose $\mathcal{M}_{f_i,f_j}(X^\ell, Y^\ell)$ takes set values $\mathcal{M}^1_{f_i,f_j}, \mathcal{M}^2_{f_i,f_j}, \mathcal{M}^3_{f_i,f_j}$ and so on. Then, for fixed i and j, the sets $\mathcal{M}^k_{f_i,f_j}$ are intersections of the pre-images of the two messages by Encoders f_i and f_j, and form a partition over the set of source sequences. For a fixed $\mathcal{M}^k_{f_i,f_j}$, define $\mathcal{G}_{\mathcal{M}^k_{f_i,f_j}} := \{(x^\ell, y^\ell) \in \mathcal{M}^k_{f_i,f_j} : (x^\ell, y^\ell) \notin \mathcal{E}_{\mathbf{f},g}\}$.

Lemma 7. *For $1 \leq i < j \leq 3$, and for any k, $|\mathcal{G}_{\mathcal{M}^k_{f_i,f_j}}| \leq 1$.*

Proof. Suppose (X^ℓ, Y^ℓ) is the observed source sequence. For every possible set of three messages that the decoder receives, it must output reconstructions for X^ℓ and Y^ℓ. For a given triple of messages, the decoder may either decide that there is no traitor and output $(\hat{X}^\ell, \hat{Y}^\ell)$ as its reconstruction, or it may decide that a traitor is present and output $(\hat{X}^\ell, \emptyset)$. The messages from Encoders f_1 and f_2, assuming they are not corrupted, tell the decoder that the observed source sequence is in one of the sets $\mathcal{M}^k_{f_1,f_2}$. Suppose Encoder f_3 transmits a message C_3. For each set $\mathcal{M}^k_{f_1,f_2}$, define the set $\mathcal{S}_k := \{(x^\ell, y^\ell) \in \mathcal{M}^k_{f_1,f_2} : (\hat{X}^\ell, \hat{Y}^\ell) = (x^\ell, y^\ell) \text{ if } C_3 = f_3(x^\ell, y^\ell)\}$. We claim that

1. if $(x^\ell, y^\ell) \notin \mathcal{S}_k$, then $(x^\ell, y^\ell) \in \mathcal{E}_{\mathbf{f},g}$, and
2. if $|\mathcal{S}_k| > 1$, then $\mathcal{S}_k \subset \mathcal{E}_{\mathbf{f},g}$.

To see why Claim 1 is true, consider $(x^\ell, y^\ell) \notin \mathcal{S}_k$. Then if $C_3 = f_3(x^\ell, y^\ell)$, $(\hat{X}^\ell, \hat{Y}^\ell) \neq (x^\ell, y^\ell)$. Therefore, if (x^ℓ, y^ℓ) is the true source sequence and the traitor takes over Encoder f_3 and transmits $f_3(x^\ell, y^\ell)$, the decoder will output $(\hat{X}^\ell, \hat{Y}^\ell) \neq (x^\ell, y^\ell)$, resulting in an error. Thus $(x^\ell, y^\ell) \in \mathcal{E}_{\mathbf{f},g}$. To see why Claim 2 is true, suppose $|\mathcal{S}_k| > 1$ and consider $(x^\ell, y^\ell) \in \mathcal{S}_k$ and $(\tilde{x}^\ell, \tilde{y}^\ell) \in \mathcal{S}_k$. Then if (x^ℓ, y^ℓ) is the true source sequence, the traitor can transmit $f_3(\tilde{x}^\ell, \tilde{y}^\ell)$, causing the decoder to output $(\hat{X}^\ell, \hat{Y}^\ell) = (\tilde{x}^\ell, \tilde{y}^\ell)$ which would result in an error. Thus $(x^\ell, y^\ell) \in \mathcal{E}_{\mathbf{f},g}$. Similarly, if $(\tilde{x}^\ell, \tilde{y}^\ell)$ is the true source sequence, then the traitor can transmit $f_3(x^\ell, y^\ell)$ which would result in an error. Thus $(\tilde{x}^\ell, \tilde{y}^\ell) \in \mathcal{E}_{\mathbf{f},g}$. Therefore, $\mathcal{S}_k \subset \mathcal{E}_{\mathbf{f},g}$.

Claims 1 and 2 imply that if $\mathcal{S}_k = \emptyset$ or $|\mathcal{S}_k| > 1$, then $\mathcal{G}_{\mathcal{M}^k_{f_i,f_j}} = \emptyset$, and if $|\mathcal{S}_k| = 1$, then since $\mathcal{G}_{\mathcal{M}^k_{f_i,f_j}} \subset \mathcal{S}_k$, and therefore $|\mathcal{G}_{\mathcal{M}^k_{f_i,f_j}}| \leq |\mathcal{S}_k| = 1$. □

Lemma 8. *Given a code (f_1, f_2, f_3, g) for the problem at hand, there exists an MLD code $(f_1, f_2, f_3, T_\mathbf{v}, \mathbf{v} \in \Omega_3)$, with the same encoders, such that for all $\mathbf{v} \in \Omega_3$, $\mathcal{E}_{\mathbf{f},T_\mathbf{v}} \subset \mathcal{E}_{\mathbf{f},g}$.*

Proof. Consider a code (f_1, f_2, f_3, g) for the problem at hand. We construct an MLD code that has the same encoders f_1, f_2, and f_3. We define the three weight-1 MLD decoders as follows. Let $\mathcal{M}^j_{f_i}$ be the pre-images of the messages sent by Encoder f_i, $i \in \{1, 2, 3\}$, and let $\mathcal{G}_{\mathcal{M}^j_{f_i}}$ be as defined previously. The weight-1 decoder corresponding to Encoder f_i operates as follows. If $\mathcal{M}^j_{f_i}$ is the pre-image of the message it receives from Encoder f_i, then the decoder looks at the set $\mathcal{G}_{\mathcal{M}^j_{f_i}}$. If $\mathcal{G}_{\mathcal{M}^j_{f_i}} \neq \emptyset$, Lemma 6 guarantees that all sequences in $\mathcal{G}_{\mathcal{M}^j_{f_i}}$ will have the same X component, say x^ℓ. The decoder outputs x^ℓ as its reconstruction for X^ℓ. If $\mathcal{G}_{\mathcal{M}^j_{f_i}}$ is empty, then it randomly picks a sequence from $\mathcal{M}^j_{f_i}$ and outputs its X component as its reconstruction for X^ℓ.

We define the weight-2 MLD decoders similarly. Let $\mathcal{M}^k_{f_i,f_j}$ be the intersection of the pre-images of the messages sent by Encoders f_i and f_j, $1 \leq i < j \leq 3$, and let $\mathcal{G}_{\mathcal{M}^k_{f_i,f_j}}$ be as defined previously. The weight-2 decoder corresponding to Encoders f_i and f_j operates as follows. If $\mathcal{M}^k_{f_i,f_j}$ is the intersection of the pre-images of the messages it receives from Encoders f_i and f_j, then the decoder looks at the set $\mathcal{G}_{\mathcal{M}^k_{f_i,f_j}}$. If $\mathcal{G}_{\mathcal{M}^k_{f_i,f_j}} \neq \emptyset$, Lemma 7 guarantees that there will only be a single sequence in $\mathcal{G}_{\mathcal{M}^k_{f_i,f_j}}$. The decoder outputs this sequence as its reconstruction for (X^ℓ, Y^ℓ). If $\mathcal{G}_{\mathcal{M}^k_{f_i,f_j}}$ is empty, the decoder outputs a randomly chosen sequence from $\mathcal{M}^k_{f_i,f_j}$ as its reconstruction.

The weight-3 decoder receives all three messages and operates by outputting the same reconstruction as the weight-2 encoder that receives messages from Encoders f_1 and f_2.

Consider now the error sets of these MLD decoders. The weight-1 decoders output the X component of the sequences in $\mathcal{G}_{\mathcal{M}^j_{f_i}} \neq \emptyset$, if this set is not empty. An error therefore occurs for all sequences not in $\mathcal{G}_{\mathcal{M}^j_{f_i}}$, *i.e.*, all of these sequences are in $\mathcal{E}_{\mathbf{f},T_{\mathrm{v}}}$. Note that all of these sequences are also in $\mathcal{E}_{\mathbf{f},g}$, since if $(x^\ell, y^\ell) \notin \mathcal{G}_{\mathcal{M}^j_{f_i}}$, then $(x^\ell, y^\ell) \in \mathcal{E}_{\mathbf{f},g}$ by definition. If $\mathcal{G}_{\mathcal{M}^j_{f_i}}$ is empty, then the decoder outputs a randomly chosen sequence from $\mathcal{M}^j_{f_i}$, which means that all other sequences in $\mathcal{M}^j_{f_i}$ are in $\mathcal{E}_{\mathbf{f},T_{\mathrm{v}}}$. Note that in this case, all sequences in $\mathcal{M}^j_{f_i}$ are in $\mathcal{E}_{\mathbf{f},g}$, since $\mathcal{G}_{\mathcal{M}^j_{f_i}}$ is empty. In both cases, therefore, if $(x^\ell, y^\ell) \in \mathcal{E}_{\mathbf{f},T_{\mathrm{v}}}$, then $(x^\ell, y^\ell) \in \mathcal{E}_{\mathbf{f},g}$.

The weight-2 decoders output the solitary sequence in $\mathcal{G}_{\mathcal{M}^k_{f_i,f_j}}$ if it is not empty. An error therefore occurs for all sequences not in $\mathcal{G}_{\mathcal{M}^k_{f_i,f_j}}$, *i.e.*, all of these sequences are in $\mathcal{E}_{\mathbf{f},T_{\mathrm{v}}}$. Note that all of these sequences are also in $\mathcal{E}_{\mathbf{f},g}$, since if $(x^\ell, y^\ell) \notin \mathcal{G}_{\mathcal{M}^k_{f_i,f_j}}$, then $(x^\ell, y^\ell) \in \mathcal{E}_{\mathbf{f},g}$ by definition. If $\mathcal{G}_{\mathcal{M}^k_{f_i,f_j}}$ is empty, then the decoder outputs a randomly chosen sequence from $\mathcal{M}^k_{f_i,f_j}$, which means that all other sequences in $\mathcal{M}^k_{f_i,f_j}$ are in $\mathcal{E}_{\mathbf{f},T_{\mathrm{v}}}$. Note that in this case, all sequences in

$\mathcal{M}^k_{f_i,f_j}$ are in $\mathcal{E}_{\mathbf{f},g}$, since $\mathcal{G}_{\mathcal{M}^k_{f_i,f_j}}$ is empty. In both cases, therefore, if $(x^\ell, y^\ell) \in \mathcal{E}_{\mathbf{f},T_{\mathbf{v}}}$, then $(x^\ell, y^\ell) \in \mathcal{E}_{\mathbf{f},g}$.

The error set of the weight-3 decoder is the same as the error set of the weight-2 decoder corresponding to Encoders f_1 and f_2, and therefore is a subset of $\mathcal{E}_{\mathbf{f},g}$. □

Theorem 21 now follows straightforwardly from Lemmas 5 and 8. We omit the details.

APPENDIX A

CHAPTER 2: PROOFS

A.1 Preliminaries

We define a multi-variable mutual information as follows:

$$\begin{aligned} I_K(X_1; X_2; \ldots; X_K) &= D\left(p(X_1, \ldots, X_K) || \prod_{i=1}^{K} p(X_i)\right) \\ &= \sum_{i=1}^{K} H(X_i) - H(X_1, \ldots, X_K). \end{aligned}$$

In particular, $I_1(X) = 0$. The multi-variable mutual information, as defined above, is a measure of the mutual dependence among K random variables and is different from McGill's multivariate mutual information [41]. We note the following properties of $I_K(X_1; X_2; \ldots; X_K)$.

1. $I_K(\mathbf{X}_1^l; \ldots; \mathbf{X}_K^l) \geq 0$.
2. $I_K(X_1; \ldots; X_K) \geq I_m(X_1; \ldots; X_m) + I_{(K-m+1)}(f(X_1, \ldots, X_m); X_{m+1}; \ldots; X_K)$, where $f(X_1, \ldots, X_m)$ is a function of the random variables $X_1, \ldots, X_m$, $m < K$.

   Remark: This property holds by symmetry for the general case when $f(\cdot)$ is a function of any size-m subset of $X_1, \ldots, X_K$.

   Proof.

$$\begin{aligned} & I_K(X_1; \ldots; X_K) \\ &= \sum_{i=1}^{m} H(X_i) + \sum_{i=m+1}^{K} H(X_i) - H(X_1, \ldots, X_m) \end{aligned}$$

$$
\begin{aligned}
&\qquad - H(X_{m+1}, \ldots, X_K | X_1, \ldots, X_m) \\
&= I_m(X_1; \ldots; X_m) + \sum_{i=m+1}^{K} H(X_i) \\
&\qquad - H(X_{m+1}, \ldots, X_K | X_1, \ldots, X_m) \\
&= I_m(X_1; \ldots; X_m) + \sum_{i=m+1}^{K} H(X_i) \\
&\qquad - H(X_{m+1}, \ldots, X_K | X_1, \ldots, X_m, f(X_1, \ldots, X_m)) \\
&\geq I_m(X_1; \ldots; X_m) + \sum_{i=m+1}^{K} H(X_i) \\
&\qquad - H(X_{m+1}, \ldots, X_K | f(X_1, \ldots, X_m)) \\
&= I_m(X_1; \ldots; X_m) \\
&\qquad + I_{(K-m+1)}(f(X_1, \ldots, X_m); X_{m+1}; \ldots; X_K),
\end{aligned}
$$

where the solitary inequality holds because conditioning never increases entropy. □

3. $I(X_1; \ldots; X_i; \ldots; X_K) \geq I(X_1; \ldots; f(X_i); \ldots; X_K)$, where $f(X_i)$ is a function of the random variable X_i. This is the data processing inequality for the multi-variable mutual information and is a special case of Property 2.

A.2 Proof of Theorem 3

Let $D_k < 1 - \frac{k}{n}$ and rational. Let f_i, $i \in \mathcal{N}$ and $g_{\mathcal{K}}$, $\mathcal{K} \subseteq \mathcal{N}, \mathcal{K} \neq \emptyset$, be a code that achieves $(R_k(D_k), D_1, \ldots, D_k, \ldots, D_n)$. Let $R_k(D_k)$ be the rate of f_i, $i \in \mathcal{N}$. Consider endowing the source with an *i.i.d.* uniform distribution over $\mathcal{X}^l$

for analysis purposes. From the proof of Theorem 7 (*cf.* (A.3) and (A.4)) and using the fact that the worst-case distortion is no lower than the average-case distortion, we obtain $I_k(f_{s_1}; \ldots ; f_{s_k}) = 0$.

Let $\hat{\mathbf{X}}^l_{s_i}$ be the reconstructed source string when the decoder has access to the s_i^{th} description only. By Property 3 of the multi-variable mutual information, $I_k(\hat{\mathbf{X}}^l_{s_1}; \ldots ; \hat{\mathbf{X}}^l_{s_k}) \leq I_k(f_{s_1}; \ldots ; f_{s_k}) = 0$ for all $S \subset \mathcal{N},\ |S| = k$. By Property 2 of the multi-variable mutual information, $I(\hat{\mathbf{X}}^l_i; \hat{\mathbf{X}}^l_j) = 0$ for all $i, j \in \mathcal{N},\ i \neq j$, and thus $I(\hat{X}_{it}; \hat{X}_{jt}) = 0$ for all $i, j \in \mathcal{N},\ i \neq j$, and $t = 1, \ldots , l$. Now if any two of the $\hat{\mathbf{X}}^l_{s_i}$ disagree in a source symbol they reveal, then the resulting single-message distortion is going to be ∞ and the result follows trivially, so suppose that the $\hat{\mathbf{X}}^l_{s_i}$ are consistent. Then by Lemma 1, we have

$$\sum_{i=1}^{n} \max_{\mathbf{x}^l \in \mathcal{X}^l} \left[\frac{1}{l} \sum_{t=1}^{l} d(x_t, \hat{X}_{it}) \right] \geq n - 1,$$

which implies

$$D_1 = \max_{i \in \mathcal{N}} \max_{\mathbf{x}^l \in \mathcal{X}^l} \left[\frac{1}{l} \sum_{t=1}^{l} d(x_t, \hat{X}_{it}) \right] \geq \frac{n-1}{n} = 1 - \frac{1}{n}.$$

This completes the proof.

A.3 Proof of Theorem 4

If $R' < R_k(D_k)$, then the sum rate of any k descriptions is strictly less than $1 - D_k$, and the source string cannot be reconstructed with distortion D_k. Thus the rate of each description must be at least $R_k(D_k)$. Now, in light of the previous theorem, it suffices to show that for any $(R_k(D_k), D_1, \ldots , D_k, \ldots , D_n) \in \mathcal{RD}_{worst}$, if $D_1 = 1 - \frac{1}{n}$, then $D_m \geq 1 - \frac{m}{n}$ for $m < k$. Let $S = \{s_1, \ldots , s_k\}$ and

$\mathcal{M} = \{s_1, \ldots, s_m\}$. Let $\mathbf{X}^l_{\mathcal{M}}$ be the source reconstruction when the decoder has access to set of descriptions indexed by the elements in $\mathcal{M}$. Then from (A.4) and Properties 2 and 3 of the multi-variable mutual information, it follows that

$$\begin{aligned} I(\mathbf{X}^l_{\mathcal{M}}; \mathbf{X}^l_{s_{m+1}}; \ldots; \mathbf{X}^l_{s_k}) &\leq I(\mathbf{X}^l_{\mathcal{M}}; f_{s_{m+1}}; \ldots; f_{s_k}) \\ &\leq I_k(f_{s_1}; \ldots; f_{s_k}) = 0, \end{aligned}$$

and thus $I(X_{\mathcal{M},t}; X_{s_{m+1},t}; \ldots, X_{s_k,t}) = 0$ for $t = 1, \ldots, l$. This implies that for each t, the $(n-m+1)$ random variables $\{X_{\mathcal{M},t}; X_{s_{m+1},t}; \ldots; X_{s_n,t}\}$ are pairwise independent, and therefore by Lemma 1,

$$\begin{aligned} \max_{\mathbf{x}^l \in \mathcal{X}^l} \left[\frac{1}{l} \sum_{t=1}^{l} d(x_t, X_{\mathcal{M},t}) \right] + \sum_{i=m+1}^{n} \max_{\mathbf{x}^l \in \mathcal{X}^l} \left[\frac{1}{l} \sum_{t=1}^{l} d(x_t, X_{s_i,t}) \right] \\ \geq n - m. \end{aligned}$$

Since $D_1 = 1 - \frac{1}{n}$, we have

$$\max_{\mathbf{x}^l \in \mathcal{X}^l} \left[\frac{1}{l} \sum_{t=1}^{l} d(x_t, X_{s_i,t}) \right] \leq 1 - \frac{1}{n}$$

for $m + 1 \leq i \leq n$, and thus

$$\begin{aligned} &\max_{\mathbf{x}^l \in \mathcal{X}^l} \left[\frac{1}{l} \sum_{t=1}^{l} d(x_t, X_{\mathcal{M},t}) \right] \\ &\geq n - m - \sum_{i=m+1}^{n} \max_{\mathbf{x}^l \in \mathcal{X}^l} \left[\frac{1}{l} \sum_{t=1}^{l} d(x_t, X_{s_i,t}) \right] \\ &\geq n - m - (n-m)\left(1 - \frac{1}{n}\right) \\ &= \frac{n-m}{n} = 1 - \frac{m}{n}, \end{aligned}$$

which implies

$$D_m = \max_{\substack{\mathcal{M} \subset \mathcal{N} \\ |\mathcal{M}| = m}} \max_{\mathbf{x}^l \in \mathcal{X}^l} \left[\frac{1}{l} \sum_{t=1}^{l} d(x_t, X_{\mathcal{M},t}) \right] \geq 1 - \frac{m}{n}.$$

This completes the proof.

A.4 Proof of Theorem 5

Since m divides n, we can form n/m sets consisting of m messages each. Denote these sets by $\mathcal{M}_1, \ldots, \mathcal{M}_{n/m}$, where $\mathcal{M}_i \subset \{f_1, \ldots, f_n\}$, $|\mathcal{M}_i| = m$, and $\mathcal{M}_i \cap \mathcal{M}_j = \emptyset$, $i, j \in \{1, \ldots, n/m\}$, $i \neq j$. Since $m \leq k/2$, there exists a set $S = \{s_1, \ldots, s_k\}$ of k messages containing $\mathcal{M}_i$ and $\mathcal{M}_j$ for some $i, j \in \{1, \ldots, n/m\}$, $i \neq j$. Let $\mathbf{X}^l_{\mathcal{M}_i}$ be the source reconstruction when the decoder has access to the messages in $\mathcal{M}_i$ only. By Property 2 of the multi-variable mutual information, it follows that for the set S containing $\mathcal{M}_i$ and $\mathcal{M}_j$,

$$\begin{aligned} I(\mathbf{X}^l_{\mathcal{M}_i}; \mathbf{X}^l_{\mathcal{M}_j}) &\leq I_{(k-2m+2)}(\mathbf{X}^l_{\mathcal{M}_i}; \mathbf{X}^l_{\mathcal{M}_j}; f_r; \ldots; f_{r+k-2m-1}) \\ &\leq I_k(f_{s_1}; \ldots; f_{s_k}) = 0, \end{aligned}$$

where $f_r, \ldots, f_{r+k-2m-1} \in \{f_{s_1}, \ldots, f_{s_k}\} \setminus \{\mathcal{M}_i, \mathcal{M}_j\}$. By Lemma 1, we have

$$\sum_{i=1}^{n/m} \max_{\mathbf{x}^l \in \mathcal{X}^l} \left[\frac{1}{l} \sum_{t=1}^{l} d(x_t, X_{\mathcal{M}_i,t}) \right] \geq \frac{n}{m} - 1,$$

and thus

$$\begin{aligned} D_m &= \max_{\substack{\mathcal{M} \subset \mathcal{N} \\ |\mathcal{M}|=m}} \max_{\mathbf{x}^l \in \mathcal{X}^l} \left[\frac{1}{l} \sum_{t=1}^{l} d(x_t, X_{\mathcal{M},t}) \right] \\ &\geq \max_{i \in \{1, \ldots, n/m\}} \max_{\mathbf{x}^l \in \mathcal{X}^l} \left[\frac{1}{l} \sum_{t=1}^{l} d(x_t, X_{\mathcal{M}_i,t}) \right] \\ &\geq \frac{\frac{n}{m} - 1}{\frac{n}{m}} = 1 - \frac{m}{n}. \end{aligned}$$

This completes the proof.

A.5 Proof of Theorem 7

The proof of the first part of Theorem 7 is simple. Let $D_k \geq 1 - \frac{k}{n}$. No excess rate for every k descriptions implies that every description has rate $R_k(D_k)$. If the decoder receives m descriptions, then it receives a sum-rate of $mR_k(D_k)$ bits per source symbol. Using the point-to-point rate-distortion function for a binary source with erasure distortion, we get $D_m \geq 1 - mR_k(D_k)$.

The proof of the second part of Theorem 7 is less trivial. We begin with a lemma which is similar in spirit to Lemma 1 for worst-case distortion.

Lemma 9. *Let $X_1, \ldots, X_n$ be erased versions (Definition 4) of a uniform binary random variable X taking values in $\{+, -\}$. If $\left(1 - \frac{1}{n}\right)^k \leq \frac{1}{2}$ and $I_k(X_{s_1}; \ldots; X_{s_k}) = 0 \quad \forall\, S = \{s_1, \ldots, s_k\}, S \subset \mathcal{N}, |S| = k$, then $\sum_{i=1}^{n} \Pr(X_i = 0) \geq n - 1$.*

Proof. $\left(1 - \frac{1}{n}\right)^k \leq \frac{1}{2} \Rightarrow \left(\frac{1}{2}\right)^{\frac{1}{k}} \geq 1 - \frac{1}{n}$. We have the following four cases:

Case I: There exists $i \in \mathcal{N}$ such that $\Pr(X_i = +) > 0$ and $\Pr(X_i = -) > 0$.

Assume $i = 1$ without loss of generality. Since $X_1, \ldots, X_n$ are erased versions of the same variable, they can never disagree in the source symbol they reveal (*i.e.*, if $X_i = +$ for some $i \in \mathcal{N}$, then the rest cannot be $-$, and if $X_i = -$, then the rest cannot be $+$). Thus $\Pr(X_1 = +, X_j = -) = 0$, $j \in \{2, \ldots, n\}$. Since $I_k(X_{s_1}; \ldots; X_{s_k}) = 0$ for any set of k variables containing X_1 and X_j, X_1 and X_j must be independent. Thus

$$\begin{aligned} \Pr(X_1 = +) \cdot \Pr(X_j = -) &= \Pr(X_1 = +, X_j = -) = 0 \\ &\Rightarrow \Pr(X_j = -) = 0. \end{aligned} \tag{A.1}$$

Likewise, $\Pr(X_1 = -, X_j = +) = 0 \Rightarrow \Pr(X_j = +) = 0$. Thus $\Pr(X_j = 0) = 1$ and so $\sum_{i=1}^{n} \Pr(X_i = 0) \geq n - 1$.

Case II: There exists $i \in \mathcal{N}$ such that $\Pr(X_i = +) > 0$ and $\Pr(X_i = -) = 0$, and Case I does not hold.

Let $S = \{s_1, \ldots, s_k\}$ be a size-k subset of $\mathcal{N}$. For all $\mathcal{T} \subset S$, denote by $E_{\mathcal{T}}$ the event that $X_{s_j} = - \ \forall \ s_j \in \mathcal{T}$, and $X_{s_j} = 0 \ \forall \ s_j \notin \mathcal{T}, \ s_j \in S$. Now since $\Pr(X_{s_j} = -) = 0$ from (A.1), $\Pr(E_{\mathcal{T}}) = 0 \ \forall \ \mathcal{T} \neq \emptyset$. Thus

$$\begin{aligned} \Pr(X = -) &\leq \sum_{\mathcal{T} \subset S} \Pr(E_{\mathcal{T}}) \\ &= \Pr(X_{s_1} = X_{s_2} = \ldots = X_{s_k} = 0). \end{aligned} \tag{A.2}$$

Since $\Pr(X = -) = 1/2$ and $(X_{s_1}, \ldots, X_{s_k})$ are independent, (A.2) yields

$$\prod_{j=1}^{k} \Pr(X_{s_j} = 0) = \Pr(X_{s_1} = X_{s_2} = \ldots = X_{s_k} = 0) \geq \frac{1}{2}.$$

In order to lower bound $\sum_{i=1}^{n} \Pr(X_i = 0)$, we solve

$$\begin{array}{lll} \min & \sum_{j=1}^{n} & \Pr(X_j = 0) \\ \text{s.t.} & \prod_{j=1}^{k} & \Pr(X_{s_j} = 0) \geq \dfrac{1}{2} \quad \forall \, S = \{s_1, \ldots, s_k\} \subset \mathcal{N}. \end{array}$$

This is a convex optimization problem, as can be readily seen by substituting $\alpha_j = \log \Pr(X_j = 0)$, and can therefore be solved by choosing $\Pr(X_j = 0) = \left(\frac{1}{2}\right)^{\frac{1}{k}}$ for $j = 1, \ldots, n$. Thus $\sum_{j=1}^{n} \Pr(X_j = 0) \geq n \left(\frac{1}{2}\right)^{\frac{1}{k}} \geq n(1 - 1/n) = n - 1$.

Case III: There exists $i \in \mathcal{N}$ such that $\Pr(X_i = -) > 0$ and $\Pr(X_i = +) = 0$, and Case I does not hold.

This case is symmetric to Case II.

Case IV: For all $i \in \mathcal{N}$, $\Pr(X_i = +) = \Pr(X_i = -) = 0$.

We have $\sum_{j=1}^{n} \Pr(X_j = 0) > \sum_{j=2}^{n} \Pr(X_j = 0) = n - 1$. □

We are now in a position to prove the second part of Theorem 7. Let $D_k < 1 - \frac{k}{n}$, D_k rational, and $\left(1 - \frac{1}{n}\right)^k \leq \frac{1}{2}$, and let f_i, $i \in \mathcal{N}$ and $g_{\mathcal{K}}$, $\mathcal{K} \subseteq \mathcal{N}$, $\mathcal{K} \neq \emptyset$ be a code that achieves the rate-distortion vector $(R_k(D_k), D_1, \ldots, D_k, \ldots, D_n)$. Let $f_i,\ i \in \mathcal{N}$ have rate $R_k(D_k)$. We have

$$lR_k(D_k) \geq H(f_i),\ i \in \mathcal{N}. \tag{A.3}$$

Let $\hat{\mathbf{X}}^l_{\mathcal{S}}$ be the reconstruction when the source $\mathbf{X}^l$ is reconstructed from a set $\mathcal{S}$ of descriptions. Since D_k is finite, the decoder cannot make errors in its reconstruction (which would incur infinite distortion). Thus $\hat{\mathbf{X}}^l_{\mathcal{S}}$ must be an erased version of $\mathbf{X}^l$, *i.e.*, for all $t \in \{1, \ldots, l\}$, $\hat{X}_{\mathcal{S},t} = X_t$ or $\hat{X}_{\mathcal{S},t} = 0$. Then $\forall\ S = \{s_1, \ldots, s_k\} \subset \mathcal{N}, |S| = k$, we have

$$\begin{aligned}
H(f_{s_1} \ldots f_{s_k}) &\geq H(\hat{\mathbf{X}}^l_{\mathcal{S}}) \\
&\geq I(\mathbf{X}^l; \hat{\mathbf{X}}^l_{\mathcal{S}}) \\
&= H(\mathbf{X}^l) - H(\mathbf{X}^l | \hat{\mathbf{X}}^l_{\mathcal{S}}) \\
&= l - \sum_{t=1}^{l} H(X_t | \hat{\mathbf{X}}^l_{\mathcal{S}}, X_1, \ldots, X_{t-1}) \\
&\geq l - \sum_{t=1}^{l} H(X_t | \hat{X}_{\mathcal{S},t}) \\
&= l - \sum_{t=1}^{l} H(X_t | \hat{X}_{\mathcal{S},t} = 0) \cdot \Pr(\hat{X}_{\mathcal{S},t} = 0) \\
&= l - \sum_{t=1}^{l} \Pr(\hat{X}_{\mathcal{S},t} = 0) \\
&= l - \mathbf{E}\left[\sum_{t=1}^{l} \mathbf{1}_{\{\hat{X}_{\mathcal{S},t} = 0\}}\right] \\
&\geq l - lD_k = l(1 - D_k).
\end{aligned} \tag{A.4}$$

Thus

$$\begin{aligned}
I_k(f_{s_1}; \ldots; f_{s_k}) &= \sum_{j=1}^{k} H(f_{s_j}) - H(f_{s_1} \ldots f_{s_k}) \\
&\leq klR_k(D_k) - l(1 - D_k) = 0.
\end{aligned}$$

Let $\hat{\mathbf{X}}^l_{s_i}$ be the reconstruction when the decoder receives the s_i^{th} description only. Then $I_k(\hat{\mathbf{X}}^l_{s_1};\ldots;\hat{\mathbf{X}}^l_{s_k}) \leq I_k(f_{s_1};\ldots;f_{s_k}) = 0$ (Property 3) and so $I_k(\hat{X}_{s_1,t};\ldots;\hat{X}_{s_k,t}) = 0$, $t \in \{1,\ldots,l\}$. By Lemma 9, $\sum_{i=1}^n \Pr(\hat{X}_{it} = 0) \geq (n-1)$ for $t \in \{1,\ldots,l\}$. Thus

$$\begin{aligned} \frac{1}{l}\sum_{t=1}^{l}\sum_{i=1}^{n}\Pr(\hat{X}_{it} = 0) &\geq n-1 \\ \Rightarrow \max_i \left(\frac{1}{l}\sum_{t=1}^{l}\Pr(\hat{X}_{it} = 0)\right) &\geq 1 - \frac{1}{n}. \end{aligned}$$

This completes the proof.

A.6 Proof of Theorem 8

We establish two lemmas before proving Theorem 8.

Lemma 10. *Let $X_1, X_2,$ and X_3 be Bernoulli random variables such that $I(X_i;X_j) = 0, \forall\, i,j \in \{1,2,3\}, i \neq j$, and $\Pr(X_1 = X_2 = X_3 = 0) \geq \frac{1}{2}$. Let $p = \max(\Pr(X_1 = 0), \Pr(X_2 = 0))$. Then*

$$\Pr(X_3 = 0) \geq \frac{1}{2} + \frac{p(1-p)}{2p-1}.$$

Proof. If $p = 1$, then the conclusion follows directly from the hypothesis, so suppose that $p < 1$. Let p_i denote $\Pr(X_i = 0)$, $p(x_1,x_2,x_3)$ denote $\Pr(X_1 = x_1, X_2 = x_2, X_3 = x_3)$, and $p_{x_3|x_1,x_2}$ denote $\Pr(X_3 = x_3 | X_1 = x_1, X_2 = x_2)$. Let $q_0 = p_{0|0,0}$, $q_1 = p_{0|0,1}$, and $q_2 = p_{0|1,1}$. We thus have $p(0,0,0) = p_1p_2q_0$, $p(0,1,0) = p_1(1-p_2)q_1$, and $p(1,1,0) = (1-p_1)(1-p_2)q_2$. Then

$$\begin{aligned} \Pr(X_1 = 0, X_3 = 0) &= p(0,0,0) + p(0,1,0) \\ &= p_1(p_2q_0 + (1-p_2)q_1) \qquad \text{(A.5)} \end{aligned}$$

$$\begin{aligned}\Pr(X_2 = 1, X_3 = 0) &= p(0,1,0) + p(1,1,0) \\ &= (1-p_2)(p_1 q_1 + (1-p_1)q_2). \qquad \text{(A.6)}\end{aligned}$$

Since (X_1, X_3) and (X_2, X_3) are pairwise independent, we have, from (A.5) and (A.6),

$$\begin{aligned}\Pr(X_1 = 0, X_3 = 0) &= p_1 p_3 = p_1(p_2 q_0 + (1-p_2)q_1) \\ \Rightarrow p_3 &= p_2 q_0 + (1-p_2)q_1, \qquad \text{(A.7)} \\ \Pr(X_2 = 1, X_3 = 0) &= (1-p_2)p_3 \\ &= (1-p_2)(p_1 q_1 + (1-p_1)q_2) \\ \Rightarrow p_3 &= p_1 q_1 + (1-p_1)q_2. \qquad \text{(A.8)}\end{aligned}$$

From (A.7) and (A.8),

$$\begin{aligned}p_1 q_1 + (1-p_1)q_2 &= p_2 q_0 + (1-p_2)q_1 \\ \Rightarrow q_2 &= \frac{p_2 q_0 - (p_1 + p_2 - 1)q_1}{1-p_1}. \qquad \text{(A.9)}\end{aligned}$$

Since $p(0,0,0) \geq 1/2$ by hypothesis, we have $p_1 p_2 \geq 1/2$, and thus $p_1 + p_2 - 1 > 0$. Now since $q_2 \leq 1$, (A.9) gives

$$1 \geq \frac{p_2 q_0 - (p_1 + p_2 - 1)q_1}{1-p_1} \Rightarrow q_1 \geq \frac{p_2 q_0 - (1-p_1)}{p_1 + p_2 - 1}. \qquad \text{(A.10)}$$

Now

$$p(0,0,0) = p_1 p_2 q_0 \geq \frac{1}{2} \Rightarrow p_2 q_0 \geq \frac{1}{2p_1}. \qquad \text{(A.11)}$$

Assume without loss of generality that $p_1 \geq p_2$. Then $p_1 + p_2 \leq 2p_1$. Substituting this and (A.11) into (A.10) yields

$$q_1 \geq \frac{\frac{1}{2p_1} - 1 + p_1}{2p_1 - 1} = \frac{p_1}{2p_1 - 1} - \frac{1}{2p_1}. \qquad \text{(A.12)}$$

Upon substituting (A.11) and (A.12) into (A.7), we get

$$p_3 \geq \frac{1}{2p_1} + (1-p_2)\left(\frac{p_1}{2p_1 - 1} - \frac{1}{2p_1}\right)$$

$$\begin{aligned} &\geq \frac{1}{2p_1} + (1-p_1)\left(\frac{p_1}{2p_1-1} - \frac{1}{2p_1}\right) \\ &= \frac{1}{2} + \frac{p_1(1-p_1)}{2p_1-1} \end{aligned}$$

where the last inequality follows because $p_2 \leq p_1$ and $\frac{p_1}{2p_1-1} - \frac{1}{2p_1} > 0$. □

Corollary 1. *Let X_1, X_2, X_3 and X_4 be Bernoulli random variables such that $I(X_i; X_j) = 0$, $\forall\, i, j \in \{1,2,3,4\}, i \neq j$, and $\Pr(X_1 = X_2 = X_3 = X_4 = 0) \geq \frac{1}{2}$. Then*

$$\sum_{i=1}^{4} \Pr(X_i = 0) \geq 3.$$

Proof. Let $p_i = \Pr(X_i = 0)$. Assume WLOG that $p_1 \geq p_2 \geq p_3 \geq p_4$. Now $p_3 p_4 = \Pr(X_3 = X_4 = 0) \geq 1/2$ by hypothesis, which implies $p_3 \geq 1/\sqrt{2}$ and $p_4 \geq 1/2p_3$. Applying Lemma 10 to X_2, X_3, and X_4 gives $p_2 \geq \frac{1}{2} + \frac{p_3(1-p_3)}{2p_3-1}$. Thus

$$\begin{aligned} \sum_{i=1}^{4} p_i &= p_1 + p_2 + p_3 + p_4 \\ &\geq 2p_2 + p_3 + p_4 \\ &\geq 2\max\left(p_3, \frac{1}{2} + \frac{p_3(1-p_3)}{2p_3-1}\right) + p_3 + \frac{1}{2p_3} \\ &\geq \min_{x \in [\frac{1}{\sqrt{2}}, 1]} 2\max\left(x, \frac{1}{2} + \frac{x(1-x)}{2x-1}\right) + x + \frac{1}{2x}. \end{aligned}$$

Since $\frac{1}{2} + \frac{p_3(1-p_3)}{2p_3-1}$ is monotonically decreasing in p_3 for $p_3 \in (1/2, 1]$, it is easy to verify that

$$\max\left(x, \frac{1}{2} + \frac{x(1-x)}{2x-1}\right) = \begin{cases} x & \text{if } x \geq \frac{1}{2} + \frac{1}{\sqrt{12}} \\ \frac{1}{2} + \frac{x(1-x)}{2x-1} & \text{if } x \leq \frac{1}{2} + \frac{1}{\sqrt{12}}, \end{cases}$$

where $\frac{1}{2} + \frac{1}{\sqrt{12}}$ is the admissible solution to the equation $x = \frac{1}{2} + \frac{x(1-x)}{2x-1}$. Thus

$$\sum_{i=1}^{4} p_i \geq \min\left(\min_{x \in [\frac{1}{\sqrt{2}}, \frac{1}{2} + \frac{1}{\sqrt{12}}]} 2\left(\frac{1}{2} + \frac{x(1-x)}{2x-1}\right) + x + \frac{1}{2x},\right.$$

$$
\begin{aligned}
&\qquad\qquad\qquad \left.\min_{x\in[\frac{1}{2}+\frac{1}{\sqrt{12}},1]} 2x+x+\frac{1}{2x}\right) \\
&= \min\left(\min_{x\in[\frac{1}{\sqrt{2}},\frac{1}{2}+\frac{1}{\sqrt{12}}]} 1+\frac{1}{2x}+\frac{x}{2x-1},\right. \\
&\qquad\qquad\qquad \left.\min_{x\in[\frac{1}{2}+\frac{1}{\sqrt{12}},1]} 3x+\frac{1}{2x}\right) \\
&= \min(3,3) = 3,
\end{aligned}
$$

where the penultimate equality follows from the fact that $1+\frac{1}{2x}+\frac{x}{2x-1}$ is a monotonically decreasing in x for $x \in [\frac{1}{\sqrt{2}}, \frac{1}{2}+\frac{1}{\sqrt{12}}]$ and takes a minimum value of 3 at $x = \frac{1}{2}+\frac{1}{\sqrt{12}}$, and that $3x+\frac{1}{2x}$ is monotonically increasing in x for $x \in [\frac{1}{2}+\frac{1}{\sqrt{12}}, 1]$ and takes a minimum value of 3 at $x = \frac{1}{2}+\frac{1}{\sqrt{12}}$. □

The following lemma is similar to Lemma 9, but is adapted to the $n = 4$, $k = 2$ case, which is not covered by Lemma 9. Lemma 9 requires that n and k satisfy the inequality $\left(1-\frac{1}{n}\right)^k \leq \frac{1}{2}$, which is violated when $n = 4$ and $k = 2$. Indeed, much of the following proof is similar to that of Lemma 9, except for Cases II and III, where we use Corollary 1 to bypass the condition $\left(1-\frac{1}{n}\right)^k \leq \frac{1}{2}$ which is needed in Case II of the proof of Lemma 9.

Lemma 11. *Let $X_1, \ldots, X_4$ be erased versions of a uniform binary random variable X taking values in $\{+,-\}$. If $I(X_i; X_j) = 0,\ i,j \in \{1,\ldots,4\},\ i \neq j$, then*

$$\sum_{i=1}^{4} \Pr(X_i = 0) \geq 3.$$

Proof. The proof is very similar to that of Lemma 9, so we only summarize the argument here.

Case I: There exists $i \in \{1,2,3,4\}$ such that $\Pr(X_i = +) > 0$ and $\Pr(X_i = -) > 0$. Just as in the proof of Lemma 9, we have $\sum_{j=1}^{4} \Pr(X_j = 0) \geq 4-1 = 3$.

Case II: There exists $i \in \{1,2,3,4\}$ such that $\Pr(X_i = +) > 0$ and $\Pr(X_i = -) =$

0, and Case I does not hold.

Assume $i = 1$ WLOG. Then from (A.1), $\Pr(X_j = -) = 0$ for $j \in \{2, 3, 4\}$. Thus the X_j are effectively binary random variables such that $\Pr(X_1 = \ldots = X_4 = 0) \geq 1/2$. By Corollary 1, $\sum_{j=1}^{4} \Pr(X_j = 0) \geq 3$.

Case III: There exists $i \in \{1, 2, 3, 4\}$ such that $\Pr(X_i = -) > 0$ and $\Pr(X_i = +) = 0$, and Case I does not hold.

This case is analogous to Case II.

Case IV: For all $i \in \{1, 2, 3, 4\}$, $\Pr(X_i = +) = \Pr(X_i = -) = 0$.

We have $\sum_{j=1}^{4} \Pr(X_j = 0) > \sum_{j=2}^{4} \Pr(X_j = 0) = 4 - 1 = 3$. □

We are now in a position to prove Theorem 8. Let f_i, $i \in \mathcal{N}$ and $g_{\mathcal{K}}$, $\mathcal{K} \subseteq \mathcal{N}$ be a code that achieves $(\frac{1-D_2}{2}, D_1, D_2, D_3, D_4)$. Using the same argument as that in the proof of the second part of Theorem 7, we have for $i, j \in \{1, 2, 3, 4\}$, $i \neq j$ that $I(\mathbf{X}_i^l; \mathbf{X}_j^l) \leq I(f_i; f_j) = 0$ and thus $I(X_{it}; X_{jt}) = 0$ for all $t \in \{1, \ldots, l\}$. By Lemma 13, $\sum_{i=1}^{4} \Pr(X_{it} = 0) \geq 3$ for $t \in \{1, \ldots, l\}$. It follows that

$$\begin{aligned} \frac{1}{l} \sum_{t=1}^{l} \sum_{i=1}^{4} \Pr(X_{it} = 0) &\geq 3 \\ \Rightarrow \max_i \left(\frac{1}{l} \sum_{t=1}^{l} \Pr(X_{it} = 0) \right) &\geq \frac{3}{4}. \end{aligned}$$

This completes the proof.

A.7 Proof of Theorem 9

We establish two lemmas before proving Theorem 9.

Lemma 12. *Let $X_1, \ldots, X_n$ be Bernoulli random variables such that $I(X_i; X_j) = 0$*

$\forall\, i, j \in \mathcal{N},\ i \neq j$, and $\Pr(X_1 = X_2 = \ldots = X_n = 0) \geq \frac{1}{2}$. Then

$$\frac{1}{n}\sum_{i=1}^{n} \Pr(X_i = 0) \geq 1 - \frac{2}{n}.$$

Proof. Let p_i denote $\Pr(X_i = 0)$ and let $q_i = \Pr(X_i = 1) = 1 - p_i$. Since the X_i's are pairwise independent, we have

$$\begin{aligned} \mathbf{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] &= \frac{1}{n}\sum_{i=1}^{n} q_i \\ \text{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] &= \frac{1}{n^2}\sum_{i=1}^{n} \text{Var}(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} p_i q_i. \end{aligned}$$

Let $\alpha > \sqrt{\frac{2}{n^2}(\sum_{i=1}^{n} p_i q_i)}$. Then, by Chebyshev's inequality,

$$\begin{aligned} \Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} q_i\right| > \alpha\right) &\leq \frac{\text{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]}{\alpha^2} \\ &= \frac{\sum_{i=1}^{n} p_i q_i}{n^2 \alpha^2} < \frac{1}{2}. \end{aligned}$$

Let E_1 and E_2 be the events $|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} q_i| \leq \alpha$ and $X_1 = X_2 = \ldots = X_n = 0$, respectively. Then $\Pr(E_1) > \frac{1}{2}$, and $\Pr(E_2) \geq \frac{1}{2}$ by hypothesis. Since $\Pr(E_1) + \Pr(E_2) > 1$, $\Pr(E_1 \cap E_2) > 0$. This implies that

$$\frac{1}{n}\sum_{i=1}^{n} q_i \leq \alpha \Rightarrow \frac{1}{n}\sum_{i=1}^{n} p_i \quad \geq \quad 1 - \alpha.$$

Since α was arbitrary, this implies

$$\frac{1}{n}\sum_{i=1}^{n} p_i \quad \geq \quad 1 - \sqrt{\frac{2}{n^2}(\sum_{i=1}^{n} p_i q_i)}. \tag{A.13}$$

Moreover,

$$\frac{1}{n}\sum_{i=1}^{n} p_i q_i \quad \leq \quad \frac{1}{n}\sum_{i=1}^{n} q_i \leq \sqrt{\frac{2}{n^2}(\sum_{i=1}^{n} p_i q_i)}.$$

A little algebra gives

$$\sum_{i=1}^{n} p_i q_i \leq \sqrt{2\sum_{i=1}^{n} p_i q_i} \Rightarrow \sum_{i=1}^{n} p_i q_i \leq 2. \tag{A.14}$$

Substituting (A.14) into (A.13) yields

$$\frac{1}{n}\sum_{i=1}^{n} p_i \quad \geq \quad 1-\sqrt{\frac{2}{n^2}\cdot 2} = 1-\frac{2}{n}.$$

□

Lemma 13. *Let $X_1,\ldots,X_n$ be erased versions of a uniform binary random variable X taking values in $\{+,-\}$. If $I(X_i;X_j)=0,\ i,j\in\mathcal{N},\ i\neq j$, then*

$$\sum_{i=1}^{n}\Pr(X_i=0)\geq n-2.$$

Proof. We have Cases I, II, III, and IV as in the proof of Lemma 9. Cases I and IV are the same as those in Lemma 9, so we will only mention Cases II and III.

Case II: There exists $i\in\mathcal{N}$ such that $\Pr(X_i=+)>0$ and $\Pr(X_i=-)=0$ and Case I does not hold.

Assume $i=1$ WLOG. Then from (A.1), $\Pr(X_j=-)=0$ for $j\in\{2,\ldots,n\}$. Thus the X_j's are always erased when the binary source $X=-$, and so $\Pr(X_1=\ldots=X_n=0)\geq 1/2$. By Lemma 12, $\sum_{i=1}^{n}\Pr(X_i=0)\geq n-2$. The proof of Case III is analogous to the proof of Case II. □

We are now in a position to prove Theorem 9. Let f_i, $i\in\mathcal{N}$ and $g_{\mathcal{K}}$, $\mathcal{K}\subseteq\mathcal{N}$ be a code that achieves $(\frac{1-D_2}{2},D_1,D_2,\ldots,D_n)$. Using the same argument as that in the proof of the second part of Theorem 7, we have for $i,j\in\mathcal{N}$, $i\neq j$ that $I(\mathbf{X}_i^l;\mathbf{X}_j^l)\leq I(f_i;f_j)=0$ and thus $I(X_{it};X_{jt})=0$ for $t\in\{1,\ldots,l\}$. By Lemma 13, $\sum_{i=1}^{n}\Pr(X_{it}=0)\geq n-2$ for $t\in\{1,\ldots,l\}$. It follows that

$$\begin{aligned}\frac{1}{l}\sum_{t=1}^{l}\sum_{i=1}^{n}\Pr(X_{it}=0) &\geq n-2.\\ \Rightarrow \max_i\left(\frac{1}{l}\sum_{t=1}^{l}\Pr(X_{it}=0)\right) &\geq 1-\frac{2}{n}.\end{aligned}$$

This completes the proof.

A.8 A Random Coding Proof of Theorem 6

Like the MDS coding scheme for worst-case distortion, the random coding scheme consists of two parts - uncoded bits and an random binning component. The uncoded component is similar to the uncoded component of the MDS coding scheme. The difference lies in the encoded component; instead of encoding an erased version using an (n, k) systematic MDS code, the average-case distortion encoder randomly bins an erased version of the source and then sends bin indices to the decoder. The decoder outputs the uncoded bits as the source reconstruction if less than k descriptions are received. If k or more descriptions are received, the decoder uses the uncoded bits and the bin indices to decode the encoded erased version using typicality considerations. A formal description of the scheme follows.

Case I: $D_k \geq 1 - \frac{k}{n}$
Assume without loss of generality that D_k is rational (if D_k is irrational, then we can prove achievability for a sequence of rational distortions in $[1 - k/n, 1]$ converging to D_k and take limits). Then there exists a positive integer l' such that $l'R_k(D_k)$ is a positive integer. Choose a blocklength $l = \alpha n l'$, where α is any positive integer. Observe a length-l source sequence $\mathbf{X}^l$, and divide $\mathbf{X}^l$ into n disjoint parts such that each part contains $l/n = \alpha l'$ bits. (The division is the same regardless of the source realization.) Label the parts $\mathbf{X}_i,\ i \in \mathcal{N}$. Choose $lR_k(D_k)$ bits from each of the n parts (since $D_k \geq 1 - \frac{k}{n}$, $lR_k(D_k) \leq \frac{l}{n}$ and therefore $lR_k(D_k)$ bits can be chosen from each part). Denote by $\mathbf{Y}_i$ the set of $lR_k(D_k)$ bits chosen from $\mathbf{X}_i$. Transmit $\mathbf{Y}_i$ uncoded over the i^{th} channel.

The decoding is trivial. If m descriptions, say $(\mathbf{Y}_1, \ldots, \mathbf{Y}_m)$, are received, output $\hat{\mathbf{X}}^l_{\mathbf{m}}$ as the reconstruction of $\mathbf{X}^l$, where $\hat{\mathbf{X}}^l_{\mathbf{m}}$ is such that the $mlR_k(D_k)$ bits corresponding to $(\mathbf{Y}_1, \ldots, \mathbf{Y}_m)$ are non-erased and the other $(l - mlR_k(D_k))$ bits are erasures. The distortion, therefore, is $(l - mlR_k(D_k))/l = 1 - mR_k(D_k)$. When k descriptions are received, the distortion is $1 - kR_k(D_k) = D_k$. Thus $\tilde{\mathbf{R}} \in \mathcal{RD}_{avg}$, and therefore also lies in $\overline{\mathcal{RD}}_{avg}$.

Case II: $D_k < 1 - \frac{k}{n}$

The scheme for this case is an extension of the scheme for Case I. It has two components; random binning and transmission of uncoded source bits. An erased version of every source sequence is binned separately at each encoder. The observed source string is divided into n disjoint parts. Each uncoded part is then sent on one of the n channels along with the corresponding bin index of the erased version of the source. If less than k descriptions are received, the decoder outputs a partial reconstruction based solely on the uncoded parts; if k or more descriptions are received, the decoder outputs a reconstruction based on the uncoded parts and the bin indices.

Assume again that D_k is rational. Choose $\epsilon > 0$, and define $R' = R_k(D_k) - 1/n + \epsilon$. Since D_k is rational, there exists a positive integer l' such that $l'D_k/(n-k)$ is an integer. Choose a blocklength $l = \alpha n l'$, where α is any positive integer.

Random binning: Construct n sets of bins such that every set contains $2^{lR'}$ bins. For every length-l source string $\mathbf{x}^l \in \mathcal{X}^l$, construct an erased version as follows. Divide $\mathbf{x}^l$ into n disjoint parts such that each part contains $l/n = \alpha l'$ bits (the division is done identically for all source sequences). For each part, replace the last $lD_k/(n-k)$ bits by erasures (since $D_k < 1 - \frac{k}{n}$, each part contains $l/n > lD_k/(n-k)$ bits). Assign the resulting erased version $\mathbf{x_e}^l$ uniformly at random, and independently from other strings, to one of the $2^{lR'}$ bins in the i^{th} set, for all $i \in \mathcal{N}$. The assignment is done only once for each erased version. This is important because multiple source strings can have the same erased version. Denote the assignments by Γ_i.

Encoding: Let $\mathbf{X}^l$ be the observed source sequence. Divide $\mathbf{X}^l$ into n disjoint parts each containing l/n bits as described above. Label the parts $\mathbf{X}_i$, $i \in \mathcal{N}$. Let $B_i = \Gamma_i(\mathbf{X}^l)$ be the index of the bin containing the erased version of $\mathbf{X}^l$ in the i^{th} bin set. Transmit $(\mathbf{X}_i, B_i)$ over the i^{th} channel.

Decoding: If m descriptions, say $\{(\mathbf{X}_1, B_1), \ldots, (\mathbf{X}_m, B_m)\}$, are received, where $m < k$, output $\hat{\mathbf{X}}^l_{\mathbf{m}}$ as the reconstruction of $\mathbf{X}^l$, where $\hat{\mathbf{X}}^l_{\mathbf{m}}$ is such that the ml/n bits corresponding to $(\mathbf{X}_1, \ldots, \mathbf{X}_m)$ are non-erased and the other $(l - ml/n)$ bits are erasures. If $m > k$ descriptions are received, say $\{(\mathbf{X}_1, B_1), \ldots, (\mathbf{X}_m, B_m)\}$, choose any k descriptions, say $\{(\mathbf{X}_1, B_1), \ldots, (\mathbf{X}_k, B_k)\}$, and search the bins $(B_1, \ldots, B_k)$ for a sequence $\mathbf{Y}$ such that $\Gamma_i(\mathbf{Y}) = B_i$, $i = 1, \ldots, k$, and $\mathbf{Y}$ is consistent with the partially revealed source string $(\mathbf{X}_1, \ldots, \mathbf{X}_k)$. Output $\hat{\mathbf{X}}^l_{\mathbf{m}} = \{(\mathbf{X}_1, \ldots, \mathbf{X}_m)\} \cup \{\mathbf{Y}\}$ as the reconstruction of $\mathbf{X}^l$. (Thus the non-erased bits in $\hat{\mathbf{X}}^l_{\mathbf{m}}$ are the bits revealed by $(\mathbf{X}_1, \ldots, \mathbf{X}_m)$ or by the erased version $\mathbf{Y}$, or both.) There is guaranteed to be at least one such sequence $\mathbf{Y}$ in the bins indexed by $B_1, \ldots, B_k$. If there is more than one such sequence, output the non-erased portion $(\mathbf{X}_1, \ldots, \mathbf{X}_m)$ as the reconstruction of $\mathbf{X}^l$.

Error analysis: We say an error $E_{\mathcal{S}}$ has occurred at the decoder if, for a set $\mathcal{S} = \{s_1, \ldots, s_k\}$ of k descriptions, there exists an erased version $\mathbf{Y} \neq \mathbf{X_e}^l$ such that $\Gamma_{s_i}(\mathbf{Y}) = \Gamma_{s_i}(\mathbf{X_e}^l)$ for all $s_i \in \mathcal{S}$ and $\mathbf{Y}$ is consistent with $(\mathbf{X}_{\mathbf{s_1}}, \ldots, \mathbf{X}_{\mathbf{s_k}})$. Let $\mathcal{C}_{\mathcal{S}}$ be the set of erased versions that are consistent with $(\mathbf{X}_{\mathbf{s_1}}, \ldots, \mathbf{X}_{\mathbf{s_k}})$. Define

$E = \bigcup_{\mathcal{S},|\mathcal{S}|=k} E_{\mathcal{S}}$. We bound $\Pr(E)$ as follows.

$$
\begin{aligned}
&\Pr(E) \\
&\leq \sum_{\mathcal{S},|\mathcal{S}|=k} \Pr(E_{\mathcal{S}}) \\
&= \sum_{\mathcal{S},|\mathcal{S}|=k} \Pr(\exists \mathbf{Y} \neq \mathbf{X_e}^l, \mathbf{Y} \in \mathcal{C}_{\mathcal{S}} : \Gamma_{s_i}(\mathbf{Y}) = \Gamma_{s_i}(\mathbf{X_e}^l) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall s_i \in \mathcal{S}) \\
&= \sum_{\mathbf{x}^l} p(\mathbf{x}^l) \sum_{\mathcal{S},|\mathcal{S}|=k} \Pr(\exists \mathbf{Y} \neq \mathbf{x_e}^l, \mathbf{Y} \in \mathcal{C}_{\mathcal{S}} : \\
&\qquad\qquad\qquad\qquad \Gamma_{s_i}(\mathbf{Y}) = \Gamma_{s_i}(\mathbf{x_e}^l) \forall s_i \in \mathcal{S} | \mathbf{X}^l = \mathbf{x}^l) \\
&\leq \sum_{\mathbf{x}^l} p(\mathbf{x}^l) \sum_{\mathcal{S},|\mathcal{S}|=k} \sum_{\substack{\mathbf{y} \neq \mathbf{x_e}^l \\ \mathbf{y} \in \mathcal{C}_{\mathcal{S}}}} \Pr(\Gamma_{s_i}(\mathbf{y}) = \Gamma_{s_i}(\mathbf{x_e}^l) \\
&\qquad\qquad\qquad\qquad\qquad\qquad \forall s_i \in \mathcal{S} | \mathbf{X}^l = \mathbf{x}^l) \\
&\leq \sum_{\mathbf{x}^l} p(\mathbf{x}^l) \sum_{\mathcal{S},|\mathcal{S}|=k} 2^{-klR'} |\mathcal{C}_{\mathcal{S}}| \\
&= \sum_{\mathbf{x}^l} p(\mathbf{x}^l) \sum_{\mathcal{S},|\mathcal{S}|=k} 2^{-kl(\frac{1-D_k}{k} - \frac{1}{n} + \epsilon)} \cdot 2^{(n-k)(\frac{l}{n} - l\frac{D_k}{n-k})} \\
&= \sum_{\mathbf{x}^l} p(\mathbf{x}^l) \sum_{\mathcal{S},|\mathcal{S}|=k} 2^{-lk\epsilon} \\
&\leq \binom{n}{k} 2^{-lk\epsilon}.
\end{aligned}
$$

We now show that for any $\epsilon > 0$, the $(n+1)$-tuple $(R_k(D_k)+\epsilon, 1-\frac{1}{n}+\epsilon, 1-\frac{2}{n}+\epsilon, \ldots, 1-\frac{k-1}{n}+\epsilon, D_k+\epsilon, (\frac{n-k-1}{n-k})D_k+\epsilon, (\frac{n-k-2}{n-k})D_k+\epsilon, \ldots, (\frac{1}{n-k})D_k+\epsilon, \epsilon)$ is achievable, and thus $\hat{\mathbf{R}} \in \overline{\mathcal{RD}}_{avg}$. Fix $\epsilon > 0$ and define R' as above. In our scheme, any description $(\mathbf{X}_i, B_i)$ has rate $R = 1/n + R'$, where $1/n$ is the rate due to $\mathbf{X}_i$ and R' is the rate due to binning. Thus $R = 1/n + (R_k(D_k) - 1/n + \epsilon) = R_k(D_k) + \epsilon$. Moreover, if $m < k$ descriptions are received, the decoder outputs ml/n bits as revealed by the m descriptions and the other $(l - ml/n)$ bits as erasures. Thus $D_m = 1 - m/n < 1 - m/n + \epsilon$. If k descriptions are received, say $\mathcal{S} = \{s_1, \ldots, s_k\}$, the decoder either outputs an erased version of the correct source sequence if $E_{\mathcal{S}}^c$ occurs, or outputs $(\mathbf{X_{s_1}}, \ldots, \mathbf{X_{s_k}})$ if $E_{\mathcal{S}}$ occurs. If $E_{\mathcal{S}}^c$ occurs, then the decoder receives kl/n bits uncoded from the k descriptions, and is able to figure out a further $(n-k)(l/n - lD_k/(n-k)) = l(1 - k/n - D_k)$ bits by using the bin indices to decode the erased version of the source sequence. Hence the maximum per-letter distortion over sets of k descriptions is $1 - (k/n + 1 - k/n - D_k) = D_k$ if E^c occurs, and $1 - k/n$ if E occurs. Let $d_{\mathcal{S},\mathbf{x}}$ be the per-letter distortion achieved using the set $\mathcal{S}$ of descriptions if the observed source string is $\mathbf{x}^l$. Thus

$$\begin{aligned}
&\mathbf{E}_{f,g} \max_{\mathcal{S},|\mathcal{S}|=k} \mathbf{E}_{\mathbf{X}}[d_{\mathcal{S}},\mathbf{x}] \\
&\leq \mathbf{E}_{f,g}\mathbf{E}_{\mathbf{X}}[\max_{\mathcal{S},|\mathcal{S}|=k} d_{\mathcal{S}},\mathbf{x}] \\
&= \mathbf{E}_{f,g}\mathbf{E}_{\mathbf{X}}\left[\left(1 - \frac{k}{n}\right) \cdot 1_E + D_k \cdot 1_{E^c}\right] \\
&= \left(1 - \frac{k}{n}\right) \Pr(E) + D_k(1 - \Pr(E)) \\
&= \left(1 - \frac{k}{n} - D_k\right) \Pr(E) + D_k \\
&\leq \left(1 - \frac{k}{n} - D_k\right) \left[\binom{n}{k} 2^{-kl\epsilon}\right] + D_k,
\end{aligned}$$

which can be made smaller than $D_k + \epsilon$ by letting $\alpha \to \infty$. Thus $D_k + \epsilon$ is achievable for some sufficiently large l. If $m > k$ descriptions are received, then the decoder receives ml/n bits uncoded, and is able to figure out a further $(n-m)(l/n - lD_k/(n-k))$ bits by decoding the binned erased version. Thus, if E^c occurs, the maximum per-letter distortion is $1 - m/n - ((n-m)/n - (n - m)D_k/(n-k)) = (\frac{n-m}{n-k})D_k$, and by the same analysis as above, a distortion of $(\frac{n-m}{n-k})D_k + \epsilon$ can be achieved for some sufficiently large l. This completes the proof.

A.9 Proof of Lemma 1

For any $t \in \{1, \ldots, l\}$, we have exactly one of the following four cases:

Case I: $\exists\, i \in \mathcal{N}$ s.t. $\Pr(\tilde{X}_{it}(X) = +) > 0$ and $\Pr(\tilde{X}_{it}(X) = -) > 0$.

Case II: $\exists\, i \in \mathcal{N}$ s.t. $\Pr(\tilde{X}_{it}(X) = +) > 0$ and $\Pr(\tilde{X}_{it}(X) = -) = 0$, and Case I does not hold.

Case III: $\exists\, i \in \mathcal{N}$ s.t. $\Pr(\tilde{X}_{it}(X) = -) > 0$ and $\Pr(\tilde{X}_{it}(X) = +) = 0$, and Case I does not hold.

Case IV: $\forall\, i \in \mathcal{N},\ \Pr(\tilde{X}_{it}(X) = +) = \Pr(\tilde{X}_{it}(X) = -) = 0$.

Let $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$ and $\mathcal{B}_4$ be the sets of $t \in \{1, \ldots, l\}$ satisfying Cases I, II, III and IV, respectively. Moreover, let $|\mathcal{B}_1| = b_1$, $|\mathcal{B}_2| = b_2$, $|\mathcal{B}_3| = b_3$ and $|\mathcal{B}_4| = b_4$. Then $b_1 + b_2 + b_3 + b_4 = l$. Now consider a source string $(\mathbf{x}^*)^l$ such that $x_t^* = -$ if $t \in \mathcal{B}_2$ and $x_t^* = +$ if $t \in \mathcal{B}_3$. We have

$$\begin{aligned}
&\max_{\mathbf{x}^l \in \mathcal{X}^l} \sum_{i=1}^{n} \left[\frac{1}{l} \sum_{t=1}^{l} d(x_t, \tilde{X}_{it}(x)) \right] \\
&\geq \sum_{i=1}^{n} \frac{1}{l} \sum_{t=1}^{l} d(x_t^*, \tilde{X}_{it}(x^*)) \\
&= \frac{1}{l} \sum_{t \in \mathcal{B}_1} \sum_{i=1}^{n} d(x_t^*, \tilde{X}_{it}(x^*)) + \frac{1}{l} \sum_{t \in \mathcal{B}_2} \sum_{i=1}^{n} d(x_t^*, \tilde{X}_{it}(x^*)) \\
&\quad + \frac{1}{l} \sum_{t \in \mathcal{B}_3} \sum_{i=1}^{n} d(x_t^*, \tilde{X}_{it}(x^*)) + \frac{1}{l} \sum_{t \in \mathcal{B}_4} \sum_{i=1}^{n} d(x_t^*, \tilde{X}_{it}(x^*)).
\end{aligned}$$

Consider now $t \in \mathcal{B}_1$. Since $\tilde{X}_{1t}(X), \ldots, \tilde{X}_{nt}(X)$ are erased versions of the same binary random variable X_t, they can never disagree in the source symbol they reveal. We therefore have $\Pr(\tilde{X}_{it}(X) = +, \tilde{X}_{jt}(X) = -) = 0$, $j \in \mathcal{N},\ j \neq i$. Since $\tilde{X}_{it}(X)$ and $\tilde{X}_{jt}(X)$, $i, j \in \mathcal{N}$, $i \neq j$, are pairwise independent, we have

$$\begin{aligned}
&\Pr(\tilde{X}_{it}(X) = +) \cdot \Pr(\tilde{X}_{jt}(X) = -) \\
&= \Pr(\tilde{X}_{it}(X) = +, \tilde{X}_{jt}(X) = -) = 0
\end{aligned}$$

$$\Rightarrow \Pr(\tilde{X}_{jt}(X) = -) = 0, \tag{A.15}$$

since $\Pr(\tilde{X}_{it}(X) = +) > 0$. Repeating the same analysis with $\Pr(\tilde{X}_{it}(X) = -, \tilde{X}_{jt}(X) = +)$ yields $\Pr(\tilde{X}_{jt}(X) = +) = 0$. Thus $\Pr(\tilde{X}_{jt}(X) = 0) = 1$ for all $j \in \mathcal{N},\ j \neq i$, and therefore $\tilde{X}_{jt}(x^*) = 0$ for all $j \in \mathcal{N},\ j \neq i$. Similarly, it follows from (A.15) that $\Pr(\tilde{X}_{jt}(X) = -) = 0$ for $j \in \mathcal{N},\ j \neq i$ if $t \in \mathcal{B}_2$ and $\Pr(\tilde{X}_{jt}(X) = +) = 0$ for $j \in \mathcal{N},\ j \neq i$ if $t \in \mathcal{B}_3$. Thus by construction, $\tilde{X}_i^l(x^*)$, $i \in \mathcal{N}$, must have $\tilde{X}_{it}(x^*) = 0$ for $t \in \mathcal{B}_2 \cup \mathcal{B}_3 \cup \mathcal{B}_4$. It follows that

$$\begin{aligned}
&\max_{\mathbf{x}^l \in \mathcal{X}^l} \sum_{i=1}^{n} \left[\frac{1}{l} \sum_{t=1}^{l} d(x_t, \tilde{X}_{it}(x)) \right] \\
&\geq \frac{1}{l} \sum_{t \in \mathcal{B}_1} \sum_{i=1}^{n} 1_{(\tilde{X}_{it}(x^*)=0)} + \frac{1}{l} \sum_{t \in \mathcal{B}_2} \sum_{i=1}^{n} 1_{(\tilde{X}_{it}(x^*)=0)} \\
&\qquad + \frac{1}{l} \sum_{t \in \mathcal{B}_3} \sum_{i=1}^{n} 1_{(\tilde{X}_{it}(x^*)=0)} + \frac{1}{l} \sum_{t \in \mathcal{B}_4} \sum_{i=1}^{n} 1_{(\tilde{X}_{it}(x^*)=0)} \\
&\geq \frac{1}{l} b_1(n-1) + \frac{1}{l} b_2 n + \frac{1}{l} b_3 n + \frac{1}{l} b_4 n \\
&= \frac{1}{l}(nl - b_1) \\
&= n - \frac{b_1}{l} \geq n - 1.
\end{aligned}$$

This completes the proof.

APPENDIX B

CHAPTER 3: PROOFS

B.1 Proof of Theorem 13

This bound differs only slightly from the outer bound proposed in [44] and much of the proof is similar to that in [44]. Suppose $(\mathbf{R}, \mathbf{D})$ is achievable. Let $f_1^{(l)}, \ldots, f_n^{(l)}$ be encoders and $(g_{\mathcal{K}}^j)^l,\ \mathcal{K} \subseteq \mathcal{N}$ be decoders satisfying (3.3). Take any Z in ψ and augment the sample space to include Z^l so that $(Z_t, Y_{0,t}, \mathbf{Y}_{\mathcal{N},t}, Y_{n+1,t})$ is independent over $t \in \{1, \ldots, l\}$. Next let T be uniformly distributed over $\{1, \ldots, l\}$ and independent of Z^l, Y_0^l, $\mathbf{Y}_{\mathcal{N}}^l$ and Y_{n+1}^l. Then define

$$\begin{aligned}
Z &= Z_T \\
Y_0 &= Y_{0,T} \\
Y_i &= Y_{i,T} \text{ for } i \in \mathcal{N} \\
Y_{n+1} &= Y_{n+1,T} \\
U_i &= \left(f_i^{(l)}(Y_i^l), Z_{1:T-1}, \{Y_{n+1}^l\} \backslash \{Y_{n+1,T}\}\right) \text{ for } i \in \mathcal{N} \\
V_j &= V_{j,T} \text{ for } j = 1, \ldots, J \\
W &= (\{Z^l\} \backslash \{Z_T\}, \{Y_{n+1}^l\} \backslash \{Y_{n+1,T}\}).
\end{aligned}$$

It can be verified that $\gamma = (\mathbf{U}_{\mathcal{N}}, V_1, \ldots, V_j, W, T)$ is in Γ_o and that, together with Y_0, $\mathbf{Y}_{\mathcal{N}}$, Y_{n+1}, and Z, it satisfies the Markov coupling. It suffices to show that $(\mathbf{R}, \mathbf{D})$ is in $\mathcal{RD}_o(Z, \gamma)$. Note that (3.3) implies, for $j = 1, \ldots, J$,

$$D_{k,j} \geq \max_{\mathcal{K}:|\mathcal{K}|=k} \mathbf{E}[d_j(Y_{0,T}, \mathbf{Y}_{\mathcal{K},T}, Y_{n+1,T}, V_{j,T})],$$

i.e.,

$$D_{k,j} \geq \max_{\mathcal{K}:|\mathcal{K}|=k} \mathbf{E}[d_j(Y_0, \mathbf{Y}_{\mathcal{K}}, Y_{n+1}, V_j)].$$

Second, by the cardinality bound on entropy and the fact that conditioning never increases entropy,

$$\begin{aligned} l\sum_{i\in\mathcal{K}} R_i &\geq H\left(\left(f_i^{(l)}(Y_i^l)\right)_{i\in\mathcal{K}}\right) \\ &= I\left(Z^l, \mathbf{Y}_{\mathcal{K}}^l; \left(f_i(Y_i^l)\right)_{i\in\mathcal{K}} \middle| Y_{n+1}^l\right). \end{aligned} \tag{B.1}$$

By the chain rule for mutual information,

$$\begin{aligned} &I\left(Z^l, \mathbf{Y}_{\mathcal{K}}^l; \left(f_i(Y_i^l)\right)_{i\in\mathcal{K}} \middle| Y_{n+1}^l\right) \\ &\quad = I\left(Z^l; \left(f_i(Y_i^l)\right)_{i\in\mathcal{K}} \middle| Y_{n+1}^l\right) \\ &\qquad + I\left(\mathbf{Y}_{\mathcal{K}}^l; \left(f_i(Y_i^l)\right)_{i\in\mathcal{K}} \middle| Z^l, Y_{n+1}^l\right). \end{aligned}$$

The rest of the proof is similar to that in [44]. The main difference between this proof and the proof in [44] is that here we do not condition on $\left(f_i(Y_i^l)\right)_{i\in\mathcal{K}^c}$ in (B.1). Taking the maximum over this bound and the bound in [44] yields the desired outer bound.

B.2 Proof of Lemma 2

Assume WLOG that $\mathcal{K} = \{1, \ldots, m\}$. For each possible realization (w,t) of (W,T), let

$$D_{w,t} = E[d^\lambda(X, \hat{X}_{\mathcal{K}})|W = w, T = t].$$

Let $S = \{(w,t) : D_{w,t} \leq \sqrt{\lambda}\}$. Then by Markov's inequality,

$$\Pr((W,T) \notin S) \leq \frac{\tilde{D}}{\sqrt{\lambda}} \leq \delta. \tag{B.2}$$

In particular, $\Pr((W,T) \in S) > 0$. Also, for any $(w,t) \in S$,

$$\frac{32m}{p(1-p)}\left(\frac{2D_{w,t}}{\lambda}\right)^{1/m} \leq \delta.$$

Thus, by Lemma 6 in [44], if $(w, t) \in S$,

$$\frac{1}{m}\sum_{i=1}^{m} I(Y_i; U_i | X, W = w, T = t) \\ \geq g\left((D_{w,t} + \delta)^{1/m}\right) + 2\delta \log \frac{\delta}{5}.$$

By averaging over $(w, t) \in S$ and invoking Corollary 1 in [44], we obtain

$$\sum_{(w,t)\in S} \frac{1}{m}\sum_{i=1}^{m} I(Y_i; U_i | X, W = w, T = t) \cdot \frac{\Pr(W = w, T = t)}{\Pr((W, T) \in S)} \\ \geq g((\tilde{D} + \delta)^{1/m}) + 2\delta \log \frac{\delta}{5}.$$

Therefore, $\frac{1}{m}\sum_{i=1}^{m} I(Y_i; U_i | X, W, T)$

$$\begin{aligned} &\geq \left[g((\tilde{D} + \delta)^{1/m}) + 2\delta \log \frac{\delta}{5}\right] \cdot \Pr((W, T) \in S) \\ &\geq \left[g((\tilde{D} + \delta)^{1/m}) + 2\delta \log \frac{\delta}{5}\right] (1 - \delta) \\ &= g((\tilde{D} + \xi(\tilde{D}, \delta))^{1/m}) \end{aligned}$$

for some continuous $\xi \geq 0$ satisfying $\xi(\tilde{D}, 0) = 0$. It follows from this and constraint *(iii)* of the lemma that $g(D^{1/m}) \geq g((\tilde{D} + \xi(\tilde{D}, \delta))^{1/m})$. From the monotonicity of $g(D^{1/m})$ in D (Corollary 1 in [44]), we obtain $\tilde{D} + \xi(\tilde{D}, \delta) \geq D$. Thus $\tilde{D} \geq D - \xi(\tilde{D}, \delta)$, completing the proof.

APPENDIX C

CHAPTER 5: PROOFS

C.1 Proof: $\mathcal{R}_{MLD} = \mathcal{R}_{MLD,blk}$

It can be readily shown that $\mathcal{R}_{MLD,blk} \subset \mathcal{R}_{MLD}$, since block-error probability dominates symbol-error probability. Consider now the other direction. Suppose $(R_1, R_2, R_3) \in \mathcal{R}_{MLD}$. Fix $\epsilon > 0$, and consider a sequence of codes, indexed by the blocklength ℓ, that satisfies Conditions (5.3) and (5.4). Condition (5.4) implies that the codes achieve a Hamming distortion of less than ϵ, which in turn implies that the encoders can transmit a minuscule amount of rate to allow the decoders to recover the source completely (and hence achieve $P_{\mathrm{v}} \leq \epsilon$). Consider how much additional rate Encoder f_i needs to transmit to allow the decoder to reconstruct X with a vanishingly small block-error probability. We can compute $H(X^\ell | f_i(X^\ell, Y^\ell))$ as follows. Consider $H(X_t | f_i(X^\ell, Y^\ell))$, $t \in \{1, \ldots, \ell\}$. Let $\epsilon_t = \Pr(\hat{X}_t \neq X_t | f(X^\ell, Y^\ell))$. Then, by Fano's inequality,

$$H(X_t | f_i(X^\ell, Y^\ell)) \leq H(X_t | \hat{X}_t) \leq h(\epsilon_t) + \epsilon_t \log(|\mathcal{X}| - 1).$$

Thus

$$\begin{aligned}
\ell^{-1} H(X^\ell | f_i(X^\ell, Y^\ell)) &\leq \ell^{-1} \sum_{t=1}^{\ell} H(X_t | f_i(X^\ell, Y^\ell)) \\
&\leq \ell^{-1} \sum_{t=1}^{\ell} [h(\epsilon_t) + \epsilon_t \log(|\mathcal{X}| - 1)] \\
&\leq h\left(\ell^{-1} \sum_{t=1}^{\ell} \epsilon_t\right) + \ell^{-1} \sum_{t=1}^{\ell} \epsilon_t \log(|\mathcal{X}| - 1) \\
&\leq h(\epsilon) + \epsilon \log(|\mathcal{X}| - 1),
\end{aligned}$$

where the penultimate inequality follows from the concavity of $h(\cdot)$ and the last inequality follows from (5.4). By the same analysis, we have

$$\ell^{-1}H(Y^\ell|f_i(X^\ell, Y^\ell), f_j(X^\ell, Y^\ell)) \leq h(\epsilon) + \epsilon \log(|\mathcal{Y}| - 1)$$

for $1 \leq i < j \leq 3$, where f_i and f_j are the original symbol-error MLD encoders.

Define $U = X^\ell$ and $V = f_i(X^\ell, Y^\ell)$. Consider a block-error encoder $f_{i,blk}$ and the corresponding decoder. If the decoder has access to V, then the block-error encoder can use a fixed-rate lossless code for U (treating V as side-information at both the encoder and decoder) that communicates U to the decoder with vanishing block-error probability at rate $H(U|V)$. Suppose the block-error encoder observes m length-ℓ sequences of the source (say $U^m = (X_1^\ell, \ldots, X_m^\ell)$), and uses the original symbol-error encoder f_i to generate m messages $V^m = (f_i(X_1^\ell, Y_1^\ell), \ldots, f_i(X_m^\ell, Y_m^\ell))$. The block-error encoder transmits V_m to the decoder, and then uses a fixed-rate lossless code with rate $H(U|V)$ and blocklength m to encode and transmit U^m, treating V^m as side-information. By making m sufficiently large, the block-error probability of losslessly decoding U^m can be made less than ϵ. The block-error encoder therefore transmits $m+1$ messages to the decoder: m messages generated from the symbol-error encoder, and an additional message generated from a fixed-rate lossless code.

Likewise, every pair of encoders can encode $Y^{m\ell}$, separately from $X^{m\ell}$, using a fixed-rate lossless code, and then each encoder can transmit disjoint halves of the resulting codeword. With this scheme, the rate of Encoder $f_{i,blk}$ will be

$$(R_i + \epsilon) + (h(\epsilon) + \epsilon \log(|\mathcal{X}| - 1)) + \delta_X + (h(\epsilon) + \epsilon \log(|\mathcal{Y}| - 1)) + \delta_Y.$$

Making m sufficiently large would guarantee that at this rate the block-error probability of decoding $X^{m\ell}$ and $Y^{m\ell}$ can be made less than ϵ for all blocklengths

$m\ell$. For blocklengths that are not a multiple of ℓ, the source symbols that do not form a full length-ℓ block can be transmitted uncoded. The extra rate incurred will tend to zero as m becomes large. In this way, the block-error probability can be made less than ϵ for all sufficiently large blocklengths.

BIBLIOGRAPHY

[1] eDonkey. http://www.edonkey2000.com/index.html.

[2] KaZaA. http://www.kazaa.com/.

[3] B. Cohen, "Incentives Build Robustness in BitTorrent," May 2003. http://bitconjurer.org/BitTorrent/bittorrentecon.pdf.

[4] F. Mathieu and J. Reynier, "Missing Piece Issue Missing Piece Issue and Upload Strategies in Flashcrowds and P2P-assisted Filesharing," in *Proc. AICT/ICIW*, Guadeloupe, French Caribbean, February 2006.

[5] J. A. Pouwelse, P. Garbacki, D. H. J. Epema, and H. J. Sips, "A Measurement Study of the BitTorrent Peer-to-peer File-sharing System," Technical Report PDS-2004-003, Delft University of Technology, The Netherlands, April 2004.

[6] M. Izal, G. Urvoy-Keller, E. W. Biersack, P. Felber, A. Al Hamra, and L. Garcés-Erice, "Dissecting BitTorrent: Five Months in a Torrents Lifetime," *PANM*, pp. 1–11, April 2004.

[7] D. Qiu and R. Srikant, "Modeling and Performance Analysis of BitTorrent-like Peer-to-Peer Networks," *SIGCOMM*, vol. 34, no. 4, pp. 367-378, October 2004.

[8] A. Bharambe and C. Herley, "Analyzing and Improving a BitTorrent Network's Performance Mechanisms," in *Proc. IEEE Infocom*, Barcelona, Spain, pp. 1–12, April 2006.

[9] T. Ho, R. Koetter, M. Médard, D. R. Karger, and M. Effros, "The Benefits of Coding over Routing in a Randomized Setting,", in *Proc. IEEE ISIT*, Yokohama, Japan, pp. 442–447, June 2003.

[10] C. Fragouli and E. Soljanin, "Network Coding Applications," *Foundations and Trends in Networking*, vol. 2, no. 2, pp. 135–269, 2007.

[11] C. Wu and B. Li, "rStream: Resilient Peer-to-Peer Streaming with Rateless Codes," in *Proc. ACM Multimedia*, November 2005.

[12] S. Deb, M. Médard, and C. Choute, "Algebraic Gossip: A Network Coding Approach to Optimal Multiple Rumor Mongering," *IEEE/ACM Trans.*

Networking, Special Issue on Networking and Information Theory, vol. 14, pp. 2486–2507, June 2006.

[13] K. Nguyen, T. Nguyen, and S. Cheung, “Peer-to-Peer Streaming with Hierarchical Network Coding,” in *Proc. IEEE ICME*, Beijing, China, pp. 396-399, July 2007.

[14] M. Wang and B. Li, “R2: Random push with Random Network Coding in Live Peer-to-Peer Streaming,” *IEEE JSAC Special Issue on Advances in Peer-to-Peer Streaming Systems*, vol. 25, no. 9, pp. 1655–1666, December 2007.

[15] C. Gkantsidis and P. Rodriguez, “Network Coding for Large Scale Content Distribution,” in *Proc. IEEE INFOCOM*, Miami, Florida, March 2005.

[16] M. Luby, “Lt Codes,” in *IEEE FOCS*, pp. 271-282, 2002.

[17] A. Shokrollahi, “Raptor Codes,” *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2551-2567, 2006.

[18] P. Maymounkov and D. Mazieres, “Rateless Codes and Big Downloads,” in *Proc. IPTPS*, Berkeley, California, February, 2003.

[19] P. Maymounkov, “Online Codes,” *New York University Technical Report TR2002-833*, October 2002.

[20] S. Sanghavi, “Intermediate Performance of Rateless Codes,” in *Proc. IEEE ITW*, Lake Tahoe, California, vol. 2, no. 6, pp. 478-482, 2007.

[21] S. Kim and S. Lee, “Improved Intermediate Performance of Rateless Codes,” in *Proc. ICACT*, Gangwon-Do, South Korea, pp. 1682–1686, February 2009.

[22] A. Talari and N. Rahnavard, “Rateless Codes with Optimum Intermediate Performance,” in *Proc. Globecom*, Honolulu, Hawaii, pp. 1-6, December 2009.

[23] A. A. El Gamal and T. M. Cover, “Achievable Rates for Multiple Descriptions,” *IEEE Trans. Inform. Theory*, vol. 28, pp. 851–857, November 1982.

[24] L. Ozarow, “On a Source Coding Problem with Two Channels and Three Receivers,” *Bell Syst. Tech. J.*, vol. 59, pp. 1909-1921, 1980.

[25] R. Ahlswede, “The Rate-distortion Region for Multiple Descriptions without Excess Rate,” *IEEE Trans. Inform. Theory*, vol. 31, pp. 721–726, November 1985.

[26] Z. Zhang and T. Berger, “New Results in Binary Multiple Descriptions,” *IEEE Trans. Inform. Theory*, vol. 33, pp. 502–521, July 1987.

[27] R. Venkataramani, G. Kramer, and V. K. Goyal, “Multiple Description Coding with Many Channels,” *IEEE Trans. Inform. Theory*, vol. 49, pp. 2106–2114, September 2003.

[28] S. S. Pradhan, R. Puri, and K. Ramchandran, “n-channel Symmetric Multiple Descriptions - Part I: (n, k) Source-channel Erasure Codes,” *IEEE Trans. Inform. Theory*, vol. 50, no. 1, pp. 47–61, January 2004.

[29] R. Puri, S. S. Pradhan, and K. Ramchandran, “n-channel Symmetric Multiple Descriptions - Part II: An Achievable Rate-distortion Region,” *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1377-1392, April 2005.

[30] C. Tian and J. Chen, “New coding schemes for the symmetric K-description problem,” *IEEE Trans. Inform. Theory*, vol. 56, no. 6, pp. 5344-5365, October 2010.

[31] H. Wang and P. Viswanath, “Vector Gaussian Multiple Description with Individual and Central Receivers,” *IEEE Trans. Inform. Theory*, vol. 53, pp. 2133-2153, June 2007.

[32] J. Chen, “Rate Region of Gaussian Multiple Description Coding with Individual and Central Distortion Constraints,” *IEEE Trans. Inform. Theory*, vol. 55, pp. 3991–4005, September 2009.

[33] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. 2nd ed. Hoboken, NJ: Wiley, 2006.

[34] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.

[35] E. Ahmed and A. B. Wagner, “Binary Erasure Multiple Descriptions: Average-case Distortion,” in *Proc. IEEE ITW*, Volos, Greece, pp. 166–170, June 2009.

[36] E. Ahmed and A. B. Wagner, “Binary Erasure Multiple Descriptions:

Worst-case Distortion," in *Proc. IEEE ISIT*, Seoul, Korea, pp. 55–59, June 2009.

[37] E. Ahmed and A. B. Wagner, "Erasure Multiple Descriptions," *IEEE Trans. Inform Theory*, vol. 58, no. 3, pp. 1328–1344, March 2012.

[38] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, "Priority Encoding Transmission," *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 1737-1744, November 1996.

[39] J. R. Roche, R. W. Yeung, and K. P. Hau, "Symmetrical Multilevel Diversity Coding", *IEEE Trans. Inform. Theory*, vol. 43, no. 5, pp. 1059-1064, May 1997.

[40] J. M. Walsh, S. Weber, and C. wa Maina, "Optimal Rate Delay Tradeoffs and Delay Mitigating Codes for Multipath Routed and Network Coded Networks," *IEEE Trans. Inform. Theory*, vol. 55, no. 12, pp. 5491–5510, December 2009.

[41] W. J. McGill, "Multivariate Information Transmission," *IEEE Trans. Inform. Theory*, vol. 4, no. 4, pp. 93–111, 1954.

[42] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*. New Jersey: Prentice Hall, 1995.

[43] H. Wang and P. Viswanath, "Vector Gaussian Multiple Description with Two Levels of Receivers," *IEEE Trans. Inform. Theory*, vol. 55, pp. 401–410, January 2009.

[44] A. B. Wagner and V. Anantharam, "An Improved Outer Bound for Multiterminal Source Coding," *IEEE Trans. Inform. Theory*, vol. 54, no. 5, pp. 1919–1937, May 2008.

[45] N. Thomos and P. Frossard, "Raptor Network Video Coding," in *Proc. ACM International Workshop on Mobile Video*, Augsburg, Germany, September 2007.

[46] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inform. Theory*, vol. 42, no. 3, pp. 887-902, May 1996.

[47] J. Chen and T. Berger, "Robust Distributed Source Coding," *IEEE Trans. Inform. Theory*, vol. 54, pp. 3385–3398, August 2008.

[48] P. Ishwar, R. Puri, K. Ramchandran, and S. S. Pradhan, “On rate-constrained distributed estimation in unreliable sensor networks,” *IEEE JSAC*, vol. 23, no. 4, pp. 765–775, April 2005.

[49] V. Prabhakaran, “Two Multi-Terminal Communication Problems: Distributed Estimation and Source-Channel Broadcast,” Ph.D. thesis, UC Berkeley, 2007.

[50] Y. Oohama, “The Rate-Distortion Function for the Quadratic Gaussian CEO Problem,” *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.

[51] A. D. Wyner, The Rate-distortion Function for Source Coding with Side Information at the Decoder II: General sources, *Inform. Contr.*, vol. 38, pp. 6080, July 1978.

[52] Y. Oohama, “Gaussian Multi-terminal Source Coding,” *IEEE Trans. Inform. Theory*, vol. 43, no. 6, pp. 1912–1923, November 1997.

[53] R. W. Yeung and N. Cai, “Network error correction, part I: Basic concepts and upper bounds, *Comm. in Inf. and Syst.*, vol. 6, no. 1, pp. 19–36, 2006.

[54] N. Cai and R. W. Yeung, “Secure network coding,” in *Proc. IEEE ISIT*, 2002.

[55] T. Cui, T. Ho, and J. Kliewer, “Achievable strategies for general secure network coding,” in *Information Theory and Applications Workshop*, pp. 1–6, San Diego, CA, 2010.

[56] T. Ho, B. Leong, R. Koetter, M. Medard, M. Effros, and D. R. Karger, “Byzantine modification detection in multicast networks using randomized network coding,” in *Proc. IEEE ISIT*, 2004.

[57] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, and M. Medard, “Resilient network coding in the presence of Byzantine adversaries,” in *Proc. INFOCOM*, pp. 616–624, 2007.

[58] R. Koetter and F. R. Kschischang, “Coding for errors and erasures in random network coding,” *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3579–3591, 2008.

[59] M. Kim, M. Medard, J. Barros, and R. Koetter, “An algebraic watchdog for

wireless network coding," in *Proc. IEEE ISIT*, Seoul, Korea, pp. 1159–1163, July 2009.

[60] O. Kosut and L. Tong, "Distributed source coding in the presence of Byzantine sensors," *IEEE Trans. Inform. Theory*, vol. 54, no. 6, pp. 2550–2565, June 2008.

[61] O. Kosut and L. Tong, "A Characterization of the Error Exponent for the Byzantine CEO Problem", in *Proc. 46th Allerton Conference on Communication, Control and Computing*, pp. 1207–1214, September 2008.

[62] O. Kosut and L. Tong, "The quadratic Gaussian CEO problem with Byzantine agents," in *Proc. IEEE ISIT*, Seoul, Korea, pp. 1145–1149, 2009.

[63] T. M. Cover, A. El Gamal and M. Salehi, "Multiple Access Channels with Arbitrarily Correlated Sources," *IEEE Trans. Inform. Theory*, vol. 26, no. 6, pp. 648-657, Nov. 1980.

[64] M. Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian 'sensor' network," *IEEE Trans. Inform. Theory*, vol. 54, no. 11, pp. 5247–5251, Nov. 2008.

[65] T. J. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. Inform. Theory*, vol. 11, no. 4, pp. 558–567, Oct. 1965.

[66] D. J. Kleitman, "On a combinatorial conjecture of Erdős," *J. Combinatorial Theory*, vol. 1, no. 2, pp. 209–214, September 1966.

[67] O. Kosut, "Adversaries in networks," *Ph.D. thesis*, Cornell University, August 2010.

[68] R. W. Yeung and Z. Zhang, "On symmetrical multilevel diversity coding, *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 609–621, March 1999.