KTH Electrical Engineering

Communication With Reconstruction and Privacy Constraints

KITTIPONG KITTICHOKECHAI

Doctoral Thesis in Telecommunications
Stockholm, Sweden 2014

KTH, School of Electrical Engineering
Communication Theory Laboratory
SE-100 44 Stockholm
SWEDEN

TRITA-EE 2014:026
ISSN 1653-5146
ISBN 978-91-7595-162-1

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen i telekommunikation tisdagen den 3 june 2014 klockan 14.15 i hörsal F3, Lindstedtsvägen 26, Stockholm.

© 2014 Kittipong Kittichokechai, unless otherwise noted.

Tryck: Universitetsservice US AB

Abstract

Communication networks are an integral part of the Internet of Things (IoT) era. They enable endless opportunities for connectivity in a wide range of applications, leading to advances in efficiency of day-to-day life. While creating opportunities, they also incur several new challenges. In general, we wish to design a system that performs optimally well in all aspects. However, there usually exist competing objectives which lead to tradeoffs. In this thesis, driven by several applications, new features and objectives are included into the system model, making it closer to reality and needs. The results presented in this thesis aim at providing insight into the fundamental tradeoff of the system performance which can serve as a guideline for the optimal design of real-world communication systems.

The thesis is divided into two parts. The first part considers the aspect of signal reconstruction requirement as a new objective in the source and channel coding problems. In this part, we consider the framework where the quality and/or availability of the side information can be influenced by a cost-constrained action sequence. In the source coding problem, we impose a constraint on the reconstruction sequence at the receiver that it should be reproduced at the sender, and characterize the fundamental tradeoff in the form of the rate-distortion-cost region, revealing the optimal relation between compression rate, distortion, and action cost. The channel coding counterpart is then studied where a reconstruction constraint is imposed on the channel input sequence such that it should be reconstructed at the receiver. An extension to the multi-stage channel coding problem is also considered where inner and outer bounds to the capacity region are given. The result on the channel capacity reveals interesting consequence of imposing an additional reconstruction requirement on the system model which has a causal processing structure.

In the second part, we consider the aspect of information security and privacy in lossy source coding problems. The sender wishes to compress the source sequence in order to satisfy a distortion criterion at the receiver, while revealing only limited knowledge about the source to an unintended user. We consider three different aspects of information privacy. First, we consider privacy of the source sequence against the eavesdropper in the problem of source coding with action-dependent side information. Next, we study privacy of the source sequence due to the presence of a public helper in distributed lossy source coding problems. The public helper is assumed to be either a user who provides side information over a public link which can be eavesdropped, or a legitimate user in the network who helps to relay information to the receiver, but may not ignore the information that is not intended for it. Lastly, we take on a new perspective of information privacy in the source coding problem. That is, instead of protecting the source sequence, we are interested in the privacy of the reconstruction sequence with respect to a user in the system. For above settings, we provide the complete characterization of the rate-distortion(-cost)-leakage/equivocation region or corresponding inner and outer bounds for discrete memoryless systems.

Acknowledgments

I would like to take this opportunity to acknowledge all those who have supported me in the development of this thesis.

First and foremost, I would like to express my deepest gratitude to my advisor and co-advisor, Prof. Mikael Skoglund and Prof. Tobias Oechtering for their attentive supervision and careful guidance. Mikael gave me the opportunity to join the communication theory lab and the freedom to pursue my research interests. I am grateful for his openness to ideas, insightful suggestions and feedbacks, and supports through my years of study. Tobias has always been an encouraging and motivating mentor. I am very thankful to him for invaluable discussions and lessons both on research and life. Also, I want to thank Prof. Ragnar Thobaben for kindly introducing me to the lab and for his encouragement and important guidance in the early stage of my PhD. In addition, I am grateful to Ananda Mahidol Foundation and Ericsson Thailand for providing me the opportunity to come to study in Sweden.

I would like to thank Prof. Tsachy Weissman at Stanford University for allowing me to visit his group. This opportunity has broadened my knowledge and view on research tremendously. I wish to thank to Dr. Yeow-Khiang Chia for sharing his knowledge and experience with me when we work together. I am also grateful for all friends in the lab who made my time enjoyable and rewarding. Special thanks to Bobbie Chern for taking good care of me during the visit. I would also like to acknowledge the John och Karin Engbloms Stipendiefond for supporting the visit.

I would like to take the opportunity to thank Prof. Sennur Ulukus from University of Maryland for acting as the opponent for this thesis. I also thank Prof. Johan Håstad from KTH, Prof. Abdellatif Zaidi from Université Paris-Est Marne la Vallée, and Prof. Giacomo Como from Lund University for acting on the grading committee, and Prof. Ming Xiao for the thesis review.

I thank my colleagues on floors 4 and 3 who make the working environment enjoyable. Especially, I am thankful to my friend, Dr. Hieu Do. Our lunch and dinner time have stimulated a lot of interesting discussions on life, family, tech., and politics and are most enjoyable time. Thanks also to Prof. Lars Rasmussen, Prof. Ming Xiao, and Prof. James Gross for giving me suggestions and sharing some experiences with me on various occasions. Special thanks to Nan Li, Tai Do, Dr. Ali Zaidi, and many other friends in CT for many wonderful moments and events we have had together. It has been a pleasure for me to share an office with Leefke Grosjean and Dr. Zhongwei Si. I enjoyed your laugh and often got inspired by your hardworking attitude. I am indebted to Mikael, Tobias, Hieu, Tai, Zhao Wang, Farshad Naghibi, Sheng Huang, and Yuwa Chompoobutrgool for their time in proofreading part of the thesis and for giving me helpful comments and feedbacks. I also wish to thank Zuxing Li and Derek Xu with whom I have enjoyable experience working. Also, I would like to thank Raine Tiivel, Irene Kindblom, and Annika Augustsson for their kindness in helping out with administrative matters.

Being far away from home is sometimes tough. I would like to thank all my friends in Stockholm who make my life enjoyable and memorable. Special thanks to P'Niya, P'Jaae, P'Jo, P'Bee, Big, P'Mu, P'Lek, P'Dome, P'Donut, and P'Kla for all their helps and supports. I feel fortunate to take the Swedish course, elementary level at KTH which allowed me to learn some basic Swedish and meet my friend, Seitaro. I am also grateful to all of my teachers and friends in Thailand for their continuous support and encouragement. I thank P'Nan for always being with me, and making me feel happy. Most importantly, I thank my parents and sisters for their love, care, and encouragement.

Kittipong Kittichokechai
Stockholm, June 2014

Contents

Chapter 1

Introduction

Communication systems have extensive impact on our society. At present, they are not only designed for human-to-human communication, but also evolve to serve more complex, machine-to-machine type communication. It is evident that communications play essential roles in various applications, e.g., automation systems, health care, electric power grid, and emergency monitoring. With the emergence of the Internet, they enable endless possibilities to connect *things* together, pushing towards the era of the Internet of Things (IoT) [MF10], [MSDPC12], and are expected to play an indispensable role in transforming and further improving our society and individuals (see, for example, the vision of a future ecosystem driven by high connectivity of individuals and communities, called Networked Society [Eri13]). While creating new opportunities, they also incur several challenges to the traditional system design.

If the demand for communication keeps increasing, we can imagine that one day almost everything around us, ranging from mobile phones, wearable devices, household appliances to all kinds of monitoring sensors, will be connected and form highly interconnected networks. To understand such networks, we need reasonable models which can capture important characteristics and features, yet are tractable to study. One possible approach is to study them from a *local view*, i.e., to consider only a part of the system with a few terminals and abstractly treat other external interactions in the form of *side information*. This local model can then serve as a building block and provide a basis for better understanding of a more general network. Moreover, with the increasing number of devices, it is inevitable that there will be significant amount of data exchanged over the network. Technique such as *data compression* will therefore become an integral part of the commination. Efficient data compression can reduce redundancy of the data to be communicated and thus the amount of physical resources such as bandwidth and energy consumption. If the goal of the system is to compute some function of the data over the network, it is, for example, more efficient to communicate the value of the function of interest rather than the data itself.

Traditionally, the main focus of communication system is on reliability. How-

ever, with a wide range of applications, there exist several other *objectives* that need to be addressed. For example, in some applications, the message of the sender is not the only information that needs to be communicated reliably to the receiver. Other signals in the system may be required simultaneously to improve the possibility of cooperation. In communication scenarios which involve sensitive information such as e-health and online banking, the information exchanged over the network need to have some guaranteed level of privacy and security, i.e., not leaking information to the unintended party. Although high levels of reliability, efficiency, and information security are all desirable objectives in the communication systems, we usually cannot achieve all of them at once as there exist some tradeoffs in the system design. For example, compressing the data to a small amount can save some communication resources, making the system more efficient. However, if the data is overly compressed, the original content may not be reconstructed reliably. On the other hand, sending more information over the network tends to improve quality of the received message, but at the same time increases the chance of leaking information to the unintended users. So the natural questions arise: Is there a fundamental tradeoff of the system's performance metrics, and what do we know about it?

In this thesis, we take some of the new objectives and features driven by several applications into account to model the system closer to reality and needs. We then aim to understand these system models by characterizing their fundamental limits or tradeoffs which shed light on how good the real-world system can be under the optimal design. Below we discuss more specifically how the new features and objectives are considered in the thesis.

Feature to capture interaction among nodes: To support a massive number of devices, the system should have features to adapt to several types of interactions among them. For example, in sensor networks where several sensors are deployed for different tasks, it is desirable that the sensors can operate for a reasonably long time. To achieve that, one can design the system such that each task may require only a small number of sensors to provide measurements and the set of active sensors can be different from task to task. In other words, measurement from each sensor may be acquired differently according to the task. In this thesis, we touch upon this aspect by considering the system models which capture the flexible information acquisition. It is flexible in the sense that the acquisition of information can be controlled by a node in the network. In particular, we consider source networks where a node can take a cost-constrained *action* to influence the quality and/or availability of the side information at certain nodes, termed as *action-dependent side information*. Although not considered directly in the thesis, this cost-constrained information acquisition model can be related to an energy-efficient approach when the cost is defined in terms of energy consumption.

Objectives on additional reconstruction and information privacy[1]: In

[1]We sometimes use the term *privacy, secrecy*, and *security* interchangeably. However, in our context, secrecy and security are intended for scenarios where we wish to protect our information against an external eavesdropper, while the term privacy is used in scenarios against an unintended, legitimate user who is also part of the system.

the world full of connectivity, huge amount of information will be shared among several parties. Exchanging information is of benefit when parties can cooperate with one another to improve the system performance. Imposing some additional reconstruction requirements in the system will enhance the possibility for cooperation as it essentially allows parties to know more about the other. On the other hand, when sharing information, it is of significant concern that some sensitive information should be kept private or secret from the unintended party. This leads to a natural tradeoff between degree of cooperation and level of information privacy when designing the system. It is important in modern communication and compression systems that the system design offers both reliability and security. In this thesis, we include some additional reconstruction requirements into the system model and try to understand their effects by characterizing the corresponding fundamental limits or tradeoffs. In addition, we consider information privacy in the compression system where a signal of interest (e.g., source sequence) should be communicated reliably (satisfying a distortion criterion) and also securely with respect to a certain measure of privacy. We often assume that in the network there is an eavesdropper who observes the transmission over the public channel and therefore can learn about the signal of interest. In this thesis, we consider only a passive type of eavesdropper who does not tamper with the transmission. In some cases, we are also interested in information privacy against an unintended, legitimate user who is friendly in the sense that it processes information as the protocol requests, but may be curious and not ignore the information that is not intended for it.

1.1 Application Scenario

To visualize the networking system with aforementioned objectives and features, we consider a potential application scenario in the area of e-health, depicted in Fig. 1.1. Although the term e-health has a broad definition, in the following, we focus on its aspect in *enabling information exchange and communication in a standardized way between health care establishments, leading to improved possibilities for institution-to-institution transmissions of data* [Eys01].

In the e-health scenario, some medical data of patients such as scanned images and medical care records can be shared among several institutions or doctors to facilitate the medical treatment. For example, senders S in Fig. 1.1 represent doctors who have access to different databases[2] of medical images of the patient. They wish to send relevant information over the network to the one who requests it, depicted as receiver R in Fig. 1.1. It is common to assume that communication links between senders and receivers are of limited capacity. Therefore, the transmission has to be done in an efficient way, i.e., using (distributed) data compression. In some cases, the sender or receiver might take some action at a certain cost to request some extra information from another doctor/institution, depicted as helper H in Fig. 1.1. This extra acquisition is important especially when the quality of the received

[2]For privacy reason, institutions might not be allowed to share the whole databases to others.

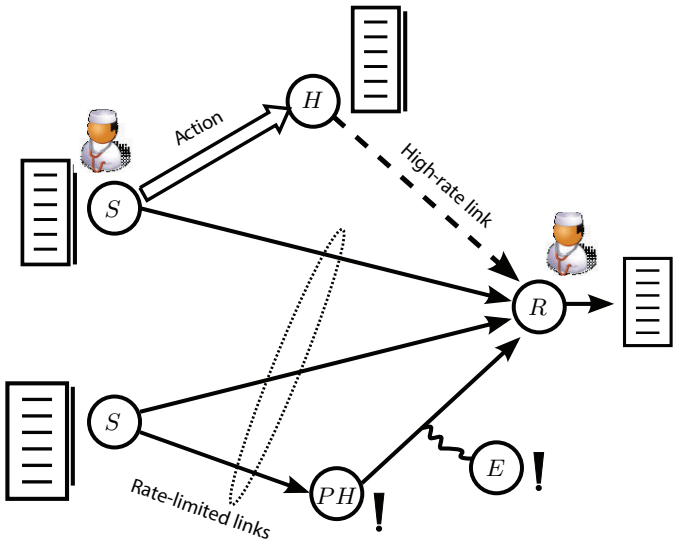

Figure 1.1: A node with symbol S represents a sender who has access to some medical database, while a node with symbol R represents a receiver. A node with symbol H represents a helper who can facilitate the transmission by providing some side information to the receiver. Symbol PH denotes a public helper which could be a networking device such as a router. It obeys the communication protocol by relaying information to the receiver, but may not ignore the information that is not intended for it. Lastly, a node with symbol E represents an external eavesdropper who can observe the transmission sent over the public communication link.

image is crucial for the treatment. Another important functionality in e-health is online medical consultation. Considering for example an important surgery which requires a consultation among several doctors. The consultation is done over the network based on some references such as the medical image sent by the consulting doctors. In this scenario, the senders may want to know exactly what the received image looks like since it may get distorted and thus affect the outcome of the consultation and surgical operation significantly. When the medical data is sent over the network, e.g., the Internet, it is possible that it passes through several intermediate terminals before reaching the intended receiver. As they are often sensitive data, it is important that they are protected against disclosure at any unintended terminal (the public helper PH or eavesdropper E) who may receive or intercept some transmission.

With a reasonable model for the networking system in Fig. 1.1, we can study some relevant problems and are able to understand the fundamental tradeoffs of the system. These tradeoffs can serve as a guideline for the optimal design of a

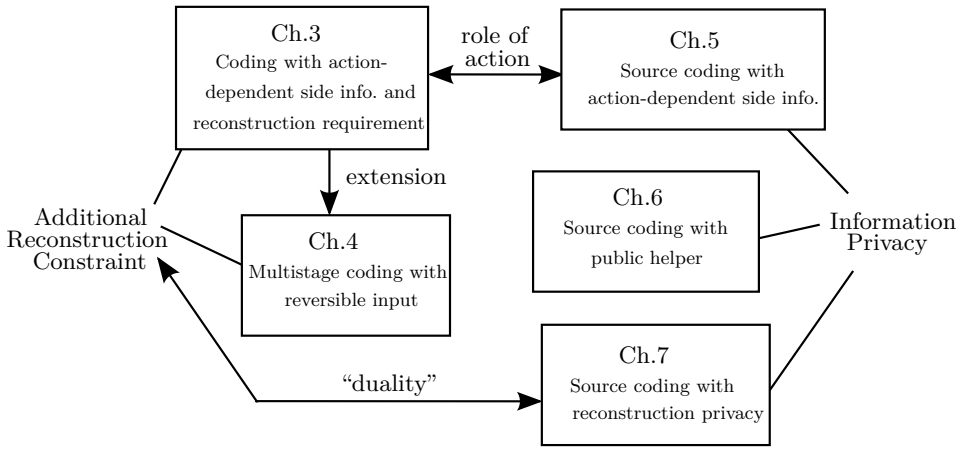

Figure 1.2: Dependency graph.

practical networking system for e-health.

1.2 Organization and Contributions of the Thesis

The remaining of the thesis is divided into seven chapters. We summarize the contents in each of them along with the contributions below. Relations among the chapters are shown by a dependency graph in Fig. 1.2.

Chapter 2

In this chapter, we provide some fundamental concepts in information theory which serve as a background for studying the problems in the subsequent chapters. For more detailed treatment of the subject, readers are referred to standard textbooks such as [CT06] or [EK11]. The chapter also introduces some aspects of the problem formulation which include features and objectives that will be considered later in the thesis.

Chapters 3 and 4 consider the problems of source and channel coding with additional reconstruction constraints where the side/state information can be controlled by a node in the network. In the source coding problem, the additional reconstruction requirement is on the receiver's reconstruction sequence, i.e., it should also be locally estimated at the transmitter. In the channel coding problem, we consider both basic and multi-stage settings, where the additional reconstruction requirement is on the channel input sequence, i.e., it should also be decoded at the receiver.

Chapter 3

In Chapter 3, we consider two classes of source and channel coding problems involving action-dependent side information/states and an additional reconstruction requirement, namely, a problem of source coding with two-sided action dependent side information under common reconstruction constraint, and a problem of channel coding with two-sided action-dependent state under reversible input constraint. We characterize the respective rate-distortion-cost region, and the channel capacity. One interesting observation lies in the fact that the capacity is expressed with an additional condition which restricts the set of feasible input distributions. We term such condition as *two-stage coding condition* as it arises essentially from the two-stage structure of the coding which requires an additional reconstruction of a signal generated in the second stage.

The contribution of this chapter is based on

- [KOS12a] K. Kittichokechai, T. J. Oechtering, and M. Skoglund, "Coding with action-dependent side information and additional reconstruction requirements," submitted to *IEEE Trans. Inf. Theory,* Feb. 2012 (revised July 2013).

Shorter versions also appeared in

- [KOST10] K. Kittichokechai, T. J. Oechtering, M. Skoglund, and R.Thobaben, "Source and channel coding with action-dependent partially known two-sided state information," in *Proc. of IEEE International Symposium on Information Theory (ISIT),* Austin, TX, USA, June 2010.

- [KOS10] K. Kittichokechai, T. J. Oechtering, and M. Skoglund, "Source coding with common reconstruction and action-dependent side information," in *Proc. of IEEE Information Theory Workshop (ITW)* Dublin, Ireland, Sep. 2010.

- [KOS11b] K. Kittichokechai, T. J. Oechtering, and M. Skoglund, "On the capacity of a channel with action-dependent state and reversible input," in *Proc. of IEEE International Symposium on Information Theory (ISIT),* St. Petersburg, Russia, July 2011.

- [KOS11a] K. Kittichokechai, T. J. Oechtering, and M. Skoglund, "Capacity of the channel with action-dependent state and reversible input," in *Proc. of IEEE Swedish Communication Technologies Workshop (Swe-CTW),* Stockholm, Sweden, Oct. 2011.

Chapter 4

In Chapter 4, we extend the channel coding model in Chapter 3 to the multi-stage case. Motivation of the study is to explore further the role of the additional reconstruction requirement in the multi-stage setting. We characterize the capacity

region for the two-stage case and draw a connection to the *two-stage coding condition* observed in Chapter 3. For the case of three or more stages, we provide inner and outer bounds to the capacity region, and show that they match under certain special channel assumptions.

The contribution of this chapter is based on

- [KOS12b] K. Kittichokechai, T. J. Oechtering, and M. Skoglund, "Multi-stage coding for channels with a rewrite option and reversible input," in *Proc. of IEEE International Symposium on Information Theory (ISIT),* Boston, MA, USA, July 2012.

- [KOS12a] K. Kittichokechai, T. J. Oechtering, and M. Skoglund, "Coding with action-dependent side information and additional reconstruction requirements," submitted to *IEEE Trans. Inf. Theory,* Feb. 2012 (revised July 2013).

Chapters 5 to 7 consider the problem of source coding under information security and privacy constraints. Three different aspects of privacy are considered, namely, 1) source privacy with respect to an eavesdropper when the side information can be controlled by a node in the network, 2) source privacy in the presence of a legitimate helper in the network, and 3) privacy of the reconstruction signal with respect to a user in the network. We use normalized information leakage or equivocation as a measure of information privacy and are interested in characterizing the optimal tradeoff among the system parameters such as compression rate, incurred distortion at the intended receiver, and privacy level at the unintended user.

Chapter 5

In this chapter, we study a lossy source coding problem with action-dependent side information under source secrecy constraint. Several aspects of action-dependent side information are considered, i.e., the cases where the action is taken at either the encoder or decoder, and the case where the action is taken to generate a common two-sided side information. We characterize the rate-distortion-cost-leakage regions or the corresponding inner and outer bounds to the rate-distortion-cost-leakage region for the mentioned problems.

The contribution of this chapter is based on

- [KOS$^+$14b] K. Kittichokechai, T. J. Oechtering, M. Skoglund, Y.-K. Chia, and T. Weissman, "Secure source coding with action-dependent side information," to be submitted to *IEEE Trans. Inf. Theory,* 2014.

A shorter version appeared in

- [KOS11c] K. Kittichokechai, T. J. Oechtering, and M. Skoglund, "Secure source coding with action-dependent side information," in *Proc. of IEEE International Symposium on Information Theory (ISIT),* St. Petersburg, Russia, July 2011.

Chapter 6

In Chapter 6, we consider source coding problem in the presence of a public helper. We are interested in how the system can utilize an extra node, called helper, to support the transmission, given that the helper is a terminal which can leak information through its public link or that the helper itself is a public terminal which we cannot trust and therefore reveal too much information. We study several classes of the problems and characterize the complete rate-distortion-leakage regions both for the general case and under special assumptions on the Markov structure of the side information or distortion measure.

The contribution of this chapter is based on

- [KCO+13a] K. Kittichokechai, Y.-K. Chia, T. J. Oechtering, M. Skoglund, and T. Weissman, "Secure source coding with a public helper," submitted to *IEEE Trans. Inf. Theory,* July 2013.

A shorter version appeared in

- [KCO+13b] K. Kittichokechai, Y.-K. Chia, T. J. Oechtering, M. Skoglund, and T. Weissman, "Secure source coding with a public helper," in *Proc. of IEEE International Symposium on Information Theory (ISIT),* Istanbul, Turkey, July 2013.

Chapter 7

In this chapter, we study the problem of lossy source coding with a privacy constraint on the reconstruction sequence at the receiver, which we term as *end-user privacy*. This situation is an extension of the Wyner-Ziv problem to include a privacy constraint on the receiver's reconstruction sequence such that it should be protected against any inference from the other node. From a problem formulation point of view, end-user privacy can be viewed as a dual constraint to the reconstruction constraint considered in Chapter 3. We characterize inner and outer bounds to the rate-distortion-equivocation regions for different cases, and show that under the memoryless reconstruction assumption, the complete rate-distortion-equivocation region can be obtained for the case of end-user privacy at the eavesdropper.

The contribution of this chapter is based on

- [KOS14a] K. Kittichokechai, T. J. Oechtering, and M. Skoglund, "Lossy source coding with reconstruction privacy," to appear in *Proc. of IEEE International Symposium on Information Theory (ISIT),* Honolulu, HI, USA, June 2014.

- [KOS14c] K. Kittichokechai, T. J. Oechtering, and M. Skoglund, "Lossy source coding with reconstruction privacy," in preparation, 2014.

Chapter 8

In the last chapter, we summarize our contributions in the thesis, and discuss potential directions for future research.

Contributions Outside the Scope of the Thesis

Besides the contributions listed above, the author of this thesis has also contributed to some other related works which are published in the papers listed below. For consistency of the thesis structure, these are not included in the thesis.

- [CK13a] Y.-K. Chia and K. Kittichokechai, "On secure source coding with side information at the encoder," in *Proc. of IEEE International Symposium on Information Theory (ISIT),* Istanbul, Turkey, July 2013.
- [LOK14] Z. Li, T. J. Oechtering, and K. Kittichokechai, "Parallel distributed bayesian detection with privacy constraints," to appear in *Proc. of IEEE International Conference on Communications (ICC),* Sydney, Australia, June 2014.
- [XKOS14] D. Xu, K. Kittichokechai, T. J. Oechtering, and M. Skoglund, "Secure successive refinement with degraded side information," to appear in *Proc. of IEEE International Symposium on Information Theory (ISIT),* Honolulu, HI, USA, June 2014.

1.3 Copyright Notice

As specified in Section 1.2, parts of the material presented in this thesis are partly verbatim based on the thesis author's joint works which are previously published or submitted to conferences and journals held by or sponsored by the Institute of Electrical and Electronics Engineer (IEEE). IEEE holds the copyright of the published papers and will hold the copyright of the submitted papers if they are accepted. Materials (e.g., figure, graph, table, or textual material) are reused in this thesis with permission.

1.4 Notations and Abbreviations

We denote the real-valued random variables, their corresponding realizations or deterministic values, and their alphabets by the upper case, lower case, and calligraphic letters, respectively. The term X_m^n denotes the sequence $\{X_m, \ldots, X_n\}$ when $m \leq n$, and the empty set otherwise. Also, we use the shorthand notation X^n for X_1^n. The term $X^{n\setminus i}$ denotes the set $\{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$. We denote that random variable X is a constant by $X = \emptyset$. Cardinality of the set $\mathcal{X}$ is denoted by $|\mathcal{X}|$. All logarithms in the thesis are to the base 2, unless stated otherwise. The following two tables summarize notations and abbreviations used in this thesis.

Notations

$P_X(x), P_X, p(x)$	Probability distribution of discrete random variable X
$X \sim P_X$	Random variable X is distributed according to P_X
$X - Y - Z$	Markov chain (Definition 2.2)
$E[X]$	Expectation of random variable X
$H(X)$	Entropy of random variable X
$H(X,Y)$	Joint entropy of random variables X and Y
$H(X\|Y)$	Conditional entropy of random variable X given Y
$I(X;Y)$	Mutual information between random variables X and Y
$I(X;Y\|Z)$	Conditional mutual information between X and Y given Z
$\Pr(\mathcal{E})$	Probability of the event $\mathcal{E}$
$\mathcal{T}_\epsilon^{(n)}(X)$	Typical set with respect to P_X (Definition 2.7)
$\mathcal{C}_n$	Random codebook containing codewords of length-n
δ_ϵ	Small value that depends on $\epsilon > 0$ such that $\delta_\epsilon \to 0$ as $\epsilon \to 0$
δ_n	Small value that depends on n such that $\delta_n \to 0$ as $n \to \infty$
$\mathrm{BSC}(a)$	Binary symmetric channel with transition probability a
$\mathrm{Bernoulli}(a)$	Bernoulli distribution with success probability a
$\mathcal{N}(\mu, \sigma^2)$	Gaussian distribution with mean μ and variance σ^2
$\mathbb{R}_+^n$	The set of n-dimensional non-negative real vectors
$1_{\{X=a\}}(x)$	Indicator function, i.e., $1_{\{X=a\}}(x) = 1$ if $x = a$, and zero else.
$\sum_x$	Summation over all $x \in \mathcal{X}$
$[1 : 2^r]$	The set $\{1, 2, \ldots, 2^r\}$
$[a]^+$	$\max(0, a)$
$h(\cdot)$	Binary entropy function (see also Lemma 2.4)

Abbreviations

CR constraint	Common reconstruction constraint
DMC	Discrete memoryless channel
DMS	Discrete memoryless source
EPI	Entropy power inequality
i.i.d.	independent and identically distributed
LLN	Law of large number
log-loss	Logarithmic loss
MMSE	Minimum mean square error
PMF	Probability mass function
RI constraint	Reversible input constraint

Chapter 2

Background

This chapter provides some fundamental concepts from probability and information theory which serve as a background for studying the problems in the subsequent chapters. It also introduces some basic problem settings and interesting aspects of the problem formulation that will be considered and discussed in more detail later in the thesis.

2.1 Preliminaries

We start with a review of some basic notations and properties of a discrete random variable and its probability distribution. Then we introduce some important information measures and discuss their properties. Finally, we present definitions and important properties of jointly typical set and sequences.

2.1.1 Discrete Random Variable

A random variable is a basic element in probability theory as it is a real-valued function of an outcome of a random experiment which represents some numerical quantity associated with the outcome [MW12, Chapter 7]. In the study of communication systems, probabilistic models are commonly used to describe the system in which the signals are represented in the form of random variables. In the following, some basic notations and properties of discrete random variable and its probability distribution are given.

Let X be a discrete random variable with alphabet $\mathcal{X}$. We denote the probability mass function (PMF) of X by $P_X(x)$ or in short by P_X or $p(x)$. As for multiple random variables, similar notations are used, for example, $P_{X,Y}$ for a pair of random variables (X, Y), and P_{X^n} for a vector X^n.

Definition 2.1 (Independence)**.** Two random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ are called *independent* if

$$P_{X,Y}(x, y) = P_X(x)P_Y(y), \tag{2.1}$$

for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$, where $P_X(x)$ and $P_Y(y)$ are the marginal PMFs of X and Y, respectively.

Definition 2.2 (Markov chain)**.** We use $X - Y - Z$ to denote that (X, Y, Z) forms a Markov chain, that is, their joint PMF factorizes as $P_{X,Y,Z}(x, y, z) = P_{X,Y}(x, y)P_{Z|Y}(z|y)$ or $P_{X,Y,Z}(x, y, z) = P_{X|Y}(x|y)P_{Y,Z}(y, z)$.

Markov chain is related to the concept of *conditional independence*. For example, $X - Y - Z$ implies that X is conditionally independent of Z given Y. We can see that it has a symmetrical property, i.e., $X - Y - Z$ implies $Z - Y - X$. In the following, we state a few other properties satisfied by a Markov chain or conditional independence.

Lemma 2.1 (Properties satisfied by Markov chain [Pea00])**.** *Let (X, Y, Z, W) be discrete random variables.*

- *Decomposition: $X - Y - (Z, W)$ implies $X - Y - Z$.*
- *Weak union: $X - Y - (Z, W)$ implies $X - (Y, Z) - W$.*
- *Contraction: $X - Y - Z$ and $X - (Y, Z) - W$ imply $X - Y - (Z, W)$.*
- *Intersection: $X - (Y, Z) - W$ and $X - (Y, W) - Z$ imply $X - Y - (Z, W)$, where the intersection is valid only for strictly positive probability distributions.*

2.1.2 Shannon's Information Measures

The solutions to various information theoretic problems are often expressed in terms of certain quantities which are related to information measures. The most common and important ones are entropy and mutual information which are defined based on the probability distributions associated with related random variables. In the following, we define these quantities along with some basic properties. More details can be found, e.g., in [CT06] or [EK11].

Definition 2.3 (Entropy)**.** Let X be a discrete random variable with finite alphabet $\mathcal{X}$ and a PMF $P_X(x)$. The *entropy* of X, denoted by $H(X)$, is defined as

$$H(X) \triangleq -\sum_{x \in \mathcal{X}} P_X(x) \log P_X(x). \tag{2.2}$$

The unit of $H(X)$ depends on the base of the logarithm. $H(X)$ is measured in *bits* if the logarithm is to the base 2, and is measured in *nats* if the logarithm is to the base e. Entropy can be interpreted as a measure of uncertainty, for example, $H(X)$ represents amount of uncertainty in X.

Definition 2.4 (Conditional entropy)**.** Let X and Y be two discrete random variables with finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, and a joint PMF $P_{X,Y}(x,y)$. The conditional entropy of Y given X, denoted by $H(Y|X)$, is defined as

$$H(Y|X) \triangleq -\sum_{x,y} P_{X,Y}(x,y) \log P_{Y|X}(y|x). \tag{2.3}$$

Conditional entropy $H(Y|X)$ can be interpreted as amount of uncertainty in Y remained after observing X.

Definition 2.5 (Mutual information)**.** Let X and Y be two discrete random variables with finite alphabets $\mathcal{X}$, $\mathcal{Y}$ and a joint PMF $P_{X,Y}(x,y)$. The mutual information between X and Y, denoted by $I(X;Y)$, is defined as

$$I(X;Y) \triangleq \sum_{x,y} P_{X,Y}(x,y) \log \frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)}, \tag{2.4}$$

where $P_X(x)$ and $P_Y(y)$ are the marginal PMFs of X and Y, respectively. Mutual information can be interpreted as a measure of information shared between random variables. For example, $I(X;Y)$ represents amount of information about X one can infer from observing Y, or equivalently amount of information about Y one can infer from observing X.

Definition 2.6 (Conditional mutual information)**.** Let X, Y, and Z be discrete random variables with finite alphabets $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ and a joint PMF $P_{X,Y,Z}(x,y,z)$. The conditional mutual information between X and Y conditioned on Z, denoted by $I(X;Y|Z)$, is defined as

$$I(X;Y|Z) \triangleq \sum_{x,y,z} P_{X,Y,Z}(x,y,z) \log \frac{P_{X,Y|Z}(x,y|z)}{P_{X|Z}(x|z)P_{Y|Z}(y|z)}. \tag{2.5}$$

Conditional mutual information $I(X;Y|Z)$ represents amount of information about X one can infer from observing Y given that Z is already known.

Above definitions for entropy and mutual information are given for a pair or triple of random variables. However, similar definitions apply for multiple random variables with corresponding joint distributions.

Lemma 2.2 (Properties of H and I)**.** *The following are some basic properties of entropy and mutual information.*

- *Non-negativity:* $H(X) \geq 0$.
- *Conditioning reduces entropy:* $H(X|Y) \leq H(X)$.
- *Chain rule:* $H(X^n) = \sum_{i=1}^{n} H(X_i|X^{i-1}) = \sum_{i=1}^{n} H(X_i|X_{i+1}^n)$.

- *Upper bound:* $H(X) \leq \log|\mathcal{X}|$*, with equality iff* X *is distributed uniformly over* $\mathcal{X}$*.*
- *Non-negativity:* $I(X;Y) \geq 0$*.*
- *Relation to entropy:* $I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$*.*
- *Self information:* $I(X;X) = H(X)$*.*
- *Chain rule:* $I(X^n;Y^n|Z^n) = \sum_{i=1}^{n} I(X_i;Y^n|Z^n,X^{i-1})$*.*

Lemma 2.3 (Data processing inequality [CT06])**.** *If* $X - Y - Z$*, then* $I(X;Y) \geq I(X;Z)$*. In particular, if* $Z = f(Y)$*, we have* $I(X;Y) \geq I(X;f(Y))$*. This implies that any processing of the data* Y *cannot increase information about* X*.*

Lemma 2.4 (Fano's inequality [CT06], [BB11])**.** *Let* $X \in \mathcal{X}$ *be a discrete random variable and* $Y \in \mathcal{Y}$ *be any estimate of* X *with* $P_e = Pr(X \neq Y)$*. Then we have*

$$H(X|Y) \leq h(P_e) + P_e \log|\mathcal{X}| \leq 1 + P_e \log|\mathcal{X}|, \tag{2.6}$$

where $h(\cdot)$ *is the binary entropy function defined as* $h(p) \triangleq -p\log p - (1-p)\log(1-p)$*,* $p \in [0,1]$ *where* $h(0) = h(1) = 0$*.*

If $\mathcal{X} = \mathcal{Y}$*, we have that*

$$H(X|Y) \leq h(P_e) + P_e \log(|\mathcal{X}| - 1). \tag{2.7}$$

Fano's inequality is the key ingredient of many proofs in information theory as it relates an information theoretic quantity (conditional entropy) to an operational quantity (error probability of estimation). For example, for any estimate Y, we can obtain a lower bound on the error probability of estimating X expressed in the form of the conditional entropy $H(X|Y)$. In the thesis, we often write Fano's inequality in the form $H(X|Y) \leq \delta(P_e)$ to emphasize that $H(X|Y) \to 0$ as $P_e \to 0$, or when $P_e \to 0$ as $n \to \infty$, we write $H(X|Y) \leq \delta_n$, where $\delta_n \to 0$ as $n \to \infty$.

Lemma 2.5 (Csiszár's sum identity [CK11], [EK11])**.** *Let* X^n *and* Y^n *be two random vectors with arbitrary joint probability distribution, then*

$$\sum_{i=1}^{n} I(X_i;Y^{i-1}|X_{i+1}^n) - I(Y_i;X_{i+1}^n|Y^{i-1}) = 0, \tag{2.8}$$

where we assume that $Y_0 = \varnothing$[1]*.*

[1]We use notation $Y = \varnothing$ to denote the event that a variable Y is a constant.

2.1.3 Typical Sequences

In the following, we present definitions and properties of jointly typical set and sequences which play an important role in information theory. In particular, all achievability proofs of the results in the thesis are based on random coding arguments using joint typicality to describe the encoding and decoding. We note that there exist several notions and definitions of typicality. In this thesis, we follow the strong typicality or ϵ-typicality defined in [OR01], [EK11].

Definition 2.7 (Jointly typical set)**.** Let $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ where $\mathcal{X}$ and $\mathcal{Y}$ are finite. The number of occurrences of a pair of symbol $(a, b) \in \mathcal{X} \times \mathcal{Y}$ in the tuple (x^n, y^n) is denoted by $N(a, b; x^n, y^n)$. The set of jointly typical sequences (x^n, y^n) for $\epsilon > 0$ with respect to the joint distribution $P_{X,Y}$ denoted by $\mathcal{T}_\epsilon^{(n)}(X, Y)$ or in short by $\mathcal{T}_\epsilon^{(n)}$ is defined as

$$\mathcal{T}_\epsilon^{(n)}(X, Y) \triangleq \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \Big|\frac{1}{n}N(a, b; x^n, y^n) - P_{X,Y}(a, b)\Big| \leq \epsilon P_{X,Y}(a, b), \text{ for all } (a, b) \in \mathcal{X} \times \mathcal{Y}\}. \tag{2.9}$$

Since the term $\frac{N(a,b;x^n,y^n)}{n}$ is the empirical joint distribution of (x^n, y^n), the jointly typical set contains all pairs of sequences which have empirical distribution *close* to the true distribution $P_{X,Y}$. As a pair of random variables can be defined as a new random variable, similar definitions and corresponding properties of (jointly) typical set and sequences also apply for one or multiple random variables.

In the following, we state some important and useful properties of (joint) typicality known as asymptotic equipartition property (AEP).

Theorem 2.1.1 (Joint AEP)**.** *Let (x^n, y^n) be a pair of sequences of i.i.d. random variables with joint PMF $P_{X,Y}$. Then*

- *For all $(x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y)$, $2^{-n(H(X,Y)+\delta_\epsilon)} \leq p(x^n, y^n) \leq 2^{-n(H(X,Y)-\delta_\epsilon)}$,*
- *$Pr\big((X^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y)\big) \geq 1 - \delta_\epsilon$ for n sufficiently large,*
- *$|\mathcal{T}_\epsilon^{(n)}(X, Y)| \leq 2^{n(H(X,Y)+\delta_\epsilon)}$,*
- *$|\mathcal{T}_\epsilon^{(n)}(X, Y)| \geq (1 - \epsilon)2^{n(H(X,Y)-\delta_\epsilon)}$ for n sufficiently large.*

Definition 2.8 (Conditional typical set)**.** Let $x^n \in \mathcal{T}_\epsilon^{(n)}(X)$ for $\epsilon > 0$ and $Y^n \sim \prod_{i=1}^n P_{Y|X}(y_i|x_i)$. The conditional typical set with respect to x^n is defined as

$$\mathcal{T}_\epsilon^{(n)}(Y|x^n) \triangleq \{y^n \in \mathcal{Y}^n : (x^n, y^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y)\}. \tag{2.10}$$

Theorem 2.1.2 (Conditional AEP)**.** *For $0 < \epsilon' < \epsilon$, let $x^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)$ and $Y^n \sim \prod_{i=1}^n P_{Y|X}(y_i|x_i)$. Then*

- *For all* $y^n \in \mathcal{T}_\epsilon^{(n)}(Y|x^n)$, $2^{-n(H(Y|X)+\delta_\epsilon)} \leq p(y^n|x^n) \leq 2^{-n(H(Y|X)-\delta_\epsilon)}$,
- $Pr\big(Y^n \in \mathcal{T}_\epsilon^{(n)}(Y|x^n)\big) \geq 1-\delta_{\epsilon,\epsilon'}$ *for* n *sufficiently large. This statement is also known as conditional typicality lemma,*
- $|\mathcal{T}_\epsilon^{(n)}(Y|x^n)| \leq 2^{n(H(Y|X)+\delta_\epsilon)}$,
- $|\mathcal{T}_\epsilon^{(n)}(Y|x^n)| \geq (1-\epsilon)2^{n(H(Y|X)-\delta_\epsilon)}$ *for* n *sufficiently large.*

As shown in [EK11, Chapter 2], a stronger statement also holds, namely, for $x^n \in \mathcal{T}_{\epsilon'}^{(n)}(X)$ *and* n *sufficiently large,* $|\mathcal{T}_\epsilon^{(n)}(Y|x^n)| \geq 2^{n(H(Y|X)-\delta_\epsilon)}$ *where* $\delta_\epsilon \to 0$ *as* $\epsilon \to 0$.

Lemma 2.6 (Covering lemma [EK11])**.** *Let* $\epsilon' < \epsilon$. *Let* (U^n, X^n) *be a pair of arbitrarily distributed random sequences (not necessarily according to* $\prod_{i=1}^n P_{U,X}$*) such that* $Pr\big((U^n, X^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U,X)\big) \to 1$ *as* $n \to \infty$ *and let* $\hat{X}^n(w), w \in \mathcal{A}$, *where* $|\mathcal{A}| \geq 2^{nR}$, *be random sequences, conditionally independent of each other and of* X^n, *given* U^n, *where each is distributed according to* $\prod_{i=1}^n P_{\hat{X}|U}(\hat{x}_i|u_i)$. *Then, there exists* $\delta_\epsilon \to 0$ *as* $\epsilon \to 0$ *such that*

$$Pr\big((U^n, X^n, \hat{X}^n(w)) \notin \mathcal{T}_\epsilon^{(n)}(U,X,\hat{X}) \text{ for all } w \in \mathcal{A}\big) \to 0$$

as $n \to \infty$ *if* $R > I(X;\hat{X}|U) + \delta_\epsilon$.

Lemma 2.7 (Packing lemma [EK11])**.** *Let* (U^n, Y^n) *be a pair of arbitrarily distributed random sequences (not necessarily according to* $\prod_{i=1}^n P_{U,Y}$*). Let* $X^n(m)$, $m \in \mathcal{A}$, *where* $|\mathcal{A}| \leq 2^{nR}$, *be random sequences, each distributed according to the distribution* $\prod_{i=1}^n P_{X|U}(x_i|u_i)$. *Assume that* $X^n(m), m \in \mathcal{A}$ *is pairwise conditionally independent of* Y^n, *given* U^n. *Then, there exists* $\delta_\epsilon \to 0$ *as* $\epsilon \to 0$ *such that*

$$Pr\big((U^n, Y^n, X^n(m)) \in \mathcal{T}_\epsilon^{(n)}(U,Y,X) \text{ for some } m \in \mathcal{A}\big) \to 0$$

as $n \to \infty$ *if* $R < I(Y;X|U) - \delta_\epsilon$.

2.2 Basic Problem Settings

Information theory in the classical sense deals with two fundamental problems, namely the problems of efficient compression and reliable communication. Here we review some of the basic problem settings that serve as the basis for our problems in the thesis. We start with the point-to-point compression and communication problems introduced by Shannon [Sha48], [Sha59]. We will introduce the problem formulations, define some important concepts such as achievability, and state the main results. Later on, we review the network models involving multiple terminals, namely the Wyner-Ziv's rate-distortion with side information, and the Gel'fand-Pinsker's channel with random states. Lastly, we discuss extensions of aforementioned problems which include several interesting aspects into the problem formulation.

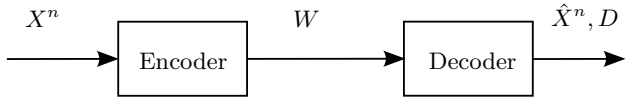

Figure 2.1: Point-to-point lossy source coding.

2.2.1 Point-to-point Lossy Source Coding

In the lossy compression problem, we wish to compress a source sequence into a source description such that a decoder based on the description can reconstruct the source within the prescribed distortion level. For the basic setting, the source is modeled as a discrete memoryless source (DMS) and the goal of the compression is to achieve the lowest possible compression rate for a given distortion. Below we present the problem formulation for a point-to-point model of lossy source coding problem in Fig. 2.1.

Given a length-n source sequence X^n which is distributed according to the distribution $\prod_{i=1}^{n} P_X(x_i)$, an encoder generates a source description $W \in \mathcal{W}^{(n)}$ and sends it to the decoder over a noise-free rate-limited link. The decoder based on the description reconstructs the source as a sequence $\hat{X}^n$ such that the average distortion between the source and the reconstruction is below value D. In order to achieve the objective of the problem, the encoder and decoder need to agree on a *code* which is basically described by the encoding and decoding processes.

Definition 2.9 (Code)**.** A $(|\mathcal{W}^{(n)}|, n)$ code for the point-to-point lossy source coding for a DMS consists of

- an encoding function $f^{(n)} : \mathcal{X}^n \to \mathcal{W}^{(n)}$,
- a decoding function $g^{(n)} : \mathcal{W}^{(n)} \to \hat{\mathcal{X}}^n$,

where $\mathcal{W}^{(n)}$ is a finite set.

Definition 2.10 (Distortion)**.** Let $d : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty)$ be the single-letter distortion measure. The distortion between the source sequence and its reconstruction at the decoder is defined as

$$d^{(n)}\big(X^n, \hat{X}^n\big) \triangleq \frac{1}{n} \sum_{i=1}^{n} d(X_i, \hat{X}_i), \tag{2.11}$$

where $d^{(n)}(\cdot)$ is the distortion function.

Definition 2.11 (Achievability)**.** Definition of the code together with the source distribution induce the joint distribution of the form

$$\prod_{i=1}^{n} P_X(x_i) 1_{\{W = f^{(n)}(x^n)\}}(w) 1_{\{\hat{X}^n = g^{(n)}(w)\}}(\hat{x}^n). \tag{2.12}$$

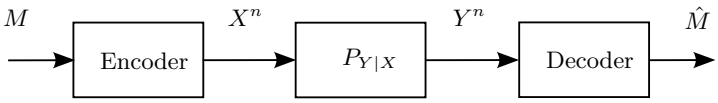

Figure 2.2: Point-to-point discrete memoryless channel.

A rate-distortion pair (R, D) is said to be *achievable* if for any $\delta > 0$ and all sufficiently large n there exists a $(|\mathcal{W}^{(n)}|, n)$ code such that $\frac{1}{n}\log\left|\mathcal{W}^{(n)}\right| \leq R + \delta$ and $E[d^{(n)}(X^n, \hat{X}^n)] \leq D + \delta$.

Definition 2.12 (Rate-distortion region)**.** The *rate-distortion region* $\mathcal{R}$ is the set of all achievable rate-distortion pairs. Due to the time-sharing argument [CT06, Chapter 15], the rate-distortion region of any DMS is a convex set.

Theorem 2.2.1 (Rate-distortion theorem)**.** *The rate-distortion region of the point-to-point lossy source coding problem for a given DMS is described by the set of all* (R, D) *that satisfy*

$$R \geq I(X; \hat{X}), \tag{2.13}$$

$$D \geq E[d(X, \hat{X})], \tag{2.14}$$

for some joint distributions of the form $P_X(x)P_{\hat{X}|X}(\hat{x}|x)$*.*

2.2.2 Point-to-point Discrete Memoryless Channel

The main goal of communication is to transmit a message reliably from one point to another (or more) over a communication channel. In practice, communication channels can be corrupted by noise or other disturbances. To capture such characteristics, we commonly use probabilistic model to describe the communication channel. One of the simplest models is the point-to-point discrete memoryless channel (DMC) shown in Fig. 2.2.

In the point-to-point system, a transmitter (encoder) wants to send a *message* to a receiver (decoder) over a DMC governed by the conditional probability distribution $P_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^{n} P_{Y|X}(y_i|x_i)$, where X^n is the channel input sequence, Y^n is the channel output sequence, and n is the sequence length. Assume that the transmitter chooses the message m uniformly from the set $\{1, \ldots, |\mathcal{M}^{(n)}|\}$. In order to achieve reliable communication, the transmitter and receiver need to agree on a *code* which is basically described by the encoding and decoding processes.

Definition 2.13 (Code)**.** An $(|\mathcal{M}^{(n)}|, n)$ code for the point-to-point DMC consists of

- a message set $\mathcal{M}^{(n)} = \{1, \ldots, |\mathcal{M}^{(n)}|\}$,
- an encoding function $f^{(n)} : \mathcal{M}^{(n)} \to \mathcal{X}^n$,

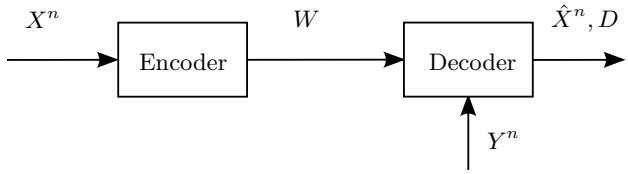

Figure 2.3: Rate-distortion with side information at the decoder.

- a decoding function $g^{(n)} : \mathcal{Y}^n \to \hat{\mathcal{M}}^{(n)}$.

Definition 2.14 (Achievability)**.** Assuming that M is distributed uniformly over the set $\mathcal{M}^{(n)}$. Together with the code, this induces the joint distribution of the form

$$\frac{1}{|\mathcal{M}^{(n)}|} 1_{\{X^n = f^{(n)}(m)\}}(x^n) \prod_{i=1}^{n} P_{Y|X}(y_i|x_i) 1_{\{\hat{M} = g^{(n)}(y^n)\}}(\hat{m}). \tag{2.15}$$

A rate R is said to be *achievable* if for any $\delta > 0$ and all sufficiently large n there exists an $(|\mathcal{M}^{(n)}|, n)$ code such that $\frac{1}{n} \log \left|\mathcal{M}^{(n)}\right| \geq R - \delta$ and the average error probability $P_e^{(n)} = \Pr(\hat{M} \neq M) \leq \delta$.

Definition 2.15 (Capacity)**.** The *capacity* C of a DMC is the supremum of all achievable rates.

Theorem 2.2.2 (Channel coding theorem)**.** *The capacity of the point-to-point DMC is given by*

$$C = \sup_{P_X(x)} I(X;Y), \tag{2.16}$$

where supremum is equal to maximum if the set $\mathcal{X}$ is finite, i.e., $|\mathcal{X}| < \infty$.

2.2.3 Rate-distortion With Side Information

In this section, we review the problem of lossy source coding with side information at the decoder, also known as the Wyner-Ziv problem [WZ76], as shown in Fig. 2.3. It is an extension of the point-to-point lossy source coding problem in Fig. 2.1 in that another sequence Y^n correlated with the source is assumed to be available at the decoder. This sequence is called *side information* as it provides some information about the source to the terminal that observes it. An encoder wishes to compress the source X^n into a source description W and send it to the decoder over a rate-limited link. The decoder, based on the description and the side information, reconstructs the source as $\hat{X}^n$ such that the constraint on the average distortion between the source and reconstruction is satisfied. When compared with the point-to-point lossy source coding in Fig. 2.1, we can see that with side information available at the decoder, the encoder can potentially improve the compression rate by sending a

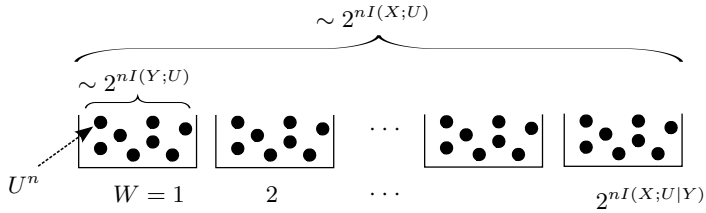

Figure 2.4: Binning for Wyner-Ziv coding.

source description with a lower rate to the decoder. Below we present the main result of the problem and discuss the coding scheme that can achieve it.

Theorem 2.2.3 (Wyner-Ziv theorem [WZ76])**.** *The rate-distortion region of the problem of lossy source coding with side information at decoder for a given DMS is described by the set of all* (R, D) *that satisfy*

$$R \geq I(X;U|Y), \tag{2.17}$$
$$D \geq E[d(X, \tilde{g}(U,Y))], \tag{2.18}$$

for some joint distributions of the form $P_X(x)P_{U|X}(u|x)P_{Y|X}(y|x)$ *with* $|\mathcal{U}| \leq |\mathcal{X}|+1$*, and a function* $\tilde{g} : \mathcal{U} \times \mathcal{Y} \to \hat{\mathcal{X}}$*.*

The common method of proving the result above (and also other network information theoretic problems) is to derive an achievable (rate-distortion) region which serves as an *inner bound* to the optimal rate-distortion region, and then show that the inner bound cannot be improved further by establishing a matching outer bound. The key idea used in the achievable scheme of the Wyner-Ziv theorem is to construct a code in such a way that it can exploit the knowledge of side information at the decoder. In particular, one uses a technique called *binning* illustrated in Fig. 2.4 which involves generating many codewords and partitioning them into a smaller number of equal-sized bins. The size of each bin is defined by the *quality* of side information at the decoder. To send the source description, the encoder only sends an index indicating the bin of the chosen codeword. One can see that upon receiving the bin index alone it is not sufficient for the decoder to decide which codeword U^n is chosen. However, together with the side information, the decoder can uniquely identify the chosen codeword. This method of sending only the bin index essentially reduces the rate of the source description and is shown to be optimal for the Wyner-Ziv problem.

2.2.4 Capacity of Channel With States

In this section, we review the problem of communicating a message over a state-dependent channel where the channel state is known at the encoder, as shown in

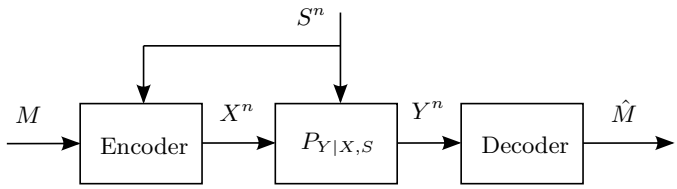

Figure 2.5: State-dependent channel with states known at the transmitter.

Fig. 2.5. The problem is known as the Gel'fand-Pinsker problem [GP80]. It is an extension of the point-to-point channel coding problem in the sense that if the state sequence is constant, it reduces to the point-to-point channel coding problem. In this setting, the channel output Y^n depends on both the state S^n and the input X^n with the distribution $P_{Y^n|X^n,S^n} = \prod_{i=1}^{n} P_{Y|X,S}(y_i|x_i, s_i)$. The encoder wishes to send a message M reliably over the channel to the decoder. The state S^n is distributed according to $\prod_{i=1}^{n} P_S(s_i)$ and is assumed to be known noncausally at the encoder, and the channel input X^n is a function of M and S^n. Here we see that the knowledge of S^n at the encoder can potentially be helpful to adjust the channel input X^n such that it *aligns* well with the channel.

Theorem 2.2.4 (Gel'fand-Pinsker theorem [GP80])**.** *The capacity of the discrete memoryless state-dependent channel with state known noncausally at the transmitter is given by*

$$C = \max[I(U;Y) - I(U;S)], \tag{2.19}$$

where the maximization is over $P_{U|S}(u|s)$ with $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{S}|$, and a function $\tilde{f} : \mathcal{U} \times \mathcal{S} \to \mathcal{X}$.

The achievable scheme for the Gel'fand-Pinsker theorem is based on the code construction that exploits the knowledge of channel state at the encoder. In particular, one uses a technique called *multicoding* which is *dual* to binning, i.e., we generate a set of codewords for each message. To send a message, we select one of the codewords corresponding to that message which satisfies a desired property defined by the channel state, i.e., one that is jointly typical with the channel state. The channel input is then chosen as a function of the selected codeword and channel state. This method of generating many codewords for each message is shown to be optimal for the Gel'fand-Pinsker problem.

2.3 Additional Aspects Introduced in the Problem Formulation

Abstract models in network information theory aim to represent real-world problems as accurate as possible, yet still tractable to solve. With this in mind, several

other aspects such as different features and objectives are additionally introduced into the problem settings, extending the basic problems, for example those introduced in Section 2.2, to the more relevant and application-oriented settings. In the following, we discuss some interesting aspects of the problem formulation that will be considered later in the thesis. The aim of this section is to discuss the extension ideas and their connections to corresponding chapters in the thesis, and also to give an overview of related works.

2.3.1 Action-dependent Side/State Information

Information availability at a certain node in the network is essential in network information theory problems. As we saw in Section 2.2.3, when side information is available at the decoder in the source coding problem, the minimum rate needed for lossy source compression can be decreased. Conventionally, side information is modeled as one given by nature. In practice however, it might be reasonable that some node in the network may have some control over the side information. After Weissman [Wei10] introduced action-dependent states in the problem of coding for channel with states in which we will discuss next, *action-dependent side information* was considered by Permuter and Weissman in [PW11] to model side information that can be controlled by a cost-constrained action sequence, i.e., the side information Y^n is distributed according to the distribution $P_{Y^n|X^n,A^n} = \prod_{i=1}^{n} P_{Y|X,A}(y_i|x_i, a_i)$. This framework provides another degree of freedom to the model and captures scenarios where a node can take actions with some cost to influence quality and/or availability of side information at the other node. This is relevant especially in controlled sensing when information acquisition is not available *for free*. Fig. 2.6 depicts a basic model of action-dependent side information in source coding problem which is essentially an extension of the Wyner-Ziv source coding problem to include a cost-constrained action sequence A^n. Below we state one of the main results in [PW11] for the lossy source coding with action-dependant side information.

Theorem 2.3.1 ([PW11]). *The rate-distortion-cost region of lossy source coding with action-dependent side information at the decoder for a DMS in Fig. 2.6 is given by the set of all (R, D, C) that satisfy*

$$R \geq I(X; A) + I(X; U|A, Y), \tag{2.20}$$

$$D \geq E[d(X, \tilde{g}(U, Y))], \tag{2.21}$$

$$C \geq E[\Lambda(A)], \tag{2.22}$$

for some joint distributions $P_X(x)P_{A|X}(a|x)P_{U|X,A}(u|x, a)P_{Y|X,A}(y|x, a)$ with $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{A}| + 1$, and a function $\tilde{g} : \mathcal{U} \times \mathcal{Y} \to \hat{\mathcal{X}}$, where $\Lambda(\cdot)$ is a bounded, single-letter cost measure.

Other extensions of action-dependent side information in source coding problems include [KOST10, CAW13, AS13, ZCW14]. In [KOST10], we considered a variant

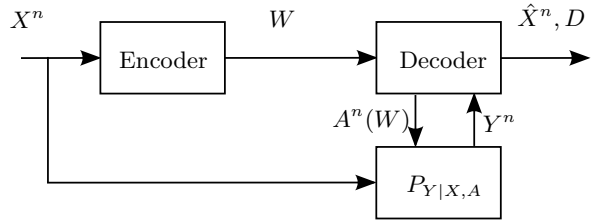

Figure 2.6: Source coding with action-dependent side information.

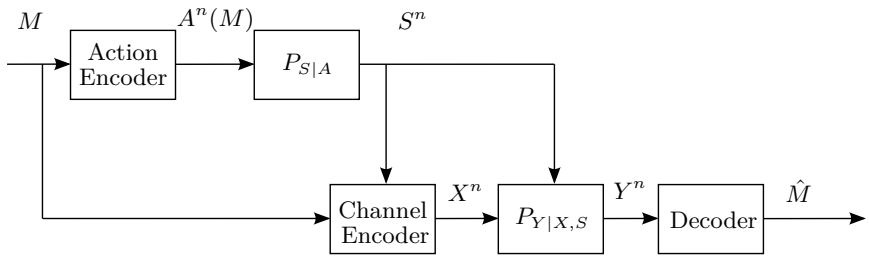

Figure 2.7: Channel with action-dependent states known at the transmitter.

of the problem with two-sided action-dependent side information at the encoder and decoder and gave the complete rate-distortion-cost region. This will also be presented in Chapter 3 of the thesis. Action-dependent side information in multi-terminal source coding was considered in [CAW13], while the distributed and cascade counterparts were studied in [AS13]. [ZCW14] recently considered action in a different compression model.

As mentioned before, Weissman in [Wei10] also introduced the notion of *action-dependent states* in the channel coding problem. In this setting the channel states can be influenced by a message-dependent action sequence A^n, i.e., $P_{S^n|A^n} = \prod_{i=1}^{n} P_{S|A}(s_i|a_i)$. It has several relevant applications, for example, in networked control where a message to be communicated affects the channel state via the control action, or in memory storage with rewrite option where the action sequence can be seen as an input to the first-stage writing. Fig 2.7 depicts a basic model of action-dependent states in channel coding problem which is essentially an extension of the Gel'fand-Pinsker problem to include an action sequence.

Theorem 2.3.2 ([Wei10]). *The capacity of a discrete memoryless state-dependent channel with action-dependent state known at transmitter is given by*

$$C = \max[I(A, U; Y) - I(U; S|A)], \tag{2.23}$$

where the maximization is over $P_A(a)P_{U|S,A}(u|s, a)$ *with* $|\mathcal{U}| \leq |\mathcal{A}||\mathcal{X}||\mathcal{S}| + 1$, *and a function* $f : \mathcal{U} \times \mathcal{S} \to \mathcal{X}$.

Other extensions of action-dependent states in channel coding problem include [KOST10, APW11, AASP13, SW12, Ste13, CM12a, CM12b]. In [KOST10], we considered a variant of the problem setting with two-sided action-dependent states at the encoder and decoder and characterized the capacity of such channel. [APW11] considered a variant of the problem setting where the channel states is given by nature, but can be acquired by an action probing. [AASP13] considered an extension of [Wei10] to include an information embedding on action into the model. [SW12] and [Ste13] extended action-dependent states to the broadcast channel and obtained the capacity for the case where states is known causally at the encoder. [CM12a] and [CM12b] considered the adaptive action case where the action is taken based on the past states and showed that allowing feedback does not improve the capacity.

2.3.2 Additional Reconstruction Requirement

Conventional source and channel coding problems are concerned with a single reconstruction objective. For example, in lossy source coding, an encoder wishes to compress a source so that it can be reconstructed at a decoder within a certain distortion, as specified by a distortion criterion, while in channel coding, a transmitter wishes to transmit a message over a noisy channel such that it can be reconstructed reliably at a receiver. In some scenarios though, it may be reasonable to introduce additional reconstruction requirements into the model of the system. There are several possible such requirements one can impose. In the following, we present some of the interesting ones which could play important roles in future network as discussed in Chapter 1, and which will be considered later in the thesis.

Common Reconstruction

In a lossy source coding/compression problem, the receiver reconstructs the source sequence based on the source description and possibly with the help of side information so that the reconstruction satisfies a distortion criterion. This means that the reconstruction at the receiver is not necessary the same as the source sequence at the sender. In some scenarios, especially those where the reconstruction sequence is sensitive information, the sender might want to know exactly which reconstruction sequence the receiver has produced. For example, when an online medical consultation is modeled as a lossy source coding problem, a source sequence could be a Magnetic Resonance Imaging (MRI) scan relating to a patient. In this case, the sender may want to know exactly what the reconstruction image at the receiver is so that he/she can prepare for a retransmission if necessary. The requirement that the sender can estimate the reconstruction sequence at the receiver is termed as *common reconstruction* and was first considered in the context of lossy source coding with side information in [Ste09]. Below we present one of the main results in [Ste09] on the Wyner-Ziv problem with common reconstruction, as shown in Fig. 2.8.

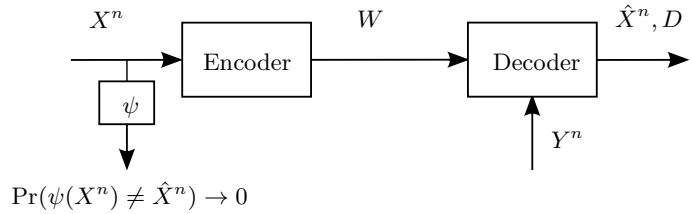

Figure 2.8: Source coding with common reconstruction constraint.

Theorem 2.3.3 ([Ste09]). *The rate-distortion region for the problem of lossy source coding with side information at decoder under a common reconstruction constraint for a DMS is given by the set of all* (R, D) *that satisfy*

$$R \geq I(X; \hat{X}|Y), \tag{2.24}$$

$$D \geq E[d(X, \hat{X})], \tag{2.25}$$

for some joint distributions of the form $P_X(x)P_{\hat{X}|X}(\hat{x}|x)P_{Y|X}(y|x)$.

Let $\mathcal{R}_{CR}$ and $\mathcal{R}_{WZ}$ denote the rate-distortion regions for lossy source coding with side information at the decoder under a common reconstruction in Theorem 2.3.3, and for the Wyner-Ziv problem in Theorem 2.2.3, respectively. Due to the additional common reconstruction constraint, $\mathcal{R}_{CR}$ is generally smaller than $\mathcal{R}_{WZ}$. This can be seen from the fact that the set of input distributions in Theorem 2.3.3 is smaller than that of Theorem 2.2.3. For example, when we set $\hat{X} = U$ in Theorem 2.2.3, $\mathcal{R}_{WZ}$ reduces to $\mathcal{R}_{CR}$. This restriction on the input distribution in the rate-distortion region in Theorem 2.3.3 reflects the fact that a random sequence Y^n which is used for reconstruction at the decoder is not known at the encoder. For the achievable scheme, we utilize Y^n for binning to reduce the rate as in the Wyner-Ziv coding. However, in order to satisfy the common reconstruction requirement, one needs to restrict the reconstruction symbol such that it does not depend on the side information symbol. Essentially, we can conclude that it is optimal to use the side information Y^n only for binning to reduce the rate, but not for generating the reconstruction symbols.

We note that the common reconstruction requirement is already satisfied for any lossless settings, and for the point-to-point lossy source coding since there is no additional randomness to be used for signal reconstruction at the receiver that is not known at the transmitter.

Reversible Input

Similarly as in the case of a common reconstruction constraint in the lossy source coding problem, we can consider an additional reconstruction requirement for the channel input sequence in the problem of coding for channel with states. Fig. 2.9

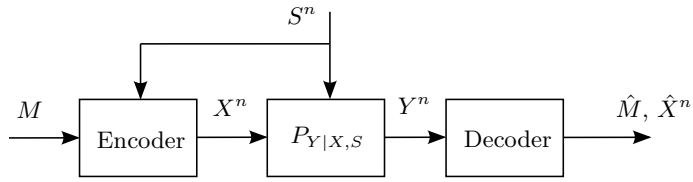

Figure 2.9: State-dependent channel with reversible input constraint.

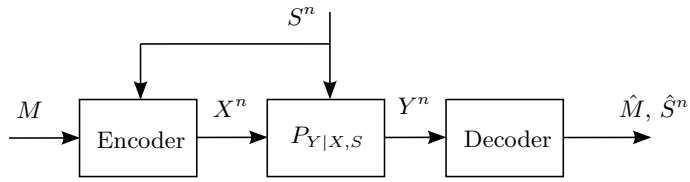

Figure 2.10: State-dependent channel with state amplification constraint.

illustrates the problem of coding for channel with states known at the transmitter under *reversible input* requirement. It has been studied in [SS09] and the channel capacity was found in the context of information embedding where the message is considered as a watermark message, the state as a host signal, and the channel input as a composite or stegotext signal. The setting with reversible input requirement can be motivated for example in scenarios where it is desirable for the receiver to be able to decode also the composite signal and reuse it for the next stage transmission or for possible retransmission [SS09]. In Chapter 4, we study the aspect of reversible input requirement in the problem of multi-stage writing on memory with rewrite option where the reversible input enables the tracking capability of the system, i.e., tracking what have been written in the past stages.

State Amplification

In the problem of coding for channel with states known at the transmitter, in addition to decoding the message, the receiver may want to decode as well the channel state sequence. This is motivated by the fact that the channel state affects the quality of the channel, and having its knowledge at the receiver could be useful in the next stage transmission. The requirement that the state should be decoded within a list at the receiver is termed as *state amplification* and is studied in [KSC08]. We can see that the problem of state amplification is closely related to the reversible input requirement discussed earlier. For example, if the encoding function is deterministic, having decoded the message and the state correctly implies that the channel input is also decoded correctly. Fig. 2.10 illustrates the problem of coding for channel with states known at the transmitter under state amplification

requirement.

2.3.3 Information Privacy Requirement

Instead of requiring additional information to be reconstructed at a certain node, we might want to *hide* or *protect* the information from a particular node in the network. The concept of information privacy is highly relevant already today and expectedly even more so in future networks where almost everyone and everything will be connected. Nodes want to share information with each other for possible cooperation, but at the same time they wish to reveal as little information to a certain node as possible. Information security can be addressed at different layers of the system. For example, we use encryption to protect information in the link layer. Nevertheless, an important concern of security in communication networks is at the physical layer where the communication channel is vulnerable to eavesdropping. It is therefore natural to consider also a *physical layer security* which can complement the level of security in general. More recently, due to potential applications in areas such as privacy in sensor networks and databases (see, e.g., [SRP13]), and privacy of distributed storage of genomic data (see, e.g., [GMG$^+$13], [KBLV13]), an idea of physical layer security from the compression perspective has also been studied. In the following, we present some of the information theoretic security and privacy requirements which are considered in the source coding problems. We use the Wyner-Ziv problem as a basis and assume that there exists an eavesdropper who can observe the source description and its own side information. An additional privacy constraint is imposed at the *eavesdropper* where we wish to protect the source sequence, decoder's side information sequence, or decoder's reconstruction sequence. Throughout the thesis, we assume that the eavesdropper is of passive type, i.e., it does not tamper with the transmission, and we use *information leakage* or *equivocation* as a measure of secrecy or privacy of our information. Information leakage is defined as a normalized mutual information between the signal targeted by the eavesdropper and those known at the eavesdropper. Similarly, equivocation is defined as a normalized conditional entropy of the signal targeted by the eavesdropper conditioned on the signals which are known at the eavesdropper.

For the case of active eavesdropper or attacker, there exist several works which consider different objectives and approaches in the problem formulation. For example, a game theoretic approach can be applied to model and investigate the scenarios where the legitimate users and attacker in the network act with conflicting goals. Readers are referred to the existing literature on this topic, e.g., [AB10]. Another approach is to model malicious behavior by associating attackers with a different type of a utility function which represents a gain of attacker at the expense of performance degradation of the legitimate party. This can be done by formulating the problem with a joint payoff function with respect to the legitimate users. For example, the legitimate users may want to maximize the minimum payoff achieved by the attacker. This line of studies was considered for example in [Yam97, Cuf10, SC13], etc.

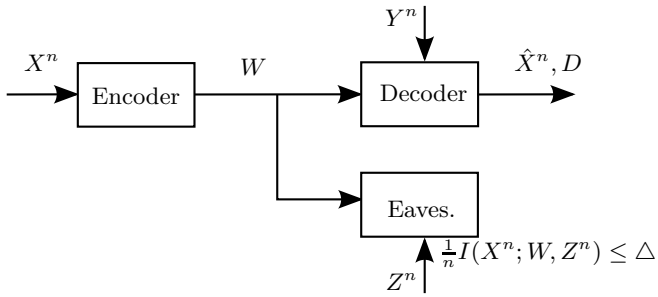

Figure 2.11: Secure source coding with side information.

Source Privacy

In the lossy source coding problem, the encoder wishes to compress the source into a source description and send it over a rate-limited link to the decoder. Now we assume that there is an eavesdropper who can observe the source description and thus can learn about the source, as shown in Fig. 2.11. In this problem, our goal is to communicate the source over a rate-limited link to the decoder satisfying the distortion criterion, and at the same time, reveal only limited knowledge about the source to the eavesdropper. Clearly, there is a tradeoff between the reconstruction quality and information privacy. For example, the encoder may want to send a high-rate source description to the decoder in order to achieve a low distortion. However, this in turn leads to large amount of information leakage to the eavesdropper. The objective of the problem described above (also called *secure source coding* problem) is to characterize the optimal tradeoff between minimum rate needed to describe the source, incurred distortion at the decoder, and the leakage rate at the eavesdropper. The problem of secure distributed lossless source coding was considered in [PR07, GEP08, TUR13], etc. [VP13] then extended to the lossy case and characterized a complete optimal tradeoff between rate, distortion, and equivocation for some special cases. In [EU11] the authors considered a new measure for source secrecy using the relative equivocation at the eavesdropper with respect to the legitimate decoder, and showed that the achievable scheme for the basic setting in [VP13] is also optimal for this new measure. The new measure is considered to be a natural generalization of the equivocation in a wiretap channel to a secure lossy source coding context. Several other works considered source privacy/secrecy in the source networks with an eavesdropper, for example, [CK13a] considered source secrecy in a setting with side information at the encoder, [NSTS13] considered source secrecy in the lossless CEO problem [BZV96]. In some cases, it is also important to impose information privacy constraint at the legitimate user. [Cou12] considered multi-terminal source coding with a requirement of amplifying one source and masking another at the decoder. We in [KCO$^+$13b] considered a source network under a

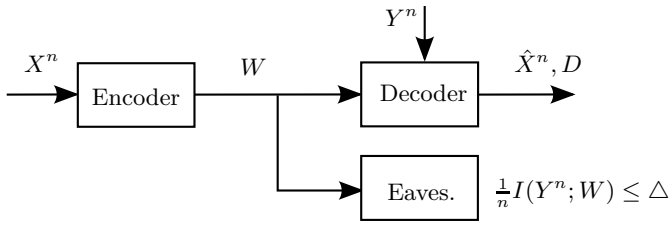

Figure 2.12: Source coding with side information privacy.

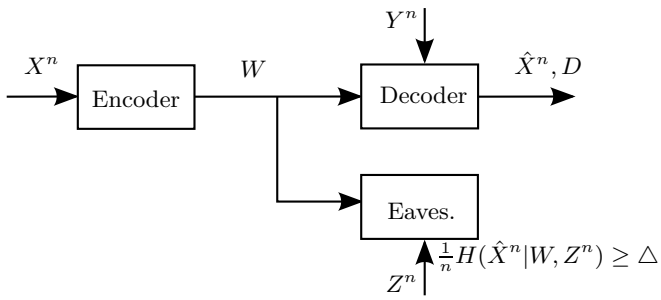

Figure 2.13: Source coding with reconstruction privacy.

source privacy constraint at the public helper which will be presented in Chapter 6. Another line of works considered source secrecy in the settings with explicit common secrete key between the encoder and decoder [Sha49, Yam97, Mer08, Cuf10, SC13], etc. In Chapters 5 and 6, we study some related settings where the encoder and decoder can use the common side information to generate a secret key and use it for *encrypting* the source description.

Side Information Privacy

One application of the Wyner-Ziv problem is in databases where we may view the source and side information as different but correlated databases available at the sender and receiver. When the transmitter shares its database over the rate-limited public link, the eavesdropper observing the link can try to learn about both databases. The privacy constraint on the database at the receiver in this model, as shown in Fig. 2.12, is termed as *side information privacy* in [TSP13].

Reconstruction Privacy

In some scenarios, it is the estimate of the source (reconstruction sequence) at the receiver rather than the source itself that should be protected against the

eavesdropper. For example, let us consider a distributed cloud services. An end-user (decoder) receives information over rate-limited links from the cloud service providers (encoders), and makes a final action/decision (reconstruction sequence) such that it satisfies a certain distortion criterion. It is of interest that this final action/decision should be kept private from the providers, at least to a certain extent. We introduce the notion of *end-user privacy* as a privacy measure on the receiver's reconstruction at the other node in the network, and study it in the Wyner-Ziv setting, as shown in Fig. 2.13, in Chapter 7. The idea of protecting the reconstruction sequence against an eavesdropper was first considered in the context of watermarking in [Mer06a], [Mer06b]. It was also considered in the setting where the encoder and decoder share a common secret key [SC13].

2.3.4 Summary

Below we give a summary of how each chapter in the thesis is related to different aspects of the general problem formulations discussed in Section 2.3.

- Chapter 3 considers two main problems involving *action-dependent side information/states* and *additional reconstruction requirement*, namely, a problem of source coding with two-sided action dependent side information under common reconstruction constraint, and a problem of channel coding with two-sided action-dependent state under reversible input constraint.

- Chapter 4 extends the channel coding problem part in Chapter 3 to a multi-stage setting.

- Chapter 5 studies lossy source coding problem with *action-dependent side information* under *source privacy constraint*. Several aspects of the action-dependent side information are considered, namely, the cases where action is taken at either the encoder or decoder, and the case where action is taken to generate a common two-sided side information.

- Chapter 6 considers the source coding problem with a *source privacy constraint* in the presence of a public helper. In this chapter, we are interested in how the system can utilize a helper to support the transmission in a scenario that the helper is a terminal which can leak information through its public link or that the helper itself is a public terminal to which we do not want to reveal much information.

- Chapter 7 studies the problem of lossy source coding with *reconstruction privacy*. In this chapter, we extend the Wyner-Ziv problem to include a privacy constraint on the reconstruction sequence at the decoder, termed as end-user privacy, at different nodes in the system, e.g., at the eavesdropper, encoder, and helper.

Chapter 3

Coding With Action-dependent Side Information and Reconstruction Requirement

This chapter studies two classes of source/channel coding problems, namely coding with action-dependent side information, and coding with additional signal reconstruction, in a unified fashion. In the source coding setting, a decoder wishes to reconstruct the source subject to a distortion constraint, while an encoder is required to estimate the decoder's reconstruction reliably. Side information is action-dependent in the sense that its quality and/or availability at the encoder or decoder can be influenced by a cost-constrained action sequence. In the channel coding "dual," the decoder wishes to decode both message and channel input sequence reliably, and the channel state information available at the encoder or decoder are assumed to depend on the action sequence. We consider discrete memoryless systems and characterize single letter expressions of the rate-distortion-cost function and channel capacity for the respective source and channel coding problems. The dual relation between the two problems and the *two-stage coding condition* which appears in the channel capacity expression are also discussed.

3.1 Introduction

Problems involving source and channel coding with side or state information were recently generalized to capture a new interaction between nodes in the form of side or state information acquisition. Instead of assuming that side information/channel state are given by nature, Weissman and Permuter [Wei10], [PW11] introduced coding with *action-dependent* side/state information in which the availability and/or quality of side/state information can be influenced by other nodes in the system via a cost-constrained action sequence. This novel action-dependent coding framework introduces new interesting features to the general system model involving

cost-constrained communication. The action sequence has dual roles, namely, communicating the message/source description, and controlling the channel state/side information. It is therefore highly relevant to many applications including sensor networking and control, and multi-stage coding for memories [Wei10], [PW11].

Meanwhile, problems of coding with side/state information were also generalized to include a new reconstruction requirement. For example, lossy source coding with side information under the additional requirement that a sender should be able to locally reproduce an exact copy of the receiver's reconstruction was introduced by Steinberg [Ste09]. The additional reconstruction requirement is termed as the *common reconstruction* (CR) constraint. As with action-dependent coding, also the framework of additional reconstruction requirement provides new useful features of simultaneous signal transmission in the general system model.

In this chapter, we unify the problem of action-dependent side information and the additional reconstruction framework by studying problems involving source and channel coding with action-dependent partial side information known non-causally at the encoder and decoder under additional reconstruction requirements. Our source coding problem can be viewed as an extension of Wyner-Ziv lossy source coding [WZ76], where the encoder is additionally required to estimate the decoder's reconstruction reliably (CR constraint) and the available two-sided partial side information depends on a cost-constrained action sequence. Clearly, there exists a tradeoff between the quality of the decoder's reconstruction and the CR constraint, based on the presence of side information at the decoder, i.e., side information at the decoder can be used for reconstructing the source, while it also introduces some randomness into the reconstruction sequence that is supposed to be estimated at the encoder. Since the side information can be influenced with some cost, this gives rise to a new interesting tension between the action cost and the mentioned tradeoff between the attainable distortion and CR constraint. Settings with CR constraints can be relevant in communication scenarios where sensitive information is involved. For example, in online medical consultation as discussed in Chapter 1, a sender might want to know exactly which version of the medical image a receiver reconstructs [Ste09]. In our considered setting, the sender has an additional degree of freedom to control the quality of receiver's reconstruction through the choice of cost-constrained action.

On the other hand, the channel coding dual is an extension of the Gel'fand-Pinsker problem [GP80], where the two-sided channel states are allowed to depend on an action sequence and the decoder is additionally required to reconstruct the channel input signal reliably. This setting captures the idea of simultaneously transmitting the message and the channel input sequence reliably over the channel. To be consistent with the terminology used in [SS09], which studied a similar problem without action-dependent states, we refer to the channel input reconstruction requirement as the *reversible input* (RI) constraint. This setup is for example relevant in the problem of multi-stage coding on memory storage where the user is interested in both decoding an embedded message and in tracking what has been written in previous stages.

3.1.1 Related Work

The problems of source coding with side information and channel coding with state information have received considerable attention due to their broad set of applications, e.g., in high-definition television where the noisy analog version of the TV signal is the side information at the receiver, in cognitive radio where the secondary user has knowledge of the message to be transmitted by the primary user, or in digital watermarking where the host signal plays a role of state information available at the transmitter [CC02], [KSM07], etc. In [WZ76] Wyner and Ziv considered rate-distortion coding for a source with side information available at the receiver, while the problem of coding for channels with noncausal state information available at the transmitter was solved by Gel'fand and Pinsker in [GP80]. In practice, the transmitter and/or the receiver may not have full knowledge of the channel state information. Heegard and El Gamal in [HE83] studied the channel with rate-limited noncausal state information available at the encoder and/or the decoder. Further, Cover and Chiang provided in [CC02] a unifying framework to characterize channel capacity and rate-distortion functions for systems with two-sided partial state information, and they also discuss aspects of duality between the source and channel coding problems. Different kinds of duality between various source and channel coding problems with side information have been recognized earlier. For example, *formula duality* between the two-sided extensions of the Wyner-Ziv and Gel'fand-Pinsker problems was discussed in [CC02] where there exists one-to-one correspondence between random variables in the expressions of the respective rate-distortion function and capacity. In [PCR03], *functional duality* between source and channel coding with side information was established under a condition on the joint distributions. This formulates the conditions under which the optimal encoder for one problem is functionally identical to the optimal decoder for the other problem such that the rate-distortion function is equal to the capacity-cost function. Similarly, [BCW03] considered duality between information embedding problem and the Wyner-Ziv problem in the sense that the optimal encoder-decoder for one problem is the optimal decoder-encoder pair for the other. Recently, the duality between classical lossy source coding and channel coding was explored in [GV11] in the *operational* sense, i.e., whether a capacity-achieving encoder-decoder sequence achieves the rate-distortion function of the dual problem when the channel decoder (encoder) is the source encoder (decoder, resp.), and vice versa.

As mentioned before in Section 2.3, Weissman studied first a problem of coding for a channel with action-dependent state [Wei10] which is an extension of the Gel'fand-Pinsker problem to allow the states to depend on the action sequence. It is closely related to the problem of multiple access channel (MAC) with common message and states at one transmitter [SBSV08] where the message is available at both action encoder and main channel encoder. It also has some connection to the relay with unlimited lookahead channel [EHM07] where the channel encoder plays a role of the relay encoder with noncausal observation of the states. The source coding dual to [Wei10] was investigated by Permuter and Weissman [PW11], where

a node in the system can take action to influence the quality/availability of the side information, extending the Wyner-Ziv problem. Additional works on coding with action include [APW11], where it is natural to consider action probing as a means for channel state acquisition, [CM12a], [CM12b], where the adaptive causal action-dependent state with feedback are considered, and [CAW13, ZCW14, AS13, AASP13, ASCM12, SW12, Ste13, DPS12], etc., where the problems of coding with action-dependent side information are extended to several multi-terminal cases such as multi-terminal source coding, distributed and cascade source coding, broadcast channel, and MAC.

The common reconstruction requirement was introduced by Steinberg [Ste09] in the context of lossy source coding with side information. The general case of additional reconstruction subject to the distortion constraint was later studied in [LMW11]. The channel coding dual is also investigated in the context of information embedding by Sumszyk and Steinberg in [SS09], where the decoder is interested in decoding both an embedded message and a stegotext signal. Recent works on common reconstruction in multi-terminal information theoretic problems include [TGK13, ATSP13, TASP12]. Some closely related works on additional signal reconstruction include [KSC08] where the decoder is required to decode the message reliably and decode the encoder's state information within a list, and [WK03] where the decoder is interested in decoding both the message and the encoder's state information reliably.

3.1.2 Summary of Results and Organization

Below we provide an organization of the chapter along with a summary of our results.

- Section 3.2 considers the source coding problem depicted in Fig. 3.1 and provides a complete characterization of the rate-distortion-cost function for the discrete memoryless source. An example illustrating the rate penalty induced by the additional reconstruction requirement and the new tension between the action cost and the rate-distortion tradeoff is provided.

- Section 3.3 considers the channel coding problem depicted in Fig. 3.4 and provides a complete characterization of the channel capacity for the discrete memoryless channel. The obtained capacity expression is in the form under an additional *two-stage coding condition* which arises essentially from the two-stage operational structure of the setting that requires channel input reconstruction. We show by example that there exists a case where the two-stage coding condition can be active in restricting the set of capacity achieving input distributions.

- Section 3.4 discusses dual relationship between the two studied problems. Our definition of duality simply follows the notion of "formula" duality in [CC02]. Although it is not based on a strict operational definition, it is appealing that

one might be able to anticipate the optimal solution of a new problem from its dual problem. Based on the obtained rate-distortion-cost function and channel capacity of the problems, we conclude that the formula duality between our problems does not hold in general. This is partly due to some fundamental differences in their operational structures such as causal processing based on available signals.

- Section 3.5 concludes the chapter.

3.2 Source Coding With Action-dependent Side Information Under CR Constraint

In this section, we study source coding with action-dependent side information and CR constraint, as depicted in Fig. 3.1. The side information is generated based on the source and cost-constrained action sequences, and is given at both encoder and decoder. The decoder reconstructs the source sequence subject to the distortion constraint. Meanwhile, the encoder is required to locally produce an exact copy of the decoder's reconstruction. This setting captures the idea of simultaneously controlling the quality of the decoder's reconstruction via action-dependent side information, and monitoring the resulting performance via common reconstruction. It can also be considered as a combination of Permuter and Weissman's source coding with side information "vending machine" [PW11] and Steinberg's coding and common reconstruction [Ste09].

In the following, we present the problem formulation, characterize the rate-distortion-cost function for a discrete memoryless source, and also present some other related results. Finally, a binary example is given to illustrate an implication of the action and common reconstruction on the rate-distortion-cost tradeoff.

3.2.1 Problem Formulation

We consider finite alphabets for the source, action, side information, and reconstruction sets, i.e., $\mathcal{X}$, $\mathcal{A}$, $\mathcal{S}_e$, $\mathcal{S}_d$, and $\hat{\mathcal{X}}$ are finite. Let X^n be a source sequence of length n with i.i.d. elements according to P_X. Given a source sequence X^n, an encoder generates an index representing the source sequence and sends it over a noise-free, rate-limited link to an action encoder and a decoder. An action sequence is then selected based on the index. With input (X^n, A^n) whose current symbols do not depend on the previous channel output, the side information (S_e^n, S_d^n) is generated as an output of the memoryless channel with transition probability

$$P_{S_e^n,S_d^n|X^n,A^n}(s_e^n, s_d^n|x^n, a^n) = \prod_{i=1}^{n} P_{S_e,S_d|X,A}(s_{e,i}, s_{d,i}|x_i, a_i).$$

The side information is then mapped to the partial side information for the encoder and the decoder by the mappings $l_e^{(n)}(S_e^n, S_d^n) = S_e^n$ and $l_d^{(n)}(S_e^n, S_d^n) = S_d^n$.

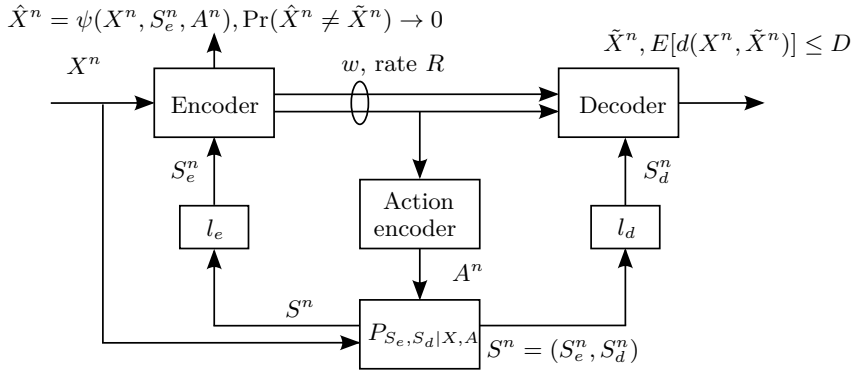

Figure 3.1: Rate distortion with action-dependent partial side information and CR constraint.

Next, the encoder uses the knowledge about S_e^n to generate another index and sends it to the decoder. Given the indices and the side information S_d^n, the decoder reconstructs the source sequence as $\tilde{X}^n$. Under the CR constraint, the encoder also estimates the decoder's reconstruction as $\hat{X}^n$.

Definition 3.1. A $(|\mathcal{W}^{(n)}|, n)$-code for a memoryless source with partially known two-sided action-dependent side information under a CR constraint consists of the following functions

- encoder one $f_1^{(n)} : \mathcal{X}^n \to \mathcal{W}_1^{(n)}$,
- an action encoder $f_a^{(n)} : \mathcal{W}_1^{(n)} \to \mathcal{A}^n$,
- encoder two $f_2^{(n)} : \mathcal{X}^n \times \mathcal{S}_e^n \to \mathcal{W}_2^{(n)}$,
- a decoder $g^{(n)} : \mathcal{W}_1^{(n)} \times \mathcal{W}_2^{(n)} \times \mathcal{S}_d^n \to \hat{\mathcal{X}}^n$, and
- a CR mapping $\psi^{(n)} : \mathcal{X}^n \times \mathcal{S}_e^n \times \mathcal{A}^n \to \hat{\mathcal{X}}^n$,

where $\mathcal{W}_1^{(n)}$ and $\mathcal{W}_2^{(n)}$ are finite sets, and $|\mathcal{W}^{(n)}| = |\mathcal{W}_1^{(n)}| \cdot |\mathcal{W}_2^{(n)}|$. We assume here that $|\tilde{\mathcal{X}}| = |\hat{\mathcal{X}}|$.

Let $d : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty)$ and $\Lambda : \mathcal{A} \to [0, \infty)$ be the single-letter distortion and cost measures. The distortion between a length-n source sequence and its reconstruction at the decoder, and the cost of action are defined as

$$d^{(n)}\big(X^n, \tilde{X}^n\big) \triangleq \frac{1}{n} \sum_{i=1}^{n} d(X_i, \tilde{X}_i),$$

$$\Lambda^{(n)}(A^n) \triangleq \frac{1}{n} \sum_{i=1}^{n} \Lambda(A_i),$$

where $d^{(n)}(\cdot)$ and $\Lambda^{(n)}(\cdot)$ are the distortion and cost functions, respectively.

The average probability of error in estimating the decoder's reconstruction sequence is defined by

$$P_{\mathrm{CR}}^{(n)} = \Pr\big(\psi^{(n)}(X^n, S_e^n, A^n) \neq g^{(n)}(f_1^{(n)}(X^n), f_2^{(n)}(X^n, S_e^n), S_d^n)\big).$$

Definition 3.2. A rate-distortion-cost triple (R, D, C) is said to be *achievable* if for any $\delta > 0$, there exists for all sufficiently large n a $(|\mathcal{W}^{(n)}|, n)$-code such that $\frac{1}{n}\log|\mathcal{W}^{(n)}| \leq R+\delta$, $E\big[d^{(n)}(X^n, \tilde{X}^n)\big] \leq D+\delta$, $E\big[\Lambda^{(n)}(A^n)\big] \leq C+\delta$, and $P_{\mathrm{CR}}^{(n)} \leq \delta$. The *rate-distortion-cost function* $R_{\mathrm{ac,cr}}(D, C)$ is the infimum of the achievable rates at distortion level D and cost C.

3.2.2 Main Result

Theorem 3.2.1 (Rate-distortion-cost function)**.** *The rate-distortion-cost function for the source coding problem in Fig. 3.1 is given by*

$$R_{ac,cr}(D, C) = \min[I(X; A) + I(\hat{X}; X, S_e|A) - I(\hat{X}; S_d|A)], \tag{3.1}$$

$$= \min[I(X; A) + I(\hat{X}; X, S_e|A, S_d)], \tag{3.2}$$

where the joint distribution of $(X, A, S_e, S_d, \hat{X})$ *is of the form*

$$P_X(x)P_{A|X}(a|x)P_{S_e,S_d|X,A}(s_e, s_d|x, a)P_{\hat{X}|X,S_e,A}(\hat{x}|x, s_e, a)$$

and the minimization is over all $P_{A|X}$ *and* $P_{\hat{X}|X,S_e,A}$ *subject to*

$$E\big[d(X, \hat{X})\big] \leq D, \qquad E[\Lambda(A)] \leq C.$$

Proof. The proof follows similar arguments as in [PW11] with some modifications in which we extend the side information transition probability to the two-sided side information $P_{S_e,S_d|X,A}$, and consider the additional CR constraint at the encoder as in [Ste09]. In the following, we give a sketch of the achievability proof: An action codebook $\{a^n\}$ of size $2^{n(I(X;A)+\delta_\epsilon)}$ is generated i.i.d. $\sim P_A$. For each a^n another codebook $\{\hat{x}^n\}$ of size $2^{(n(I(\hat{X};X,S_e|A)+\delta_\epsilon))}$ is generated i.i.d. $\sim P_{\hat{X}|A}$. These codewords are then distributed at random into $2^{n(I(\hat{X};X,S_e|A)-I(\hat{X};S_d|A)+2\delta_\epsilon)}$ equal-sized bins. Given the source sequence x^n the encoder in the first step uses $n(I(X; A)+\delta_\epsilon)$ bits to transmit an index representing the action codeword a^n which is jointly typical with x^n to the decoder. Then the action-dependent side information is generated based on x^n and a^n. Given x^n, s_e^n and previously chosen a^n, the encoder in the second step uses another $n(I(\hat{X}; X, S_e|A) - I(\hat{X}; S_d|A) + 2\delta_\epsilon)$ bits to communicate the bin index of the jointly typical codeword $\hat{x}^n$. In addition, the encoder produces this jointly typical $\hat{x}^n$ as an estimate of the decoder's reconstruction (common reconstruction). Given the identity of a^n, the bin index of $\hat{x}^n$, and the side information s_d^n, the decoder will find with high probability the unique codeword $\hat{x}^n$ in its bin that is jointly typical with s_d^n and a^n so that the decoder reconstructs $\tilde{x}^n = \hat{x}^n$ with high probability. The complete proof is provided in Appendix 3.A. □

Lemma 3.1 (Convexity of $R_{\mathrm{ac,cr}}(D,C)$)**.** *The rate-distortion-cost function given in Theorem 3.2.1 is a non-increasing convex function of distortion D and cost C.*

Proof. Since the domain of minimization in (3.1) increases with D and C, we have that $R_{\mathrm{ac,cr}}(D,C)$ is non-increasing in D and C. For convexity, the proof follows a standard time-sharing argument (see, e.g., [CT06, Lemma 15.9.1]) and is therefore omitted. □

Remark 3.1 (Other results)**.** Theorem 3.2.1 relates to other known results in the literature:

(i) When the CR constraint is omitted, our setting resembles the source coding with action-dependent side information setup of [PW11]. The rate-distortion-cost function in this case can be derived along the lines of Theorem 3.2.1 and is given by

$$R_{\mathrm{ac}}(D,C) = \min[I(X;A) + I(U;X,S_e|A) - I(U;S_d|A)], \tag{3.3}$$

where the minimization is over all $P_{A|X}, P_{U|X,S_e,A}$ and $\tilde{g} : \mathcal{U} \times \mathcal{S}_d \to \hat{\mathcal{X}}$ subject to $E\big[d\big(X, \tilde{g}(U,S_d)\big)\big] \leq D$, and $E[\Lambda(A)] \leq C$, and U is the auxiliary random variable with $|\mathcal{U}| \leq |\mathcal{A}||\mathcal{X}| + 3$. The achievability proof is a straightforward modification of that of Theorem 3.2.1 where the codeword U^n is used instead of $\hat{X}^n$ and the decoding function $\tilde{g}$ is introduced. The cardinality bound for $\mathcal{U}$ follows from standard arguments using the support lemma [CK11, Lemma 15.4], [EK11, Appendix C].

Unlike in the rate-distortion-cost function above, the optimization domain in Theorem 3.2.1 is restricted to a smaller set where the reconstruction symbol does not depend on the decoder's side information. This indicates the fact that the side information at the decoder introduces some randomness into the decoder's reconstruction sequence which adversely affects the CR constraint. In our setting in Fig. 3.1, the side information at the decoder can be influenced by the cost-constrained action sequence. Therefore, the rate-distortion-cost function in Theorem 3.2.1 provides new interesting insight into the tension between the action cost and the existing tradeoff between the decoder's reconstruction performance and CR constraint.

(ii) When we have no control over the side information, i.e., setting A constant, our setting simply recovers source coding with common reconstruction [Ste09].

Remark 3.2 (General rate-distortion-cost region)**.** In some scenarios where the links between encoder and decoder are different physical links with different budget (rate) constraints, we might be interested in characterizing the individual rate constraint in the form of a rate-distortion-cost region. Here we consider the same setting as in Fig. 3.1, but more generally we assume that the rate on the link used for generating the action sequence is denoted by R_1, and the remaining rate from

the encoder to the decoder is denoted by R_2. The rate-distortion-cost region in this case, defined as the set of all achievable tuples (R_1, R_2, D, C), is given by the set of all (R_1, R_2, D, C) satisfying

$$R_1 \geq I(X; A) \tag{3.4}$$

$$R_1 + R_2 \geq I(X; A) + I(\hat{X}; X, S_e|A, S_d) \tag{3.5}$$

$$D \geq E[d(X, \hat{X})] \tag{3.6}$$

$$C \geq E[\Lambda(A)], \tag{3.7}$$

for some the joint distributions of the form

$$P_X(x)P_{A|X}(a|x)P_{S_e,S_d|X,A}(s_e, s_d|x, a)P_{\hat{X}|X,S_e,A}(\hat{x}|x, s_e, a).$$

The result is related to the successive refinement rate-distortion region [EK11, Chapter 13], [TD07] where we might consider the action sequence as a reconstruction sequence in the first stage, and the refinement stage involves the side information available at the encoder and the decoder (S_e, S_d). We also see that the rate-distortion-cost function in Theorem 3.2.1 is simply a constraint on the total rate $R = R_1 + R_2$ for a given distortion D and cost C. The proof is a modification of that of Theorem 3.2.1 where we consider instead the individual rate constraints. More specifically, the achievable scheme of Theorem 3.2.1 is modified so that the index W_1 is split into two independent parts $(W_{1,1}, W_{1,2})$, and the action sequence is selected based on only $W_{1,1}$. In the converse, the sum-rate constraint is the same as in the converse proof of Theorem 3.2.1, while the constraint on R_1 can be derived straightforwardly using the techniques from point-to-point lossy source coding [CT06, Chapter 10].

3.2.3 Binary Example

We consider an example of the rate-distortion-cost function in Theorem 3.2.1 for the special case where side information at the encoder is absent. Our example is a combination of examples in [PW11] and [Ste09] which are based on the Wyner-Ziv example [WZ76] and illustrate nicely the expected behavior of the rate-distortion tradeoff due to the implication of action-dependent side information with cost [PW11] and common reconstruction constraint [Ste09].

Let us consider binary source, reconstruction, action, and side information alphabets, i.e., $\mathcal{X} = \hat{\mathcal{X}} = \mathcal{A} = \mathcal{S}_d = \{0, 1\}$, where the source X is distributed according to Bernoulli(1/2). We assume that taking action $A = 1$ corresponds to observing the side information symbol S_d as an output of a binary symmetric channel with input X and crossover probability p_0 (BSC(p_0)), and $A = 0$ corresponds to not observing the side information. Also, an observation is assumed to have unit cost, i.e., $\Lambda(A) = A$ and $E[\Lambda(A)] = P_A(1) = C$, and the Hamming distance is considered as a distortion measure, i.e., $d(x, \hat{x}) = 1$ if $x \neq \hat{x}$, and 0 otherwise.

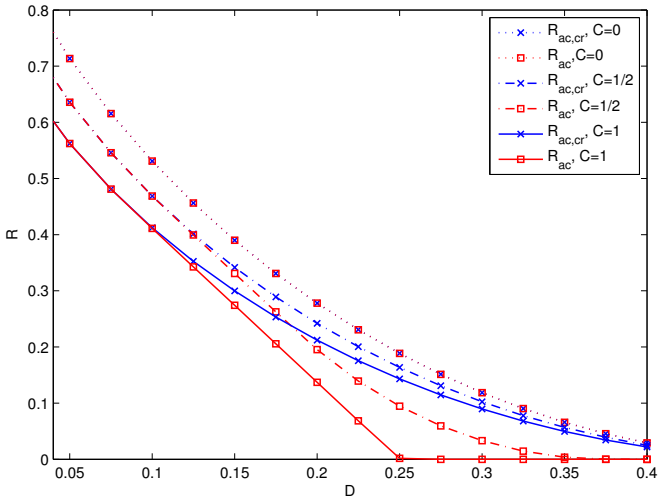

Figure 3.2: Rate-distortion curves for the binary symmetric source with common reconstruction and action-dependent side information available at the decoder. The markers $\times/\square$ correspond to the cases with/without CR constraint ($R_{\mathrm{ac,cr}}(D,C)/R_{\mathrm{ac}}(D,C)$). Different line styles correspond to different costs: dotted $C=0$, dashed-dotted $C=1/2$, and solid $C=1$. The rate-distortion tradeoff changes according to the action-cost C. Interestingly, at "low" D, the CR constraint does not increase the rate R.

The rate-distortion-cost function for this case is given as a special case of Theorem 3.2.1 when S_e is constant. We note that the second mutual information term in (3.2) neglecting S_e corresponds to the CR rate-distortion function [Ste09, Equation (8)] conditioned on A. Let D_i be the contribution to the average distortion given $A=i$, $i=0,1$, i.e., $(1-C)D_0+CD_1=D$. Thus, for this case,

$$R_{\mathrm{ac,cr}}(D,C)=\min_{P_{A|X},P_A(1)=C,(1-C)D_0+CD_1=D} I(X;A)+(1-C)\cdot R(P_{X|A=0},D_0) \\ +C\cdot R_{\mathrm{cr}}(P_{X,S_d|A=1},D_1), \tag{3.8}$$

where $R(P_X,D)$ denotes the rate-distortion function of the source P_X without side information and $R_{\mathrm{cr}}(P_{X,S_d},D)$ denotes the CR rate-distortion function defined in [Ste09] when source and side information are jointly distributed according to P_{X,S_d}.

It is interesting to compare $R_{\mathrm{ac,cr}}(D,C)$ to the rate-distortion-cost function of the case without the CR constraint $R_{\mathrm{ac}}(D,C)$ (a special case of (3.3) when neglecting S_e) to see how much we have to "pay" for satisfying the additional CR

constraint. In this case,

$$R_{\mathrm{ac}}(D,C) = \min_{P_{A|X},P_A(1)=C,(1-C)D_0+CD_1=D} I(X;A) + (1-C)\cdot R(P_{X|A=0},D_0) + C\cdot R_{\mathrm{wz}}(P_{X,S_d|A=1},D_1), \tag{3.9}$$

where $R_{\mathrm{wz}}(P_{X,S_d},D)$ denotes the Wyner-Ziv rate-distortion function when source and side information are jointly distributed according to P_{X,S_d}. The difference between (3.8) and (3.9) is only in their last terms.

In [Ste09, Example 1], the author computes the CR rate-distortion function for this source,

$$R_{\mathrm{cr}}(P_{X,S_d|A=1},D) = h(p_0 \star D) - h(D), \quad 0 \le D \le 1/2,$$

where $h(\cdot)$ is the binary entropy function and $p_0 \star D \triangleq p_0(1-D) + (1-p_0)D$. As known from [WZ76], the Wyner-Ziv rate-distortion function for this source is given by

$$R_{\mathrm{wz}}(P_{X,S_d|A=1},D) = \inf_{\theta,\beta} \big[\theta\big(h(p_0 \star \beta) - h(\beta)\big)\big],$$

for $0 \le D \le p_0$, where the infimum is with respect to all θ, β, where $0 \le \theta \le 1$ and $0 \le \beta \le p_0$ such that $D = \theta\beta + (1-\theta)p_0$. In addition, we know that for this source [CT06, Chapter 10]

$$R(P_{X|A=0},D) = 1 - h(D).$$

Using these results, we can compute (3.8) and (3.9), and compare $R_{\mathrm{ac,cr}}(D,C)$ and $R_{\mathrm{ac}}(D,C)$ to illustrate the consequences of the CR constraint. For a given $C = 0, 1/2$, and 1, and $p_0 = 1/4$, we plot the rate-distortion tradeoffs in Fig. 3.2. The plot shows that for given D and C there is a rate penalty in general as expected when the CR constraint is imposed since the optimization domain of $R_{\mathrm{ac,cr}}(D,C)$ is restricted to a smaller set such that the reconstruction symbol does not depend on the decoder's side information.

As discussed before, our setting with action-dependent side information at the decoder provides new insight on the tension between the action cost and the existing tradeoff between the decoder's reconstruction and CR constraint. In this example, we see that the rate-distortion tradeoff changes according to the action-cost as shown in Fig. 3.2 for different costs. At the extreme case when $C = 0$, $R_{\mathrm{ac,cr}}(D,C) = R_{\mathrm{ac}}(D,C)$ since there is no side information available at the decoder, and thus the decoder's reconstruction is a function of the source description which is known at the encoder. On the other extreme when $C = 1$, the side information at the decoder is always present. In this case, the plot resembles that in [Ste09, Example 1]. We also see that for a given $C > 0$ there is no rate penalty when imposing the CR constraint, i.e., $R_{\mathrm{ac,cr}}(D,C) = R_{\mathrm{ac}}(D,C)$, for "sufficiently low" values of D. This observation might be interpreted as follows. In the low distortion region, the side information at the decoder is so coarse that it is not helpful for the reconstruction function as

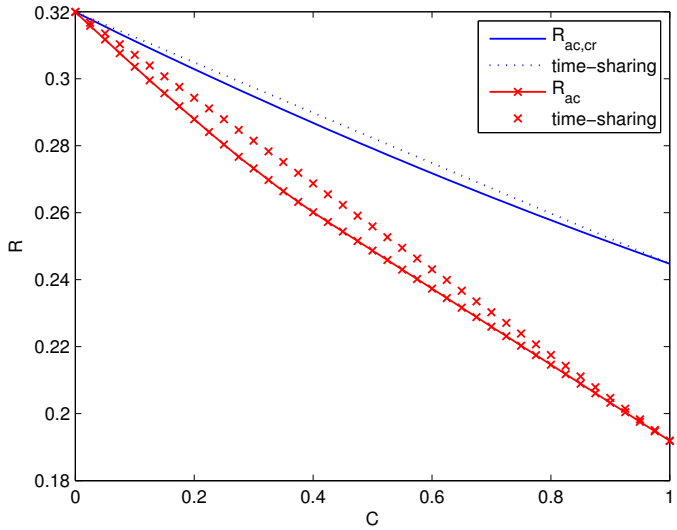

Figure 3.3: Rate-cost tradeoff for a given $D = 0.18$. The rate penalty of imposing the CR constraint increases with cost. The rate-cost tradeoff obtained from the time-sharing between $C = 0$ and $C = 1$ is suboptimal

compared to the source description. This observation and the characterization of D_c where $R_{ac,cr}(D, C) = R_{ac}(D, C)$ for $0 \leq D \leq D_c$ are also discussed in details in [Ste09, Example 1] for the case where $C = 1$. In general case where $0 < C < 1$, D_c can be obtained by solving the equation (3.8) = (3.9). The rate-distortion-cost functions in (3.8) and (3.9) have the form consisting of an optimized weighted-sum between the rate-distortion functions with and without side information which prevents an explicit expression for D_c.

In addition, for a given D, we can see the tradeoff between the rate R and cost C in $R_{\text{ac,cr}}(D, C)$ and $R_{\text{ac}}(D, C)$. In Fig. 3.3, we see that for the case $D = 0.18$ the rate penalty of imposing the CR constraint increases with cost. We note also that the rate-cost tradeoff for a given D obtained from the time-sharing between $C = 0$ and $C = 1$ is suboptimal.

3.3 Channel Coding With Action-dependent State and Reversible Input

In this section, we consider channel coding with action-dependent state, where the state is known partially and noncausally at the encoder and the decoder, as depicted in Fig. 3.4. In addition to decoding the message, the channel input X^n is

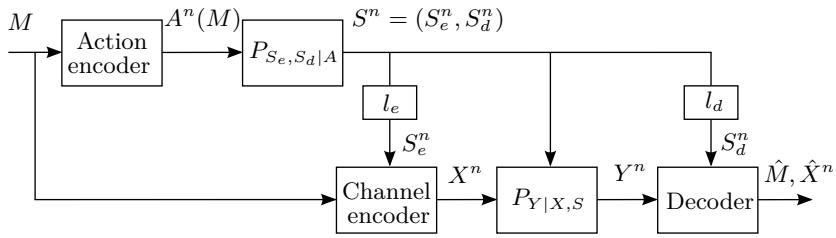

Figure 3.4: Channel with action-dependent state information and reversible channel input.

reconstructed with arbitrarily small error probability at the decoder (RI constraint). This setup captures the idea of simultaneously transmitting both the message and channel input sequence reliably over the channel. Our setup can be considered as a combination of Weissman's channel with action-dependent state [Wei10], and Sumszyk and Steinberg's information embedding with reversible stegotext [SS09]. It is also closely related to the problems of reversible information embedding [WK03] and state amplification [KSC08].

In the following, we present the problem formulation, characterize the capacity of a discrete memoryless channel, and also present some other related results. The channel capacity is given as a solution to a constrained optimization problem in a form having a constraint on the set of input distributions. We term this constraint the *two-stage coding condition* since it arises essentially from the two-stage structure of the encoding as well as the additional reconstruction constraint of a signal generated in the second stage. Also, we show in one example that such a constraint can be active in some cases, i.e., it actively restricts the set of capacity achieving input distributions, and when it is active, it will be satisfied with equality.

3.3.1 Problem Formulation

Let n denote the block length and $\mathcal{A}, \mathcal{S}_e, \mathcal{S}_d, \mathcal{X}$, and $\mathcal{Y}$ be finite sets. The system consists of two encoders, namely, an action encoder and a channel encoder, and one decoder. A message M chosen uniformly from the set $\mathcal{M}^{(n)} = \{1, 2, \ldots, |\mathcal{M}^{(n)}|\}$ is given to both encoders. An action sequence A^n is chosen based on the message M and is the input to the *state information channel*, described by a triple $(\mathcal{A}, P_{S_e,S_d|A}, \mathcal{S}_e \times \mathcal{S}_d)$, where $\mathcal{A}$ is the action alphabet, $\mathcal{S}_e$ and $\mathcal{S}_d$ are the state alphabets, and $P_{S_e,S_d|A}$ is the transition probability from $\mathcal{A}$ to $(\mathcal{S}_e \times \mathcal{S}_d)$. The channel state $S^n = (S_e^n, S_d^n)$ is mapped to the partial state information for the encoder and the decoder by the mappings $l_e^{(n)}(S_e^n, S_d^n) = S_e^n$ and $l_d^{(n)}(S_e^n, S_d^n) = S_d^n$. The input to the *state-dependent channel* is denoted by X^n. This channel is described by a quadruple $(\mathcal{X}, \mathcal{S}_e \times \mathcal{S}_d, P_{Y|X,S_e,S_d}, \mathcal{Y})$, where $\mathcal{X}$ is the input alphabet, $\mathcal{Y}$ is the output alphabet and $P_{Y|X,S_e,S_d}$ is the transition probability from $(\mathcal{X} \times \mathcal{S}_e \times \mathcal{S}_d)$ to $\mathcal{Y}$. The decoder, which might be considered as two separate decoders, i.e., a mes-

sage decoder and a channel input decoder, decodes the message and the channel input based on channel output Y^n and state information S_d^n. We assume that both state information and state-dependent channels are discrete memoryless and used without feedback with transition probabilities,

$$P_{S_e^n,S_d^n|A^n}(s_e^n, s_d^n|a^n) = \prod_{i=1}^{n} P_{S_e,S_d|A}(s_{e,i}, s_{d,i}|a_i),$$

$$P_{Y^n|X^n,S_e^n,S_d^n}(y^n|x^n, s_e^n, s_d^n) = \prod_{i=1}^{n} P_{Y|X,S_e,S_d}(y_i|x_i, s_{e,i}, s_{d,i}).$$

Definition 3.3. An $(|\mathcal{M}^{(n)}|, n)$ code for the channels $P_{S_e,S_d|A}$ and $P_{Y|X,S_e,S_d}$ consists of the following functions

- an action encoder $f_a^{(n)} : \mathcal{M}^{(n)} \to \mathcal{A}^n$,
- a channel encoder $f^{(n)} : \mathcal{M}^{(n)} \times \mathcal{S}_e^n \to \mathcal{X}^n$,
- a message decoder $g_m^{(n)} : \mathcal{Y}^n \times \mathcal{S}_d^n \to \mathcal{M}^{(n)}$, and
- a channel input decoder $g_x^{(n)} : \mathcal{Y}^n \times \mathcal{S}_d^n \to \mathcal{X}^n$.

The average probabilities of error in decoding the message M and the channel input X^n are defined by

$$P_{m,e}^{(n)} = \frac{1}{|\mathcal{M}^{(n)}|} \sum_{m,s_e^n,s_d^n,y^n : g_m^{(n)}(y^n,s_d^n) \neq m} p(y^n|f^{(n)}(m, s_e^n), s_e^n, s_d^n) \cdot p(s_e^n, s_d^n|f_a^{(n)}(m)),$$

$$P_{x,e}^{(n)} = \frac{1}{|\mathcal{M}^{(n)}|} \sum_{m,s_e^n,s_d^n,y^n : g_x^{(n)}(y^n,s_d^n) \neq f^{(n)}(m,s_e^n)} p(y^n|f^{(n)}(m, s_e^n), s_e^n, s_d^n) \cdot$$
$$p(s_e^n, s_d^n|f_a^{(n)}(m)).$$

Definition 3.4. A rate R is said to be *achievable* if for any $\delta > 0$ there exists for all sufficiently large n an $(|\mathcal{M}^{(n)}|, n)$-code such that $\frac{1}{n} \log |\mathcal{M}^{(n)}| \geq R - \delta$, $P_{m,e}^{(n)} \leq \delta$, and $P_{x,e}^{(n)} \leq \delta$. The *capacity* of the channel is the supremum of all achievable rates.

3.3.2 Main Result

Theorem 3.3.1 (Capacity)**.** *The capacity of the channel with action-dependent state available noncausally to the encoder and the decoder under RI constraint shown in Fig. 3.4 is given by*

$$C = \max[I(A, X; Y, S_d) - I(X; S_e|A)], \tag{3.10}$$

where the joint distribution of (A, S_e, S_d, X, Y) *is of the form*

$$P_A(a)P_{S_e,S_d|A}(s_e, s_d|a)P_{X|A,S_e}(x|a, s_e)P_{Y|X,S_e,S_d}(y|x, s_e, s_d)$$

and the maximization is over all P_A *and* $P_{X|A,S_e}$ *such that*

$$0 \leq I(X; Y, S_d|A) - I(X; S_e|A). \tag{3.11}$$

Proof. The achievability proof follows arguments in [Wei10] with a modification in which we use the channel input codeword x^n directly instead of the auxiliary codeword. In the following, we give a sketch of the achievability proof: An action codebook $\{a^n\}$ of size $2^{n(I(A;Y,S_d)-\delta_\epsilon)}$ is generated i.i.d. $\sim P_A$. For each a^n, another codebook $\{x^n\}$ of size $2^{n(I(X;Y,S_d|A)-\delta_\epsilon)}$ is generated i.i.d. $\sim P_{X|A}$. Then the codewords are distributed uniformly into $2^{n(I(X;Y,S_d|A)-I(X;S_e|A)-2\delta_\epsilon)}$ equal-sized bins. Given the message $m = (m_1, m_2)$, the action codeword $a^n(m_1)$ is selected. Then the channel states (s_e^n, s_d^n) are generated as an output of the memoryless channel with transition probability $P_{S_e^n,S_d^n|A^n}(s_e^n, s_d^n|a^n) = \prod_{i=1}^n P_{S_e,S_d|A}(s_{e,i}, s_{d,i}|a_i)$. The encoder looks for x^n that corresponds to m_1 and is in the bin m_2 such that it is jointly typical with the selected a^n and s_e^n. For sufficiently large n, with arbitrarily high probability, there exists such a codeword because there are approximately $2^{n(I(X;S_e|A)+\delta_\epsilon)}$ codewords in the bin. Then the selected x^n is transmitted over the channel $P_{Y|X,S_e,S_d}$. Given y^n and s_d^n, the decoder in the first step looks for codeword a^n that is jointly typical with y^n and s_d^n. With high probability, it will find one and it is the one chosen by the encoder since the codebook size is $2^{n(I(A;Y,S_d)-\delta_\epsilon)}$. Then, given the correctly decoded m_1, the decoder in the second step looks for x^n that is jointly typical with y^n, s_d^n, and a^n. Again, with high probability, it will find one and it is the one chosen by the encoder since the size of the codebook is $2^{n(I(X;Y,S_d|A)-\delta_\epsilon)}$. The corresponding bin index is then decoded as $\hat{m}_2$. In total, $I(A; Y, S_d) + I(X; Y, S_d|A) - I(X; S_e|A) - 3\delta_\epsilon$ bits per channel use can be used to transmit the message m such that both m and x^n are decoded correctly at the decoder. Note that the above coding scheme which splits the message into two parts and decodes them sequentially works successfully when we have a proper positive number of bins for codewords x^n, i.e., $I(X; Y, S_d|A) - I(X; S_e|A) - 2\delta_\epsilon > 0$. The complete proof is given in Appendix 3.B. □

Remark 3.3 (More general channel)**.** It is possible to consider the action symbol as another input to the memoryless channel $P_{Y|X,S_e,S_d}$. The capacity expression for this more general channel $P_{Y|X,S_e,S_d,A}$ remains unchanged. This can be shown by defining the new state $S'_e \triangleq (S_e, A)$ and then applying the result of Theorem 3.3.1.

Remark 3.4 (Other results)**.** Theorem 3.3.1 relates to other known results in the literature:

(i) When the reversible input constraint is omitted, the problem reduces to an extension of Weissman's channel with action-dependent states [Wei10]. The

capacity of the channel in this case can be derived along the lines of [Wei10] where the achievability proof are modified such that the state $S^n = (S_e^n, S_d^n)$, and (Y^n, S_d^n) are considered as the new channel output, and a set of distributions is restricted to satisfy the Markov relations $U - (A, S_e) - S_d$ and $X - (U, S_e) - (A, S_d)$. The capacity is given by

$$C_M = \max[I(A, U; Y, S_d) - I(U; S_e|A)], \tag{3.12}$$

where the maximization is over $P_A, P_{U|A,S_e}$ and $\tilde{f} : \mathcal{U} \times \mathcal{S}_e \to \mathcal{X}$, and U is the auxiliary random variable with $|\mathcal{U}| \leq |\mathcal{A}||\mathcal{S}_e||\mathcal{X}| + 1$. Unlike C_M in (3.12), the optimization domain in Theorem 3.3.1 is restricted to a smaller set where the channel input symbol does not depend on the encoder's state information, and additionally it has to satisfy the condition $I(X; Y, S_d|A) - I(X; S_e|A) \geq 0$. More discussion on this condition is given in Section 3.3.3.

(ii) When the state information at the decoder is absent and the channel state is given by nature, i.e., the action alphabet size is one, we recover a special case of the results on information embedding with reversible stegotext [SS09] when there is no distortion constraint between X^n and S_e^n.

Let us consider for a moment a new and slightly different communication problem where the decoder is interested in decoding both the message M and the state S_e^n instead. Due to a deterministic encoding function, the channel input signal can be retrieved based on the decoded message and the encoder's state information. This communication problem has therefore a seemingly more demanding reconstruction constraint than our main problem considered in Fig. 3.4 since it essentially requires that the decoder can decode the message, the encoder's state, and the channel input signal, all reliably. Proposition 3.3.1 below states the capacity of this new channel.

Proposition 3.3.1 (Reconstruct S_e^n)**.** *The capacity of a new channel considered above is given by*

$$C_{S_e} = \max[I(A, S_e, X; Y, S_d) - H(S_e|A)], \tag{3.13}$$

where the joint distribution of (A, S_e, S_d, X, Y) *is of the form*

$$P_A(a)P_{S_e,S_d|A}(s_e, s_d|a)P_{X|A,S_e}(x|a, s_e)P_{Y|X,S_e,S_d}(y|x, s_e, s_d)$$

and the maximization is over all P_A *and* $P_{X|A,S_e}$ *such that*

$$0 \leq I(S_e, X; Y, S_d|A) - H(S_e|A). \tag{3.14}$$

Proof. The achievable scheme in this case is different from the previous case of decoding M and X^n in that the state codebook is introduced and it has to "cover" all possible generated S_e^n *losslessly*. That is, the size of the state codebook should be sufficiently large so that the encoder is able to find an exact S_e^n from the codebook. Similarly to Theorem 3.3.1, in the capacity expression, we also have a similar

restricting condition $0 \leq I(S_e, X; Y, S_d|A) - H(S_e|A)$ on the set of input distributions. Besides the rate constraint, this condition can be considered as a necessary and sufficient condition for the process of losslessly compressing S_e^n through X^n and then transmitting them reliably over the channel in our two-stage communication problem. The detailed achievability proof and the converse proof are given in Appendix 3.C. □

Remark 3.5 (Reconstruct S_e^n is more demanding)**.** Since the channel input sequence can be retrieved based on the decoded message, the encoder's state information, and a known deterministic encoding function, it is natural to compare the capacity C in Theorem 3.3.1 with C_{S_e} in Proposition 3.3.1. As expected, for a given channel $P_{S_e,S_d|A}, P_{Y|X,S_e,S_d}$, we have that $C \geq C_{S_e}$.

Proof. From standard properties of entropy function, we have that $I(S_e, X; Y, S_d|A) - H(S_e|A) \leq I(X; Y, S_d|A) - I(X; S_e|A)$ for all joint distributions factorized in the form of $P_A P_{S_e,S_d|A} P_{X|A,S_e} P_{Y|X,S_e,S_d}$. This implies that C_{S_e} is evaluated over a smaller set than that of C. In addition, it follows in a similar fashion that the terms $I(A, S_e, X; Y, S_d) - H(S_e|A) \leq I(A, X; Y, S_d) - I(X; S_e|A)$. This therefore concludes that $C \geq C_{S_e}$. □

This new communication problem is closely related to the problems of state amplification [KSC08], and reversible information embedding [WK03]. The main difference is that, in our setting, channel states are generated based on the action sequence. In [KSC08] the decoder is interested in decoding the message reliably and in decoding the encoder's state information within a list, while in [WK03], the decoder is interested in decoding both the message and the encoder's state information reliably. The result in Remark 3.5 is also analogous to that in information embedding with reversible stegotext [SS09] in which the authors showed that if the objective is to decode only M and X^n, then decoding M and S_e^n first and re-encoding X^n using a deterministic encoding function is suboptimal.

3.3.3 Two-stage Coding Condition

We term the condition $I(X; Y, S_d|A) - I(X; S_e|A) \geq 0$ which appears in Theorem 3.3.1 the *two-stage coding condition* since it represents the underlining sufficient condition for successful two-stage coding in our achievable scheme. From the result in Theorem 3.3.1, the condition is not only a sufficient condition obtained from the successive decoding, but also a necessary condition for the additional reconstruction requirement. It is also useful for understanding the dual relation between the source and channel coding problems to be discussed in the next section. We note that a similar condition also appears as an extra constraint resulted from the additional reconstruction requirements in the two-stage communication setting later observed in [CM12a], [ZPS12].

As the two-stage coding condition plays a role in restricting the set of input distributions in Theorem 3.3.1, it is natural to wonder whether the condition can

really be active in the optimal design. In the following subsection, we show by an example that there exists a case where the condition is active. Also, we show in the following proposition that if the condition is active, it is satisfied with equality, i.e., the capacity is obtained with $I(X;Y,S_d|A) - I(X;S_e|A) = 0$.

Proposition 3.3.2. *If the two-stage coding condition is ignored, and the solution to the unconstrained optimization problem in* (3.10) *results in* $I(X;Y,S_d|A) - I(X;S_e|A) < 0$ *(the two-stage coding condition would be active), then the actual channel capacity will be obtained with* $I(X;Y,S_d|A) - I(X;S_e|A) = 0$.

Proof. Let us define $f(p) \triangleq I(A,X;Y,S_d) - I(X;S_e|A)$, and $g(p) \triangleq I(X;Y,S_d|A) - I(X;S_e|A)$, where $p \in \mathcal{P}$, and $\mathcal{P}$ is the set of all input probability distributions of form $P_A(a)P_{X|A,S_e}(x|a,s_e)$. It can be shown that both $f(p)$ and $g(p)$ are concave functions in p. Therefore, the optimization problem in Theorem 3.3.1 is convex. Then, we consider the perturbation function

$$\begin{aligned} C(y) = &\max_{p\in\mathcal{P}} f(p) \\ &\text{s.t. } g(p) \geq y. \end{aligned} \tag{3.15}$$

Due to convexity of the original problem, we have that $C(y)$ is a concave function of y [BV04, Chapter 5]. Thus, if the two-stage coding condition is active, i.e., there exists $y^- < 0$ s.t. $C(y^-) > C(y)$ for all $y \geq 0$, the capacity $C = C(0)$ is obtained with $p = p^*$ s.t. $g(p^*) = 0$. □

3.3.4 Examples

In the following, we show two examples to illustrate the role of the two-stage coding condition in restricting a set of capacity achieving input distributions. Example 1 shows that the two-stage coding condition can be active when computing the capacity, while Example 2 shows that there also exists a case where such a condition is not active in the optimal design.

Example 1: Memory Cell With a Rewrite Option

For simplicity, we consider a special case of Theorem 3.3.1 where S_d^n is absent and the channel is in the more general form $P_{Y|X,S_e,A}$ as in Remark 3.3. We consider a binary example where $A, X, S_e, Y \in \{0,1\}$, and the scenario of writing on a memory cell with a rewrite option. The first writing is done through a BSC(δ) with input A, and output S_e. Then, assuming that there is a perfect feedback of the output S_e to the second encoder. The second encoder has an option to rewrite or not to rewrite on the memory (indicated by a value of X). If the rewrite value $X = 1$ which corresponds to "rewrite" is selected, then Y is given as the output of BSC(δ) with input A (rewrite using the old input). If $X = 0$ which corresponds to "no rewrite," we simply get $Y = S_e$. In this case, the decoder is interested in decoding both the embedded message and the rewrite signal. See Fig. 3.5 for an illustration of this rewrite channel.

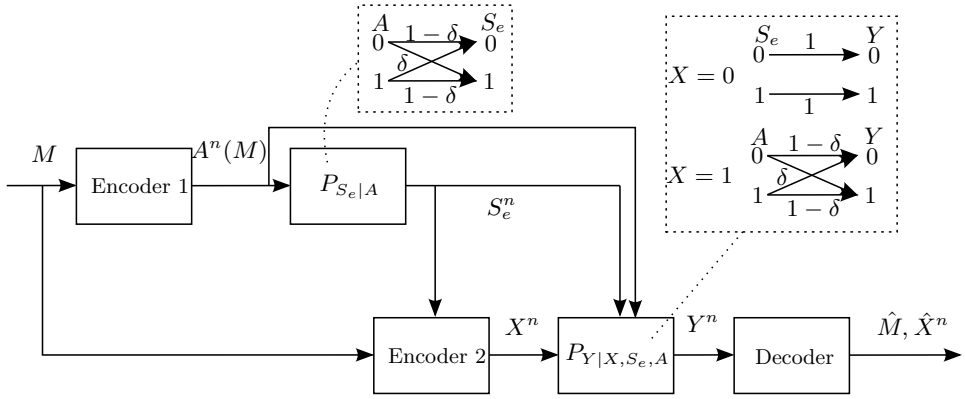

Figure 3.5: Two-stage writing on a memory cell with a rewrite option.

From Theorem 3.3.1 and Remark 3.3, we know that the capacity of this channel is given by

$$C = \max[I(A, X; Y) - I(X; S_e|A)], \tag{3.16}$$

where the maximization is over all P_A and $P_{X|A,S_e}$ such that $0 \leq I(X; Y|A) - I(X; S_e|A)$. Letting $A \sim \text{Bernoulli}(p_a)$, and

$$\begin{aligned} p(x = 0|s_e = 0, a = 0) &= p \\ p(x = 0|s_e = 0, a = 1) &= q \\ p(x = 0|s_e = 1, a = 0) &= r \\ p(x = 0|s_e = 1, a = 1) &= s. \end{aligned}$$

By straightforward manipulation and performing numerical optimization with $\delta = 0.1$, we obtain that the capacity of the channel equals to 0.5310 bits per channel use. The optimal input distributions in this case are those in which $X - A - S_e$ forms a Markov chain, i.e., $p = r, q = s$, and in the end, p_a is the only remaining optimization variable. This corresponds to the capacity of the effective channel $P_{Y|A}$. If we instead neglect the two-stage coding condition and solve the unconstrained optimization problem, we would obtain the maximum value of 0.6690 which is strictly larger than the true capacity. Therefore, this example shows that there exists a case where the two-stage coding condition is active. In fact, the corresponding two-stage coding condition in this case is satisfied with equality as expected from Proposition 3.3.2.

Example 2: Inactive Two-stage Coding Condition

In other cases, the two-stage coding condition in the capacity expression might not be active. One trivial example is when $S_e - A - S_d$ forms a Markov chain for the action-dependent state channel $P_{S_e,S_d|A}$, and $Y - (X, S_d) - S_e$ forms a Markov chain for the state-dependent channel $P_{Y|X,S_e,S_d}$. In this case, it can be shown that for any

joint distribution $\left(P_A^{(1)}(a), P_{X|A,S_e}^{(1)}(x|a,s_e)\right)$ such that $I(X;Y,S_d|A)-I(X;S_e|A) < 0$, there always exists another joint distribution $\left(P_A^{(2)}(a), P_{X|A,S_e}^{(2)}(x|a,s_e)\right)$ which satisfies $I(X;Y,S_d|A) - I(X;S_e|A) \geq 0$ and achieves a higher rate. One possible choice is to let $P_A^{(2)} = P_A^{(1)}$ and $P_{X|A,S_e}^{(2)} = \sum_{s_e} P_{S_e|A} P_{X|A,S_e}^{(1)}$, which corresponds to the distribution that satisfies the Markov chain $X - A - S_e$, and thus $I(X;Y,S_d|A) - I(X;S_e|A) = I(X;Y,S_d|A) \geq 0$. Since $P_{A,Y,S_d}^{(1)} = P_{A,Y,S_d}^{(2)}$, we have that this new distribution achieves a higher rate $I(A,X;Y,S_d)-I(X;S_e|A)$. Consequently, the maximizing input distribution in this case will result in the condition $I(X;Y,S_d|A) - I(X;S_e|A) \geq 0$ and the capacity of such a channel is given by $C = \max_{P_A, P_{X|A,S_e}} [I(A,X;Y,S_d) - I(X;S_e|A)]$.

3.4 Duality

In this section, we discuss the potential duality between the source coding (SC) and channel coding (CC) problems considered in Section 3.2 and 3.3 where we illustrate the "dual" relations between input-output of elements in the source and channel coding systems in Fig. 3.6. Similar dual relations also appear in other related problems, as listed in Table 3.1, where (♯) denotes our settings studied in Sections 3.2 and 3.3.

We are interested in investigating *formula duality* of a set of problems [CC02]. Table 3.2 summarizes the rate-distortion(-cost) function and the channel capacity expressions of the interested problems, neglecting the optimization variables (input distributions). As in [CC02], the formula duality of the rate-distortion(-cost) function and the channel capacity can be recognized by the following correspondence,

$$\begin{aligned}
\underline{\text{Rate-distortion-cost}} &\leftrightarrow \underline{\text{Channel capacity}} \\
\text{minimization} &\leftrightarrow \text{maximization} \\
X(\text{source symbol}) &\leftrightarrow Y(\text{received symbol}) \\
\hat{X}(\text{decoded symbol}) &\leftrightarrow X(\text{transmitted symbol}) \\
S_e(\text{state at the encoder}) &\leftrightarrow S_d(\text{state at the decoder}) \\
S_d(\text{state at the decoder}) &\leftrightarrow S_e(\text{state at the encoder}) \\
U(\text{auxiliary}) &\leftrightarrow U(\text{auxiliary}) \\
A(\text{action}) &\leftrightarrow A(\text{action}).
\end{aligned}$$

We see that the first three dualities in Table 3.2 are obvious from the expressions of the rate-distortion(-cost) function and the channel capacity, while the last duality (Sec. 3.2 and 3.3) does not hold in general due to the fundamental differences between the two problems. We now give some justification based on the dual roles of the encoder/decoder in the source coding problem and the decoder/encoder in the channel coding problem.

- The first reason that the last duality in Table 3.1 does not hold in general is the presence of the two-sided side/state information. That is, the processing is sequential at the *encoder* in the source coding setup, i.e., the action-dependent side information is generated first and then the side information S_e^n is used in compressing the source sequence. However, this sequential processing is not required in the *decoding* process of the channel coding problem since the state information for both encoder and decoder are generated in the beginning, and both Y^n and S_d^n are available at the decoder noncausally. The effect of this fundamental difference can be seen from the difference in the terms $I(A;X)$ and $I(A;Y,S_d)$ in the rate-distortion-cost function and channel capacity expressions in Table 3.2.

- The second reason is due to the additional reconstruction constraint imposed on the two communication problems. First consider the channel coding problem where we require to decode as well the channel input sequence (reversible input constraint). In our problem, the *encoder* has a causal structure; that is, S_e^n is generated first, then followed by X^n. When we require to decode X^n which is the signal generated in the second stage, the two-stage coding condition, apart from the rate constraint, is necessary to ensure reliable transmission of the channel input X^n. Now we consider the source coding counterpart where we require the encoder to estimate the decoder's reconstruction (common reconstruction constraint). Although there seems to be a similar two-stage structure in the *decoder* of this setup, the two-stage coding condition is not relevant here. This is because the common reconstruction is performed in the beginning at the encoder side and the choice of action sequence is in fact known at both sides due to the noiseless link between the encoder and the decoder.

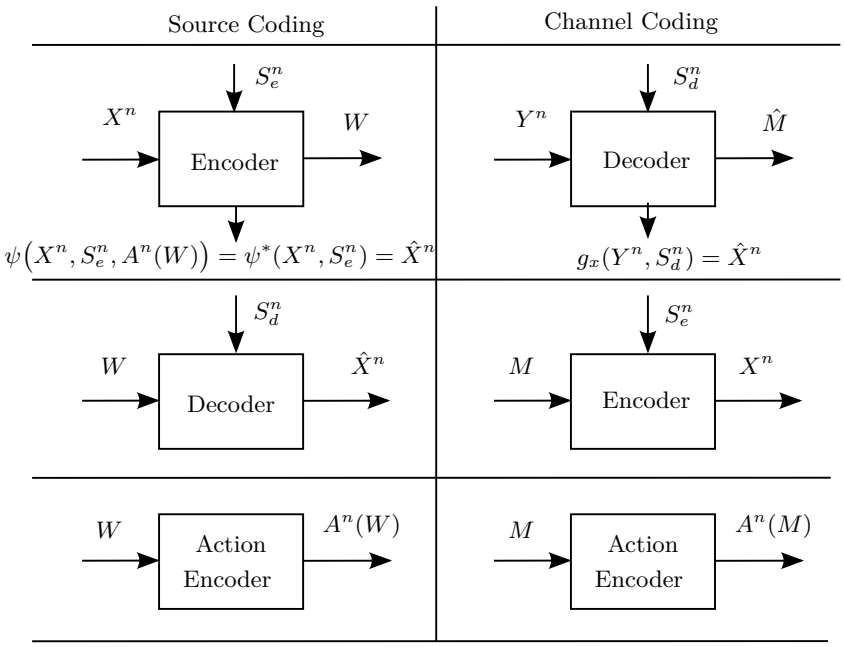

Figure 3.6: Duality between the source coding with action-dependent side information and common reconstruction (Fig. 3.1) and channel coding with action-dependent states and reversible input (Fig. 3.4).

Table 3.1: Dual problems.

Wyner-Ziv's SC (WZ, [WZ76]) $\leftrightarrow$ Gel'fand-Pinsker's CC (GP, [GP80])

Permuter-Weissman's SC with action (PW, [PW11]) $\leftrightarrow$ Weissman's CC with action (W, [Wei10])

Steinberg's SC with CR (S, [Ste09]) $\leftrightarrow$ Sumszyk-Steinberg's CC with RI (SS, [SS09])

Section 3.2 $\overset{(\sharp)}{\leftrightarrow}$ Section 3.3

Table 3.2: Rate-distortion(-cost) function and channel capacity.

Problems	Rate-distortion(-cost) function	Channel capacity
WZ and GP	$R_{\mathrm{WZ}}(D) = \min[I(U;X) - I(U,S_d)]$	$C_{\mathrm{GP}} = \max[I(U;Y) - I(U;S_e)]$
PW and W	$R_{\mathrm{PW}}(D,C) = \min[I(A;X) + I(U;X\|A) - I(U,S_d\|A)]$	$C_{\mathrm{W}} = \max[I(A;Y) + I(U;Y\|A) - I(U;S_e\|A)]$
S and SS	$R_{\mathrm{S}}(D) = \min[I(\hat{X};X) - I(\hat{X};S_d)]$	$C_{\mathrm{SS}} = \max[I(X;Y) - I(X;S_e)]$
Sec. 3.2 and 3.3	$R(D,C) = \min[I(A;X) + I(\hat{X};X,S_e\|A) - I(\hat{X};S_d\|A)]$ s.t. $I(\hat{X};X,S_e\|A) - I(\hat{X};S_d\|A) \geq 0$ [1]	$C = \max[I(A;Y,S_d) + I(X;Y,S_d\|A) - I(X;S_e\|A)]$ s.t. $I(X;Y,S_d\|A) - I(X;S_e\|A) \geq 0$

[1] This condition is only included for the discussion of the formula duality. It is straightforward to show that the condition is not active (always satisfied) under the considered set of input distributions.

3.5 Conclusion

In this chapter, we studied a class of problems that extend Wyner-Ziv source coding and Gel'fand-Pinsker channel coding with action-dependent side information. The extension involves having two-sided action-dependent partial side information, and also additional reconstruction requirements. In the source coding problem, we solved the rate-distortion-cost function for the memoryless source with two-sided action-dependent side information and common reconstruction, while in the channel coding problem, the capacity of the discrete memoryless channel with two-sided action-dependent state and reversible input is derived. The two-stage coding condition in the capacity expression arises from the additional reconstruction constraint and the causal structure of the coding, i.e., the channel input signal to be reconstructed is generated in the second stage transmission. Besides the message rate constraint, it can be considered as a necessary and sufficient condition for reliable transmission of channel input signal over the channel given that the action is communicated.

In this chapter, we also investigated the formula duality between rate-distortion-cost function and channel capacity of the source and channel coding problems. Although the problems considered in Sections 3.2 and 3.3 seem to retain the dual structure as seen in the Wyner-Ziv and Gel'fand-Pinsker problems, they are not dual in general. In fact, there is "operational mismatch" caused by causal processing in the system. It is interesting that the two-stage coding condition which appears in the capacity expression can be active when computing the capacity, as shown in one example.

Appendix for Chapter 3

3.A Proof of Theorem 3.2.1

3.A.1 Achievability Proof of Theorem 3.2.1

The proof follows from a standard random coding argument where we use the definitions and properties of ϵ-typicality as in Definition 2.7.

Codebook Generation: Fix $P_{A|X}, P_{\hat{X}|X,S_e,A}$. Let $\mathcal{W}_1^{(n)} = \{1, 2, \ldots, |\mathcal{W}_1^{(n)}|\}$, $\mathcal{W}_2^{(n)} = \{1, 2, \ldots, |\mathcal{W}_2^{(n)}|\}$, and $\mathcal{V}^{(n)} = \{1, 2, \ldots, |\mathcal{V}^{(n)}|\}$. For all $w_1 \in \mathcal{W}_1^{(n)}$, the action codewords $a^n(w_1)$ are generated i.i.d. each according to $\prod_{i=1}^n P_A(a_i)$, and for each $w_1 \in \mathcal{W}_1^{(n)}$, $|\mathcal{W}_2^{(n)}||\mathcal{V}^{(n)}|$ codewords $\{\hat{x}^n(w_1, w_2, v)\}_{w_2 \in \mathcal{W}_2^{(n)}, v \in \mathcal{V}^{(n)}}$ are generated i.i.d. each according to $\prod_{i=1}^n P_{\hat{X}|A}\big(\hat{x}_i | a_i(w_1)\big)$. The codebooks are then revealed to the encoder, the action decoder, and the decoder. Let $0 < \epsilon_0 < \epsilon_1 < \epsilon_2 < \epsilon_3 < \epsilon < 1$.

Encoding: Given a source realization x^n, the encoder first looks for the smallest $w_1 \in \mathcal{W}_1^{(n)}$ such that $a^n(w_1)$ is jointly typical with x^n. Then the side information are generated as outputs of the memoryless channel with transition probability $P_{S_e^n, S_d^n | X^n, A^n}(s_e^n, s_d^n | x^n, a^n) = \prod_{i=1}^n P_{S_e, S_d | X, A}(s_{e,i}, s_{d,i} | x_i, a_i)$, and the encoder in the second stage looks for the smallest $w_2 \in \mathcal{W}_2^{(n)}$ and $v \in \mathcal{V}^{(n)}$ such that $\big(x^n, \hat{x}^n(w_1, w_2, v),\ \ s_e^n, a^n(w_1)\big) \in T_{\epsilon_3}^{(n)}(X, \hat{X}, S_e, A)$. If successful, the encoder produces $\hat{x}^n(w_1, w_2, v)$ as a common reconstruction at the encoder and transmits indices (w_1, w_2) to the decoder. If not successful, the encoder transmits $w_1 = 1, w_2 = 1$ and produces $\hat{x}^n(1, 1, 1)$.

Decoding: Given the indices w_1 and w_2, and the side information s_d^n the decoder reconstructs $\tilde{x}^n = \hat{x}^n(w_1, w_2, \tilde{v})$ if there exists a unique $\tilde{v} \in \mathcal{V}^{(n)}$ such that $\big(s_d^n, \hat{x}^n(w_1, w_2, \tilde{v}), a^n(w_1)\big) \in T_{\epsilon}^{(n)}(S_d, \hat{X}, A)$. Otherwise, the decoder puts out $\tilde{x}^n = \hat{x}^n(w_1, w_2, 1)$.

Analysis of Probability of Error: Let (W_1, W_2, V) denote the corresponding indices of the chosen codewords A^n and $\hat{X}^n$ at the encoder. We define the "error" events as follows.

$$
\begin{aligned}
\mathcal{E}_0 &= \big\{X^n \notin T_{\epsilon_0}^{(n)}(X)\big\} \\
\mathcal{E}_{1a} &= \big\{(X^n, A^n(w_1)) \notin T_{\epsilon_1}^{(n)}(X, A) \text{ for all } w_1 \in \mathcal{W}_1^{(n)}\big\} \\
\mathcal{E}_{1b} &= \big\{(X^n, A^n(W_1), S_e^n, S_d^n) \notin T_{\epsilon_2}^{(n)}(X, A, S_e, S_d)\big\} \\
\mathcal{E}_2 &= \big\{\big(X^n, \hat{X}^n(W_1, w_2, v), S_e^n, A^n(W_1)\big) \notin T_{\epsilon_3}^{(n)}(X, \hat{X}, S_e, A) \\
&\qquad \text{for all } (w_2, v) \in \mathcal{W}_2^{(n)} \times \mathcal{V}^{(n)}\big\} \\
\mathcal{E}_3 &= \big\{\big(S_d^n, \hat{X}^n(W_1, W_2, V), A^n(W_1)\big) \notin T_{\epsilon}^{(n)}(S_d, \hat{X}, A)\big\} \\
\mathcal{E}_4 &= \big\{\big(S_d^n, \hat{X}^n(W_1, W_2, \tilde{v}), A^n(W_1)\big) \in T_{\epsilon}^{(n)}(S_d, \hat{X}, A) \text{ for some } \tilde{v} \in \mathcal{V}^{(n)}, \tilde{v} \neq V\big\}.
\end{aligned}
$$

The total "error" probability is bounded by

$$\Pr(\mathcal{E}) \leq \Pr(\mathcal{E}_0) + \Pr(\mathcal{E}_{1a} \cap \mathcal{E}_0^c) + \Pr(\mathcal{E}_{1b} \cap \mathcal{E}_{1a}^c) + \Pr(\mathcal{E}_2 \cap \mathcal{E}_{1b}^c) + \Pr(\mathcal{E}_3 \cap \mathcal{E}_2^c) + \Pr(\mathcal{E}_4),$$

where $\mathcal{E}_i^c$ denotes the complement of the event $\mathcal{E}_i$.

0) By the law of large numbers (LLN), $\Pr\big(X^n \in T_{\epsilon_0}^{(n)}(X)\big) \geq 1 - \delta_{\epsilon_0}$. Since δ_{ϵ_0} can be made arbitrarily small with increasing n if $\epsilon_0 > 0$, we have $\Pr(\mathcal{E}_0) \to 0$ as $n \to \infty$.

1a) By the covering lemma (Lemma 2.6), $\Pr(\mathcal{E}_{1a} \cap \mathcal{E}_0^c) \to 0$ as $n \to \infty$ if $\frac{1}{n} \log |\mathcal{W}_1^{(n)}| > I(X; A) + \delta_{\epsilon_1}$.

1b) By the conditional typicality lemma [EK11] where (S_d^n, S_e^n) is i.i.d. according to $\prod_{i=1}^n P_{S_d, S_e|X,A}\big(s_{d,i}, s_{e,i}|x_i, a_i(w_1)\big)$, we have $\Pr(\mathcal{E}_{1b} \cap \mathcal{E}_{1a}^c) \to 0$ as $n \to \infty$.

2) Averaging over all $W_1 = w_1$, by the covering lemma, where each $\hat{X}^n$ is drawn independently according to $\prod_{i=1}^n P_{\hat{X}|A}\big(\hat{x}_i|a_i(w_1)\big)$, we have that $\Pr(\mathcal{E}_2 \cap \mathcal{E}_{1b}^c) \to 0$ as $n \to \infty$ if $\frac{1}{n} \log |\mathcal{W}_2^{(n)}| + \frac{1}{n} \log |\mathcal{V}^{(n)}| > I(X, S_e; \hat{X}|A) + \delta_{\epsilon_3}$.

3) Consider the event $\mathcal{E}_2^c$ in which there exists (W_1, W_2, V) such that the tuple $\big(X^n, \hat{X}^n(W_1, W_2, V), A^n(W_1), S_e^n\big) \in T_{\epsilon_3}^{(n)}(X, \hat{X}, A, S_e)$. Since we have the Markov chain $\hat{X} - (X, S_e, A) - S_d$, and S_d^n is distributed according to the distribution $\prod_{i=1}^n P_{S_d|X,S_e,A}\big(s_{d,i}|x_i, s_{e,i}, a_i(w_1)\big)$, by using the conditional typicality lemma, we have $\Pr\big((X^n, \hat{X}^n(W_1, W_2, V), \quad A^n(W_1), S_e^n, S_d^n) \in T_\epsilon^{(n)}(X, \hat{X}, A, S_e, S_d)\big) \to 1$ as $n \to \infty$. This implies that $\Pr(\mathcal{E}_3 \cap \mathcal{E}_2^c) \to 0$ as $n \to \infty$.

4) Averaging over all $W_1 = w_1, W_2 = w_2$, and $V = v$ [EK11, Chapter 11, Lemma 11.1], by the packing lemma (Lemma 2.7) where each $\hat{X}^n$ is drawn independently according to $\prod_{i=1}^n P_{\hat{X}|A}\big(\hat{x}_i|a_i\big)$, we have that $\Pr(\mathcal{E}_4) \to 0$ as $n \to \infty$ if $\frac{1}{n} \log |\mathcal{V}^{(n)}| < I(\hat{X}; S_d|A) - \delta_\epsilon$.

Finally, we consider the case where there is no error, i.e.,

$$(X^n, \hat{X}^n(W_1, W_2, V), A^n(W_1), S_e^n, S_d^n) \in T_\epsilon^{(n)}(X, \hat{X}, A, S_e, S_d).$$

By the law of total expectation, the averaged distortion (over all codebooks $\mathcal{C}_n$ containing codewords $(\hat{X}^n, A^n)$) is given by

$$E_{\mathcal{C}_n, X^n}[d^{(n)}(X^n, \tilde{X}^n)] = \Pr(\mathcal{E}) \cdot E_{\mathcal{C}_n, X^n}[d^{(n)}(X^n, \hat{X}^n)|\mathcal{E}] + \Pr(\mathcal{E}^c) \cdot E_{\mathcal{C}_n, X^n}[d^{(n)}(X^n, \hat{X}^n)|\mathcal{E}^c] \leq \Pr(\mathcal{E}) \cdot d_{max} + \Pr(\mathcal{E}^c) \cdot E_{\mathcal{C}_n, X^n}[d^{(n)}(X^n, \hat{X}^n)|\mathcal{E}^c],$$

where d_{max} is assumed to be the maximal average distortion incurred by the "error" events.

Given $\mathcal{E}^c$, the distortion is bounded by

$$d^{(n)}(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^n d\big(x_i, \hat{x}_i\big)$$

$$= \frac{1}{n}\sum_{a,b} N(a,b|x^n,\hat{x}^n)d\big(a,b\big)$$
$$\overset{(*)}{\leq} \sum_{a,b} P_{X,\hat{X}}(a,b)\,(1+\epsilon)\,d\big(a,b\big) = E\big[d\big(X,\hat{X}\big)\big]\,(1+\epsilon)\,,$$

where $(*)$ follows from the definition of jointly typical set (Definition 2.7).

Therefore, we have

$$E_{\mathcal{C}_n,X^n}[d^{(n)}(X^n,\tilde{X}^n)] \leq \Pr(\mathcal{E})\cdot d_{max} + \Pr(\mathcal{E}^c)\cdot E\big[d\big(X,\hat{X}\big)\big](1+\epsilon).$$

Similarly, we have for the average cost

$$E_{\mathcal{C}_n}[\Lambda^{(n)}(A^n)] \leq \Pr(\mathcal{E})\cdot c_{max} + \Pr(\mathcal{E}^c)\cdot E\big[\Lambda(A)](1+\epsilon),$$

where c_{max} is assumed to be the maximal average cost incurred by the "error" events.

By combining the bounds on the code rates that make $\Pr(\mathcal{E})\to 0$ as $n\to\infty$ and considering the constraint $\frac{1}{n}\log|\mathcal{W}^{(n)}| = \frac{1}{n}\log|\mathcal{W}_1^{(n)}||\mathcal{W}_2^{(n)}| \leq R+\delta$, for any $\delta>0$, we have

$$R+\delta \geq \frac{1}{n}\log|\mathcal{W}^{(n)}| > I(X;A) + I(X,S_e;\hat{X}|A) - I(\hat{X};S_d|A) + \delta'_\epsilon,$$

where δ'_ϵ can be made arbitrarily small, i.e., $\delta'_\epsilon\to 0$ as $\epsilon\to 0$[2].

Thus, for any $\delta>0$, if $R \geq I(X;A)+I(X,S_e;\hat{X}|A)-I(\hat{X};S_d|A)$, $E\big[d\big(X,\hat{X}\big)\big] \leq D$ and $E\big[\Lambda(A)\big]\leq C$, then we have $\Pr(\mathcal{E})\to 0$ as $n\to\infty$, and for all sufficiently large n,

$$E_{\mathcal{C}_n,X^n}[d^{(n)}(X^n,\tilde{X}^n)] \leq D+\delta_\epsilon \leq D+\delta,$$
$$E_{\mathcal{C}_n}[\Lambda^{(n)}(A^n)] \leq C+\delta_\epsilon \leq C+\delta.$$

Lastly, with $\Pr(\mathcal{E})\to 0$ as $n\to\infty$, it follows that with high probability the decoded codeword $\tilde{X}^n = \hat{X}^n(W_1,W_2,\tilde{v})$ at the decoder is the correct one which was chosen at the encoder. We recall the encoding process which determines the codeword $\hat{X}^n$ based on X^n, S_e^n and A^n, and puts out $\hat{X}^n$ as a common reconstruction at the encoder. That is, there exists a mapping $\psi^{(n)}(\cdot)$ such that $\hat{X}^n = \psi^{(n)}(X^n,S_e^n,A^n)$. Thus, for any $\delta>0$, we can have $\Pr\big(\psi^{(n)}(X^n,S_e^n,A^n)\neq \tilde{X}^n\big)\leq\delta$ for all sufficiently large n.

The average distortion, cost, and common reconstruction error probability (over all codebooks) are upper-bounded by $D+\delta, C+\delta$ and δ, respectively. Therefore, from the selection lemma [BB11, Lemma 2.2], there must exist at least one code such that, for sufficiently large n, the average distortion, cost, and common reconstruction error probability are upper-bounded by $D+\delta, C+\delta$ and δ.

Thus, any (R,D,C) such that $R \geq I(X;A)+I(X,S_e;\hat{X}|A)-I(\hat{X};S_d|A)$, $E\big[d\big(X,\hat{X}\big)\big]\leq D$, and $E[\Lambda(A)]\leq C$ for some $P_X(x)P_{A|X}(a|x)P_{S_e,S_d|X,A}(s_e,s_d|x,a)$ $P_{\hat{X}|X,S_e,A}(\hat{x}|x,s_e,a)$ is achievable. This concludes the achievability proof.

[2]We often combine the *small* terms into one term denoted by δ'_ϵ where $\delta'_\epsilon\to 0$ as $\epsilon\to 0$.

3.A.2 Converse Proof of Theorem 3.2.1

Let us assume the existence of a specific sequence of $(|\mathcal{W}^{(n)}|, n)$ codes such that for $\delta_n > 0$, $\frac{1}{n}\log|\mathcal{W}_1^{(n)}||\mathcal{W}_2^{(n)}| \leq R+\delta_n$, $\frac{1}{n}E\left[\sum_{i=1}^n d(X_i, g_i)\right] \leq D+\delta_n$, $\frac{1}{n}E\left[\sum_{i=1}^n \Lambda(A_i)\right] \leq C + \delta_n$, and $\Pr\big(\psi^{(n)}(X^n, S_e^n, A^n) \neq g^{(n)}(W_1, W_2, S_d^n)\big) \leq \delta_n$, where g_i denotes the i^{th} symbol of $g^{(n)}(W_1, W_2, S_d^n)$ and $\lim_{n\to\infty}\delta_n = 0$. Then we will show that $R \geq R_{\mathrm{ac,cr}}(D, C)$, where $R_{\mathrm{ac,cr}}(D, C)$ is the rate-distortion-cost function defined as

$$R_{\mathrm{ac,cr}}(D, C) = \min_{P_{A|X}, P_{\hat{X}|X,S_e,A}} [I(X; A) + I(\hat{X}; X, S_e|A, S_d)]. \tag{3.17}$$

With $\Pr\big(\psi(X^n, S_e^n, A^n) \neq g^{(n)}(W_1, W_2, S_d^n)\big) = \delta_n' \leq \delta_n$, and $|\hat{\mathcal{X}}| = |\tilde{\mathcal{X}}|$, the Fano inequality can be applied to bound

$$H\big(\psi^{(n)}(X^n, S_e^n, A^n)\big|g^{(n)}(W_1, W_2, S_d^n)\big) \leq h(\delta_n') + \delta_n' \log(|\hat{\mathcal{X}}|^n - 1) \triangleq n\epsilon_n, \tag{3.18}$$

where $h(\delta_n')$ is the binary entropy function, and $\epsilon_n \to 0$ as $\delta_n' \to 0$.

Then the standard properties of the entropy function give

$$\begin{aligned}
&n(R+\delta_n) \geq \log\left(|\mathcal{W}_1^{(n)}| \cdot |\mathcal{W}_2^{(n)}|\right) \geq H(W_1, W_2)\\
&\overset{(*)}{=} H(W_1, W_2, A^n) = H(A^n) + H(W_1, W_2|A^n)\\
&\geq [H(A^n) - H(A^n|X^n, S_e^n)] + [H(W_1, W_2|A^n, S_d^n) - H(W_1, W_2|A^n, X^n, S_e^n, S_d^n)]\\
&= \underbrace{H(X^n, S_e^n) - H(X^n, S_e^n|A^n) + H(X^n, S_e^n|A^n, S_d^n)}_{=P} \underbrace{-H(X^n, S_e^n|A^n, S_d^n, W_1, W_2)}_{=Q},
\end{aligned} \tag{3.19}$$

where in $(*)$ we used the fact that $A^n = f_a^{(n)}(W_1)$, and $f_a^{(n)}(\cdot)$ is a deterministic function. Further,

$$\begin{aligned}
P &= H(X^n, S_e^n) - H(X^n, S_e^n|A^n) + H(X^n, S_e^n|A^n, S_d^n)\\
&= H(X^n, S_e^n) + H(S_d^n|X^n, S_e^n, A^n) - H(S_d^n|A^n)\\
&= H(X^n) + H(S_e^n|X^n) + H(S_e^n, S_d^n|X^n, A^n) - H(S_e^n|X^n, A^n) - H(S_d^n|A^n)\\
&\overset{(\star)}{\geq} \sum_{i=1}^n H(X_i) + H(S_{e,i}, S_{d,i}|X_i, A_i) - H(S_{d,i}|A_i)\\
&= \sum_{i=1}^n H(X_i) + H(S_{e,i}|X_i, A_i) + H(S_{d,i}|X_i, S_{e,i}, A_i) - H(S_{d,i}|A_i)\\
&= \sum_{i=1}^n H(X_i) + H(S_{e,i}|X_i, A_i) - H(X_i, S_{e,i}|A_i) + H(X_i, S_{e,i}|A_i, S_{d,i})\\
&= \sum_{i=1}^n I(X_i; A_i) + H(X_i, S_{e,i}|A_i, S_{d,i}),
\end{aligned} \tag{3.20}$$

where $(\star)$ holds due to the i.i.d. property of P_{X^n} and $P_{S_e^n,S_d^n|X^n,A^n}$,

$$\begin{aligned}
Q &= -H\big(X^n, S_e^n|A^n, S_d^n, W_1, W_2, g^{(n)}(W_1, W_2, S_d^n)\big)\\
&\geq -H\big(X^n, S_e^n|A^n, S_d^n, g^{(n)}(W_1, W_2, S_d^n)\big)\\
&= -H\big(\psi^{(n)}(X^n, S_e^n, A^n), X^n, S_e^n|A^n, S_d^n, g^{(n)}(W_1, W_2, S_d^n)\big)\\
&\quad + H(\psi^{(n)}(X^n, S_e^n, A^n)|A^n, S_d^n, g^{(n)}(W_1, W_2, S_d^n), X^n, S_e^n)\\
&\geq -H\big(\psi^{(n)}(X^n, S_e^n, A^n)|g^{(n)}(W_1, W_2, S_d^n)\big)\\
&\quad - H\big(X^n, S_e^n|A^n, S_d^n, \psi^{(n)}(X^n, S_e^n, A^n)\big)\\
&\overset{(a)}{\geq} -n\epsilon_n - \sum_{i=1}^{n} H(X_i, S_{e,i}|A^n, S_d^n, \psi^{(n)}(X^n, S_e^n, A^n), X^{i-1}, S_e^{i-1})\\
&\overset{(b)}{\geq} -n\epsilon_n - \sum_{i=1}^{n} H\big(X_i, S_{e,i}|A_i, S_{d,i}, \psi_i^{(n)}(X^n, S_e^n, A^n)\big),
\end{aligned} \tag{3.21}$$

where (a) follows from (3.18) and $\psi_i^{(n)}(X^n, S_e^n, A^n)$ in (b) corresponds to the i^{th} symbol of $\psi^{(n)}(X^n, S_e^n, A^n)$.

Define $\hat{X}_i = \psi_i^{(n)}(X^n, S_e^n, A^n)$. Then the Markov chain $\hat{X}^n - (X^n, S_e^n, A^n) - S_d^n$ holds. Together with the memoryless property, $P_{S_e^n,S_d^n|X^n,A^n}(s_e^n, s_d^n|x^n, a^n) = \prod_{i=1}^n P_{S_e,S_d|X,A}(s_{e,i}, s_{d,i}|x_i, a_i)$, we also have that $(S_d^{i-1}, X^{n\setminus i}, S_e^{n\setminus i}, A^{n\setminus i}, \hat{X}^n) - (X_i, S_{e,i}, A_i) - S_{d,i}$ forms a Markov chain. Combining (3.19)-(3.21), we have

$$\begin{aligned}
R + \delta_n &\geq \frac{1}{n}\log\big(|\mathcal{W}_1^{(n)}| \cdot |\mathcal{W}_2^{(n)}|\big)\\
&\geq \frac{1}{n}\sum_{i=1}^{n} I(X_i; A_i) + I(\hat{X}_i; X_i, S_{e,i}|A_i, S_{d,i}) - \epsilon_n\\
&\overset{(a)}{\geq} \frac{1}{n}\sum_{i=1}^{n} R_{\mathrm{ac,cr}}\Big(E\big[d\big(X_i, \hat{X}_i\big)\big], E[\Lambda(A_i)]\Big) - \epsilon_n\\
&\overset{(b)}{\geq} R_{\mathrm{ac,cr}}\left(\frac{1}{n}\sum_{i=1}^{n} E\big[d\big(X_i, \hat{X}_i\big)\big], \frac{1}{n}\sum_{i=1}^{n} E[\Lambda(A_i)]\right) - \epsilon_n,
\end{aligned} \tag{3.22}$$

where (a) follows from the definition of $R_{\mathrm{ac,cr}}(D, C)$ in (3.17), and the fact that $\hat{X}_i - (X_i, S_{e,i}, A_i) - S_{d,i}$ forms a Markov chain, (b) follows from Jensen's inequality [CT06, Theorem 2.6.2] and convexity of $R_{\mathrm{ac,cr}}(D, C)$ (Lemma 3.1).

Let β be the event that the reconstruction at the encoder is not equal to that at the decoder, i.e., $\beta = \{\psi^{(n)}(X^n, S_e^n, A^n) \neq g^{(n)}(W_1, W_2, S_d^n)\}$. It then follows that

$$\begin{aligned}
E[d(X_i, g_i)] &= E[d(X_i, g_i)|\beta^c] \cdot \Pr(\beta^c) + E[d(X_i, g_i)|\beta] \cdot \Pr(\beta)\\
&\overset{(\star)}{\geq} E[d(X_i, \hat{X}_i)|\beta^c] \cdot \Pr(\beta^c),
\end{aligned} \tag{3.23}$$

where $(\star)$ holds because we have $g_i = \hat{X}_i$ for given β^c. Thus

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} E[d(X_i,\hat{X}_i)] &= \frac{1}{n}\sum_{i=1}^{n}[E[d(X_i,\hat{X}_i)|\beta^c]\cdot\Pr(\beta^c) + E[d(X_i,\hat{X}_i)|\beta]\cdot\Pr(\beta)] \\ &\overset{(a)}{\leq} \frac{1}{n}\sum_{i=1}^{n} E[d(X_i,\hat{X}_i)|\beta^c]\cdot\Pr(\beta^c) + \tilde{d}_{max}\delta_n \\ &\overset{(b)}{\leq} \frac{1}{n}\sum_{i=1}^{n} E[d(X_i,g_i)] + \tilde{d}_{max}\delta_n \\ &\overset{(c)}{\leq} D + (1+\tilde{d}_{max})\delta_n, \end{aligned} \tag{3.24}$$

where (a) follows from the assumption that $\tilde{d}_{max}$ is the maximum average distortion incurred by the error event β and that $\Pr\big(\psi^{(n)}(X^n, S_e^n, A^n) \neq g^{(n)}(W_1, W_2, S_d^n)\big) \leq \delta_n$, (b) follows from (3.23), and (c) follows from the assumption that $\frac{1}{n}E[\sum_{i=1}^{n} d(X_i, g_i)] \leq D + \delta_n$ in the beginning.

Finally, we substitute (3.24) into (3.22). With $\lim_{n\to\infty}\delta_n = 0$, and $\lim_{n\to\infty}\epsilon_n = 0$, we thus get $R \geq R_{\mathrm{ac,cr}}(D,C)$ by using the assumption that $\frac{1}{n}E\left[\sum_{i=1}^{n}\Lambda(A_i)\right] \leq C + \delta_n$ and the non-increasing property of $R_{\mathrm{ac,cr}}(D,C)$. This concludes the proof of converse.

3.B Proof of Theorem 3.3.1

3.B.1 Achievability Proof of Theorem 3.3.1

Similarly to the previous achievability proof in Theorem 3.2.1, the proof follows from a standard random coding argument. We use the technique of rate splitting, i.e., the message M of rate R is split into two messages M_1 and M_2 of rates R_1 and R_2. Two-stage coding is then considered, i.e., the first stage for communicating the identity of the action sequence, and the second stage for communicating the identity of X^n based on the known action sequence.

For given channels with transition probabilities $P_{S_e,S_d|A}$ and $P_{Y|X,S_e,S_d}$ we can assign the joint probability to any random vector (A, X, S_e) by

$$\begin{aligned}P_{A,S_e,S_d,X,Y}(a,s_e,s_d,x,y) = P_A(a)P_{S_e,S_d|A}(s_e,s_d|a)P_{X|A,S_e}(x|a,s_e) \\ P_{Y|X,S_e,S_d}(y|x,s_e,s_d).\end{aligned}$$

Codebook Generation: Fix P_A and $P_{X|A,S_e}$. Let $\mathcal{M}_1^{(n)} = \{1,2,\ldots,|\mathcal{M}_1^{(n)}|\}$, $\mathcal{M}_2^{(n)} = \{1,2,\ldots,|\mathcal{M}_2^{(n)}|\}$ and $\mathcal{J}^{(n)} = \{1,2,\ldots,|\mathcal{J}^{(n)}|\}$. For all $m_1 \in \mathcal{M}_1^{(n)}$, randomly generate $a^n(m_1)$ i.i.d. according to $\prod_{i=1}^{n} P_A(a_i)$. For each $m_1 \in \mathcal{M}_1^{(n)}$, generate $|\mathcal{M}_2^{(n)}|\cdot|\mathcal{J}^{(n)}|$ codewords, $\{x^n(m_1,m_2,j)\}_{m_2\in\mathcal{M}_2^{(n)}, j\in\mathcal{J}^{(n)}}$ i.i.d. each according to $\prod_{i=1}^{n} P_{X|A}(x_i|a_i(m_1))$. Then the codebooks are revealed to the action encoder, the channel encoder and the decoder. Let $0 < \epsilon_0 < \epsilon_1 < \epsilon_2 < \epsilon < 1$.

Encoding: Given the message $m = (m_1, m_2) \in \mathcal{M}^{(n)}$, the action codeword $a^n(m_1)$ is chosen and the channel state information (s_e^n, s_d^n) is generated as an output of the memoryless channel $P_{S_e^n, S_d^n|A^n}(s_e^n, s_d^n|a^n) = \prod_{i=1}^n P_{S_e,S_d|A}(s_{e,i}, s_{d,i}|a_i)$. The encoder looks for the smallest value of $j \in \mathcal{J}^{(n)}$ such that the tuple $\big(s_e^n, a^n(m_1), x^n(m_1, m_2, j)\big) \in T_{\epsilon_2}^{(n)}(S_e, A, X)$. If no such j exists, set $j = 1$. The channel input sequence is then chosen to be $x^n(m_1, m_2, j)$.

Decoding: Upon receiving y^n and s_d^n, the decoder in the first step looks for the smallest $\tilde{m}_1 \in \mathcal{M}_1^{(n)}$ such that $\big(y^n, s_d^n, a^n(\tilde{m}_1)\big) \in T_\epsilon^{(n)}(Y, S_d, A)$. If successful, then set $\hat{m}_1 = \tilde{m}_1$. Otherwise, set $\hat{m}_1 = 1$. Then, based on the known $a^n(\hat{m}_1)$, the decoder looks for a pair $(\tilde{m}_2, \tilde{j})$ with the smallest $\tilde{m}_2 \in \mathcal{M}_2^{(n)}$ and $\tilde{j} \in \mathcal{J}^{(n)}$ such that $\big(y^n, s_d^n, a^n(\hat{m}_1), x^n(\hat{m}_1, \tilde{m}_2, \tilde{j})\big) \in T_\epsilon^{(n)}(Y, S_d, A, X)$. If there exists such a pair, the decoded message is set to be $\hat{m} = (\hat{m}_1, \tilde{m}_2)$, and $\hat{x}^n = x^n(\hat{m}_1, \tilde{m}_2, \tilde{j})$. Otherwise, $\hat{m} = (1, 1)$ and $\hat{x}^n = x^n(1, 1, 1)$.[3]

Analysis of Probability of Error: Due to the symmetry of the random code construction, the error probability does not depend on which message was sent. Assuming that $M = (M_1, M_2)$ and J were sent and chosen at the encoder. We define the error events as follows.

$$\begin{aligned}
\mathcal{E}_1 &= \{A^n(M_1) \notin T_{\epsilon_0}^{(n)}(A)\} \\
\mathcal{E}_2 &= \big\{\big(S_e^n, S_d^n, A^n(M_1)\big) \notin T_{\epsilon_1}^{(n)}(S_e, S_d, A)\big\} \\
\mathcal{E}_3 &= \big\{\big(S_e^n, A^n(M_1), X^n(M_1, M_2, j)\big) \notin T_{\epsilon_2}^{(n)}(S_e, A, X) \text{ for all } j \in \mathcal{J}^{(n)}\big\} \\
\mathcal{E}_{4a} &= \big\{\big(Y^n, S_d^n, A^n(M_1)\big) \notin T_\epsilon^{(n)}(Y, S_d, A)\big\} \\
\mathcal{E}_{4b} &= \big\{\big(Y^n, S_d^n, A^n(\tilde{m}_1)\big) \in T_\epsilon^{(n)}(Y, S_d, A) \text{ for some } \tilde{m}_1 \in \mathcal{M}_1^{(n)}, \tilde{m}_1 \neq M_1\big\} \\
\mathcal{E}_{5a} &= \big\{\big(Y^n, S_d^n, A^n(M_1), X^n(M_1, M_2, J)\big) \notin T_\epsilon^{(n)}(Y, S_d, A, X)\big\} \\
\mathcal{E}_{5b} &= \big\{\big(Y^n, S_d^n, A^n(M_1), X^n(M_1, \tilde{m}_2, \tilde{j})\big) \in T_\epsilon^{(n)}(Y, S_d, A, X) \text{ for some} \\
&\qquad (\tilde{m}_2, \tilde{j}) \in \mathcal{M}_2^{(n)} \times \mathcal{J}^{(n)}, (\tilde{m}_2, \tilde{j}) \neq (M_2, J)\big\}.
\end{aligned}$$

The probability of error events can be bounded by

$$\begin{aligned}
\Pr(\mathcal{E}) \leq \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2 \cap \mathcal{E}_1^c) + \Pr(\mathcal{E}_3 \cap \mathcal{E}_2^c) + \Pr(\mathcal{E}_{4a} \cap \mathcal{E}_3^c) + \Pr(\mathcal{E}_{4b}) + \\
\Pr(\mathcal{E}_{5a} \cap \mathcal{E}_3^c) + \Pr(\mathcal{E}_{5b}),
\end{aligned}$$

where $\mathcal{E}_i^c$ denotes the complement of event $\mathcal{E}_i$.

1) Since $A^n(M_1)$ is i.i.d. according to P_A, by the LLN we have $\Pr(\mathcal{E}_1) \to 0$ as $n \to \infty$.

2) Consider the event $\mathcal{E}_1^c$ where we have $A^n(M_1) \in T_{\epsilon_0}^{(n)}(A)$. Since (S_d^n, S_e^n) is distributed according to $\prod_{i=1}^n P_{S_e,S_d|A}\big(s_{e,i}, s_{d,i}|a_i\big)$, by the conditional typicality lemma, we have that $\Pr\big(\mathcal{E}_2 \cap \mathcal{E}_1^c\big) \to 0$ as $n \to \infty$.

[3]We note that although the simultaneous joint typicality decoding gives us different constraints on the individual rate as compared to the sequential two-stage decoding considered in this proof, it gives the same constraints on the total transmission rate in which we are interested.

3) By the covering lemma where X^n is i.i.d. according to $\prod_{i=1}^n P_{X|A}(x_i|a_i)$, we have $\Pr\big(\mathcal{E}_3 \cap \mathcal{E}_2^c\big) \to 0$ as $n \to \infty$ if $\frac{1}{n}\log|\mathcal{J}^{(n)}| > I(X;S_e|A) + \delta_{\epsilon_2}$, where $\delta_{\epsilon_2} \to 0$ as $\epsilon_2 \to 0$.

4a) Consider the event $\mathcal{E}_3^c$ where we have the tuple $\big(S_e^n, A^n(M_1), X^n(M_1, M_2, J)\big) \in T_{\epsilon_2}^{(n)}(S_e, A, X)$. Since we have $S_d - (A, S_e) - X$ forms a Markov chain and S_d^n is distributed according to $\prod_{i=1}^n P_{S_d|A,S_e}\big(s_{d,i}|a_i, s_{e,i}\big)\big)$, we have that by the conditional typicality lemma, $\Pr\big((S_d^n, S_e^n, A^n(M_1), X^n(M_1, M_2, J)) \in T_\epsilon^{(n)}(S_d, S_e, A, X)\big) \to 1$ as $n \to \infty$. And since we have the Markov chain $A - (X, S_e, S_d) - Y$ and Y^n is distributed according to $\prod_{i=1}^n P_{Y|X,S_e,S_d}(y_i|x_i, s_{e,i}, s_{d,i})$, by using once again the conditional typicality lemma, it follows that $\Pr\big(\big(Y^n, S_e^n, S_d^n, A^n(M_1), X^n(M_1, M_2, J)\big)$ $\in T_\epsilon^{(n)}(Y, A, S_e, S_d, X)\big) \to 1$ as $n \to \infty$. This also implies that $\Pr(\mathcal{E}_{4a} \cap \mathcal{E}_3^c) \to 0$ as $n \to \infty$.

4b) By the packing lemma, we have $\Pr\big(\mathcal{E}_{4b}\big) \to 0$ as $n \to \infty$ if $\frac{1}{n}\log|\mathcal{M}_1^{(n)}| < I(A;Y,S_d) - \delta_\epsilon$, where $\delta_\epsilon \to 0$ as $\epsilon \to 0$.

5a) As in $\mathcal{E}_{4a}$) we have $\Pr(\mathcal{E}_{5a} \cap \mathcal{E}_3^c) \to 0$ as $n \to \infty$.

5b) Averaging over all $J = j$, by the packing lemma where X^n is i.i.d. according to $\prod_{i=1}^n P_{X|A}(x_i|a_i)$, we have $\Pr\big(\mathcal{E}_{5b}\big) \to 0$ as $n \to \infty$ if $\frac{1}{n}\log|\mathcal{M}_2^{(n)}| + \frac{1}{n}\log|\mathcal{J}^{(n)}| < I(X;Y,S_d|A) - \delta_\epsilon$.

Finally, by combining the bounds on the code rates that make $\Pr\big(\mathcal{E}\big) \to 0$ as $n \to \infty$,

$$\begin{aligned}
\frac{1}{n}\log|\mathcal{J}^{(n)}| &> I(X;S_e|A) + \delta_{\epsilon_2} \\
\frac{1}{n}\log|\mathcal{M}_1^{(n)}| &< I(A;Y,S_d) - \delta_\epsilon \\
\frac{1}{n}\log|\mathcal{M}_2^{(n)}| + \frac{1}{n}\log|\mathcal{J}^{(n)}| &< I(X;Y,S_d|A) - \delta_\epsilon,
\end{aligned}$$

we have

$$\begin{aligned}
\frac{1}{n}\log|\mathcal{M}^{(n)}| = \frac{1}{n}\log|\mathcal{M}_1^{(n)}||\mathcal{M}_2^{(n)}| &< I(A,X;Y,S_d) - I(X;S_e|A) - \delta_\epsilon' \\
\frac{1}{n}\log|\mathcal{M}_2^{(n)}| &< I(X;Y,S_d|A) - I(X;S_e|A) - \delta_\epsilon'',
\end{aligned}$$

where $\delta_\epsilon', \delta_\epsilon'' \to 0$ as $\epsilon \to 0$.

Since, for any $\delta > 0$, the achievable rate R satisfies $\frac{1}{n}\log|\mathcal{M}^{(n)}| \geq R - \delta$, and we know that $\frac{1}{n}\log|\mathcal{M}_2^{(n)}| \geq 0$, then we get

$$\begin{aligned}
R - \delta \leq \frac{1}{n}\log|\mathcal{M}^{(n)}| &< I(A,X;Y,S_d) - I(X;S_e|A) - \delta_\epsilon' \\
\text{and} \quad 0 \leq \frac{1}{n}\log|\mathcal{M}_2^{(n)}| &< I(X;Y,S_d|A) - I(X;S_e|A) - \delta_\epsilon''. \qquad (3.25)
\end{aligned}$$

Since ϵ can be made arbitrarily small for increasing n, and by a standard random coding argument, we have that

$$\begin{aligned} & R \leq I(A,X;Y,S_d) - I(X;S_e|A) \\ \text{and} \quad & 0 < I(X;Y,S_d|A) - I(X;S_e|A), \end{aligned}$$

for some $P_A(a)P_{S_e,S_d|A}(s_e,s_d|a)P_{X|A,S_e}(x|a,s_e)P_{Y|X,S_e,S_d}(y|x,s_e,s_d)$ is achievable.

Note that the latter condition is for the two-stage coding to be successful, i.e., we can split the message into two parts with positive rates. This concludes the achievability proof.

3.B.2 Converse Proof of Theorem 3.3.1

We will show that for any achievable rate R, it follows that $R \leq I(A,X;Y,S_d) - I(X;S_e|A)$ and $0 \leq I(X;Y,S_d|A) - I(X;S_e|A)$ for some $P_A(a)P_{S_e,S_d|A}(s_e,s_d|a)$ $P_{X|A,S_e}(x|a,s_e)P_{Y|X,S_e,S_d}(y|x,s_e,s_d)$. From the problem formulation, we can write the joint probability mass function,

$$\begin{aligned} & P_{M,A^n,S_e^n,S_d^n,X^n,Y^n,\hat{M},\hat{X}^n}(m,a^n,s_e^n,s_d^n,x^n,y^n,\hat{m},\hat{x}^n) \\ & = \frac{1_{\{A^n=f_a^{(n)}(m)\}}(a^n)1_{\{X^n=f^{(n)}(m,s_e^n)\}}(x^n)1_{\{\hat{X}^n=g_x^{(n)}(y^n,s_d^n)\}}(\hat{x}^n)1_{\{\hat{M}=g_m^{(n)}(y^n,s_d^n)\}}(\hat{m})}{|\mathcal{M}^{(n)}|} \\ & \quad \cdot \prod_{i=1}^{n} P_{S_e,S_d|A}(s_{e,i},s_{d,i}|a_i)P_{Y|X,S_e,S_d}(y_i|x_i,s_{e,i},s_{d,i}), \end{aligned} \tag{3.26}$$

where M is chosen uniformly at random from the set $\mathcal{M}^{(n)} = \{1,2,\ldots,|\mathcal{M}^{(n)}|\}$. For the joint PMF in (3.26), it can be shown that $S_{d,i} - (A_i,S_{e,i},X_i) - (M,A^{n\setminus i},S_{e,i+1}^n, X^n,Y^{i-1},S_d^{i-1})$ forms a Markov chain.

Let us assume that a specific sequence of $(|\mathcal{M}^{(n)}|,n)$ codes exists such that the average error probabilities $P_{m,e}^{(n)} = \delta_n' \leq \delta_n$ and $P_{x,e}^{(n)} = \delta_n' \leq \delta_n$, and $\log|\mathcal{M}^{(n)}| = n(R-\delta_n') \geq n(R-\delta_n)$, with $\lim_{n\to\infty}\delta_n = 0$. Then standard properties of the entropy function give

$$\begin{aligned} n(R-\delta_n) \leq \log|\mathcal{M}^{(n)}| &= H(M) \\ &= H(M) - H(X^n,M|Y^n,S_d^n) + H(M|Y^n,S_d^n) + H(X^n|M,Y^n,S_d^n) \\ &\leq H(M) - H(X^n,M|Y^n,S_d^n) + H(M|Y^n,S_d^n) + H(X^n|Y^n,S_d^n). \end{aligned}$$

Consider the last two terms in the above inequality. By Fano's inequality, we get

$$\begin{aligned} H(M|Y^n,S_d^n) &\leq h(\delta_n') + \delta_n' \cdot \log(2^{n(R-\delta_n')} - 1) = n\epsilon_n^{(m)}, \\ H(X^n|Y^n,S_d^n) &\leq h(\delta_n') + \delta_n' \cdot \log(|\mathcal{X}|^n - 1) = n\epsilon_n^{(x)}, \end{aligned}$$

where $h(\cdot)$ is the binary entropy function, and $\epsilon_n^{(m)} \to 0, \epsilon_n^{(x)} \to 0$ as $\delta_n \to 0$.

Let $n\epsilon_n^{(m)} + n\epsilon_n^{(x)} \triangleq n\epsilon_n$, where ϵ_n satisfies $\lim_{n\to\infty} \epsilon_n = 0$. Now we continue the chain of inequalities and get

$$\begin{aligned}
n(R-\delta_n) &\le H(M) - H(X^n, M|Y^n, S_d^n) + n\epsilon_n \\
&= H(M, S_e^n) - H(S_e^n|M) - H(X^n, M|Y^n, S_d^n) + n\epsilon_n \\
&\overset{(a)}{=} H(M, A^n, S_e^n, X^n) - H(S_e^n|M, A^n) - H(X^n, M, A^n|Y^n, S_d^n) + n\epsilon_n \\
&\overset{(b)}{=} H(M, A^n, S_e^n, X^n) - H(S_e^n|A^n) - H(X^n, M, A^n|Y^n, S_d^n) + n\epsilon_n \\
&= I(X^n, M, A^n; Y^n, S_d^n) - I(X^n, M; S_e^n|A^n) + n\epsilon_n, \qquad (3.27)
\end{aligned}$$

where (a) follows from the fact that $X^n = f^{(n)}(M, S_e^n)$ and $A^n = f_a^{(n)}(M)$, and (b) holds since S_e^n is independent of M given A^n. Continuing the chain of inequalities, we get

$$\begin{aligned}
&n(R-\delta_n-\epsilon_n) \\
&\le \sum_{i=1}^n I(X^n, M, A^n; Y_i, S_{d,i}|Y^{i-1}, S_d^{i-1}) - I(X^n, M; S_{e,i}|S_{e,i+1}^n, A^n) \\
&= \sum_{i=1}^n [I(X^n, M, A^n, S_{e,i+1}^n; Y_i, S_{d,i}|Y^{i-1}, S_d^{i-1}) \\
&\qquad - I(S_{e,i+1}^n; Y_i, S_{d,i}|X^n, M, A^n, Y^{i-1}, S_d^{i-1})] \\
&\qquad - [I(X^n, M, Y^{i-1}, S_d^{i-1}; S_{e,i}|S_{e,i+1}^n, A^n) \\
&\qquad - I(Y^{i-1}, S_d^{i-1}; S_{e,i}|X^n, M, S_{e,i+1}^n, A^n)] \\
&\overset{(a)}{=} \sum_{i=1}^n I(A_i, Z_i; Y_i, S_{d,i}|Y^{i-1}, S_d^{i-1}) - I(A_i, Z_i; S_{e,i}|S_{e,i+1}^n, A^n) \\
&\overset{(b)}{\le} \sum_{i=1}^n H(Y_i, S_{d,i}) - H(Y_i, S_{d,i}|Z_i, A_i) - H(S_{e,i}|A_i) + H(S_{e,i}|Z_i, A_i) \\
&= \sum_{i=1}^n I(A_i, Z_i; Y_i, S_{d,i}) - I(Z_i; S_{e,i}|A_i) \\
&\overset{(c)}{=} \sum_{i=1}^n I(A_i; Y_i, S_{d,i}) + I(Z_i, X_i; Y_i, S_{d,i}|A_i) - I(Z_i, X_i; S_{e,i}|A_i),
\end{aligned}$$

where (a) follows from the Csiszár's sum identity (Lemma 2.5), where we have that the sum of the second and fourth mutual information terms cancel out, and by defining $Z_i \triangleq (M, A^{n\setminus i}, S_{e,i+1}^n, X^n, Y^{i-1}, S_d^{i-1})$, (b) follows from the fact that $(S_{e,i+1}^n, A^{n\setminus i}) - A_i - S_{e,i}$ forms a Markov chain, and (c) follows from the definition of Z_i.

Consider the sum of the last two terms,

$$\begin{aligned}
&\sum_{i=1}^{n} I(Z_i, X_i; Y_i, S_{d,i}|A_i) - I(Z_i, X_i; S_{e,i}|A_i) \\
&= \sum_{i=1}^{n} [I(X_i; Y_i, S_{d,i}|A_i) - I(X_i; S_{e,i}|A_i)] \\
&\quad - [I(Z_i; Y_i, S_{d,i}|X_i, A_i) - I(Z_i; S_{e,i}|X_i, A_i)] \\
&\overset{(a)}{=} \sum_{i=1}^{n} I(X_i; Y_i, S_{d,i}|A_i) - I(X_i; S_{e,i}|A_i) - I(Z_i; S_{e,i}|X_i, A_i, Y_i, S_{d,i}) \\
&\leq \sum_{i=1}^{n} I(X_i; Y_i, S_{d,i}|A_i) - I(X_i; S_{e,i}|A_i),
\end{aligned}$$

where (a) follows from the memoryless property of the channel where $(Z_i, A_i) - (X_i, S_{e,i}, S_{d,i}) - Y_i$ forms a Markov chain and also the Markov chain $S_{d,i} - (A_i, S_{e,i}, X_i) - Z_i$ (derived from (3.26) and definition of Z_i) which together imply $Z_i - (A_i, X_i, S_{e,i}) - (S_{d,i}, Y_i)$ (Contraction property in Lemma 2.1). Finally, we get

$$n(R - \delta_n - \epsilon_n) \leq \sum_{i=1}^{n} I(A_i; Y_i, S_{d,i}) + I(X_i; Y_i, S_{d,i}|A_i) - I(X_i; S_{e,i}|A_i). \quad (3.28)$$

Next, we prove the two stage coding condition. It can be considered as the restriction imposed on the set of input distribution in a similar flavor as the dependence balance bound in [HW89]. From the standard properties of the entropy function, we observe that

$$\begin{aligned}
0 &\leq H(M|A^n) \\
&= H(M|A^n) - H(X^n, M|Y^n, S_d^n, A^n) + H(M|Y^n, S_d^n, A^n) \\
&\quad + H(X^n|M, Y^n, S_d^n, A^n) \\
&\overset{(a)}{\leq} H(M|A^n) - H(X^n, M|Y^n, S_d^n, A^n) + n\epsilon_n^{(m)} + n\epsilon_n^{(x)},
\end{aligned}$$

where (a) follows from Fano's inequality. Let $n\epsilon_n^{(m)} + n\epsilon_n^{(x)} \triangleq n\epsilon_n$, where ϵ_n satisfies $\lim_{n\to\infty} \epsilon_n = 0$. Now we continue the chain of inequalities and get

$$\begin{aligned}
-n\epsilon_n &\leq H(M|A^n) - H(X^n, M|Y^n, S_d^n, A^n) \\
&= H(M, S_e^n|A^n) - H(S_e^n|M, A^n) - H(X^n, M|Y^n, S_d^n, A^n) \\
&\overset{(a)}{=} H(M, S_e^n, X^n|A^n) - H(S_e^n|M, A^n) - H(X^n, M|Y^n, S_d^n, A^n) \\
&\overset{(b)}{=} H(M, S_e^n, X^n|A^n) - H(S_e^n|A^n) - H(X^n, M|Y^n, S_d^n, A^n) \\
&\overset{(c)}{\leq} \sum_{i=1}^{n} I(X_i; Y_i, S_{d,i}|A_i) - I(X_i; S_{e,i}|A_i), \quad (3.29)
\end{aligned}$$

where (a) follows from the fact that $X^n = f^{(n)}(M, S_e^n)$, (b) holds since S_e^n is independent of M given A^n, and (c) follows by repeating similar steps after (3.27) with every term conditioned on A^n.

Let Q be a random variable uniformly distributed over $\{1, \ldots, n\}$ and independent of $(M, A^n, X^n, S_e^n, S_d^n, Y^n)$, we can rewrite (3.28) and (3.29) as

$$\begin{aligned} R &\leq \frac{1}{n}\sum_{i=1}^{n} I(A_i, X_i; Y_i, S_{d,i}|Q=i) - I(X_i; S_{e,i}|A_i, Q=i) + \delta_n + \epsilon_n \\ &= I(A_Q, X_Q; Y_Q, S_{d,Q}|Q) - I(X_Q; S_{e,Q}|A_Q, Q) + \delta_n + \epsilon_n \end{aligned}$$

and similarly

$$0 \leq I(X_Q; Y_Q, S_{d,Q}|A_Q, Q) - I(X_Q; S_{e,Q}|A_Q, Q) + \epsilon_n.$$

Now since we have that $P_{S_{e,Q}, S_{d,Q}|A_Q} = P_{S_e,S_d|A}$, $P_{Y_Q|X_Q,S_{e,Q},S_{d,Q}} = P_{Y|X,S_e,S_d}$, and $A_Q - (X_Q, S_{e,Q}, S_{d,Q}) - Y_Q$ forms a Markov chain, we identify $A \triangleq A_Q$, $S_e \triangleq S_{e,Q}$, $S_d \triangleq S_{d,Q}$, $X \triangleq X_Q$, and $Y \triangleq Y_Q$ to finally obtain

$$\begin{aligned} R &\leq I(A, X; Y, S_d|Q) - I(X; S_e|A, Q) + \delta_n + \epsilon_n \\ \text{and } \ 0 &\leq I(X; Y, S_d|A, Q) - I(X; S_e|A, Q) + \epsilon_n, \end{aligned}$$

for some joint distribution of the form $P_Q P_{A|Q} P_{S_e,S_d|A} P_{X|A,S_e,Q} P_{Y|X,S_e,S_d}$.

To this end, under the joint distribution of the form above, we have that $(Y, S_d) - (X, A, S_e) - Q$ and $S_e - A - Q$ form Markov chains, and

$$\begin{aligned} I(A, X; Y, S_d|Q) - I(X; S_e|A, Q) &= I(A, X, S_e; Y, S_d|Q) - I(X, Y, S_d; S_e|A, Q) \\ &\leq H(Y, S_d) - H(Y, S_d|A, X, S_e, Q) - H(S_e|A, Q) + H(S_e|X, Y, S_d, A) \\ &\overset{(*)}{=} H(Y, S_d) - H(Y, S_d|A, X, S_e) - H(S_e|A) + H(S_e|X, Y, S_d, A) \\ &= I(A, X, S_e; Y, S_d) - I(X, Y; S_e|A) \\ &= I(A, X; Y, S_d) - I(X; S_e|A), \end{aligned}$$

and similarly

$$I(X; Y, S_d|A, Q) - I(X; S_e|A, Q) \leq I(X; Y, S_d|A) - I(X; S_e|A),$$

where the equality $(*)$ follows from the Markov chains $(Y, S_d) - (X, A, S_e) - Q$ and $S_e - A - Q$, and the joint distribution of (A, S_e, S_d, X, Y) is of the form

$$\begin{aligned} &\sum_{q \in \mathcal{Q}} P_Q(q) P_{A|Q}(a|q) P_{S_e,S_d|A}(s_e, s_d|a) P_{X|A,S_e,Q}(x|a, s_e, q) P_{Y|X,S_e,S_d}(y|x, s_e, s_d) \\ &\quad = P_A(a) P_{S_e,S_d|A}(s_e, s_d|a) P_{X|A,S_e}(x|a, s_e) P_{Y|X,S_e,S_d}(y|x, s_e, s_d). \end{aligned}$$

The proof is concluded by taking the limit $n \to \infty$.

3.C Proof of Proposition 3.3.1

3.C.1 Achievability Proof of Proposition 3.3.1

The proof follows from a standard random coding argument. We use the technique of rate splitting, i.e., the message M of rate R is split into two messages M_1 and M_2 of rates R_1 and R_2. Two-stage coding is then considered, i.e., the first stage for communicating the identity of the action sequence, and the second stage for communicating the identity of S_e^n based on the known action sequence.

For given channels with transition probabilities $P_{S_e,S_d|A}$ and $P_{Y|X,S_e,S_d}$ we can assign the joint probability to any random vector (A, X, S_e) by

$$P_{A,S_e,S_d,X,Y}(a, s_e, s_d, x, y) = P_A(a)P_{S_e,S_d|A}(s_e, s_d|a)P_{X|A,S_e}(x|a, s_e) \\ P_{Y|X,S_e,S_d}(y|x, s_e, s_d).$$

Codebook Generation: Fix P_A and $P_{X|A,S_e}$. Let $\mathcal{M}_1^{(n)} = \{1, 2, \ldots, |\mathcal{M}_1^{(n)}|\}$, $\mathcal{M}_2^{(n)} = \{1, 2, \ldots, |\mathcal{M}_2^{(n)}|\}$ and $\mathcal{J}^{(n)} = \{1, 2, \ldots, |\mathcal{J}^{(n)}|\}$. For all $m_1 \in \mathcal{M}_1^{(n)}$, generate $a^n(m_1)$ i.i.d. according to $\prod_{i=1}^n P_A(a_i)$. Then for each m_1, generate $|\mathcal{M}_2^{(n)}||\mathcal{J}^{(n)}|$ codewords $\{\check{s}_e^n(m_1, m_2, j)\}_{m_2 \in \mathcal{M}_2^{(n)}, j \in \mathcal{J}^{(n)}}$ i.i.d. each according to the distribution $\prod_{i=1}^n P_{S_e|A}(\check{s}_{e,i}|a_i(m_1))$. Finally, for each $(a^n, \check{s}_e^n)$ pair, generate x^n i.i.d. according to $\prod_{i=1}^n P_{X|S_e,A}(x_i|\check{s}_{e,i}, a_i(m_1))$. Then the codebooks are revealed to the action encoder, the channel encoder, and the decoder. Let $0 < \epsilon_0 < \epsilon_1 < \epsilon_2 < \epsilon < 1$.

Encoding: Given the message $m = (m_1, m_2) \in \mathcal{M}^{(n)}$, the action codeword $a^n(m_1)$ is chosen and the channel state information (s_e^n, s_d^n) is generated as an output of the memoryless channel, $P_{S_e^n,S_d^n|A^n}(s_e^n, s_d^n|a^n) = \prod_{i=1}^n P_{S_e,S_d|A}(s_{e,i}, s_{d,i}|a_i)$. The encoder looks for the smallest value of $j \in \mathcal{J}^{(n)}$ such that $\check{s}_e^n(m_1, m_2, j) = s_e^n$. The channel input sequence is then chosen to be $x^n(m_1, m_2, j)$. If no such j exists, set $j = 1$.

Decoding: Upon receiving y^n and s_d^n, the decoder in the first step looks for the smallest $\tilde{m}_1 \in \mathcal{M}_1^{(n)}$ such that $\big(y^n, s_d^n, a^n(\tilde{m}_1)\big) \in T_\epsilon^{(n)}(Y, S_d, A)$. If successful, then set $\hat{m}_1 = \tilde{m}_1$. Otherwise, set $\hat{m}_1 = 1$. Then, based on the known $a^n(\hat{m}_1)$, the decoder looks for a pair $(\tilde{m}_2, \tilde{j})$ with the smallest $\tilde{m}_2 \in \mathcal{M}_2^{(n)}$ and $\tilde{j} \in \mathcal{J}^{(n)}$ such that $\big(y^n, s_d^n, a^n(\hat{m}_1), \check{s}_e^n(\hat{m}_1, \tilde{m}_2, \tilde{j}), x^n(\hat{m}_1, \tilde{m}_2, \tilde{j})\big) \in T_\epsilon^{(n)}(Y, S_d, A, S_e, X)$. If there exists such a pair, the decoded message is set to be $\hat{m} = (\hat{m}_1, \tilde{m}_2)$, and the decoded state $\hat{s}_e^n = \check{s}_e^n(\hat{m}_1, \tilde{m}_2, \tilde{j})$. Otherwise, $\hat{m} = (1, 1)$ and $\hat{s}_e^n = \check{s}_e^n(1, 1, 1)$.[4]

Analysis of Probability of Error: Due to the symmetry of the random code construction, the error probability does not depend on which message was sent. Assuming that $M = (M_1, M_2)$ and J were sent and chosen at the encoder. We

[4]Similarly as in Appendix 3.B, both the sequential two-stage decoding considered in this proof and the simultaneous joint typicality decoding give the same constraints on the total transmission rate in which we are interested.

define the error events as follows.

$$\mathcal{E}_1 = \{A^n(M_1) \notin T_{\epsilon_0}^{(n)}(A)\}$$
$$\mathcal{E}_2 = \big\{\big(S_e^n, S_d^n, A^n(M_1)\big) \notin T_{\epsilon_1}^{(n)}(S_e, S_d, A)\big\}$$
$$\mathcal{E}_{3a} = \big\{S_e^n \neq \check{S}_e^n(M_1, M_2, j) \text{ for all } j \in \mathcal{J}^{(n)}\big\}$$
$$\mathcal{E}_{3b} = \big\{\big(S_e^n, A^n(M_1), X^n(M_1, M_2, J)\big) \notin T_{\epsilon_2}^{(n)}(S_e, A, X)\big\}$$
$$\mathcal{E}_{4a} = \big\{\big(Y^n, S_d^n, A^n(M_1)\big) \notin T_{\epsilon}^{(n)}(Y, S_d, A)\big\}$$
$$\mathcal{E}_{4b} = \big\{\big(Y^n, S_d^n, A^n(\tilde{m}_1)\big) \in T_{\epsilon}^{(n)}(Y, S_d, A) \text{ for some } \tilde{m}_1 \in \mathcal{M}_1^{(n)}, \tilde{m}_1 \neq M_1\big\}$$
$$\mathcal{E}_{5a} = \big\{\big(Y^n, S_d^n, A^n(M_1), \check{S}_e^n(M_1, M_2, J), X^n(M_1, M_2, J)\big) \notin T_{\epsilon}^{(n)}(Y, S_d, A, S_e, X)\big\}$$
$$\mathcal{E}_{5b} = \big\{\big(Y^n, S_d^n, A^n(M_1), \check{S}_e^n(M_1, \tilde{m}_2, \tilde{j}), X^n(M_1, \tilde{m}_2, \tilde{j})\big) \in T_{\epsilon}^{(n)}(Y, S_d, A, S_e, X)$$
$$\text{for some } (\tilde{m}_2, \tilde{j}) \in \mathcal{M}_2^{(n)} \times \mathcal{J}^{(n)}, (\tilde{m}_2, \tilde{j}) \neq (M_2, J)\big\}.$$

The probability of error events can be bounded by

$$\Pr(\mathcal{E}) \leq \Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2 \cap \mathcal{E}_1^c) + \Pr(\mathcal{E}_{3a} \cap \mathcal{E}_2^c) + \Pr(\mathcal{E}_{3b} \cap \mathcal{E}_{3a}^c \cap \mathcal{E}_2^c) + \Pr(\mathcal{E}_{4a} \cap \mathcal{E}_3^c) + \Pr(\mathcal{E}_{4b}) + \Pr(\mathcal{E}_{5a} \cap \mathcal{E}_3^c) + \Pr(\mathcal{E}_{5b}),$$

where $\mathcal{E}_3 \triangleq \mathcal{E}_{3a} \cup \mathcal{E}_{3b}$ and $\mathcal{E}_i^c$ denotes the complement of event $\mathcal{E}_i$.

1) Since $A^n(M_1)$ is i.i.d. according to P_A, by the LLN we have $\Pr(\mathcal{E}_1) \to 0$ as $n \to \infty$.

2) Consider the event $\mathcal{E}_1^c$ where we have $A^n(M_1) \in T_{\epsilon_0}^{(n)}(A)$. Since (S_d^n, S_e^n) is distributed according to $\prod_{i=1}^n P_{S_e,S_d|A}\big(s_{e,i}, s_{d,i}|a_i\big)$, by the conditional typicality lemma, we have that $\Pr\big(\mathcal{E}_2 \cap \mathcal{E}_1^c\big) \to 0$ as $n \to \infty$.

3a) Consider the event $\mathcal{E}_2^c$ where we have $\big(S_e^n, S_d^n, A^n(M_1)\big) \in T_{\epsilon_1}^{(n)}(S_e, S_d, A)$. It follows from the property of typical sequences (see Theorem 2.1.2) that $P_{S_e^n|A}(s_e^n|a^n) \geq 2^{-n[H(S_e|A)+\delta\epsilon_1]}$. Since both S_e^n and $\check{S}_e^n$ are i.i.d. according to $P_{S_e|A}$, we have $\Pr\big(\mathcal{E}_{31} \cap \mathcal{E}_2^c\big) \to 0$ as $n \to \infty$ if $\frac{1}{n}\log|\mathcal{J}^{(n)}| > H(S_e|A) + \delta_{\epsilon_1}$, where $\delta_{\epsilon_1} \to 0$ as $\epsilon_1 \to 0$.

3b) Consider the event $\mathcal{E}_{3a}^c$ where J is selected and $S_e^n = \check{S}_e^n(M_1, M_2, J)$. Since X^n is i.i.d. according to $\prod_{i=1}^n P_{X|S_e,A}(x_i|s_{e,i}, a_i)$, by the conditional typicality lemma, we have that $\Pr\big(\mathcal{E}_{3b} \cap \mathcal{E}_{3a}^c \cap \mathcal{E}_2^c\big) \to 0$ as $n \to \infty$.

4a) Consider the event $\mathcal{E}_3^c$ where we have the tuple $\big(S_e^n, A^n(M_1), X^n(M_1, M_2, J)\big)$ $\in T_{\epsilon_2}^{(n)}(S_e, A, X)$. Since we have $S_d - (A, S_e) - X$ forms a Markov chain and S_d^n is distributed according to $\prod_{i=1}^n P_{S_d|A,S_e}\big(s_{d,i}|a_i, s_{e,i}\big)$, by the conditional typicality lemma, we have that $\Pr\big((S_d^n, S_e^n, A^n, X^n) \in T_{\epsilon}^{(n)}(S_d, S_e, A, X)\big) \to 1$ as $n \to \infty$. And since we have the Markov chain $A - (X, S_e, S_d) - Y$ and Y^n is distributed according to $\prod_{i=1}^n P_{Y|X,S_e,S_d}(y_i|x_i, s_{e,i}, s_{d,i})$, by using once again the conditional typicality lemma, it follows that $\Pr\big(\big(Y^n, S_e^n, S_d^n, A^n(M_1), X^n(M_1, M_2, J)\big) \in T_{\epsilon}^{(n)}(Y, A, S_e, S_d, X)\big) \to 1$ as $n \to \infty$. This also implies that $\Pr(\mathcal{E}_{4a} \cap \mathcal{E}_3^c) \to 0$ as $n \to \infty$.

4*b*) By the packing lemma, we have $\Pr(\mathcal{E}_{4b}) \to 0$ as $n \to \infty$ if $\frac{1}{n}\log|\mathcal{M}_1^{(n)}| < I(A;Y,S_d) - \delta_\epsilon$, where $\delta_\epsilon \to 0$ as $\epsilon \to 0$.

5*a*) As in $\mathcal{E}_{4a}$) we have $\Pr(\mathcal{E}_{5a} \cap \mathcal{E}_3^c) \to 0$ as $n \to \infty$.

5*b*) Averaging over all $J = j$, by the packing lemma where $\check{S}_e^n$ is i.i.d. according to $\prod_{i=1}^n P_{S_e|A}(\check{s}_{e,i}|a_i)$ and X^n is i.i.d. according to $\prod_{i=1}^n P_{X|S_e,A}(x_i|\check{s}_{e,i},a_i)$, we have $\Pr(\mathcal{E}_{5b}) \to 0$ as $n \to \infty$ if $\frac{1}{n}\log|\mathcal{M}_2^{(n)}| + \frac{1}{n}\log|\mathcal{J}^{(n)}| < I(S_e,X;Y,S_d|A) - \delta_\epsilon$.

Finally, by combining the bounds on the code rates,

$$\begin{aligned}\frac{1}{n}\log|\mathcal{J}^{(n)}| &> H(S_e|A) + \delta_{\epsilon_1}\\ \frac{1}{n}\log|\mathcal{M}_1^{(n)}| &< I(A;Y,S_d) - \delta_\epsilon\\ \frac{1}{n}\log|\mathcal{M}_2^{(n)}| + \frac{1}{n}\log|\mathcal{J}^{(n)}| &< I(S_e,X;Y,S_d|A) - \delta_\epsilon,\end{aligned}$$

where $\epsilon > 0$ can be made arbitrarily small with increasing block length n, we have shown that, for any $\delta > 0$, with n sufficiently large, $\Pr(\mathcal{E}) < \delta$ when $R \le I(A;Y,S_d) + I(S_e,X;Y,S_d|A) - H(S_e|A)$ and $I(S_e,X;Y,S_d|A) - H(S_e|A) > 0$ for some $P_A(a)P_{S_e,S_d|A}(s_e,s_d|a)P_{X|A,S_e}(x|a,s_e)P_{Y|X,S_e,S_d}(y|x,s_e,s_d)$. Again, we note that the latter condition is for the successful two-stage coding, i.e., we can split the message into two parts with positive rates. This together with a random coding argument concludes the achievability proof.

3.C.2 Converse Proof of Proposition 3.3.1

We show that, for any achievable rate R, it follows that $R \le I(A,X,S_e;Y,S_d) - H(S_e|A)$ and $0 \le I(S_e,X;Y,S_d|A) - H(S_e|A)$ for some $P_A(a)P_{S_e,S_d|A}(s_e,s_d|a)$ $P_{X|A,S_e}(x|a,s_e)P_{Y|X,S_e,S_d}(y|x,s_e,s_d)$. From the problem formulation, we can write the joint probability mass function,

$$\begin{aligned}&P_{M,A^n,S_e^n,S_d^n,X^n,Y^n,\hat{M},\hat{S}_e^n}(m,a^n,s_e^n,s_d^n,x^n,y^n,\hat{m},\hat{s}_e^n)\\ &= \frac{1_{\{A^n=f_a^{(n)}(m)\}}(a^n)1_{\{X^n=f^{(n)}(m,s_e^n)\}}(x^n)1_{\{\hat{S}_e^n=g_{s_e}^{(n)}(y^n,s_d^n)\}}(\hat{s}_e^n)1_{\{\hat{M}=g_m^{(n)}(y^n,s_d^n)\}}(\hat{m})}{|\mathcal{M}^{(n)}|}\\ &\quad\cdot\prod_{i=1}^n P_{S_e,S_d|A}(s_{e,i},s_{d,i}|a_i)P_{Y|X,S_e,S_d}(y_i|x_i,s_{e,i},s_{d,i}),\end{aligned}$$

where M is chosen uniformly at random from the set $\mathcal{M}^{(n)} = \{1,2,\ldots,|\mathcal{M}^{(n)}|\}$.

Let us assume that a specific sequence of $(|\mathcal{M}^{(n)}|,n)$ codes exists such that the average error probabilities $P_{m,e}^{(n)} = \delta_n' \le \delta_n$, $P_{s_e,e}^{(n)} = \delta_n' \le \delta_n$, and $\log|\mathcal{M}^{(n)}| = n(R-\delta_n') \ge n(R-\delta_n)$, with $\lim_{n\to\infty}\delta_n = 0$. Then standard properties of the entropy function give

$$\begin{aligned}n(R-\delta_n) &\le \log|\mathcal{M}^{(n)}| = H(M)\\ &\overset{(*)}{\le} H(M) - H(X^n,S_e^n,M|Y^n,S_d^n) + H(M|Y^n,S_d^n) + H(S_e^n|Y^n,S_d^n)\end{aligned}$$

where $(*)$ holds since $X^n = f^{(n)}(M, S_e^n)$ and $f^{(n)}(\cdot)$ is a deterministic function. By Fano's inequality, we have that

$$H(M|Y^n, S_d^n) \le h(\delta_n') + \delta_n' \cdot \log(2^{n(R-\delta_n')} - 1) = n\epsilon_n^{(m)},$$
$$H(S_e^n|Y^n, S_d^n) \le h(\delta_n') + \delta_n' \cdot \log(|\mathcal{S}_e|^n - 1) = n\epsilon_n^{(s_e)},$$

where $h(\cdot)$ is the binary entropy function, and $\epsilon_n^{(m)} \to 0, \epsilon_n^{(s_e)} \to 0$ as $n \to \infty$.

Let $n\epsilon_n^{(m)} + n\epsilon_n^{(s_e)} \triangleq n\epsilon_n$, where ϵ_n satisfies $\lim_{n\to\infty} \epsilon_n = 0$. Now we continue the chain of inequalities and get

$$\begin{aligned}
n(R - \delta_n) &\le H(M) - H(X^n, S_e^n, M|Y^n, S_d^n) + n\epsilon_n \\
&= H(M, S_e^n) - H(S_e^n|M) - H(X^n, S_e^n, M|Y^n, S_d^n) + n\epsilon_n \\
&\overset{(a)}{=} H(M, S_e^n, X^n) - H(S_e^n|M, A^n) - H(X^n, S_e^n, M|Y^n, S_d^n) + n\epsilon_n \\
&\overset{(b)}{=} H(M, S_e^n, X^n) - H(S_e^n|A^n) - H(X^n, S_e^n, M|Y^n, S_d^n) + n\epsilon_n \\
&= I(X^n, S_e^n, M; Y^n, S_d^n) - H(S_e^n|A^n) + n\epsilon_n, \qquad (3.30)
\end{aligned}$$

where (a) follows from the fact that $X^n = f^{(n)}(M, S_e^n)$ and $A^n = f_a^{(n)}(M)$, (b) follows since S_e^n is independent of M given A^n.

Continuing the chain of inequalities, we get

$$\begin{aligned}
&n(R - \delta_n - \epsilon_n) \\
&\overset{(a)}{\le} \sum_{i=1}^n I(X^n, S_e^n, M; Y_i, S_{d,i}|Y^{i-1}, S_d^{i-1}) - H(S_{e,i}|A_i) \\
&\overset{(b)}{=} \sum_{i=1}^n H(Y_i, S_{d,i}|Y^{i-1}, S_d^{i-1}) - H(Y_i, S_{d,i}|Y^{i-1}, S_d^{i-1}, X^n, S_e^n, M, A_i) - H(S_{e,i}|A_i) \\
&\overset{(c)}{\le} \sum_{i=1}^n H(Y_i, S_{d,i}) - H(Y_i, S_{d,i}|X_i, S_{e,i}, A_i) - H(S_{e,i}|A_i) \\
&= \sum_{i=1}^n I(A_i, S_{e,i}, X_i; Y_i, S_{d,i}) - H(S_{e,i}|A_i), \qquad (3.31)
\end{aligned}$$

where (a) follows from the memoryless property of the channel $P_{S_e|A}$, (b) follows from the fact that $A^n = f_a^{(n)}(M)$, and (c) follows from the Markov chain $(Y^{i-1}, S_d^{i-1}, X^{n\setminus i}, S_e^{n\setminus i}, M) - (X_i, S_{e,i}, A_i) - (Y_i, S_{d,i})$ and that conditioning reduces entropy.

Next we prove the "two-stage coding condition." From the standard properties of the entropy function, we observe that

$$\begin{aligned}
&0 \le H(M|A^n) \\
&\overset{(*)}{=} H(M|A^n) - H(X^n, S_e^n, M|Y^n, S_d^n, A^n) + H(M|Y^n, S_d^n, A^n)
\end{aligned}$$

$$
\begin{aligned}
&\quad + H(S_e^n|M, Y^n, S_d^n, A^n) \\
&\leq H(M|A^n) - H(X^n, S_e^n, M|Y^n, S_d^n, A^n) + H(M|Y^n, S_d^n) + H(S_e^n|Y^n, S_d^n),
\end{aligned}
$$

where $(*)$ follows from the fact that $X^n = f^{(n)}(M, S_e^n)$.

Again, applying Fano's inequality to last two terms in the above inequality, we get

$$
\begin{aligned}
-n\epsilon_n &\leq H(M|A^n) - H(X^n, S_e^n, M|Y^n, S_d^n, A^n) \\
&= H(M, S_e^n|A^n) - H(S_e^n|M, A^n) - H(X^n, S_e^n, M|Y^n, S_d^n, A^n) \\
&\overset{(a)}{=} H(M, S_e^n, X^n|A^n) - H(S_e^n|M, A^n) - H(X^n, S_e^n, M|Y^n, S_d^n, A^n) \\
&\overset{(b)}{=} H(M, S_e^n, X^n|A^n) - H(S_e^n|A^n) - H(X^n, S_e^n, M|Y^n, S_d^n, A^n) \\
&= I(X^n, S_e^n, M; Y^n, S_d^n|A^n) - H(S_e^n|A^n) \\
&\overset{(c)}{\leq} \sum_{i=1}^{n} I(S_{e,i}, X_i; Y_i, S_{d,i}|A_i) - H(S_{e,i}|A_i),
\end{aligned} \tag{3.32}
$$

where (a) follows from the fact that $X^n = f^{(n)}(M, S_e^n)$, (b) holds since S_e^n is independent of M given A^n, and (c) follows by repeating similar steps after (3.30) with every term conditioned on A^n.

To this end, we can follow similar steps as in the end of the converse proof of Theorem 3.3.1 for the single letterization.

Chapter 4

Multi-stage Coding for Channels With a Rewrite Option and Reversible Input

In Chapter 3, we have studied basic models for source and channel coding with action-dependent side information under additional reconstruction requirement. The rate-distortion-cost region and the channel capacity for the respective source and channel problems were characterized. One interesting observation is the presence of the so called *two-stage coding condition* in the capacity expression of the channel with action-dependent states and reversible input. It arises from the coding structure of the system in which we require to reconstruct the signal generated in the second stage. Inspired by the two-stage coding condition, this chapter explores a similar impact of the input reconstruction requirement in the multi-stage setting. In particular, we consider a problem of *constrained* multi-stage coding for channels with a rewrite option. It is a natural extension of the problem in Section 3.3 in the previous chapter to the multi-stage case where an encoder in each stage observes its own message as well as all previous-stage messages, inputs, and outputs, and the decoder is required to reconstruct all the messages and channel input sequences reliably. The complete characterization of the channel capacity region is given for the two-stage case, while the inner and outer bounds to the capacity regions for the cases of three or more stages are provided. For the two-stage case, a discussion regarding the rate constraint of the message in the second stage is also given in which we can draw a connection to the two-stage coding condition.

4.1 Introduction

The problem of coding for channels with random states has received considerable attention due to its broad set of applications, e.g., in a computer memory with defective cells, digital watermarking, etc. In [GP80] Gel'fand and Pinsker solved the problem of coding for channels with states where the states are known noncausally at the encoder. Several extensions are then considered such as the cases with causal, two-sided, and partial state information [KSM07].

As discussed in Sections 2.3.1 and 2.3.2, Weissman in [Wei10] considered an extension in which the channel state is allowed to depend on a message-dependent action sequence, while Sumszyk and Steinberg [SS09] studied an extension in the context of information embedding where apart from decoding the embedded message, the decoder has an additional constraint on reconstructing the channel input signal (stegotext) reliably (reversible stegotext). For discussions regarding the action-dependent state and additional input reconstruction, the readers are referred to Section 3.1 in Chapter 3.

In Section 3.3 of Chapter 3, we considered the combined setting of Weissman [Wei10] and Sumszyk and Steinberg [SS09] where the channel state is allowed to depend on the action sequence and the decoder is interested in decoding both message and channel input signal reliably. We provided a complete characterization of the channel capacity in which we showed that the *two-stage coding condition* plays a role in restricting the set of admissible input distributions. In fact, this condition arises from the coding structure of the system in which we require to reconstruct the signal generated in the second stage. We also showed by example that the condition can be active when computing the capacity.

To gain more insight into the structure of the problem as well as the role of the two-stage coding condition, in this chapter, we study an extension of the previously considered problem where we add more encoding stages (K stages) and in each stage a channel encoder observes its own message and all previous-stage messages. The channel output of each transmission stage will play a role of channel state in the next-stage transmission and it is assumed to be available noncausally to the next-stage encoder as a channel state information. The encoder in each stage is assumed to have *memory*, i.e., all previous channel inputs and outputs are known. At the final stage, the decoder is interested in decoding all the messages and channel input sequences. See Fig. 4.1 for the case of $K = 3$. This setup can be seen as a problem of multi-stage coding for channels with a "rewrite option" based on the noise-free feedback, i.e., the writing output is given noncausally to the next-stage encoder as a channel state information, and the encoder has an option to make another pass where it may rewrite at whichever locations it chooses [Wei10]. Our setting is motivated by scenarios with *multiple* writing on a memory cell with a rewrite option where the decoder is interested in both decoding the embedded messages and in tracking what have been written in the previous stages (reversible input).

We give a single-letter characterization of the capacity region for the case of two-stage coding. Interestingly but perhaps not surprisingly, this helps to establish a connection to the result in our previously studied problem on channel with action-dependent state and reversible input in Chapter 3. That is, the two-stage coding condition noted in Theorem 3.3.1 can simply be seen as a degenerate rate constraint derived from the underlying rate constraint in this multi-stage coding setup. For $K \geq 3$, we provide inner and outer bounds to the capacity region. The inner bound can be achieved by a straightforward extension of the coding scheme for the case of $K = 2$. As for proving the tightness of the bounds, we find it challenging due to the structure of our problem formulation which prevents us from having the

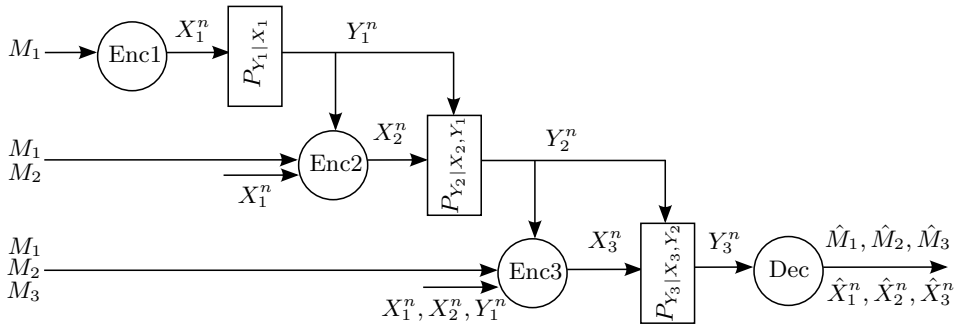

Figure 4.1: Three-stage channel with rewrite option and reversible input. (© 2012 IEEE)

desired Markov chains to account for the dependency of the channel states on the channel outputs. For some special cases, e.g., where the channels are deterministic, the inner and outer bounds coincide, establishing the capacity region.

The rest of the chapter is organized as follows. In Section 4.2 we formulate the problem and present the capacity region for the two-stage case. The discussion on how the result is related to another variation of the problem is then given. Section 4.3 discusses the connection between the rate constraint in the second stage transmission and the two-stage coding condition noted in the previous chapter. Lastly, inner and outer bounds to the capacity region for $K \geq 3$ are given in Section 4.4, where in some special cases, they are shown to coincide.

Notation: We use the shorthand notation X_i^n or $\boldsymbol{X}_i$ for a length-n sequence $X_{i,1}^n = (X_{i,1}, \ldots, X_{i,n})$. The term $\boldsymbol{X}_j^k$ denotes the vector of length-n sequences $(\boldsymbol{X}_j, \ldots, \boldsymbol{X}_k)$ when $j \leq k$, and the empty set otherwise, where $\boldsymbol{X}_j = X_{j,1}^n = (X_{j,1}, \ldots, X_{j,n})$. The term $(X_j^k)_l$ denotes the vector of the l^{th} elements of $\boldsymbol{X}_j^k$, i.e., $(X_j^k)_l = (X_{j,l}, \ldots, X_{k,l})$, and similarly $(X_j^k)^{n\setminus l}$ denotes $(X_j^{n\setminus l}, \ldots, X_k^{n\setminus l})$, where $X_j^{n\setminus l} = (X_{j,1}, \ldots, X_{j,l-1}, X_{j,l+1}, \ldots, X_{j,n})$.

4.2 Capacity Region for Two-stage Case, $K = 2$

We first provide the problem formulation for the general K-stage setting. Then we will focus on the case of $K = 2$ where we are able to derive the capacity result. The discussion on how the result is related to other problems is also given.

4.2.1 Problem Formulation

Let n denote the block length and $\mathcal{X}_i, \mathcal{Y}_i, i = 1, \ldots, K$ be finite sets. The setting consists of K encoders and one decoder. Message M_i chosen uniformly from the set $\mathcal{M}_i^{(n)} = \{1, 2, \ldots, |\mathcal{M}_i^{(n)}|\}$ is given to the encoders $i, i+1, \ldots, K$, for all $i = 1, \ldots, K$. The encoder in each stage is assumed to observe as well all previous-stage inputs and

outputs. A set of channels is described by a set of tuples $\{(\mathcal{X}_i, \mathcal{Y}_{i-1}, P_{Y_i|X_i,Y_{i-1}}, \mathcal{Y}_i)\}$, for $i = 1, \ldots, K$, where $\mathcal{X}_i$, $\mathcal{Y}_i$, and $P_{Y_i|X_i,Y_{i-1}}$ are the input alphabet, the output alphabet, and the transition probability from $\mathcal{X}_i \times \mathcal{Y}_{i-1}$ to $\mathcal{Y}_i$ in stage i, respectively. The decoder, which might be considered as two separate decoders, i.e., message decoder and channel input decoder, decodes all the messages and channel inputs based on the final-stage channel output Y_K^n. The channels in all stages $i = 1, \ldots, K$ are assumed to be memoryless and used without feedback with transition probabilities,

$$P_{Y_i^n|X_i^n,Y_{i-1}^n}(y_i^n|x_i^n, y_{i-1}^n) = \prod_{l=1}^{n} P_{Y_i|X_i,Y_{i-1}}(y_{i,l}|x_{i,l}, y_{i-1,l}).$$

Definition 4.1. An $(|\mathcal{M}_1^{(n)}|, \ldots, |\mathcal{M}_K^{(n)}|, n)$ code for the channels $\{P_{Y_i|X_i,Y_{i-1}}\}, i = 1, \ldots, K$ consists of the following functions

- encoders
$$f_i^{(n)} : \mathcal{M}_1^{(n)} \times \ldots \times \mathcal{M}_i^{(n)} \times \mathcal{X}_1^n \times \ldots \times \mathcal{X}_{i-1}^n \times \mathcal{Y}_1^n \times \ldots \times \mathcal{Y}_{i-1}^n \to \mathcal{X}_i^n, \quad i = 1, \ldots, K,$$
- a message decoder $g_m^{(n)} : \mathcal{Y}_K^n \to \mathcal{M}_1^{(n)} \times \mathcal{M}_2^{(n)} \times \ldots \times \mathcal{M}_K^{(n)}$, and
- a channel input decoder $g_x^{(n)} : \mathcal{Y}_K^n \to \mathcal{X}_1^n \times \mathcal{X}_2^n \times \ldots \times \mathcal{X}_K^n$.

Definition 4.2. A rate tuple $(R_1, \ldots, R_K) \in \mathbb{R}_+^K$ is said to be *achievable* if for any $\delta > 0$ there exists for all sufficiently large n an $(|\mathcal{M}_1^{(n)}|, \ldots, |\mathcal{M}_K^{(n)}|, n)$-code such that $\frac{1}{n} \log |\mathcal{M}_i^{(n)}| \geq R_i - \delta$, for $i = 1, \ldots, K$, $P_{m,e}^{(n)} \leq \delta$, and $P_{x,e}^{(n)} \leq \delta$, where $P_{m,e}^{(n)}$ and $P_{x,e}^{(n)}$ are the average error probabilities that $(M_1, M_2, \ldots, M_K) \neq g_m^{(n)}(Y_K^n)$ and $(X_1^n, X_2^n, \ldots, X_K^n) \neq g_x^{(n)}(Y_K^n)$, respectively. The *capacity region* $\mathcal{C}^{(K)}$ is the set of all achievable rate tuples.

We first consider the case of $K = 2$ for which we can derive the capacity region. In this case, the message M_1 is given to both encoder 1 and 2, and M_2 is given only to the encoder 2. The first-stage channel input sequence X_1^n is chosen based on the message M_1 and is the input to the *first-stage channel* which is described by a triple $(\mathcal{X}_1, P_{Y_1|X_1}, \mathcal{Y}_1)$. The output Y_1^n then becomes the channel state of the *second-stage channel* which is described by a quadruple $(\mathcal{X}_2, \mathcal{Y}_1, P_{Y_2|X_2,Y_1}, \mathcal{Y}_2)$, and is also given to the encoder 2 noncausally. The decoder decodes the messages (M_1, M_2) and the inputs (X_1^n, X_2^n) based on the final-stage channel output Y_2^n.

4.2.2 Main Result

Theorem 4.2.1 (Capacity region)**.** *The capacity region $\mathcal{C}^{(2)}$ of the two-stage setup is given by the set of all $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying*

$$R_1 + R_2 \leq I(X_1, X_2; Y_2) - I(X_2; Y_1|X_1), \tag{4.1}$$
$$R_2 \leq I(X_2; Y_2|X_1) - I(X_2; Y_1|X_1), \tag{4.2}$$

for some joint distributions of the form

$$P_{X_1}(x_1)P_{Y_1|X_1}(y_1|x_1)P_{X_2|X_1,Y_1}(x_2|x_1,y_1)P_{Y_2|X_2,Y_1}(y_2|x_2,y_1).$$

Proof. The achievability proof is based on a standard random coding argument using Gel'fand-Pinsker coding, and is a modification of that in Section 3.3, Chapter 3. For example, we follow the proof of an achievable rate pair (R_1, R_2) in Appendix 3.B, with random variables X_1, Y_1, X_2, and Y_2 replacing A, S_e, X, and Y, respectively. Alternatively, one can refer to a special case of the achievability proof for a general K-stage problem of which the sketch is given in Appendix 4.A. For the converse proof, we refer readers to the one given for the K-stage problem in Section 4.4 in which we have to apply with some modifications at the end of the proof, i.e., we use similar steps as in the case $j = K$ for both $j = 1$ and $j = 2$. □

4.2.3 Related Result

So far we have characterized the capacity region of the two-stage problem formulated with the reversible input requirement. It is also natural to consider a new and slightly different communication problem where the decoder is interested in decoding the messages (M_1, M_2) and the "channel state" Y_1^n instead. Due to the deterministic encoding functions, the channel input (X_1^n, X_2^n) can be retrieved based on the decoded messages and state information. We note that this communication problem has a more demanding reconstruction constraint than the previous problem studied in Section 4.2.1 for $K = 2$ since it essentially requires that the decoder can decode the message, the state, and the channel input signal, all reliably. We show that if the objective is to decode only the message and the channel input, then decoding the message and the state first, and then re-encoding the channel input is suboptimal.

Proposition 4.2.1 (Reconstruct Y_1^n)**.** *Consider the new two-stage communication problem in which the decoder is required to decode the messages (M_1, M_2) and the "channel state" Y_1^n reliably. The capacity region $\mathcal{C}_{Y_1}^{(2)}$ of such a channel is given by the set of all $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying*

$$R_1 + R_2 \le I(X_1, Y_1, X_2; Y_2) - H(Y_1|X_1), \tag{4.3}$$

$$R_2 \le I(Y_1, X_2; Y_2|X_1) - H(Y_1|X_1), \tag{4.4}$$

for some joint distributions of the form

$$P_{X_1}(x_1)P_{Y_1|X_1}(y_1|x_1)P_{X_2|X_1,Y_1}(x_2|x_1,y_1)P_{Y_2|X_2,Y_1}(y_2|x_2,y_1). \tag{4.5}$$

Proof. Since decoding (M_1, M_2) and Y_1^n implies that (X_1^n, X_2^n) is also decoded from the deterministic encoding functions, one can substitute (Y_1, X_2) in place of X_2 in Theorem 4.2.1 and obtain the new capacity region. More specifically, the achievable scheme in this case is different from the previous case of decoding (M_1, M_2) and

(X_1^n, X_2^n) in that the state information codebook is introduced and it should "cover" all possible generated Y_1^n *losslessly*. The proof of this proposition follows similarly as that of Proposition 3.3.1 in Chapter 3. □

Remark 4.1. We know that the channel input sequence can be retrieved based on the decoded message, the state information, and a known deterministic encoding function. Therefore, it is natural to compare the capacity region $\mathcal{C}^{(2)}$ in Theorem 4.2.1 with $\mathcal{C}_{Y_1}^{(2)}$ in Proposition 4.2.1. For a given channel $P_{Y_1|X_1}, P_{Y_2|X_2,Y_1}$, we have that $\mathcal{C}^{(2)} \supseteq \mathcal{C}_{Y_1}^{(2)}$.

Proof. We can show that both terms on the right hand side of (4.1) and (4.2) are greater than or equal to those of (4.3) and (4.4) for all joint distributions factorized as in (4.5), i.e.,

$$\begin{aligned} I(X_1, X_2; Y_2) - I(X_2; Y_1|X_1) &= I(X_1, Y_1, X_2; Y_2) - I(X_2, Y_2; Y_1|X_1) \\ &\geq I(X_1, Y_1, X_2; Y_2) - H(Y_1|X_1) \quad \text{and} \\ I(X_2; Y_2|X_1) - I(X_2; Y_1|X_1) &= I(Y_1, X_2; Y_2|X_1) - I(X_2, Y_2; Y_1|X_1) \\ &\geq I(Y_1, X_2; Y_2|X_1) - H(Y_1|X_1). \end{aligned}$$

We conclude that $\mathcal{C}^{(2)} \supseteq \mathcal{C}_{Y_1}^{(2)}$. □

4.3 Connection to Two-stage Coding Condition

In this section, we relate the main result for the case of $K = 2$ to the previous result on the capacity of channels with action-dependent state and reversible input considered in Section 3.3, Chapter 3. It is obvious from the problem formulation that if the rate of the message M_2 is zero, i.e., no new message to be transmitted at the encoder 2, then the two-stage setting in this chapter simply reduces to the problem of channel with action-dependent state and reversible input. In Theorem 3.3.1, the channel capacity is given in the form with a constraint on the set of input distributions $I(X_2; Y_2|X_1) - I(X_2; Y_1|X_1) \geq 0$ which is called the *two-stage coding condition*. This condition arises essentially from the two-stage coding structure of the problem and the extra requirement that the decoder should be able to reconstruct the signal generated in later stages. In addition to the rate constraint on the message, the two-stage coding condition can be interpreted as a necessary and sufficient condition for reliable transmission of the channel input signal over the channel in second stage transmission.

By studying the multi-stage coding problem in this chapter, we can straightforwardly connect the two-stage coding condition to the result for $K = 2$ in this chapter. That is, the two-stage coding condition can be seen as a degenerate rate constraint derived from the underlying rate constraint of the message M_2 in the multi-stage coding setting. In retrospect, it is therefore not surprising that this condition is present in the capacity expression in Theorem 3.3.1, and in fact it can be active when computing the capacity as shown in an example in Section 3.3.

4.4 Bounds to Capacity Region for the Case $K \geq 3$

We now consider an extension to the multi-stage setting in which we characterize inner and outer bounds to the capacity region. The inner bound can be derived as a straightforward extension of the two-stage setting, while we find it challenging to generalize the converse proof for $K \geq 3$. Based on the problem formulation, the main difficulty in deriving tight bounds lies in the structure of the coding which prevents us from having the desired Markov chains to account for the dependency between the channel outputs and the channel states.

4.4.1 Main Results

Theorem 4.4.1 (Inner bound to the capacity region)**.** *A rate tuple* $(R_1, \ldots, R_K) \in \mathbb{R}_+^K$ *is achievable if*

$$\sum_{i=j}^{K} R_i \leq \sum_{i=j}^{K} I(X_i; Y_K | X_1^{i-1}) - I(X_i; Y_1^{i-1} | X_1^{i-1}), \tag{4.6}$$

for all $j = 1, \ldots, K$*, and some joint distributions of the form*

$$\prod_{i=1}^{K} \big(P_{X_i|X_1,\ldots,X_{i-1},Y_1,\ldots,Y_{i-1}}(x_i|x_1,\ldots,x_{i-1},y_1,\ldots,y_{i-1}) P_{Y_i|X_i,Y_{i-1}}(y_i|x_i,y_{i-1}) \big). \tag{4.7}$$

Proof. The sketch of the proof is given in Appendix 4.A. □

Theorem 4.4.2 (Outer bound to the capacity region)**.** *For any achievable rate tuple* $(R_1, \ldots, R_K)$*,* $K \geq 3$*, it follows that*

$$\sum_{i=j}^{K} R_i \leq I(X_j^K, Y_{K-1}; Y_K | X_1^{j-1}, Y_1^{j-2}) - \sum_{i=j}^{K} I(X_i; Y_{i-1} | X_1^{i-1}, Y_1^{i-2}), \tag{4.8}$$

for all $j = 1, \ldots, K-1$*, and*

$$R_K \leq I(X_K; Y_K | X_1^{K-1}, Y_1^{K-2}) - I(X_K; Y_{K-1} | X_1^{K-1}, Y_1^{K-2}), \tag{4.9}$$

for some joint distributions of the same form as in (4.7)*.*

Proof. The proof is given in Appendix 4.B. It essentially involve basic properties of entropy function, Fano's inequality, and inspired applications of the Csiszár's sum identity (Lemma 2.5). □

Remark 4.2 (Conditionally independent channel)**.** In general we do not know how good the inner and outer bounds are for the case of $K \geq 3$. However, for a degenerate case where in each stage the channel output is conditionally independent

of the state given the input, the bounds will coincide and provide the result of the point-to-point case as expected from inspection, i.e., the optimal input distribution turns out to be the one which is independent of all previous states and inputs.

Remark 4.3 (Deterministic channel[1])**.** For the deterministic channel case where $Y_i = f(X_i, Y_{i-1}), i = 1, \ldots, K$, the inner and outer bounds coincide and give the capacity region as shown in Corollary 4.4.1. In this case, the problem of decoding both messages and channel inputs reduces to that of decoding messages only because we can recover all channel inputs from the messages, the deterministic encoders, and the channel functions. For the two-stage case, this corresponds to a class of multiple access channel (MAC) with common message and cribbing encoder [WvdM85].

Corollary 4.4.1 (Capacity region for the deterministic channel)**.** *The capacity region of the deterministic multi-stage channel with rewrite option and reversible input is given by a set of all rate tuples* $(R_1, \ldots, R_K) \in \mathbb{R}_+^K$ *satisfying*

$$\sum_{i=j}^{K} R_i \leq H(Y_K | X_1^{j-1}), \tag{4.10}$$

for all $j = 1, \ldots, K$*, and some joint distributions of the form*

$$P_{X_1,\ldots,X_K}(x_1, \ldots, x_K) \prod_{i=1}^{K} 1_{\{Y_i = f(x_i, y_{i-1})\}}(y_i).$$

Proof. The achievability proof follows directly from the result in Theorem 4.4.1 with specialization to the deterministic channel, and is given below. For the deterministic channel case where $Y_i = f(X_i, Y_{i-1}), i = 1, \ldots, K$, we have

$$\begin{aligned}\sum_{i=j}^{K} I(X_i; Y_K | X_1^{i-1}) - I(X_i; Y_1^{i-1} | X_1^{i-1}) &= I(X_j^K; Y_K | X_1^{j-1}) \\ &\quad - \sum_{i=j}^{K} I(X_i; Y_1^{i-1} | X_1^{i-1}) \overset{(a)}{=} H(Y_K | X_1^{j-1}),\end{aligned}$$

where (a) follows from the fact that $Y_1 = f(X_1)$, and $Y_i = f(X_i, Y_{i-1}), i = 2, \ldots, K$.

The converse proof is given as follows. For any achievable tuple $(R_1, \ldots, R_K)$, we have that for all $j = 1, \ldots, K$,

$$\begin{aligned}&n \sum_{i=j}^{K} R_i - n\epsilon_n^{(j)} \\ &\overset{(a)}{\leq} H(M_j^K | M_1^{j-1}, \boldsymbol{X}_1^{j-1}, \boldsymbol{Y}_1^{j-2}) - H(M_j^K | \boldsymbol{Y}_K, M_1^{j-1}, \boldsymbol{X}_1^{j-1}, \boldsymbol{Y}_1^{j-2})\end{aligned}$$

[1] **Acknowledgment**: The author wishes to thank Lele Wang @UCSD for helpful suggestions on the deterministic channel case.

$$
\begin{aligned}
&= I(M_j^K; \boldsymbol{Y}_K | M_1^{j-1}, \boldsymbol{X}_1^{j-1}, \boldsymbol{Y}_1^{j-2}) \\
&= H(\boldsymbol{Y}_K | M_1^{j-1}, \boldsymbol{X}_1^{j-1}, \boldsymbol{Y}_1^{j-2}) - H(\boldsymbol{Y}_K | M_1^K, \boldsymbol{X}_1^{j-1}, \boldsymbol{Y}_1^{j-2}) \\
&\overset{(b)}{=} H(\boldsymbol{Y}_K | M_1^{j-1}, \boldsymbol{X}_1^{j-1}, \boldsymbol{Y}_1^{j-2}) \\
&= \sum_{l=1}^{n} H(Y_{K,l} | Y_K^{l-1}, M_1^{j-1}, \boldsymbol{X}_1^{j-1}, \boldsymbol{Y}_1^{j-2}) \\
&\leq \sum_{l=1}^{n} H(Y_{K,l} | (X_1^{j-1})_l),
\end{aligned}
$$

where (a) follows from the independency $M_j^K \perp W_j$, where we define the variable $W_j \triangleq (M_1^{j-1}, \boldsymbol{X}_1^{j-1}, \boldsymbol{Y}_1^{j-2})$, and from Fano's inequality, i.e., $H(M_j^K | \boldsymbol{Y}_K, W_j) \leq n\epsilon_n^{(j)}$, $\lim_{n\to\infty} \epsilon_n^{(j)} = 0$, and (b) follows from the deterministic encoding and deterministic channel functions. □

4.5 Conclusion

In this chapter, we studied an extension of the problem of coding with action-dependent state and reversible input to the multi-stage case. For the two-stage case, the input X_1^n plays a role of action sequence A^n, and the output Y_1^n corresponds to the action-dependent state. The capacity region for the two-stage case is found, and inner and outer bounds to the capacity region are given in the case of three stages or more. We were not able to establish bounds that match in general for the case $K \geq 3$ due to the structure of our problem which prevents us from having the desired Markov chains to account for the dependency of the channel states on the channel outputs. This can be seen from the proof of and corresponding mutual information terms in Theorem 4.4.2. For example, when compared with Theorem 4.4.1, there are additional variables in (4.8) such as Y_{K-1} in the first term, and Y_1^{i-2} in the conditioning of the second term. These variables appear due to the dependency between the channel state and the channel output. For the two-stage problem, it turned out that the capacity result helps us to establish a connection to the previous result on the two-stage coding condition. It shows that the two-stage coding condition is simply the degenerate rate constraint derived from the underlying rate constraint on the message in the second stage. The inner and outer bounds derived in this chapter also suggest the presence of similar two-stage coding conditions in the multi-stage setting, i.e., the rate constraint on the message in later stages implicitly includes the condition for reliable transmission of the channel input sequences over the channels in later stages.

Appendix for Chapter 4

4.A Sketch of the Proof of Theorem 4.4.1

Fix $\prod_{i=1}^{K} P_{X_i|X_1,\ldots,X_{i-1},Y_1,\ldots,Y_{i-1}}$. The codebook generation consists of randomly generating 2^{nR_1} codewords $X_1^n(m_1)$ each i.i.d. according to $\prod_{l=1}^{n} P_{X_1}(x_{1,l}(m_1))$, and for each $X_1^n(m_1)$, generating $2^{n(R_2+R_2')}$ codewords $X_2^n(m_1, m_2, m_2')$ each i.i.d. according to $\prod_{l=1}^{n} P_{X_2|X_1}(x_{2,l}|x_{1,l}(m_1))$. Next, for each pair of (X_1^n, X_2^n), generate $2^{n(R_3+R_3')}$ codewords $X_3^n(m_1, m_2, m_2', m_3, m_3')$ each i.i.d. according to the distribution $\prod_{l=1}^{n} P_{X_3|X_1,X_2}(x_{3,l}|x_{1,l}, x_{2,l})$. We continue this process until we have generated $2^{n(R_K+R_K')}$ codewords $X_K^n(m_1, m_2, m_2', \ldots, m_K, m_K')$ for all $(X_1^n, \ldots, X_{K-1}^n)$.

Given the messages $(m_1, \ldots, m_i)$ and the previous inputs and outputs $(X_1^n, \ldots, X_{i-1}^n)$, $(Y_1^n, \ldots, Y_{i-1}^n)$, $i = 1, \ldots, K$, the encoder in stage i looks for and transmits the codeword X_i^n that is jointly typical with $(X_1^n, \ldots, X_{i-1}^n, Y_1^n, \ldots, Y_{i-1}^n)$. From the covering lemma (Lemma 2.6), we ensure the vanishing probability of encoding error as $n \to \infty$ if $R_i' > I(X_i; Y_1, \ldots, Y_{i-1}|X_1, \ldots, X_{i-1})$ for all $i = 2, \ldots, K$.

Given the final-stage output Y_K^n, the decoder looks for the unique $(X_1^n, \ldots, X_K^n)$ that is jointly typical with it. From the packing lemma (Lemma 2.7), we ensure the vanishing probability of decoding error as $n \to \infty$ if $R_1 + R_2 + R_2' + \ldots + R_K + R_K' < I(X_1, \ldots, X_K; Y_K)$, and $\sum_{i=j}^{K} R_i + R_i' < I(X_j, \ldots, X_K; Y_K|X_1, \ldots, X_{j-1})$ for all $j = 2, \ldots, K$.

By these results together with a random coding argument, we conclude that a rate tuple $(R_1, \ldots, R_K)$ is achievable if for all $j = 1, \ldots, K$, it satisfies $\sum_{i=j}^{K} R_i < \sum_{i=j}^{K} I(X_i; Y_K|X_1^{i-1}) - I(X_i; Y_1^{i-1}|X_1^{i-1})$.

4.B Proof of Theorem 4.4.2

For any achievable rate tuple $(R_1, \ldots, R_K)$, we have that for all $j = 1, \ldots, K$,

$$
\begin{aligned}
& n\sum_{i=j}^{K} R_i - n\epsilon_n^{(j)} \\
& \overset{(a)}{\leq} H(M_j^K|W_j) - H(M_j^K, \boldsymbol{X}_j^K|\boldsymbol{Y}_K, W_j) \\
& = H(M_j^K, \boldsymbol{Y}_{j-1}^{K-1}|W_j) - H(\boldsymbol{Y}_{j-1}^{K-1}|M_j^K, W_j) - H(M_j^K, \boldsymbol{X}_j^K|\boldsymbol{Y}_K, W_j) \\
& \overset{(b)}{=} H(M_j^K, \boldsymbol{Y}_{j-1}^{K-1}, \boldsymbol{X}_j^K|W_j) - H(\boldsymbol{Y}_{j-1}^{K-1}|M_j^K, W_j) - H(M_j^K, \boldsymbol{X}_j^K|\boldsymbol{Y}_K, W_j) \\
& \overset{(c)}{\leq} \sum_{i=j}^{K} H(M_i, \boldsymbol{Y}_{i-1}, \boldsymbol{X}_i|W_i) - H(\boldsymbol{Y}_{i-1}|W_i) - H(M_i, \boldsymbol{X}_i|\boldsymbol{Y}_K, W_i) \\
& = \sum_{i=j}^{K} I(M_i, \boldsymbol{X}_i; \boldsymbol{Y}_K|W_i) - I(M_i, \boldsymbol{X}_i; \boldsymbol{Y}_{i-1}|W_i),
\end{aligned}
$$

where (a) follows from the independency $M_j^K \perp W_j$ with $W_j \triangleq (M_1^{j-1}, \boldsymbol{X}_1^{j-1}, \boldsymbol{Y}_1^{j-2})$, and from Fano's inequality, i.e., $H(M_j^K, \boldsymbol{X}_j^K | \boldsymbol{Y}_K, W_j) \leq n\epsilon_n^{(j)}$, $\lim_{n\to\infty} \epsilon_n^{(j)} = 0$, (b) follows from the deterministic encoding function $\boldsymbol{X}_i = f_i^{(n)}(M_1^i, \boldsymbol{X}_1^{i-1}, \boldsymbol{Y}_1^{i-1})$, $i = j, \ldots, K$, (c) holds since $\boldsymbol{X}_i = f_i^{(n)}(M_1^i, \boldsymbol{X}_1^{i-1}, \boldsymbol{Y}_1^{i-1})$, and $\boldsymbol{Y}_{i-1} - W_i - M_i^K$ forms a Markov chain with $W_i = (M_1^{i-1}, \boldsymbol{X}_1^{i-1}, \boldsymbol{Y}_1^{i-2})$, and conditioning reduces entropy. Note that the term W_i is introduced for the sake of readability and it essentially captures the previous variables known before the encoding stage i.

Continuing the chain of inequalities, we get

$$
\begin{aligned}
& n\sum_{i=j}^{K} R_i - n\epsilon_n^{(j)} \\
& \leq \sum_{l=1}^{n}\sum_{i=j}^{K} I(M_i, \boldsymbol{X}_i; Y_{K,l} | W_i, Y_{K,1}^{l-1}) - I(M_i, \boldsymbol{X}_i; Y_{i-1,l} | W_i, Y_{i-1,l+1}^{n}) \\
& \overset{(d)}{=} \sum_{l=1}^{n}\sum_{i=j}^{K} I(M_i, \boldsymbol{X}_i, Y_{i-1,l+1}^{n}; Y_{K,l} | W_i, Y_{K,1}^{l-1}) - I(M_i, \boldsymbol{X}_i, Y_{K,1}^{l-1}; Y_{i-1,l} | W_i, Y_{i-1,l+1}^{n}),
\end{aligned}
$$

where (d) follows from Csiszár's sum identity (Lemma 2.5), i.e., for $i = j, \ldots, K$, $\sum_{l=1}^{n} I(Y_{i-1,l+1}^{n}; Y_{K,l} | W_i, M_i, \boldsymbol{X}_i, Y_{K,1}^{l-1}) - I(Y_{K,1}^{l-1}; Y_{i-1,l} | W_i, M_i, \boldsymbol{X}_i, Y_{i-1,l+1}^{n}) = 0$.

Next, we replace W_i by $(M_1^{i-1}, \boldsymbol{X}_1^{i-1}, \boldsymbol{Y}_1^{i-2})$ again and continue the chain of inequalities.

$$
\begin{aligned}
& n\sum_{i=j}^{K} R_i - n\epsilon_n^{(j)} \\
& \overset{(e)}{\leq} \sum_{l=1}^{n} \Big(I(M_j^K, \boldsymbol{X}_j^K, \boldsymbol{Y}_{j-1}^{K-2}, Y_{K-1,l+1}^{n}; Y_{K,l} | M_1^{j-1}, \boldsymbol{X}_1^{j-1}, \boldsymbol{Y}_1^{j-2}, Y_{K,1}^{l-1}) \\
& \qquad - \sum_{i=j}^{K} I(M_i, \boldsymbol{X}_i, Y_{K,1}^{l-1}; Y_{i-1,l} | M_1^{i-1}, \boldsymbol{X}_1^{i-1}, \boldsymbol{Y}_1^{i-2}, Y_{i-1,l+1}^{n}) \Big) \\
& \overset{(f)}{\leq} \sum_{l=1}^{n} \Big(I(Z_K, (X_j^K)_l, (Y_{j-1}^{K-2})_l; Y_{K,l} | (X_1^{j-1})_l, (Y_1^{j-2})_l) \\
& \qquad - \sum_{i=j}^{K} I(Z_i, X_{i,l}; Y_{i-1,l} | (X_1^{i-1})_l, (Y_1^{i-2})_l) \Big),
\end{aligned}
$$

where (e) holds using a telescoping series to bound the sum from $i = j$ to $i = K$ of the first mutual information term, (f) follows since conditioning reduces entropy, and from the Markov chain $Y_{i-1,l} - ((X_1^{i-1})_l, (Y_1^{i-2})_l) - (M_1^{i-1}, (X_1^{i-1})^{n\setminus l}, (Y_1^{i-2})^{n\setminus l}, Y_{i-1,l+1}^{n})$ and by defining $Z_i \triangleq (M_1^i, (X_1^i)^{n\setminus l}, (Y_1^{i-2})^{n\setminus l}, Y_{i-1,l+1}^{n}, Y_{K,1}^{l-1})$, $i = 2, \ldots, K$.

For $j = 1, \ldots, K-1$, we continue the chain of inequalities.

$$n\sum_{i=j}^{K} R_i - n\epsilon_n^{(j)} \overset{(\sharp)}{\leq} \sum_{l=1}^{n} \big(I((X_j^K)_l, Y_{K-1,l}; Y_{K,l} | (X_1^{j-1})_l, (Y_1^{j-2})_l) - \sum_{i=j}^{K} I(X_{i,l}; Y_{i-1,l} | (X_1^{i-1})_l, (Y_1^{i-2})_l) \big),$$

where $(\sharp)$ holds since conditioning reduces entropy and we have the Markov chain $Y_{K,l} - ((X_1^K)_l, Y_{K-1,l}) - ((Y_1^{K-2})_l, Z_K)$. As for $j = K$, we have

$$\begin{aligned} nR_K - n\epsilon_n^{(K)} &\leq \sum_{l=1}^{n} I(Z_K, X_{K,l}; Y_{K,l} | (X_1^{K-1})_l, (Y_1^{K-2})_l) \\ &\quad - I(Z_K, X_{K,l}; Y_{K-1,l} | (X_1^{K-1})_l, (Y_1^{K-2})_l) \\ &= \sum_{l=1}^{n} I(Z_K, X_{K,l}; Y_{K,l} | (X_1^{K-1})_l, (Y_1^{K-2})_l, Y_{K-1,l}) \\ &\quad - I(Z_K, X_{K,l}; Y_{K-1,l} | (X_1^{K-1})_l, (Y_1^{K-2})_l, Y_{K,l}) \\ &\overset{(*)}{\leq} \sum_{l=1}^{n} I(X_{K,l}; Y_{K,l} | (X_1^{K-1})_l, (Y_1^{K-2})_l, Y_{K-1,l}) \\ &\quad - I(X_{K,l}; Y_{K-1,l} | (X_1^{K-1})_l, (Y_1^{K-2})_l, Y_{K,l}) \\ &= \sum_{l=1}^{n} I(X_{K,l}; Y_{K,l} | (X_1^{K-1})_l, (Y_1^{K-2})_l) \\ &\quad - I(X_{K,l}; Y_{K-1,l} | (X_1^{K-1})_l, (Y_1^{K-2})_l), \end{aligned}$$

where $(*)$ follows from the Markov chain $Y_{K,l} - ((X_1^K)_l, (Y_1^{K-1})_l) - Z_K$. Note that when $K = 2$, the above steps for $j = 1$ will not result in the desired expression. To fix this, we can use similar steps as in the case $j = K$ for both $j = 1, 2$ instead.

To this end, we introduce a random variable Q distributed uniformly over $\{1, \ldots, n\}$ and independent of all others. We have $P_{Y_{i,Q}|X_{i,Q},Y_{i-1,Q}} = P_{Y_i|X_i,Y_{i-1}}$ for all Q, and $i = 1, \ldots, K$, and we can identify $X_{i,Q} \triangleq X_i$ and $Y_{i,Q} \triangleq Y_i$. Since $Y_i - (X_i, Y_{i-1}) - (X_1^{i-1}, Y_1^{i-2}, Q)$ forms a Markov chain for all $i = 1, \ldots, K$,

$$\sum_{i=j}^{K} R_i - \epsilon_n^{(j)} \leq I(X_j^K, Y_{K-1}; Y_K | X_1^{j-1}, Y_1^{j-2}) - \sum_{i=j}^{K} I(X_i; Y_{i-1} | X_1^{i-1}, Y_1^{i-2}), \text{ for all } j = 1, \ldots, K-1, \text{ and}$$

$$R_K - \epsilon_n^{(K)} \leq I(X_K; Y_K | X_1^{K-1}, Y_1^{K-2}) - I(X_K; Y_{K-1} | X_1^{K-1}, Y_1^{K-2}).$$

The proof is concluded by letting $n \to \infty$.

Chapter 5

Secure Source Coding With Action-dependent Side Information

So far we have studied problems of source and channel coding with action-dependent side information (SI) and additional reconstruction requirements in both basic and multi-stage settings. As discussed in Chapter 2, information privacy is another important requirement that is of significant interest in today's and future's networks. In this chapter, we study the information security/privacy constraint in source coding problems. In particular, we consider problems of secure lossy source coding with side information in the presence of a passive eavesdropper who has access to the source description. The encoder wishes to compress the source sequence in order to satisfy a distortion criterion at the decoder, while revealing only limited knowledge about the source to the eavesdropper. The side information available to the encoder, the legitimate decoder, or the eavesdropper can be influenced by a cost-constrained action sequence.

Three different settings are studied. In the first two settings, we are interested in the influence of the action sequence on the rate-distortion-leakage tradeoff where the action is taken either by the decoder or by the encoder to influence side information at the decoder and eavesdropper. Next, we consider a setting where common action-dependent side information is available securely to both encoder and decoder, and thus can be used for secret key generation. We characterize the optimal rate-distortion-cost-leakage region or corresponding inner and outer bounds for a discrete memoryless source for above settings. The results are useful in characterizing fundamental limits for example in secure sensor networking.

5.1 Introduction

As discussed in Section 2.3.3, Chapter 2, the problem of physical layer security currently arises as one of the major and interesting problems in designing the corresponding communication protocols. We not only require the communication to be reliable, but also we need to guarantee a certain level of security or privacy of

our information.

In this chapter, we study a *secure source coding* problem which is essentially the source coding problem with additional secrecy constraint. An encoder wishes to compress the source sequence X^n into the source description W and transmit it over a noiseless rate-limited link to a decoder. An eavesdropper is assumed to have access to the source description and correlated side information Z^n, and therefore can learn about the source sequence. The information leakage at the eavesdropper is defined as a normalized mutual information $\frac{1}{n}I(X^n; W, Z^n)$. The objective of the system design is to reveal as little knowledge about the source as possible to the eavesdropper, and at the same time to satisfy the distortion constraint at the intended decoder. These are clearly conflicting objectives which lead to an interesting tradeoff between rate, distortion, and leakage levels.

More specifically, in this chapter, we apply the notion of action-dependent side information studied in Chapter 3 to the *secure source coding* problem, and characterize the fundamental tradeoff between the minimum rate needed to describe the source, the resulting distortion at the legitimate decoder, the cost associated with the taken actions, and the leakage rate at the eavesdropper, in the form of the *rate-distortion-cost-leakage region*. In secure lossy source coding, secrecy is characterized by amount of information leaked through the source description and the eavesdropper's side information. By modelling the system with action-dependent side information, we essentially allow another degree of freedom to control the side information available at each node in the network, which in turn affects the amount of information needed to describe the source, leading to a more flexible and efficient approach to handle the secrecy constrained system. Relevant applications of this work include for example secure transmission for sensor networking, where a sensor node may take actions with some costs to enhance the side information at the intended receiver (or worsen that at the eavesdropper), with the goal of achieving higher security and reconstruction quality.

5.1.1 Related Work

Physical layer security was introduced based on the fact that the signals available at the legitimate receiver and the eavesdropper usually possess different characteristics. It has gained significant attention due to its advantages in some applications, for example, those where complex cryptographic protocols cannot be implemented. Information theoretic study of physical layer security was pioneered by Wyner in [Wyn75], which introduces and solves the problem of coding for the Wiretap channel. Wyner has shown that perfect secure communication is possible when the channel from the sender to the legitimate receiver is "stronger" than the channel to the eavesdropper. Later generalizations include the case of a broadcast channel with confidential messages by Csiszár and Körner [CK78]. Several extensions to other multiuser channels including secure coding for multiple access channels, channels with states, etc. are considered in [LPSS08]. An idea of physical layer security from the source coding perspective has also been studied, i.e.,

source coding with side information subject to an additional secrecy constraint. As before, security/privacy of the source relies mainly on the different characteristics of signals (side information) which are available at the legitimate receiver and the eavesdropper. Secure lossless distributed source coding was studied by Prabhakaran and Ramchandran [PR07], Gündüz et al. [GEP08], and Tandon et al. [TUR13], and an extension to the lossy case was considered by Villard and Piantanida [VP13] and Ekrem and Ulukus [EU11]. The closely related work which characterized the tradeoff between amplifying information about one source and masking another was recently studied in [Cou12]. Another line of work considers *explicit* secret key sharing in the system model that is based on the Shannon cipher system [Sha49, Yam97, Mau93, AC93, Mer08, Cuf10, BB11]. We note that we do not assume any explicit secret key sharing in this chapter. Nevertheless, in some scenarios, we may be able to exploit some common randomness for secrecy by *implicitly* generating a secret key using the common side information (Section 5.4).

5.1.2 Overview of Problem Settings and Organization

In this chapter, we study several relevant aspects of action-dependent side information and fundamental tradeoffs in the following three settings.

- Section 5.2 considers secure source coding problem with action-dependent side information where *action is taken at the decoder*, as shown in Fig. 5.1. The action sequence is taken at the decoder based on the source description, and it influences side information at the decoder and eavesdropper through the side information channel $P_{Y,Z|X,A}$. We provide a complete characterization of the rate-distortion-cost-leakage region for the discrete memoryless source. Extension to include a private link between the encoder and decoder is also studied. A binary example of taking actions to enhance or to suppress the side information at the legitimate decoder is considered at the end of the section.

- Section 5.3 considers a variant of the problem in Section 5.2, where *action is taken at the encoder*, as shown in Fig. 5.7. The action is taken at the encoder based on the source sequence X^n, i.e., $A^n \sim P_{A^n|X^n}$ to influence side information at the decoder and eavesdropper. We give inner and outer bounds for the rate-distortion-cost-leakage region. In the special case under lossless reconstruction and $Z = \emptyset$, the complete rate-distortion-cost-leakage region is given. A lossless example illustrating the leakage-cost tradeoff for different rates is provided.

- Section 5.4 considers the problem of secure source coding with common action-dependent side information, as depicted Fig. 5.9. The action sequence is assumed to be taken at the decoder based on the source description to influence the generation of common side information at the encoder and decoder. Side information at the eavesdropper is assumed to be a degraded version of the

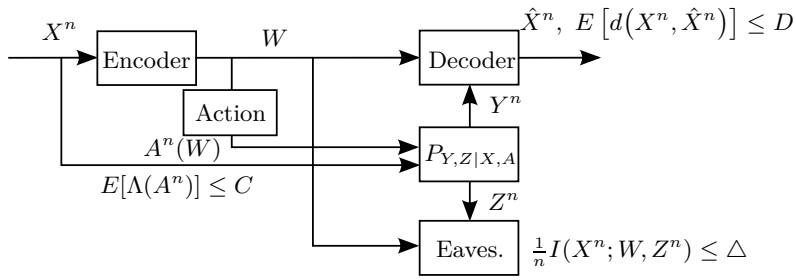

Figure 5.1: Secure source coding with action-dependent side information where action is taken at the decoder.

common side information. We are interested in how the common side information can support the transmission, and additionally enable the secret key generation. We give inner and outer bounds to the rate-distortion-cost-leakage region and show that they coincide under a special case of logarithmic loss distortion. This setting is inspired by the problem of source coding with two-sided action-dependent side information studied in Chapter 3. The secret key generation concept is also studied in [CK13a] and [KCO+13b] which will be presented in Chapter 6.

5.2 Secure Source Coding With Action-dependent SI: Action Taken at Decoder

In this section, we consider the problem setting where an action sequence is taken at the decoder based on the source description. It then influences the side information at the decoder and eavesdropper through the side information channel $P_{Y,Z|X,A}$. We characterize the complete rate-distortion-cost-leakage region for the discrete memoryless source.

5.2.1 Problem Formulation

Consider the problem in Fig. 5.1. Let n denote the sequence length and $\mathcal{X}, \mathcal{A}, \mathcal{Y}$, and $\mathcal{Z}$ be finite source, action, and side information alphabets. Let X^n be the source sequence which is i.i.d. according to P_X. The action sequence A^n is generated based on the source description which is an output of the encoder mapping. Side information (Y^n, Z^n) available to the decoder and the eavesdropper are the outputs of the discrete memoryless channel $P_{Y,Z|X,A}$.

Definition 5.1. A $(|\mathcal{W}^{(n)}|, n)$-code for secure source coding with action-dependent side information where *action is taken at decoder* consists of

- a stochastic encoder $F^{(n)}$ which takes X^n as input and generates $W \in \mathcal{W}^{(n)}$ according to a conditional PMF $p(w|x^n)$,
- a deterministic action encoder $f_a^{(n)} : \mathcal{W}^{(n)} \to \mathcal{A}^n$, and
- a deterministic decoder $g^{(n)} : \mathcal{W}^{(n)} \times \mathcal{Y}^n \to \hat{\mathcal{X}}^n$,

where $\mathcal{W}^{(n)}$ is a finite set.

The *information leakage* at the eavesdropper who has access to W and Z^n is measured by the normalized mutual information $\frac{1}{n}I(X^n; W, Z^n)$.

Let $d : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty)$ and $\Lambda : \mathcal{A} \to [0, \infty)$ be the single-letter distortion and cost measures. The *distortion* between the source and its reconstruction at the decoder, and the *cost* of action are defined as

$$d^{(n)}\big(X^n, \hat{X}^n\big) \triangleq \frac{1}{n}\sum_{i=1}^{n} d(X_i, \hat{X}_i),$$
$$\Lambda^{(n)}(A^n) \triangleq \frac{1}{n}\sum_{i=1}^{n} \Lambda(A_i),$$

where $d^{(n)}(\cdot)$ and $\Lambda^{(n)}(\cdot)$ are distortion and cost functions, respectively.

Definition 5.2. A rate-distortion-cost-leakage tuple $(R, D, C, \triangle)$ is said to be *achievable* if for any $\delta > 0$ and all sufficiently large n there exists a $(|\mathcal{W}^{(n)}|, n)$ code such that

$$\frac{1}{n}\log\big|\mathcal{W}^{(n)}\big| \le R + \delta,$$
$$E\Big[d^{(n)}\big(X^n, g^{(n)}(W, Y^n)\big)\Big] \le D + \delta,$$
$$E\Big[\Lambda^{(n)}(A^n)\Big] \le C + \delta,$$
$$\text{and} \quad \frac{1}{n}I(X^n; W, Z^n) \le \triangle + \delta.$$

The *rate-distortion-cost-leakage* region $\mathcal{R}$ is the set of all achievable tuples.

5.2.2 Main Result

Theorem 5.2.1. *The rate-distortion-cost-leakage region $\mathcal{R}$ is the set of all tuples $(R, D, C, \triangle) \in \mathbb{R}_+^4$ that satisfy*

$$R \ge I(X; A) + I(X; V|Y, A), \tag{5.1a}$$
$$D \ge E\big[d\big(X, \tilde{g}(V, Y)\big)\big], \tag{5.1b}$$
$$C \ge E[\Lambda(A)], \tag{5.1c}$$
$$\triangle \ge I(X; V, Y, A) - [I(X; Y|U, A) - I(X; Z|U, A)], \tag{5.1d}$$

for some joint distributions of the form

$$P_X(x)P_{A|X}(a|x)P_{V|X,A}(v|x,a)P_{U|V}(u|v)P_{Y,Z|X,A}(y,z|x,a)$$

with $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{A}| + 2,\ |\mathcal{V}| \leq |\mathcal{U}|(|\mathcal{X}||\mathcal{A}| + 1)$*, and a function* $\tilde{g} : \mathcal{V} \times \mathcal{Y} \to \hat{\mathcal{X}}$.

Proof. The proof of Theorem 5.2.1 is provided in Appendix 5.A. □

Remark 5.1 (Layered coding)**.** In the achievable scheme for Theorem 5.2.1, after communicating the action sequence, layered coding (superposition type coding) is used, i.e., we utilize both codewords U^n and V^n to carry the description of X^n. This layering is required to adapt to the eavesdropper in this general setting.

Remark 5.2 (Interpretation of leakage rate constraint)**.** The right-hand side of the leakage rate constraint (5.1d) consists of contributions from the remaining uncertainty at the decoder $H(X|V,Y,A)$ which could be exploited to reduce the leakage from the maximum level $H(X)$, and the extra gain due to different side information available at the decoder and the eavesdropper $I(X;Y|U,A) - I(X;Z|U,A)$ which can be tuned by the choice of codeword U and action A.

Remark 5.3 (Stochastic action encoder)**.** In this section, we restrict the action encoder to be a deterministic mapping in the problem setting. This might not be the best choice in the secure compression scenario since the eavesdropper who observes the source description W knows exactly which action sequence is chosen. It would be interesting to see if the rate-distortion-cost-leakage region can be improved if we allow the use of a stochastic action encoder. As a simple illustrating example showing that a stochastic action encoder can enlarge the rate-distortion-cost-leakage region, we assume that the action sequence is generated at the decoder (co-located) and thus known at the decoder. Also, the side information at the decoder and the eavesdropper are assumed to be equal, i.e., $Y = Z = X \oplus A$. Since the decoder can recover the source from its side information and its knowledge of the action, the encoder does not need to send anything over the rate-limited link. As for the leakage rate, in the deterministic action encoder case, the eavesdropper knows the action sequence and can therefore recover the exact source sequence. This results in the maximum leakage rate $H(X)$. On the other hand, when the action encoder is stochastic, i.e., a stochastic mapping from the source description W to A^n, we may set A^n to be i.i.d. where $A \sim \text{Bernoulli}(1/2)$ independent of X. This makes Z independent of X, and therefore zero leakage rate is achieved.

5.2.3 Extension

In this section, we consider an extension in which the communication from an encoder to a decoder consists of public and private rate-limited links, as shown in Fig. 5.2. In the scenario that information can be leaked to the eavesdropper, we are interested in how the encoder sends the source description over private and

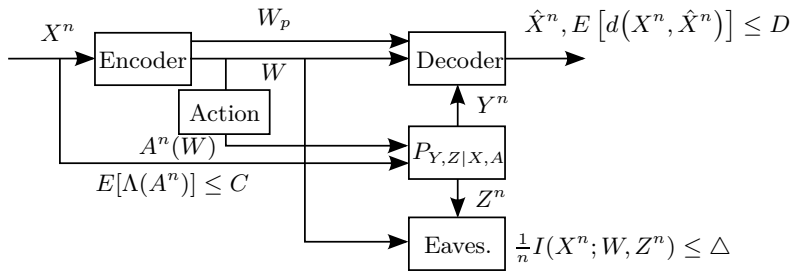

Figure 5.2: Secure source coding with action-dependent side information and private transmission.

public links where it is possible to influence the side information at the decoder and the eavesdropper via an action sequence over the public link. We characterize the rate-distortion-cost-leakage region which provides the optimal balance of using the public link, i.e., tradeoff between the gain obtained by influencing the side information via action and the amount of leaked information over the public link. The result gives insight on how much the rate on the private link is necessary if certain performance (distortion) and privacy are needed.

The problem formulation of the extension follows similarly as in Section 5.2.1, except that we also have a stochastic *private* encoder $F_p^{(n)}$ which takes X^n as input and generates $W_p \in \mathcal{W}_p^{(n)}$ according to a conditional PMF $p(w_p|x^n)$, and the decoder becomes $g^{(n)} : \mathcal{W}^{(n)} \times \mathcal{W}_p^{(n)} \times \mathcal{Y}^n \to \hat{\mathcal{X}}^n$. Let $\mathcal{R}_{pri}$ be the rate-distortion-cost-leakage region for this extended setting. We provide a complete characterization of $\mathcal{R}_{pri}$ below.

Corollary 5.2.1 (Private and public links)**.** *The rate-distortion-cost-leakage region $\mathcal{R}_{pri}$ is the set of all tuples $(R, R_p, D, C, \triangle) \in \mathbb{R}_+^5$ that satisfy*

$$R \geq I(X; A) + I(X; V|Y, A), \tag{5.2a}$$
$$R_p \geq I(X; W|U, V, Y, A), \tag{5.2b}$$
$$D \geq E\big[d\big(X, \tilde{g}(V, W, Y)\big)\big], \tag{5.2c}$$
$$C \geq E[\Lambda(A)], \tag{5.2d}$$
$$\triangle \geq I(X; V, Y, A) - [I(X; Y|U, A) - I(X; Z|U, A)], \tag{5.2e}$$

for some joint distributions of the form

$$P_X(x)P_{A|X}(a|x)P_{V,W|X,A}(v, w|x, a)P_{U|V}(u|v)P_{Y,Z|X,A}(y, z|x, a)$$

with $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{A}| + 3$, $|\mathcal{V}| \leq |\mathcal{U}|(|\mathcal{X}||\mathcal{A}| + 2)$, $|\mathcal{W}| \leq |\mathcal{U}||\mathcal{V}|(|\mathcal{X}||\mathcal{A}| + 1)$, and a function $\tilde{g} : \mathcal{V} \times \mathcal{W} \times \mathcal{Y} \to \hat{\mathcal{X}}$.

Proof. The corollary follows from Theorem 5.2.1 with an inclusion of the rate constraint on the private link. The private link rate constraint can be proved using

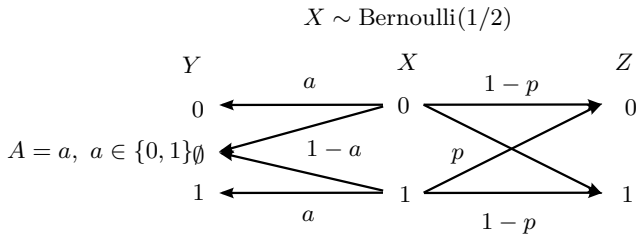

Figure 5.3: Action-dependent side information distribution.

the Wyner-Ziv type proof where U^n, V^n, A^n, and Y^n are considered as side information at the decoder. That is, for achievability, superposition codeword W^n is used for communicating the private source description after A^n, U^n, and V^n are communicated to the decoder. The converse proof follows similarly as that of Theorem 5.2.1 with auxiliary random variables defined as $U_i \triangleq (W, A^{n\setminus i}, Y_{i+1}^n, Z^{i-1})$, $V_i \triangleq (W, X^{i-1}, A^{n\setminus i}, Y^{n\setminus i}, Z^{i-1})$, and $W_i \triangleq W_p$, for $i = 1, \ldots, n$ satisfying the Markov chains $U_i - V_i - (W_i, X_i, A_i)$ and $(Y_i, Z_i) - (X_i, A_i) - (U_i, V_i, W_i)$. □

Remark 5.4 (Separate coding over public and private links)**.** Separate encoding using codewords A^n, U^n, and V^n to communicate over the public link, followed by using codeword W^n over the private link turns out to be optimal in this case. The scheme results in individual rate constraints on R and R_p in the rate-distortion-cost-leakage region. The joint encoding over public and private links, e.g., the successive refinement type coding [EK11, Chapter 13], is not necessary since the eavesdropper is a passive terminal which does not decode any sequences.

5.2.4 Example: Taking Action to Enhance or to Suppress the Side Information at the Decoder

To illustrate the tradeoff in Theorem 5.2.1, we discuss a simple binary example. We are especially interested in characterizing the tradeoff between leakage rate and distortion when the action is taken at different costs. We assume that the actions are taken to enhance or to switch off the side information at the legitimate decoder. That is, taking action $A = 1$ corresponds to enhancing the side information and the enhanced side information is actually equal to the source, i.e., $Y = X$, while taking $A = 0$ corresponds to suppressing it, i.e., $Y = \emptyset$. We assume a unit cost function and the Hamming distortion measure, i.e., $\Lambda(a) = a$, and $d(x, \hat{x}) = 1$ if $x \neq \hat{x}$ and zero otherwise. The binary source X is assumed to be distributed according to Bernoulli(1/2), and the side information Z is simply given as an output of a binary symmetric channel with input X and crossover probability p (BSC(p)), as shown in Fig. 5.3.

It can be shown that the rate-distortion-leakage region for some fixed costs which is a specialization of Theorem 5.2.1 is given by the set of all tuples $(R, D, C, \triangle)$

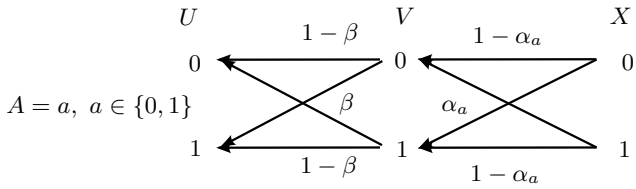

Figure 5.4: Achieving input distribution conditioned on action.

such that there exist $\alpha_0, \alpha_1, \beta \in [0, 1/2]$ satisfying

$$\begin{aligned} R &\geq \big(1 - h(\alpha_0)\big) \cdot (1 - C), \\ D &\geq \alpha_0 \cdot (1 - C), \\ \triangle &\geq 1 - \big[(1 - C) \cdot \big(h(\alpha_0) - h(p \star \alpha_0 \star \beta) + h(p)\big) \\ &\qquad + C \cdot \big(h(\alpha_1 \star \beta) - h(p \star \alpha_1 \star \beta) + h(p)\big)\big], \end{aligned} \tag{5.3}$$

where $h(\cdot)$ is the binary entropy function and $a \star b \triangleq a(1-b) + (1-a)b$.

The region above is achieved by the following input distribution. We set A independent of X where $\Pr(A = 1) = C$. Conditioned on $A = a$ for $a \in \{0, 1\}$, we model U and V as in Fig. 5.4, i.e., letting U be an output of a BSC(β) with input V, and letting V be an output of a BSC(α_a) with input X. Finally, we define a function $\tilde{g}(V, Y) = V$ if $A = 0$, and $\tilde{g}(V, Y) = Y = X$ otherwise. The converse can be shown using similar techniques as in [WZ76] and [VP13]. For completeness, we give the proof in Appendix 5.D.

It is interesting to see the tradeoff between the distortion at the decoder and the leakage rate at the eavesdropper when the actions are taken with different costs. Fig. 5.5 shows such a tradeoff where R is set to be sufficiently high such that the rate constraint is not active. It shows that for a given cost (less than 1) the minimum leakage rate at the eavesdropper can be reduced with a small increase in distortion, especially in the low distortion region. There also exists a distortion level D^* such that for any $D > D^*$ the minimum leakage rate cannot be improved further. This occurs when the distortion constraint is satisfied even without any information sent over the public link. The remaining leakage is due to the side information available at the eavesdropper. In addition, for a fixed leakage rate, the minimum distortion can be improved when the cost is increased, and similarly for a fixed distortion, the minimum leakage rate will decrease with costs.

For the case where rate R is not high enough, e.g., $R \in [0, 1 - C)$, the rate constraint is active and the minimum distortion will be limited by the rate such that $D_{\min} = (1-C) * h^{-1}(1 - \frac{R}{1-C})$. As shown in Fig. 5.6 for $C = 0.4$ and $R = 0.3168$, the distortion cannot decrease beyond $D_{\min} = 0.06$ with the increasing leakage rate. For $D \geq D_{\min}$, the tradeoff between the leakage rate and distortion remains the same as in Fig. 5.5.

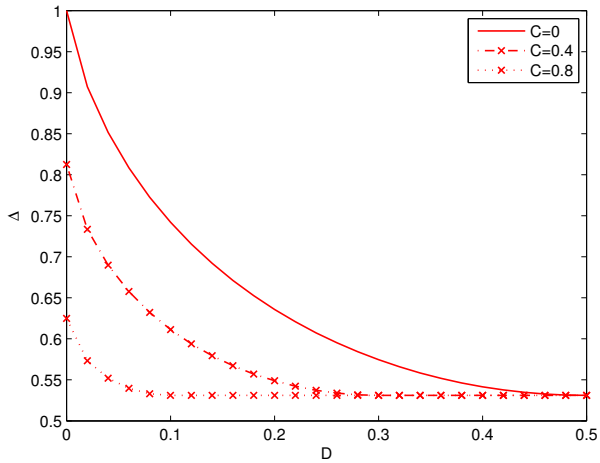

Figure 5.5: Leakage-distortion tradeoff for different costs and $p = 1/10$. The different line styles correspond to different costs.

Remark 5.5. The example in this chapter is slightly different from that in [VP13]. In [VP13], Y is modeled as an output of a binary erasure channel with input X and it is not known at the encoder, while in our case the encoder can use the knowledge about binary action A to deduce the exact Y. This knowledge about the decoder's side information leads to a larger leakage-distortion region. For example, the leakage-distortion region shown in Fig. 5.5 is larger than that in [VP13] where the erasure probability is set to be equal to $1 - C$.

5.3 Secure Source Coding With Action-dependent SI: Action Taken at Encoder

In this section, we consider a variant of the problem setting involving action-dependent side information in secure source coding. The main difference from the previous section is that now the action sequence is generated randomly based on the source sequence, i.e., $A^n \sim P_{A^n|X^n}$, rather than as a function of the source description. Therefore, neither the decoder nor the eavesdropper knows which action sequence is taken. This setting is seen as the case where *action is taken at the encoder*. The rate-limited communication from the encoder to the decoder is assumed to be eavesdropped and potentially leaks information to the eavesdropper. We characterize inner and outer bounds to the rate-distortion-cost-leakage region for a general setting, and show that the inner bound is tight under a special case of lossless reconstruction and no side information at the eavesdropper.

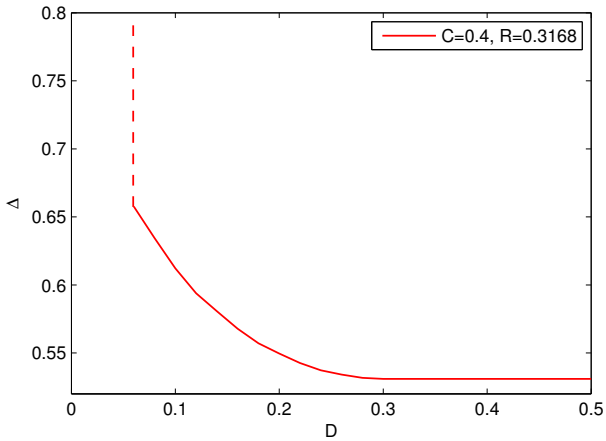

Figure 5.6: Leakage-distortion tradeoff when the rate constraint is active. Minimum distortion cannot decrease beyond $D_{\min} = 0.06$ with the increasing leakage rate.

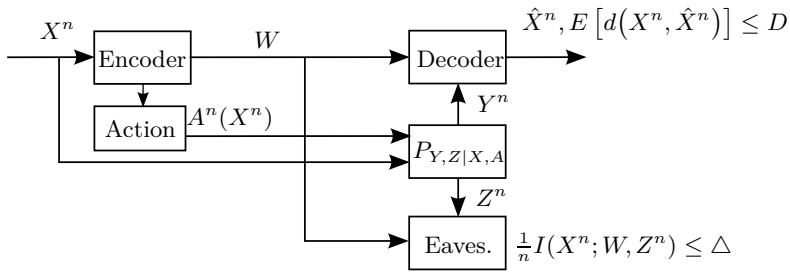

Figure 5.7: Secure source coding with action-dependent side information where action is taken at the encoder.

5.3.1 Problem Formulation

Consider the problem in Fig. 5.7. Let n denote the sequence length and $\mathcal{X}, \mathcal{A}, \mathcal{Y}$, and $\mathcal{Z}$ be finite source, action, and side information alphabets. Let X^n be the source sequence which is i.i.d. according to P_X. The action sequence A^n is generated based on the source sequence, i.e., $A^n \sim P_{A^n|X^n}$, and the side information (Y^n, Z^n) available to the decoder and the eavesdropper are the output of the discrete memoryless channel $P_{Y,Z|X,A}$. The definitions of $(|\mathcal{W}^{(n)}|, n)$-code, achievability of a rate-distortion-cost-leakage tuple, and the rate-distortion-cost-leakage region are the same as in Section 5.2.1, except that a stochastic action encoder $F_a^{(n)}$ which takes X^n as input and generates A^n according to a conditional PMF $p(a^n|x^n)$ replaces the deterministic action encoder in the definition of the code.

5.3.2 Main Result

Theorem 5.3.1 (Inner bound)**.** *Let X^n be a discrete memoryless source (DMS) $\sim P_X$, (Y^n, Z^n) be side information $\sim P_{Y,Z|X,A}$, and $d(x,\hat{x})$ and $\Lambda(a)$ be distortion and cost measures. A tuple $(R, D, C, \triangle) \in \mathbb{R}_+^4$ is achievable for secure source coding with action-dependent side information where action is taken at the encoder if*

$$R \geq I(X;U) - I(Y;U) + I(X,A;V|U,Y), \tag{5.4a}$$

$$D \geq E\big[d\big(X, \tilde{g}(V,Y)\big)\big], \tag{5.4b}$$

$$C \geq E[\Lambda(A)], \tag{5.4c}$$

$$\triangle \geq I(Z;X|U) + \min\{I(Y;U), I(Z;U)\} + I(X;U) - I(Y;U) + I(X,A;V|U,Y), \tag{5.4d}$$

$$= I(X;U,V,Y) + \min\{I(Z;X|U) - I(Y;X|U), I(Z,X,U) - I(Y;X,U)\} + I(A;V|U,X,Y), \tag{5.4e}$$

$$\triangle \geq I(Z;X|U) + \min\{I(Y;U), I(Z;U)\}, \tag{5.4f}$$

for some joint distributions of the form

$$P_X(x)P_{U|X}(u|x)P_{A|U,X}(a|u,x)P_{V|U,A,X}(v|u,a,x)P_{Y,Z|X,A}(y,z|x,a),$$

with $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{A}| + 5, |\mathcal{V}| \leq |\mathcal{U}|(|\mathcal{X}||\mathcal{A}| + 1)$, and a function $\tilde{g} : \mathcal{V} \times \mathcal{Y} \to \hat{\mathcal{X}}$.

Proof. The proof of Theorem 5.3.1 is provided in Appendix 5.E. □

Remark 5.6 ("Hybrid" coding scheme)**.** The achievable scheme used to prove the inner bound in Theorem 5.3.1 is of a hybrid type which combines elements from layered (superposition-type) coding, Wyner-Ziv coding, and Gel'fand-Pinsker coding. We generate superposition codewords U^n and V^n, with U^n as a basis codeword. Depending on the value of $I(X;U) - I(Y;U)$, U^n can be communicated either over the rate-limited link using the Wyner-Ziv coding (when $I(X;U) - I(Y;U) \geq 0$), or over the channel $P_{Y|X,A}$ using the Gel'fand-Pinsker type coding (when $I(X;U) - I(Y;U) < 0$). Then, an action sequence A^n is i.i.d. according to $P_{A|X,U}$. In addition, we perform the Wyner-Ziv coding using codeword V^n, regardless of the operation w.r.t U^n. We note that, unlike the previous case in Section 5.2 where A^n is a function of the source description W, here both decoder and eavesdropper who have access to W do not know which action sequence is taken. The scheme is an extension of that considered in [PW11, Section III] where U^n is set to be equal to A^n.

Remark 5.7 (Interpretation of leakage rate constraint)**.** Inner bound in Theorem 5.3.1 is expressed with two leakage rate bounds. The first one in (5.4d) is derived from the case where some source description with positive rate is sent over the rate-limited link, while the second bound in (5.4f) is derived from the case where no information with positive rate is sent. In the first bound, the term

$I(X;U) - I(Y;U) + I(X,A;V|U,Y)$ corresponds to the leakage rate due to the source description which is observed by the eavesdropper. The remaining terms can be interpreted as additional leakage due to the side information Z^n which is an output of the channel $P_{Z,Y|X,A}$, given that the description over the rate-limited link is observed. The $\min\{\cdot\}$ in the expression in Theorem 5.3.1 appears due to two different bounding methods used in the proof.

Theorem 5.3.2 (Outer bound)**.** *If a tuple $(R, D, C, \triangle) \in \mathbb{R}_+^4$ is achievable for secure source coding with action-dependent side information where action is taken at the encoder, then it must satisfy*

$$R \geq I(X;A) - I(Y;A) + H(X|Y,A) - H(X|V,Y), \tag{5.5a}$$

$$D \geq E\big[d\big(X, \tilde{g}(V,Y)\big)\big], \tag{5.5b}$$

$$C \geq E[\Lambda(A)], \tag{5.5c}$$

$$\triangle \geq I(X;V,Y) + I(Z;X,K|U) - I(Y;X,K|U), \tag{5.5d}$$

$$\triangle \geq I(X;Z), \tag{5.5e}$$

for some joint distributions of the form

$$P_X(x)P_{A,U,V,K|X}(a,u,v,k|x)P_{Y,Z|X,A}(y,z|x,a),$$

and a function $\tilde{g} : \mathcal{V} \times \mathcal{Y} \to \hat{\mathcal{X}}$.

Proof. The proof of Theorem 5.3.2 is provided in Appendix 5.F. □

Remark 5.8 (Special cases)**.** In the following, we present a couple of *extreme* special cases which show that our results can reduce to the optimal ones.

- (i) State masking with zero rate [MS07]: If we set $D = \infty$, there is no distortion constraint and any non-negative rate is achievable. By setting $U = V = \varnothing$ in Theorem 5.3.1, the minimum achievable leakage rate is $I(X;Z)$ which coincides with the result for state masking with zero rate in [MS07].
- (ii) Lossless reconstruction with $Z = \varnothing$: In this case, the only leakage comes from the source description. By setting $Z = \varnothing, V = X$, and $U = A$ in Theorem 5.3.1, we obtain the optimal rate-cost-leakage region in Corollary 5.3.1 below. The converse proof for the rate constraint follows similarly as in Theorem 5.3.2, while the proof for the leakage rate constraint is given in Appendix 5.G.

  Corollary 5.3.1. *The rate-cost-leakage region for lossless secure source coding with action-dependent side information where action is taken at encoder and $Z = \varnothing$ is given by the set of all $(R, C, \triangle) \in \mathbb{R}_+^3$ satisfying*

$$R \geq I(X;A) - I(Y;A) + H(X|Y,A), \tag{5.6a}$$

$$C \geq E[\Lambda(A)], \tag{5.6b}$$

$$\triangle \geq I(X;A) - I(Y;A) + H(X|Y,A), \tag{5.6c}$$

for some joint distributions of the form $P_X(x)P_{A|X}(a|x)P_{Y|X,A}(y|x,a)$.

5.3.3 Example: Taking Action to Observe or not to Observe the Side Information at Decoder

In this section, we discuss a binary example for the special case of lossless reconstruction of Theorem 5.3.1. The inner bound for the case of lossless reconstruction is obtained from Theorem 5.3.1 by setting $V = X$. We are interested in characterizing the tradeoff between achievable leakage rate and cost when different rates are used. We assume that the actions are taken at the encoder to observe or not to observe the side information at the legitimate decoder. That is, taking action $A = 1$ corresponds to observing the binary side information at the decoder as an output of a BSC(p_y) and input X, while taking $A = 0$ corresponds to not observing the side information, i.e., $Y = \emptyset$. We assume a unit cost function, i.e., $\Lambda(a) = a, a \in \{0, 1\}$. The binary source X is assumed to be distributed according to Bernoulli(1/2), and the side information Z is simply given as an output of a BSC(p_z) with input X.

By assuming a general input distribution $p_{A=1|X=0} = p_0$ and $p_{A=0|X=1} = p_1$, where $p_0, p_1 \in [0, 1]$ and set $U = A$ ($P_{A|X}$ is the only variable to be optimized over), the inner bound to the rate-cost-leakage region in Theorem 5.3.1 reduces to the set of all tuples $(R, C, \triangle)$ that satisfy

$$\begin{aligned} R &\geq q, \\ C &\geq 1/2(1 + p_0 - p_1), \\ \triangle &\geq 1 - h(p_z) + [q]^+, \end{aligned}$$

for some $p_0, p_1 \in [0, 1]$, where $h(\cdot)$ is the binary entropy function and

$$\begin{aligned} q = 1 - [&-1/2(p_0(1-p_y) + (1-p_1)p_y)\log_2(1/2(p_0(1-p_y) + (1-p_1)p_y)) \\ &- 1/2(p_0p_y + (1-p_1)(1-p_y))\log_2(1/2(p_0p_y + (1-p_1)(1-p_y))) \\ &- 1/2(1-p_0+p_1)\log_2(1/2(1-p_0+p_1))] + 1/2(1+p_0-p_1)h(p_y). \end{aligned}$$

It is interesting to see the tradeoff between the leakage rate at the eavesdropper and the action cost. Fig. 5.8 shows such a tradeoff for different rates where $p_z = 0.2$ and $p_y = 0.3$. It shows that for a given rate the minimum leakage rate at the eavesdropper can be reduced with an increasing cost. Interestingly, we see that the minimum achievable leakage rate can be achieved with only a fraction of cost, e.g., $C = 0.38$ which suggests that an action can be taken efficiently. The fact that increasing cost C beyond certain value does not improve the leakage rate is because at sufficiently high C the side information Y^n is of high quality, and the lossless reconstruction can be satisfied without any communication over a rate-limited link. In the low cost region, we observe that the rate R is a limiting factor for the minimum achievable cost required for lossless reconstruction. For example, when the rate $R = 0.6$, we could achieve a lower leakage rate than the maximum leakage rate at cost $C = 0.1$. However, this same level of leakage rate would require a higher cost when $R = 0.4$.

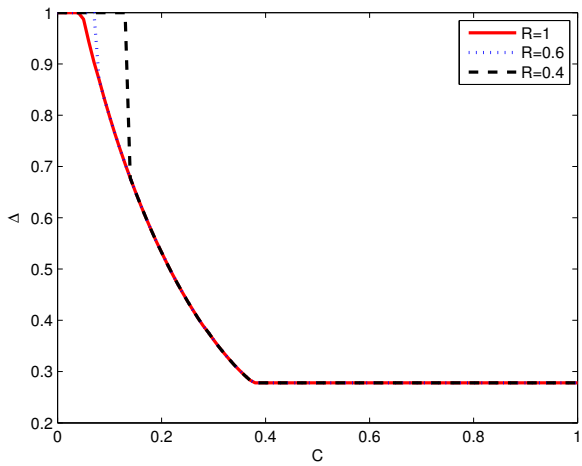

Figure 5.8: Leakage-cost tradeoff for different rates and $p_z = 0.2, p_y = 0.3$. The different line styles correspond to different rates.

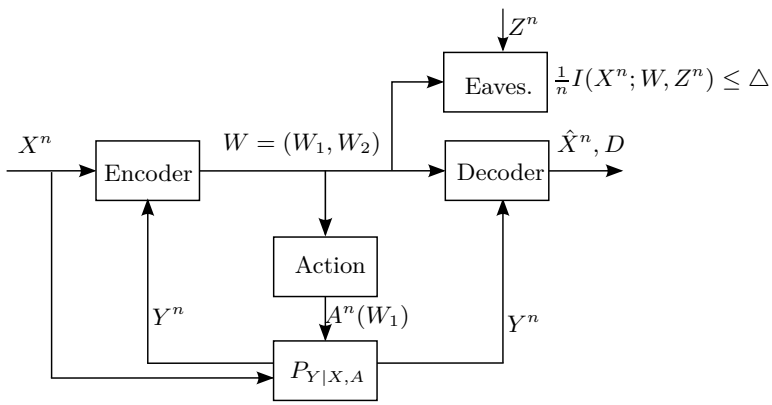

Figure 5.9: Secure source coding with common two-sided action-dependent.

5.4 Secure Source Coding With Common Two-sided Action-dependent Side Information: Secret Key Generation

This section focuses on another aspect of action-dependent side information in secure source coding problem, i.e., enabling the secret key generation. The setting is shown in Fig. 5.9 where we assume that there exists a costly secure action-dependent

side information channel which provides side information to both encoder and decoder, but not to the eavesdropper. When the side information at the encoder and decoder are identical, we are interested in how this common randomness can support the source transmission, and at the same time helps the system to reduce the leakage level. The setting is relevant in scenarios where it is possible for the decoder (action taken at the decoder) to generate secure side information with some costs while assuming that the communication links between the encoder and the decoder is public. Our goal is to characterize the fundamental tradeoff between the transmission rate, incurred distortion, associated action cost, and resulting information leakage rate in the form of the rate-distortion-cost-leakage region. We first characterize inner and outer bounds to the rate-distortion-cost-leakage region and then show that they are tight under the logarithmic loss distortion measure recently introduced in the context of multiterminal source coding in [CW14].

5.4.1 Problem Formulation

We consider finite sets for source, action, side information, and reconstruction alphabets, i.e., $\mathcal{X}, \mathcal{A}, \mathcal{Y}, \hat{\mathcal{X}}$ are finite. Let X^n be the n-length source sequence which is i.i.d. according to P_X. Given a source sequence X^n, an encoder generates a source description $W_1 \in \mathcal{W}_1^{(n)}$ and sends it over a noise-free, rate-limited link to a decoder and an action encoder. The action sequence is chosen based on the source description. Then the action-dependent side information is generated as an output of the memoryless channel $P_{Y|X,A}$ and is available to both encoder and decoder. Based on the side information Y^n and the source sequence X^n, the encoder in the second stage generates another source description $W_2 \in \mathcal{W}_2^{(n)}$ and sends it over the noise-free, rate-limited link to the decoder. Let $W = (W_1, W_2)$. Upon receiving the source descriptions W and the side information, the decoder reconstructs the source sequence as $\hat{X}^n$ subject to the distortion constraint. We note that an eavesdropper also observes W, and the degraded side information Z^n which is generated as an output of the memoryless channel $P_{Z|Y}$, i.e., $Z^n - Y^n - (X^n, A^n)$ forms a Markov chain.

Definition 5.3. A $(|\mathcal{W}^{(n)}|, n)$-code for secure source coding with common two-sided action-dependent side information consists of

- a stochastic encoder $F_1^{(n)}$ which takes X^n as input and generates $W_1 \in \mathcal{W}_1^{(n)}$ according to a conditional PMF $p(w_1|x^n)$,
- a deterministic action encoder $f_a^{(n)} : \mathcal{W}_1^{(n)} \to \mathcal{A}^n$,
- a stochastic encoder $F_2^{(n)}$ which takes X^n and Y^n as inputs and generates $W_2 \in \mathcal{W}_2^{(n)}$ according to a conditional PMF $p(w_2|x^n, y^n)$, and
- a deterministic decoder $g^{(n)} : \mathcal{W}_1^{(n)} \times \mathcal{W}_2^{(n)} \times \mathcal{Y}^n \to \hat{\mathcal{X}}^n$,

where $\mathcal{W}^{(n)} = \mathcal{W}_1^{(n)} \times \mathcal{W}_2^{(n)}$, and $\mathcal{W}_1^{(n)}$ and $\mathcal{W}_2^{(n)}$ are finite sets.

The *information leakage* at the eavesdropper, who has access to W and the degraded side information Z^n, is measured by the normalized mutual information $\frac{1}{n} I(X^n; W, Z^n)$.

Definition 5.4. A rate-distortion-cost-leakage tuple $(R, D, C, \triangle)$ is said to be *achievable* if for any $\delta > 0$ and all sufficiently large n there exists a $(|\mathcal{W}^{(n)}|, n)$ code such that $\frac{1}{n} \log \left|\mathcal{W}^{(n)}\right| \leq R + \delta$, $E[d^{(n)}(X^n, g^{(n)}(W, Y^n))] \leq D + \delta$, $E\left[\Lambda^{(n)}(A^n)\right] \leq C + \delta$, and $\frac{1}{n} I(X^n; W, Z^n) \leq \triangle + \delta$, where $d^{(n)}(\cdot)$ and $\Lambda^{(n)}(\cdot)$ are distortion and cost functions, respectively. The *rate-distortion-cost-leakage* region $\mathcal{R}_{commonSI}$ is the set of all achievable tuples.

5.4.2 Inner and Outer Bounds

Proposition 5.4.1 (Inner bound)**.** *Let X^n be a DMS $\sim P_X$, (Y^n, Z^n) be side information $\sim P_{Y|X,A}$ and $P_{Z|Y}$, and $d(x, \hat{x})$ and $\Lambda(a)$ be distortion and cost measures. A tuple $(R, D, C, \triangle) \in \mathbb{R}_+^4$ is achievable for secure source coding with common two-sided action-dependent side information if*

$$R \geq I(X; A) + I(X; U|A, Y), \tag{5.7a}$$
$$D \geq E[d(X, \tilde{g}(U, A, Y))], \tag{5.7b}$$
$$C \geq E[\Lambda(A)], \tag{5.7c}$$
$$\triangle \geq I(X; A, Z) + [I(X; U|A, Y) - H(Y|X, A, Z)]^+, \tag{5.7d}$$

for some joint distributions of the form

$$P_X(x) P_{A|X}(a|x) P_{Y|X,A}(y|x, a) P_{U|X,A}(u|x, a) P_{Z|Y}(z|y)$$

with $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{A}| + 1$, and a function $\tilde{g} : \mathcal{U} \times \mathcal{A} \times \mathcal{Y} \to \hat{\mathcal{X}}$.

Proof. The proof is provided in Appendix 5.H where we extend a similar proof technique used in the problem of secure source coding with side information at the encoder [CK13a]. That is, the secret key is generated based on the side information Y^n and is used to scramble part of the source description sent in the second stage. This method can effectively decrease the amount of relevant information leaked to the eavesdropper as seen by the term $-H(Y|X, A, Z)$ in (5.7d). □

Proposition 5.4.2 (Outer bound)**.** *If a tuple $(R, D, C, \triangle) \in \mathbb{R}_+^4$ is achievable for secure source coding with* common two-sided action-dependent side information*, then it must satisfy*

$$R \geq I(X; A) + I(X; U|A, Y), \tag{5.8a}$$
$$D \geq E[d(X, \tilde{g}(U, A, Y))], \tag{5.8b}$$
$$C \geq E[\Lambda(A)], \tag{5.8c}$$
$$\triangle \geq I(X; A, Z) + [I(X; U|A, Y) - H(Y|X, A, Z)]^+, \tag{5.8d}$$

for some joint distributions of the form

$$P_X(x)P_{A|X}(a|x)P_{Y|X,A}(y|x,a)P_{U|X,A,Y}(u|x,a,y)P_{Z|Y}(z|y)$$

and a function $\tilde{g}: \mathcal{U} \times \mathcal{A} \times \mathcal{Y} \to \hat{\mathcal{X}}$.

Proof. The proof is provided in Appendix 5.I. □

Remark 5.9. Inner and outer bounds in Propositions 5.4.1 and 5.4.2 are different only in the set of input distributions. However, we do not expect the inner bound in Proposition 5.4.1 to be tight in general as the action-dependent side information Y^n at the encoder is not used for encoding U^n, cf. $Y - (X, A) - U$ forms a Markov chain. For a more general inner bound, one might follow the similar scheme proposed in [CK13a].

5.4.3 Special Case: Logarithmic Loss Distortion

We show that the inner and outer bounds in Propositions 5.4.1 and 5.4.2 are tight for the problem under a *logarithmic loss* distortion measure. Before stating the main result, we restate the definition and highlight some important properties of the logarithmic loss distortion for which more details can be found in [CW14]. Logarithmic loss has the interesting property that, when used as a distortion measure in the Wyner-Ziv (like) problem [WZ76], the side information at the encoder does not improve the rate-distortion region. This property is a reminiscence of what is known for the Gaussian Wyner-Ziv problem [Wyn78], and is essential in establishing a complete result in this section by using the achievable schemes which neglect the side information for encoding at the encoder. The assumption of logarithmic loss distortion also allows us to establish several complete results in the next chapter where the public helper is present in the model.

Definition 5.5 (Logarithmic loss)**.** For logarithmic loss distortion measure, we let the reconstruction alphabet $\hat{\mathcal{X}}$ be the set of probability distributions over the source alphabet $\mathcal{X}$, i.e., $\hat{\mathcal{X}} = \{p | p \text{ is a PMF on } \mathcal{X}\}$. For a sequence $\hat{X}^n \in \hat{\mathcal{X}}^n$, we denote $\hat{X}_i$, $i = 1, \ldots, n$, the i^{th} element of $\hat{X}^n$. Then $\hat{X}_i, i = 1, \ldots, n$ is a probability distribution on $\mathcal{X}$, i.e., $\hat{X}_i : X \to [0, 1]$, and $\hat{X}_i(x)$ is a probability distribution on $\mathcal{X}$ evaluated for the outcome $x \in \mathcal{X}$. In other words, the decoder generates "soft" estimates of the source sequence. The logarithmic loss distortion measure is defined as

$$d(x, \hat{x}) \triangleq \log(\frac{1}{\hat{x}(x)}) = D_{KL}(\mathbf{1}_{\{x\}} || \hat{x}),$$

where $\mathbf{1}_{\{x\}} : \mathcal{X} \to \{0, 1\}$ is an indicator function such that, for $a \in \mathcal{X}$, $\mathbf{1}_{\{x\}}(a) = 1$ if $a = x$, and $\mathbf{1}_{\{x\}}(a) = 0$ otherwise. That is, $d(x, \hat{x})$ is the Kullback-Leibler divergence between the empirical distribution of the event $X = x$ and the estimate $\hat{x}$. By using this definition for the symbol-wise distortion, the distortion between sequences is then defined as $d^{(n)}(x^n, \hat{x}^n) \triangleq \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i)$.

In the following, we present a couple of lemmas which appear in [CW14] and are essential in proving the result under the logarithmic loss distortion. Lemma 5.1 is used in the achievability proof (inner bound argument), while Lemma 5.2 is used for upper bounding the conditional entropy in the converse proof (outer bound argument). Both follow quite directly from the definition of logarithmic loss, but we include their proofs [CW14] in Appendices 5.J and 5.K for completeness.

Lemma 5.1 (Inner bound argument)**.** *Let U be the argument of the reconstruction function $g(\cdot)$, then under the log-loss distortion measure, we get $E[d(X, g(U))] = H(X|U)$.*

Lemma 5.2 (Outer bound argument)**.** *Let Z be the argument of the reconstruction function $g^{(n)}(\cdot)$, then under the log-loss distortion measure, we get $E[d^{(n)}(X^n, g^{(n)}(Z))] \geq \frac{1}{n} H(X^n|Z)$.*

We now state the main result which is the rate-distortion-cost-leakage region for secure source coding with common two-sided action-dependent side information under logarithmic loss distortion measure.

Theorem 5.4.1. *The rate-distortion-cost-leakage region $\mathcal{R}_{\text{common SI, log-loss}}$ for secure source coding with common two-sided action-dependent side information under logarithmic loss and degraded side information at eavesdropper is the set of all tuples $(R, D, C, \triangle) \in \mathbb{R}_+^4$ that satisfy*

$$R \geq I(X; A) + [H(X|A, Y) - D]^+, \tag{5.9a}$$

$$C \geq E[\Lambda(A)], \tag{5.9b}$$

$$\triangle \geq I(X; A, Z) + [H(X|A, Y) - D - H(Y|X, A, Z)]^+, \tag{5.9c}$$

for some joint distributions of the form $P_X(x)P_{A|X}(a|x)P_{Y|X,A}(y|x, a)P_{Z|Y}(z|y)$.

Proof. The proof is given in Appendix 5.L. It follows from those of Propositions 5.4.1 and 5.4.2 where we utilize the properties of logarithmic loss distortion such as Lemmas 5.1 and 5.2. □

Remark 5.10. We note that the proof of the inner bound in Theorem 5.4.1 holds only for bounded distortion measures. However, the logarithmic loss distortion measure is not bounded. To address this issue, we refer to [CK13b, Remark 3.4] where the proof of achievability can be extended to logarithmic loss distortion by perturbing the reconstruction probability distribution. That is, we assign a small positive value to the reconstruction probability distribution that takes value zero. By this perturbation, the maximum distortion incurred can be upperbounded, and the proof of the inner bound in Theorem 5.4.1 can then be applied with this perturbed reconstruction function.

Remark 5.11 (Side information at encoder under log-loss distortion)**.** We note that the encoding consists of two processing stages. That is, the encoder first transmits the source description used for generating the action sequence, and then utilizes the newly generated action-dependent side information for the second-stage encoding. This causal structure essentially makes it difficult for us to characterize the complete rate-distortion-cost-leakage region for a general distortion case. In the general case, the side information Y^n at the encoder could be exploited for both encoding and enabling the secret key generation. However, these dual roles are coupled and they lead to a certain Markov condition on the set of input distributions that is difficult to satisfy in the converse argument (see, e.g, a related work that focuses on the role of side information at the encoder in secure source coding [CK13a]). When considering the logarithmic loss distortion, the availability of side information Y^n at the encoder does not improve the rate-distortion-cost tradeoff. That is, we can neglect the side information at the encoder in the second stage encoding and therefore avoid the causality issue. Effectively, the encoding process reduces to the one-step joint encoding which allows us to characterize the complete rate-distortion-cost-leakage region.

Remark 5.12 (Interpretation of leakage rate constraint)**.** The fact that the side information Y^n at the encoder is not helpful in terms of rate-distortion-cost tradeoff under log-loss distortion is a reminiscence of what is known for the Gaussian Wyner-Ziv problem [Wyn78]. However, this action-dependent side information still has a role in enabling the secret key generation at the encoder and the decoder. We can see the effect of leakage reduction from the term $-H(Y|X, A, Z)$ in the leakage rate constraint in (5.9c). Basically, the expression of leakage rate constraint (5.9c) consists of three parts. The term $I(X; A, Z)$ is the leakage due to the side information Z^n and the description sent for generating action sequence and intercepted by the eavesdropper; the term $H(X|A, Y) - D$ is due to the additional source description (second-stage) sent to the decoder and intercepted by the eavesdropper; and finally the term $-H(Y|X, A, Z)$ is the leakage reduction due to the use of secret key to scramble parts of the second-stage description. We see that contribution of each term can be controlled by the choice of action. For the case when $H(X|A, Y) - D \leq 0$, we do not need to send any second-stage description and thus the secret key generation is not required.

Remark 5.13 (Tradeoff between dual roles of side information)**.** The action-dependent side information plays dual roles in our setting: 1) reducing the rate needed for source reconstruction, and 2) explicitly reducing the leakage rate by enabling secret key generation at the encoder and the decoder. We argue that there exists a tradeoff between these two roles. The tradeoff essentially depends on the correlation between the side information Y^n and the source sequence X^n which can be tuned by the choice of actions. For example, in the extreme case, if the side information Y^n is equal to X^n, the role of rate reduction totally dominates the role of secret key generation. That is, we can reduce the rate so that we do not send any second-stage description, while the side information at the encoder cannot be used

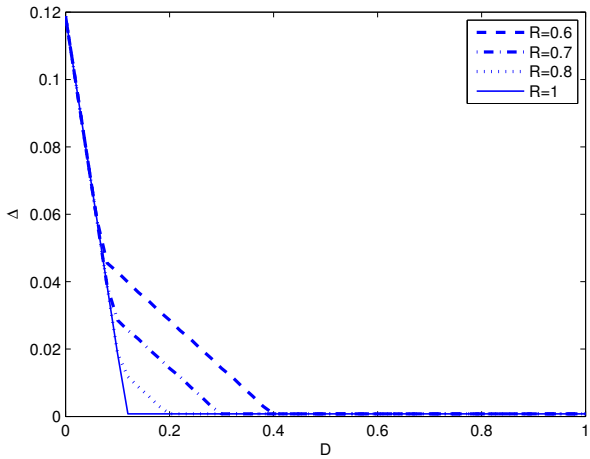

Figure 5.10: Leakage-distortion tradeoff for different rates and $C = 0.8, p_y = 0.3, p_z = 0.1$. The different line styles correspond to different rates.

for secret key generation since it is just equal to the source. On the other extreme, if the side information Y^n is conditionally independent of X^n given A^n, it provides no rate reduction, but it can be fully exploited for secret key generation.

5.4.4 Example: Taking Action to Improve the Quality of Common Side Information at Encoder and Decoder

In this section, we discuss a binary example of Theorem 5.4.1. We are interested in characterizing the optimal tradeoff between minimum leakage rate and distortion when different rates are used at a certain fixed cost. The actions are taken at the decoder based on the source description to improve the quality of common side information Y. That is, taking action $A = 1$ corresponds to observing the binary side information Y as an output of a BSC(p_y), $p_y < 0.5$, with input X, while taking $A = 0$ corresponds to observing an independent side information, i.e., Y is an output of a BSC(0.5) with input X. We assume a unit cost function, i.e., $\Lambda(a) = a, a \in \{0, 1\}$. The binary source X is assumed to be distributed according to Bernoulli(1/2), and the degraded side information Z is simply given as an output of a BSC(p_z) with input Y.

In Theorem 5.4.1, $P_{A|X}$ is the only variable to be optimized. By assuming a general input distribution $p_{A=1|X=0} = p_0$ and $p_{A=0|X=1} = p_1$, $p_0, p_1 \in [0, 1]$, one can numerically evaluate the rate-distortion-cost-leakage region in Theorem 5.4.1 and plot the optimal tradeoff between the leakage rate and distortion for certain rates R and a fixed cost $C = 0.8$ in Fig. 5.10. Fig. 5.10 shows that for fixed

$C = 0.8$, the minimum leakage rate is a piecewise linear function of distortion. At a very high rate, e.g., $R = 1$, there is no additional restriction on the feasible input distribution $P_{A|X}$ as the rate constraint is always satisfied for all D. The minimum leakage rate is therefore a linear function of D up to certain value where the term $H(X|A,Y) - D - H(Y|X,A,Z)$ becomes negative. When the rate is lower, e.g., $R = 0.8$, we still have no additional restriction on $P_{A|X}$ if D is sufficiently low. However, for a moderate D, the rate constraint essentially restricts the feasible $P_{A|X}$ which in turn affects the minimum value of $I(X;A,Z)+H(X|A,Y)-H(Y|X,A,Z)$. This causes the minimum leakage rate to linearly decrease with D at a different slope from that of the low D region.

5.5 Conclusion

In this chapter, we apply the notion of action-dependent side information to the secure source coding problem. It is motivated by the fact that secrecy in secure source coding is characterized by amount of information leaked through the source description and the eavesdropper's side information. Modelling the system with action-dependent side information essentially provides another degree of freedom to control the side information available at the decoder and possibly at the eavesdropper, which in turn affects the amount of information needed to describe the source, leading to a more flexible and efficient approach to handle a secrecy constrained system. We consider different aspects of action-dependent side information in the problem and characterize the fundamental tradeoffs between the minimum information rate needed for describing the source, the resulting distortion at the legitimate decoder, the leakage rate at the eavesdropper, and the cost associated with the actions taken, in the form of the rate-distortion-cost-leakage region. The scenario where the common side information is available at both encoder and decoder, but not at the eavesdropper, gives rise to a new, interesting scheme involving secret key generation, and also reveals some connection to secure network coding. It is interesting to note that, in some cases, the availability of side information at the encoder is not useful in terms of rate-distortion tradeoff, but it can be used for secret key generation and thus helps to improve the leakage rate. The results in this chapter can serve as guidelines in designing for example secure transmission for sensor networks, where a sensor node may take actions with some costs to enhance the side information at the intended receiver (or worsen that at the eavesdropper), with the goal of achieving higher security and reconstruction quality.

Appendix for Chapter 5

5.A Proof of Theorem 5.2.1

5.A.1 Sketch of Achievability

The proof follows a standard random coding arguments where we show the existence of a code that satisfies the rate, distortion, cost, and leakage constraints. The outline of the proof is given in the following.

Codebook generation: Fix $P_{A|X}P_{V|X,A}P_{U|V}$, and the function $\tilde{g} : \mathcal{V} \times \mathcal{Y} \to \mathcal{X}$.

- Randomly and independently generating $2^{n(I(X;A)+\delta_\epsilon)}$ codewords $a^n(w_a) \sim \prod_{i=1}^n P_A(a_i(w_a)), w_a \in [1 : 2^{n(I(X;A)+\delta_\epsilon)}]$.
- For each $w_a \in [1 : 2^{n(I(X;A)+\delta_\epsilon)}]$, randomly and conditionally independently generate $2^{n(I(U;X|A)+\delta_\epsilon)}$ codewords $u^n(w_a, k)$ each according to the distribution $\prod_{i=1}^n P_{U|A}(u_i|a_i(w_a))$, for $k \in [1 : 2^{n(I(U;X|A)+\delta_\epsilon)}]$, and distribute these codewords uniformly at random into $2^{n(I(U;X|A)-I(U;Y|A)+2\delta_\epsilon)}$ equal-sized bins $b_U(w_u), w_u \in [1 : 2^{n(I(U;X|A)-I(U;Y|A)+2\delta_\epsilon)}]$.
- Then for each pair of (w_a, k), randomly and conditionally independently generate $2^{n(I(V;X|U,A)+\delta_\epsilon)}$ codewords $v^n(w_a, k, l)$ each according to the distribution $\prod_{i=1}^n P_{V|U,A}(v_i|u_i(k), a_i(w_a)), l \in [1 : 2^{n(I(V;X|U,A)+\delta_\epsilon)}]$, and distribute them uniformly at random into $2^{n(I(V;X|U,A)-I(V;Y|U,A)+2\delta_\epsilon)}$ bins $b_V(w_v)$, where $w_v \in [1 : 2^{n(I(V;X|U,A)-I(V;Y|U,A)+2\delta_\epsilon)}]$.

The codebooks are revealed to the encoder, the action encoder, the decoder, and the eavesdropper.

Encoding:

- For given source sequence x^n, the encoder looks for $a^n(w_a)$ which is jointly typical with x^n. Since there are more than $2^{nI(X;A)}$ codewords a^n generated, by the covering lemma (Lemma 2.6), there exists such an a^n with high probability. If there are more than one, we choose one uniformly at random and send the corresponding index w_a to the decoder.
- Next, the encoder looks for $u^n(w_a, k)$ which is jointly typical with (x^n, a^n). Since there are more than $2^{nI(X;U|A)}$ codewords u^n generated, by the covering lemma, there exists such a u^n with high probability. If there are more than one, we choose one uniformly at random and send the corresponding bin index w_u to the decoder.
- Again, the encoder looks for $v^n(w_a, k, l)$ which is jointly typical with (x^n, a^n, u^n). Since there are more than $2^{nI(X;V|U,A)}$ codewords v^n generated, by the covering lemma, there exists such a v^n with high probability. If there are more than one, we choose one uniformly at random and send the corresponding bin index w_v to the decoder.

This gives the total rate of $I(X;A) + I(U;X|A) - I(U;Y|A) + I(V;X|U,A) - I(V;Y|U,A) + 5\delta_\epsilon = I(X;A) + I(V;X|Y,A) + 5\delta_\epsilon$. Then action-dependent side information (y^n, z^n) are generated as the output of the memoryless channel $P_{Y,Z|X,A}$.

Decoding:

- Upon receiving the indices w_a, w_u, w_v and the side information y^n, the decoder looks for the unique u^n which is jointly typical with (y^n, a^n). Since there are less than $2^{nI(Y;U|A)}$ sequences in the bin $b_U(w_u)$, by the packing lemma (Lemma 2.7), it will find the unique and correct u^n with high probability.
- Then the decoder looks for the unique v^n which is jointly typical with the tuple (y^n, a^n, u^n). Since there are less than $2^{nI(Y;V|U,A)}$ sequences in the bin $b_V(w_v)$, by the packing lemma, it will find the unique and correct v^n with high probability.
- The decoder puts out $\hat{x}^n$ as a reconstruction of x^n, where the i^{th} element $\hat{x}_i = \tilde{g}(v_i, y_i)$.

Analysis of distortion and cost: As we have that (x^n, a^n, v^n, y^n) are jointly typical with high probability, the average distortion and cost constraints are satisfied provided $E[d(X, \tilde{g}(V,Y))] \le D$ and $E[\Lambda(A)] \le C$.

Based on the condition that all sequences are jointly typical with high probability, we can bound some conditional entropy terms of interest into a single letter form using the following two lemmas. These lemmas will be useful in the analyses of the leakage rate, and their proofs are given in Appendix 5.B and 5.C.

Lemma 5.3. *Assume that (X^n, A^n) are jointly typical with high probability. Let Z^n i.i.d. $\sim P_{Z|X,A}$, we have that $H(Z^n|X^n, A^n) \ge n(H(Z|X,A) - \delta_\epsilon)$, where $\delta_\epsilon \to 0$ as $\epsilon \to 0$, and $\epsilon \to 0$ as $n \to \infty$.*

Lemma 5.4. *Assume that (X^n, U^n, Z^n) are jointly typical with high probability. Then we have that $H(Z^n|A^n, U^n, \mathcal{C}_n) \le n(H(Z|U,A) + \delta_\epsilon)$.*

Analysis of leakage rate: The normalized mutual information averaged over randomly chosen codebook $\mathcal{C}_n$ can be bounded as follows.

$$
\begin{aligned}
&I(X^n; W_a, W_u, W_v, Z^n|\mathcal{C}_n) \le I(X^n; W_a, K, W_v, Z^n|\mathcal{C}_n) \\
&= H(X^n|\mathcal{C}_n) - H(X^n, W_a, K, W_v, Z^n|\mathcal{C}_n) + H(W_a, K, W_v|\mathcal{C}_n) \\
&\qquad + H(Z^n|W_a, K, W_v, \mathcal{C}_n) \\
&= -H(Z^n|X^n, \mathcal{C}_n) - H(W_a, K, W_v|X^n, Z^n, \mathcal{C}_n) + H(W_a, K, W_v|\mathcal{C}_n) \\
&\qquad + H(Z^n|W_a, K, W_v, \mathcal{C}_n) \\
&\overset{(a)}{\le} -H(Z^n|X^n, A^n) + H(W_a, K, W_v|\mathcal{C}_n) + H(Z^n|W_a, K, W_v, \mathcal{C}_n) \\
&\le -H(Z^n|X^n, A^n) + H(W_a|\mathcal{C}_n) + H(K|\mathcal{C}_n) + H(W_v|\mathcal{C}_n) + H(Z^n|A^n, U^n, \mathcal{C}_n) \\
&\overset{(b)}{\le} n[-H(Z|X,A) + I(X;A) + \delta_\epsilon + I(U;X|A) + \delta_\epsilon + (I(V;X|U,A)
\end{aligned}
$$

$$
\begin{aligned}
&\quad - I(V;Y|U,A) + 2\delta_\epsilon) + H(Z|A,U) + \delta_\epsilon] \\
&\overset{(c)}{=} n[I(X;U,A,Z) + I(V;X|U,A) - I(V;Y|U,A) + \delta'_\epsilon] \\
&\overset{(c)}{=} n[I(X;V,A,Y) - (I(X;Y|U,A) - I(X;Z|U,A)) + \delta'_\epsilon] \leq n[\triangle + \delta'_\epsilon]
\end{aligned}
$$

if $\triangle \geq I(X;V,A,Y) - (I(X;Y|U,A) - I(X;Z|U,A))$, where (a) follows from the facts that conditioning reduces entropy, that (W_a, K, W_v) is a deterministic function of X^n from the encoding process, and that source sequence and side information channel are independent of the codebook, (b) follows from the codebook generation, from the memoryless properties of the source and the side information channel, from Lemma 5.3 in which we bound the term $H(Z^n|X^n, A^n)$, and from Lemma 5.4 in which we bound the term $H(Z^n|A^n, U^n, \mathcal{C}_n)$, and (c) follows from the Markov chain $(Y,Z) - (X,A) - V - U$.

From the random coding argument, we have that a tuple $(R, D, C, \triangle) \in \mathbb{R}_+^4$ which satisfies

$$
\begin{aligned}
&R \geq I(X;A) + I(V;X|Y,A), \\
&D \geq E\big[d\big(X, \tilde{g}(V,Y)\big)\big], \\
&C \geq E[\Lambda(A)], \\
&\triangle \geq I(X;V,A,Y) - (I(X;Y|U,A) - I(X;Z|U,A)),
\end{aligned}
$$

for some $P_{A|X} P_{V|X,A} P_{U|V}$ and a function $\tilde{g} : \mathcal{V} \times \mathcal{Y} \to \mathcal{X}$ is achievable.

5.A.2 Proof of Converse

For any achievable tuple $(R, D, C, \triangle)$, by standard properties of the entropy function, it follows that

$$
\begin{aligned}
&n(R + \delta_n) \geq \log |\mathcal{W}^{(n)}| \geq H(W) \\
&\overset{(a)}{=} H(W) + H(A^n|W) = H(A^n) + H(W|A^n) \\
&\geq [H(A^n) - H(A^n|X^n, Z^n)] + [H(W|A^n, Y^n) - H(W|A^n, X^n, Y^n, Z^n)] \\
&= [H(X^n, Z^n) - H(X^n, Z^n|A^n)] + [H(X^n, Z^n|A^n, Y^n) - H(X^n, Z^n|A^n, Y^n, W)] \\
&= H(X^n) + H(Z^n|X^n) - H(Y^n|A^n) + H(Y^n, Z^n|X^n, A^n) - H(Z^n|X^n, A^n) \\
&\qquad - H(X^n, Z^n|A^n, Y^n, W) \\
&\geq \sum_{i=1}^n H(X_i) - H(Y_i|A_i) + H(Y_i, Z_i|X_i, A_i) - H(X_i, Z_i|A^n, Y^n, W, X^{i-1}, Z^{i-1}) \\
&\overset{(b)}{=} \sum_{i=1}^n H(X_i) - H(Y_i|A_i) + H(Y_i|X_i, A_i, Z_i) + H(Z_i|X_i, A_i) \\
&\qquad - H(X_i, Z_i|A_i, Y_i, V_i)
\end{aligned}
$$

$$\geq \sum_{i=1}^{n} I(X_i; A_i) + I(V_i; X_i|Y_i, A_i),$$

where (a) follows from the deterministic action encoder and (b) follows by defining the auxiliary random variables $U_i \triangleq (W, A^{n\setminus i}, Y_{i+1}^n, Z^{i-1})$, and $V_i \triangleq (W, X^{i-1}, A^{n\setminus i}, Y^{n\setminus i}, Z^{i-1})$.

The leakage rate

$$\begin{aligned}
n(\triangle + \delta_n) &\geq I(X^n; W, Z^n) = I(X^n; W) + I(X^n; Z^n|W) \\
&\overset{(a)}{=} I(X^n; W, A^n) + I(X^n; Z^n|W, A^n) = H(X^n) - H(X^n|W, A^n, Y^n) \\
&\quad - I(X^n; Y^n|W, A^n) + I(X^n; Z^n|W, A^n) \\
&\overset{(b)}{=} \sum_{i=1}^{n} H(X_i) - H(X_i|W, A^n, Y^n, X^{i-1}) + H(Y_i|X_i, A_i) - H(Z_i|X_i, A_i) \\
&\quad - H(Y_i|W, A^n, Y_{i+1}^n) + H(Z_i|W, A^n, Z^{i-1}) \\
&\overset{(c)}{=} \sum_{i=1}^{n} H(X_i) - H(X_i|W, A^n, Y^n, X^{i-1}, Z^{i-1}) - I(X_i; Y_i|A_i) + H(Y_i|A_i) \\
&\quad + I(X_i; Z_i|A_i) - H(Z_i|A_i) - H(Y_i|W, A^n, Y_{i+1}^n) + H(Z_i|W, A^n, Z^{i-1}) \\
&\overset{(d)}{=} \sum_{i=1}^{n} H(X_i) - H(X_i|V_i, A_i, Y_i) - I(X_i; Y_i|A_i) + H(Y_i|A_i) + I(X_i; Z_i|A_i) \\
&\quad - H(Z_i|A_i) - H(Y_i|W_1, A^n, Y_{i+1}^n) + H(Z_i|W_1, A^n, Z^{i-1}) \\
&= \sum_{i=1}^{n} \underbrace{I(X_i; V_i, A_i, Y_i) - I(X_i; Y_i|A_i) + I(X_i; Z_i|A_i)}_{\triangleq P_i} + I(W, Y_{i+1}^n, A^{n\setminus i}; Y_i|A_i) \\
&\quad - I(W, Z^{i-1}, A^{n\setminus i}; Z_i|A_i),
\end{aligned}$$

where (a) follows from the deterministic action encoder, (b) from the Markov chain $(W, A^{n\setminus i}, X^{n\setminus i}, Y_{i+1}^n, Z^{i-1}) - (A_i, X_i) - (Y_i, Z_i)$, (c) from the Markov chain $(X_i, W, A_i^n, Y_i^n) - (A^{i-1}, X^{i-1}) - (Z^{i-1}, Y^{i-1})$, and (d) follows from the definition of V_i, and the deterministic action encoder.

Next, by adding $\sum_{i=1}^{n} I(Y_i; Z^{i-1}|A^n, W, Y_{i+1}^n) - I(Z_i; Y_{i+1}^n|A^n, W, Z^{i-1}) = 0$, which is similar to the Csiszár's sum identity (Lemma 2.5), to the right hand side and continuing the chain of inequalities, we get

$$\begin{aligned}
n(\triangle + \delta_n) &\geq \sum_{i=1}^{n} P_i + I(W, Y_{i+1}^n, Z^{i-1}, A^{n\setminus i}; Y_i|A_i) - I(W, Y_{i+1}^n, Z^{i-1}, A^{n\setminus i}; Z_i|A_i) \\
&\overset{(a)}{=} \sum_{i=1}^{n} I(X_i; V_i, A_i, Y_i) - I(X_i; Y_i|A_i) + I(X_i; Z_i|A_i) + I(U_i; Y_i|A_i) \\
&\quad - I(U_i; Z_i|A_i)
\end{aligned}$$

$$\overset{(b)}{=} \sum_{i=1}^{n} I(X_i; V_i, A_i, Y_i) - (I(X_i; Y_i|U_i, A_i) - I(X_i; Z_i|U_i, A_i)),$$

where (a) follows from definition of $U_i \triangleq (W, A^{n\setminus i}, Y_{i+1}^n, Z^{i-1})$, and (b) follows from the Markov chain $U_i - (A_i, X_i) - (Y_i, Z_i)$. Note that we have $U_i - V_i - (A_i, X_i) - (Y_i, Z_i)$ forms a Markov chain.

Let Q be a random variable uniformly distributed over the set $\{1, 2, \ldots, n\}$ and independent of X_i and (Y_i, Z_i) given (A_i, X_i), $1 \leq i \leq n$. We consider the joint distribution of new random variables (X, A, V, U, Y, Z), where $X \triangleq X_Q, Y \triangleq Y_Q, Z \triangleq Z_Q$, $A \triangleq A_Q$, $V \triangleq (Q, V_Q)$, and $U \triangleq (Q, U_Q)$. Note that we have $P_X = P_{X_Q}$, $P_{Y,Z|X,A} = P_{Y_Q,Z_Q|X_Q,A_Q}$ and $U - V - (A, X) - (Y, Z)$ forms a Markov chain, for $Q \in \{1, 2, \ldots, n\}$.

By introducing Q in above expressions, it is straightforward to show that rate and leakage constraints above can be bounded further by

$$\begin{aligned} R + \delta_n &\geq I(X; A) + I(X; V|Y, A) \\ \triangle + \delta_n &\geq I(X; V, A, Y) - (I(X; Y|U, A) - I(X; Z|U, A)). \end{aligned}$$

For the distortion constraint, we have

$$\begin{aligned} D + \delta_n &\geq E\big[d^{(n)}\big(X^n, g^{(n)}(W, Y^n)\big)\big] \overset{(a)}{=} \frac{1}{n}\sum_{i=1}^{n} E\big[d\big(X_i, \tilde{g}_i(V_i, Y_i)\big)\big] \\ &= \frac{1}{n}\sum_{i=1}^{n} E\big[d\big(X_Q, \tilde{g}_Q(V_Q, Y_Q)\big)\big|Q = i\big] \overset{(b)}{=} E\big[d\big(X, \tilde{g}(V, Y)\big)\big], \end{aligned}$$

where (a) follows from the definitions of the auxiliary random variable V_i, and that there exists some function $\tilde{g}_i$ such that $\tilde{g}_i(V_i, Y_i) = g_i^{(n)}(W, Y^n)$, and (b) follows from the definition of $\tilde{g}_Q(V_Q, Y_Q) \triangleq \tilde{g}(Q, V_Q, Y_Q) = \tilde{g}(V, Y)$.

Similarly, for the cost constraints,

$$C + \delta_n \geq E[\Lambda^{(n)}(A^n)] = \frac{1}{n}\sum_{i=1}^{n} E[\Lambda(A_i)] = E[\Lambda(A)].$$

For the bounds on the cardinalities of the sets $\mathcal{U}$ and$\mathcal{V}$, it can be shown by using the support lemma [CK11, Lemma 15.4] that $\mathcal{U}$ should have $|\mathcal{X}||\mathcal{A}| - 1$ elements to preserve $P_{X,A}$, plus three more for $H(X|V, Y, A)$, $I(X; Y|U, A) - I(X; Z|U, A)$, the distortion constraint. And similarly to [CK78, Appendix], the set $\mathcal{V}$ should have at most $(|\mathcal{X}||\mathcal{A}| + 2)(|\mathcal{X}||\mathcal{A}| + 1)$ elements. The proof is concluded by letting $n \to \infty$.

5.B Proof of Lemma 5.3

Let T be a binary random variable taking value 1 if (X^n, A^n) are jointly typical, and 0 otherwise. Note that $\Pr(T = 0) \leq \delta_\epsilon$ for n sufficiently large (provided the rate

constraints are satisfied). To bound the term $H(Z^n|X^n, A^n)$ for Z^n i.i.d. $\sim P_{Z|X,A}$, consider

$$\begin{aligned}
&H(Z^n|X^n, A^n) \geq H(Z^n|X^n, A^n, T) \\
&\geq (1-\delta_\epsilon) H(Z^n|X^n, A^n, T=1) \\
&\overset{(a)}{=} (1-\delta_\epsilon) \sum_{(x^n,a^n)} p(x^n, a^n|T=1) \sum_{i=1}^{n} H(Z_i|X_i = x_i, A_i = a_i) \\
&\overset{(b)}{=} (1-\delta_\epsilon) \sum_{(x^n,a^n)} p(x^n, a^n|T=1) \sum_{x,a} N(a,x) H(Z|X=x, A=a) \\
&\overset{(c)}{\geq} (1-\delta_\epsilon) \sum_{(x^n,a^n)} p(x^n, a^n|T=1) \sum_{x,a} np(a,x)(1-\epsilon) H(Z|X=x, A=a) \\
&\geq n(H(Z|X,A) - \delta'_\epsilon),
\end{aligned}$$

where (a) follows from memoryless property of the channel $P_{Z|X,A}$, (b) follows from the fact that $N(a,x)$ is the occurrences of $(x_i, a_i) = (x, a)$, and (c) follows from definition of joint typical set (Definition 2.7). Note that $\delta_\epsilon, \delta'_\epsilon \to 0$ as $\epsilon \to 0$, and $\epsilon \to 0$ as $n \to \infty$.

5.C Proof of Lemma 5.4

Consider the term $H(Z^n|A^n, U^n, \mathcal{C}_n)$. Let T be a binary random variable taking value 1 if (Z^n, A^n, U^n) are jointly typical, and 0 otherwise. Note that $\Pr(T=0) \leq \delta_\epsilon$ for n sufficiently large (provided the rate constraints are satisfied).

$$\begin{aligned}
&H(Z^n|A^n, U^n, \mathcal{C}_n) \leq H(Z^n|A^n, U^n, T) + H(T) \\
&\leq \Pr(T=0) \cdot H(Z^n|A^n, U^n, T=0) + \Pr(T=1) \cdot H(Z^n|A^n, U^n, T=1) + h(\delta_\epsilon) \\
&\leq n\delta_\epsilon \log|\mathcal{Z}| + H(Z^n|A^n, U^n, T=1) + h(\delta_\epsilon) \\
&= \sum_{(a^n,u^n)} p(a^n, u^n|T=1) H(Z^n|A^n = a^n, U^n = u^n, T=1) + n\delta_\epsilon \log|\mathcal{Z}| + h(\delta_\epsilon) \\
&\leq \sum_{(a^n,u^n)} p(a^n, u^n|T=1) \log|T_\epsilon^{(n)}(Z|a^n, u^n)| + n\delta_\epsilon \log|\mathcal{Z}| + h(\delta_\epsilon) \\
&\leq n(H(Z|U,A) + \delta'_\epsilon),
\end{aligned}$$

where the last inequality follows from properties of typical sequences (see, e.g., Theorem 2.1.2).

5.D Proof of the Rate-distortion-cost-leakage Region in (5.3)

Achievability: Recall that the side information at the decoder in this example is such that $Y = X$ if $A = 1$, and $Y = \emptyset$ if $A = 0$. Let the random variable A be independent of X where $\Pr(A = 1) = C$. Also, conditioned on $A = a$ for $a \in \{0, 1\}$, we model U and V as in Fig. 5.4, i.e., letting U be an output of a BSC(β) with input V, and letting V be an output of a BSC(α_a) with input X. Finally, we define a function $\tilde{g}(V, Y) = V$ if $A = 0$, and $\tilde{g}(V, Y) = Y = X$ otherwise. Then it follows from Theorem 5.2.1 that

$$\begin{aligned}(I) \qquad R &\geq I(X; A) + I(V; X|Y, A) \\ &\overset{(a)}{=} (1 - C) \cdot [H(X) - H(X|V, A = 0)] \\ &\overset{(b)}{=} (1 - C) \cdot [1 - h(\alpha_0)],\end{aligned}$$

where (a) follows since A is independent of X, and $Y = X$ if $A = 1$ and $Y = \emptyset$ if $A = 0$, and (b) follows since $X \sim$Bernoulli(0.5), and from the distribution of U and V in Fig. 5.4,

$$\begin{aligned}(II) \qquad D &\geq E[d(X, \tilde{g}(V, Y))] \\ &\overset{(a)}{=} (1 - C) \cdot E[d(X, V)|A = 0] \\ &= (1 - C) \cdot \Pr(X \neq V|A = 0) \\ &= (1 - C) \cdot \alpha_0,\end{aligned}$$

where (a) follows since given $A = 1$, $\tilde{g}(V, Y) = Y = X$, and

$$\begin{aligned}(III) \qquad \triangle &\geq I(X; V, Y, A) - [I(X; Y|U, A) - I(X; Z|U, A)] \\ &\overset{(a)}{=} 1 - \Pr(A = 0)[H(X|V, A = 0) - I(X; Z|U, A = 0)] \\ &\quad - \Pr(A = 1)[H(X|U, A = 1) - I(X; Z|U, A = 1)] \\ &\overset{(b)}{=} 1 - (1 - C) \cdot [h(\alpha_0) - h(p \star \alpha_0 \star \beta) + h(p)] \\ &\quad - C \cdot [h(\alpha_1 \star \beta) - h(p \star \alpha_1 \star \beta) + h(p)],\end{aligned}$$

where (a) follows since $Y = X$ if $A = 1$, and $Y = \emptyset$ if $A = 0$, and (b) follows from the Markov chain $Z - X - (A, U)$ and the distribution in Fig. 5.4.

Converse: For some cost $C = E[\Lambda(A)] = \Pr(A = 1)$, let $(R, D, \triangle)$ be an achievable tuple. We now prove that there exist α_0, α_1, and $\beta \in [0, 1/2]$ satisfying the inequalities in (5.3). From the converse proof of Theorem 5.2.1, we have the following bounds.

$$\begin{aligned}(I) \qquad R &\geq I(X; A) + I(V; X|Y, A) \\ &= (1 - C) \cdot [H(X) - H(X|A = 0) + I(V; X|Y, A = 0)]\end{aligned}$$

$$
\begin{aligned}
&+ C \cdot [H(X) - H(X|A=1) + I(V;X|Y,A=1)] \\
&\overset{(a)}{=} (1-C) \cdot [H(X) - H(X|V,A=0)] + C \cdot [H(X) - H(X|A=1)] \\
&\geq (1-C) \cdot [1 - H(X|V,A=0)],
\end{aligned}
$$

where (a) follows since A is independent of X, and $Y = X$ if $A = 1$ and $Y = \emptyset$ if $A = 0$. Since $0 \leq H(X|V,A=0) \leq H(X) = 1$, and $h(\cdot)$ is a continuous one-to-one mapping from $[0,1/2]$ to $[0,1]$, there exists $\alpha_0 \in [0,1/2]$ s.t. $H(X|V,A=0) = h(\alpha_0)$, and thus $R \geq (1-C) \cdot [1 - h(\alpha_0)]$.

$$
\begin{aligned}
(II) \qquad D &\geq E[d(X, \tilde{g}(V,Y))] \\
&\geq \Pr(A=0) E[d(X, \tilde{g}(V,Y))|A=0] \\
&= (1-C) \cdot \Pr(X \neq \tilde{g}(V,Y))|A=0).
\end{aligned}
$$

From Fano's inequality in (2.7) where $\Pr(X \neq \tilde{g}(V,Y)|A=0) \triangleq P_e$, we have that $H(X|V,Y,A=0) \leq H(X|\tilde{g}(V,Y),A=0) \leq h(P_e) + P_e \cdot \log(|\mathcal{X}|-1) = h(P_e)$. Since $h(\cdot)$ is an increasing function on $[0,1/2]$, we have that $\Pr(X \neq \tilde{g}(V,Y)|A=0) = P_e \geq h^{-1}(H(X|V,Y,A=0)) = h^{-1}(H(X|V,A=0)) = \alpha_0$, where the last equality follows from part (I). Thus, $D \geq (1-C) \cdot \alpha_0$.

$$
\begin{aligned}
(III) \qquad \triangle &\geq I(X;V,Y,A) - [I(X;Y|U,A) - I(X;Z|U,A)] \\
&\overset{(a)}{=} 1 - \Pr(A=0)[H(X|V,A=0) - I(X;Z|U,A=0)] \\
&\quad - \Pr(A=1)[H(X|U,A=1) - I(X;Z|U,A=1)] \\
&= 1 - (1-C) \cdot [h(\alpha_0) - H(Z|U,A=0) + h(p)] \\
&\quad - C \cdot [H(X|U,A=1) - H(Z|U,A=1) + h(p)], \qquad (5.10)
\end{aligned}
$$

where (a) follows since $Y = X$ if $A = 1$, and $Y = \emptyset$ if $A = 0$.

Now we use Mrs. Gerber's lemma [WZ73] to bound the remaining conditional entropy terms. Define a random variable $\hat{V}$ on $\{0,1\}$ as the output of a BSC(α_a) with input X conditioned on $A = a$, where $a = 0, 1$. Since X is uniformly distributed on $\{0,1\}$, X is also an output of a BSC(α_a) with input $\hat{V}$ conditioned on $A = a$, $a = 0, 1$, i.e., $\Pr(X = 1|\hat{V} = 0, A = a) = \alpha_a$. Note that this new variable $\hat{V}$ preserves the value $H(X|\hat{V},A=0) = h(\alpha_0)$.

Consider the term $H(X|U,A=1)$, where

$$
\begin{aligned}
H(X|U=u,A=1) &= h(\Pr(X=1|U=u,A=1)) \\
&= h\Big(\sum_{\hat{v}=0,1} \Pr(X=1|\hat{V}=\hat{v},U=u,A=1) \cdot \Pr(\hat{V}=\hat{v}|U=u,A=1)\Big) \\
&= h\Big(\sum_{\hat{v}=0,1} \Pr(X=1|\hat{V}=\hat{v},A=1) \cdot \Pr(\hat{V}=\hat{v}|U=u,A=1)\Big) \\
&= h(\alpha_1 \star \Pr(\hat{V}=1|U=u,A=1)).
\end{aligned}
$$

Since $\hat{V} \in \{0,1\}$, we have that $H(\hat{V}|U=u,A=1) = h(\Pr(\hat{V}=1|U=u,A=1))$. Thus, $\Pr(\hat{V}=1|U=u,A=1) = h^{-1}(H(\hat{V}|U=u,A=1))$, and $H(X|U=u,A=1) = h(\alpha_1 \star h^{-1}(H(\hat{V}|U=u,A=1)))$.

Next we consider the term $H(Z|U,A=a)$, $a=0,1$, where

$$\begin{aligned} H(Z|U=u,A=a) &= h(\Pr(Z=1|U=u,A=a)) \\ &= h\Big(\sum_{\hat{v},x\in\{0,1\}} \Pr(Z=1|X=x)\cdot p(x|\hat{v},a)\cdot p(\hat{v}|u,a)\Big) \\ &= h(p\star\alpha_a\star\Pr(\hat{V}=1|U=u,A=a)) \\ &= h(p\star\alpha_a\star h^{-1}(H(\hat{V}|U=u,A=a))). \end{aligned}$$

Substituting $H(X|U=u,A=1)$ and $H(Z|U=u,A=a)$ into (5.10), we get

$$\begin{aligned} \triangle \geq 1 - \sum_{u\in\mathcal{U}} &(1-C)\cdot \\ &[h(\alpha_0) - h(p\star\alpha_0\star h^{-1}(H(\hat{V}|U=u,A=0))) + h(p)]\cdot p(u|A=0) \\ &- \sum_{u\in\mathcal{U}} C\cdot[h(\alpha_1\star h^{-1}(H(\hat{V}|U=u,A=1))) \\ &\quad - h(p\star\alpha_1\star h^{-1}(H(\hat{V}|U=u,A=1))) + h(p)]\cdot p(u|A=1). \end{aligned}$$

Now, since $0 \leq H(\hat{V}|U=u,A=a) \leq H(\hat{V}) \leq 1$, and $h(\cdot)$ is a continuous one-to-one mapping from $[0,1/2]$ to $[0,1]$, there exist $\beta_{u,a} \in [0,1/2]$ s.t. $H(\hat{V}|U=u,A=a) = h(\beta_{u,a})$ for each $u\in\mathcal{U}$, $a\in\{0,1\}$. Thus,

$$\begin{aligned} \triangle &\geq 1 - \sum_{u_0\in\mathcal{U}} (1-C)\cdot[h(\alpha_0) - h(p\star\alpha_0\star\beta_{u_0,0}) + h(p)]\cdot p(u_0|A=0) \\ &\quad - \sum_{u_1\in\mathcal{U}} C\cdot[h(\alpha_1\star\beta_{u_1,1}) - h(p\star\alpha_1\star\beta_{u_1,1}) + h(p)]\cdot p(u_1|A=1) \\ &\overset{(a)}{\geq} 1 - (1-C)\cdot[h(\alpha_0) - h(p\star\alpha_0\star\beta_0) + h(p)] \\ &\quad - C\cdot[h(\alpha_1\star\beta_1) - h(p\star\alpha_1\star\beta_1) + h(p)] \\ &\overset{(b)}{=} 1 - (1-C)\cdot[h(\alpha_0) - h(p\star\alpha_0\star\beta_0) + h(p)] \\ &\quad - C\cdot[h(\alpha_1\star\beta_0) - h(p\star\alpha_1\star\beta_0) + h(p)], \end{aligned}$$

where (a) follows since there exist $u_0 = u_0^*$ and corresponding $\beta_{u_0^*,0} \triangleq \beta_0$, and $u_1 = u_1^*$ and corresponding $\beta_{u_1^*,1} \triangleq \beta_1$ s.t. the * terms are smaller than or equal to the averages, and (b) holds since the optimization does not change whether it is over $\alpha_1\star\beta_1$ or $\alpha_1\star\beta_0$.

5.E Proof of Theorem 5.3.1

The achievable scheme follows from a combination of layered (superposition-type) coding, Wyner-Ziv-coding, and Gel'fand-Pinsker coding. However, some care needs to be taken for the analysis of an achievable leakage rate when there is no information with positive rate sent over the rate-limited link. As in [PW11, Section III], we divide the proof into two main cases.

I) $I(X;U) - I(Y;U) \geq 0$

Key coding idea: The achievable scheme consists of two-stage Wyner-Ziv coding for communicating the codewords U^n and V^n over the rate-limited link to the decoder.

Codebook generation: Fix $P_{U|X}P_{A|U,X}P_{V|U,A,X}$, and a function $\tilde{g} : \mathcal{V} \times \mathcal{Y} \to \mathcal{X}$.

- Randomly and independently generate $2^{n(I(X;U)+\delta_\epsilon)}$ codewords $u^n(k)$ each according to $\sim \prod_{i=1}^n P_U(u_i), k \in [1 : 2^{n(I(X;U)+\delta_\epsilon)}]$, and distribute them uniformly at random into $2^{n(I(X;U)-I(Y;U)+2\delta_\epsilon)}$ equal-sized bins $b_U(w_u)$, where $w_u \in [1 : 2^{n(I(X;U)-I(Y;U)+2\delta_\epsilon)}]$.

- For each $k \in [1 : 2^{n(I(X;U)+\delta_\epsilon)}]$, randomly and conditionally independently generate $2^{n(I(X,A;V|U)+\delta_\epsilon)}$ codewords $v^n(k,l)$ each according to the distribution $\prod_{i=1}^n P_{V|U}(v_i|u_i(k)), l \in [1 : 2^{n(I(X,A;V|U)+\delta_\epsilon)}]$, and distribute them uniformly at random into $2^{n(I(X,A;V|U)-I(Y;V|U)+2\delta_\epsilon)}$ bins $b_V(k, w_v), w_v \in [1 : 2^{n(I(X,A;V|U)-I(Y;V|U)+2\delta_\epsilon)}]$.

The codebooks are revealed to the encoder, the action encoder, the decoder, and the eavesdropper.

Encoding:

- Given a source sequence x^n, the encoder looks for $u^n(k)$ which is jointly typical with x^n. Since there are more than $2^{nI(X;U)}$ codewords u^n generated, from the covering lemma (Lemma 2.6), we know that there exists such a u^n with high probability. If there are more than one, we choose one uniformly at random and send the corresponding bin index w_u over the rate-limited link. Then the action sequence a^n is generated randomly according to $\prod_{i=1}^n P_{A|U,X}(a_i|u_i(k), x_i)$.

- Next, the encoder looks for $v^n(k,l)$ which is jointly typical with (x^n, a^n, u^n). Since there are more than $2^{nI(X,A;V|U)}$ codewords v^n generated, from the covering lemma, we know that there exists such a v^n with high probability. If there are more than one, we choose one uniformly at random and send the corresponding bin index w_v over the rate-limited link.

Then the action-dependent side information (y^n, z^n) are generated as the output of the memoryless channel $P_{Y,Z|X,A}$.

Decoding:

- Upon receiving the indices (w_u, w_v) and the side information y^n, the decoder looks for the unique u^n which is jointly typical with y^n. Since there are less than $2^{nI(Y;U)}$ sequences in the bin $b_U(w_u)$, from the packing lemma (Lemma 2.7), it will find the unique and correct u^n with high probability.

- Then the decoder looks for the unique v^n which is jointly typical with (y^n, u^n). Since there are less than $2^{nI(Y;V|U)}$ sequences in the bin $b_V(k, w_v)$, from the packing lemma, it will find the unique and correct v^n with high probability.

- The decoder puts out $\hat{x}^n$ as a reconstruction of x^n, where the i^{th} element $\hat{x}_i = \tilde{g}(v_i, y_i)$.

Analysis of distortion and cost: As we have $(x^n, u^n, a^n, v^n, y^n)$ jointly typical, the average distortion and cost constraints are satisfied provided $E[d(X, \tilde{g}(V, Y))] \leq D$ and $E[\Lambda(A)] \leq C$.

Analysis of leakage rate at the eavesdropper: We consider the information leakage averaged over randomly chosen codebook. As there can be several ways of bounding the leakage term, we consider here two cases, namely, (Ia) "$I(Y;U)$ bound" and (Ib) "$I(Z;U)$ bound."

(Ia) "$I(Y;U)$ bound":

$$
\begin{aligned}
&I(X^n; W_u, W_v, Z^n | \mathcal{C}_n) \\
&= H(X^n|\mathcal{C}_n) - H(X^n|W_u, W_v, Z^n, \mathcal{C}_n) \\
&\leq H(X^n|\mathcal{C}_n) - H(X^n|W_u, Z^n, \mathcal{C}_n) + H(W_v|\mathcal{C}_n) \\
&\leq H(X^n|\mathcal{C}_n) - H(X^n|K, Z^n, \mathcal{C}_n) + H(W_v|\mathcal{C}_n) \\
&= H(X^n|\mathcal{C}_n) - H(X^n, K, Z^n|\mathcal{C}_n) + H(K|\mathcal{C}_n) + H(Z^n|K, \mathcal{C}_n) + H(W_v|\mathcal{C}_n) \\
&\overset{(a)}{\leq} -H(Z^n|X^n, U^n) + H(K|\mathcal{C}_n) + H(Z^n|U^n, \mathcal{C}_n) + H(W_v|\mathcal{C}_n) \\
&\overset{(b)}{\leq} n[-H(Z|X,U) + I(X;U) + \delta_\epsilon + H(Z|U) + \delta_\epsilon + I(X,A;V|U) \\
&\qquad - I(Y;V|U) + 2\delta_\epsilon] \\
&\overset{(c)}{=} n[I(Z;X|U) + I(Y;U) + I(X;U) - I(Y;U) + I(X,A;V|U,Y) + \delta'_\epsilon] \\
&\leq n[\triangle + \delta'_\epsilon]
\end{aligned}
$$

if $\triangle \geq I(Z;X|U) + I(Y;U) + [I(X;U) - I(Y;U) + I(X,A;V|U,Y)]$, where (a) follows from the facts that conditioning reduces entropy, that K is a deterministic function of X^n from the encoding process, and that side information channel is independent of the codebook, (b) follows from the codebook generation, from the memoryless properties of the side information channel where $A^n \sim \prod_{i=1}^n P_{A|U,X}$, from bounding the term $H(Z^n|X^n, U^n)$ where $Z^n \sim \prod_{i=1}^n P_{Z|X,U}$ whose proof is similar to that of Lemma 5.3, and from bounding the term $H(Z^n|U^n, \mathcal{C}_n)$ whose proof is similar to that of Lemma 5.4, and (c) follows from the Markov chain $Y - (X, A) - (U, V)$.

(Ib) "$I(Z;U)$ bound":

$$
\begin{aligned}
&I(X^n; W_u, W_v, Z^n|\mathcal{C}_n)\\
&= H(X^n|\mathcal{C}_n) - H(X^n|W_u, W_v, Z^n, \mathcal{C}_n)\\
&\le H(X^n|\mathcal{C}_n) - H(X^n|W_u, Z^n, \mathcal{C}_n) + H(W_v|\mathcal{C}_n)\\
&= H(X^n|\mathcal{C}_n) - H(X^n, W_u, Z^n|\mathcal{C}_n) + H(W_u|\mathcal{C}_n) + H(Z^n|W_u, \mathcal{C}_n) + H(W_v|\mathcal{C}_n)\\
&\overset{(a)}{\le} -H(Z^n|X^n, U^n) + H(W_u|\mathcal{C}_n) + H(Z^n|\mathcal{C}_n) + H(W_v|\mathcal{C}_n)\\
&\overset{(b)}{\le} n[-H(Z|X,U) + I(X;U) - I(Y;U) + \delta_\epsilon + H(Z) + \delta_\epsilon + (I(X,A;V|U)\\
&\qquad - I(Y;V|U) + 2\delta_\epsilon)]\\
&\overset{(c)}{=} n[I(Z;X|U) + I(Z;U) + I(X;U) - I(Y;U) + I(X,A;V|U,Y) + \delta'_\epsilon]\\
&\le n[\triangle + \delta'_\epsilon]
\end{aligned}
$$

if $\triangle \ge I(Z;X|U) + I(Z;U) + [I(X;U) - I(Y;U) + I(X,A;V|U,Y)]$, where (a), (b), and (c) follow similarly as those in (Ia) "$I(Y;U)$ bound" above.

II) $I(X;U) - I(Y;U) < 0$

Key coding idea: Under this case, we do not use binning for the first-stage codeword U^n. Instead, we utilize the fact that $I(X;U) - I(Y;U) < 0$ and that U^n can be reliably communicated over the channel $P_{Y|X,A}$ to generate an oversized codebook $\{U^n\}$ in order to reduce the rate needed for transmitting the source description over the rate-limited link in the second stage.

Codebook generation: Fix $P_{U|X}P_{A|U,X}P_{V|U,A,X}$ and a function $\tilde{g} : \mathcal{V} \times \mathcal{Y} \to \mathcal{X}$.

- Randomly and independently generate $2^{n(I(Y;U)-\delta_\epsilon)}$ codewords $u^n(j,k) \sim \prod_{i=1}^n P_U$, where $j \in [1 : 2^{n(I(Y;U)-I(X;U)-2\delta_\epsilon)}]$, and $k \in [1 : 2^{n(I(X;U)+\delta_\epsilon)}]$.
- For each (j,k), randomly and conditionally independently generate codewords $v^n(j,k,w,w')$ each according to $\prod_{i=1}^n P_{V|U}(v_i|u_i(j,k))$, where $w \in [1 : 2^{n([I(X,A;V|U)-I(Y;V|U)-I(Y;U)+I(X;U)]^+ + 4\delta_\epsilon)}]$, and $w' \in [1 : 2^{n(I(Y;V|U)-\delta_\epsilon)}]$.

The codebooks are revealed to the encoder, the action encoder, the decoder, and the eavesdropper.

Encoding:

- Given a source sequence x^n, for all $j \in [1 : 2^{n(I(Y;U)-I(X;U)-2\delta_\epsilon)}]$, the encoder looks for k which makes $u^n(j,k)$ jointly typical with x^n. Since for each j there are more than $2^{nI(X;U)}$ codewords u^n generated, by the covering lemma, there exists such a $u^n(j,k)$ and the corresponding k with high probability. If there are more than one such k, the encoder chooses one uniformly at random. Then for all $j \in [1 : 2^{n(I(Y;U)-I(X;U)-2\delta_\epsilon)}]$, the action sequences $a^n(j)$ are generated according to $a^n(j) \sim \prod_{i=1}^n P_{A|U,X}(a_i(j)|u_i(j,k), x_i)$. From the conditional typicality lemma [EK11, Chapter 2], we have that the tuple $(x^n, u^n(j,k)a^n(j))$ are jointly typical with high probability for all $j \in [1 : 2^{n(I(Y;U)-I(X;U)-2\delta_\epsilon)}]$.

Note here that more than one action sequences are generated, but only one will be selected and used for generating side information in the next step.

- Next, given that for all $j \in [1 : 2^{n(I(Y;U)-I(X;U)-2\delta_\epsilon)}]$, the tuple $(x^n, u^n(j,k), a^n(j))$ are jointly typical with the selected k, the encoder looks for (j, w, w') such that the codeword $v^n(j,k,w,w')$ is also jointly typical with the tuple $(x^n, u^n(j,k), a^n(j))$. It can be shown that the probability that there exists such triple (j, w, w') is arbitrarily high if $\frac{1}{n}\log|\mathcal{J}||\mathcal{W}||\mathcal{W}'| > I(X,A;V|U)+\delta_\epsilon$. The proof is given below.

$$
\begin{aligned}
&\Pr((X^n, U^n(j,K), A^n(j)) \in T_\epsilon^{(n)}(X,U,A), (X^n, U^n(j,K), A^n(j),\\
&\qquad V^n(j,K,w,w')) \notin T_\epsilon^{(n)}(X,U,A,V) \text{ for all } (j,w,w') \in \mathcal{J}\times\mathcal{W}\times\mathcal{W}')\\
&= \sum_k p(k) \sum_{x^n \in T_\epsilon^{(n)}(X)} p(x^n|k) \cdot \prod_j \sum_{(u^n,a^n)\in T_\epsilon^{(n)}(U,A|x^n)} p_{U^n(j,K),A^n(j)|X^n,K}(u^n,\\
&\qquad a^n|x^n,k) \cdot \Pr((x^n,u^n,a^n,V^n(j,K,w,w')) \notin T_\epsilon^{(n)}(X,U,A,V)\\
&\qquad \text{for all } (w,w') \in \mathcal{W}\times\mathcal{W}' | X^n = x^n, U^n(j,k) = u^n, A^n(j) = a^n, K = k)\\
&= \sum_k p(k) \sum_{x^n \in T_\epsilon^{(n)}(X)} p(x^n|k) \cdot \prod_j \sum_{(u^n,a^n)\in T_\epsilon^{(n)}(U,A|x^n)} p_{U^n(j,K),A^n(j)|X^n,K}(u^n,\\
&\qquad a^n|x^n,k) \cdot \prod_{w,w'} [1 - \Pr((x^n,u^n,a^n,V^n(j,k,w,w')) \in\\
&\qquad\qquad T_\epsilon^{(n)}(X,U,A,V)|U^n(j,k) = u^n)]\\
&\overset{(a)}{\le} \sum_k p(k) \sum_{x^n \in T_\epsilon^{(n)}(X)} p(x^n|k) \cdot \prod_j \sum_{(u^n,a^n)\in T_\epsilon^{(n)}(U,A|x^n)} p_{U^n(j,K),A^n(j)|X^n,K}(u^n,\\
&\qquad a^n|x^n,k) \cdot [1 - 2^{-n(I(X,A;V|U)+\delta_\epsilon)}]^{|\mathcal{W}||\mathcal{W}'|}\\
&\le [1 - 2^{-n(I(X,A;V|U)+\delta_\epsilon)}]^{|\mathcal{J}||\mathcal{W}||\mathcal{W}'|}\\
&\overset{(b)}{\le} \exp(-2^{n(\frac{1}{n}\log|\mathcal{J}||\mathcal{W}||\mathcal{W}'|-I(X,A;V|U)-\delta_\epsilon)}),
\end{aligned}
$$

where (a) follows from joint typicality lemma derived from Theorem 2.1.2, and (b) from the inequality $(1-x)^a \le \exp(-ax)$, for $0 \le x \le 1, a > 0$ [CT06, Lemma 10.5.3].

Since for each k, the number of codewords v^n generated are approximately $2^{n(I(Y;U)-I(X;U)+I(Y;V|U)+[I(X,A;V|U)-I(Y;V|U)-I(Y;U)+I(X;U)]^+)}$, there exists such (j,w,w') with high probability. If there are more than one triple, we choose one uniformly at random. The encoder sends the corresponding index w over the rate-limited link.

Then action-dependent side information (y^n, z^n) are generated as the output of the memoryless channel $P_{Y,Z|X,A}$ with input x^n and $a^n(j)$.

Decoding:

- Upon receiving the side information y^n, the decoder looks for the unique $u^n(j,k)$ which is jointly typical with y^n. Since there are less than $2^{nI(Y;U)}$ sequences u^n generated, by the packing lemma, it will find the unique and correct u^n with high probability.
- Then the decoder looks for the unique v^n which is jointly typical with (y^n, u^n). Based on the indices w and decoded index (j,k), there are less than $2^{nI(Y;V|U)}$ remaining sequences v^n, by the packing lemma, it will find the unique and correct v^n with high probability.
- The decoder puts out $\hat{x}^n$ as a reconstruction of x^n, where the i^{th} element $\hat{x}_i = \tilde{g}(v_i, y_i)$.

Analysis of distortion and cost: As we have $(x^n, u^n, a^n, v^n, y^n)$ jointly typical, the average distortion and cost constraints are satisfied provided $E[d(X, \tilde{g}(V,Y))] \leq D$ and $E[\Lambda(A)] \leq C$.

Analysis of leakage rate at the eavesdropper: As in the previous case, we consider two different ways of bounding the leakage averaged over randomly chosen codebook, namely, (IIa) "$I(Y;U)$ bound" and (IIb) "$I(Z;U)$ bound."

(IIa) "$I(Y;U)$ bound":

$$
\begin{aligned}
&I(X^n; W, Z^n|\mathcal{C}_n) \\
&= H(X^n|\mathcal{C}_n) - H(X^n|W, Z^n, \mathcal{C}_n) \\
&\leq H(X^n|\mathcal{C}_n) - H(X^n|J, K, W, Z^n, \mathcal{C}_n) \\
&\leq H(X^n|\mathcal{C}_n) - H(X^n|J, K, Z^n, \mathcal{C}_n) + H(W|\mathcal{C}_n) \\
&= H(X^n|\mathcal{C}_n) - H(X^n, J, K, W, Z^n|\mathcal{C}_n) + H(J, K|\mathcal{C}_n) + H(Z^n|J, K, \mathcal{C}_n) \\
&\quad + H(W|\mathcal{C}_n) \\
&\overset{(a)}{\leq} -H(Z^n|X^n, U^n) + H(J, K|\mathcal{C}_n) + H(Z^n|U^n, \mathcal{C}_n) + H(W|\mathcal{C}_n) \\
&\overset{(b)}{\leq} n[-H(Z|X,U) + I(Y;U) - \delta_\epsilon) + H(Z|U) + \delta_\epsilon + [I(X,A;V|U) - I(Y;V|U) \\
&\quad - I(Y;U) + I(X;U)]^+ + 4\delta_\epsilon] \\
&\overset{(c)}{=} n[I(Z;X|U) + I(Y;U) + [I(X;U) - I(Y;U) + I(X,A;V|U,Y)]^+ + \delta'_\epsilon] \\
&\leq n[\triangle + \delta'_\epsilon]
\end{aligned}
$$

if $\triangle \geq I(Z;X|U) + I(Y;U) + [I(X;U) - I(Y;U) + I(X,A;V|U,Y)]^+$, where (a) follows from the facts that conditioning reduces entropy and that side information channel is independent of the codebook, (b) follows from the codebook generation, from the memoryless properties of the side information channel where $A^n \sim \prod_{i=1}^n P_{A|U,X}$, from bounding the term $H(Z^n|X^n, U^n)$ where $Z^n \sim \prod_{i=1}^n P_{Z|X,U}$ whose proof is similar to that of Lemma 5.3, and from bounding the term $H(Z^n|U^n, \mathcal{C}_n)$ whose proof is similar to that of Lemma 5.4, and (c) follows from the Markov chain $Y - (X,A) - (U,V)$.

(IIb) "$I(Z;U)$ bound":

$$
\begin{aligned}
&I(X^n;W,Z^n|\mathcal{C}_n)\\
&= H(X^n|\mathcal{C}_n) - H(X^n|W,Z^n,\mathcal{C}_n)\\
&= H(X^n|\mathcal{C}_n) - H(X^n,W,Z^n|\mathcal{C}_n) + H(W|\mathcal{C}_n) + H(Z^n|W,\mathcal{C}_n)\\
&\overset{(a)}{\leq} -H(Z^n|X^n,U^n) + H(W|\mathcal{C}_n) + H(Z^n|\mathcal{C}_n)\\
&\overset{(b)}{\leq} n[-H(Z|X,U) + [I(X,A;V|U) - I(Y;V|U) - I(Y;U) + I(X;U)]^+ + 4\delta_\epsilon\\
&\qquad + H(Z) + \delta_\epsilon]\\
&\overset{(c)}{=} n[I(Z;X|U) + I(Z;U) + [I(X;U) - I(Y;U) + I(X,A;V|U,Y)]^+ + \delta'_\epsilon]\\
&\leq n[\triangle + \delta'_\epsilon]
\end{aligned}
$$

if $\triangle \geq I(Z;X|U) + I(Z;U) + [I(X;U) - I(Y;U) + I(X,A;V|U,Y)]^+$, where (a), (b), and (c) follow similarly as those in (IIa) "$I(Y;U)$ bound" above.

Based on all above results, from the random coding argument, we have that a tuple $(R,D,C,\triangle) \in \mathbb{R}_+^4$ which satisfies

$$
\begin{aligned}
&R \geq I(X;U) - I(Y;U) + I(X,A;V|U,Y),\\
&D \geq E\big[d\big(X,\tilde{g}(V,Y)\big)\big],\\
&C \geq E[\Lambda(A)],\\
&\triangle \geq I(Z;X|U) + \min\{I(Y;U), I(Z;U)\} + [I(X;U) - I(Y;U) + I(X,A;V|U,Y)]^+\\
&\triangle \geq I(Z;X|U) + \min\{I(Y;U), I(Z;U)\},
\end{aligned}
$$

for some $P_X P_{U,A,V|X} P_{Y,Z|A,X}$ and a function $\tilde{g} : \mathcal{V} \times \mathcal{Y} \to \mathcal{X}$ is achievable.

5.F Proof of Theorem 5.3.2

Here we only provide the proof for the rate and leakage rate constraints. The converse proofs for distortion and cost constraints follow similarly as in the proof of Theorem 5.2.1 . Let us define the auxiliary random variables $V_i \triangleq (W, Y^{n\setminus i})$, $U_i \triangleq (W, Y_{i+1}^n, Z^{i-1})$, and $K_i \triangleq (W, X^{n\setminus i}, Y_{i+1}^n, Z^{i-1})$ which satisfy $(Y_i,Z_i) - (X_i,A_i) - (V_i,U_i,K_i)$ for all $i = 1,\ldots,n$.

Rate constraint:

$$
\begin{aligned}
n(R+\delta_n) &\geq H(W)\\
&\geq H(W|Y^n)\\
&= H(W,X^n,Y^n) - H(Y^n) - H(X^n|W,Y^n)\\
&\geq H(X^n) + H(Y^n|X^n,A^n) - H(Y^n) - H(X^n|W,Y^n)\\
&\overset{(a)}{\geq} \sum_{i=1}^n H(X_i) + H(Y_i|X_i,A_i) - H(Y_i) - H(X_i|V_i,Y_i)
\end{aligned}
$$

$$= \sum_{i=1}^{n} I(X_i; A_i) - I(Y_i; A_i) + H(X_i|Y_i, A_i) - H(X_i|V_i, Y_i),$$

where (a) follows from the definition of V_i.

Leakage rate constraint:

$$\begin{aligned}
&n(\triangle + \delta_n) \geq I(X^n; W, Z^n) = I(X^n; W) + I(X^n; Z^n|W) \\
&= H(X^n) - H(X^n|W, Y^n) - I(X^n; Y^n|W) + I(X^n; Z^n|W) \\
&= \sum_{i=1}^{n} H(X_i) - H(X_i|W, Y^n, X^{i-1}) - H(Y_i|W, Y_{i+1}^n) + H(Z_i|W, Z^{i-1}) \\
&\quad + H(Y_i|W, X^n, Y_{i+1}^n) - H(Z_i|W, X^n, Z^{i-1}) \\
&= \sum_{i=1}^{n} H(X_i) - H(X_i|W, Y^n, X^{i-1}) + I(Y_i; W, Y_{i+1}^n) - I(Z_i; W, Z^{i-1}) \\
&\quad - I(Y_i; W, X^n, Y_{i+1}^n) + I(Z_i; W, X^n, Z^{i-1}) \\
&\overset{(a)}{=} \sum_{i=1}^{n} H(X_i) - H(X_i|W, Y^n, X^{i-1}) + I(Y_i; W, Y_{i+1}^n, Z^{i-1}) - I(Z_i; W, Y_{i+1}^n, Z^{i-1}) \\
&\quad - I(Y_i; W, X^n, Y_{i+1}^n, Z^{i-1}) + I(Z_i; W, X^n, Y_{i+1}^n, Z^{i-1}) \\
&\overset{(b)}{\geq} \sum_{i=1}^{n} I(X_i; V_i, Y_i) + I((Y_i; U_i) - I(Z_i; U_i) - I(Y_i; X_i, U_i, K_i) + I(Z_i; X_i, U_i, K_i) \\
&= \sum_{i=1}^{n} I(X_i; V_i, Y_i) - I(Y_i; X_i, K_i|U_i) + I(Z_i; X_i, K_i|U_i),
\end{aligned}$$

where (a) follows from adding the Csiszár's sum identity, $\sum_{i=1}^{n} I(Y_i; Z^{i-1}|W, Y_{i+1}^n) - I(Z_i; Y_{i+1}^n|W, Z^{i-1}) = 0$ and $\sum_{i=1}^{n} I(Y_i; Z^{i-1}|W, X^n, Y_{i+1}^n) - I(Z_i; Y_{i+1}^n|W, X^n, Z^{i-1}) = 0$, and (b) follows from the definition of V_i, U_i, and K_i.

In addition, we have

$$\begin{aligned}
n(\triangle + \delta_n) &\geq I(X^n; W, Z^n) \\
&\geq I(X^n; Z^n) \\
&\geq \sum_{i=1}^{n} H(X_i) - H(X_i|Z_i).
\end{aligned}$$

The proof ends by introducing an independent time-sharing variable Q and defining new corresponding random variables as in the converse proof of Theorem 5.2.1.

5.G Converse Proof of Leakage Constraint in Corollary 5.3.1

When there is no side information at the eavesdropper, the information leakage is expressed by $I(X^n;W)$. For any achievable leakage rate $\triangle$, we have that

$$\begin{aligned}
n(\triangle+\delta_n) &\geq I(X^n;W)\\
&= H(W)-H(W|X^n)\\
&\geq H(W|Y^n)-H(W|X^n)\\
&= H(W,X^n|Y^n)-H(X^n|W,Y^n)-H(W|X^n)\\
&\overset{(a)}{\geq} H(W,X^n,Y^n)-H(Y^n)-n\delta_n-H(W|X^n)\\
&\overset{(b)}{\geq} H(X^n)+H(Y^n|X^n,A^n)-H(Y^n)\\
&\geq \sum_{i=1}^{n} H(X_i)+H(Y_i|X_i,A_i)-H(Y_i)\\
&= \sum_{i=1}^{n} I(X_i;A_i)-I(Y_i;A_i)+H(X_i|A_i,Y_i),
\end{aligned}$$

where (a) follows from Fano's inequality $H(X^n|W,Y^n)\leq n\delta_n$, (b) follows from the fact that conditioning reduces entropy and that $Y^n-(X^n,A^n)-W$ forms a Markov chain.

5.H Proof of Proposition 5.4.1

The achievable scheme consists of a successive refinement like coding scheme and a secret key generation scheme.

Codebook Generation: For fixed $P_{A|X}, P_{U|X,A}$, and $\tilde{g}:\mathcal{U}\times\mathcal{A}\times\mathcal{Y}\to\hat{\mathcal{X}}$,

- Randomly and independently generate $2^{n(I(X;A)+\delta_\epsilon)}$ $a^n(w_a)\sim\prod_{i=1}^{n}P_A(a_i)$, $w_a\in[1:2^{n(I(X;A)+\delta_\epsilon)}]$.
- For each $w_a\in[1:2^{n(I(X;A)+\delta_\epsilon)}]$, randomly and conditionally independently generate $2^{n(I(X;U|A)+\delta_\epsilon)}$ $u^n(w_a,m)\sim\prod_{i=1}^{n}P_{U|A}(u_i|a_i(w_a))$. Then distribute them uniformly at random into $2^{n(I(X;U|A,Y)+2\delta_\epsilon)}$ bins and label these bins $b_U(w_u)$, $w_u\in[1:2^{n(I(X;U|A,Y)+2\delta_\epsilon)}]$. We further split the bin indices w_u into $w_{u,k}\in[1:2^{nR_k}]$ and $w_{u,l}\in[1:2^{n(I(X;U|A,Y)-R_k+2\delta_\epsilon)}]$. Note that w_u can be deduced from $(w_{u,k},w_{u,l})$.
- For secret key generation codebook, we randomly and uniformly partition the set of sequences $\mathcal{Y}^n$ into 2^{nR_k} bins $b_K(k)$, $k\in[1:2^{nR_k}]$, where $R_k=\min\{H(Y|X,A,Z),I(X;U|A,Y)\}-2\delta_\epsilon$.

The codebooks are revealed to the encoder, the action encoder, the decoder, and the eavesdropper.

Encoding:

- Given a sequence x^n, the encoder looks for a^n that is jointly typical with x^n. By the covering lemma, with high probability, there exists such an a^n since there are $2^{n(I(X;A)+\delta_\epsilon)}$ codewords a^n generated. If there are more than one, we choose one uniformly at random and send the corresponding index w_a to the decoder and the action encoder.

- Then, the encoder looks for u^n which is jointly typical with (x^n, a^n). By the covering lemma, with high probability, there exists such a u^n since there are $2^{n(I(X;U|A)+\delta_\epsilon)}$ u^n generated. If there are more than one, we choose one uniformly at random.

- The side information is generated as an output of the memoryless channel $P_{Y|X,A}$ with inputs x^n and a^n. To generate a secret key, the encoder looks for an index k for which $y^n \in b_K(k)$. Further the encoder transmits the encrypted bin index corresponding to the chosen u^n by splitting and sending $w_{u,l}$ and $w_{u,k} \oplus k$ to the decoder, where $w_{u,k} \oplus k$ denotes the modulo operation, $(w_{u,k} + k) \mathrm{mod} 2^{nR_k}$. This results in the total additional rate of $I(X;U|A,Y)+2\delta_\epsilon$ over the rate-limited link.

Decoding: Upon receiving w_a, $w_{u,l}$, and $w_{u,k} \oplus k$, the decoder uses its side information y^n to generate its own key and decrypt the index $w_{u,k}$, and thus the bin index w_u. Then it looks for a unique u^n that is jointly typical with (a^n, y^n). By the packing lemma, with high probability, it will find the unique and correct one since there are $2^{n(I(Y;U|A)-\delta_\epsilon)}$ codewords in each bin $b_U(w_u)$. The decoder puts out $\hat{x}^n$ where $\hat{x}_i = \tilde{g}(u_i, a_i, y_i)$.

Analysis of distortion and cost: Since (x^n, a^n, y^n, u^n) are jointly typical, we can show that D and C satisfying $D \geq E[d(X, \tilde{g}(U, A, Y))]$ and $C \geq E[\Lambda(A)]$ are achievable.

Analysis of leakage rate: The leakage averaged over randomly chosen codebook $\mathcal{C}_n$ can be bounded as follows.

$$
\begin{aligned}
&I(X^n; W_a, W_{u,l}, W_{u,k} \oplus K, Z^n | \mathcal{C}_n) \\
&= I(X^n; W_a|\mathcal{C}_n) + I(X^n; Z^n|W_a, \mathcal{C}_n) + I(X^n; W_{u,l}|W_a, Z^n, \mathcal{C}_n) \\
&\quad + I(X^n; W_{u,k} \oplus K|W_a, W_{u,l}, Z^n, \mathcal{C}_n) \\
&\leq H(W_a|\mathcal{C}_n) + H(Z^n|W_a, \mathcal{C}_n) - H(Z^n|W_a, X^n, \mathcal{C}_n) + H(W_{u,l}|\mathcal{C}_n) \\
&\quad + H(W_{u,k} \oplus K|\mathcal{C}_n) - H(W_{u,k} \oplus K|W_a, W_{u,l}, X^n, Z^n, \mathcal{C}_n) \\
&\overset{(a)}{\leq} H(W_a|\mathcal{C}_n) + H(Z^n|A^n, \mathcal{C}_n) - H(Z^n|X^n, A^n, \mathcal{C}_n) + H(W_{u,l}|\mathcal{C}_n) \\
&\quad + H(W_{u,k} \oplus K|\mathcal{C}_n) - H(K|X^n, A^n, Z^n, \mathcal{C}_n)
\end{aligned}
$$

$$
\begin{aligned}
&\overset{(b)}{\leq} n[\underbrace{I(X;A)+H(Z|A)+I(X;U|A,Y)}_{\triangleq P}-R_k+R_k+\delta'_\epsilon]-H(Z^n|X^n,A^n,\mathcal{C}_n)\\
&\quad -H(K|X^n,A^n,Z^n,\mathcal{C}_n)\\
&\leq n[P+\delta'_\epsilon]-H(Z^n|X^n,A^n,\mathcal{C}_n)-I(K;Y^n|X^n,A^n,Z^n,\mathcal{C}_n)\\
&= n[P+\delta'_\epsilon]-H(Y^n,Z^n|X^n,A^n,\mathcal{C}_n)+H(Y^n|K,X^n,A^n,Z^n,\mathcal{C}_n)\\
&\overset{(c)}{\leq} n[P-H(Y,Z|X,A)+\delta''_\epsilon]+H(Y^n|K,X^n,A^n,Z^n,\mathcal{C}_n)\\
&\overset{(d)}{\leq} n[I(X;A)+H(Z|A)+I(X;U|A,Y)-H(Y,Z|X,A)+H(Y|X,A,Z)-R_k\\
&\quad +\delta'''_\epsilon]\\
&= n[I(X;A,Z)+I(X;U|A,Y)-R_k+\delta'''_\epsilon]\\
&\leq n[\triangle+\delta'''_\epsilon]
\end{aligned}
$$

if $\triangle \geq I(X;A,Z)+I(X;U|A,Y)-R_k$, where (a) follows from the facts that conditioned on $\mathcal{C}_n$, the indices $W_a, W_{u,l}, W_{u,k}$ are functions of X^n, that A^n is a function of W_a, and that $W_a-(X^n,A^n)-Y^n-Z^n$ forms a Markov chain, (b) follows from the codebook generation and from bounding the term $H(Z^n|A^n,\mathcal{C}_n)$ similarly as in Lemma 5.4, (c) follows from bounding the term $H(Y^n,Z^n|X^n,A^n,\mathcal{C}_n)$ where $(Y^n,Z^n)\sim\prod_{i=1}^n P_{Y,Z|X,A}$, similarly as in Lemma 5.3, and from the Markov chain $(X,A)-Y-Z$, and (d) follows from [EK11, Lemma 22.3] given that $R_k < H(Y|X,A,Z)-\delta_\epsilon$ which holds since $R_k=\min\{H(Y|X,A,Z),I(X;U|A,Y)\}-2\delta_\epsilon$.

That is, from the random coding argument, we have that a tuple $(R,D,C,\triangle)$ which satisfies $R\geq I(X;A)+I(X;U|A,Y)$, $D\geq E[d(X,\tilde{g}(U,A,Y))]$, $C\geq E[\Lambda(A)]$, and $\triangle\geq I(X;A,Z)+I(X;U|A,Y)-R_k$, for some $P_{A|X}P_{U|A,X}$ and $\tilde{g}:\mathcal{U}\times\mathcal{A}\times\mathcal{Y}\to\hat{\mathcal{X}}$, where $R_k=\min\{H(Y|X,A,Z),I(X;U|A,Y)\}$, is achievable.

5.I Proof of Proposition 5.4.2

For any achievable tuple $(R,D,C,\triangle)$, by standard properties of the entropy function, it follows that

$$
\begin{aligned}
n(R+\delta_n) &\geq H(W)\\
&\overset{(a)}{=} H(W,A^n)\\
&= H(A^n)+H(W|A^n)\\
&\geq I(X^n;A^n)+I(X^n;W|A^n,Y^n)\\
&= H(X^n)-H(X^n|A^n)+H(X^n|A^n,Y^n)-H(X^n|W,A^n,Y^n)\\
&= H(X^n)-H(Y^n|A^n)+H(Y^n|X^n,A^n)-H(X^n|W,A^n,Y^n)\\
&\overset{(b)}{\geq} \sum_{i=1}^n H(X_i)-H(Y_i|A_i)+H(Y_i|X_i,A_i)-H(X_i|U_i,A_i,Y_i)
\end{aligned}
$$

$$= \sum_{i=1}^{n} I(X_i; A_i) + I(X_i; U_i | A_i, Y_i),$$

where (a) follows from the deterministic action encoder, and (b) follows from the memoryless channel $P_{Y|X,A}$ and by defining $U_i \triangleq (W, X^{i-1}, A^{n\setminus i}, Y^{n\setminus i})$.

For the average distortion and cost constraint, we have

$$D + \delta_n \geq E[d^{(n)}(X^n, g^{(n)}(W, Y^n))] \overset{(a)}{=} \frac{1}{n} \sum_{i=1}^{n} E[d(X_i, \tilde{g}_i(U_i, A_i, Y_i))],$$

$$C + \delta_n \geq E[\Lambda(A^n)] = \frac{1}{n} \sum_{i=1}^{n} E[\Lambda(A_i)],$$

where (a) follows from the definitions of auxiliary random variable U_i, and that there exists some function $\tilde{g}_i$ such that $\tilde{g}_i(U_i, A_i, Y_i) = g_i^{(n)}(W, Y^n)$.

Lastly, the leakage rate

$$\begin{aligned}
n(\triangle + \delta_n) &\geq I(X^n; W, Z^n) \\
&\overset{(a)}{=} I(X^n; W, A^n, Z^n) \\
&= I(X^n; A^n, Z^n) + I(X^n; W | A^n, Z^n) \\
&= I(X^n; A^n, Z^n) + I(X^n; W, Y^n | A^n, Z^n) - I(X^n; Y^n | W, A^n, Z^n) \\
&= I(X^n; A^n, Z^n) + I(X^n; Y^n | A^n, Z^n) + I(X^n; W | A^n, Y^n, Z^n) \\
&\quad - I(X^n; Y^n | W, A^n, Z^n) \\
&\overset{(b)}{\geq} I(X^n; A^n, Z^n) + I(X^n; Y^n | A^n, Z^n) + H(X^n | A^n, Y^n) - H(X^n | W, A^n, Y^n) \\
&- I(X^n; Y^n | W, A^n, Z^n) \\
&\geq I(X^n; A^n, Z^n) + H(X^n | A^n, Y^n) - H(X^n | W, A^n, Y^n) - H(Y^n | X^n, A^n, Z^n) \\
&\overset{(b)}{=} H(X^n) - I(X^n; Y^n | A^n, Z^n) - H(X^n | W, A^n, Y^n) - H(Y^n | X^n, A^n, Z^n) \\
&= H(X^n) - H(Y^n | A^n, Z^n) - H(X^n | W, A^n, Y^n) \\
&\geq \sum_{i=1}^{n} H(X_i) - H(Y_i | A_i, Z_i) - H(X_i | W, X^{i-1}, A^n, Y^n) \\
&\overset{(c)}{=} \sum_{i=1}^{n} I(X_i; A_i, Z_i) + I(X_i; U_i | A_i, Y_i) - H(Y_i | X_i, A_i, Z_i),
\end{aligned}$$

where (a) follows from the deterministic action encoder, and (b) follows from the Markov chain $(X^n, A^n) - Y^n - Z^n$, and (c) from the Markov chain $Z_i - Y_i - (X_i, A_i)$ and the definition of U_i.

The proof for $n(\triangle + \delta_n) \geq \sum_{i=1}^{n} I(X_i; A_i, Z_i)$ is straightforward, and is therefore omitted. We end the proof by introducing an independent time-sharing variable Q and defining new corresponding random variables as in the converse proof of Theorem 5.2.1.

5.J Proof of Lemma 5.1

By definition of the reconstruction function, we get $g(u) = \hat{x}(\cdot|u)$. Then we obtain $E[d(X, g(U))] = \sum_{x\in\mathcal{X},u\in\mathcal{U}} p(x,u)d(x,\hat{x}(\cdot|u)) = \sum_{x\in\mathcal{X},u\in\mathcal{U}} p(x,u)\log(\frac{1}{p(x|u)}) = H(X|U)$.

5.K Proof of Lemma 5.2

By definition of the reconstruction alphabet, we consider the reconstruction $\hat{X}^n$ to be a probability distribution on $\mathcal{X}^n$ conditioned on Z. In particular, if $\hat{x}^n = g^{(n)}(z)$, we define $s(x^n|z) = \prod_{i=1}^n \hat{x}_i(x_i|z)$. Note that s is a probability mass function on $\mathcal{X}^n$. We obtain the following bound on the expected distortion conditioned on $Z = z$,

$$
\begin{aligned}
&E[d^{(n)}(X^n, g^{(n)}(z))|Z = z] = E[\frac{1}{n}\sum_{i=1}^n d(X_i, g_i^{(n)}(z))|Z = z] \\
&= \sum_{x^n\in\mathcal{X}^n} p(x^n|z)\frac{1}{n}\sum_{i=1}^n d(x_i, g_i^{(n)}(z)) \\
&= \sum_{x^n\in\mathcal{X}^n} p(x^n|z)\frac{1}{n}\sum_{i=1}^n \log(\frac{1}{\hat{x}_i(x_i|z)}) \\
&= \frac{1}{n}\sum_{x^n\in\mathcal{X}^n} p(x^n|z)\log(\frac{1}{s(x^n|z)}) \\
&= \frac{1}{n}\sum_{x^n\in\mathcal{X}^n} p(x^n|z)\log(\frac{p(x^n|z)}{s(x^n|z)}\cdot\frac{1}{p(x^n|z)}) \\
&= \frac{1}{n}\sum_{x^n\in\mathcal{X}^n} p(x^n|z)\log(\frac{p(x^n|z)}{s(x^n|z)}) + \frac{1}{n}\sum_{x^n\in\mathcal{X}^n} p(x^n|z)\log(\frac{1}{p(x^n|z)}) \\
&= \frac{1}{n}D(p(x^n|z)||s(x^n|z)) + \frac{1}{n}H(X^n|Z = z) \\
&\geq \frac{1}{n}H(X^n|Z = z).
\end{aligned}
$$

By averaging both sides over all $z \in \mathcal{Z}$, from the law of total expectation, we obtain the desired result.

5.L Proof of Theorem 5.4.1

The proof of Theorem 5.4.1 is an application of the proofs of Proposition 5.4.1 and 5.4.2.

5.L.1 Achievability

I) If $H(X|A,Y) - D < 0$: The encoder looks for a^n that is jointly typical with x^n and transmits the corresponding index w_a to the decoder. With high probability, there exists such a^n since there are $2^{n(I(X;A)+\delta_\epsilon)}$ codewords generated. The action-dependent side information y^n is then generated. Since the decoder knows (a^n, y^n), it can put out $\hat{x}^n$ where $\hat{x}_i = \tilde{g}(a_i, a_i, y_i)$. Since (x^n, a^n, y^n) are jointly typical, it can be shown that this is sufficient to satisfy the distortion and cost constraints. Note that, under the logarithmic loss distortion, we have $E[d(X, \tilde{g}(A, A, Y))] = H(X|A,Y)$.

II) If $H(X|A,Y) - D > 0$: We follow exactly the proof of Proposition 5.4.1 except that we specify a particular choice of U in the following way. From the property of log-loss distortion function (Lemma 5.1), we have that $E[d(X, \tilde{g}(U, A, Y))] = H(X|U,A,Y)$. We define $U = X$ with probability $p = 1 - \frac{D}{H(X|A,Y)}$ and a constant otherwise. Then, $H(X|U,A,Y) = (1-p)H(X|A,Y) = D$. The inner bound in Proposition 5.4.1 thus reduces to the desired result.

That is, we have that a tuple $(R, D, C, \triangle)$ which satisfies $R \geq I(X;A) + H(X|A,Y) - D$, $C \geq E[\Lambda(A)]$, and $\triangle \geq I(X;A,Z) + H(X|A,Y) - D - R_k$, for some $P_{A|X}$, where $R_k = \min\{H(Y|X,A,Z), H(X|A,Y) - D\}$, is achievable.

5.L.2 Converse

The converse proof follows similarly as the proof of Proposition 5.4.2 except the step that we define the auxiliary random variable U_i. Instead, we utilize the fact that, for logarithmic loss distortion, $\frac{1}{n}H(X^n|W,Y^n) \leq E[d^{(n)}(X^n, g^{(n)}(W,Y^n))]$ (Lemma 5.2). Thus, for any achievable tuple $(R, D, C, \triangle)$, it follows that $H(X^n|W,Y^n) \leq n(D + \delta_n)$. The proof for $n(R + \delta_n) \geq \sum_{i=1}^n I(X_i; A_i)$ is straightforward.

Chapter 6

Secure Source Coding With Public Helper

In the previous chapter, we have studied problems of secure source coding with action-dependent side information where the source secrecy constraint is imposed at an external eavesdropper. In this chapter, we consider new models for secure multi-terminal source coding in the presence of a public helper. Two main scenarios are studied: 1) source coding with a helper where the coded side information from the helper is eavesdropped by an external eavesdropper; 2) triangular source coding with a helper where the helper is considered as a public terminal. We are interested in how the helper can support the source transmission subject to a constraint on the amount of information leaked due to its public nature. We characterize the fundamental tradeoff between transmission rate, incurred distortion, and information leakage rate at the helper/eavesdropper in the form of the rate-distortion-leakage region for various classes of problems.

6.1 Introduction

Nowadays the Internet is an essential part of our daily life. We rely on many online services which inevitably create huge amounts of information flows in a network. With this huge amount of information, the main tasks for network designers are to ensure that the data can be transmitted reliably and also securely across the network. The latter requirement is becoming increasingly acute, especially when sensitive information is involved. Let us imagine a network in which information flows from one node to another through a number of intermediate nodes. The system design generally makes use of these intermediate nodes to help the transmission. However, these nodes might be public devices or terminals which we cannot fully trust. This scenario leads to a natural tradeoff between cooperation and secrecy in a system and motivates the study of secure communication and compression in the presence of public helpers.

In this chapter, we consider secure lossy source coding problems involving a public helper under an information leakage constraint. A summary of the problem settings and contributions of this chapter is given below (see also Table 6.1).

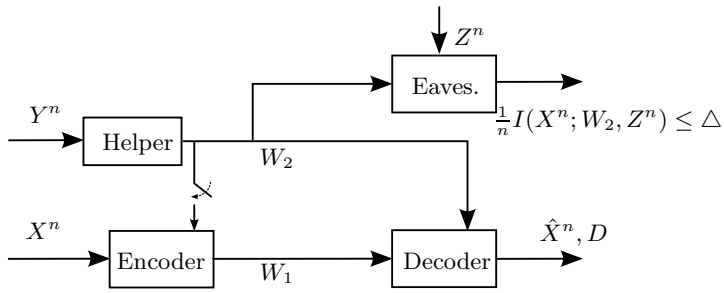

Figure 6.1: Secure source coding with one-sided/two-sided public helper. (© 2013 IEEE)

6.1.1 Overview of Problem Settings and Organization

Secure Source Coding with a Public Helper

In Section 6.2, we consider secure lossy source coding problems with a public helper, as depicted in Fig. 6.1. The setting is motivated by a scenario where the helper can only provide side information through a rate-limited communication link which is not secure due to its *public* nature, i.e., it can be eavesdropped by an external eavesdropper. In the "one-sided helper" setting, the helper communicates through a public link only to the decoder, while in the "two-sided helper" case, the helper *broadcasts* the same coded side information to both encoder and decoder. We provide an inner bound to the rate-distortion-leakage region for the one-sided helper case and show that it is tight under the logarithmic loss distortion measure and for the Gaussian case with quadratic distortion and the Markov relation $Y - X - Z$. A Gaussian example illustrating the distortion-leakage tradeoff under different rate constraints for the one-sided helper is also given. For the two-sided helper case, we solve the rate-distortion-leakage tradeoff under general distortion. We note that the one-sided/two-sided helper settings considered in Fig. 6.1 are essentially extensions of the one-helper problem [BHO$^+$79, JB08, PSW10] to include the presence of an eavesdropper. Variation of the settings where the eavesdropper sees instead the link from an encoder to a decoder was studied in [TUR13, VP13].

Secure Triangular/Cascade Source Coding with a Public Helper

In Section 6.3, we consider problems of triangular source coding with a public helper, as shown in Fig. 6.2. In contrast to the previous settings, where the focus is on leakage at an external eavesdropper, we address the problem of information leakage at a legitimate user. The setting is motivated by a scenario where the helper is a public terminal that forwards the information as the protocol requests from the encoder to the decoder. However, the helper might be curious and not ignore the data which may not be intended for him. Clearly, there exists a tradeoff between amount of information leakage to the helper and the helper's ability to support

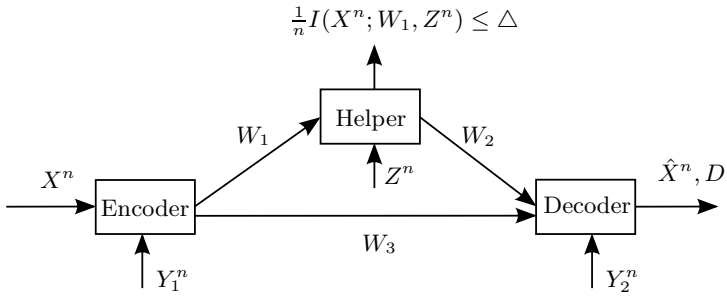

Figure 6.2: Secure triangular/cascade source coding with public helper. (© 2013 IEEE)

the source transmission. The problem of characterizing the optimal rate-distortion-leakage tradeoff in general is difficult due to the ambiguity of the helper's strategy and the role of side information at the encoder. In Section 6.3, we characterize the rate-distortion-leakage regions for various special cases based on different *side information patterns* available to the encoder, the helper, and the decoder. Our contributions are summarized below.

- *Setting (A): We assume that* Y_1 *is constant,* $Y_2 = Y$*, and that* $X-Y-Z$ *forms a Markov chain. We solve the problem under the logarithmic loss distortion and for the Gaussian sources with quadratic distortion, and show that the forwarding scheme at the helper (setting* $W_2 = W_1$*) is optimal. Note that the Markov assumption* $X - Y - Z$ *in this setting can be relevant in scenarios where the decoder is a fusion center collecting all correlated side information.*

- *Setting (B): We assume that the side information* $Y_1 = Y_2 = Y$*, and that* $X - Y - Z$ *forms a Markov chain. We solve the problem under the logarithmic loss distortion and for the Gaussian source with quadratic distortion. We show that the availability of the side information at the encoder does not improve the rate-distortion tradeoff, and that the forwarding scheme at the helper is optimal. Interestingly, we note that although the availability of the side information at the encoder does not improve the rate-distortion tradeoff, this side information can be used for a secret key generation at the encoder and the decoder. In our coding scheme, the secret key is used to scramble part of the message sent to the helper, and thus decrease the information leakage.*

- *Setting (C): We assume that* (Y_1, Z) *is constant and* $Y_2 = Y$*. It can be seen that the setting essentially reduces to the Wyner-Ziv like problem with an additional leakage constraint, and that the forwarding scheme at the helper is optimal. The Wyner-Ziv like coding achieves the whole rate-distortion-leakage region in this case.*

- *Setting (D): We assume that the side information at the helper is also available at the encoder, i.e.,* $Y_1 = Z$*, and we let* $Y_2 = Y$*. In this case, we assume*

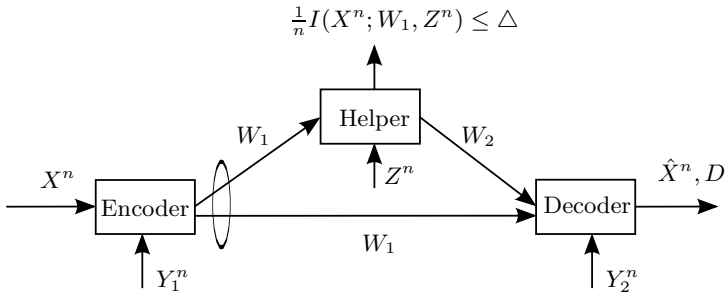

Figure 6.3: Secure triangular source coding with public helper where the encoder "broadcasts" the same message.

that $X - Z - Y$ forms a Markov chain and solve the problem under general distortion. Due to $X - Z - Y$, we show that the decode-and-reencode type scheme at the helper is optimal. That is, it is meaningful to take into account Z^n at the helper in relaying information to the decoder.

We note that our settings are different from the conventional triangular source coding problem in that the decoding constraint at the helper is replaced by the privacy constraint. Also, all the triangular settings can reduce to the cascade settings when the private link from the encoder to the decoder is removed (setting W_3 constant). The results for the cascade settings can therefore be obtained straightforwardly from the triangular settings.

Apart from the triangular settings (A)-(D) mentioned above, we also consider a slightly different setting where the encoder "broadcasts" the same source description to the helper and the decoder, as depicted in Fig. 6.3. Note that this is not a special case of previous triangular settings in Fig. 6.2 since it is more restrictive than simply setting the rate $R_1 = R_3$. Depending on the side information pattern, we will see that, in some cases (setting (A)-(C)), the helper is not helpful in terms of providing more information to the decoder, i.e., $R_2 \geq 0$ is achievable.

6.1.2 Related Work

Here we discuss some related works on multiterminal source coding problems and refer the readers to the related work section in Chapter 5 for the discussion regarding the secure source coding problems. Multi-terminal source coding problems have been studied extensively in various settings. The lossless distributed source coding problem was solved by Slepian and Wolf [SW73]. The setting has since been extended to the lossy case (cf., e.g., [Ber77]), and remains open in general. It is also unsolved even when we require only one distortion constraint, i.e., one-helper problem [BHO$^+$79], [JB08]. There exist only a few special cases which can be solved completely. This includes the Wyner-Ziv problem [WZ76], the case when one source

	Helper uses public link		Triangular/Cascade with Public Helper			
Distortion	one-sided	two-sided	(A)	(B)	(C)	(D)
general distortion	?	✓	?	?	✓	✓(X-Z-Y)
logarithmic loss	✓	✓	✓(X-Y-Z)	✓(X-Y-Z)	-	-
quadratic (Gaussian)	✓(Y-X-Z)	✓(Y-X-Z)	✓(X-Y-Z)	✓(X-Y-Z)	-	-

Table 6.1: Summary of contributions; the check mark ✓ denotes the cases that are solved in this chapter, while the question mark ? denotes the cases that are left open; the Markov assumption on the side information is stated if any.

is decoded losslessly [BY89], and the special case of a Gaussian source with quadratic distortion [Ooh97], [WTV08]. Recently, Courtade and Weissman [CW14] introduced a logarithmic loss distortion as a new and interesting distortion measure for lossy distributed source coding and solved the problem completely under this distortion measure. As for other related multi-terminal source coding problems, Yamamoto in [Yam81] studied and established the rate-distortion regions for the cascade and triangular source coding problems without side information. Variations of the cascade and triangular source coding settings have been studied in recent years (see, e.g., [VTD06, CPW12]).

6.2 Secure Source Coding with One-sided/Two-sided Public Helper

In this section, we consider source coding with a helper where the link from the helper is public and can therefore be eavesdropped by an external eavesdropper, as depicted in Fig. 6.1. We state the detailed problem formulation below. For the one-sided public helper case, we provide an inner bound to the rate-distortion-leakage region. Then we show that the inner bound is tight for some special cases including the cases under the logarithmic loss distortion measure, and the Gaussian case with quadratic distortion under $Y - X - Z$ Markov assumption. We also consider the two-sided helper case and characterize the rate-distortion-leakage region under a general distortion function.

6.2.1 One-sided Public Helper

Let us consider the setting in Fig. 6.1 when the switch is open. Source, side information, and reconstruction alphabets, $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \hat{\mathcal{X}}$ are assumed to be finite. Let (X^n, Y^n, Z^n) be the n-length sequences which are i.i.d. according to $P_{X,Y,Z}$. Given a source sequence X^n, an encoder generates a source description $W_1 \in \mathcal{W}_1^{(n)}$ and sends it over the noise-free, rate-limited link to a decoder. Meanwhile, a helper who observes the side information Y^n generates coded side information $W_2 \in \mathcal{W}_2^{(n)}$ and sends it to the decoder over another noise-free, rate-limited link. Given the source description and the coded side information, the decoder reconstructs the source sequence as $\hat{X}^n$ subject to a distortion constraint. We note that the eavesdropper also receives the coded side information and its own side information Z^n.

Definition 6.1. A $(|\mathcal{W}_1^{(n)}|, |\mathcal{W}_2^{(n)}|, n)$-code for secure source coding with one-sided public helper consists of

- a stochastic encoder $F_1^{(n)}$ which takes X^n as input and generates $W_1 \in \mathcal{W}_1^{(n)}$ according to a conditional PMF $p(w_1|x^n)$,
- a stochastic helper $F_2^{(n)}$ which takes Y^n as input and generates $W_2 \in \mathcal{W}_2^{(n)}$ according to $p(w_2|y^n)$, and

- a decoder $g^{(n)} : \mathcal{W}_1^{(n)} \times \mathcal{W}_2^{(n)} \to \hat{\mathcal{X}}^n$,

where $\mathcal{W}_1^{(n)}$ and $\mathcal{W}_2^{(n)}$ are finite sets.

Let $d : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty)$ be the single-letter distortion measure. The distortion between the source sequence and its reconstruction at the decoder is defined as

$$d^{(n)}\big(X^n, \hat{X}^n\big) \triangleq \frac{1}{n} \sum_{i=1}^{n} d(X_i, \hat{X}_i),$$

where $d^{(n)}(\cdot)$ is the distortion function.

The information leakage at the eavesdropper who has access to W_2 and Z^n is measured by the normalized mutual information $\frac{1}{n} I(X^n; W_2, Z^n)$.

Definition 6.2. A rate-distortion-leakage tuple $(R_1, R_2, D, \triangle) \in \mathbb{R}_+^4$ is said to be *achievable* if for any $\delta > 0$ and all sufficiently large n there exists a $(|\mathcal{W}_1^{(n)}|, |\mathcal{W}_2^{(n)}|, n)$ code such that

$$\frac{1}{n} \log \big|\mathcal{W}_i^{(n)}\big| \leq R_i + \delta, \ i = 1, 2,$$

$$E[d^{(n)}(X^n, g^{(n)}(W_1, W_2))] \leq D + \delta,$$

$$\text{and} \quad \frac{1}{n} I(X^n; W_2, Z^n) \leq \triangle + \delta.$$

The *rate-distortion-leakage region* $\mathcal{R}_{\text{one-sided}}$ is the set of all achievable tuples.

Inner Bound

The following theorem gives an inner bound to the region $\mathcal{R}_{\text{one-sided}}$, i.e., it defines region $\mathcal{R}_{\text{in}} \subseteq \mathcal{R}_{\text{one-sided}}$.

Theorem 6.2.1 (Inner Bound)**.** *A tuple $(R_1, R_2, D, \triangle) \in \mathbb{R}_+^4$ is achievable for secure source coding with one-sided public helper if*

$$R_2 \geq I(Y; U), \tag{6.1a}$$

$$R_1 \geq I(X; V|U), \tag{6.1b}$$

$$D \geq E[d(X, \tilde{g}(U, V))], \tag{6.1c}$$

$$\triangle \geq I(X; U, Z), \tag{6.1d}$$

for some joint distributions of the form $P_{X,Y,Z}(x, y, z) P_{U|Y}(u|y) P_{V|X}(v|x)$ with $|\mathcal{U}| \leq |\mathcal{Y}| + 4$, $|\mathcal{V}| \leq |\mathcal{X}| + 1$, and a function $\tilde{g} : \mathcal{U} \times \mathcal{V} \to \mathcal{X}$.

Region $\mathcal{R}_{in}$ is defined as the convex hull of the set of all such tuples.

Remark 6.1. Our achievable scheme is identical to that of the original one-helper problem. However, the resulting rate-distortion tradeoff is different since the set of optimizing input distributions may change due to an additional leakage constraint.

We note also that the problem of characterizing the complete rate-distortion-leakage region under general distortion remains open. This is to be expected in view of the fact that the one-helper problem (without the leakage rate constraint) is still open.

Proof of Theorem 6.2.1: The proof is based on a random coding argument. The achievable scheme follows the standard rate-distortion and Wyner-Ziv like coding scheme. That is, for fixed $P_{U|Y}, P_{V|X}$, and $\tilde{g}(\cdot)$, randomly and independently generate $2^{n(I(Y;U)+\delta_\epsilon)}$ sequences $u^n(w_2) \sim \prod_{i=1}^n P_U(u_i(w_2))$, $w_2 \in [1 : 2^{n(I(Y;U)+\delta_\epsilon)}]$. Also, randomly and independently generate $2^{n(I(X;V)+\delta_\epsilon)}$ sequences $v^n(\tilde{w}) \sim \prod_{i=1}^n P_V(v_i(\tilde{w}))$, $\tilde{w} \in [1 : 2^{n(I(X;V)+\delta_\epsilon)}]$, and distribute them uniformly into $2^{n(I(X;V|U)+2\delta_\epsilon)}$ bins $b_v(w_1)$, $w_1 \in [1 : 2^{n(I(X;V|U)+2\delta_\epsilon)}]$. For encoding, the helper looks for u^n that is jointly typical with y^n. If there is more than one, it selects one of them uniformly at random. If there is no such u^n, it selects one out of $2^{n(I(Y;U)+\delta_\epsilon)}$ uniformly at random. Then it transmits the corresponding index w_2 to the decoder. With high probability, there exists such u^n since there are $2^{n(I(Y;U)+\delta_\epsilon)}$ codewords generated. The encoder looks for v^n that is jointly typical with x^n. If there is more than one, it selects one of them uniformly at random. If there is no such v^n, it selects one out of $2^{n(I(X;V)+\delta_\epsilon)}$ uniformly at random. It then transmits the corresponding bin index w_1 to the decoder. With high probability, there exists such v^n since there are $2^{n(I(X;V)+\delta_\epsilon)}$ codewords generated. Upon receiving (w_1, w_2), the decoder looks for the unique v^n such that it is jointly typical with u^n. With high probability, it will find the unique and correct one since there are $2^{n(I(U;V)-\delta_\epsilon)}$ codewords in each bin $b_v(w_1)$. Then $\hat{x}^n$ is put out as a source reconstruction, where $\hat{x}_i = \tilde{g}(u_i, v_i)$. Since (x^n, u^n, v^n) are jointly typical, we can show that $D \geq E[d(X, \tilde{g}(U, V))]$ is achievable.

As for the analysis of leakage rate, we consider the following bound on the normalized mutual information averaged over all codebooks $\mathcal{C}_n$,

$$
\begin{aligned}
&I(X^n; W_2, Z^n|\mathcal{C}_n) \\
&= H(X^n|\mathcal{C}_n) - H(X^n, W_2, Z^n|\mathcal{C}_n) + H(W_2, Z^n|\mathcal{C}_n) \\
&= H(X^n|\mathcal{C}_n) - H(X^n, Z^n|\mathcal{C}_n) - H(W_2|X^n, Z^n, \mathcal{C}_n) + H(W_2|\mathcal{C}_n) + H(Z^n|W_2, \mathcal{C}_n) \\
&\leq H(X^n|\mathcal{C}_n) - H(X^n, Z^n|\mathcal{C}_n) - I(W_2; Y^n|X^n, Z^n, \mathcal{C}_n) + H(W_2|\mathcal{C}_n) \\
&\quad + H(Z^n|W_2, \mathcal{C}_n) \\
&\overset{(a)}{\leq} H(X^n) - H(X^n, Y^n, Z^n) + H(Y^n|U^n(W_2), X^n, Z^n, \mathcal{C}_n) + H(W_2|\mathcal{C}_n) \\
&\quad + H(Z^n|U^n(W_2), \mathcal{C}_n) \\
&\overset{(b)}{\leq} n[H(X) - H(X, Y, Z) + H(Y|U, X, Z) + \delta_\epsilon + I(Y; U) + \delta_\epsilon + H(Z|U) + \delta_\epsilon] \\
&\overset{(c)}{=} n[H(X) - H(X, Z|Y, U) - I(Y; X, Z|U) + H(Z|U) + \delta'_\epsilon] \\
&= n[I(X; U, Z) + \delta'_\epsilon],
\end{aligned}
$$

where (a) follows from the fact that (X^n, Y^n, Z^n) are independent of the codebook, (b) follows from the i.i.d. property of (X^n, Y^n, Z^n), from the codebook

generation that we have $W_2 \in [1 : 2^{n(I(Y;U)+\delta_\epsilon)}]$, and from bounding the terms $H(Y^n|U^n(W_2), X^n, Z^n, \mathcal{C}_n)$ and $H(Z^n|U^n(W_2), \mathcal{C}_n)$ similarly as in Lemma 5.4 in Chapter 5, where $\Pr((Y^n, U^n(W_2), X^n, Z^n) \in \mathcal{T}_\epsilon^{(n)}) \to 1$ as $n \to \infty$ from the codebook generation and encoding process, (c) follows from the Markov chain $U - Y - (X, Z)$.

For the bounds on the cardinalities of the sets $\mathcal{U}$ and $\mathcal{V}$, it can be shown by using the support lemma [CK11, Lemma 15.4] that it suffices that $\mathcal{U}$ should have $|\mathcal{Y}|-1$ elements to preserve P_Y, plus five more for $H(Y|U), H(X|U), H(X|U,V), H(X|U,Z)$, and the distortion constraint. And similarly, it suffices that $\mathcal{V}$ should have at most $|\mathcal{X}|+1$ elements to preserve P_X, $H(X|U,V)$, and the distortion constraint. This finally concludes the proof. □

Next, we show that the inner bound provided in Theorem 6.2.1 is tight for some special cases, namely, the setting under logarithmic loss distortion measure and the Gaussian setting under quadratic distortion for the case $Y - X - Z$.

Logarithmic Loss Distortion

Theorem 6.2.2 (Logarithmic Loss)**.** *The rate-distortion-leakage region under logarithmic loss distortion $\mathcal{R}_{\text{one-sided,log-loss}}$ is the set of all tuples $(R_1, R_2, D, \triangle) \in \mathbb{R}_+^4$ that satisfy*

$$R_2 \geq I(Y;U), \tag{6.2a}$$
$$R_1 \geq [H(X|U) - D]^+, \tag{6.2b}$$
$$\triangle \geq I(X;U,Z), \tag{6.2c}$$

for some joint distributions of the form $P_{X,Y,Z}(x,y,z)P_{U|Y}(u|y)$ with $|\mathcal{U}| \leq |\mathcal{Y}|+2$.

Proof of Theorem 6.2.2:
Sketch of Achievability: The achievable proof follows the proof of the inner bound in Theorem 6.2.1. That is, the scheme consists of the rate-distortion code for lossy transmission of y^n via the codeword u^n at rate $I(Y;U)+\delta_\epsilon$, and the Wyner-Ziv code at rate $I(X;V|U)+2\delta_\epsilon$ for lossy transmission of x^n with u^n as side information at the decoder. We can show that the distortion D and the leakage $\triangle$, satisfying $D \geq E[d(X, \tilde{g}(U,V))], \triangle \geq I(X;U,Z)$, are achievable. Due to the property of logarithmic loss distortion function (Lemma 5.1), we have $E[d(X, \tilde{g}(U,V))] = H(X|U,V)$. If $H(X|U) < D$, the encoder does not need to send anything, i.e., setting V constant. If $H(X|U) > D$, we define $V = X$ with probability $1 - \frac{D}{H(X|U)}$ and constant otherwise. Then we get $H(X|U,V) = D$ and $I(X;V|U) = H(X|U) - D$. Therefore, we obtain the desired achievable rate-distortion-leakage expressions. The converse proof uses the fact that for logarithmic loss distortion function $E[d(X^n, g^{(n)}(W_1, W_2))] \geq \frac{1}{n}H(X^n|W_1, W_2)$ (Lemma 5.2), and it is given in Appendix 6.A. □

Remark 6.2. Interestingly, Theorem 6.2.2 shows that the achievable scheme for the original one-helper problem (the one used in Theorem 6.2.1) is also optimal in

the presence of an eavesdropper who observes the link from the helper. Due to the property of logarithmic loss distortion which allows the use of Lemmas 5.1 and 5.2 in proving achievability and converse, an additional auxiliary random variable V and its associated Markov chain are not needed in characterizing the rate-distortion-leakage region. This essentially allows us to overcome the common issue faced in establishing a complete result for the lossy multi-terminal source coding problem in general.

Gaussian Setting Under Quadratic Distortion With $Y - X - Z$

In this part, we evaluate the rate-distortion regions when (X^n, Y^n, Z^n) are jointly Gaussian and the distortion function is quadratic. Let the sequences (X^n, Y^n, Z^n) be i.i.d. according to $P_{X,Y,Z}$. We will assume that $Y \sim \mathcal{N}(0, \sigma_Y^2)$, $X = Y + N_1, N_1 \sim \mathcal{N}(0, \sigma_{N_1}^2)$ independent of Y, and $Z = X + N_2, N_2 \sim \mathcal{N}(0, \sigma_{N_2}^2)$ independent of (X, Y, N_1), where $\sigma_Y^2, \sigma_{N_1}^2, \sigma_{N_2}^2 > 0$. Note that this satisfies the Markov assumption $Y - X - Z$. While our main results in previous cases were proven only for discrete memoryless sources, the extension to the quadratic Gaussian case is standard and it follows, for example, [Wyn78] and [EK11].

Theorem 6.2.3 (Gaussian, $Y - X - Z$)**.** *The rate-distortion-leakage region for a Gaussian source with quadratic distortion under the Markov assumption $Y - X - Z$, $\mathcal{R}_{\text{one-sided, Gaussian}}$, is the set of all tuples $(R_1, R_2, D, \triangle) \in \mathbb{R}_+^4$ that satisfy*

$$R_2 \geq \frac{1}{2} \log \big(1/\alpha\big), \tag{6.3a}$$

$$R_1 \geq \frac{1}{2} \log \big(\frac{\alpha \sigma_Y^2 + \sigma_{N_1}^2}{D}\big), \tag{6.3b}$$

$$\triangle \geq \frac{1}{2} \log \big(\frac{(\sigma_Y^2 + \sigma_{N_1}^2)(\alpha \sigma_Y^2 + \sigma_{N_1}^2 + \sigma_{N_2}^2)}{\sigma_{N_2}^2 (\alpha \sigma_Y^2 + \sigma_{N_1}^2)}\big), \tag{6.3c}$$

for some $\alpha \in (0, 1)$.

Proof. The proof is given in Appendix 6.B. □

Corollary 6.2.1. *The minimum achievable distortion for given rates and leakage rate $R_1, R_2, \triangle$ under the Markov assumption $Y - X - Z$ is given by*

$$D_{min}(R_1, R_2, \triangle) = \max\{2^{-2R_1}(2^{-2R_2}\sigma_Y^2 + \sigma_{N_1}^2), 2^{-2R_1}(\alpha^* \sigma_Y^2 + \sigma_{N_1}^2)\}, \tag{6.4}$$

where $0 \leq \alpha^* = \frac{2^{-2\triangle}(\sigma_Y^2 + \sigma_{N_1}^2)(\sigma_{N_1}^2 + \sigma_{N_2}^2) - \sigma_{N_1}^2 \sigma_{N_2}^2}{\sigma_Y^2 \sigma_{N_2}^2 - 2^{-2\triangle} \sigma_Y^2 (\sigma_Y^2 + \sigma_{N_1}^2)} = \frac{\sigma_{N_2}^2}{\left(\frac{2^{2\triangle} \sigma_{N_2}^2}{(\sigma_Y^2 + \sigma_{N_1}^2)} - 1\right) \sigma_Y^2} - \frac{\sigma_{N_1}^2}{\sigma_Y^2} \leq 1$, *and*

$1/2 \log(1 + \frac{\sigma_Y^2 + \sigma_{N_1}^2}{\sigma_{N_2}^2}) \leq \triangle \leq 1/2 \log(\frac{(\sigma_Y^2 + \sigma_{N_1}^2)(\sigma_{N_1}^2 + \sigma_{N_2}^2)}{\sigma_{N_1}^2 \sigma_{N_2}^2})$.

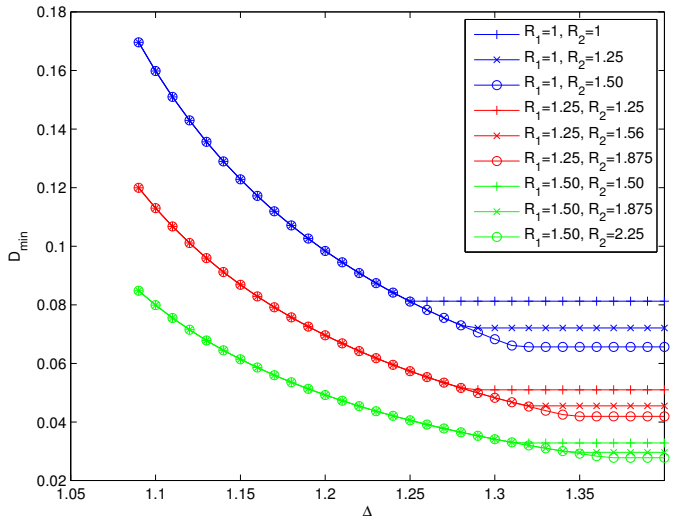

Figure 6.4: Gaussian example: Minimum achievable distortion as a function of leakage rate for given rates R_1, R_2.

Proof. The proof follows from the result in Theorem 6.2.3 where we solve for D, and the fact that the leakage rate $\triangle$ is lower- and upper-bounded by $I(X;Z)$ and $I(X;Y,Z)$, respectively, i.e., $1/2\log(1+\frac{\sigma_Y^2+\sigma_{N_1}^2}{\sigma_{N_2}^2}) = I(X;Z) \leq \triangle \leq I(X;Y,Z) = 1/2\log(\frac{(\sigma_Y^2+\sigma_{N_1}^2)(\sigma_{N_1}^2+\sigma_{N_2}^2)}{\sigma_{N_1}^2\sigma_{N_2}^2})$. □

Example 1

We evaluate the minimum achievable distortion for given rates and leakage rate in Corollary 6.2.1. For fixed $\sigma_Y^2 = 0.5, \sigma_{N_1}^2 = \sigma_{N_2}^2 = 0.2$, we plot $D_{\min}$ as a function of $\triangle$ for given R_1 and R_2 in Fig 6.4.

We can see that, in general, for given R_1 and R_2, $D_{\min}$ is decreasing when $\triangle$ becomes larger. This is because the helper is able to transmit more information to the decoder without violating the leakage constraint. However, there exists a $\triangle^*$ such that for any $\triangle > \triangle^*$ we cannot improve $D_{\min}$ further by increasing $\triangle$ since it is limited by the rate R_2. This saturation effect can be seen from the expression of $D_{\min}$ as a max function in (6.4). That is, for given R_1, R_2, when $\triangle$ is sufficiently large, we get $D_{\min} = 2^{-2R_1}(2^{-2R_2}\sigma_Y^2 + \sigma_{N_1}^2)$ which is constant. In fact, we can determine the value of $\triangle^*$ from (6.4) by solving for $\triangle$ in the equation $2^{-2R_2} = \alpha^*$. We note that $\triangle^*$ depends only on R_2 as seen also from Fig. 6.4 that when $R_2 = 1.25$, we get the same $\triangle^*$ for different R_1, e.g., $R_1 = 1$ or 1.25. We note that $D_{\min}$ still depends on R_1, i.e., it is saturated at a lower level for larger R_1. To this end, we conclude that at high $\triangle$ region, R_2 is a limiting factor of $D_{\min}$.

On the other hand, when $\triangle$ is "small," the decreasing region is active, i.e., $D_{\min} = 2^{-2R_1}(\alpha^* \sigma_Y^2 + \sigma_{N_1}^2)$, and $D_{\min}$ depends only on R_1 and $\triangle$ (not on R_2). That is, in the "small" $\triangle$ region, $D_{\min}$ is limited by $\triangle$ so that we cannot improve $D_{\min}$ by increasing R_2 further. This can be seen from the plots that, for a given R_1, three distortion-leakage curves with different R_2 coincide in the small $\triangle$ region.

6.2.2 Two-sided Public Helper

Let us consider the setting in Fig. 6.1 when the switch is closed. Since the problem setting is similar to that of the one-sided helper case, details are omitted. The main difference is that the coded side information $W_2 \in \mathcal{W}_2^{(n)}$ is given to both the encoder and the decoder. Then, based on X^n and W_2, the encoder generates the source description $W_1 \in \mathcal{W}_1^{(n)}$. That is, the encoding function becomes $F_1^{(n)}$ that takes (X^n, W_2) as input and generates W_1 according to $p(w_1|x^n, w_2)$. In the following, we characterize the rate-distortion-leakage region under general distortion function for the secure source coding with two-sided public helper problem.

Theorem 6.2.4. *The rate-distortion-leakage region $\mathcal{R}_{\textit{two-sided}}$ is the set of all tuples $(R_1, R_2, D, \triangle) \in \mathbb{R}_+^4$ that satisfy*

$$R_2 \geq I(Y;U), \tag{6.5a}$$

$$R_1 \geq I(X;\hat{X}|U), \tag{6.5b}$$

$$D \geq E[d(X,\hat{X})], \tag{6.5c}$$

$$\triangle \geq I(X;U,Z), \tag{6.5d}$$

for some joint distributions of the form $P_{X,Y,Z}(x,y,z)P_{U|Y}(u|y)P_{\hat{X}|X,U}(\hat{x}|x,u)$ with $|\mathcal{U}| \leq |\mathcal{Y}| + 3$.

Proof. The achievable scheme consists of the rate-distortion code for lossy transmission of y^n via u^n at rate $I(Y;U) + \delta_\epsilon$. Since w_2 is given to both the encoder and the decoder, source coding with side information known at both encoder and decoder at rate $I(X;\hat{X}|U) + 2\delta_\epsilon$ is used for lossy transmission of x^n with u^n as side information. The achievable leakage rate proof and the converse proof follow similarly as that of one-sided helper case, and are therefore omitted. Similarly as in [PSW10], we note that by following the converse proof of one-sided helper case, we in fact proved the outer bound which has the same rate, distortion, and leakage rate constraints as in (6.5d), but with the joint distribution satisfying $U - Y - (X,Z)$ and $\hat{X} - (U,X,Y) - Z$. Clearly this outer bound includes the achievable region due to the larger set of distributions. To show that the outer bound is also included in the achievable region, we let $(R_1, R_2, D, \triangle)$ be in the outer bound with the joint distribution of the form

$$\bar{p}(x,y,z,u,\hat{x}) = p(x,y,z)p(u|y)\bar{p}(\hat{x}|u,x,y). \tag{6.6}$$

Then we show that there exists a distribution of the form satisfying the Markov conditions in the achievable region such that the constraints on $(R_1, R_2, D, \triangle)$ in (6.5d) hold.

Let

$$p(x, y, z, u, \hat{x}) = p(x, y, z)p(u|y)\bar{p}(\hat{x}|u, x), \tag{6.7}$$

where $\bar{p}(\hat{x}|u, x)$ is induced by $\bar{p}(x, y, z, u, \hat{x})$. We now show that the terms $I(Y; U)$, $I(X; \hat{X}|U)$, $E[d(X, \hat{X})]$, and $I(X; U, Z)$ are the same whether we evaluate over $\bar{p}(x, y, z, u, \hat{x})$ in (6.6) or $p(x, y, z, u, \hat{x})$ in (6.7), and thus $(R_1, R_2, D, \triangle)$ is also in the achievable region. To do that, we show that the marginal distributions $\bar{p}(x, y, z, u)$ and $\bar{p}(x, u, \hat{x})$ induced by $\bar{p}(x, y, z, u, \hat{x})$ are equal to $p(x, y, z, u)$ and $p(x, u, \hat{x})$ induced by $p(x, y, z, u, \hat{x})$. By summing over $\hat{x}$ in (6.6) and (6.7), we have $\bar{p}(x, y, z, u) = p(x, y, z, u)$. To show that $\bar{p}(x, u, \hat{x}) = p(x, u, \hat{x})$, we consider $\bar{p}(x, u, \hat{x}) = \bar{p}(x, u)\bar{p}(\hat{x}|x, u)$. Note that, by summing over $(y, z, \hat{x})$ in (6.6) and (6.7), we get $\bar{p}(x, u) = p(x, u)$. Also, $p(\hat{x}|x, u) = \bar{p}(\hat{x}|x, u)$ since $p(\hat{x}|x, u)$ is the induced $\bar{p}(\hat{x}|x, u)$ by construction. Thus, we conclude that $\bar{p}(x, u, \hat{x}) = p(x, u, \hat{x})$. See also [PSW10] for more details. For the bound on the cardinality of the set $\mathcal{U}$, it can be shown by using the support lemma [CK11, Lemma 15.4] that it suffices that $\mathcal{U}$ should have $|\mathcal{Y}| - 1$ elements to preserve P_Y, plus four more for $H(Y|U), I(X; \hat{X}|U), H(X|U, Z)$, and the distortion constraint. □

Remark 6.3. For the cases of logarithmic loss distortion and the Gaussian source with quadratic distortion specified before, it can be shown that the rate-distortion-leakage regions for the corresponding two-sided helper cases remain the same as those of the one-sided helper cases. This is a reminiscence of the well-known result in the Wyner-Ziv source coding problem with Gaussian source and quadratic distortion that the side information Y^n at the encoder does not improve the rate-distortion function, i.e., $R_{X|Y}(D) = R^{WZ}(D) = \frac{1}{2}\log(\frac{\text{var}(X|Y)}{D})$ [Wyn78]. In our case, to prove the achievability, we simply neglect the coded side information at the encoder and achieve the same region as in the one-sided helper case. The converse proof also follows the one-sided helper case.

6.3 Secure Triangular/Cascade Source Coding With a Public Helper

In this section, we consider new settings where the data transmission involves an intermediate node, termed as *helper*. We assume that the communication through the helper is not secure, i.e., the helper itself is a public terminal to which we do not want to reveal too much information about the source sequence. We characterize the tradeoff between rate, distortion, and information leakage rate in the form of the rate-distortion-leakage region for different settings of secure triangular source coding with a public helper (settings (A)-(D)) described in Section 6.1.1, see also Fig. 6.2). In our considered settings, the strategy at the helper depends heavily on the side information available at the helper and the decoder. For example, if the side

information at the helper is "degraded" with respect to that at the decoder, then the simple forwarding scheme is optimal. On the other hand, if the side information at the decoder is "degraded," it is optimal to perform decoding and re-encoding at the helper. Since we can see the cascade settings as special cases of the triangular settings when removing the private link, i.e., setting W_3 to be constant, we only present the results and proofs for the triangular settings.

6.3.1 Problem Formulation

Let us consider the setting in Fig. 6.2. Source and side information sequences (X^n, Y_1^n, Y_2^n, Z^n) are assumed to be i.i.d. according to $P_{X,Y_1,Y_2,Z}$. Given the sequences (X^n, Y_1^n), an encoder generates a description $W_1 \in \mathcal{W}_1^{(n)}$ and sends it to the helper over a noise-free, rate-limited link. The encoder also generates a description $W_3 \in \mathcal{W}_3^{(n)}$ based on (X^n, Y_1^n) and sends it to the decoder over another noise-free, rate-limited link. Based upon the description W_1 and the side information Z^n, the helper generates a new description $W_2 \in \mathcal{W}_2^{(n)}$ and sends it to the decoder. Given W_2, W_3, and its own side information Y_2^n, the decoder reconstructs the source sequence as $\hat{X}^n$.

Definition 6.3. A $(|\mathcal{W}_1^{(n)}|, |\mathcal{W}_2^{(n)}|, |\mathcal{W}_3^{(n)}|, n)$-code for secure triangular source coding with a public helper consists of

- a stochastic encoder $F_1^{(n)}$ which takes (X^n, Y_1^n) as input and generates $W_1 \in \mathcal{W}_1^{(n)}$ according to a conditional PMF $p(w_1|x^n, y_1^n)$,
- a stochastic helper $F_2^{(n)}$ which takes (W_1, Z_1^n) as input and generates $W_2 \in \mathcal{W}_2^{(n)}$ according to $p(w_2|w_1, z^n)$,
- a stochastic encoder $F_3^{(n)}$ which takes (X^n, Y_1^n) as input and generates $W_3 \in \mathcal{W}_3^{(n)}$ according to $p(w_3|x^n, y_1^n)$, and
- a decoder $g^{(n)} : \mathcal{W}_2^{(n)} \times \mathcal{W}_3^{(n)} \times \mathcal{Y}_2^{(n)} \to \hat{\mathcal{X}}^n$,

where $\mathcal{W}_1^{(n)}, \mathcal{W}_2^{(n)}$, and $\mathcal{W}_3^{(n)}$ are finite sets.

The information leakage at the helper who has access to W_1 and Z^n is measured by $\frac{1}{n} I(X^n; W_1, Z^n)$.

Definition 6.4. A rate-distortion-leakage tuple $(R_1, R_2, R_3, D, \triangle) \in \mathbb{R}_+^5$ is said to be *achievable* if for any $\delta > 0$ and all sufficiently large n there exists a $(|\mathcal{W}_1^{(n)}|, |\mathcal{W}_2^{(n)}|, |\mathcal{W}_3^{(n)}|, n)$ code such that

$$\frac{1}{n} \log \left|\mathcal{W}_i^{(n)}\right| \leq R_i + \delta, \ i = 1, 2, 3,$$

$$E[(X^n, g^{(n)}(W_2, W_3, Y_2^n))] \leq D + \delta,$$

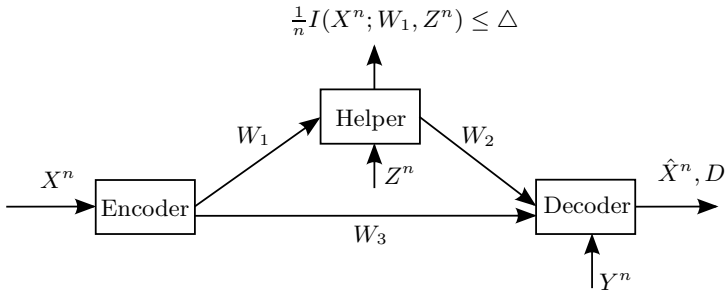

Figure 6.5: Secure triangular source coding with a public helper, setting (A).

$$\text{and} \quad \frac{1}{n}I(X^n;W_1,Z^n)\leq\triangle+\delta.$$

The *rate-distortion-leakage* region is defined as the set of all achievable tuples.

6.3.2 Triangular Setting (A)

Setting (A) assumes that the side information Y^n at a decoder is stronger than Z^n at a helper in the sense that $X-Y-Z$ forms a Markov chain. We characterize the rate-distortion-leakage region of the triangular setting (A) (with the Markov chain assumption $X-Y-Z$) under logarithmic loss distortion measure, and for the Gaussian setting under quadratic distortion.

Logarithmic Loss Distortion

Theorem 6.3.1 (Triangular (A), logarithmic loss)**.** *The rate-distortion-leakage region $\mathcal{R}_{tri(A),X\text{-}Y\text{-}Z,log\text{-}loss}$ under logarithmic loss distortion and $X-Y-Z$ assumption is the set of all tuples $(R_1,R_2,R_3,D,\triangle)\in\mathbb{R}_+^5$ that satisfy*

$$R_1 \geq [H(X|Y)-D-R_3]^+, \tag{6.8a}$$
$$R_2 \geq [H(X|Y)-D-R_3]^+, \tag{6.8b}$$
$$\triangle \geq I(X;Z)+[H(X|Y)-D-R_3]^+. \tag{6.8c}$$

Proof of Theorem 6.3.1:
Sketch of Achievability: The Wyner-Ziv coding at rate of $I(X;U|Y)+2\delta_\epsilon = H(X|Y)-D+2\delta_\epsilon$ is performed to satisfy the distortion constraint, where the equality is due to the choice of U and the property of logarithmic loss distortion. If $H(X|Y)-D>R_3$, we perform rate-splitting on the Wyner-Ziv index. That is, we split the index into two parts, namely $w_1\in[1:2^{n(H(X|Y)-D-R_3+\delta_\epsilon)}]$, and $w_3\in[1:2^{n(R_3+\delta_\epsilon)}]$. The indices w_1 and w_3 are sent over the cascade link and the private (triangular) link, respectively. Then the helper forwards the index w_1 to the decoder. It can be seen that the rate and distortion constraints are satisfied. As for the analysis of

leakage rate, we consider the following bound on the normalized mutual information averaged over all codebooks $\mathcal{C}_n$,

$$\begin{aligned} & I(X^n; W_1, Z^n|\mathcal{C}_n) \\ & = I(X^n; Z^n|\mathcal{C}_n) + I(X^n; W_1|Z^n, \mathcal{C}_n) \\ & \leq I(X^n; Z^n|\mathcal{C}_n) + H(W_1|Z^n, \mathcal{C}_n) \\ & \overset{(a)}{\leq} n[I(X;Z) + H(X|Y) - D - R_3 + \delta_\epsilon], \end{aligned}$$

where (a) follows from the facts that (X^n, Z^n) are i.i.d. and independent of the codebook, and from the codebook generation that $W_1 \in [1 : 2^{n(H(X|Y)-D-R_3+\delta_\epsilon)}]$.

On the other hand, if $H(X|Y) - D < R_3$, we send the Wyner-Ziv index over the private link, and send nothing over the cascade links, i.e., $R_1 \geq 0, R_2 \geq 0$ are achievable. The corresponding leakage rate is $\frac{1}{n} I(X^n; W_1, Z^n|\mathcal{C}_n) = \frac{1}{n} I(X^n; Z^n) = I(X;Z)$. The converse proof is given in Appendix 6.C. □

Remark 6.4. Since we assume that $X - Y - Z$ forms a Markov chain, it is optimal to perform the Wyner-Ziv coding with Y^n as side information at the receiver, and ignore the side information Z^n by simply forwarding the index received at the helper. This results in the same rate constraints on R_1 and R_2. Moreover, with this forwarding scheme at hand, rate-splitting of the index over the cascade and private links turns out to be optimal. Terms on the right hand side of the leakage rate constraint are simply the leakage rate due to the correlated side information Z^n, and the index received at the helper.

Gaussian Source Under Quadratic Distortion With $X - Y - Z$

Let the sequences (X^n, Y^n, Z^n) be i.i.d. according to $P_{X,Y,Z}$. We assume that X has a Gaussian distribution with zero mean and variance σ_X^2, i.e., $X \sim \mathcal{N}(0, \sigma_X^2)$. Let $Y = X + N_1, N_1 \sim \mathcal{N}(0, \sigma_{N_1}^2)$ independent of X, and $Z = Y + N_2, N_2 \sim \mathcal{N}(0, \sigma_{N_2}^2)$ independent of (X, Y, N_1), where $\sigma_X^2, \sigma_{N_1}^2, \sigma_{N_2}^2 > 0$. This satisfies the Markov assumption $X - Y - Z$.

Theorem 6.3.2 (Triangular (A), Gaussian)**.** *The rate-distortion-leakage region for a Gaussian source with quadratic distortion under the Markov assumption $X - Y - Z$, $\mathcal{R}_{tri(A),X\text{-}Y\text{-}Z,Gaussian}$, is the set of all tuples $(R_1, R_2, R_3, D, \triangle) \in \mathbb{R}_+^5$ that satisfy*

$$R_1 \geq [\frac{1}{2} \log \big(\sigma^2/D\big) - R_3]^+, \tag{6.9a}$$

$$R_2 \geq [\frac{1}{2} \log \big(\sigma^2/D\big) - R_3]^+, \tag{6.9b}$$

$$\triangle \geq \frac{1}{2} \log \big(1 + \frac{\sigma_X^2}{\sigma_{N_1}^2 + \sigma_{N_2}^2}\big) + [\frac{1}{2} \log \big(\sigma^2/D\big) - R_3]^+, \tag{6.9c}$$

where $\sigma^2 = \frac{\sigma_X^2 \sigma_{N_1}^2}{\sigma_X^2 + \sigma_{N_1}^2}$.

Proof. The proof is given in Appendix 6.D. □

Based on the triangular setting in Fig. 6.5, one might consider a related scenario where the encoder can only "broadcast" (BC) the same source description to the helper and the decoder over the rate-limited digital links, in the sense of Fig. 6.3, i.e., $W_3 = W_1$. Based on the source description and some side information, the helper generates a new description and sends it to the decoder. The rest of the problem formulation of this triangular setting is similar to that of Fig. 6.5. We characterize the rate-distortion-leakage region under the logarithmic loss distortion and the Markov assumption $X - Y - Z$.

Theorem 6.3.3 (Triangular (A), logarithmic loss, BC)**.** *The rate-distortion-leakage region $\mathcal{R}_{tri(A),X\text{-}Y\text{-}Z,log\text{-}loss,BC}$ under logarithmic loss distortion is the set of all tuples $(R_1, R_2, D, \triangle) \in \mathbb{R}_+^4$ that satisfy*

$$R_1 \geq [H(X|Y) - D]^+, \tag{6.10a}$$

$$R_2 \geq 0, \tag{6.10b}$$

$$\triangle \geq I(X;Z) + [H(X|Y) - D]^+. \tag{6.10c}$$

Proof. If $H(X|Y) - D > 0$, the achievability proof follows the Wyner-Ziv coding for the encoder/decoder pair at rate above $I(X;U|Y) = H(X|Y) - D$. Since the index W_1 is also available at the decoder, the helper does not need to send anything to the decoder (due to the data processing inequality (Lemma 2.3)). The leakage proof follows similarly as the proof of Theorem 6.3.1. On the other hand, if $H(X|Y) - D < 0$, the encoder does not need to send anything, i.e., $R_1 \geq 0, R_2 \geq 0$ are achievable. The corresponding leakage rate is $\frac{1}{n} I(X^n; W_1, Z^n|\mathcal{C}_n) = \frac{1}{n} I(X^n; Z^n) = I(X;Z)$. Converse proofs for R_1 and $\triangle$ constraints follow similarly as the proof of Theorem 6.3.1, while $R_2 \geq 0$ is trivial. □

Remark 6.5. In this case, the helper does not help to provide any additional information to the decoder due to the Markov relation $X^n - Y^n - Z^n$. That is, given (W_1, Y^n), the decoder already has all information about the source available. Note also that this setting is similar to the source coding setting considered in [VP13] with the cooperation link from the helper to the decoder. However, the cooperation link does not provide any extra information to the decoder.

6.3.3 Triangular Setting (B)

Setting (B) assumes that the common side information Y^n is available at both encoder and decoder. This allows the encoder and decoder a possibility to generate a secret key for protecting the source description sent through the public helper. We characterize the rate-distortion-leakage region of the triangular setting (B) (with the Markov chain assumption $X - Y - Z$) under logarithmic loss distortion measure, and for the Gaussian setting under quadratic distortion.

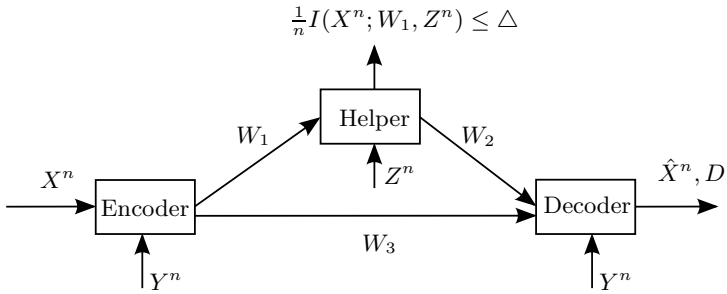

Figure 6.6: Secure triangular source coding with a public helper, setting (B).

Logarithmic Loss Distortion

Theorem 6.3.4 (Triangular (B), logarithmic loss)**.** *The rate-distortion-leakage region $\mathcal{R}_{tri(B),X\text{-}Y\text{-}Z,log\text{-}loss}$ under logarithmic loss distortion and $X-Y-Z$ is the set of all tuples $(R_1,R_2,R_3,D,\triangle)\in\mathbb{R}_+^5$ that satisfy*

$$R_1 \geq [H(X|Y)-D-R_3]^+, \tag{6.11a}$$

$$R_2 \geq [H(X|Y)-D-R_3]^+, \tag{6.11b}$$

$$\triangle \geq I(X;Z)+[H(X|Y)-D-R_3-H(Y|X,Z)]^+. \tag{6.11c}$$

Remark 6.6. We first note that the availability of side information Y^n at the encoder does not improve the rate-distortion tradeoff under a logarithmic loss distortion, with respect to the Wyner-Ziv setting [WZ76] (like in the Gaussian case [Wyn78]). Similarly as in Section 5.4, the common side information at the encoder helps to reduce the leakage rate at the helper by allowing the encoder and the decoder to generate a secret key. We can see this from the leakage constraint (6.11c) above where the leakage rate consists of contributions from the eavesdropper's side information $I(X;Z)$ and from the source description which is partially protected by the secret key of rate $\min\{H(Y|X,Z),H(X|Y)-D-R_3\}$ (cf. (6.8c) in Theorem 6.3.1 where there is no leakage reduction from the secret key). This role of side information at the encoder and the decoder in another secure source coding setting is also studied in [CK13b].

Proof of Theorem 6.3.4:
Sketch of Achievability: The proof follows similarly as in previous triangular case with the additional steps of secret key generation using y^n. That is, the Wyner-Ziv coding at rate $I(X;U|Y)+2\delta_\epsilon = H(X|Y)-D+2\delta_\epsilon$ is performed to satisfy the distortion constraint. Then we perform rate-splitting on the Wyner-Ziv index by splitting it into two parts, namely $w_1\in[1:2^{n(H(X|Y)-D-R_3+\delta_\epsilon)}]$, and $w_3\in[1:2^{n(R_3+\delta_\epsilon)}]$. Next we distinguish between two cases where we further split the index w_1 and where the key rate is sufficient for scrambling the whole index w_1.

If $H(X|Y) - D - R_3 > H(Y|X,Z)$, we further split index w_1 into indices $w_{11} \in [1 : 2^{n(H(X|Y)-D-R_3-H(Y|X,Z)+\delta_\epsilon)}]$ and $w_{12} \in [1 : 2^{nH(Y|X,Z)}]$. Then the secret key k is generated by randomly and independently partitioning sequences in $\mathcal{Y}^n$ into $2^{nH(Y|X,Z)}$ bins and choosing k as the corresponding bin index of the given y^n. The encoder sends w_{11} and $w_{12} \oplus k$ over the cascade link, and w_3 over the private link, where $w_{12} \oplus k$ denotes the modulo operation, $(w_{12}+k) \mathrm{mod} 2^{nH(Y|X,Z)}$[1]. The helper forwards the index w_{11} and $w_{12} \oplus k$ to the decoder. The decoder can recover w_{12} from its key generated by y^n. We can show that the tuples satisfying (6.11) where $[a]^+ = a$ in (6.11c) are achievable.

If $H(X|Y) - D - R_3 < H(Y|X,Z)$, the secret key is generated by randomly and independently partitioning sequences in $\mathcal{Y}^n$ into $2^{n(H(X|Y)-D-R_3+\delta_\epsilon)}$ bins and choosing the corresponding bin index of given y^n as a key. The encoder sends $w_1 \oplus k$ over the cascade link, and w_3 over the private link. The helper forwards $w_1 \oplus k$ to the decoder. We can show that the tuples satisfying (6.11) where $[a]^+ = 0$ in (6.11c) are achievable. For the detailed achievability proof and converse proof, please see Appendix 6.E. □

Gaussian Source Under Quadratic Distortion With $X - Y - Z$

Let the sequences (X^n, Y^n, Z^n) be i.i.d. according to $P_{X,Y,Z}$. We assume that X has a Gaussian distribution with zero mean and variance σ_X^2, i.e., $X \sim \mathcal{N}(0, \sigma_X^2)$. Let $Y = X + N_1, N_1 \sim \mathcal{N}(0, \sigma_{N_1}^2)$ independent of X, and $Z = Y + N_2, N_2 \sim \mathcal{N}(0, \sigma_{N_2}^2)$ independent of (X, Y, N_1), where $\sigma_X^2, \sigma_{N_1}^2, \sigma_{N_2}^2 > 0$. This satisfies the Markov assumption $X - Y - Z$.

Theorem 6.3.5 (Triangular (B), Gaussian)**.** *The rate-distortion-leakage region for a Gaussian source with quadratic distortion under the Markov assumption $X - Y - Z$, $\mathcal{R}_{tri(B),X\text{-}Y\text{-}Z,Gaussian}$, is the set of all tuples $(R_1, R_2, R_3, D, \triangle) \in \mathbb{R}_+^5$ that satisfy*

$$R_1 \geq [\frac{1}{2} \log \left(\sigma^2/D\right) - R_3]^+, \tag{6.12a}$$

$$R_2 \geq [\frac{1}{2} \log \left(\sigma^2/D\right) - R_3]^+, \tag{6.12b}$$

$$\triangle \geq \frac{1}{2} \log \left(1 + \frac{\sigma_X^2}{\sigma_{N_1}^2 + \sigma_{N_2}^2}\right). \tag{6.12c}$$

Proof. The availability of Gaussian side information Y^n at the encoder and decoder allows us to generate a *discrete* secret key at arbitrarily high rate. This implies that we can essentially protect the whole source description sent over the rate limited link to the helper, and the only leakage to the eavesdropper is due to the eavesdropper's correlated side information Z^n. The detailed proof is given in Appendix 6.F. □

[1]Here, we have w_{12} and $k \in [1 : 2^{nH(Y|X,Z)}]$. Thus, in the modulo operation, 0 is mapped to $2^{nH(Y|X,Z)}$.

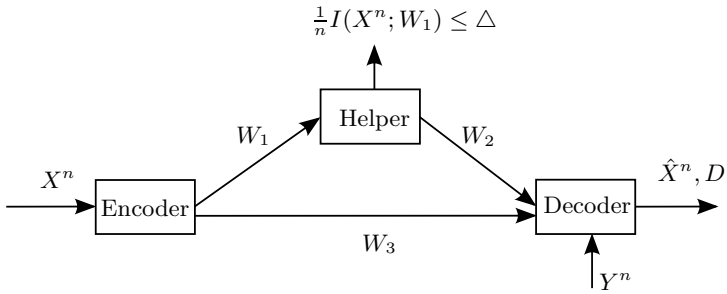

Figure 6.7: Secure triangular source coding with a public helper, setting (C).

Again, based on the triangular setting in Fig. 6.6, one can consider the case where the encoder can only "broadcast" the same source description to the helper and the decoder over the rate-limited digital link, in the sense of Fig. 6.3. We characterize the rate-distortion-leakage region under the logarithmic loss distortion. Similarly to the result in Theorem 6.3.3, the helper is not helpful in terms of helping the transmission due to the Markov assumption $X - Y - Z$. In other words, W_2 does not provide any extra information to the decoder.

Theorem 6.3.6 (Triangular (B), logarithmic loss, BC)**.** *The rate-distortion-leakage region $\mathcal{R}_{tri(B),X\text{-}Y\text{-}Z,log\text{-}loss,BC}$ under logarithmic loss distortion is the set of all tuples $(R_1, R_2, D, \triangle) \in \mathbb{R}_+^4$ that satisfy*

$$R_1 \geq [H(X|Y) - D]^+, \tag{6.13a}$$

$$R_2 \geq 0, \tag{6.13b}$$

$$\triangle \geq I(X;Z) + [H(X|Y) - D - H(Y|X,Z)]^+. \tag{6.13c}$$

Proof. Since we have the Markov assumption $X - Y - Z$ and the index W_1 is also available at the decoder, the helper does not need to send anything to the decoder. Hence, the problem turns into a standard secure source coding with side information at both encoder and decoder. The proof follows similarly as that of Theorem 6.3.4. □

6.3.4 Triangular Setting (C)

Setting (C) assumes that the helper has no side information. We characterize the rate-distortion-leakage region for the triangular setting (C) under general distortion.

Theorem 6.3.7 (Triangular (C))**.** *The rate-distortion-leakage region for the triangular setting (C), $\mathcal{R}_{tri(C)}$ is the set of all tuples $(R_1, R_2, R_3, D, \triangle) \in \mathbb{R}_+^5$ that*

satisfy

$$R_1 \geq [I(X;U|Y) - R_3]^+, \tag{6.14a}$$
$$R_2 \geq [I(X;U|Y) - R_3]^+, \tag{6.14b}$$
$$D \geq E[d(X, \tilde{g}(U,Y))], \tag{6.14c}$$
$$\triangle \geq [I(X;U|Y) - R_3]^+, \tag{6.14d}$$

for some joint distributions of the form $P_{X,Y}(x,y)P_{U|X}(u|x)$ *with* $|\mathcal{U}| \leq |\mathcal{X}| + 1$, *and a function* $\tilde{g} : \mathcal{U} \times \mathcal{Y} \to \hat{\mathcal{X}}$.

Remark 6.7. Since there is no side information at the helper, it is obvious that the optimal scheme at the helper is to simply forward the source description, i.e., setting $W_2 = W_1$. In this case, unlike setting (A) in Fig. 6.5, we are able to solve the problem under a general distortion measure since the problem essentially reduces to the Wyner-Ziv problem with an additional leakage rate constraint.

Proof of Theorem 6.3.7:
Sketch of Achievability: The proof is similar to that of triangular setting (A) where we use rate splitting. The Wyner-Ziv coding at rate of $I(X;U|Y) + 2\delta_\epsilon$ is performed to satisfy the distortion constraint. Then we perform rate-splitting on the Wyner-Ziv index. That is, we split the index into two parts, namely $w_1 \in [1, 2^{n(I(X;U|Y)-R_3+\delta_\epsilon)}]$, and $w_3 \in [1, 2^{n(R_3+\delta_\epsilon)}]$. The indices w_1 and w_3 are sent over the cascade link and the private (triangular) link, respectively. The helper forwards the index w_1 to the decoder. The analysis of distortion follows from the analysis for the Wyner-Ziv setting in [EK11, Chapter 11]. As for the analysis of leakage rate, we consider the following bound on the normalized mutual information averaged over all codebooks,

$$I(X^n; W_1|\mathcal{C}_n) \leq H(W_1|\mathcal{C}_n) \overset{(a)}{\leq} n[I(X;U|Y) - R_3 + \delta_\epsilon]$$

where (a) follows from the codebook generation that $W_1 \in [1 : 2^{n(I(X;U|Y)-R_3+\delta_\epsilon)}]$.

The converse proof follows similarly as in the triangular setting (A) and is given in Appendix 6.G. □

As before, based on the triangular setting in Fig. 6.7, one can consider a related scenario where the encoder broadcasts the same source description to the helper and the decoder, in the sense of Fig. 6.3. We characterize the rate-distortion-leakage region for a general distortion. Similarly to the results in Theorems 6.3.3 and 6.3.6, the helper is not helpful in terms of providing additional information to the decoder.

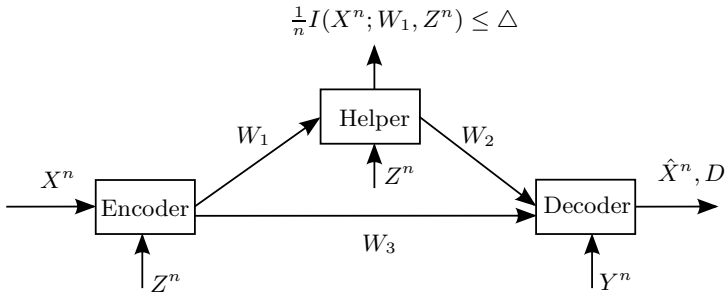

Figure 6.8: Secure triangular source coding with a public helper with $X - Z - Y$, setting (D).

Theorem 6.3.8 (Triangular (C), BC)**.** *The rate-distortion-leakage region $\mathcal{R}_{tri(C),BC}$ is the set of all tuples $(R_1, R_2, D, \triangle) \in \mathbb{R}_+^4$ that satisfy*

$$R_1 \geq I(X;U|Y), \tag{6.15a}$$
$$R_2 \geq 0, \tag{6.15b}$$
$$D \geq E[d(X, \tilde{g}(U,Y))], \tag{6.15c}$$
$$\triangle \geq I(X;U|Y), \tag{6.15d}$$

for some joint distributions of the form $P_{X,Y}(x,y)P_{U|X}(u|x)$ with $|\mathcal{U}| \leq |\mathcal{X}| + 1$, and a function $\tilde{g} : \mathcal{U} \times \mathcal{Y} \to \hat{\mathcal{X}}$.

Proof. Since the index W_1 is also available at the decoder, the helper does not need to send anything to the decoder. The proof essentially follows the Wyner-Ziv coding proof [EK11, Chapter 11] and the proof of information leakage constraint follows similarly as in Theorem 6.3.7. □

6.3.5 Triangular Setting (D)

In setting (D), we consider the case where side information Z^n at the helper is also available to the encoder, as depicted in Fig. 6.8, under the Markov assumption $X - Z - Y$. This setting is "dual" to the setting (B) in the sense that we switch the order of side information degradedness and the availability of helper's side information or decoder's side information at the encoder. We characterize the rate-distortion-leakage region for triangular setting (D) under general distortion.

Theorem 6.3.9 (Triangular (D))**.** *The rate-distortion-leakage region $\mathcal{R}_{tri(D),X\text{-}Z\text{-}Y}$ is the set of all tuples $(R_1, R_2, R_3, D, \triangle) \in \mathbb{R}_+^5$ that satisfy*

$$R_1 \geq I(X;U|Z), \tag{6.16a}$$
$$R_2 \geq I(X,Z;U|Y), \tag{6.16b}$$

$$R_3 \geq I(X, Z; V|U, Y), \tag{6.16c}$$
$$D \geq E[d(X, \tilde{g}(U, V, Y))], \tag{6.16d}$$
$$\triangle \geq I(X; U, Z), \tag{6.16e}$$

for some joint distributions of the form

$$P_{X,Z}(x, z)P_{Y|Z}(y|z)P_{U|X,Z}(u|x, z)P_{V|X,Z,U}(v|x, z, u)$$

with $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{Z}| + 3$ *and* $|\mathcal{V}| \leq (|\mathcal{X}||\mathcal{Z}| + 3)(|\mathcal{X}||\mathcal{Z}| + 1)$*, and a function* $\tilde{g} : \mathcal{U} \times \mathcal{V} \times \mathcal{Y} \to \hat{\mathcal{X}}$*.*

Remark 6.8. Since we assume a new order of side information degradedness $X - Z - Y$, it is optimal for the helper to perform decoding and re-encoding. In other words, the side information Z^n at the helper is useful in providing extra information to the decoder. We note that if the leakage constraint at the helper is replaced by the decoding constraint under some distortion, the problem turns into the original triangular/cascade source coding problem in [CPW12].

Proof of Theorem 6.3.9:
Sketch of Achievability: The achievable scheme follows the *decode and re-bin* scheme of [CPW12]. That is, for fixed $P_{U|X,Z}, P_{V|X,Z,U}$, and $\tilde{g} : \mathcal{U} \times \mathcal{V} \times \mathcal{Y} \to \hat{\mathcal{X}}$, randomly and independently generate $2^{n(I(X,Z;U)+\delta_\epsilon)}$ sequences $u^n(\tilde{w}_1) \sim \prod_{i=1}^n P_U(u_i(\tilde{w}_1))$, $\tilde{w}_1 \in [1 : 2^{n(I(X,Z;U)+\delta_\epsilon)}]$. Then distribute them uniformly into $2^{n(I(X;U|Z)+2\delta_\epsilon)}$ bins $b_{u_1}(w_1)$, $w_1 \in [1 : 2^{n(I(X;U|Z)+2\delta_\epsilon)}]$. In addition, we distribute them also uniformly into another $2^{n(I(X,Z;U|Y)+2\delta_\epsilon)}$ bins $b_{u_2}(w_2)$, $w_2 \in [1 : 2^{n(I(X,Z;U|Y)+2\delta_\epsilon)}]$. Also, for each $\tilde{w}_1$, randomly and conditionally independently generate $2^{n(I(X,Z;V|U)+\delta_\epsilon)}$ sequences $v^n(\tilde{w}_1, \tilde{w}_3) \sim \prod_{i=1}^n P_{V|U}(\cdot|u_i(\tilde{w}_1))$, and distribute them uniformly into $2^{n(I(X,Z;V|U,Y)+2\delta_\epsilon)}$ bins $b_v(w_3)$, $w_3 \in [1 : 2^{n(I(X,Z;V|U,Y)+2\delta_\epsilon)}]$. For encoding, the encoder looks for a sequence u^n that is jointly typical with (x^n, z^n). If there is more than one such sequence, it selects one of them uniformly at random. If there is no such u^n, it selects one out of $2^{n(I(X,Z;U)+\delta_\epsilon)}$ uniformly at random. With high probability, there exists such u^n since there are $2^{n(I(X,Z;U)+\delta_\epsilon)}$ codewords generated. Then it transmits the corresponding bin index w_1 to the helper. Also, the encoder looks for v^n that is jointly typical with (x^n, z^n, u^n). If there is more than one, it selects one of them uniformly at random. If there is no such v^n, it selects one out of $2^{n(I(X,Z;V|U)+\delta_\epsilon)}$ uniformly at random. With high probability, there exists such v^n since there are $2^{n(I(X,Z;V|U)+\delta_\epsilon)}$ codewords generated. Then it transmits the corresponding bin index w_3 to the decoder over the private link. Upon receiving the bin index w_1, the helper node looks for the unique u^n such that it is jointly typical with the side information z^n. With high probability, it will find the unique and correct one since there are $2^{n(I(U;Z)-\delta_\epsilon)}$ codewords in each bin $b_{u_1}(w_1)$. After that the helper looks for the corresponding bin $b_{u_2}(w_2)$ such that the decoded $u^n \in b_{u_2}(w_2)$, and transmit the bin index w_2 to the decoder. The decoder, with high probability, will successively find the unique and correct u^n and v^n that are jointly typical with y^n since there are $2^{n(I(U;Y)-\delta_\epsilon)}$ codewords in each bin $b_{u_2}(w_2)$,

and there are $2^{n(I(V;Y|U)-\delta_\epsilon)}$ codewords in each bin $b_v(w_3)$. Then $\hat{x}^n$ is put out as a source reconstruction, where $\hat{x}_i = \tilde{g}(u_i, v_i, y_i)$. Since (x^n, u^n, v^n, y^n) are jointly typical, we can show that $D \geq E[d(X, \tilde{g}(U, V, Y))]$ is achievable.

As for the analysis of leakage rate, we consider the following bound on the normalized mutual information averaged over all codebooks,

$$\begin{aligned}
&I(X^n; W_1, Z^n|\mathcal{C}_n) \\
&= I(X^n; Z^n|\mathcal{C}_n) + I(X^n; W_1|Z^n, \mathcal{C}_n) \\
&\leq I(X^n; Z^n|\mathcal{C}_n) + H(W_1|Z^n, \mathcal{C}_n) \\
&\overset{(a)}{\leq} n[I(X; Z) + I(X; U|Z) + \delta_\epsilon] \\
&= n[I(X; U, Z) + \delta_\epsilon],
\end{aligned}$$

where (a) follows from the fact that (X^n, Z^n) are i.i.d. and independent of the codebook, and from the codebook generation that we have $W_1 \in [1 : 2^{n(I(X;U|Z)+\delta_\epsilon)}]$. The converse proof is given in Appendix 6.H. □

We also consider the scenario where the encoder broadcasts the source description to the helper and the decoder, in the sense of Fig. 6.3. We characterize the rate-distortion-leakage region for a general distortion. In this case, unlike the previous three cases under settings (A)-(C), the helper is useful in terms of supporting the transmission since its side information Z^n is "stronger" than Y^n at the decoder due to the assumption that $X - Z - Y$ forms a Markov chain.

Theorem 6.3.10 (Triangular (D), BC)**.** *The rate-distortion-leakage region for triangular setting (D) where the encoder broadcasts the description, $\mathcal{R}_{tri(D),X\text{-}Z\text{-}Y,BC}$ is the set of all tuples $(R_1, R_2, D, \triangle) \in \mathbb{R}_+^4$ that satisfy*

$$R_1 \geq I(X; U|Z), \tag{6.17a}$$
$$R_1 + R_2 \geq I(X, Z; U|Y), \tag{6.17b}$$
$$D \geq E[d(X, \tilde{g}(U, Y))], \tag{6.17c}$$
$$\triangle \geq I(X; U, Z), \tag{6.17d}$$

for some joint distributions of the form $P_{X,Z}(x, z)P_{Y|Z}(y|z)P_{U|X,Z}(u|x, z)$ with $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{Z}| + 2$, and a function $\tilde{g} : \mathcal{U} \times \mathcal{Y} \to \hat{\mathcal{X}}$.

Proof. Since both indices W_1 and W_2 are available at the decoder, we can perform joint coding which leads to a sum-rate constraint $R_1 + R_2$ instead of the individual rate R_2. For achievability, the encoder sends the partial bin index of the selected codeword at rate R_1 and then the helper performs *decode and re-bin* and sends the rest of the bin index at rate R_2 to the decoder. The detailed proof is given in Appendix 6.I. □

Remark 6.9. Here we discuss the optimal operation at the helper in all considered settings. Since the private link in the triangular setting can only provide additional

information subject to its rate constraint, the processing ambiguity lies only in the cascade transmission, i.e., what is the best relaying strategy at the helper? For ease of discussion, we will for now neglect the private link, and argue that when the side information at the helper is degraded with respect to that at the decoder, forwarding scheme at the helper is optimal; otherwise, it is optimal to employ a decode-and-reencode type scheme.

- Let us consider the setting (A) in which we assume that $X - Y - Z$ forms a Markov chain (the discussion for settings (B) and (C) follows similarly). On the cascade link, in order to attain low distortion at the decoder, we wish to compress the source so that the decoder, upon receiving (W_2, Y^n), can extract as much information about the source as possible, i.e., maximizing $I(X^n; W_2, Y^n)$. The joint PMF of this setting after summing out the reconstruction sequences is given by $P_{X^n,Y^n} P_{Z^n|Y^n} P_{W_1|X^n} P_{W_2|W_1,Z^n}$. Data processing inequality implies that $I(X^n; W_2, Y^n) \leq I(X^n; W_1, Y^n)$. This suggests that the forwarding scheme at the helper (setting $W_2 = W_1$) is a good strategy for this setting, and it is in fact optimal in this case.

- On the other case (setting (D)) where we assume that $X - Z - Y$ forms a Markov chain, the joint PMF after summing out the reconstruction sequences is given by $P_{X^n,Z^n} P_{Y^n|Z^n} P_{W_1|X^n,Z^n} P_{W_2|W_1,Z^n}$. To see if the forwarding scheme is still optimal, we consider the following inequality (derived from the joint PMF using the data processing inequality), $I(X^n; W_2, Y^n) \leq I(X^n; W_1, Z^n)$. The inequality suggests that, based on information available, the helper can extract more information about X^n than the decoder does, regardless of what the helper scheme is. Since W_2 is generated based on (W_1, Z^n), it is reasonable that the helper takes into account the knowledge about Z^n in relaying the information, rather than just forwarding W_1. It turns out that the decode-and-reencode type scheme is optimal in this case.

6.4 Conclusion

We study secure source coding problems with a public helper that supports the transmission while there is a risk for information leakage. Two classes of problems are considered, namely secure source coding with a helper where the helper link is eavesdropped, and secure triangular source coding with a public helper who is friendly but curious. We are interested in how the helper can facilitate transmission in these unsecured scenarios. We characterize the rate-distortion-leakage regions for different cases. In the first class of the problems, we present the rate-distortion-leakage regions for one-sided and two-sided helper cases under some specific distortion measure, and show that a standard coding scheme is optimal. We found that, for the logarithmic loss distortion case and the case of a Gaussian source with quadratic distortion under a Markov relation, the region is the same for both the one-sided and two-sided settings. This observation provides evidence that the

availability of (coded) side information at the encoder does not improve the rate-distortion-leakage tradeoff. Furthermore, in triangular settings, we solve several special cases and observe that the optimal operation at the helper in our coding scheme depends heavily on the order of side information degradedness, i.e., when $X - Y - Z$ forms a Markov chain, the forwarding scheme is optimal, and when $X - Z - Y$ forms a Markov chain, the *decode and re-bin* scheme is optimal.

Appendices for Chapter 6

6.A Proof of Converse for One-sided Helper

Proof of Converse: For any achievable tuple $(R_1, R_2, \triangle, D)$, by standard properties of the entropy function, it follows that

$$\begin{aligned} n(R_1+\delta_n) &\geq \log|\mathcal{W}_1^{(n)}| \\ &\geq H(W_1) \geq H(W_1|W_2) \\ &= H(X^n, W_1|W_2) - H(X^n|W_1, W_2) \\ &\overset{(a)}{\geq} H(X^n, W_1|W_2) - nD \\ &\geq \sum_{i=1}^{n} H(X_i|W_2, X^{i-1}) - nD \\ &\overset{(b)}{=} \sum_{i=1}^{n} H(X_i|U_i) - nD, \end{aligned}$$

where (a) follows from the fact that under logarithmic loss distortion we have that $D \geq E[d(X^n, g^{(n)}(W_1, W_2))] \geq \frac{1}{n}H(X^n|W_1, W_2)$ (Lemma 5.2) and (b) follows by defining $U_i \triangleq (W_2, X^{i-1})$.

Next,

$$\begin{aligned} n(R_2+\delta_n) &\geq H(W_2) \\ &\geq I(W_2; X^n, Y^n) \\ &= \sum_{i=1}^{n} H(X_i, Y_i) - H(X_i, Y_i|W_2, X^{i-1}, Y^{i-1}) \\ &\geq \sum_{i=1}^{n} H(X_i, Y_i) - H(X_i, Y_i|U_i) \\ &\geq \sum_{i=1}^{n} I(Y_i; U_i). \end{aligned}$$

Lastly, the leakage rate

$$\begin{aligned} n(\triangle+\delta_n) &\geq I(X^n; W_2, Z^n) \\ &= \sum_{i=1}^{n} H(X_i) - H(X_i|W_2, X^{i-1}, Z^n) \\ &\geq \sum_{i=1}^{n} H(X_i) - H(X_i|U_i, Z_i). \end{aligned}$$

We proceed by using the standard time-sharing argument. Let Q be a random variable uniformly distributed over the set $\{1, 2, \ldots, n\}$ and independent of X_i, Y_i, Z_i, $1 \leq i \leq n$. We consider the joint distribution of new random variables (X, Y, Z, U), where $X \triangleq X_Q, Y \triangleq Y_Q, Z \triangleq Z_Q$, and $U \triangleq (Q, U_Q)$. Note that we have $P_{X,Y,Z} = P_{X_Q,Y_Q,Z_Q}$ and $U - Y - (X, Z)$ forms a Markov chain due to the i.i.d. property of the source and side information sequences.

By introducing Q in above expressions, it is straightforward to show that rate and leakage rate constraints above can be bounded further by

$$\begin{aligned} R_1 + \delta_n &\geq H(X|U) - D \\ R_2 + \delta_n &\geq I(Y;U) \\ \triangle + \delta_n &\geq I(X;U,Z), \end{aligned}$$

for some $P_{X,Y,Z}P_{U|Y}$. The proof is concluded by letting $n \to \infty$.

For the bound on the cardinality of the set $\mathcal{U}$, it can be shown by using the support lemma [CK11, Lemma 15.4] that it suffices that $\mathcal{U}$ should have $|\mathcal{Y}| - 1$ elements to preserve P_Y, plus three more for $H(Y|U), H(X|U)$, and $H(X|U,Z)$.

6.B Proof of Theorem 6.2.3

With the assumption that $Y \sim \mathcal{N}(0, \sigma_Y^2)$, $X = Y + N_1, N_1 \sim \mathcal{N}(0, \sigma_{N_1}^2)$ independent of Y, and $Z = X + N_2, N_2 \sim \mathcal{N}(0, \sigma_{N_2}^2)$ independent of X, Y, N_1, we will prove that the inner bound given in Theorem 6.2.1 is tight for this case.

Proof of Achievability: Let us choose $U = Y + Q, Q \sim \mathcal{N}(0, \frac{\alpha}{1-\alpha}\sigma_Y^2)$ independent of Y, and $V = X + P, P \sim \mathcal{N}(0, \sigma_P^2)$ independent of X, where $\alpha \in (0, 1)$ and $\sigma_P^2 = \frac{(\alpha\sigma_Y^2 + \sigma_{N_1}^2)D}{\alpha\sigma_Y^2 + \sigma_{N_1}^2 - D}$ for $D < \alpha\sigma_Y^2 + \sigma_{N_1}^2$, otherwise setting V constant. Also, choose $\tilde{g}(U, V)$ to be an MMSE estimate of X given U and V.

With these choices of U, V and $\tilde{g}(\cdot)$, it can be shown that

$$\begin{aligned} I(Y;U) &= h(U) - h(U|Y) \\ &= \frac{1}{2}\log(2\pi e(\sigma_Y^2 + \frac{\alpha}{1-\alpha}\sigma_Y^2)) - \frac{1}{2}\log(2\pi e(\frac{\alpha}{1-\alpha}\sigma_Y^2)) \\ &= \frac{1}{2}\log(1/\alpha), \end{aligned}$$

where $h(U)$ and $h(U|Y)$ are the differential entropy and conditional differential entropy, as defined in [CT06, Chapter 8], and

$$\begin{aligned} I(X;V|U) &= h(X|U) - h(X|U,V) \\ &= \frac{1}{2}\log(\frac{\mathrm{var}(X|U)}{\mathrm{var}(X|U,V)}) \\ &= \frac{1}{2}\log(\frac{\alpha\sigma_Y^2 + \sigma_{N_1}^2}{D}), \end{aligned}$$

for $D < \alpha\sigma_Y^2 + \sigma_{N_1}^2$, where $\mathrm{var}(X|U) = \alpha\sigma_Y^2 + \sigma_{N_1}^2$ and $\mathrm{var}(X|U,V) = \frac{\mathrm{var}(X|U)\sigma_P^2}{\mathrm{var}(X|U)+\sigma_P^2}$, and

$$\begin{aligned} I(X;U,Z) &= h(X) - h(X|U,Z) \\ &= \frac{1}{2}\log(\frac{\sigma_X^2}{\mathrm{var}(X|U,Z)}) \\ &= \frac{1}{2}\log\big(\frac{(\sigma_Y^2+\sigma_{N_1}^2)(\alpha\sigma_Y^2+\sigma_{N_1}^2+\sigma_{N_2}^2)}{\sigma_{N_2}^2(\alpha\sigma_Y^2+\sigma_{N_1}^2)}\big), \end{aligned}$$

where $\mathrm{var}(X|U,Z) = \frac{\mathrm{var}(X|U)\sigma_{N_2}^2}{\mathrm{var}(X|U)+\sigma_{N_2}^2}$, and lastly,

$$\begin{aligned} E[d(X,\tilde{g}(U,V))] &= E[(X-\tilde{g}(U,V))^2] \\ &= \mathrm{var}(X|U,V) \\ &= D. \end{aligned}$$

Proof of Converse: From the problem formulation, the joint PMF is given by

$$P_{X^n,Y^n}P_{Z^n|X^n}P_{W_1|X^n}P_{W_2|Y^n}1_{\{\hat{X}^n=g^{(n)}(W_1,W_2)\}}.$$

It follows that

$$\begin{aligned} n(R_2+\delta_n) &\geq H(W_2) \\ &= I(W_2;Y^n) \\ &= h(Y^n) - h(Y^n|W_2) \\ &= n/2\log(2\pi e\sigma_Y^2) - h(Y^n|W_2) \\ &\overset{(a)}{\geq} n/2\log(2\pi e\sigma_Y^2) - n/2\log(2^{\frac{2}{n}h(X^n|W_2)} - 2^{\frac{2}{n}h(N_1^n|W_2)}) \\ &= n/2\log(2\pi e\sigma_Y^2) - n/2\log(2^{\frac{2}{n}h(X^n|W_2)} - 2\pi e\sigma_{N_1}^2), \end{aligned}$$

where (a) follows from the conditional EPI [EK11, Chapter 2] and the fact that $X^n = Y^n + N_1^n$, Y^n conditionally independent of N_1^n given W_2.

Next, consider the Markov chain $W_2 - Y^n - X^n - Z^n$, we have that

$$n/2\log(2\pi e\sigma_X^2) = h(X^n) \geq h(X^n|W_2) \geq h(X^n|Y^n) = h(N_1^n) = n/2\log(2\pi e\sigma_{N_1}^2).$$

Then there must exists $\alpha \in [0,1]$ such that $h(X^n|W_2) = n/2\log(2\pi e(\alpha\sigma_X^2 + (1-\alpha)\sigma_{N_1}^2)) = n/2\log(2\pi e(\alpha\sigma_Y^2 + \sigma_{N_1}^2))$. Thus, we have

$$n(R_2+\delta_n) \geq n/2\log(2\pi e\sigma_Y^2) - n/2\log(2\pi e\alpha\sigma_Y^2) = n/2\log(1/\alpha).$$

Next,

$$\begin{aligned}
n(R_1+\delta_n) &\geq H(W_1)\\
&\geq I(W_1;X^n|W_2)\\
&= h(X^n|W_2)-h(X^n|W_1,W_2)\\
&\geq h(X^n|W_2)-\sum_{i=1}^{n}h(X_i|W_1,W_2)\\
&\geq h(X^n|W_2)-\sum_{i=1}^{n}\frac{1}{2}\log(2\pi e\mathrm{var}(X_i|W_1,W_2))\\
&\overset{(a)}{\geq} h(X^n|W_2)-\sum_{i=1}^{n}\frac{1}{2}\log(2\pi eE[(X_i-\hat{X}_i(W_1,W_2))^2])\\
&\overset{(b)}{\geq} n/2\log(2\pi e(\alpha\sigma_Y^2+\sigma_{N_1}^2))-n/2\log(\frac{2\pi e}{n}\sum_{i=1}^{n}E[(X_i-\hat{X}_i(W_1,W_2))^2])\\
&\geq n/2\log(2\pi e(\alpha\sigma_Y^2+\sigma_{N_1}^2))-n/2\log(2\pi eD)\\
&= n/2\log(\frac{\alpha\sigma_Y^2+\sigma_{N_1}^2}{D}),
\end{aligned}$$

where (a) follows from the fact that $\mathrm{var}(X_i|W_1,W_2)$ is the MMSE over all possible estimator of X_i for each $i=1,\ldots,n$, (b) follows from substituting $h(X^n|W_2)=n/2\log(2\pi e(\alpha\sigma_X^2+(1-\alpha)\sigma_{N_1}^2))$, and using Jensen's inequality [CT06, Theorem 2.6.2] and the fact that $\log(\cdot)$ is a concave function.

Lastly,

$$\begin{aligned}
n(\triangle+\delta_n) &\geq I(X^n;W_2,Z^n)\\
&= h(X^n)-h(X^n|W_2,Z^n)\\
&= h(X^n)-h(X^n,Z^n|W_2)+h(Z^n|W_2)\\
&\overset{(a)}{=} h(X^n)-h(X^n|W_2)-h(Z^n|X^n)+h(Z^n|W_2)\\
&\overset{(b)}{\geq} h(X^n)-h(X^n|W_2)-h(Z^n|X^n)+n/2\log(2^{\frac{2}{n}h(X^n|W_2)}+2^{\frac{2}{n}h(N_2^n|W_2)})\\
&\overset{(c)}{=} n/2\log\big(\frac{(\sigma_Y^2+\sigma_{N_1}^2)(\alpha\sigma_Y^2+\sigma_{N_1}^2+\sigma_{N_2}^2)}{\sigma_{N_2}^2(\alpha\sigma_Y^2+\sigma_{N_1}^2)}\big),
\end{aligned}$$

where (a) follows from the Markov chain $Z^n-X^n-W_2$, (b) follows from the conditional EPI and the fact that $Z^n=X^n+N_2^n$, X^n conditionally independent of N_2^n given W_2, (c) follows from substituting $h(X^n|W_2)$.

6.C Proof of Converse for Triangular Setting (A)

Proof of Converse: For any achievable tuple $(R_1, R_2, R_3, D, \triangle)$, by standard properties of the entropy function, it follows that

$$\begin{aligned} n(R_1 + R_3 + \delta_n) &\geq H(W_1, W_3) \\ &\geq I(X^n; W_1, W_3 | Y^n, Z^n) \\ &= H(X^n|Y^n, Z^n) - H(X^n|W_1, W_3, Y^n, Z^n) \\ &\overset{(a)}{=} H(X^n|Y^n) - H(X^n|W_1, W_2, W_3, Y^n, Z^n) \\ &\geq H(X^n|Y^n) - H(X^n|W_2, W_3, Y^n) \\ &\overset{(b)}{\geq} H(X^n|Y^n) - nD \\ &= \sum_{i=1}^{n} H(X_i|Y_i) - nD, \end{aligned}$$

where (a) follows from the Markov chains $W_2 - (W_1, Z^n) - (W_3, X^n, Y^n)$ and $X^n - Y^n - Z^n$, (b) follows from the fact that $D \geq E[d(X^n, g^{(n)}(W_2, W_3, Y^n))] \geq \frac{1}{n} H(X^n|W_2, W_3, Y^n)$ under the logarithmic loss distortion (Lemma 5.2).

Next,

$$\begin{aligned} n(R_2 + R_3 + \delta_n) &\geq H(W_2, W_3) \\ &\geq I(W_2, W_3; X^n|Y^n) \\ &= H(X^n|Y^n) - H(X^n|W_2, W_3, Y^n) \\ &\geq \sum_{i=1}^{n} H(X_i|Y_i) - nD, \end{aligned}$$

and the leakage rate

$$\begin{aligned} n(\triangle + \delta_n) &\geq I(X^n; W_1, Z^n) \\ &= I(X^n; Z^n) + I(X^n; W_1|Z^n) \\ &\overset{(a)}{=} I(X^n; Z^n) + H(W_1|Z^n) - H(W_1|X^n, Y^n, Z^n) \\ &\geq I(X^n; Z^n) + H(W_1|Y^n, Z^n) - H(W_1|X^n, Y^n, Z^n) \\ &= I(X^n; Z^n) + I(X^n; W_1|Y^n, Z^n) \\ &= I(X^n; Z^n) + I(X^n; W_1, W_3|Y^n, Z^n) - I(X^n; W_3|W_1, Y^n, Z^n) \\ &\geq I(X^n; Z^n) + I(X^n; W_1, W_3|Y^n, Z^n) - H(W_3) \\ &\overset{(b)}{\geq} \sum_{i=1}^{n} I(X_i; Z_i) + H(X_i|Y_i) - D - n(R_3 + \delta_n), \end{aligned}$$

where (a) follows from the Markov chain $W_1 - X^n - (Y^n, Z^n)$ and (b) follows from the steps used to bound $R_1 + R_3$, and lastly

$$\begin{aligned} n(\triangle + \delta_n) &\geq I(X^n; W_1, Z^n) \\ &\geq I(X^n; Z^n) \\ &= \sum_{i=1}^{n} I(X_i; Z_i). \end{aligned}$$

We end the proof by following the standard time-sharing argument and letting $n \to \infty$.

6.D Proof of Theorem 6.3.2

Since $X - Y - Z$ forms a Markov chain, we let the helper simply forward the index. Also, in the Gaussian setting with quadratic distortion, it is known that the side information at the encoder does not improve the rate distortion region, we neglect this side information in encoding. It is straightforward to show that a set of all tuples $(R_1, R_2, R_3, D, \triangle)$ satisfying the conditions below is the achievable region,

$$\begin{aligned} R_1 &\geq [I(X;U|Y) - R_3]^+, \\ R_2 &\geq [I(X;U|Y) - R_3]^+, \\ D &\geq E[d(X, \tilde{g}(U,Y))], \\ \triangle &\geq I(X;Z) + [I(X;U|Y) - R_3]^+, \end{aligned}$$

for some $P_{X,Y}P_{Z|Y}P_{U|X}$ and $\tilde{g}(\cdot)$.

With the assumption that $Y = X + N_1, N_1 \sim \mathcal{N}(0, \sigma^2_{N_1})$ independent of X, and $Z = Y + N_2, N_2 \sim \mathcal{N}(0, \sigma^2_{N_2})$ independent of X, Y, N_1, we will prove that the achievable region above is tight for this case.

Proof of Achievability: Let us choose $U = X + Q, Q \sim \mathcal{N}(0, \sigma^2_Q)$ independent of X, where $\sigma^2_Q = \frac{\sigma^2 D}{\sigma^2 - D}, \sigma^2 = \frac{\sigma^2_X \sigma^2_{N_1}}{\sigma^2_X + \sigma^2_{N_1}}$. Also, choose $\tilde{g}(U,Y)$ to be an MMSE estimate of X given U and Y.

With these choices of U and $\tilde{g}(\cdot)$, it can be shown that

$$\begin{aligned} I(X;U|Y) &= h(U|Y) - h(U|X,Y) \\ &= h(U|Y) - h(U|X) \\ &= \frac{1}{2}\log(\frac{\sigma^2_Q + \sigma^2}{\sigma^2_Q}) \\ &= \frac{1}{2}\log(\frac{\sigma^2}{D}), \end{aligned}$$

where $h(U|Y)$ and $h(U|X,Y)$ are the differential entropy and conditional differential entropy, as defined in [CT06, Chapter 8], and

$$\begin{aligned} I(X;Z) &= h(Z) - h(Z|X) \\ &= \frac{1}{2}\log(\frac{\sigma_X^2 + \sigma_{N_1}^2 + \sigma_{N_2}^2}{\sigma_{N_1}^2 + \sigma_{N_2}^2}), \end{aligned}$$

and lastly

$$\begin{aligned} E[d(X, \tilde{g}(U,Y))] &= E[(X - \tilde{g}(U,Y))^2] \\ &= \mathrm{var}(X|U,Y) \\ &= \frac{\sigma^2\sigma_Q^2}{\sigma^2 + \sigma_Q^2} = D, \end{aligned}$$

where $\mathrm{var}(X|U,Y) = \frac{\mathrm{var}(X|Y)\sigma_Q^2}{\mathrm{var}(X|Y)+\sigma_Q^2}$.

Proof of Converse: From the problem formulation the joint PMF is given by

$$P_{X^n,Y^n} P_{Z^n|Y^n} P_{W_1|X^n} P_{W_3|X^n} P_{W_2|W_1,Z^n} 1_{\{\hat{X}^n = g^{(n)}(W_2,W_3,Y^n)\}}.$$

It follows that

$$\begin{aligned} n(R_1 + R_3 + \delta_n) &\geq H(W_1, W_3) \\ &\geq I(W_1, W_3; X^n|Y^n, Z^n) \\ &\overset{(a)}{=} h(X^n|Y^n) - h(X^n|W_1, W_3, W_2, Y^n, Z^n) \\ &\geq h(X^n|Y^n) - \sum_{i=1}^n h(X_i|W_2, W_3, Y^n) \\ &\geq h(X^n|Y^n) - \sum_{i=1}^n \frac{1}{2}\log(2\pi e \mathrm{var}(X_i|W_2, W_3, Y^n)) \\ &\overset{(b)}{\geq} h(X^n|Y^n) - \sum_{i=1}^n \frac{1}{2}\log(2\pi e E[(X_i - \hat{X}_i(W_2, W_3, Y^n))^2]) \\ &\overset{(c)}{\geq} n/2\log(2\pi e\sigma^2) - n/2\log(\frac{2\pi e}{n}\sum_{i=1}^n E[(X_i - \hat{X}_i(W_2, W_3, Y^n))^2]) \\ &\geq n/2\log(2\pi e\sigma^2) - n/2\log(2\pi eD) \\ &= n/2\log(\frac{\sigma^2}{D}), \end{aligned}$$

where (a) follows from the Markov chain $X^n - Y^n - Z^n$ and the Markov chain $W_2 - (W_1, Z^n) - (W_3, X^n, Y^n)$, (b) follows from the fact that $\mathrm{var}(X_i|W_2, W_3, Y^n)$ is the MMSE over all possible estimator of X_i for each $i = 1, \ldots, n$, (c) follows from Jensen's inequality and the fact that $\log(\cdot)$ is a concave function.

$$\begin{aligned} n(R_2+R_3+\delta_n) &\geq H(W_2,W_3) \\ &\geq I(W_2,W_3;X^n|Y^n) \\ &\geq h(X^n|Y^n)-\sum_{i=1}^{n} h(X_i|W_2,W_3,Y^n) \\ &\overset{(a)}{\geq} n/2\log(\frac{\sigma^2}{D}), \end{aligned}$$

where (a) follows from steps used to prove the constraint on R_1+R_3.

Lastly,

$$\begin{aligned} n(\triangle+\delta_n) &\geq I(X^n;W_1,Z^n) \\ &= I(X^n;Z^n)+I(X^n;W_1|Z^n) \\ &\overset{(a)}{=} I(X^n;Z^n)+I(X^n,Y^n;W_1|Z^n) \\ &\geq I(X^n;Z^n)+I(X^n;W_1|Y^n,Z^n) \\ &= I(X^n;Z^n)+I(X^n;W_1,W_3|Y^n,Z^n)-I(X^n;W_3|W_1,Y^n,Z^n) \\ &\geq I(X^n;Z^n)+I(X^n;W_1,W_3|Y^n,Z^n)-H(W_3) \\ &\overset{(b)}{\geq} n/2\log(\frac{\sigma_X^2+\sigma_{N_1}^2+\sigma_{N_2}^2}{\sigma_{N_1}^2+\sigma_{N_2}^2})+n/2\log(\frac{\sigma^2}{D})-n(R_3+\delta_n), \end{aligned}$$

where (a) follows from the Markov chain $W_1-(X^n,Z^n)-Y^n$, (b) follows from steps used to prove the constraint on R_1+R_3.

The constraint $\triangle+\delta_n \geq 1/2\log(1+\frac{\sigma_X^2}{\sigma_{N_1}^2+\sigma_{N_2}^2})$ follows straightforwardly from $n(\triangle+\delta_n)\geq I(X^n;Z^n)$.

6.E Proof of Theorem 6.3.4 for Triangular Setting (B)

Proof of Achievabiltiy: The proof follows standard random coding arguments where we show the existence of a code that satisfies the rate, distortion, and leakage rate constraints. The outline of the proof is given in the following.

Codebook generation: Fix $P_{U|X}$, and the function $\tilde{g}:\mathcal{U}\times\mathcal{Y}\to\hat{\mathcal{X}}$.

- Randomly and independently generating $2^{n(I(X;U)+\delta_\epsilon)}$ codewords $u^n(w)\sim\prod_{i=1}^n P_U(u_i(w)), w\in[1:2^{n(I(X;U)+\delta_\epsilon)}]$.
- Then distributed them uniformly at random into $2^{n(I(X;U|Y)+2\delta_\epsilon)}$ bins $b_U(w_u)$, where $w_u\in[1:2^{n(I(X;U|Y)+2\delta_\epsilon)}]$.
- We split the bin indices w_u into $w_{u,1}\in[1:2^{n(I(X,U|Y)-R_3+\delta_\epsilon)}]$ and $w_{u,3}\in[1:2^{n(R_3+\delta_\epsilon)}]$.

- For secret key generation codebook, we randomly and uniformly partition the set of sequences $\mathcal{Y}^n$ into 2^{nR_k} bins $b_K(k)$, $k \in [1 : 2^{nR_k}]$, where $R_k = \min\{H(Y|X,Z), H(X|Y) - D - R_3\} - 2\delta_\epsilon$.

The codebooks are revealed to the encoder, the helper, the decoder, and the eavesdropper. We consider the following two cases.

I) If $H(X|Y) < D$, we do not need to send anything over the rate-limited links. Since the decoder knows y^n, it can generate $\hat{x}^n$ based on y^n. Since (x^n, y^n) are jointly typical, it can be shown that this is sufficient to satisfy the distortion, i.e., under the logarithmic loss distortion, we have $E[d(X, \tilde{g}(Y))] = H(X|Y)$.

II) If $H(X|Y) - D > 0$: We further split the bin indices $w_{u,1}$ into $w_{u,1k} \in [1 : 2^{nR_k}]$ and $w_{u,1l} \in [1 : 2^{n(I(X,U|Y)-R_3-R_k+\delta_\epsilon)}]$. Note that this is possible if $R_k \le I(X,U|Y) - R_3 + \delta_\epsilon$. Note also that $w_{u,1}$ can be deduced from $(w_{u,1k}, w_{u,1l})$.

Encoding at the encoder and helper:

- Given sequences (x^n, y^n), the encoder looks for u^n that is jointly typical with x^n. If there is more than one, it selects one of them uniformly at random. If there is no such u^n, it selects one out of $2^{n(I(X;U)+\delta_\epsilon)}$ uniformly at random. With high probability, there exists such u^n since there are $2^{n(I(X;U)+\delta_\epsilon)}$ codewords u^n generated.

- Then the encoder wishes to transmit the corresponding bin index w_u to the helper and decoder in a secure way by incorporating the secret key, e.g., using "one-time pad" based on the key. To generate a secret key, the encoder looks for an index k for which $y^n \in b_K(k)$. Then the encoder transmits $w_{u,1k} \oplus k$ and $w_{u,1l}$ to the helper over the cascade link, where $w_{u,1k} \oplus k$ denotes the modulo operation, $(w_{u,1k} + k) \mathrm{mod} 2^{nR_k}$, and also transmits $w_{u,3}$ to the decoder over the private (triangular) link. The helper simply forwards the indices $w_{u,1k} \oplus k$ and $w_{u,1l}$ to the decoder.

Decoding at the decoder: Upon receiving $w_{u,1k} \oplus k$, $w_{u,1l}$, and $w_{u,3}$, the decoder uses its side information y^n to generate its own key and decrypt the index $w_{u,1k}$, and thus the bin index w_u. Then it looks for a unique u^n that is jointly typical with y^n. With high probability, it will find the unique and correct one since there are $2^{n(I(Y;U)-\delta_\epsilon)}$ codewords in each bin $b_U(w_u)$. The decoder puts out $\hat{x}^n$ where $\hat{x}_i = \tilde{g}(u_i, y_i)$.

Analysis of distortion: Since (x^n, y^n, u^n) are jointly typical, we can show that D satisfying $D \ge E[d(X, \tilde{g}(U,Y))]$ is achievable. Also, due to the property of log-loss distortion function (Lemma 5.1), we have that $E[d(X, \tilde{g}(U,Y))] = H(X|U,Y)$. We define $U = X$ with probability $p = 1 - \frac{D}{H(X|Y)}$ and a constant otherwise. This gives us $H(X|U,Y) = (1-p)H(X|Y) = D$.

Analysis of leakage rate: The leakage averaged over all codebooks $\mathcal{C}_n$ can be

bounded as follows.

$$
\begin{aligned}
&I(X^n; W_{u,1l}, W_{u,1k} \oplus K, Z^n|\mathcal{C}_n) \\
&= I(X^n; Z^n) + I(X^n; W_{u,1l}|Z^n, \mathcal{C}_n) + I(X^n; W_{u,1k} \oplus K|W_{u,1l}, Z^n, \mathcal{C}_n) \\
&\leq I(X^n; Z^n) + H(W_{u,1l}|\mathcal{C}_n) + H(W_{u,1k} \oplus K|\mathcal{C}_n) \\
&\quad - H(W_{u,1k} \oplus K|W_{u,1l}, X^n, Z^n, \mathcal{C}_n) \\
&\overset{(a)}{=} I(X^n; Z^n) + H(W_{u,1l}|\mathcal{C}_n) + H(W_{u,1k} \oplus K|\mathcal{C}_n) - H(K|X^n, Z^n, \mathcal{C}_n) \\
&\overset{(b)}{\leq} n[I(X; Z) + H(X|Y) - D - R_3 - R_k + \delta_\epsilon + R_k] - H(K|X^n, Z^n, \mathcal{C}_n) \\
&\overset{(c)}{=} n[I(X; Z) + H(X|Y) - D - R_3 - R_k + \delta_\epsilon + R_k] - I(K; Y^n|X^n, Z^n, \mathcal{C}_n) \\
&\overset{(d)}{\leq} n[I(X; Z) + H(X|Y) - D - R_3 - R_k + \delta'_\epsilon] \\
&\leq n[\triangle + \delta'_\epsilon]
\end{aligned}
$$

if $\triangle \geq I(X; Z) + H(X|Y) - D - R_3 - R_k$, where (a) follows from the fact that $W_{u,1l}, W_{u,1k}$ are functions of X^n and $\mathcal{C}_n$, (b) follows from the codebook generation, and (c) follows from the fact that K is a function of Y^n and $\mathcal{C}_n$, and (d) follows from bounding the term $H(Y^n|X^n, Z^n, K, \mathcal{C}_n)$ ([EK11, lemma 22.3] when setting $\tilde{R} = 0$), given that $R_k < H(Y|X, Z) - \delta_\epsilon$ which holds due to the assumption that $R_k = \min\{H(Y|X, Z), H(X|Y) - D - R_3\} - 2\delta_\epsilon$ in the beginning.

Proof of Converse: For any achievable tuple $(R_1, R_2, R_3, D, \triangle)$, the constraints on $R_1 + R_3$ and $R_2 + R_3$ follow the proof of triangular setting (A) (Appendix 6.C). As for the leakage rate, we have

$$
\begin{aligned}
n(\triangle + \delta_n) &\geq I(X^n; W_1, Z^n) \\
&= I(X^n; Z^n) + I(X^n; W_1|Z^n) \\
&= I(X^n; Z^n) + I(X^n; W_1, Y^n|Z^n) - I(X^n; Y^n|W_1, Z^n) \\
&= I(X^n; Z^n) + I(X^n; Y^n|Z^n) + I(X^n; W_1|Y^n, Z^n) - I(X^n; Y^n|W_1, Z^n) \\
&\geq I(X^n; Z^n) - H(Y^n|X^n, Z^n) + I(X^n; W_1|Y^n, Z^n) \\
&= I(X^n; Z^n) - H(Y^n|X^n, Z^n) + I(X^n; W_1, W_3|Y^n, Z^n) \\
&\quad - I(X^n; W_3|W_1, Y^n, Z^n) \\
&\geq I(X^n; Z^n) - H(Y^n|X^n, Z^n) + I(X^n; W_1, W_3|Y^n, Z^n) - H(W_3) \\
&\overset{(a)}{\geq} \sum_{i=1}^{n} I(X_i; Z_i) - H(Y_i|X_i, Z_i) + H(X_i|Y_i) - D - n(R_3 + \delta_n),
\end{aligned}
$$

where (a) follows from the steps used to bound $R_1 + R_3$.

Also,

$$
\begin{aligned}
n(\triangle + \delta_n) &\geq I(X^n; W_1, Z^n) \\
&= I(X^n; Z^n)
\end{aligned}
$$

$$= \sum_{i=1}^{n} I(X_i; Z_i).$$

We end the proof by following the standard time-sharing argument and letting $n \to \infty$.

6.F Proof of Theorem 6.3.5

Proof of Achievability: The rate and distortion constraints are the same as in the Gaussian triangular example in Setting (A). The proof follows Wyner's partitioning approach for the Gaussian Wyner-Ziv problem [Wyn78]. The leakage rate constraint however requires some new analysis. We will first show that the leakage rate $\triangle$ satisfying $\triangle > I(X, Z)$ is achievable.

We note that the side information Y^n is distributed according to Gaussian distribution on $\mathbb{R}^n$. Let $\mathcal{X}_p, \mathcal{Y}_p, \mathcal{Z}_p$ be discrete sets corresponding to the partitioned version of $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ to mutually exclusive sets whose union is the entire set. Following the argument in [CT06, Chapter 8], [Gal68], as we can partition $\mathbb{R}$ as fine as we wish, there exist finite partitions on $\mathcal{Y}, \mathcal{X}, \mathcal{Z}$ such that $H(Y_p|X_p, Z_p)$ can be made arbitrarily large. For example, there exist finite partitions such that

$$H(Y_p|X_p, Z_p) \geq I(X_p; U_p|Y_p) - R_3, \tag{6.18}$$

where $I(X_p; U_p|Y_p) - R_3$ is a term associated with the source description rate on the cascade link (Wyner-Ziv rate) in the discrete case.

The remaining proof steps follows similarly to those of the proof for Theorem 6.3.4. To generate a secret key, we randomly and uniformly partition the set $\mathcal{Y}_p^n$ into 2^{nR_k} bins $\mathcal{B}(k)$, where k is the bin index, and we set $R_k = I(X_p; U_p|Y_p) - R_3 - 2\delta_\epsilon$. At the encoder and the decoder, given $y^n \in \mathcal{Y}^n$ which is mapped to $y_p^n \in \mathcal{Y}_p^n$, the secret key is chosen to be the bin index k where $y_p^n \in \mathcal{B}(k)$. Note that, with this key rate, we are able to scramble *essentially* the whole source description w_1. For example, we may consider splitting the source description (Wyner-Ziv index) $w_1 \in [1 : 2^{n(I(X_p;U_p|Y_p)-R_3+\delta_\epsilon)}]$ into two parts, $w_{1,l} \in [1 : 2^{3n\delta_\epsilon}]$, and $w_{1,k} \in [1 : 2^{nR_k}]$, and transmit $w_{1,l}$ and $w_{1,k} \oplus k$ to the helper, where $w_{1,k} \oplus k$ denotes the modulo operation $(w_{1,k} + k) \mathrm{mod} 2^{nR_k}$.

To analyze the leakage rate averaged over all codebooks $\frac{1}{n} I(X^n; W_{1,l}, W_{1,k} \oplus K, Z^n|C_n)$, we first argue that, for any $\epsilon' > 0$, there exist finite partitions of $\mathcal{X}, \mathcal{Y}$, and $\mathcal{Z}$ such that $\frac{1}{n} I(X_p^n; W_{1,l}, W_{1,k} \oplus K, Z_p^n|\mathcal{C}_n) \geq \frac{1}{n} I(X^n; W_{1,l}, W_{1,k} \oplus K, Z^n|\mathcal{C}_n) - \epsilon'$. The analysis can then be done using the similar discrete proof as in the achievability proof of Theorem 6.3.4, i.e.,

$$\begin{aligned}
&I(X^n; W_{1,l}, W_{1,k} \oplus K, Z^n|\mathcal{C}_n) \\
&\leq I(X_p^n; W_{1,l}, W_{1,k} \oplus K, Z_p^n|\mathcal{C}_n) + n\epsilon' \\
&= I(X_p^n; Z_p^n) + I(X_p^n; W_{1,l}, W_{1,k} \oplus K|Z_p^n, \mathcal{C}_n) + n\epsilon'
\end{aligned}$$

$$
\begin{aligned}
&\leq I(X_p^n; Z_p^n) + H(W_{1,l}, W_{1,k} \oplus K|\mathcal{C}_n) - H(W_{1,l}, W_{1,k} \oplus K|X_p^n, Z_p^n, \mathcal{C}_n) + n\epsilon' \\
&\overset{(a)}{=} I(X_p^n; Z_p^n) + H(W_{1,l}, W_{1,k} \oplus K|\mathcal{C}_n) - H(K|X_p^n, Z_p^n, \mathcal{C}_n) + n\epsilon' \\
&\overset{(b)}{\leq} n[I(X_p; Z_p) + R_k + 3\delta_\epsilon + \epsilon'] - H(K|X_p^n, Z_p^n, \mathcal{C}_n) \\
&\overset{(c)}{=} n[I(X_p; Z_p) + R_k + 3\delta_\epsilon + \epsilon'] - I(K; Y_p^n|X_p^n, Z_p^n, \mathcal{C}_n) \\
&\overset{(d)}{\leq} n[I(X_p; Z_p) + \delta'_\epsilon] \\
&\overset{(e)}{\leq} n[I(X; Z) + \delta'_\epsilon],
\end{aligned}
$$

where (a) follows from the fact that $(W_{1,l}, W_{1,k})$ is a function of X_p^n and $\mathcal{C}_n$, (b) follows from the codebook generation, (c) follows from the fact that K is a function of Y_p^n and $\mathcal{C}_n$, (d) follows from bounding the term $H(Y_p^n|X_p^n, Z_p^n, K, \mathcal{C}_n) \leq n(H(Y_p|X_p, Z_p) - R_k + \delta_\epsilon)$ ([EK11, lemma 22.3] when setting $\tilde{R} = 0$), given that $R_k < H(Y_p|X_p, Z_p) - \delta_\epsilon$ which holds due to the assumption that $R_k = I(X_p; U|Y_p) - R_3 - 2\delta_\epsilon$ and (6.18) in the beginning, and (e) follows from the Markov chain $X_p - X - Z - Z_p$. With the same choice of U_p as U in Theorem 6.3.2, we have proved the achievability part.

Proof of Converse: The converse part also follows similarly that of Theorem 6.3.2, where the constraint $\triangle + \delta_n \geq 1/2 \log(1 + \frac{\sigma_X^2}{\sigma_{N_1}^2 + \sigma_{N_2}^2})$ follows straightforwardly from $n(\triangle + \delta_n) \geq I(X^n; W_1, Z^n) \geq I(X^n; Z^n)$.

6.G Proof of Converse for Triangular Setting (C)

Proof of Converse: We define $U_i \triangleq (W_2, W_3, X^{i-1}, Y^{n\setminus i})$ which satisfies $U_i - X_i - Y_i$ for all $i = 1, \ldots, n$. For any achievable tuple $(R_1, R_2, R_3, D, \triangle)$, by standard properties of the entropy function, it follows that

$$
\begin{aligned}
n(R_1 + R_3 + \delta_n) &\geq H(W_1, W_3) \\
&\geq I(X^n, W_1, W_3|Y^n) \\
&= H(X^n|Y^n) - H(X^n|W_1, W_3, Y^n) \\
&\overset{(a)}{=} H(X^n|Y^n) - H(X^n|W_1, W_2, W_3, Y^n) \\
&\geq H(X^n|Y^n) - H(X^n|W_2, W_3, Y^n) \\
&= \sum_{i=1}^n H(X_i|Y_i) - H(X_i|W_2, W_3, X^{i-1}, Y^n) \\
&\overset{(b)}{=} \sum_{i=1}^n H(X_i|Y_i) - H(X_i|U_i, Y_i) \\
&= \sum_{i=1}^n I(X_i; U_i|Y_i),
\end{aligned}
$$

where (a) follows from the Markov chain $W_2 - W_1 - (W_3, X^n, Y^n)$ and (b) follows from the definition of U_i.

Next,

$$\begin{aligned} n(R_2 + R_3 + \delta_n) &\geq H(W_2, W_3) \\ &\geq I(W_2, W_3; X^n|Y^n) \\ &= H(X^n|Y^n) - H(X^n|W_2, W_3, Y^n) \\ &\geq \sum_{i=1}^{n} H(X_i|Y_i) - H(X_i|U_i, Y_i) \\ &= \sum_{i=1}^{n} I(X_i; U_i|Y_i). \end{aligned}$$

For the bound on distortion, we have

$$\begin{aligned} D + \delta_n &\geq \frac{1}{n} \sum_{i=1}^{n} E[d(X_i, g_i^{(n)}(W_2, W_3, Y^n))] \\ &\geq \frac{1}{n} \sum_{i=1}^{n} E[d(X_i, g_i(U_i, Y_i))], \end{aligned}$$

and lastly, the leakage rate

$$\begin{aligned} n(\triangle + \delta_n) &\geq I(X^n; W_1) \\ &\overset{(a)}{=} I(X^n, Y^n; W_1) \\ &= I(X^n, Y^n; W_1, W_3) - I(X^n, Y^n; W_3|W_1) \\ &\geq I(X^n; W_1, W_3|Y^n) - H(W_3) \\ &\overset{(b)}{\geq} \sum_{i=1}^{n} I(X_i; U_i|Y_i) - R_3 - \delta_n, \end{aligned}$$

where (a) follows from the Markov chain $W_1 - X^n - Y^n$ and (b) follows from the steps used to bound $R_1 + R_3$. We end the proof by following the standard time-sharing argument and letting $n \to \infty$.

For the bound on the cardinality of the set $\mathcal{U}$, it can be shown by using the support lemma [CK11, Lemma 15.4] that it suffices that $\mathcal{U}$ should have $|\mathcal{X}| - 1$ elements to preserve P_X, plus two more for $H(X|U, Y)$ and the distortion constraint.

6.H Proof of Converse for Triangular Setting (D)

Proof of Converse: Define $U_i \triangleq (W_2, X^{i-1}, Z^{i-1}, Y^{n\setminus i})$ and $V_i \triangleq W_3$ which satisfies $(U_i, V_i) - (X_i, Z_i) - Y_i$ for all $i = 1, \ldots, n$. For any achievable tuple $(R_1, R_2, R_3, D, \triangle)$,

by standard properties of the entropy function, it follows that

$$\begin{aligned}
n(R_1+\delta_n) &\geq H(W_1)\\
&\geq I(X^n, W_1|Y^n, Z^n)\\
&= H(X^n|Y^n, Z^n) - H(X^n|W_1, Y^n, Z^n)\\
&\overset{(a)}{=} H(X^n|Z^n) - H(X^n|W_1, W_2, Y^n, Z^n)\\
&\geq \sum_{i=1}^{n} H(X_i|Z_i) - H(X_i|W_2, X^{i-1}, Z^{i-1}, Y^{n\setminus i}, Z_i)\\
&\overset{(b)}{=} \sum_{i=1}^{n} H(X_i|Z_i) - H(X_i|U_i, Z_i)\\
&= \sum_{i=1}^{n} I(X_i, U_i|Z_i),
\end{aligned}$$

where (a) follows from the Markov chains $W_2 - (W_1, Z^n) - (X^n, Y^n)$ and $X^n - Z^n - Y^n$, (b) follows from the definition of U_i.

Next,

$$\begin{aligned}
n(R_2+\delta_n) &\geq H(W_2) \geq I(W_2; X^n, Z^n|Y^n)\\
&= H(X^n, Z^n|Y^n) - H(X^n, Z^n|W_2, Y^n)\\
&= \sum_{i=1}^{n} H(X_i, Z_i|Y_i) - H(X_i, Z_i|W_2, X^{i-1}, Z^{i-1}, Y^n)\\
&= \sum_{i=1}^{n} H(X_i, Z_i|Y_i) - H(X_i, Z_i|U_i, Y_i)\\
&= \sum_{i=1}^{n} I(X_i, Z_i; U_i|Y_i),
\end{aligned}$$

and

$$\begin{aligned}
n(R_3+\delta_n) &\geq H(W_3) \geq I(W_3; X^n, Z^n|W_2, Y^n)\\
&= H(X^n, Z^n|W_2, Y^n) - H(X^n, Z^n|W_2, W_3, Y^n)\\
&= \sum_{i=1}^{n} H(X_i, Z_i|W_2, X^{i-1}, Z^{i-1}, Y^n)\\
&\qquad - H(X_i, Z_i|W_2, W_3, X^{i-1}, Z^{i-1}, Y^n)\\
&= \sum_{i=1}^{n} H(X_i, Z_i|U_i, Y_i) - H(X_i, Z_i|U_i, V_i, Y_i)\\
&= \sum_{i=1}^{n} I(X_i, Z_i; V_i|U_i, Y_i).
\end{aligned}$$

For the bound on distortion, we have

$$\begin{aligned} D + \delta_n &\geq \frac{1}{n}\sum_{i=1}^{n} E[d(X_i, g_i^{(n)}(W_2, W_3, Y^n))] \\ &\geq \frac{1}{n}\sum_{i=1}^{n} E[d(X_i, g_i(U_i, V_i, Y_i))], \end{aligned}$$

and lastly, the leakage rate

$$\begin{aligned} n(\triangle + \delta_n) &\geq I(X^n; W_1, Z^n) \\ &\overset{(a)}{=} H(X^n) - H(X^n|W_1, Z^n, Y^n) \\ &\overset{(b)}{=} H(X^n) - H(X^n|W_1, W_2, Z^n, Y^n) \\ &= \sum_{i=1}^{n} H(X_i) - H(X_i|W_1, W_2, X^{i-1}, Z^n, Y^n) \\ &\geq \sum_{i=1}^{n} I(X_i; U_i, Z_i), \end{aligned}$$

where (a) follows from the Markov chain $(W_1, X^n) - Z^n - Y^n$, (b) follows from the Markov chain $W_2 - (W_1, Z^n) - (X^n, Y^n)$. We end the proof by following the standard time-sharing argument and letting $n \to \infty$.

For the bounds on the cardinalities of the sets $\mathcal{U}$ and $\mathcal{V}$, it can be shown by using the support lemma [CK11, Lemma 15.4] that it suffices that $\mathcal{U}$ should have $|\mathcal{X}||\mathcal{Z}| - 1$ elements to preserve $P_{X,Z}$, plus four more for $H(X|U, Z)$, $I(X, Z; U|Y)$, $I(X, Z; V|U, Y)$, and the distortion constraint. The new variable U induces the new V, and for each $U = u$, it suffices to consider $|\mathcal{V}| \leq |\mathcal{X}||\mathcal{Z}| + 1$ so that $P_{X,Z|U=u}$, $I(X, Z; V|U = u, Y)$, and the distortion constraint are preserved. Thus, the overall cardinality bound for $\mathcal{V}$ is $|\mathcal{V}| \leq |\mathcal{U}|(|\mathcal{X}||\mathcal{Z}| + 1) \leq (|\mathcal{X}||\mathcal{Z}| + 3)(|\mathcal{X}||\mathcal{Z}| + 1)$.

6.I Proof of Theorem 6.3.10

Proof of Achievability: The achievability proof follows from a standard random coding argument.

Codebook Generation: Fix $P_{U|X,Z}$ and $\tilde{g} : \mathcal{U} \times \mathcal{Y} \to \hat{\mathcal{X}}$. Let $\mathcal{W}_1^{(n)} = [1 : 2^{nR_1}]$, $\mathcal{W}_2^{(n)} = [1 : 2^{nR_2}]$, and $\mathcal{W}'^{(n)} = [1 : 2^{nR'}]$. The codewords $u^n(w_1, w_2, w')$ are generated i.i.d. each according to $\prod_{i=1}^{n} P_U(u_i)$, for $(w_1, w_2, w') \in \mathcal{W}_1^{(n)} \times \mathcal{W}_2^{(n)} \times \mathcal{W}'^{(n)}$. The codebook is then revealed to the encoder, helper, and decoder.

Encoder: Given a source sequence x^n, and side information z^n the encoder first looks for $u^n(w_1, w_2, w')$ that is jointly typical with (x^n, z^n). If there exists such a codeword, the encoder transmits the smallest w_1 to the helper and the decoder. If

not successful, the encoder transmits $w_1 = 1$. By the covering lemma (Lemma 2.6), the encoder is successful if $R_1 + R_2 + R' > I(X, Z; U) + \delta_\epsilon$.

Helper: Given an index w_1, and side information z^n the helper looks for a unique (w_2, w') such that $u^n(w_1, w_2, w')$ is jointly typical with z^n. If successful, the helper transmits the corresponding w_2 to the decoder. If not successful, the helper transmits $w_2 = 1$. By the packing lemma (Lemma 2.7), the helper is successful if $R_2 + R' < I(Z; U) - \delta_\epsilon$.

Decoder: Given the indices w_1 and w_2, and the side information y^n the decoder looks for a unique $u^n(w_1, w_2, w')$ such that it is jointly typical with y^n. If successful, the decoder reconstructs the source as $\hat{x}^n$ where $\hat{x}_i = \tilde{g}(u_i(w_1, w_2, w'), y_i)$. Otherwise, the decoder puts out $\hat{x}^n$ where $\hat{x}_i = \tilde{g}(u_i(w_1, w_2, 1), y_i)$. By the packing lemma, the decoder is successful if $R' < I(Y; U) - \delta_\epsilon$.

By combining the bounds on the code rates above, we obtain $R_1 > I(X; U|Z) + 2\delta_\epsilon$, and $R_1 + R_2 > I(X, Z; U|Y) + 2\delta_\epsilon$. Analysis of the distortion constraint follows standard arguments using the fact that (X^n, U^n, Y^n) are jointly typical. The leakage analysis follows similarly as in the proof of Theorem 6.3.9. This concludes the achievability proof.

Proof of Converse: Define $U_i \triangleq (W_1, W_2, X^{i-1}, Z^{i-1}, Y^{n\setminus i})$ which satisfies the Markov chains $(U_i, X_i) - Z_i - Y_i$ for all $i = 1, \ldots, n$. The proof of constraints on R_1, D and $\triangle$ follow similarly as in that of Theorem 6.3.9. As for the sum rate $R_1 + R_2$, it follows that

$$\begin{aligned}
n(R_1 + R_2 + \delta_n) &\geq H(W_1, W_2) \geq I(W_1, W_2; X^n, Z^n|Y^n) \\
&= H(X^n, Z^n|Y^n) - H(X^n, Z^n|W_1, W_2, Y^n) \\
&= \sum_{i=1}^{n} H(X_i, Z_i|Y_i) - H(X_i, Z_i|W_1, W_2, X^{i-1}, Z^{i-1}, Y^n) \\
&= \sum_{i=1}^{n} H(X_i, Z_i|Y_i) - H(X_i, Z_i|U_i, Y_i) \\
&= \sum_{i=1}^{n} I(X_i, Z_i; U_i|Y_i).
\end{aligned}$$

For the bounds on the cardinalities of the sets $\mathcal{U}$, it can be shown by using the support lemma [CK11, Lemma 15.4] that it suffices that $\mathcal{U}$ should have $|\mathcal{X}||\mathcal{Z}| - 1$ elements to preserve $P_{X,Z}$, plus three more for $H(X|U, Z)$, $I(X, Z; U|Y)$, and the distortion constraint.

Chapter 7

Lossy Source Coding With Reconstruction Privacy

In Chapters 5 and 6, we studied source privacy in various lossy source coding problems such as source coding with action-dependent side information and source coding with a public helper. We characterized the rate-distortion-(cost)-leakage regions which exhibit the optimal tradeoff among system performances. In this chapter, we consider a new aspect of privacy in the lossy source coding problem, namely privacy of the reconstruction sequence. In particular, we consider the problem of lossy source coding with side information under a privacy constraint that the reconstruction sequence at a decoder should be kept secret to a certain extent from another terminal such as an eavesdropper, a sender, or a helper. We are interested in how the reconstruction privacy constraint at a particular terminal affects the rate-distortion tradeoff. We allow the decoder to use a random mapping, and give inner and outer bounds to the rate-distortion-equivocation region for the cases where the side information is available noncausally and causally at the decoder. In the special case where each reconstruction symbol depends only on the source description and current side information symbol, the complete rate-distortion-equivocation region is characterized. A binary example illustrating a new tradeoff due to the new privacy constraint, and a gain from the use of a stochastic decoder is given.

7.1 Introduction

With the growing predominance of the Internet and the advance of cloud computing, significant amount of data will be exchanged among users and service providers. which inevitably leads to a privacy concern. A user in the network could receive different versions of certain information from different sources. Apart from being able to process the information efficiently, the user may also wish to protect the privacy of his/her action which is taken based on the received information. In this chapter, we address the privacy concern of the final action/decision taken

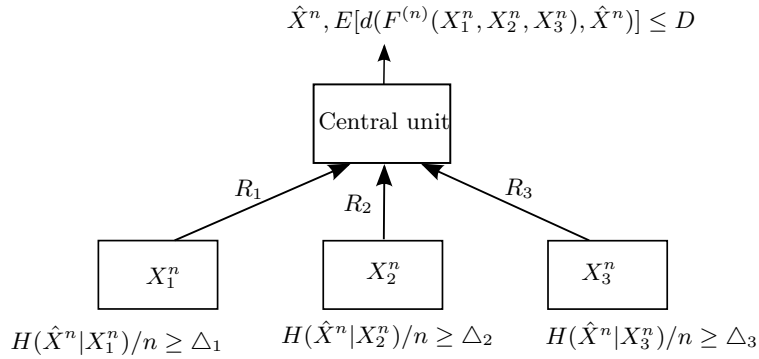

Figure 7.1: Multiterminal source coding with end-user privacy.

at the end-user in an information theoretic setting. More specifically, we consider the problem of lossy source coding under the privacy constraint of the end-user (decoder) whose goal is to reconstruct a sequence subject to a distortion criterion. The privacy concern of the end-user may arise due to the presence of an external eavesdropper or a legitimate terminal such as a sender or a helper who is curious about the final reconstruction. We term the privacy criterion as *end-user privacy*, and use the normalized equivocation of the reconstruction sequence at a particular terminal as a privacy measure.

Let us consider Fig. 7.1 where there exist several agents collecting information for the central unit. Assuming that the agents communicate efficient representations of the correlated sources to the central unit through rate-limited noiseless links so that the central unit is able to estimate a value of some function of the sources $F^{(n)}(X_1^n, X_2^n, X_3^n)$ satisfying the distortion criterion. However, there is a privacy concern regarding the reconstruction sequence (final decision/action) at the central unit, that it should be kept secret from the agents. This gives rise to a new tradeoff between the achievable rate-distortion pair and privacy of the reconstruction sequence. That is, the central unit should reconstruct a sequence in such a way that it satisfies both distortion and equivocation constraints which can be contradicting. Potential applications of the illustrated setting include those in the area of distributed cloud services where the end-user (central unit) can process information received from the cloud service providers (agents), while guaranteeing that his/her final action will be kept private from the providers, at least to a certain extent.

In this chapter, we study a special case of the problem in Fig. 7.1 where there are two sources, one of which is available directly at the decoder. For example (see Fig. 7.2), we let X^n be the source to be encoded, and Y^n be the uncoded source available at the decoder. Alternatively, we may view Y^n as correlated side information provided by a *helper*[1]. The reconstruction sequence $\hat{X}^n$ is an estimate

[1]Here we term a node who only has access to Y^n as a *helper* because it connects to the setting

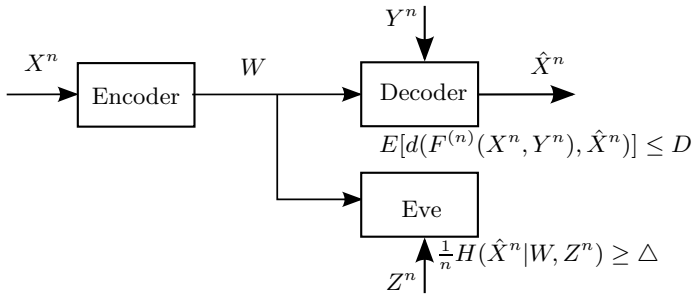

Figure 7.2: Source coding with end-user privacy at eavesdropper.

of the value of some component-wise function $F^{(n)}(X^n, Y^n)$, where the i^{th} component $F_i^{(n)}(X^n, Y^n) = F(X_i, Y_i)$ for $i = 1, \ldots, n$. Without the end-user privacy constraint, this corresponds to the problem of source coding with side information at the decoder or the Wyner-Ziv problem [WZ76], [Yam82]. We consider three scenarios where the end-user privacy constraint is imposed at different nodes, namely the eavesdropper, the encoder, and the helper, as shown in Fig. 7.2, 7.3, and 7.4. Since the goal of end-user privacy is to protect the reconstruction sequence generated at the decoder against any unwanted inferences, we allow the decoder mapping to be a random mapping. It can be shown by an example that a stochastic decoder can enlarge the rate-distortion-equivocation region as compared to the one derived for deterministic decoders.[2]

7.1.1 Overview of Problem Settings and Organization

We study an implication of the end-user privacy constraint on the rate-distortion tradeoff where the privacy constraint is imposed at different nodes in the system. A summary of contribution is given below.

- Section 7.2 considers end-user privacy at the eavesdropper, as depicted in Fig. 7.2. It corresponds to a scenario where there is an eavesdropper observing the source description and its side information, and we wish to prevent it from inferring the final reconstruction. We give inner and outer bounds to the rate-distortion-equivocation region for the cases where the side information is available noncausally and causally at the decoder. In a special case of causal side information where the decoder has no *memory*, that is, each

in Fig. 7.1 in a broader sense.

[2]Although the use of a stochastic *encoder* might also help especially if the decoder is deterministic, we restrict ourself to the deterministic encoder here. Conservatively, it might be reasonable to assume that only the end-user is willing to implement a new coding scheme (stochastic decoder) to improve his/her privacy. For the case of memoryless reconstruction in Fig. 7.2, it can be shown that allowing the use of a stochastic encoder does not improve the rate-distortion-equivocation region.

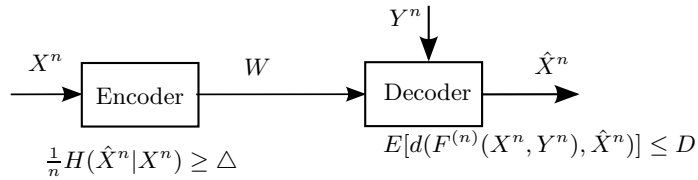

Figure 7.3: Source coding with end-user privacy at encoder.

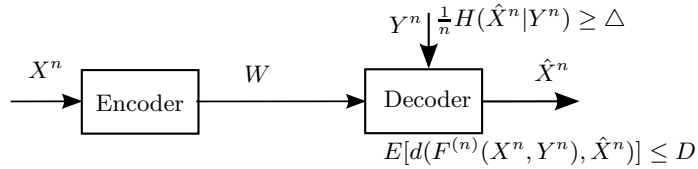

Figure 7.4: Source coding with end-user privacy at helper.

reconstruction symbol depends only on the source description and current side information symbol, the complete characterization of the rate-distortion-equivocation region is given. A binary example illustrating the potential gain from allowing the use of a stochastic decoder is also given at the end of the chapter.

We note that the case of end-user privacy at the encoder in Fig. 7.3 is included Fig. 7.2 when $Z^n = X^n$ since the encoder is a deterministic encoder. The results can therefore be obtained straightforwardly from those of the setting shown in Fig. 7.2.

- Section 7.3 considers end-user privacy at the helper as shown in Fig. 7.4. It corresponds to a scenario where we wish to prevent the helper from inferring the final reconstruction. Inner and outer bounds to the rate-distortion-equivocation region are given.

7.1.2 Related Work

The idea of protecting the reconstruction sequence against an eavesdropper was first considered as an additional secrecy constraint in the context of coding for watermarking and encryption by Merhav in [Mer06a] where the author considered a watermarking setting using a secret key sequence to protect the (watermark) message and reconstruction sequences. It was also considered in a related Shannon cipher system where the secret key is distributed through a capacity-limited channel in [Mer06b]. Recently, Schieler and Cuff in [SC13] considered a lossy source coding setting with common secret key and the objective is to maximize a payoff function based on the source, legitimate's and eavesdropper's reconstruction

sequences. Under certain assumptions, the payoff function can reduce to the equivocation of the reconstruction sequence. With the focus on *source secrecy*, it was discussed in [EU11] that the end-user privacy constraint which is the equivocation bound of the reconstruction sequence might be an inconsistent secrecy measure. However, it is still a reasonable measure from an end-user's secrecy point of view as it measures amount of the remaining uncertainty of the reconstruction sequence at the eavesdropper. Closely related to the end-user privacy, Tandon et. al in [TSP13] considered the setting of Heegard-Berger lossy source coding [HB85] where the degraded decoder has an additional privacy constraint on the side information of the stronger decoder.

7.2 End-user Privacy at Eavesdropper

7.2.1 Problem Formulation

We consider a setting in the presence of an external eavesdropper, as shown in Fig. 7.2. Source, side information, and reconstruction alphabets, $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \hat{\mathcal{X}}$ are assumed to be finite. Let (X^n, Y^n, Z^n) be n-length sequences which are i.i.d. according to $P_{X,Y,Z}$. A function $F^{(n)}(X^n, Y^n)$ is assumed to be a component-wise function, where the i^{th} component $F_i^{(n)}(X^n, Y^n) = F(X_i, Y_i)$ with $F : \mathcal{X} \times \mathcal{Y} \to \mathcal{F}$, for $i = 1, \ldots, n$ (cf., e.g., [Yam82]). Given a source sequence X^n, an encoder generates a source description $W \in \mathcal{W}^{(n)}$ and sends it over the noise-free, rate-limited link to a decoder. Given the source description and the side information Y^n, the decoder randomly generates $\hat{X}^n$ as an estimate of the value of the function $F^{(n)}(X^n, Y^n)$ such that it satisfies a distortion criterion. The eavesdropper has access to the source description and its own side information Z^n. The end-user privacy at the eavesdropper is then measured by the normalized conditional entropy $H(\hat{X}^n|W, Z^n)/n$. We are interested in characterizing the optimal tradeoff between rate, distortion, and equivocation of the reconstruction sequence in terms of the rate-distortion-equivocation region.

The model in Fig. 7.2 is similar to the secure source coding with side information in [VP13], except that the end-user privacy is imposed instead of the source privacy. The setting is also closely related to the model of side information privacy studied in [TSP13] where the authors are interested in the privacy of side information at the second decoder who is also required to decode the source subject to a distortion constraint. As for the end-user privacy, [Mer06a] considered a similar constraint in the context of coding for watermarking and encryption. The main differences to our setting are that the author considered the case where there exists a common secret key sequence independent of the message sequence at both encoder and decoder, and that the use of a stochastic decoder was not considered. From the problem formulation point of view, the end-user privacy constraint can also be considered as a complement to the *common reconstruction constraint* in lossy source coding problems [Ste09], or [LMW11] (see also Section 3.2, Chapter 3) where the

reconstruction sequence is instead required to be reproduced at another node.

Definitions of code, achievability, and the rate-distortion-equivocation region are given below.

Definition 7.1. A $(|\mathcal{W}^{(n)}|, n)$-code for source coding with end-user privacy consists of

- an encoder $f^{(n)} : \mathcal{X}^n \to \mathcal{W}^{(n)}$,
- a stochastic decoder $G^{(n)}$ which maps $w \in \mathcal{W}^{(n)}$ and $y^n \in \mathcal{Y}^n$ to $\hat{x}^n \in \hat{\mathcal{X}}^n$ according to $p(\hat{x}^n|w, y^n)$,

where $\mathcal{W}^{(n)}$ is a finite set.

Let $d : \mathcal{F} \times \hat{\mathcal{X}} \to [0, \infty)$ be the single-letter distortion measure[3]. The distortion between the value of the function of source sequence and side information and its estimate at the decoder is defined as

$$d^{(n)}(F^{(n)}(X^n, Y^n), \hat{X}^n) \triangleq \frac{1}{n} \sum_{i=1}^{n} d(F(X_i, Y_i), \hat{X}_i),$$

where $d^{(n)}(\cdot)$ is the distortion function.

Definition 7.2. A rate-distortion-equivocation tuple $(R, D, \triangle) \in \mathbb{R}_+^3$ is said to be *achievable* if for any $\delta > 0$ and all sufficiently large n there exists a $(|\mathcal{W}^{(n)}|, n)$ code such that

$$\frac{1}{n} \log \left|\mathcal{W}^{(n)}\right| \leq R + \delta,$$

$$E[d^{(n)}(F^{(n)}(X^n, Y^n), \hat{X}^n)] \leq D + \delta,$$

$$\text{and} \quad \frac{1}{n} H(\hat{X}^n|W, Z^n) \geq \triangle - \delta.$$

The *rate-distortion-equivocation region* $\mathcal{R}_{\text{eve}}$ is the set of all achievable tuples.

Definition 7.3. Let $\mathcal{R}_{\text{in}}^{(\text{eve})}$ be the set of all tuples $(R, D, \triangle) \in \mathbb{R}_+^3$ such that

$$R \geq I(X; U|Y) \tag{7.1}$$

$$D \geq E[d(F(X, Y), \hat{X})] \tag{7.2}$$

$$\triangle \leq H(\hat{X}|U, Y) + I(\hat{X}; Y|T) - I(\hat{X}; Z|T) - I(U; Z|T, Y, \hat{X}), \tag{7.3}$$

for some joint distributions of form $P_{X,Y,Z}(x, y, z)P_{U|X}(u|x)P_{T|U}(t|u)P_{\hat{X}|U,Y}(\hat{x}|u, y)$ with $|\mathcal{T}| \leq |\mathcal{X}| + 5, |\mathcal{U}| \leq (|\mathcal{X}| + 5)(|\mathcal{X}| + 4)$.

[3]Note that here $\hat{\mathcal{X}}$ does not denote an alphabet of the reconstruction of X, but of the outcome of the function $F(X, Y)$.

In addition, let $\mathcal{R}_{\text{out}}^{(\text{eve})}$ be the same set as $\mathcal{R}_{\text{in}}^{(\text{eve})}$ except that the equivocation bound is replaced by

$$\triangle \leq H(\hat{X}|U,Y) + I(V,\hat{X};Y|T) - I(V,\hat{X};Z|T), \tag{7.4}$$

for some joint distributions $P_{X,Y,Z}(x,y,z)P_{U|X}(u|x)P_{T|U}(t|u)P_{V,\hat{X}|U,Y}(v,\hat{x}|u,y)$, where $H(T|V) = H(T|U) = 0$.

7.2.2 Result

Proposition 7.2.1 (Inner and outer bounds)**.** *The rate-distortion-equivocation region $\mathcal{R}_{eve}$ for the problem in Fig. 7.2 satisfies $\mathcal{R}_{in}^{(eve)} \subseteq \mathcal{R}_{eve} \subseteq \mathcal{R}_{out}^{(eve)}$.*

Proof. The proof is given in Appendix 7.A. The achievable scheme is based on layered coding and Wyner-Ziv binning in which the former aims to provide some degree of freedom to adapt amount of information accessible to the eavesdropper by utilizing two layers of codewords T^n and U^n, and the latter is used to reduce the rate needed for transmission. In addition, we allow for a stochastic decoder where the final reconstruction sequence is generated randomly based on the selected codeword U^n and the side information Y^n. □

In the equivocation bound of $\mathcal{R}_{\text{in}}^{(\text{eve})}$, the first term corresponds to uncertainty of $\hat{X}^n$ due to the use of a stochastic decoder. The difference $I(\hat{X};Y|T) - I(\hat{X};Z|T)$ can be considered as an additional uncertainty due to the fact that the eavesdropper observes Z^n, but not Y^n which is used for generating $\hat{X}^n$. The last mutual information term is related to the leakage of the second layer codeword U^n. However, the fact that it is not clear to interpret might be an indication that the bound is not optimal. From the proof of the outer bound $\mathcal{R}_{\text{out}}^{(\text{eve})}$, random variable V is related to certain reconstruction symbols and it appears since the reconstruction symbol depends on the source description and the whole side information Y^n (see, e.g., (7.14) where we cannot simplify further the terms with $\hat{X}^n$ in the conditioning.).

Remark 7.1. We can relate our result to those of other settings where the function $F^{(n)}(X^n,Y^n) = X^n$. For example, the inner bound $\mathcal{R}_{\text{in}}^{(\text{eve})}$ can resemble the optimal result of the secure lossless source coding problem considered in [VP13]. To obtain the rate-equivocation region, we set $\hat{X} = U = X$ in $\mathcal{R}_{\text{in}}^{(\text{eve})}$.

7.2.3 Causal Side Information

Next, we consider the variant of the problem depicted in Fig. 7.2 where the side information Y^n is available only causally at the decoder. This could be relevant in delay-constrained applications as mentioned in [WE06] and references therein. We consider the following types of reconstructions.

- *Causal reconstruction*: $\hat{X}_i \sim p(\hat{x}_i|w,y^i,\hat{x}^{i-1})$ for $i = 1,\ldots,n$.

- *Memoryless reconstruction*: $\hat{X}_i \sim p(\hat{x}_i|w, y_i)$ for $i = 1, \ldots, n$.

Definition 7.4. Let $\mathcal{R}_{\text{in}}^{(\text{eve,causal})}$ be the set of all tuples $(R, D, \triangle) \in \mathbb{R}_+^3$ such that

$$R \geq I(X;U) \tag{7.5}$$
$$D \geq E[d(F(X,Y), \hat{X})] \tag{7.6}$$
$$\triangle \leq H(\hat{X}|U, Z), \tag{7.7}$$

for some joint distributions of the form $P_{X,Y,Z}(x,y,z)P_{U|X}(u|x)P_{\hat{X}|U,Y}(\hat{x}|u,y)$ with $|\mathcal{U}| \leq |\mathcal{X}| + 3$.

In addition, let $\mathcal{R}_{\text{out}}^{(\text{eve,causal})}$ be the same set as $\mathcal{R}_{\text{in}}^{(\text{eve,causal})}$ except that the equivocation bound is replaced by

$$\triangle \leq H(\hat{X}|T, Z), \tag{7.8}$$

for some joint distributions $P_{X,Y,Z}(x,y,z)P_{U|X}(u|x)P_{T|U}(t|u)P_{\hat{X}|U,Y}(\hat{x}|u,y)$ where $H(T|U) = 0$.

Causal Reconstruction

Proposition 7.2.2 (Inner and outer bounds)**.** *The rate-distortion-equivocation region $\mathcal{R}_{eve}$ for the problem in Fig. 7.2 with* causal reconstruction *satisfies the relation $\mathcal{R}_{in}^{(eve,causal)} \subseteq \mathcal{R}_{eve} \subseteq \mathcal{R}_{out}^{(eve,causal)}$.*

Proof. Since the side information is only available causally at the decoder, it cannot be used for binning to reduce the rate. The achievable scheme follows that of source coding with causal side information [WE06] with the additional use of a stochastic decoder. The proof is given in Appendix 7.B. □

The entropy term in the equivocation bound of $\mathcal{R}_{\text{in}}^{(\text{eve,causal})}$ corresponds to uncertainty of the reconstruction sequence given that the eavesdropper can decode the codeword U^n and has access to the side information Z^n.

Memoryless Reconstruction

Proposition 7.2.3 (Rate-distortion-equivocation region)**.** *The rate-distortion-equivocation region $\mathcal{R}_{eve}$ for the problem in Fig. 7.2 with* memoryless reconstruction *is given by $\mathcal{R}_{in}^{(eve,causal)}$, i.e., $\mathcal{R}_{eve} = \mathcal{R}_{in}^{(eve,causal)}$.*

Proof. The achievability proof follows the same as in the case of causal reconstruction. As for the converse proof, let $U_i \triangleq W$ which satisfies $U_i - X_i - (Y_i, Z_i)$ and

$\hat{X}_i - (U_i, Y_i) - (X_i, Z_i)$ for all $i = 1, \ldots, n$. It then follows that

$$\begin{aligned} n(R + \delta_n) &\geq H(W) \\ &\geq I(X^n; W) \\ &= \sum_{i=1}^{n} H(X_i) - H(X_i|W, X^{i-1}) \\ &\geq \sum_{i=1}^{n} I(X_i; U_i), \end{aligned}$$

$$\begin{aligned} D + \delta_n &\geq E[d^{(n)}(F^{(n)}(X^n, Y^n), \hat{X}^n)] \\ &= \frac{1}{n} \sum_{i=1}^{n} E[d(F(X_i, Y_i), \hat{X}_i)], \end{aligned}$$

and

$$\begin{aligned} n(\triangle - \delta_n) &\leq H(\hat{X}^n|W, Z^n) \\ &\leq \sum_{i=1}^{n} H(\hat{X}_i|U_i, Z_i). \end{aligned}$$

The proof ends using the standard time-sharing argument. The cardinality bound on the set $\mathcal{U}$ in $R_{\text{in}}^{(\text{eve,causal})}$ can be proved using the support lemma [CK11, Lemma 15.4] that $\mathcal{U}$ should have $|\mathcal{X}|-1$ elements to preserve P_X, plus four more for $H(X|U)$, $H(\hat{X}|U, Z)$, $E[d(F(X, Y), \hat{X}]$, and the Markov relation $\hat{X} - (U, Y) - (X, Z)$. □

Remark 7.2. For the special case where $Y = \varnothing$, the rate-distortion-equivocation region is given by $\mathcal{R}_{\text{in}}^{(\text{eve,causal})}$ with the corresponding set of distributions such that $Y = \varnothing$. We can see that if the decoder is a deterministic mapping, the achievable equivocation rate is zero since the eavesdropper observes everything the decoder does. However, for some positive D, by using the stochastic decoder, we can achieve the equivocation rate of $H(\hat{X}|U, Z)$ which can be strictly positive. This shows that there exist cases where stochastic decoder strictly enlarges the rate-distortion-equivocation region.

Remark 7.3. Proposition 7.2.3 resembles the result of the special case in [SC13, Corollary 5] where there is no shared secret key.

7.2.4 Special Case: End-user Privacy at the Encoder

Fig. 7.2 includes the setting of end-user privacy at the encoder in Fig. 7.3 as a special case by setting $Z^n = X^n$ since the source description is a deterministic function of X^n. The above results can readily reduce to the corresponding results for the problems in Fig. 7.3 as follows.

- Inner bound: The inner bound for the setting in Fig. 7.3 is obtained readily from $\mathcal{R}_{\text{in}}^{(\text{eve})}$ by setting $Z = X$ and $T = U$.
- Inner and outer bounds for *causal reconstruction* are obtained from $\mathcal{R}_{\text{in}}^{(\text{eve,causal})}$ and $\mathcal{R}_{\text{out}}^{(\text{eve,causal})}$ by setting $Z = X$.
- The rate-distortion-equivocation region for *memoryless reconstruction* is obtained from $\mathcal{R}_{\text{in}}^{(\text{eve,causal})}$ by setting $Z = X$.

7.3 End-user Privacy at Helper

In this section, we consider the setting in Fig. 7.4 where the end-user privacy constraint is imposed at the helper who provides side information Y^n to the decoder. We are interested in how the decoder should utilize the correlated side information in the reconstruction while keeping the reconstruction sequence secret/private from the helper.

7.3.1 Problem Formulation

The problem formulation and definition of the code are similar as before, except that the end-user privacy constraint is now at the helper.

Definition 7.5. A rate-distortion-equivocation tuple $(R, D, \triangle) \in \mathbb{R}_+^3$ is said to be *achievable* if for any $\delta > 0$ and all sufficiently large n there exists a $(|\mathcal{W}^{(n)}|, n)$ code such that

$$\frac{1}{n} \log \left|\mathcal{W}^{(n)}\right| \leq R + \delta,$$
$$E[d^{(n)}(F^{(n)}(X^n, Y^n), \hat{X}^n)] \leq D + \delta,$$
$$\text{and} \quad \frac{1}{n} H(\hat{X}^n | Y^n) \geq \triangle - \delta.$$

The *rate-distortion-equivocation region* $\mathcal{R}_{\text{helper}}$ is the set of all achievable tuples.

Definition 7.6. Let $\mathcal{R}_{\text{in}}^{(\text{help})}$ be the set of all tuples $(R, D, \triangle) \in \mathbb{R}_+^3$ such that

$$R \geq I(X; U|Y) \tag{7.9}$$
$$D \geq E[d(F(X, Y), \hat{X})] \tag{7.10}$$
$$\triangle \leq H(\hat{X}|U, Y) + I(X; \hat{X}|Y), \tag{7.11}$$

for some joint distributions of the form $P_{X,Y}(x, y)P_{U|X}(u|x)P_{\hat{X}|U,Y}(\hat{x}|u, y)$ with $|\mathcal{U}| \leq |\mathcal{X}| + 3$.

In addition, let $\mathcal{R}_{\text{out}}^{(\text{help})}$ be the same set as $\mathcal{R}_{\text{in}}^{(\text{help})}$ except that the equivocation bound is replaced by

$$\triangle \leq H(\hat{X}|U, Y) + I(X; V, \hat{X}|Y), \tag{7.12}$$

and the joint distributions is factorized as $P_{X,Y}(x, y)P_{U|X}(u|x)P_{V,\hat{X}|U,Y}(v, \hat{x}|u, y)$.

7.3.2 Result

Proposition 7.3.1 (Inner and outer bounds)**.** *The rate-distortion-equivocation region $\mathcal{R}_{help}$ for the problem in Fig. 7.4 satisfies $\mathcal{R}_{in}^{(help)} \subseteq \mathcal{R}_{help} \subseteq \mathcal{R}_{out}^{(help)}$.*

Proof. The proof is given in Appendix 7.C in which the achievable scheme implements Wyner-Ziv type coding with the additional use of a stochastic decoder. We note that since there is no eavesdropper in this setting, no layering is used in the achievable scheme. As for the outer bound, the presence of random variable V in $\mathcal{R}_{\text{out}}^{(\text{help})}$ can be argued similarly as in the proof of Proposition 7.2.1. □

Remark 7.4. One example showing that stochastic decoder can enlarge the rate-distortion-equivocation region is when $Y = X$ in Fig. 7.4. Since the source is available completely at the decoder, we do not need to send any description over the rate-limited link and the zero rate is achievable. In this case, we have that $\mathcal{R}_{\text{help}}$ is given by the inner bound $\mathcal{R}_{\text{in}}^{(\text{help})}$ where $X = Y$ and $U = \varnothing$. For any positive D, the stochastic decoder could randomly put out a reconstruction sequence that still satisfies the distortion level D, and achieve a positive equivocation rate as opposed to the zero equivocation in the case of a deterministic decoder.

7.4 Binary Example

In this section, we consider an example illustrating the potential gain from allowing the use of a stochastic decoder. Specifically, we consider the setting in Fig. 7.2 under memoryless reconstruction and assumptions that $Z = \emptyset$ and $F(X, Y) = X$. Then we evaluate the corresponding result in Proposition 7.2.3.

Let $\mathcal{X} = \hat{\mathcal{X}} = \{0, 1\}$ be binary source and reconstruction alphabets. We assume that the source symbol X is distributed according to Bernoulli(1/2), and side information $Y \in \{0, 1, e\}$ is an erased version of the source with an erasure probability p_e. The Hamming distortion measure is assumed, i.e., $d(x, \hat{x}) = 1$ if $x \neq \hat{x}$, and zero otherwise. Inspired by the optimal choice of U in the Wyner-Ziv result [WZ76], we let U be the output of a BSC(p_u), $p_u \in [0, 1/2]$ with input X. The reconstruction symbol generated from a stochastic decoder is chosen s.t. $\hat{X} = Y$ if $Y \neq e$, otherwise $\hat{X} \sim P_{\hat{X}|U}$, where $P_{\hat{X}|U}$ is modelled as a BSC(p_2), $p_2 \in [0, 1/2]$. With these assumptions at hand, the inner bound to the rate-distortion-equivocation region in Proposition 7.2.3 can be expressed as

$$\begin{aligned}\mathcal{R}_{\text{in,random}} = \{(R, D, \triangle) | R &\geq 1 - h(p_u) \\ D &\geq p_e(p_u \star p_2) \\ \triangle &\leq h(p_u(1 - p_e) + p_2 p_e) \\ \text{for some } p_u, p_2 &\in [0, 1/2]\},\end{aligned}$$

where $h(\cdot)$ is a binary entropy function and $a \star b \triangleq a(1 - b) + (1 - a)b$.

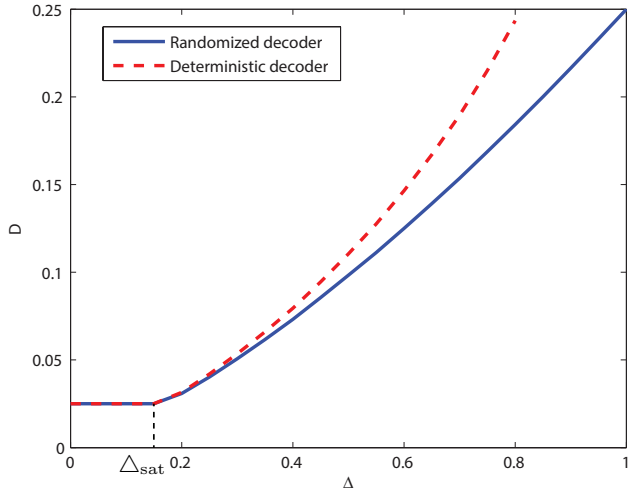

Figure 7.5: Achievable minimum distortion w.r.t. equivocation for a fixed rate $R = 0.7136$, and $p_e = 0.5$. (© 2014 IEEE)

For comparison, we also evaluate the inner bound for the case of the Wyner-Ziv optimal deterministic decoder by setting $p_2 = 0$. We plot the achievable minimum distortion as a function of equivocation rate for a fixed $R = 0.7136$, where $p_e = 0.5$. Fig. 7.5 shows the tradeoff between achievable minimum distortion and equivocation rate for a fixed rate R. We can see that in general the minimum distortion is sacrificed for a higher equivocation. For the same particular structure of $P_{U|X}$ and the given deterministic decoder in this setting, it shows that, for a given rate R and distortion D, a higher equivocation rate $\triangle$ can be achieved by using a stochastic decoder.[4] As for the low equivocation region, we observe a saturation of distortion because the minimum distortion is limited by the rate. The value $\triangle_{\text{sat}}$ at which the minimum distortion cannot be lowered by decreasing $\triangle$ can be specified as $\triangle_{\text{sat}} = h((1-p_e)h^{-1}(1-R))$, and the corresponding $D_{\min}(R, \triangle_{\text{sat}}) = p_e h^{-1}(1-R)$ is the minimum distortion according to the Wyner-Ziv rate-distortion function. It could also be interesting to see how good the inner bounds are by evaluating the outer bound result. However, it involves the optimization over an auxiliary random variable which is not straightforward and will be left for future work.

Special case: In the special case where $Y = \emptyset$, the gain can be shown as follows (cf. Remark 7.2). If the decoder is a deterministic mapping, the achievable equivocation rate is always zero since the eavesdropper is as strong as the decoder. The corresponding distortion-rate function for this example is given by

[4]Here we only evaluate and compare inner bounds on the rate-distortion-equivocation regions to illustrate a potential gain of allowing the use of a stochastic decoder.

$D \geq h^{-1}(1-R)$ [EK11, Chapter 3]. However, by using the stochastic decoder as above, we can achieve $D \geq h^{-1}(1-R) \star h^{-1}(\triangle)$ (by letting $p_e = 1$ in $\mathcal{R}_{\text{in,random}}$). For $D = h^{-1}(1-R) \star c$, where $c \in (0, 1/2]$, we can achieve strictly positive equivocation rate $h(c)$.

7.5 Conclusion

In this chapter, we introduced a new privacy metric (end-user privacy constraint) in problems of lossy source coding with side information. We considered several problems where the end-user privacy constraint is imposed at different nodes, namely the eavesdropper, the encoder, and the helper. Since the goal of end-user privacy is to protect the reconstruction sequence generated at the decoder against any unwanted inferences, we allow the decoder mapping to be a random mapping, and it was shown by example that there exist cases where a stochastic decoder strictly enlarges the rate-distortion-equivocation region as compared to the one derived for deterministic decoders. In general, characterizing the complete rate-distortion-equivocation region for the setting with end-user privacy is difficult since conditioned on the source description, the reconstruction process is not necessarily memoryless. As seen in a special case of end-user privacy at the eavesdropper, when we restrict the reconstruction symbol to depend only on the source description and the current side information symbol, the complete rate-distortion-equivocation region can be given.

Appendix for Chapter 7

7.A Proof of Proposition 7.2.1

The inner bound proofs for the rate and distortion constraints follow from the coding scheme which utilizes layered coding and Wyner-Ziv binning. That is, we have two layers of codewords T^n and U^n forming the codebook, and after encoding, only the bin indices of the chosen codewords are transmitted to the decoder. Also, instead of using the deterministic function at the decoder, we allow stochastic decoder to generate the reconstruction sequence, i.e., the decoder puts out $\hat{X}^n$, where $\hat{X}_i \sim P_{\hat{X}|U,Y}$ for each $i = 1, \ldots, n$. The outline of the proof is given below.

Fix $P_{U|X}$, $P_{T|U}$, and $P_{\hat{X}|U,Y}$. Randomly and independently generate $2^{n(I(X;T)+\delta_\epsilon)}$ $t^n(j)$ sequences, each i.i.d. according to $\prod_{i=1}^n P_T(t_i)$, $j \in [1 : 2^{n(I(X;T)+\delta_\epsilon)}]$. Then distribute them uniformly at random into $2^{n(I(X;T|Y)+2\delta_\epsilon)}$ equal-sized bins $b_T(w_1)$, $w_1 \in [1 : 2^{nI(X;T|Y)+2\delta_\epsilon}]$. For each j, randomly and conditionally independently generate $2^{n(I(X;U|T)+\delta_\epsilon)}$ $u^n(j,k)$ sequences, each i.i.d. according to $\prod_{i=1}^n P_{U|T}(u_i|t_i)$, $k \in [1 : 2^{n(I(X;U|T)+\delta_\epsilon)}]$, and distribute these sequences uniformly at random into $2^{n(I(X;U|T,Y)+2\delta_\epsilon)}$ equal-sized bins $b_U(j,w_2)$, $w_2 \in [1 : 2^{nI(X;U|T,Y)+2\delta_\epsilon}]$. For encoding, the encoder looks for $t^n(j)$ and $u^n(j,k)$ jointly typical with x^n. With high probability, it will find such codewords and then send the corresponding bin indices w_1 and w_2 to the decoder. The total rate is thus equal to $I(X;T|Y)+I(X;U|T,Y)+4\delta_\epsilon = I(X;U|Y) + 4\delta_\epsilon$. Based on the received bin indices, the decoder, with high probability, will find the unique sequences $t^n(j) \in b_T(w_1)$ and $u^n(j,k) \in b_U(j,w_2)$ such that they are jointly typical with y^n. Then it puts out $\hat{x}^n$ where $\hat{x}_i$ is randomly generated according to $P_{\hat{X}|U,Y}(\hat{x}_i|u_i,y_i)$, $i = 1, \ldots, n$.

Let $T^n(J)$ and $U^n(J,K)$ be the codewords chosen at the encoder, and W_1 and W_2 be the corresponding bin indices of the bins which $T^n(J)$ and $U^n(J,K)$ belong to. Then W_1 and W_2 are functions of J and K. Since we have that the tuple $(X^n, T^n(J), U^n(J,K), Y^n, \hat{X}^n) \in \mathcal{T}_\epsilon^{(n)}(X,T,U,Y,\hat{X})$ with high probability, it can be shown that the distortion constraint is satisfied if $E[d(F(X,Y),\hat{X})] \le D$.

Next, we give a sketch of the proof for the equivocation constraint. Let $\mathcal{C}_n$ be a random variable representing the randomly chosen codebook. By Fano's inequality, we have that for any $\mathcal{C}_n = \mathfrak{C}_n$, $H(J,K|W_1,W_2,Y^n,\mathcal{C}_n = \mathfrak{C}_n) \le 1 + \Pr(\mathcal{E})\log(|\mathcal{J}||\mathcal{K}|)$, where $\Pr(\mathcal{E})$ is the probability that (J,K) cannot be determined from (W_1,W_2,Y^n). From the decoding process, we have that $\Pr(\mathcal{E}) \to 0$ as $n \to \infty$. Then, it follows that

$$\begin{aligned}
\frac{1}{n}H(J,K|W_1,W_2,Y^n,\mathcal{C}_n) &= \sum_{\mathfrak{C}_n} p(\mathfrak{C}_n)\frac{1}{n}H(J,K|W_1,W_2,Y^n,\mathcal{C}_n = \mathfrak{C}_n) \\
&\le \sum_{\mathfrak{C}_n} p(\mathfrak{C}_n)\left(\frac{1}{n} + \Pr(\mathcal{E})\frac{1}{n}\log(|\mathcal{J}||\mathcal{K}|)\right) \\
&\le \epsilon_n, \qquad (7.13)
\end{aligned}$$

where $\epsilon_n \to 0$ as $n \to \infty$.

The equivocation averaged over all possible codebooks can be bounded as follows.

$$
\begin{aligned}
&H(\hat{X}^n|W_1, W_2, Z^n, \mathcal{C}_n) \\
&= H(\hat{X}^n|J, K, Y^n, Z^n, \mathcal{C}_n) + I(\hat{X}^n; J, K, Y^n|W_1, W_2, Z^n, \mathcal{C}_n) \\
&\overset{(a)}{\geq} H(\hat{X}^n|U^n, Y^n, \mathcal{C}_n) + I(\hat{X}^n; Y^n|W_1, W_2, Z^n, \mathcal{C}_n) \\
&\quad + H(J, K|W_1, W_2, Y^n, Z^n, \mathcal{C}_n) - n\epsilon_n \\
&= H(\hat{X}^n|U^n, Y^n, \mathcal{C}_n) + H(Y^n, Z^n) + H(W_1, W_2|Y^n, Z^n, \mathcal{C}_n) - H(W_1, W_2, Z^n|\mathcal{C}_n) \\
&\quad - H(Y^n|W_1, W_2, Z^n, \hat{X}^n, \mathcal{C}_n) + H(J, K|W_1, W_1, Y^n, Z^n, \mathcal{C}_n) - n\epsilon_n \\
&= H(\hat{X}^n|U^n, Y^n, \mathcal{C}_n) + H(Y^n, Z^n) + H(J, K|Y^n, Z^n, \mathcal{C}_n) - H(W_1, W_2, Z^n|\mathcal{C}_n) \\
&\quad - H(Y^n|W_1, W_2, Z^n, \hat{X}^n, \mathcal{C}_n) - n\epsilon_n \\
&\overset{(b)}{\geq} H(\hat{X}^n|U^n, Y^n) + H(Y^n, Z^n) + I(J, K; X^n|Y^n, Z^n, \mathcal{C}_n) - H(W_1|\mathcal{C}_n) \\
&\quad - H(W_2|\mathcal{C}_n) - H(Z^n|W_1, \mathcal{C}_n) - H(Y^n|W_1, Z^n, \hat{X}^n, \mathcal{C}_n) - n\epsilon_n \\
&= H(\hat{X}^n|U^n, Y^n) + H(Y^n, Z^n) + I(J, K; X^n|Y^n, Z^n, \mathcal{C}_n) - H(W_1|\mathcal{C}_n) \\
&\quad - H(W_2|\mathcal{C}_n) - H(Z^n|J, \mathcal{C}_n) - I(Z^n; J|W_1, \mathcal{C}_n) - H(Y^n|J, Z^n, \hat{X}^n, \mathcal{C}_n) \\
&\quad - I(Y^n; J|W_1, Z^n, \hat{X}^n, \mathcal{C}_n) - n\epsilon_n \\
&\geq H(\hat{X}^n|U^n, Y^n) + H(Y^n, Z^n) + I(J, K; X^n|Y^n, Z^n, \mathcal{C}_n) - H(J|\mathcal{C}_n) \\
&\quad - H(W_2|\mathcal{C}_n) - H(Z^n|J, \mathcal{C}_n) - H(Y^n|J, Z^n, \hat{X}^n, \mathcal{C}_n) - n\epsilon_n \\
&\overset{(c)}{\geq} n[H(\hat{X}|U, Y) + H(Y, Z) + I(X; T, U|Y, Z) - I(X; T) - I(X; U|T, Y) \\
&\quad - H(Z|T) - H(Y|T, Z, \hat{X}) - \delta'_\epsilon - \epsilon_n] \\
&\overset{(d)}{=} n[H(\hat{X}|U, Y) + I(\hat{X}; Y|T) - I(\hat{X}; Z|T) - I(U; Z|T, Y, \hat{X}) - \delta''_\epsilon] \\
&\geq n[\triangle - \delta''_\epsilon]
\end{aligned}
$$

if $\triangle \leq H(\hat{X}|U, Y) + I(\hat{X}; Y|T) - I(\hat{X}; Z|T) - I(U; Z|T, Y, \hat{X})$, where (a) follows from, conditioned on the codebook, we have the Markov chain $\hat{X}^n - (U^n(J, K), Y^n) - (J, K, Z^n)$, and from Fano's inequality in (7.13), (b) follows the from the Markov chain $\hat{X}^n - (U^n(J, K), Y^n) - \mathcal{C}_n$, from (J, K) is a function of X^n, and that conditioning reduces entropy, (c) follows from the codebook generation and from bounding the term $H(\hat{X}^n|U^n, Y^n)$ where $\hat{X}^n \sim \prod_{i=1}^n P_{\hat{X}|U,Y}$ similarly as in Lemma 5.3, and the terms $H(X^n|J, K, Y^n, Z^n, \mathcal{C}_n)$, $H(Z^n|J, \mathcal{C}_n)$, and $H(Y^n|J, Z^n, \hat{X}^n, \mathcal{C}_n)$ for which the proofs follow similarly as that of Lemma 5.4, and (d) from the Markov chains $T - U - X - (Y, Z)$ and $\hat{X} - (U, Y) - (X, Z, T)$.

The cardinality bounds on the sets $\mathcal{T}$ and $\mathcal{U}$ in $\mathcal{R}_{\text{in}}^{(\text{eve})}$ can be proved using the support lemma [CK11, Lemma 15.4] and is given in Appendix 7.D.

As for the outer bound, let $T_i \triangleq (W, Z^{i-1}, Y_{i+1}^n)$, $U_i \triangleq (W, Z^{i-1}, Y^{n\setminus i})$ and $V_i \triangleq (W, Z^{i-1}, Y_{i+1}^n, \hat{X}^{n\setminus i})$ which satisfy $T_i - U_i - X_i - (Y_i, Z_i)$ and $(V_i, \hat{X}_i) - (U_i, Y_i) - (X_i, Z_i, T_i)$ for all $i = 1, \ldots, n$. The outer bound proof for the rate and distortion constraints follows similarly as that of the Wyner-Ziv problem with the exception of the part related to stochastic decoder.

That is, we have

$$\begin{aligned} n(R+\delta_n) &\geq H(W) \\ &\geq I(X^n, Z^n; W|Y^n) \\ &= \sum_{i=1}^n H(X_i, Z_i|Y_i) - H(X_i, Z_i|W, X^{i-1}, Z^{i-1}, Y^n) \\ &\overset{(a)}{\geq} \sum_{i=1}^n H(X_i, Z_i|Y_i) - H(X_i, Z_i|U_i, Y_i) \\ &\geq \sum_{i=1}^n I(X_i; U_i|Y_i), \end{aligned}$$

where (a) follows from the definition of U_i and that conditioning reduces entropy, and

$$\begin{aligned} D + \delta_n &\geq E[d^{(n)}(F^{(n)}(X^n, Y^n), \hat{X}^n)] \\ &= \frac{1}{n}\sum_{i=1}^n E[d(F(X_i, Y_i), \hat{X}_i)]. \end{aligned}$$

The equivocation bound follows below.

$$\begin{aligned} n(\triangle - \delta_n) &\leq H(\hat{X}^n|W, Z^n) \\ &= H(\hat{X}^n|W) - I(\hat{X}^n; Z^n|W) \\ &= H(\hat{X}^n|W, Y^n) + I(\hat{X}^n; Y^n|W) - I(\hat{X}^n; Z^n|W) \\ &\leq \sum_{i=1}^n H(\hat{X}_i|W, Y^n) + H(Y_i|W, Y_{i+1}^n) - H(Y_i|W, Y_{i+1}^n, \hat{X}^n) \\ &\quad - H(Z_i|W, Z^{i-1}) + H(Z_i|W, Z^{i-1}, \hat{X}^n) \qquad (7.14) \\ &\overset{(a)}{\leq} \sum_{i=1}^n H(\hat{X}_i|W, Y^n, Z^{i-1}) - I(Y_i; W, Y_{i+1}^n) + I(Y_i; \hat{X}_i) \\ &\quad + I(Y_i; W, Y_{i+1}^n, \hat{X}^{n\setminus i}|\hat{X}_i) + I(Z_i; W, Z^{i-1}) - I(Z_i; \hat{X}_i) \\ &\quad - I(Z_i; W, Z^{i-1}, \hat{X}^{n\setminus i}|\hat{X}_i) \\ &\overset{(b)}{=} \sum_{i=1}^n H(\hat{X}_i|W, Y^n, Z^{i-1}) - I(Y_i; W, Z^{i-1}, Y_{i+1}^n) + I(Y_i; \hat{X}_i) \end{aligned}$$

$$
\begin{aligned}
&+ I(Y_i; W, Z^{i-1}, Y_{i+1}^n, \hat{X}^{n\setminus i}|\hat{X}_i) + I(Z_i; W, Z^{i-1}, Y_{i+1}^n) - I(Z_i; \hat{X}_i) \\
&- I(Z_i; W, Z^{i-1}, Y_{i+1}^n, \hat{X}^{n\setminus i}|\hat{X}_i) \\
&\overset{(c)}{=} \sum_{i=1}^n H(\hat{X}_i|U_i, Y_i) - I(Y_i; T_i) + I(Y_i; \hat{X}_i) + I(Y_i; T_i, V_i|\hat{X}_i) \\
&+ I(Z_i; T_i) - I(Z_i; \hat{X}_i) - I(Z_i; T_i, V_i|\hat{X}_i) \\
&= \sum_{i=1}^n H(\hat{X}_i|U_i, Y_i) + I(Y_i; V_i, \hat{X}_i|T_i) - I(Z_i; V_i, \hat{X}_i|T_i),
\end{aligned}
$$

where (a) follows from the Markov chain $\hat{X}_i - (W, Y^n) - Z^{i-1}$, (b) follows from the Csiszár's sum identity, $\sum_{i=1}^n I(Y_i; Z^{i-1}|W, Y_{i+1}^n) - I(Z_i; Y_{i+1}^n|W, Z^{i-1}) = 0$ and $\sum_{i=1}^n I(Y_i; Z^{i-1}|W, \hat{X}^n, Y_{i+1}^n) - I(Z_i; Y_{i+1}^n|W, \hat{X}^n, Z^{i-1}) = 0$, and (c) follows from the definitions of T_i, U_i and V_i.

Note that from the definitions of T_i, U_i, and V_i, we have that T_i is a function of U_i or V_i. So we can further restrict the set of joint distributions to satisfy $H(T_i|U_i) = H(T_i|V_i) = 0$. The proof ends using the standard time-sharing argument.

7.B Proof of Proposition 7.2.2

The inner bound proof for the rate and distortion constraints follows that of source coding with causal side information [WE06] with the additional use of a stochastic decoder. Since the side information is only available causally at the decoder, it cannot be used for binning to reduce the rate. Here, we just use the rate-distortion code with codewords U^n. The decoder then generates $\hat{X}^n$, where $\hat{X}_i \sim P_{\hat{X}|U,Y}$ for $i = 1, \ldots, n$. The proof of equivocation constraint is given below. It is different from the noncausal case in that the scheme does not utilize binning. Here W denotes the index of codeword U^n. The equivocation averaged over all possible codebooks can be bounded as follows.

$$
\begin{aligned}
&H(\hat{X}^n|W, Z^n, \mathcal{C}_n) = H(\hat{X}^n|W, Z^n, Y^n, \mathcal{C}_n) + I(\hat{X}^n; Y^n|W, Z^n, \mathcal{C}_n) \\
&= H(\hat{X}^n|W, Y^n, Z^n, \mathcal{C}_n) + H(Y^n|W, Z^n, \mathcal{C}_n) - H(Y^n|W, Z^n, \hat{X}^n, \mathcal{C}_n) \\
&\overset{(a)}{=} H(\hat{X}^n|U^n(W), Y^n, \mathcal{C}_n) + H(Y^n, Z^n) + H(W|Y^n, Z^n, \mathcal{C}_n) - H(W|\mathcal{C}_n) \\
&\quad - H(Z^n|W, \mathcal{C}_n) - H(Y^n|W, Z^n, \hat{X}^n, \mathcal{C}_n) \\
&\overset{(b)}{\geq} n[\underbrace{H(\hat{X}|U, Y) + H(Y, Z) - I(X; U)}_{\triangleq P} - \delta'_\epsilon] + H(W|Y^n, Z^n, \mathcal{C}_n) - H(Z^n|W, \mathcal{C}_n) \\
&\quad - H(Y^n|W, Z^n, \hat{X}^n, \mathcal{C}_n) \\
&\overset{(c)}{=} n[P - \delta'_\epsilon] + I(W; X^n|Y^n, Z^n, \mathcal{C}_n) - H(Z^n|W, \mathcal{C}_n) - H(Y^n|W, Z^n, \hat{X}^n, \mathcal{C}_n) \\
&= n[P - \delta'_\epsilon] + H(X^n|Y^n, Z^n) - H(X^n|W, Y^n, Z^n, \mathcal{C}_n) - H(Z^n|W, \mathcal{C}_n)
\end{aligned}
$$

$$
\begin{aligned}
&\quad - H(Y^n|W, Z^n, \hat{X}^n, \mathcal{C}_n)] \\
&\overset{(d)}{\geq} n[H(\hat{X}|U, Y) + H(Y, Z) - I(X; U) + H(X|Y, Z) - H(X|U, Y, Z) \\
&\quad - H(Z|U) - H(Y|U, Z, \hat{X}) - \delta''_\epsilon] \\
&\overset{(e)}{=} n[H(\hat{X}|U, Z) - \delta''_\epsilon] \\
&\geq n[\triangle - \delta''_\epsilon]
\end{aligned}
$$

if $\triangle \leq H(\hat{X}|U, Z)$, where (a) follows from, conditioned on the codebook, we have the Markov chain $\hat{X}^n - (U^n, Y^n) - (W, Z^n)$, (b) follows from the Markov chain $\hat{X}^n - (U^n(J, K), Y^n) - \mathcal{C}_n$, from bounding the term $H(\hat{X}^n|U^n, Y^n)$ where $\hat{X}^n \sim \prod_{i=1}^n P_{\hat{X}|U,Y}$ similarly as in Lemma 5.3 and from the codebook generation, (c) follows since W is a function of X^n, (d) follows from bounding the terms $H(X^n|W, Y^n, Z^n, \mathcal{C}_n)$, $H(Z^n|W, \mathcal{C}_n)$, and $H(Y^n|W, Z^n, \hat{X}^n, \mathcal{C}_n)$ similarly as that of Lemma 5.4, and (e) follows from the Markov chains $U - X - (Y, Z)$ and $\hat{X} - (U, Y) - (X, Z)$.

The cardinality bound on the set $\mathcal{U}$ in $\mathcal{R}_{\text{in}}^{(\text{eve,causal})}$ can be proved using the support lemma that $\mathcal{U}$ should have $|\mathcal{X}| - 1$ elements to preserve P_X, plus four more for $H(X|U)$, $H(\hat{X}|U, Z)$, $E[d(F(X, Y), \hat{X}]$, and the Markov relation $\hat{X} - (U, Y) - (X, Z)$.

For the outer bound proof, let $U_i \triangleq (W, Y^{i-1}, \hat{X}^{i-1})$, $T_i \triangleq (W, \hat{X}^{i-1})$ which satisfy $T_i - U_i - X_i - (Y_i, Z_i)$ and $\hat{X}_i - (U_i, Y_i) - (X_i, Z_i, T_i)$ for all $i = 1, \ldots, n$. It then follows that

$$
\begin{aligned}
n(R + \delta_n) &\geq H(W) \\
&\geq I(X^n; W) \\
&= \sum_{i=1}^n H(X_i) - H(X_i|W, X^{i-1}) \\
&\overset{(a)}{=} \sum_{i=1}^n H(X_i) - H(X_i|W, X^{i-1}, Y^{i-1}, \hat{X}^{i-1}) \\
&\geq \sum_{i=1}^n I(X_i; U_i),
\end{aligned}
$$

where (a) follows from the Markov chain $X_i - (W, X^{i-1}) - (Y^{i-1}, \hat{X}^{i-1})$. And

$$
\begin{aligned}
D + \delta_n &\geq E[d^{(n)}(F^{(n)}(X^n, Y^n), \hat{X}^n)] \\
&= \frac{1}{n} \sum_{i=1}^n E[d(F(X_i, Y_i), \hat{X}_i)].
\end{aligned}
$$

And

$$
n(\triangle - \delta_n) \leq H(\hat{X}^n|W, Z^n)
$$

$$= \sum_{i=1}^{n} H(\hat{X}_i|T_i, Z_i).$$

Note that from the definitions of T_i and U_i, we have that T_i is a function of U_i. So we can further restrict the set of joint distributions to satisfy $H(T_i|U_i) = 0$. The proof ends using the standard time-sharing argument.

7.C Proof of Proposition 7.3.1

The inner bound proof for the rate and distortion constraints follows from the scheme that implements Wyner-Ziv type coding with the additional use of a stochastic decoder. We only give a sketch of the proof of equivocation constraint here. The equivocation averaged over all codebooks can be bounded as follows.

$$\begin{aligned}
H(\hat{X}^n|Y^n, \mathcal{C}_n) &= H(\hat{X}^n|X^n, Y^n, \mathcal{C}_n) + I(\hat{X}^n; X^n, |Y^n, \mathcal{C}_n) \\
&\overset{(a)}{=} H(\hat{X}^n|X^n, U^n, Y^n, \mathcal{C}_n) + I(\hat{X}^n; X^n, |Y^n, \mathcal{C}_n) \\
&\overset{(b)}{\geq} n[H(\hat{X}|U, Y) + H(X|Y) - H(X|Y, \hat{X}) - \delta'_\epsilon] \\
&= n[H(\hat{X}|U, Y) + I(X; \hat{X}|Y) - \delta'_\epsilon] \\
&\geq n[\triangle - \delta'_\epsilon]
\end{aligned}$$

if $\triangle \leq H(\hat{X}|U, Y) + I(X; \hat{X}|Y)$, where (a) follows since U^n is a function of X^n, and (b) follows from the Markov chain $\hat{X}^n - (U^n, Y^n) - (X^n, \mathcal{C}_n)$, and from bounding the terms $H(X^n|Y^n, \hat{X}^n, \mathcal{C}_n)$ and $H(\hat{X}^n|U^n, Y^n)$ similarly as in Lemmas 5.4 and 5.3, respectively.

The cardinality bound on the set $\mathcal{U}$ in $\mathcal{R}_{\text{in}}^{(\text{help})}$ can be proved using the support lemma that $\mathcal{U}$ should have $|\mathcal{X}| - 1$ elements to preserve P_X, plus four more for $H(X|U, Y)$, $H(\hat{X}|U, Y)$, $E[d(F(X, Y), \hat{X}]$, and the Markov relation $\hat{X} - (U, Y) - X$.

The outer bound proof for equivocation constraint is as follows. Let $U_i \triangleq (W, X^{i-1}, Y^{n\setminus i})$ and $V_i \triangleq (X^{i-1}, Y^{n\setminus i}, \hat{X}^{n\setminus i})$ which satisfy $U_i - X_i - Y_i$ and $(V_i, \hat{X}_i) - (U_i, Y_i) - X_i$ for all $i = 1, \ldots, n$. It follows that

$$\begin{aligned}
n(\triangle - \delta_n) &\leq H(\hat{X}^n|Y^n) \\
&= H(\hat{X}^n|X^n, Y^n) + I(\hat{X}^n; X^n, |Y^n) \\
&\overset{(a)}{=} H(\hat{X}^n|X^n, W, Y^n) + I(\hat{X}^n; X^n, |Y^n) \\
&= \sum_{i=1}^{n} H(\hat{X}_i|W, \hat{X}^{i-1}, X^n, Y^n) + H(X_i|Y_i) - H(X_i|X^{i-1}, Y^n, \hat{X}^n) \\
&\overset{(b)}{\leq} \sum_{i=1}^{n} H(\hat{X}_i|U_i, Y_i) + H(X_i|Y_i) - H(X_i|V_i, \hat{X}_i, Y_i)
\end{aligned}$$

$$= \sum_{i=1}^{n} H(\hat{X}_i|U_i, Y_i) + I(X_i; V_i, \hat{X}_i|Y_i),$$

where (a) follows from the deterministic encoder, (b) follows from the definition of U_i, V_i. The proof ends using the standard time-sharing argument.

7.D Cardinality Bounds of the Sets $\mathcal{T}$ and $\mathcal{U}$ in Proposition 7.2.1

Consider the expression of $\mathcal{R}_{\text{in}}^{(\text{eve})}$ in Proposition 7.2.1:

$$\begin{aligned}
&R \geq I(X;U|Y)\\
&D \geq E[d(F(X,Y),\hat{X})]\\
&\triangle \leq H(\hat{X}|U,Y) + I(\hat{X};Y|T) - I(\hat{X};Z|T) - I(U;Z|T,Y,\hat{X}),
\end{aligned}$$

for some $U \in \mathcal{U}$, $T \in \mathcal{T}$ such that $T - U - X - (Y,Z)$ and $\hat{X} - (U,Y) - (X,Z,T)$ form Markov chains.

We can rewrite some mutual information terms in the rate and equivocation expressions above and get

$$\begin{aligned}
&R \geq H(X|Y) - H(X,Y|U) + H(Y|U),\\
&\triangle \leq H(\hat{X},Y|U) - H(Y|U) + I(Y;\hat{X},Z|T) - H(Z|T) + H(Z,Y|U) - H(Y|U).
\end{aligned}$$

We will prove that the random variables T and U may be replaced by new ones, satisfying $|\mathcal{T}| \leq |\mathcal{X}|+5$, $|\mathcal{U}| \leq (|\mathcal{X}|+5)(|\mathcal{X}|+4)$, and preserving the terms $H(X,Y|U) - H(Y|U)$, $H(\hat{X},Y|U) - H(Y|U)$, $H(Z,Y|U) - H(Y|U)$, $I(Y;\hat{X},Z|T) - H(Z|T)$, $E[d(F(X,Y),\hat{X})]$, and the Markov relations.

First we bound the cardinality of the set $\mathcal{T}$. Let us define the following $|\mathcal{X}|+5$ continuous functions of $p(u|t)$, $u \in \mathcal{U}$,

$$\begin{aligned}
f_j(p(u|t)) &= \sum_{u\in\mathcal{U}} p(u|t)p(x|u,t),\ j = 1,\ldots,|\mathcal{X}|-1,\\
f_{|\mathcal{X}|}(p(u|t)) &= H(X,Y|U,T=t) - H(Y|U,T=t)\\
&= H(X,U,Y|T=t) - H(U,Y|T=t),\\
f_{|\mathcal{X}|+1}(p(u|t)) &= H(\hat{X},Y|U,T=t) - H(Y|U,T=t)\\
&= H(\hat{X},U,Y|T=t) - H(U,Y|T=t),\\
f_{|\mathcal{X}|+2}(p(u|t)) &= H(Z,Y|U,T=t) - H(Y|U,T=t)\\
&= H(Z,U,Y|T=t) - H(U,Y|T=t),\\
f_{|\mathcal{X}|+3}(p(u|t)) &= I(Y;\hat{X},Z|T=t) - H(Z|T=t),\\
f_{|\mathcal{X}|+4}(p(u|t)) &= H(\hat{X}|U,Y,X,Z,T=t)
\end{aligned}$$

$$= H(\hat{X}, U, Y, X, Z|T = t) - H(U, Y, X, Z|T = t),$$
$$f_{|\mathcal{X}|+5}(p(u|t)) = E[d(F(X, Y), \hat{X})|T = t].$$

The corresponding averages are

$$\sum_{t\in\mathcal{T}} p(t) f_j(p(u|t)) = P_X(x),\ j = 1, \ldots, |\mathcal{X}| - 1,$$
$$\sum_{t\in\mathcal{T}} p(t) f_{|\mathcal{X}|}(p(u|t)) = H(X, U, Y|T) - H(U, Y|T),$$
$$\sum_{t\in\mathcal{T}} p(t) f_{|\mathcal{X}|+1}(p(u|t)) = H(\hat{X}, U, Y|T) - H(U, Y|T),$$
$$\sum_{t\in\mathcal{T}} p(t) f_{|\mathcal{X}|+2}(p(u|t)) = H(Z, U, Y|T) - H(U, Y|T),$$
$$\sum_{t\in\mathcal{T}} p(t) f_{|\mathcal{X}|+3}(p(u|t)) = I(Y; \hat{X}, Z|T) - H(Z|T),$$
$$\sum_{t\in\mathcal{T}} p(t) f_{|\mathcal{X}|+4}(p(u|t)) = H(\hat{X}, U, Y, X, Z|T) - H(U, Y, X, Z|T),$$
$$\sum_{t\in\mathcal{T}} p(t) f_{|\mathcal{X}|+5}(p(u|t)) = E[d(F(X, Y), \hat{X})].$$

According to the support lemma [CK11, Lemma 15.4], we can deduce that there exist a new random variable T' jointly distributed with $(X, Y, Z, U, \hat{X})$ whose alphabet size is $|\mathcal{T}'| = |\mathcal{X}| + 5$, and numbers $\alpha_i \geq 0$ with $\sum_{i=1}^{|\mathcal{X}|+5} \alpha_i = 1$ that satisfy

$$\sum_{i=1}^{|\mathcal{X}|+5} \alpha_i f_j(p(u|t)) = P_X(x),\ j = 1, \ldots, |\mathcal{X}| - 1,$$
$$\sum_{i=1}^{|\mathcal{X}|+5} \alpha_i f_{|\mathcal{X}|}(p(u|t)) = H(X, U, Y|T') - H(U, Y|T'),$$
$$\sum_{i=1}^{|\mathcal{X}|+5} \alpha_i f_{|\mathcal{X}|+1}(p(u|t)) = H(\hat{X}, U, Y|T') - H(U, Y|T'),$$
$$\sum_{i=1}^{|\mathcal{X}|+5} \alpha_i f_{|\mathcal{X}|+2}(p(u|t)) = H(Z, U, Y|T') - H(U, Y|T'),$$
$$\sum_{i=1}^{|\mathcal{X}|+5} \alpha_i f_{|\mathcal{X}|+3}(p(u|t)) = I(Y; \hat{X}, Z|T') - H(Z|T'),$$
$$\sum_{i=1}^{|\mathcal{X}|+5} \alpha_i f_{|\mathcal{X}|+4}(p(u|t)) = H(\hat{X}, U, Y, X, Z|T') - H(U, Y, X, Z|T'),$$

$$\sum_{i=1}^{|\mathcal{X}|+5} \alpha_i f_{|\mathcal{X}|+5}(p(u|t)) = E[d(F(X,Y),\hat{X})].$$

Note that

$$\begin{aligned} H(X,U,Y|T') - H(U,Y|T') &= H(X,U,Y|T) - H(U,Y|T) \\ &\overset{(\star)}{=} H(X,Y|U) - H(Y|U), \end{aligned}$$

where $(\star)$ follows from the Markov chain $T - U - X - (Y,Z)$. Similarly, from the Markov chains $T - U - X - (Y,Z)$ and $\hat{X} - (U,Y) - (X,Z,T)$, we have that $H(\hat{X},U,Y|T') - H(U,Y|T') = H(\hat{X},Y|U) - H(Y|U)$ and $H(Z,U,Y|T') - H(U,Y|T') = H(Z,Y|U) - H(Y|U)$. Since $P_X(x)$ is preserved, $P_{X,Y}(x,y)$ is also preserved. Thus, $H(X|Y)$ is preserved.

Next we bound the cardinality of the set $\mathcal{U}$. For each $t' \in \mathcal{T}'$, we define the following $|\mathcal{X}|+4$ continuous functions of $p(x|t',u)$, $x \in \mathcal{X}$,

$$\begin{aligned} &f_j(p(x|t',u)) = p(x|t',u),\ j = 1,\ldots,|\mathcal{X}|-1, \\ &f_{|\mathcal{X}|}(p(x|t',u)) = H(X,Y|T'=t',U=u) - H(Y|T'=t',U=u), \\ &f_{|\mathcal{X}|+1}(p(x|t',u)) = H(\hat{X},Y|T'=t',U=u) - H(Y|T'=t',U=u), \\ &f_{|\mathcal{X}|+2}(p(x|t',u)) = H(Z,Y|T'=t',U=u) - H(Y|T'=t',U=u), \\ &f_{|\mathcal{X}|+3}(p(x|t',u)) = H(\hat{X},Y,X,Z|T'=t',U=u) - H(Y,X,Z|T'=t',U=u), \\ &f_{|\mathcal{X}|+4}(p(x|t',u)) = E[d(F(X,Y),\hat{X})|T'=t',U=u]. \end{aligned}$$

Similarly to the previous part in bounding $|\mathcal{T}|$, there exists a new random variable $U'|\{T'=t'\} \sim p(u'|t')$ such that $|\mathcal{U}'| = |\mathcal{X}|+4$ and $p(x|t')$, $H(X,Y|T'=t',U) - H(Y|T'=t',U)$, $H(\hat{X},Y|T'=t',U) - H(Y|T'=t',U)$, $H(Z,Y|T'=t',U) - H(Y|T'=t',U)$, $H(\hat{X},Y,X,Z|T'=t',U) - H(Y,X,Z|T'=t',U)$, and $E[d(F(X,Y),\hat{X})|T'=t']$ are preserved.

By setting $U'' = (U',T')$ where $\mathcal{U}'' = \mathcal{U}' \times \mathcal{T}'$, we have that $T' - U'' - X - (Y,Z)$ forms a Markov chain. To see that the Markov chain $\hat{X} - (U'',Y) - (X,Z,T')$ also holds, we consider

$$\begin{aligned} I(\hat{X};X,Z,T'|U'',Y) &= I(\hat{X};X,Z|U',T',Y) \\ &= H(\hat{X}|U',T',Y) - H(\hat{X}|U',T',Y,X,Z) \\ &\overset{(a)}{=} H(\hat{X}|U,T',Y) - H(\hat{X}|U,T',Y,X,Z) \\ &\overset{(b)}{=} H(\hat{X}|U,T,Y) - H(\hat{X}|U,T,Y,X,Z) \\ &\overset{(c)}{=} 0, \end{aligned}$$

where (a) follows from preservation by U', (b) follows from preservation by T', and (c) from the Markov chain $\hat{X} - (U,Y) - (X,Z,T)$.

Furthermore, we have the following preservations by U'',

$$\begin{aligned} H(X,Y|U'') - H(Y|U'') &= H(X,Y|U',T') - H(Y|U',T') \\ &\overset{(a)}{=} H(X,Y|U,T') - H(Y|U,T') \\ &\overset{(b)}{=} H(X,Y|U,T) - H(Y|U,T) \\ &\overset{(c)}{=} H(X,Y|U) - H(Y|U), \end{aligned}$$

where (a) follows from preservation by U', (b) follows from preservation by T', and (c) follows from the Markov chain $T - U - X - (Y,Z)$. Similarly, from preservation by U' and T', and the Markov chain $T - U - X - (Y,Z)$ and $\hat{X} - (U,Y) - (X,Z,T)$, we have that $H(\hat{X},Y|U'') - H(Y|U'') = H(\hat{X},Y|U) - H(Y|U)$ and $H(Z,Y|U'') - H(Y|U'') = H(Z,Y|U) - H(Y|U)$.

Therefore, we have shown that $T \in \mathcal{T}$ and $U \in \mathcal{U}$ may be replaced by $T' \in \mathcal{T}'$ and $U'' \in \mathcal{U}''$ satisfying

$$\begin{aligned} |\mathcal{T}'| &= |\mathcal{X}| + 5 \\ |\mathcal{U}''| &= |\mathcal{T}'||\mathcal{U}'| = (|\mathcal{X}| + 5)(|\mathcal{X}| + 4), \end{aligned}$$

and preserving the terms $H(X,Y|U) - H(Y|U)$, $H(\hat{X},Y|U) - H(Y|U)$, $H(Z,Y|U) - H(Y|U)$, $I(Y;\hat{X},Z|T) - H(Z|T)$, $E[d(F(X,Y),\hat{X})]$, and the Markov relations.

Chapter 8

Conclusion

8.1 Concluding Remarks

In this thesis, we have investigated several problems involving source and channel coding with additional constraints. Motivated by the growing needs in various applications, we have expanded previous models of the communication system with new features and objectives such as signal reconstruction and information privacy constraints and characterized the corresponding fundamental limits and tradeoffs.

In Chapters 3 to 5, we studied the communication problem with flexible information acquisition, modeled by action-dependent side information. This framework allows us to have more degrees of freedom to control or influence the interactions among nodes in the network in the form of side information, and is relevant to many applications such as sensor networking and controlled sensing. In Chapters 3 and 4, we considered signal reconstruction requirement as a new objective in the system model and derived the corresponding fundamental tradeoff. The results reveal fundamentally how a newly imposed reconstruction constraint affects the optimal tradeoffs. They could be used as a guideline for designing a new system or for modifying existing systems with some new constraints. The duality between the source and channel coding problems considered in Chapter 3 was also explored. From the obtained rate-distortion-cost function and the channel capacity, we concluded that the formula duality does not hold in general due to some differences in the operational structure between the problems. The capacity result in Chapter 3 also revealed some interesting implication of the additional reconstruction (reversible input) requirement. That is, the channel capacity is expressed in a form involving a restriction on the set of feasible input distributions, termed as the two-stage coding condition. This condition was shown to be related to an underlying rate constraint in the multi-stage setting studied in Chapter 4.

In Chapters 5 to 7, we take into account several aspects of information privacy in the system model. When several users exchange data over a network, it is important that sensitive information is kept private or secret from unintended users. Privacy and security in compression systems have become relevant, especially

when dealing with large amount of data. They have recently found applications in databases, distributed storage of genomic data, etc. We considered three different aspects of information privacy, namely privacy of the source sequence against an external eavesdropper, privacy of the source sequence in the presence of an unintended legitimate user (public helper), and privacy of the reconstruction sequence against the eavesdropper or user in the system. To measure information privacy, we use basic information theoretic quantities such as conditional entropy (equivocation) or mutual information (leakage) which provide asymptotic guarantees on the average uncertainty or leakage of information. The general aim of the study is to understand the new fundamental tradeoff between compression efficiency, reliability, and information privacy. The obtained results provide insight into how to design the communication and compression systems that are both reliable and secure.

8.2 Future Work

In the following, we discuss some potential directions for future research.

- *Implicit feedback vs explicit feedback in lossy source coding with side information*: In Chapter 3, we studied a lossy source coding problem with common reconstruction constraint which requires that the sender can locally reproduce the exact copy of the receiver's reconstruction sequence. This may be viewed as having an implicit feedback from the receiver to the sender. It is therefore natural and interesting to compare the result in Chapter 3 with that of the system model where there exists an explicit rate-limited feedback link. For example, how does the system performance compare in terms of total sum-rate over rate-limited links when we have the CR constraint (implicit feedback) and when the decoder sends back a description of the reconstruction as an explicit feedback to the encoder?

- *Code design for common reconstruction and secure compression with side information*: We have seen that an achievable scheme which uses the binning technique is optimal for lossy source coding with common reconstruction. It would be interesting to investigate how existing practical codes for the rate-distortion problem can be modified to suit the problem with common reconstruction constraint. The study would result in a practical code that potentially allows an implicit feedback in the communication which could be useful for control-oriented applications. Similarly as with common reconstruction constraint, an achievable scheme for secure source coding with side information is based on layered binning. It would be interesting to see if the existing nested-type codes can be modified to suit the secure source coding problem as well.

- *Differential privacy in public helper model*: In Chapter 6, we have studied the problem of source coding with a public helper where we wish to protect the

source sequence from an intermediate node who helps to relay information to the receiver. In some scenarios, the source may be composed of several parts and are supposed to be reconstructed at the decoder through the help of a public helper in the network (see, e.g., [Yam94], for differential privacy in the Shannon cipher system without a helper). However, only some parts require privacy, i.e., the helper is not allowed to know about the private part of the source. We are interested in characterizing the fundamental tradeoff between the rate, distortion, and privacy of the *private part* of the sources.

- *Optimal use of side information for both compression and secret key generation*: In Chapters 5 and 6, we considered the special case under log-loss distortion of secure source coding problems with common side information at the encoder and decoder. Under this assumption, it was shown that the side information at the encoder is not helpful for compression, but can be used to generate a secret key to reduce the leakage. In general case, it would be interesting to investigate how to utilize the side information at the encoder for both compression and secret key generation in the optimal way.
- *Rate-limited randomness at the decoder for end-user privacy*: In Chapter 7, we have seen that the use of a stochastic decoder is helpful in enlarging the rate-distortion-equivocation region. As the source of randomness in practice may be limited, we might consider an explicit source of randomness which has a limited rate, and aim to characterize the tradeoff among the compression rate, distortion, equivocation, and rate of the random source.
- *Stochastic encoder for end-user privacy*: Is a stochastic encoder helpful when a stochastic decoder is already used? Can the randomness introduced by the encoder be transferred to the decoder without loss?
- *Active eavesdropper (adversary) and other privacy/security metrics*: Throughout the thesis, an eavesdropper is assumed to be passive in the sense that it only listens to the eavesdropped transmission, and we use basic information theoretic quantities such as conditional entropy (equivocation) and mutual information (leakage) to measure information privacy. It could be possible that the eavesdropper in the network is active and wishes to decode the signal of interest to increase his/her utility. Other relevant metrics can then be considered, e.g., the maximum of minimum distortion that the adversary can achieve. The problem formulation can be extended to take into account the action of the adversary and to consider a joint payoff function to be optimized with respect to the legitimate users. Alternatively, we might formulate the problem as a non-cooperative game where the legitimate users and adversary have conflicting objectives. On the other hand, the eavesdropper might turn malicious and wish to tamper with the transmission. This basically changes the nature of the problem to an authentication-type or detection-type (anomaly detection) problem in which the vulnerable transmission needs to be authenticated or verified before being utilized.

Bibliography

[AASP13] B. Ahmadi, H. Asnani, O. Simeone, and H. H. Permuter, "Information embedding on actions," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2013.

[AB10] T. Alpcan and T. Basar, *Network Security: A Decision and Game Theoretic Approach*. Cambridge University Press, 2010.

[AC93] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. i. secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, Jul 1993.

[APW11] H. Asnani, H. Permuter, and T. Weissman, "Probing capacity," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7317–7332, 2011.

[AS13] B. Ahmadi and O. Simeone, "Distributed and cascade lossy source coding with a side information "vending machine"," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6807–6819, Oct 2013.

[ASCM12] B. Ahmadi, O. Simeone, C. Choudhuri, and U. Mitra, "On cascade source coding with a side information "vending machine"," in *Proc. IEEE Information Theory Workshop (ITW)*, 2012, pp. 552–556.

[ATSP13] B. Ahmadi, R. Tandon, O. Simeone, and H. Poor, "Heegard-berger and cascade source coding problems with common reconstruction constraints," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1458–1474, March 2013.

[BB11] M. Bloch and J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge Academic Press, 2011.

[BCW03] R. J. Barron, B. Chen, and G. W. Wornell, "The duality between information embedding and source coding with side information and some applications," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1159–1180, 2003.

[Ber77] T. Berger, "Multiterminal source coding," in *The information theory approach to communications*. G. Longo, Ed. Springer-Verlag, 1977, p. 170–231.

[BHO+79] T. Berger, K. Housewright, J. Omura, S. Yung, and J. Wolfowitz, "An upper bound on the rate distortion function for source coding with partial side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 25, no. 6, pp. 664–666, 1979.

[BV04] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[BY89] T. Berger and R. W. Yeung, "Multiterminal source encoding with one distortion criterion," *IEEE Trans. Inf. Theory*, vol. 35, no. 2, pp. 228–236, 1989.

[BZV96] T. Berger, Z. Zhang, and H. Viswanathan, "The ceo problem [multiterminal source coding]," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[CAW13] Y.-K. Chia, H. Asnani, and T. Weissman, "Multiterminal source coding with action-dependent side information," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3653–3667, June 2013.

[CC02] T. M. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1629–1638, Jun. 2002.

[CK78] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, 1978.

[CK11] ——, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[CK13a] Y.-K. Chia and K. Kittichokechai, "On secure source coding with side information at the encoder," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, July 2013, pp. 2204–2208.

[CK13b] ——, "On secure lossy source coding with side information at the encoder," *arXiv 1307.0974*, 2013.

[CM12a] C. Choudhuri and U. Mitra, "Action dependent strictly causal state communication," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, July 2012, pp. 3058–3062.

[CM12b] ——, "How useful is adaptive action?" in *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2012, pp. 2251–2255.

[Cou12] T. Courtade, "Information masking and amplification: The source coding setting," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2012, pp. 189–193.

[CPW12] Y.-K. Chia, H. H. Permuter, and T. Weissman, "Cascade, triangular, and two-way source coding with degraded side information at the second user," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 189–206, 2012.

[CT06] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[Cuf10] P. Cuff, "A framework for partial secrecy," in *IEEE Global Communications Conference (GLOBECOM)*, 2010, pp. 1–5.

[CW14] T. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 740–761, Jan 2014.

[DPS12] L. Dikstein, H. Permuter, and S. Shamai, "Mac with action-dependent state information at one encoder," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2012, pp. 1687–1691.

[EHM07] A. El Gamal, N. Hassanpour, and J. Mammen, "Relay networks with delays," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3413–3431, Oct 2007.

[EK11] A. El Gamal and Y. H. Kim, *Network Information Theory.* Cambridge University Press, 2011.

[Eri13] Ericsson. (2013) Networked society. [Online]. Available: http://www.ericsson.com/thinkingahead/networked_society

[EU11] E. Ekrem and S. Ulukus, "Secure lossy source coding with side information," in *Proc. 49th Annual Allerton Conf. Communication, Control, and Computing (Allerton)*, 2011, pp. 1098–1105.

[Eys01] G. Eysenbach. (2001) What is e-health? J Med Internet Res 2001;3(2):e20. [Online]. Available: http://www.jmir.org/2001/2/e20/

[Gal68] R. G. Gallager, *Information Theory and Reliable Communication.* John Wiley and Sons, New York, 1968.

[GEP08] D. Gündüz, E. Erkip, and H. V. Poor, "Lossless compression with security constraints," *Proc. IEEE International Symposium on Information Theory (ISIT)*, pp. 111–115, 2008.

[GMG+13] M. Gymrek, A. L. McGuire, D. Golan, E. Halperin, and Y. Erlich, "Identifying personal genomes by surname inference," *Science*, vol. 339, no. 6117, pp. 321–324, 2013. [Online]. Available: http://www.sciencemag.org/content/339/6117/321.abstract

[GP80] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Contr. Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[GV11] A. Gupta and S. Verdu, "Operational duality between lossy compression and channel coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3171–3179, 2011.

[HB85] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 727–734, Nov 1985.

[HE83] C. Heegard and A. El Gamal, "On the capacity of computer memory with defects," *IEEE Trans. Inf. Theory*, vol. 29, no. 5, pp. 731–739, Sep 1983.

[HW89] A. P. Hekstra and F. M. J. Willems, "Dependence balance bounds for single-output two-way channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 44–53, 1989.

[JB08] S. Jana and R. Blahut, "Partial side information problem: Equivalence of two inner bounds," in *Proc. 42nd Annual Conf. Information Sciences and Systems (CISS)*, 2008, pp. 1005–1009.

[KBLV13] L. Kamm, D. Bogdanov, S. Laur, and J. Vilo, "A new way to protect privacy in large-scale genome-wide association studies," *Bioinformatics*, vol. 29, no. 7, pp. 886–893, Apr. 2013. [Online]. Available: http://dx.doi.org/10.1093/bioinformatics/btt066

[KCO+13a] K. Kittichokechai, Y.-K. Chia, T. Oechtering, M. Skoglund, and T. Weissman, "Secure source coding with a public helper," *submitted to IEEE Trans. Inf. Theory*, July 2013.

[KCO+13b] ——, "Secure source coding with a public helper," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, July 2013, pp. 2209–2213.

[KOS10] K. Kittichokechai, T. J. Oechtering, and M. Skoglund, "Source coding with common reconstruction and action-dependent side information," in *Proc. IEEE Information Theory Workshop (ITW)*, 2010, pp. 1–5.

[KOS11a] ——, "Capacity of the channel with action-dependent state and reversible input," in *Proc. IEEE Swedish Communication Technologies Workshop (Swe-CTW)*, 2011, pp. 24–28.

[KOS11b] ——, "On the capacity of a channel with action-dependent state and reversible input," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2011, pp. 331–335.

[KOS11c] ——, "Secure source coding with action-dependent side information," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, July 2011.

[KOS12a] ——, "Coding with action-dependent side information and additional reconstruction requirements," *submitted to IEEE Trans. Inf. Theory*, Feb. 2012, coRR, abs/1202.1484.

[KOS12b] ——, "Multi-stage coding for channels with a rewrite option and reversible input," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2012, pp. 3063–3067.

[KOS14a] ——, "Lossy source coding with reconstruction privacy," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2014, to appear.

[KOS+14b] K. Kittichokechai, T. J. Oechtering, M. Skoglund, Y.-K. Chia, and T. Weissman, "Secure source coding with action-dependent side information," *to be submitted to IEEE Trans. Inf. Theory*, 2014.

[KOS14c] K. Kittichokechai, T. J. Oechtering, and M. Skoglund, "Lossy source coding with reconstruction privacy," *in preparation*, 2014, arXiv 1402.4308.

[KOST10] K. Kittichokechai, T. J. Oechtering, M. Skoglund, and R. Thobaben, "Source and channel coding with action-dependent partially known two-sided state information," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2010, pp. 629–633.

[KSC08] Y.-H. Kim, A. Sutivong, and T. M. Cover, "State amplification," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1850–1859, 2008.

[KSM07] G. Keshet, Y. Steinberg, and N. Merhav, "Channel coding in the presence of side information," *Found. Trends Commun. Inf. Theory*, vol. 4, no. 6, pp. 445–586, 2007.

[LMW11] A. Lapidoth, A. Malar, and M. Wigger, "Constrained wyner-ziv coding," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2011, pp. 1076–1080.

[LOK14] Z. Li, T. J. Oechtering, and K. Kittichokechai, "Parallel distributed bayesian detection with privacy constraints," in *IEEE International Conference on Communications (ICC)*, 2014, to appear.

[LPSS08] Y. Liang, H. V. Poor, and S. Shamai (Shitz), "Information theoretic security," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 4–5, pp. 355–580, 2008.

[Mau93] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 733–742, 1993.

[Mer06a] N. Merhav, "On joint coding for watermarking and encription," *IEEE Trans. Inf. Theory*, vol. 52, pp. 190–205, Jan 2006.

[Mer06b] ——, "On the shannon cipher system with a capacity-limited key-distribution channel," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1269–1273, Mar 2006.

[Mer08] ——, "Shannon's secrecy system with informed receivers and its application to systematic coding for wiretapped channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2723–2734, 2008.

[MF10] F. Mattern and C. Floerkemeier, "From the internet of computers to the internet of things." Springer Berlin Heidelberg, 2010, vol. 6462, pp. 242–259. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-17226-7_15

[MS07] N. Merhav and S. Shamai, "Information rates subject to state masking," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2254–2261, June 2007.

[MSDPC12] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, "Survey internet of things: Vision, applications and research challenges," *Ad Hoc Netw.*, vol. 10, no. 7, pp. 1497–1516, Sep. 2012.

[MW12] J. McDonald and N. Weiss, *A Course in Real Analysis.* Academic Press, 2012.

[NSTS13] F. Naghibi, S. Salimi, R. Thobaben, and M. Skoglund, "The lossless ceo problem with security constraints," in *Proceedings of the Tenth International Symposium on Wireless Communication Systems (ISWCS)*, Aug 2013, pp. 1–5.

[Ooh97] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1912–1923, 1997.

[OR01] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, Mar 2001.

[PCR03] S. S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1181–1203, 2003.

[Pea00] J. Pearl, *Causality Models, Reasoning, and Inference*. Cambridge University Press, 2000.

[PR07] V. Prabhakaran and K. Ramchandran, "On secure distributed source coding," in *Proc. IEEE Information Theory Workshop (ITW)*, 2007, pp. 442–447.

[PSW10] H. Permuter, Y. Steinberg, and T. Weissman, "Two-way source coding with a helper," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, June 2010.

[PW11] H. Permuter and T. Weissman, "Source coding with a side information "vending machine"," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4530–4544, 2011.

[SBSV08] A. Somekh-Baruch, S. Shamai, and S. Verdu, "Cooperative multiple-access encoding with states available at one transmitter," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4448–4469, Oct 2008.

[SC13] C. Schieler and P. Cuff, "Rate-distortion theory for secrecy systems," *arXiv 1305.3905*, 2013.

[Sha48] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July, October 1948.

[Sha49] ——, "Communication theory of secrecy systems," *Bell System Technical Journal*, vol. 28, pp. 656—-715, Oct. 1949.

[Sha59] ——, "Coding theorems for a discrete source with a fidelity criterion," *Institute of Radio Engineers, International Convention Record*, vol. 7, 1959.

[SRP13] L. Sankar, S. Rajagopalan, and H. Poor, "Utility-privacy tradeoffs in databases: An information-theoretic approach," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 6, pp. 838–852, June 2013.

[SS09] O. Sumszyk and Y. Steinberg, "Information embedding with reversible stegotext," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, July 2009, pp. 2728 –2732.

[Ste09] Y. Steinberg, "Coding and common reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4995–5010, Nov. 2009.

[Ste13] ——, "The degraded broadcast channel with non-causal action-dependent side information," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, July 2013, pp. 2965–2969.

[SW73] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[SW12] Y. Steinberg and T. Weissman, "The degraded broadcast channel with action-dependent states," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, July 2012, pp. 596–600.

[TASP12] R. Tandon, B. Ahmadi, O. Simeone, and H. Poor, "Gaussian multiple descriptions with common and constrained reconstruction constraints," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2012, p. 1376–1380.

[TD07] C. Tian and S. N. Diggavi, "On multistage successive refinement for wyner-ziv source coding with degraded side informations," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2946–2960, Aug. 2007.

[TGK13] R. Timo, A. Grant, and G. Kramer, "Lossy broadcasting with complementary side information," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 104–131, Jan 2013.

[TSP13] R. Tandon, L. Sankar, and H. Poor, "Discriminatory lossy source coding: Side information privacy," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5665–5677, Sept 2013.

[TUR13] R. Tandon, S. Ulukus, and K. Ramchandran, "Secure source coding with a helper," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2178–2187, 2013.

[VP13] J. Villard and P. Piantanida, "Secure multiterminal source coding with side information at the eavesdropper," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3668–3692, June 2013.

[VTD06] D. Vasudevan, C. Tian, and S. N. Diggavi, "Lossy source coding for a cascade communication system with side informations," in *Proc. Allerton Conf. Commun. Control Comput*, 2006.

[WE06] T. Weissman and A. El Gamal, "Source coding with limited-look-ahead side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5218–5239, Dec 2006.

[Wei10] T. Weissman, "Capacity of channels with action-dependent states," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5396–5411, 2010.

[WK03] F. M. J. Willems and T. Kalker, "Coding theorems for reversible embedding," in *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 66, Mar. 2003, pp. 61–76.

[WTV08] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic gaussian two-encoder source-coding problem," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1938–1961, 2008.

[WvdM85] F. M. J. Willems and E. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, pp. 313–327, May 1985.

[Wyn75] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, no. 8, p. 1355–1387, 1975.

[Wyn78] ——, "The rate-distortion function for source coding with side information at the decoder—-part ii: General sources," *Inf. Control*, no. 38, pp. 60—-80, 1978.

[WZ73] A. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications–i," *IEEE Trans. Inf. Theory*, vol. 19, no. 6, pp. 769–772, Nov 1973.

[WZ76] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan 1976.

[XKOS14] D. Xu, K. Kittichokechai, T. J. Oechtering, and M. Skoglund, "Secure successive refinement with degraded side information," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2014, to appear.

[Yam81] H. Yamamoto, "Source coding theory for cascade and branching communication systems," *IEEE Trans. Inf. Theory*, vol. 27, pp. 299–308, 1981.

[Yam82] ——, "Wyner-ziv theory for a general function of the correlated sources," *IEEE Trans. Inf. Theory*, vol. 28, no. 5, pp. 803–807, Sep 1982.

[Yam94] ——, "Coding theorems for shannon's cipher system with correlated source outputs, and common information," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 85–95, 1994.

[Yam97] ——, "Rate-distortion theory for the shannon cipher system," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 827–835, 1997.

[ZCW14] L. Zhao, Y.-K. Chia, and T. Weissman, "Compression with actions," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 796–807, Feb 2014.

[ZPS12] A. Zaidi, P. Piantanida, and S. Shamai, "Wyner-ziv type versus noisy network coding for a state-dependent mac," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, July 2012.