Joint Compression and Digital Watermarking: Information-Theoretic Study and Algorithms Development

by

Wei Sun

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2006

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Wei Sun

Abstract

In digital watermarking, a watermark is embedded into a covertext in such a way that the resulting watermarked signal is robust to certain distortion caused by either standard data processing in a friendly environment or malicious attacks in an unfriendly environment. The watermarked signal can then be used for different purposes ranging from copyright protection, data authentication, fingerprinting, to information hiding. In this thesis, digital watermarking will be investigated from both an information theoretic viewpoint and a numerical computation viewpoint.

From the information theoretic viewpoint, we first study a new digital watermarking scenario, in which watermarks and covertexts are generated from a joint memoryless watermark and covertext source. The configuration of this scenario is different from that treated in existing digital watermarking works, where watermarks are assumed independent of covertexts. In the case of public watermarking where the covertext is not accessible to the watermark decoder, a necessary and sufficient condition is determined under which the watermark can be fully recovered with high probability at the end of watermark decoding after the watermarked signal is disturbed by a fixed memoryless attack channel. Moreover, by using similar techniques, a combined source coding and Gel'fand-Pinsker channel coding theorem is established, and an open problem proposed recently by Cox et al is solved. Interestingly, from the sufficient and necessary condition we can show that, in light of the correlation between the watermark and covertext, watermarks still can be fully recovered with high probability even if the entropy of the watermark source is strictly above the standard public watermarking capacity.

We then extend the above watermarking scenario to a case of joint compression and watermarking, where the watermark and covertext are correlated, and the watermarked signal has to be further compressed. Given an additional constraint of the compression rate of the watermarked signals, a necessary and sufficient condition is determined again under which the watermark can be fully recovered with high probability at the end of public

watermark decoding after the watermarked signal is disturbed by a fixed memoryless attack channel.

The above two joint compression and watermarking models are further investigated under a less stringent environment where the reproduced watermark at the end of decoding is allowed to be within certain distortion of the original watermark. Sufficient conditions are determined in both cases, under which the original watermark can be reproduced with distortion less than a given distortion level after the watermarked signal is disturbed by a fixed memoryless attack channel and the covertext is not available to the watermark decoder.

Watermarking capacities and joint compression and watermarking rate regions are often characterized and/or presented as optimization problems in information theoretic research. However, it does not mean that they can be calculated easily. In this thesis we first derive closed forms of watermarking capacities of private Laplacian watermarking systems with the magnitude-error distortion measure under a fixed additive Laplacian attack and a fixed arbitrary additive attack, respectively. Then, based on the idea of the Blahut-Arimoto algorithm for computing channel capacities and rate distortion functions, two iterative algorithms are proposed for calculating private watermarking capacities and compression and watermarking rate regions of joint compression and private watermarking systems with finite alphabets. Finally, iterative algorithms are developed for calculating public watermarking capacities and compression and watermarking rate regions of joint compression and public watermarking systems with finite alphabets based on the Blahut-Arimoto algorithm and the Shannon's strategy.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisor Professor En-hui Yang at the University of Waterloo. His broad knowledge, insightful understanding, encouragement and financial support are invaluable for me to successfully finish the thesis. It is really one of the greatest experiences in my life to be supervised by Dr. Yang, and what I learned from him will definitely benefit me in future.

Second, I gratefully acknowledge my doctoral committee members, Professor Frans M. J. Willems of the Department of Electrical Engineering, Eindhoven University of Technology, Netherlands, Professor Jiahua Chen of the Department of Statistics and Actuarial Science, University of Waterloo, Professor Jon W. Mark of the Department of Electrical and Computer Engineering, University of Waterloo, and Professor Amir K. Khandani of the Department of Electrical and Computer Engineering, University of Waterloo, for their valuable time, helpful advice and important comments. Thanks are also given to my comprehensive committee member Professor G. Gong of the Department of Electrical and Computer Engineering, University of Waterloo.

Third, I am grateful to Professor Ingemar J. Cox of the Department of Electronic and Electrical Engineering, University College London, and Professor L. L. Xie of the Department of Electrical and Computer Engineering, University of Waterloo, for their valuable discussions.

Four, I am deeply indebted to my friends of the Multimedia Communications Laboratory at the University of Waterloo, Dr. Yunwei Jia (now with Gennum Co.), Alexei Kaltchenko (now with Wilfrid Laurier University), Dake He (now with IBM T. J. Watson Research Center), Dr. Guixing Wu, Dr. Haiquan Wang, Mr. Xiang Yu, Dr. Lusheng Chen, Mr. Xudong Ma and many other friends, for their help, discussions.

Last, but certainly not the least, I am much obliged to my wife Yanqiu Yang, for her love, understanding and support during the past difficult years in Canada, and my sons Yixiong and David, for their cuteness, laugh, and happiness brought to me.

To my wife Yanqiu Yang

and

my sons Yixiong and David

Contents

List of Figures

Chapter 1

Introduction

1.1 Digital Watermarking

The development of the Internet has made it much easier to access digital data than ever as audio, videos and other works become available in digital form. With Internet connection, one can easily download and distribute perfect copies of pictures, music, and videos; with suitable softwares, one can also alter these copyright-protected digital media. All these activities can be carried out by would-be pirates without paying appropriate compensation to the actual copyright owners, resulting in a huge economic risk to content owners. Thus, there is a strong need for techniques to protect the copyright of content owners. Cryptography and digital watermarking are two complementary techniques proposed so far to protect digital content.

Cryptography is the processing of information into an encrypted form for the purpose of secure transmission. Before delivery, the digital content is encrypted by the owner by using a secret key. A corresponding decryption key is provided only to a legitimate receiver. The encrypted content is then transmitted via Internet or other public channels, and it will be meaningless to pirate without the decryption key. At the receiver end, however, once the encrypted content is decrypted, it has no protection anymore.

On the other hand, digital watermarking is a technique that can protect the digital content even after it is decrypted. In digital watermarking, a watermark is embedded into a **covertext** (the digital contents to be protected), resulting in a watermarked signal called **stegotext** which has no visible difference from the covertext. In a successful watermarking system, watermarks should be embedded in such a way that the watermarked signals are robust to certain distortion caused by either standard data processing in a friendly environment or malicious attacks in an unfriendly environment. In other words, watermarks still can be recovered from the attacked watermarked signal (called **forgery**) generated by an attacker if the attack is not too much. A watermarking system is called **private** if the covertext is available to both the watermark encoder and decoder, and **public** if the covertext is available only to the watermark encoder.

The application of digital watermarking is very broad, including copyright protection, information hiding, fingerprinting, etc. For more detailed introduction and applications of digital watermarking, please refer to [10] and [25].

1.2 Research Problems and Motivations

From an information theoretic viewpoint, a major research problem on digital watermarking is to determine best tradeoffs among the distortion between the covertext and stegotext, the distortion between the stegotext and forgery, the watermark embedding rate, the compression rate and the robustness of the stegotext. Along this direction, some information theoretic results, such as watermarking capacities and watermarking error exponents and joint compression and watermarking rate regions, have been determined. Please refer to [5, 7, 19, 25, 26, 27, 32, 33] and references therein for more information theoretic results, and [25] is an excellent summary of the state of art.

The research problems to be investigated in this thesis are:

- From the viewpoint of information theory, for public digital watermarking systems

with correlated watermarks and covertexts, what's the best tradeoff among distortion level, compression rate, robustness of stegotexts and admissibility of joint watermark and covertext sources? Or under what conditions can watermarks be conveyed successfully to watermark decoder with high probability?

- From the viewpoint of computation, how can watermarking capacities and compression and watermarking rate regions of joint compression and watermarking systems be calculated efficiently?

The motivations for the above research problems are two-fold. First, in existing information-theoretic works on digital watermarking systems, the watermark to be embedded is often assumed explicitly or implicitly independent of the covertext. In some cases, for instance, a self watermarking system in which watermarks are extracted from covertexts by feature extraction techniques, however, the watermark to be embedded is correlated with the covertext. Without utilizing this correlation, a simple scheme for embedding such a watermark is to first compress the watermark and then embed the compressed watermark into the covertext as usual. If the entropy of the watermark is less than the standard public watermarking capacity, then the watermark can be recovered with high probability after watermark decoding in the case of public watermarking. Now the question is: in light of the correlation between the watermark and covertext, can one do better in the case of public watermarking? In other words, can the watermark still be recovered with high probability if its entropy is strictly above the standard public watermarking capacity? Furthermore, in many applications, watermarked signals are stored and/or transmitted in compressed formats, and/or the reproduced watermark at the end of decoding is allowed to be within certain distortion of the original watermarks, so, in these scenarios under what conditions can watermarks be conveyed to watermark decoder with high probability of success?

Second, although watermarking capacities and compression and watermarking rate regions of joint compression and watermarking systems can be characterized as optimization problems, the characterization does not mean that they can be calculated easily. Indeed,

solving these optimization problem is often very difficult, and closed forms of watermarking capacities and joint compression and watermarking rate regions are known only to very few cases. Therefore, it is important and necessary to develop efficient algorithms for numerically computing watermarking capacities and joint compression and watermarking rate regions.

1.3 Thesis Organization and Contributions

This thesis will study digital watermarking systems from an information theoretic viewpoint and from a computational viewpoint, respectively. From the viewpoint of information theory, we investigate a digital watermarking scenario with correlated watermarks and covertexts in Chapter 2, Chapter 3 and Chapter 4; from the viewpoint of numerical computation we obtain closed-forms of private watermarking capacities for Laplacian watermarking systems in Chapter 5 and propose iterative algorithms for numerically calculating watermarking capacities and joint compression and watermarking rate regions for private watermarking and public watermarking in Chapter 6 and Chapter 7, respectively. The organization and contributions of this thesis are summarized as follows.

In Chapter 2, from the information theoretic viewpoint we study a new digital watermarking scenario with correlated watermarks and covertexts. In the case of public watermarking where the covertext is not accessible to the watermark decoder, a necessary and sufficient condition is determined under which the watermark can be fully recovered with high probability at the end of watermark decoding after the watermarked signal is disturbed by a fixed memoryless attack channel. Moreover, by using similar techniques, a combined source coding and Gel'fand-Pinsker channel coding theorem is established, and an open problem proposed recently by Cox et al is solved. Interestingly, from the sufficient and necessary condition we can show that, in light of the correlation between the watermark and covertext, watermarks still can be fully recovered with high probability even if the entropy of the watermark source is strictly above the standard public watermarking

capacity.

In Chapter 3, the watermarking scenario of Chapter 2 is extended to a case of joint compression and public watermarking, where the watermark and covertext are correlated, and the watermarked signal has to be further compressed. For a given distortion level between the covertext and the watermarked signal and a given compression rate of the watermarked signal, a necessary and sufficient condition is determined again under which the watermark can be fully recovered with high probability at the end of watermark decoding after the watermarked signal is disturbed by a fixed memoryless attack channel and the covertexts is not available to the watermark decoder.

The above two joint compression and watermarking models are further investigated in Chapter 4 under a less stringent environment where the reproduced watermark at the end of decoding is allowed to be within certain distortion of the original watermark. Sufficient conditions are determined for the case without compression of watermarked signals and for the case with compression of watermarked signals, respectively, under which watermarks can be reproduced within a given distortion level with respect to the original watermarks at the end of public watermark decoding after the watermarked signals are disturbed by a fixed memoryless attack channel.

From the viewpoint of computation, Chapter 5 derives closed-forms for watermarking capacities of private Laplacian watermarking systems with the magnitude-error distortion measure under a fixed additive Laplacian attack and a fixed arbitrary additive attack, respectively.

Based on the idea of the Blahut-Arimoto algorithm for computing channel capacities and rate distortion functions, two iterative algorithms are proposed in Chapter 6 which can be combined to calculate private watermarking capacities and joint compression and private watermarking rate regions. Similarly, based on both the Blahut-Arimoto algorithm and Shannon's strategy, in Chapter 7 iterative algorithms are proposed for calculating public watermarking capacities and joint compression and public watermarking rate regions.

The last chapter, Chapter 8, is the conclusion of the thesis, and contains some future works.

1.4 Notations

Throughout the thesis, the following notations are adopted. We use capital letter to denote random variable, lowercase letter for its realization, and script letter for its alphabet. For example, S is a random variable over its alphabet $\mathcal{S}$ and $s \in \mathcal{S}$ is a realization. We use $p_S(s)$ to denote the probability distribution of a discrete random variable S taking values over its alphabet $\mathcal{S}$, that is, $p_S(s) \stackrel{def}{=} \Pr\{S = s\}$; the same notation $p_S(s)$ also is used to denote the probability density function of a continuous random variable S. If no ambiguity, the subscript in $p_S(s)$ is omitted and write $p_S(s)$ as $p(s)$. Similarly, $S^n = (S_1, S_2, ..., S_n)$ denotes a random vector taking values over $\mathcal{S}^n$, and $s^n = (s_1, s_2, ..., s_n)$ is a realization. Also, we always assume the attack is fixed and given by a conditional probability distribution $p(y|x)$ with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. Notations frequently used in this thesis are summarized as follows.

List of notations

$\mathcal{M}$	Watermark alphabet
M, M^n	Watermarks
$\hat{\mathcal{M}}$	Reproduction watermark alphabet
$\hat{M}, \hat{M}^n$	Decoded watermarks
$\mathcal{S}$	Covertext alphabet
S^n	Covertexts
$\mathcal{X}$	Stegotext alphabet
X^n	Stegotexts
$\mathcal{Y}$	Forgery alphabet
Y^n	Forgeries
$p(y\|x)$	An attack channel with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$
$f_n(M, S^n), f_n(M^n, S^n)$	Watermark encoder
$g_n(S^n, Y^n)$	Private watermark decoder
$g_n(Y^n)$	Public watermark decoder
$p(s)$	The pmf of a covertext source S
$p(m, s)$	The joint pmf of a joint watermark and covertext source (M, S)
d, d_1	Distortion measures
D, D_1	Distortion levels
$\mathbf{E}$	The expectation operator
$C(D)$	The watermarking capacity
$H(X)$	The entropy of X
$I(X;Y)$	The mutual information between X and Y
R_c	The compression rate of stegotexts
R_w	The watermarking rate

Chapter 2

On Information Embedding When Watermarks and Covertexts Are Correlated

In this chapter, the standard model of digital watermarking is introduced first from an information theoretic viewpoint. Then, the main problem on watermarking models with correlated watermarks and covertexts is formulated and the results of this chapter are stated. Next, by employing a similar approach, a combined source-channel coding theorem on Gel'fand-Pinsker channel is obtained, and an open problem proposed by Cox et al is solved. Finally, the proofs of the main results are provided.

2.1 Basic Communication Model of Digital Watermarking

From an information theoretic viewpoint, a digital watermarking system can be modeled as a communication system with side information at the watermark transmitter, as depicted in Figure 2.1. In this model, a watermark M is assumed to be a random variable uniformly

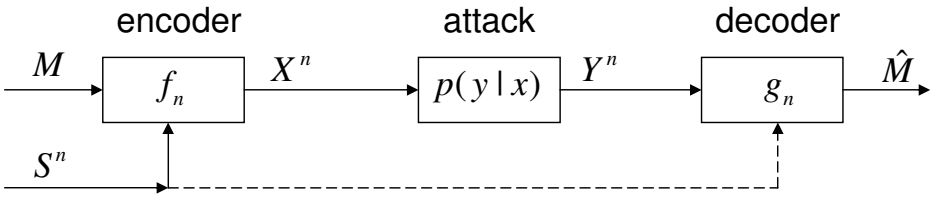

Figure 2.1: Basic communication model of digital watermarking

taking values over $\mathcal{M} = \{1, 2, ..., |\mathcal{M}|\}$, and a covertext S^n is a sequence of independent and identical drawings of a random variable S with probability distribution $p(s)$ taking values over a finite alphabet $\mathcal{S}$. If the covertext S^n is available to the watermarking decoder, then the watermarking model is called **private**; otherwise, if the covertext S^n is not available to the watermarking decoder, then the watermarking model is called **public**.

Let $\mathcal{X}$ be a finite alphabet, and define a distortion measure $d : \mathcal{S} \times \mathcal{X} \to [0, \infty)$ and let $d_{max} = \max_{s \in \mathcal{S}, x \in \mathcal{X}} d(s, x)$. Without loss of generality, assume that

$$\max_{s \in \mathcal{S}} \min_{x \in \mathcal{X}} d(s, x) = 0.$$

Let $\{d_n : n = 1, 2, ...\}$ be a single-letter fidelity criterion generated by d, where

$$d_n : \mathcal{S}^n \times \mathcal{X}^n \to [0, \infty)$$

is a mapping defined by

$$d_n(s^n, x^n) = \frac{1}{n} \sum_{i=1}^{n} d(s_i, x_i)$$

for any $s^n \in \mathcal{S}^n$ and $x^n \in \mathcal{X}^n$. Without ambiguity, the subscript n in d_n is omitted throughout the thesis.

Let $p(y|x)$ be a conditional probability distribution with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ and $p(y^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i)$ denote a fixed memoryless attack channel with input x^n and output y^n.

Definition 2.1 *A **watermarking encoder** of length n with distortion level D with respect to the distortion measure d is a mapping f_n from $\mathcal{M}\times\mathcal{S}^n$ to $\mathcal{X}^n$ such that $\mathbf{E}d(S^n, X^n) \leq D$, where the watermarked signal $x^n = f_n(m, s^n)$ is called **stegotext**. Moreover, $R = \frac{1}{n}\log|\mathcal{M}|$ is called its watermarking rate.*

Definition 2.2 *A mapping $g_n : \mathcal{S}^n \times \mathcal{Y}^n \to \mathcal{M}$, $\hat{m} = g_n(s^n, y^n)$ is called a **private watermarking decoder** of length n; A mapping $g_n : \mathcal{Y}^n \to \mathcal{M}$, $\hat{m} = g_n(y^n)$ is called a **public watermarking decoder** of length n. Here, the **forgery** y^n is generated by the attacker according to the attack channel $p(y^n|x^n)$ with input covertext x^n.*

Given a watermarking encoder and watermarking decoder pair (f_n, g_n), the error probability of watermarking is defined by

$$p_e(f_n, g_n) = \Pr\{\hat{M} \neq M\}.$$

Definition 2.3 *A rate $R \geq 0$ is called **privately (publicly) achievable** with respect to distortion level D if for arbitrary $\epsilon > 0$, there exists, for any sufficiently large n, a watermarking encoder f_n with rate $R - \epsilon$ and distortion level $D + \epsilon$ and a private (public) watermarking decoder g_n such that $p_e(f_n, g_n) < \epsilon$. The supremum of all privately (publicly) achievable rates R with respect to distortion level D is called the **private (public) watermarking capacity** of the watermarking system, and denoted by $C_{private}(D)$ and $C_{public}(D)$ respectively.*

From an information theoretic viewpoint, a major research problem is to determine the best tradeoffs among the distortion D between covertexts and stegotexts, the watermarking embedding rate R, and the robustness of the stegotext. Along this direction, some information theoretic results have been determined (see [25, 26] and references therein).

In existing information theoretic works on digital watermarking, the watermark to be embedded is often assumed independent of the covertext. In some cases, however, the watermark to be embedded is correlated with the covertext. For instance, there exist

self-watermarking systems in which watermarks are extracted from covertexts by feature extraction techniques; another application is to embed fingerprints into passport's picture for the sake of security. Obviously, without utilizing this correlation, a simple scheme for embedding such a watermark is to first compress the watermark and then embed the compressed watermark into the covertext as usual (i.e., treating the compressed watermark as being independent of the covertext even though it is not). If the entropy of the watermark is less than the standard watermarking capacity, then the watermark can be recovered with high probability after watermark decoding. Now the question is: in light of the correlation between the watermark and covertext, can one do better? In other words, can the watermark still be recovered with high probability even if its entropy is strictly above the standard watermarking capacity?

In this chapter, we shall answer the above question by determining a necessary and sufficient condition under which the watermark can be recovered with high probability at the end of watermark decoding in the case of public watermarking. It turns out that the answer is actually affirmative. When the watermark and covertext are correlated, the watermark can indeed be recovered with high probability even when its entropy is strictly above the standard public watermarking capacity.

2.2 Problem Formulation and Result Statement

2.2.1 Problem Formulation

The model studied in this chapter is designated in Figure 2.2. Suppose $\{(M_i, S_i)\}_{i=1}^{\infty}$ be a sequence of independent and identical drawings of a pair (M, S) of random variables with joint probability distribution $p(m, s)$ taking values over the finite alphabet $\mathcal{M} \times \mathcal{S}$, that is, for any n and $m^n \times s^n \in \mathcal{M}^n \times \mathcal{S}^n$,

$$p(m^n, s^n) \;=\; \prod_{i=1}^{n} p(m_i, s_i).$$

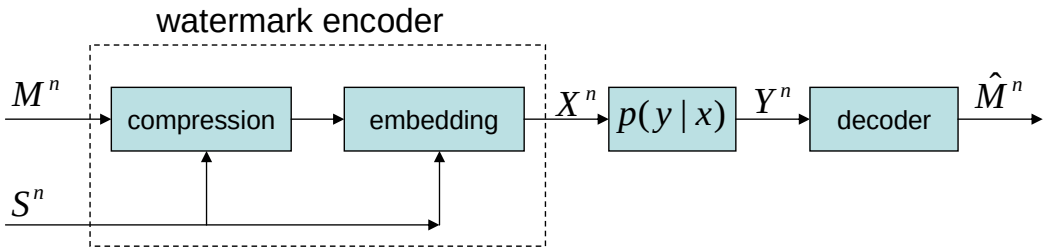

Figure 2.2: Model of watermarking system with correlated watermarks and covertexts

Here, m^n and s^n are called a watermark and a covertext respectively. As before, let $p(y|x)$ be a conditional probability distribution with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ and $p(y^n|x^n) = \prod_{i=1}^n p(y_i|x_i)$ denote a fixed memoryless attack channel with input x^n and output y^n. It is assumed that the attack channel is known to both watermark encoder and watermark decoder.

Definition 2.4 *A **watermarking encoder** of length n with distortion level D with respect to the distortion measure d is a mapping f_n from $\mathcal{M}^n \times \mathcal{S}^n$ to $\mathcal{X}^n$ such that $\mathbf{E}d(S^n, X^n) \leq D$, where the watermarked signal $x^n = f_n(m^n, s^n)$.*

Definition 2.5 *A mapping $g_n : \mathcal{Y}^n \to \mathcal{M}^n$ is called a **public watermarking decoder** of length n with $\hat{m}^n = g_n(y^n)$.*

Given a watermarking encoder and public watermarking decoder pair (f_n, g_n), the error probability of watermarking averaged over all watermarks and covertexts is defined by $p_e(f_n, g_n) = \Pr\{\hat{M}^n \neq M^n\}$.

Definition 2.6 *The joint probability distribution $p(m, s)$ of a correlated watermark and covertext source (M, S) is called **publicly admissible** with respect to distortion level D if for any $\epsilon > 0$, there exists, for any sufficiently large n, a watermarking encoder f_n with length n and distortion level $D + \epsilon$ and a public watermarking decoder g_n such that $p_e(f_n, g_n) < \epsilon$.*

An interesting problem arises naturally: under what condition is a joint probability $p(m,s)$ of a correlated watermark and covertext source (M,S) publicly admissible? It's well known [27] that the public watermarking capacity is given by

$$C_{public}(D) \stackrel{\triangle}{=} \max_{p(u,x|s):\mathbf{E}d(S,X)\leq D} [I(U;Y) - I(U;S)] \tag{2.1}$$

where the maximum is taken over all auxiliary random variables U and X jointly distributed with S and satisfying $\mathbf{E}d(S,X) \leq D$. Obviously, if $H(M) < C_{public}(D)$, then the watermark M can be recovered with high probability after watermark decoding in the case that the decoder cannot access the covertext S^n; this can be achieved by simply compressing M using $H(M)$ number of bits and then embedding the compressed M into S using standard public watermarking schemes. In other words, $H(M) < C_{public}(D)$ is a sufficient condition for $p(m,s)$ to be publicly admissible. However, it may not be necessary. Is $H(M|S) < C_{public}(D)$ a sufficient and necessary condition, where $H(M|S)$ is the conditional entropy of M given S? Note that even though S^n is available to the watermark encoder, S^n can not be fully utilized to encode M^n since S^n is not available to the watermark decoder in the public watermarking system. One of our main problems in this chapter is to determine a sufficient and necessary condition for a joint probability distribution $p(m,s)$ to be publicly admissible.

It should be pointed out that in the case of private watermarking, one can ask a similar question when watermarks and covertexts are correlated. However, since the covertext S^n is accessible to both the watermark encoder and decoder in this case, the solution to the corresponding question is straightforward; compressing M conditionally on S and then embedding the compressed watermark into S by using standard private watermarking schemes will provide one with an optimal solution.

2.2.2 Statement of Main Result

As before, let $p(m,s)$ be the joint probability distribution of a fixed correlated watermark and covertext source (M,S) taking values over $\mathcal{M} \times \mathcal{S}$. Let $D \geq 0$ be a distortion level

with respect to the distortion measure d, and $p(y|x)$ be the fixed attack channel known to watermark encoder and watermark decoder. Define

$$R_{public}^{correlated}(D) \stackrel{\triangle}{=} \sup_{p(u,x|m,s):\mathbf{E}d(S,X)\leq D} [I(U;Y) - I(U;M,S) + I(M;U,Y)] \tag{2.2}$$

where the supremum is taken over all auxiliary random variables (U, X) taking values over $\mathcal{U} \times \mathcal{X}$, jointly distributed with (M, S) with the joint probability distribution of (M, S, U, X, Y) given by $p(m, s, u, x, y) = p(m, s)p(u, x|m, s)p(y|x)$, and satisfying $\mathbf{E}d(S, X) \leq D$, and all mutual information quantities are determined by $p(m, s, u, x, y)$. It can be shown later that $|\mathcal{U}|$ can be limited by $|\mathcal{U}| \leq |\mathcal{M}||\mathcal{S}||\mathcal{X}|+1$ and so the sup in (2.2) can be replaced by max.

The following theorem is the main result, which describes the sufficient and necessary condition for public admissibility of a joint probability $p(m, s)$.

Theorem 2.1 *Let $p(m, s)$ be the fixed joint probability distribution of a watermark and covertext source pair (M, S). For any $D \geq 0$, if $R_{public}^{correlated}(D) > 0$, then the following hold:*

(a) *$p(m, s)$ is publicly admissible with respect to D if*

$$H(M) < R_{public}^{correlated}(D).$$

(b) *Conversely, $p(m, s)$ is not publicly admissible with respect to D if*

$$H(M) > R_{public}^{correlated}(D).$$

Comments:

i) The idea of the proof of Theorem 2.1 is based on the combination of Slepian-Wolf random binning technique [31] for source coding and Gel'fand-Pinsker's random binning technique [16] for channel coding. To be specific, in the decoding part of Gel'fand-Pinsker's random binning technique, in addition to correctly decoding the transmitted message, an auxiliary vector U^n correlated with the covertext S^n is obtained.

This auxiliary vector U^n can then be used as side information for the decoder of the Slepian-Wolf random binning coding scheme. Since the watermark is correlated with U^n through the covertext S^n, some gain in encoding the watermark could be obtained by exploiting this correlation. Details will be described in the following sections.

ii) Since $R_{public}^{correlated}(D) > C_{public}(D)$ in general when M and S are highly correlated, Theorem 2.1 implies that the well-known Shannon separation theorem may not be extended to the current case. Indeed, an example will be given in the next section to show that a watermark with entropy $H(M) > C_{public}(D)$ is still able to be transmitted reliably to the watermark receiver.

iii) It can be shown that (U, M) is a better auxiliary random variable than U. So, as Frans Willems pointed out to me, a question whether $U = (X, M)$ is the optimal choice for the auxiliary random variable remains open. If this would be the case then also a result from semantic coding (Willems and Kalker [41]) could be used to demonstrate admissibility, and this would then lead to the condition

$$H(M) + I(S; X|M) < I(X; Y).$$

What happens when $H(M) = R_{public}^{correlated}(D)$? In the next section we will show that as a function of D, $R_{public}^{correlated}(D)$ is concave and strictly increasing over $[0, D_{max})$, where D_{max} is the minimum D such that $R_{public}^{correlated}(D) = R_{public}^{correlated}(d_{max})$. In view of this, we have the following stronger result:

Corollary 2.1 *For any $D \in [0, D_{\max})$, if $R_{public}^{correlated}(D) > 0$, then $p(m, s)$ is publicly admissible with respect to D if and only if*

$$H(M) \;\leq\; R_{public}^{correlated}(D). \tag{2.3}$$

2.3 Evaluation and Examples

In this section we shall first investigate some properties of $R_{public}^{correlated}(D)$, and then present an example of public watermarking system with correlated watermarks and covertexts to demonstrate that transmitting a watermark reliably to the watermark receiver is still possible even when the entropy $H(M)$ of the watermark is strictly above the standard public watermarking capacity $C_{public}(D)$.

Property 2.1 *Let $p(m,s)$ be a fixed joint probability distribution of (M,S). Then*

$$R_{public}^{correlated}(D) = \max_{p(u,x|m,s):\mathbf{E}d(S,X)\leq D} [I(U;Y) - I(U;M,S) + I(M;U,Y)]$$

where the maximum is taken over all auxiliary random variables (U,X) taking values over $\mathcal{U}\times\mathcal{X}$ with $|\mathcal{U}| \leq |\mathcal{M}||\mathcal{S}||\mathcal{X}|+1$, jointly distributed with (M,S) with the joint probability distribution of (M,S,U,X,Y) given by $p(m,s,u,x,y) = p(m,s)p(u,x|m,s)p(y|x)$, and satisfying $\mathbf{E}d(S,X) \leq D$.

Proof: The proof is standard by using the Caratheodory's theorem, which can be stated as follows: Each point in the convex hull of a set $\mathcal{A}$ in $\mathbb{R}^n$ is in the convex combination of $n+1$ or fewer points of $\mathcal{A}$. Here, we follows the approach of [26].

First, we label elements (m,s,x) of $\mathcal{M}\times\mathcal{S}\times\mathcal{X}$ by $i = 1,...,t \stackrel{\triangle}{=} |\mathcal{M}||\mathcal{S}||\mathcal{X}|$. Let $\mathfrak{P}(\mathcal{M}\times\mathcal{S}\times\mathcal{X})$ be the set of all probability distributions over $\mathcal{M}\times\mathcal{S}\times\mathcal{X}$. Define a functional

$$\begin{aligned} F: \ \mathfrak{P}(\mathcal{M}\times\mathcal{S}\times\mathcal{X}) &\longrightarrow \mathbb{R}^t \\ Q &\longrightarrow (F_1(Q), F_2(Q), ..., F_t(Q)) \end{aligned}$$

where Q is a generic probability distribution over $\mathcal{M}\times\mathcal{S}\times\mathcal{X}$, and for $i = 1,2,...,t-1$

$$F_i(Q) \ = \ Q(m,s,x), \text{ if } i = (m,s,x),$$

and

$$F_t(Q) \ = \ H_Q(Y) - H_Q(M,S) - I_Q(M;Y) + H_Q(M),$$

where all information quantities are determined by $Q(m,s,x)p(y|x)$.

Next, let (U,X) be any random variables taking values over $\mathcal{U}\times\mathcal{X}$, jointly distributed with (M,S) with the joint probability distribution $p(m,s,u,x,y)=p(m,s)p(u,x|m,s)p(y|x)$ and satisfying $\mathbf{E}d(S,X)\leq D$. Then, for each $u\in\mathcal{U}$, $p(m,s,x|u)$ derived from $p(m,s,u,x,y)$ is an element of $\mathfrak{P}(\mathcal{M}\times\mathcal{S}\times\mathcal{X})$, and the set

$$\{F(p(m,s,x|u))|u\in\mathcal{U}\}\subseteq\mathbb{R}^t.$$

By the Caratheodory's theorem, there exist $t+1$ elements $u_i\in\mathcal{U}$ and $t+1$ numbers $\alpha_i\geq 0$ with $\sum_i\alpha_i=1$ such that

$$\sum_u p(u)F(p(m,s,x|u))=\sum_{i=1}^{t+1}\alpha_i F(p(m,s,x|u_i)),$$

that is,

$$\begin{aligned}&\sum_u p(u)p(m,s,x|u)=\sum_{i=1}^{t+1}\alpha_i p(m,s,x|u_i),\forall(m,s,x)\\&H(Y|U)-H(M,S|U)-I(M;Y|U)+H(M)=\\&\quad\sum_{i=1}^{t+1}\alpha_i[H(Y|u_i)-H(M,S|u_i)-I(M;Y|u_i)+H(M)].\end{aligned}$$

Now define a new random variable $U_0\in\{u_1,u_2,...,u_{t+1}\}$ with the joint probability

$$p(m,s,u_i,x,y)=\alpha_i p(m,s,x|u_i)p(y|x).$$

It is easy to check that for this new random variable $\mathbf{E}d(S,X)\leq D$ and $I(U;Y)-I(U;M,S)+I(M;U,Y)=I(U_0;Y)-I(U_0;M,S)+I(M;U_0,Y)$. This finished the proof of Property 2.1. □

Property 2.2 *$R_{public}^{correlated}(D)$ as a function of D is concave and continuous in $[0,\infty)$.*

Proof: First, for any random variables (M,S,U,X,Y), we can write

$$\begin{aligned}I(U;Y)-I(U;M,S)+I(M;U,Y)&=H(Y)-H(Y|U)-H(M,S)+H(M,S|U)\\&\quad+H(M)+H(U,Y)-H(M,U,Y)\\&=H(Y)-H(M,S)+H(M,S|U)+H(M)-H(M,Y|U).\end{aligned}$$

Now for any $D_1, D_2 \geq 0$, let $(M, S, U_i, X_i, Y_i), i = 1, 2$ be random variables achieving $R_{public}^{correlated}(D_i)$. For any $\lambda_1, \lambda_2 \geq 0$ with $\lambda_1 + \lambda_2 = 1$, let $T \in \{1, 2\}$ be a random variable independent of all other random variables with $\lambda_i = \Pr\{T = i\}$. Define new random variables

$$U = (U_T, T), X = X_T, Y = Y_T.$$

Then by the construction of (M, S, U_i, X_i, Y_i) it is easy to check that $\mathbf{E}d(S, X) \leq \lambda_1 D_1 + \lambda_2 D_2$. In view of the definition of $R_{public}^{correlated}(D)$, we then have

$$\begin{aligned}
R_{public}^{correlated}(\lambda_1 D_1 + \lambda_2 D_2) &\geq I(U;Y) - I(U;M,S) + I(M;U,Y) \\
&= H(Y) - H(M,S) + H(M,S|U) + H(M) - H(M,Y|U) \\
&\geq \lambda_1 \left(H(Y_1) - H(M,S) + H(M,S|U_1) + H(M) - H(M,Y_1|U_1)\right) \\
&\quad + \lambda_2 \left(H(Y_2) - H(M,S) + H(M,S|U_2) + H(M) - H(M,Y_2|U_2)\right) \\
&= \lambda_1 R_{public}^{correlated}(D_1) + \lambda_2 R_{public}^{correlated}(D_2)
\end{aligned}$$

where the last inequality follows from the concavity of entropy, that is, $H(Y) \geq \lambda_1 H(Y_1) + \lambda_2 H(Y_2)$. This proves that $R_{public}^{correlated}(D)$ is concave, which in turn implies the continuity of $R_{public}^{correlated}(D)$ in $(0, \infty)$. What remains is to show that $R_{public}^{correlated}(D)$ is continuous at $D = 0$.

In view of its definition, $R_{public}^{correlated}(D)$ is clearly non-decreasing in D. Therefore one has

$$R_{public}^{correlated}(0) \leq \lim_{n \to \infty} R_{public}^{correlated}(D_n) \tag{2.4}$$

for $D_n \downarrow 0$. In view of Property 1, let (M, S, U_n, X_n, Y_n), $n = 1, 2, \cdots$, denote a random vector achieving $R_{public}^{correlated}(D_n)$ and satisfying $\mathbf{E}d(S, X_n) \leq D_n$ with U_n taking values in an alphabet, say, $\mathcal{U} = \{1, 2, \cdots, |\mathcal{M}||\mathcal{S}||\mathcal{X}| + 1\}$. Consider a subsequence $\{(M, S, U_{n_i}, X_{n_i}, Y_{n_i})\}$ which converges in distribution to, say, $\{(M, S, U, X, Y)\}$. Since $D_{n_i} \to 0$, we have

$\mathbf{E}d(S,X) = \lim_{n_i\to\infty} \mathbf{E}d(S,X_{n_i}) = 0$. From the definition of $R_{public}^{correlated}(D)$, it then follows that

$$\begin{aligned} R_{public}^{correlated}(0) &\geq I(U;Y) - I(U;M,S) + I(M;U,Y) \\ &= \lim_{n_i\to\infty} [I(U_{n_i};Y_{n_i}) - I(U_{n_i};M,S) + I(M;U_{n_i},Y_{n_i})] \\ &= \lim_{n_i\to\infty} R_{public}^{correlated}(D_{n_i}) \\ &= \lim_{n\to\infty} R_{public}^{correlated}(D_n). \end{aligned} \tag{2.5}$$

Combination of (2.4) and (2.5) yields the continuity of $R_{public}^{correlated}(D)$ at $D = 0$.

□

Property 2.3 *Define*

$$D_{\max} = \min\left\{D | R_{public}^{correlated}(D) = R_{public}^{correlated}(d_{\max})\right\}.$$

Then $R_{public}^{correlated}(D)$ *as a function of* D *is strictly increasing in* $[0, D_{\max})$.

Proof: Since $R_{public}^{correlated}(D)$ is non-decreasing in D, the concavity of $R_{public}^{correlated}(D)$ guarantees that it is strictly increasing in $[0, D_{\max})$.

□

The following example shows the existence of a public watermarking system with correlated watermarks and covertexts for which transmitting watermarks M^n to the watermark receiver is successful with high probability, although the entropy $H(M)$ is strictly greater than the standard public watermarking capacity $C_{public}(D)$.

Example: Assume all alphabets are binary, that is, $\mathcal{M} = \mathcal{S} = \mathcal{X} = \mathcal{Y} = \{0,1\}$, and the covertext source S is a Bernoulli source with parameter $1/2$. The distortion measure d is the Hamming distance, and the attack channel $p(y|x)$ is a binary symmetric channel with error parameter $p = 0.01$. Let $D = 0.01$. If watermarks and covertexts are independent, Moulin and O'Sullivan [26] computed its public watermarking capacity $C_{public}(D) = 0.029$

nats/channel use, and showed that the optimal random variables $U \in \{0,1,2\}, X$ achieving the public watermarking capacity is determined by the joint probability distribution $p(s,u,x) = p(s)p_{U,X|S}(u,x|s)$, where $p_{U,X|S}(u,x|s)$ is given by

$$p_{U,X|S}(u=1,x=0|s=0) = 0.82;$$
$$p_{U,X|S}(u=2,x=0|s=0) = 0.17;$$
$$p_{U,X|S}(u=0,x=1|s=0) = 0.01;$$
$$p_{U,X|S}(u=2,x=0|s=1) = 0.01;$$
$$p_{U,X|S}(u=0,x=1|s=1) = 0.17;$$
$$p_{U,X|S}(u=1,x=1|s=1) = 0.82;$$

and all other conditional probabilities are zero.

Now we assume that the watermarking source M is binary and correlated with the covertext source S with a joint probability $p_{M,S}(m,s)$ given by

$$\begin{aligned} p_{M|S}(0|0) &= 0.996 \\ p_{M|S}(1|1) &= 0.998. \end{aligned}$$

Let U, X be the random variables as above, which are conditionally independent of M given S. Then it is not hard to see that $M \to S \to (U,X) \to Y$ forms a Markov chain in the indicated order, and

$$\begin{aligned} I(M;U,Y) - H(M) + I(U;Y) - I(U;M,S) &= I(M;U,Y) - H(M) + I(U;Y) - I(U;S) \\ &= I(M;U,Y) - H(M) + C_{public}(0.01) \\ &= 0.008 > 0, \end{aligned}$$

which in turns implies $H(M) < R_{public}^{correlated}(D)$. By Theorem 2.1, $p(m,s)$ is publicly admissible with respect to $D = 0.01$. On the other hand, $H(M) = 0.693 > C_{public}(D) = 0.029$. Thus, we can conclude that the watermark M can be transmitted reliably to the watermark decoder even though the entropy $H(M)$ is strictly above the standard public watermarking capacity.

2.4 Application to the Gel'fand and Pinsker's Channel

In this section we shall apply our techniques to the combined source and channel coding problem when the channel is Gel'fand and Pinsker's channel and the source to be transmitted is correlated with the channel state information available only to the channel encoder, and establish a combined source coding and Gel'fand and Pinsker channel coding theorem. An example is calculated to demonstrate the gain of information rate obtained by the correlation of the channel state source and the information message source.

It should be mentioned that the model considered in this section is different from that of [24], in which the message is independent of the state information of the Gel'fand and Pinsker's channel and the Gel'fand and Pinsker's channel and the Wyner-Ziv channel are separated. As a result, the separation theorem holds for the model in [24] while it does not hold for the model in this section.

To begin with, we review the Gel'fand-Pinsker's channel and the Gel'fand and Pinsker's coding theorem. In their famous paper [16], Gel'fand and Pinsker studied a communication system with channel state information non-causally available only to the transmitter, and determined its channel capacity by giving a coding theorem. To be specific, let $K = \{(p(y|x,s), p(s)) : y \in \mathcal{Y}, x \in \mathcal{X}, s \in \mathcal{S}\}$ be a stationary and memoryless channel with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and the set of channel states $\mathcal{S}$, and let the channel state source S and the message source M be independent. If the state information s^n is only available to the transmitter, then the channel capacity is equal to [16]

$$C_{G-P} = \max_{(U,X)} [I(U;Y) - I(U;S)],$$

where the maximum is taken over all random vectors $(U,X) \in \mathcal{U} \times \mathcal{X}$ such that the joint probability of (U,S,X,Y) is given by $p(u,s,x,y) = p(s)p(u,x|s)p(y|x,s)$. Moreover, $|\mathcal{U}| \leq |\mathcal{S}| + |\mathcal{X}|$.

Note that the independence between the channel state source S and the message source

M is assumed in the Gel'fand-Pinsker's coding theorem. Now we assume the channel state source S and the information message source M are correlated with a joint probability distribution $p(m,s)$ and the state information S^n is uncausally available only to the transmitter, and define

$$R_{G-P} = \max_{(U,X)}[I(U;Y) - I(U;M,S) + I(M;U,Y)]$$

where the maximum is taken over all random variables $(U,X) \in \mathcal{U} \times \mathcal{X}$ such that the joint probability of (M,S,U,X,Y) is given by

$$p(m,s,u,x,y) = p(m,s)p(u,x|m,s)p(y|x,s),$$

and $|\mathcal{U}| \leq |\mathcal{S}||\mathcal{M}| + |\mathcal{X}|$.

If the public admissibility of $p(m,s)$ is defined in a similar manner, then we have the following combined source coding and Gel'fand and Pinsker channel coding theorem.

Theorem 2.2 *If $R_{G-P} > 0$, then the following hold:*

(a) *$p(m,s)$ is publicly admissible if*

$$H(M) < R_{G-P}.$$

(b) *Conversely, $p(m,s)$ is not publicly admissible if*

$$H(M) > R_{G-P}.$$

The proof is similar to that of Theorem 2.1, so omitted here. Note that this theorem is weaker than Corollary 2.1, since we don't know what will happen for $p(m,s)$ if $H(M) = R_{G-P}$.

It is not hard to see that in general, $R_{G-P} > C_{G-P}$ when M and S are highly correlated. In the following example, we will further show the existence of a correlated message source and channel state information source, for which the message source can be transmitted

to the receiver reliably, even though $H(M)$ is strictly greater than the standard Gel'fand-Pinsker's channel capacity C_{G-P}.

Example [16] (revisited): The channel input alphabet and the output alphabet are $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, and the channel state alphabet $\mathcal{S} = \{0, 1, 2\}$. Given three parameters $0 \leq \lambda, p, q \leq 1/2$, the channel K is described in the following:

1. $p_S(0) = p_S(1) = \lambda, p_S(2) = 1 - 2\lambda$;

2. $p_{Y|XS}(y = 0|x = 0, s = 0) = p_{Y|XS}(y = 0|x = 1, s = 0) = 1 - q$,

   $p_{Y|XS}(y = 0|x = 0, s = 1) = p_{Y|XS}(y = 0|x = 1, s = 1) = q$,

   $p_{Y|XS}(y = 0|x = 0, s = 2) = 1 - p$, $p_{Y|XS}(y = 1|x = 0, s = 2) = p$.

Gel'fand and Pinsker got the capacity of K as

$$C_{G-P} = 1 - 2\lambda + 2\lambda h(\alpha_0) - h(\rho(\alpha_0)),$$

where

$$\begin{aligned} h(x) &= -x \log_2(x) - (1 - x) \log_2(1 - x), \\ \rho(\alpha) &= 2\lambda[\alpha(1 - q) + (1 - \alpha)q] + (1 - 2\lambda)(1 - p), \end{aligned} \tag{2.6}$$

and $0 \leq \alpha_0 \leq 1$ is the unique solution of the equation

$$\log_2 \frac{1 - \alpha}{\alpha} = (1 - 2q) \log_2 \frac{1 - \rho(\alpha)}{\rho(\alpha)}. \tag{2.7}$$

Now suppose the information message source M is binary and correlated with the channel state information source S by a joint probability distribution $p(m, s)$ given by

$$\begin{aligned} p_{M|S}(0|0) &= \alpha \\ p_{M|S}(0|1) &= \beta \\ p_{M|S}(0|2) &= \gamma. \end{aligned}$$

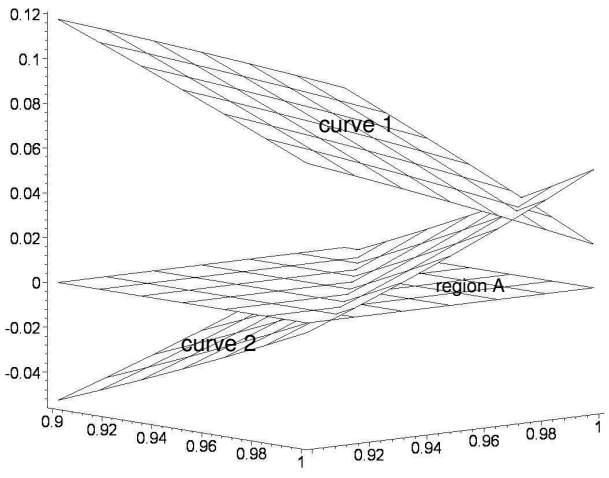

Figure 2.3: The region of (α, β)

Let U, X be the binary random variables achieving the channel capacity C_{G-P}, as described in [16], which are conditionally independent of M given S. Then $M \to (S, U, X) \to Y$ also forms a Markov chain. If (α, β, γ) satisfies

$$\begin{cases} H(M) - C_{G-P} & > & 0, \\ I(M; U, Y) - H(M) + C_{G-P} & > & 0, \end{cases} \tag{2.8}$$

then, $H(M) < R_{G-P}$, and by Theorem 2.2 the message source M can be transmitted reliably to the receiver, even though $H(M)$ is strictly greater than the Gel'fand-Pinsker's channel capacity C_{G-P}. Now we give some numerical solutions. Let $q = 0.2$, $p = 0.1$, $\lambda = 0.2$; in this case, $C_{G-P} = 0.2075$. Let $\gamma = 0.9$. Figure 2.3 shows that any point (α, β) in the region A of Figure 2.3 satisfies (2.8), where the curve 1 represents $f_1(\alpha, \beta) = H(M) - C_{G-P}$ and the curve 2 represents $f_2(\alpha, \beta) = I(M; U, Y) - H(M) + C_{G-P}$. For example, when $\alpha = \beta = 0.98$, we have $H(M) = 0.2484$ and $I(M; U, Y) + C_{G-P} - H(M) = 0.028 > 0$; thus, $H(M) < R_{G-P}$, which means that M can be transmitted reliably to the receiver, while $H(M) > C_{G-P} = 0.2075$.

2.5 Solution to the Cox's Open Problem

In [11] Cox et al proposed an open problem on how to efficiently hide watermarks into correlated covertexts, which can be stated formally as follows. Let $\mathcal{M}, \mathcal{S}, \mathcal{X}, \mathcal{Y}$ be finite alphabets, $p(m)$ a fixed probability distribution, and

$$\mathcal{P} = \left\{ p^{(i)}(m,s) | \sum_s p^{(i)}(m,s) = p(m), i = 1, 2, ..., t \right\}$$

a finite set of joint probability distributions with the fixed marginal probability $p(m)$. Let $(M, S^{(i)})$ denote an identically and independently distributed (iid) watermark and covertext source pair generated according to the probability distribution $p^{(i)}(m,s) \in \mathcal{P}$, with M serving as a watermark to be transmitted and $S^{(i)}$ serving as a covertext available only to the watermark transmitter, and let D be the fixed distortion level between covertexts and stegotexts. Assume that the fixed attack channel $p(y|x)$ with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ is memoryless, stationary and known to both the watermark transmitter and the watermark receiver. Let $e(p^{(i)}(m,s))$ be the least number of bits of information needed to be embedded into $S^{(i)}$ in order for the watermark M to be recovered with high probability after watermark decoding in the case of public watermarking if $S^{(i)}$ is chosen as a covertext. The open problem proposed by Cox et al in [11] can be reformulated as how to choose the optimal joint probability distribution $p^{(i_0)}(m,s)$ in $\mathcal{P}$ achieving $\min_{p^{(i)}(m,s) \in \mathcal{P}} e(p^{(i)}(m,s))$. With the help of Theorem 2.1 and Corollary 2.1, we are ready to solve this problem; our solution is given below in Theorem 2.3. Note that in this section, public admissibility means public admissibility with respect to D, and to emphasize on the dependence of $R_{public}^{correlated}$ on $p(m,s)$ we write $R_{public}^{correlated}(p(m,s))$ rather than $R_{public}^{correlated}(D)$.

Theorem 2.3 *Let $\mathcal{P}_1$ be the set of all publicly admissible joint probability distributions $p^{(i)}(m,s) \in \mathcal{P}$, that is,*

$$\mathcal{P}_1 = \{ p^{(i)}(m,s) \in \mathcal{P} : H(M) \leq R_{public}^{correlated}(p^{(i)}(m,s)) \}.$$

For each $p^{(i)}(m,s) \in \mathcal{P}_1$, let $\mathcal{A}(p^{(i)}(m,s))$ denote the set of all auxiliary random variables (U,X) jointly distributed with M, $S^{(i)}$, and Y with the joint probability distribution given by $p(m,s,u,x,y) = p^{(i)}(m,s)p(u,x|m,s)p(y|x)$ and satisfying

$$\mathbf{E}d(S^{(i)},X) \leq D, H(M) \leq I(U;Y) - I(U;M,S^{(i))} + I(M;U,Y)$$

where $p(u,x|m,s)$ is the conditional probability distribution of (U,X) given $(M,S^{(i)})$. Then, for each $p^{(i)}(m,s) \in \mathcal{P}_1$

$$e(p^{(i)}(m,s)) = \min_{(U,X)\in\mathcal{A}(p^{(i)}(m,s))} (H(M) - I(M;U,Y))$$

and the optimal joint probability $p^{(i_0)}(m,s)$ is given by the probability distribution achieving

$$\max_{p^{(i)}(m,s)\in\mathcal{P}_1} \max_{(U,X)\in\mathcal{A}(p^{(i)}(m,s))} I(M;U,Y).$$

Proof: In view of the definition of $\mathcal{P}_1$, it is easy to see that for each $p^{(i)}(m,s) \in \mathcal{P}_1$, the set $\mathcal{A}(p^{(i)}(m,s))$ is not empty. So,

$$\min_{(U,X)\in\mathcal{A}(p^{(i)}(m,s))} [H(M) - I(M;U,Y)]$$

is meaningful.

From the proof of Theorem 2.1 and Corollary 2.1, we know that $H(M) - I(M;U,Y)$ is the least number of bits of information needed to be embedded into $S^{(i)}$ for a fixed pair $(U,X) \in \mathcal{A}(p^{(i)}(m,s))$. Therefore

$$\min_{(U,X)\in\mathcal{A}(p^{(i)}(m,s))} (H(M) - I(M;U,Y)) \tag{2.9}$$

is the least number of bits of information needed to be embedded for a $p^{(i)}(m,s) \in \mathcal{P}_1$.

Finally, minimizing (2.9) over $\mathcal{P}_1$ yields the theorem since $H(M)$ is fixed. This completes the proof of Theorem 2.3.

In the following, an algorithm is developed to find the optimal publicly admissible joint probability described in the above theorem.

Algorithm:

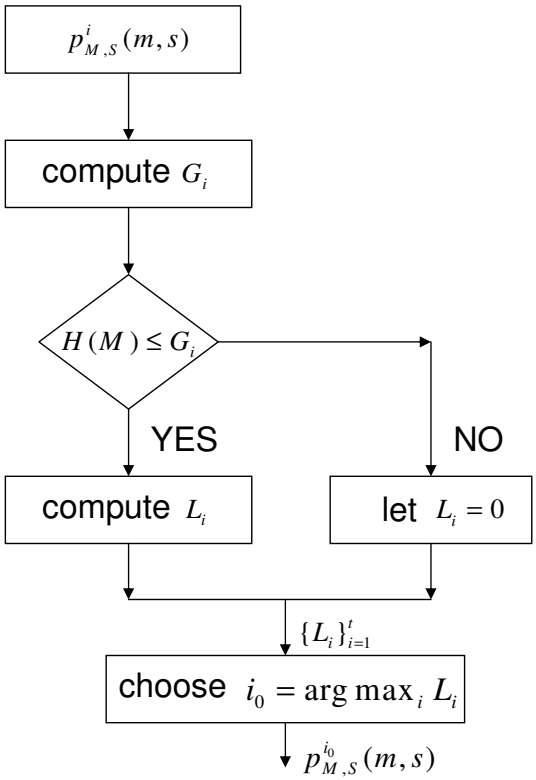

Figure 2.4: Algorithm for optimal solution to Cox's problem

Step 1 For i and $p^{(i)}(m,s) \in \mathcal{P}$, compute

$$G_i \stackrel{\triangle}{=} \max_{(U,X):\mathbf{E}d(S,X)\le D} [I(U;Y) - I(U;M,S) + I(M;U,Y)];$$

Step 2 If $H(M) \le G_i$, compute

$$L_i \stackrel{\triangle}{=} \max_{(U,X)\in\mathcal{A}(p^{(i)}(m,s))} I(M;U,Y);$$

Step 3 If $H(M) > G_i$, let $L_i = -1$;

Step 4 Let $i = i + 1$, and repeat Step 1-3;

Step 5 Let $i_0 = \arg\max_i L_i$, then $p^{(i_0)}(m,s)$ is the optimal solution.

The algorithm is designated in Figure 2.4. It shall be noted that G_i and L_i for each i can be calculated by employing numerical algorithms similar to Blahut-Arimoto algorithms [2, 1] and Willems's algorithm [42].

2.6 Proof of Direct Part

2.6.1 Preliminaries on Typicality

Typicality is an important tool in proving coding theorems, and has been studied extensively in the literature [3, 12, 9]. This section will review the definition of typicality and some basic properties needed in the following proofs.

Definition 2.7 *Let X be a random variable drawn from a finite alphabet $\mathcal{X}$ with probability distribution $p(x)$. A sequence $x^n \in \mathcal{X}^n$ is said to be ϵ-typical with respect to $p(x)$ if for all $a \in \mathcal{X}$,*

$$\left| \frac{N(a|x^n)}{n} - p(a) \right| \leq \epsilon,$$

and $N(a|x^n) = 0$ whenever $p(a) = 0$, where $N(a|x^n)$ is the number of occurrences of the symbol $a \in \mathcal{X}$ in the sequence x^n.

Definition 2.8 *Let (X, Y) be a random vector drawn from a finite alphabet $\mathcal{X} \times \mathcal{Y}$ with joint probability distribution $p(x, y)$. A pair of sequences $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ is said to be jointly ϵ-typical with respect to $p(x, y)$ if for all $(a, b) \in \mathcal{X} \times \mathcal{Y}$,*

$$\left| \frac{N(a, b|x^n, y^n)}{n} - p(a, b) \right| \leq \epsilon,$$

and $N(a, b|x^n, y^n) = 0$ whenever $p(a, b) = 0$, where $N(a, b|x^n, y^n)$ is the number of occurrences of the pair $(a, b) \in \mathcal{X} \times \mathcal{Y}$ in the pair of sequences (x^n, y^n).

Conditional typicality can be defined in a similar manner. Obviously, if (x^n, y^n) is jointly ϵ-typical with respect to $p(x, y)$, then x^n and y^n are also typical with respect to

the marginal probability mass functions $p(x)$ and $p(y)$ respectively. The set of all ϵ-typical sequences $x^n \in \mathcal{X}^n$ with respect to $p(x)$ is denoted by $A_\epsilon^{(n)}(X)$, and the set of all jointly ϵ-typical sequences $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$ with respect to $p(x, y)$ is denoted by $A_\epsilon^{(n)}(X, Y)$.

Lemma 2.1 *Let X_i, $i = 1, 2, \ldots, n$, be drawn independently and identically according to $p(x)$. Then for any given $\epsilon > 0$*

$$\Pr\{X^n \in A_\epsilon^{(n)}(X)\} \geq 1 - \epsilon,$$

for sufficiently large n.

Lemma 2.2 *Let Y_i, $i = 1, 2, \ldots, n$, be drawn independently and identically according to the marginal probability distribution $p(y)$ of $p(x, y)$. For $x^n \in A_\epsilon^{(n)}(X)$ with respect to the marginal probability distribution $p(x)$ of $p(x, y)$, denote $A_\epsilon^{(n)}(x^n, Y) = \{y^n \in \mathcal{Y}^n | (x^n, y^n) \in A_\epsilon^{(n)}(X, Y)\}$, the set of all y^n typical jointly with x^n with respect to $p(x, y)$. Then the following hold:*

(a) For sufficiently small ϵ, sufficiently large n, and any $x^n \in A_\epsilon^{(n)}(X)$

$$2^{-n(I(X;Y)+\alpha\epsilon)} \leq \Pr\{Y^n \in A_{\alpha\epsilon}^{(n)}(x^n, Y)\} \leq 2^{-n(I(X;Y)-\alpha\epsilon)};$$

where α is a constant depending only on the joint probability distribution $p(x, y)$ and the sizes of $\mathcal{X}$ and $\mathcal{Y}$.

(b) For any $\epsilon > 0$ and sufficiently large n,

$$\Pr\{Y^n \in A_\epsilon^{(n)}(x^n, Y) | x^n \in A_\epsilon^{(n)}(X)\} \geq 1 - \epsilon.$$

Lemma 2.3 ***(Markov Lemma):*** *Suppose $X \to Y \to Z$ and (X^n, Y^n) is generated identically and independently according to $p(x, y)$. Then for sufficiently small $\epsilon > 0$ and sufficiently large n.*

$$\Pr\{X^n \in A_{\alpha\epsilon}^{(n)}(z^n, X) | Y^n \in A_\epsilon^{(n)}(z^n, Y)\} \geq 1 - \epsilon$$

where $\alpha > 0$ is a constant depending only on the sizes of $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$.

2.6.2 Watermarking Coding Scheme

We now prove the direct part of the main theorem. Specifically, we will show that if

$$H(M) < R_{public}^{correlated}(D),$$

then $p(m, s)$ is publicly admissible with respect to D.

Let (M, S, U, X, Y) be the random vector with the joint probability distribution $p(m, s, u, x, y) = p(m, s)p(u, x|m, s)p(y|x)$ achieving the maximum in $R_{public}^{correlated}(D)$, that is, $\mathbf{E}d(S, X) \leq D$, and

$$R_{public}^{correlated}(D) = I(U; Y) - I(U; M, S) + I(M; U, Y).$$

Denote $\gamma \stackrel{\triangle}{=} I(U; Y) - I(U; M, S) + I(M; U, Y) - H(M) > 0$. Let $\epsilon > 0$ be an arbitrarily small but fixed number. We will show the existence of watermarking encoder and decoder pairs (f_n, g_n) for all sufficiently large n such that $\mathbf{E}d(S^n, f_n(M^n, S^n)) < D + \epsilon$ and $p_e(f_n, g_n) < \epsilon$. Note that both the watermark transmitter and the receiver know the attack channel $p(y|x)$.

Random Codes Generation: Two random codebooks C and W will be generated as follows.

- First, generate identically and independently $\exp(n[H(M) + \gamma/8])$ vectors $m^n \in \mathcal{M}^n$ according to the probability $p(m)$, and then uniformly distribute all these vectors into $t \stackrel{\triangle}{=} \exp(n[H(M) - I(M; U, Y) + \gamma/4])$ bins, each bin $C(i), i = 1, 2, ..., t$ containing $\exp(n[I(M; U, Y) - \gamma/8])$ vectors. Denote this random codebook by $C = \{C(i)\}_{i=1}^{t}$.

- Second, for each index $i = 1, ..., t$, generate a bin of vectors $W(i) = \{u^n(i, j) \in \mathcal{U}^n | j = 1, 2, ..., \exp(n[I(U; M, S) + \gamma/4])\}$, each vector $u^n(i, j)$ is generated identically and independently according to the probability $p(u)$ derived from the joint probability $p(m, s, u, x, y)$. Denote this random codebook by $W = \{W(i)\}_{i=1}^{t}$.

- The two codebooks C and W are then distributed to the watermarking decoder.

Watermarking encoding: Fix codebooks C, W. Given a watermark m^n and a covertext s^n

- if (m^n, s^n) is not jointly ϵ-typical, then an encoding error is declared;
- if (m^n, s^n) is jointly ϵ-typical, but no $C(i)$ contains m^n, $i = 1, 2, ...t$, then an encoding error is declared;
- if (m^n, s^n) is jointly ϵ-typical and $C(i)$ is the first bin in C containing m^n, but no vector $u^n \in W(i)$ such that (m^n, s^n, u^n) is jointly ϵ-typical, then an encoding error is declared;
- if (m^n, s^n) is jointly ϵ-typical, $C(i)$ is the first bin in C containing m^n, and $u^n(i, j) \in W(i)$ is the first vector in $W(i)$ such that $(m^n, s^n, u^n(i, j))$ is jointly ϵ-typical, then the encoder randomly generates a stegotext x^n according to $p(x^n|m^n, s^n, u^n(i, j))$; and finally
- if an encoding error is declared, then define a fixed x_0^n as the stegotext.

Watermarking decoding: Fix codebooks C, W. Let y^n be a forgery received by the watermarking decoder when $m^n \in C(i)$ is transmitted using s^n and $u^n(i, j) \in W(i)$.

- The decoder finds a vector in the codebook W, say $u^n(i_0, j_0) \in W(i_0)$, such that $(u^n(i_0, j_0), y^n)$ is jointly ϵ-typical with respect to $p(u, y)$.
- If no or more than one $u^n(i_0, j_0)$ are found in W such that $(u^n(i_0, j_0), y^n)$ is jointly ϵ-typical, then a decoding error is declared.
- The decoder finds a vector $\hat{m}^n \in C(i_0)$ such that $(\hat{m}^n, u^n(i_0, j_0), y^n)$ is jointly ϵ-typical with respect to $p(m, u, y)$.
- If no or more than one such $\hat{m}^n$ are found in the bin $C(i_0)$, then a decoding error is also declared.

- The decoder decodes $\hat{m}^n$ to be the watermark.

Note that in view of Lemmas 2.2 and 2.3, joint ϵ-typicality in both watermarking encoding and decoding should be understood with ϵ being replaced by ϵ multiplied by a proper constant whenever necessary.

2.6.3 Analysis on Averaged Error Probability

We shall bound the error probability $\mathbf{E}_{C,W} p_e(C, W)$ averaged over all random codebooks, watermarks and covertexts. Obviously, from the encoding scheme described above there are the following encoding error events:

- E_0: (m^n, s^n) is not jointly ϵ-typical;
- E_1: $(m^n, s^n) \in \bar{E}_0$, but $m^n \notin C$, where $\bar{E}_0$ denotes the complement of E_0;
- E_2: $(m^n, s^n) \notin E_0 \cup E_1$, but no $u^n \in W(i)$ such that (m^n, s^n, u^n) is ϵ-typical, where $i = i(m^n)$ is the smallest i such that $C(i)$ contains m^n; and
- E_3: $(m^n, s^n) \notin E_0 \cup E_1 \cup E_2$, $u^n(i, j)$ is the first vector in $W(i)$ such that $(m^n, s^n, u^n(i, j))$ is ϵ-typical—such j will be denoted by $j = j(m^n, s^n)$—but $(m^n, s^n, u^n(i, j), X^n)$ is not ϵ-typical, where $i = i(m^n)$, and X^n is generated according to $p(x^n|m^n, s^n, u^n(i, j))$.

Suppose now that encoding (m^n, s^n) is successful via C, W and the stegotext x^n is generated accordingly from $(m^n, s^n, u^n(i, j))$. Let y^n be a forgery generated by the attacker. Then, there are the following decoding error events:

- E: $(m^n, u^n(i, j), y^n)$ is not jointly ϵ-typical;
- E': more than one $u^n \in W$ such that (u^n, y^n) is ϵ-jointly typical; and
- E'': more that one $\hat{m}^n \in C(i)$ such that $(\hat{m}^n, u^n(i, j), y^n)$ is jointly ϵ-typical.

In the following, we will develop upper bounds for probabilities of these error events.

- By Lemma 2.1 there exists a large number n_0 such that for all $n > n_0$,

$$\Pr\{E_0\} \leq \epsilon. \tag{2.10}$$

- If m^n is ϵ-typical and ϵ is sufficiently small, then $p(m^n) > 2^{-n[H(M)+\gamma/16]}$ for sufficiently large n. So, there exists a large number n_1 such that for all $n > n_1$,

$$\begin{aligned}\Pr\{E_1\} &\leq \Pr\{E_1|\bar{E}_0\} \\ &\leq (1-2^{-n[H(M)+\gamma/16]})^{2^{n[H(M)-I(M;U,Y)+\gamma/4]}2^{n[I(M;U,Y)-\gamma/8]}} \\ &\leq 2^{-2^{n\gamma/16}} \leq \epsilon. \end{aligned} \tag{2.11}$$

- Given $(m^n, s^n) \notin E_0 \cup E_1$, let $i = i(m^n)$ be the smallest i such that $C(i)$ contains m^n. By the generation of $W(i)$,

$$\Pr\{(m^n, s^n, u^n) \text{ is jointly } \epsilon\text{-typical}|(m^n, s^n)\} > 2^{-n[I(U;M,S)+\gamma/8]}$$

for large n. Thus, there exists a large number n_2 such that for all $n > n_2$ and $(m^n, s^n) \notin E_0 \cup E_1$

$$\begin{aligned}\Pr\{E_2|(m^n, s^n)\} &= \Pr\{\text{no } u^n \in W(i) \text{ such that } (m^n, s^n, u^n) \text{ is jointly } \epsilon\text{-typical}\} \\ &\leq (1-2^{-n[I(U;M,S)+\gamma/8]})^{2^{n[I(U;M,S)+\gamma/4]}} \\ &\leq 2^{-2^{n\gamma/8}} \leq \epsilon \end{aligned}$$

which implies

$$\Pr\{E_2\} \leq \epsilon. \tag{2.12}$$

- Since X^n is generated according to $p(x^n|m^n, s^n, u^n(i,j))$, where $u^n(i,j)$ is the first vector in $W(i)$ such that $(m^n, s^n, u^n(i,j))$ is jointly typical, it follows from Lemma 2.2-(b) that there exists a large number n_3 such that for all $n > n_3$

$$\Pr\{E_3\} \leq \Pr\{E_3|\bar{E}_2, \bar{E}_1, \bar{E}_0\} \leq \epsilon. \tag{2.13}$$

Assume now that embedding m^n into s^n is successful via C, W and $u^n(i,j)$, resulting in a stegotext x^n, and y^n is a forgery generated by the attacker with the attack channel input x^n. We shall upper bound the error probability of watermark decoding $\Pr\{\hat{M}^n \neq M^n | m^n, s^n, u^n(i,j)\}$.

To begin with, one has

$$\begin{aligned}\Pr\{\hat{M}^n \neq M^n | m^n, s^n, u^n(i,j)\} &\leq \Pr\{E \cup E' \cup E'' | m^n, s^n, u^n(i,j)\} \\ &\leq \Pr\{E | m^n, s^n, u^n(i,j)\} + \Pr\{E' \cap \bar{E} | m^n, s^n, u^n(i,j)\} \\ &\quad + \Pr\{E'' \cap \bar{E} | m^n, s^n, u^n(i,j)\}. \end{aligned} \tag{2.14}$$

- By the Markov Lemma, $(m^n, s^n, u^n(i,j), x^n, y^n)$ is jointly typical with high probability for large n, so is $(m^n, u^n(i,j), y^n)$. Thus, there exists a large number n_4 such that for all $n > n_4$

$$\Pr\{E | m^n, s^n, u^n(i,j)\} \leq \epsilon. \tag{2.15}$$

- In light of the definition of E', one has

$$\begin{aligned}&\Pr\{E' \cap \bar{E} | m^n, s^n, u^n(i,j)\} \\ &\leq \Pr\{(u^n, y^n) \text{ is jointly } \epsilon\text{-typical for some } u^n \neq u^n(i,j), y^n \in A_\epsilon^{(n)}(Y) | m^n, s^n, u^n(i,j)\} \\ &= \sum_{y^n \in A_\epsilon^{(n)}(Y)} p(y^n | m^n, s^n, u^n(i,j)) \theta(m^n, s^n, u^n(i,j), y^n), \end{aligned} \tag{2.16}$$

where

$$\begin{aligned}&\theta(m^n, s^n, u^n(i,j), y^n) \\ &\triangleq \Pr\{(u^n, y^n) \text{ is jointly } \epsilon\text{-typical for some } u^n \neq u^n(i,j) | m^n, s^n, u^n(i,j), y^n\} \\ &\leq \sum_{u^n \in W(i'), i'=1,2,\ldots t, i' \neq i} \Pr\{(u^n, y^n) \text{ is jointly } \epsilon\text{-typical} | m^n, s^n, u^n(i,j), y^n\} \\ &\quad + \Pr\{(u^n, y^n) \text{ is jointly } \epsilon\text{-typical, } u^n \in W(i) \text{ but } u^n \neq u^n(i,j) | m^n, s^n, u^n(i,j), y^n\}. \end{aligned} \tag{2.17}$$

By the generation of W, if $u^n \in W(i')$ and $i' \neq i$, then u^n is independent of $m^n, s^n, u^n(i,j), y^n$ and hence for sufficiently large n

$$\Pr\{(u^n, y^n) \text{ is jointly } \epsilon\text{-typical} | m^n, s^n, u^n(i,j), y^n\} < 2^{-n[I(U;Y)-\gamma/4]}. \tag{2.18}$$

Therefore, there exists a large number n_5 such that for all $n > n_5$, the summation in the right side of (2.17) is less or equal to

$$\begin{aligned} & 2^{n[H(M)-I(M;U,Y)+\gamma/4]} 2^{-n[I(U;Y)-\gamma/4]} 2^{n[I(U;M,S)+\gamma/4]} \\ = \; & 2^{-n[I(U;Y)-I(U;M,S)+I(M;U,Y)-H(M)-3\gamma/4]} \\ = \; & 2^{-n\gamma/4} \leq \frac{\epsilon}{2}, \end{aligned} \tag{2.19}$$

where the equalities of (2.19) are due to the random codebooks generation and the definition of γ. To upper bound the second term in the right side of (2.17), one has

$$\begin{aligned} & \Pr\{(u^n, y^n) \text{ is jointly } \epsilon\text{-typical}, u^n \in W(i) \text{ but } u^n \neq u^n(i,j) | m^n, s^n, u^n(i,j), y^n\} \\ = \; & \sum_{l=1}^{|W(i)|} \Pr\{(u^n, y^n) \text{ is typical}, u^n \in W(i) \text{ but } u^n \neq u^n(i,j), j = l | m^n, s^n, u^n(i,j), y^n\} \\ \leq \; & \sum_{l=1}^{|W(i)|} \sum_{k=1, k\neq l}^{|W(i)|} \Pr\{(u^n_{(k)}, y^n) \text{ is jointly typical }, j = l | m^n, s^n, u^n(i,j), y^n\} \\ = \; & \sum_{l=1}^{|W(i)|} \left[\sum_{k=l+1}^{|W(i)|} \Pr\{(u^n_{(k)}, y^n) \text{ is jointly typical }, j = l | m^n, s^n, u^n(i,j), y^n\} \right. \\ & \left. + \sum_{k=1}^{l-1} \Pr\{(u^n_{(k)}, y^n) \text{ is jointly typical }, j = l | m^n, s^n, u^n(i,j), y^n\} \right] \\ \overset{(2)}{\leq} \; & \sum_{l=1}^{|W(i)|} \left[2^{-n[I(U;Y)-\gamma/4]} \sum_{k=l+1}^{|W(i)|} \Pr\{j = l | m^n, s^n, u^n(i,j), y^n\} \right. \\ & \left. + \sum_{k=1}^{l-1} \Pr\{(u^n_{(k)}, y^n) \text{ is typical }, j = l | m^n, s^n, u^n(i,j), y^n\} \right], \end{aligned} \tag{2.20}$$

where (2) is from (2.18). Continuing upper bounding (2.20) yields

$$
\begin{aligned}
&\Pr\{(u^n, y^n) \text{ is jointly } \epsilon\text{-typical}, u^n \in W(i) \text{ but } u^n \neq u^n(i,j) | m^n, s^n, u^n(i,j), y^n\} \\
&\overset{(3)}{\leq} \sum_{l=1}^{|W(i)|} \Big[2^{-n[I(U;Y)-\gamma/4]} (|W(i)| - l) \Pr\{j = l | m^n, s^n, u^n(i,j), y^n\} \\
&\quad + \sum_{k=1}^{l-1} \Pr\{(u_{(k)}^n, y^n) \text{ is typical }, (u_{(r)}^n, m^n, s^n) \text{ is not typical }, \\
&\quad r = 1, 2, ..., l-1, r \neq k, (u_{(l)}^n, m^n, s^n) \text{ is typical } | m^n, s^n, u^n(i,j), y^n\} \Big] \\
&\leq \sum_{l=1}^{|W(i)|} 2^{-n[I(U;Y)-\gamma/4]} \Big[2^{n[I(U;M,S)+\gamma/4]} \Pr\{j = l | m^n, s^n, u^n(i,j), y^n\} \\
&\quad + \sum_{k=1}^{l-1} \frac{\Pr\{(u_{(r)}^n, m^n, s^n) \text{ is not typical }, r < l, (u_{(l)}^n, m^n, s^n) \text{ is typical } | m^n, s^n, u^n(i,j), y^n\}}{\Pr\{(u_{(k)}^n, m^n, s^n) \text{ is not typical } | m^n, s^n\}} \Big] \\
&\overset{(4)}{\leq} \sum_{l=1}^{|W(i)|} 2^{-n[I(U;Y)-\gamma/4]} \Big[2^{n[I(U;M,S)+\gamma/4]} \Pr\{j = l | m^n, s^n, u^n(i,j), y^n\} \\
&\quad + \sum_{k=1}^{l-1} \frac{\Pr\{j = l | m^n, s^n, u^n(i,j), y^n\}}{1 - 2^{-n[I(U;M,S)-\gamma/4]}} \Big] \\
&\leq 2^{-n[I(U;Y)-\gamma/4]} 2^{n[I(U;M,S)+\gamma/4]} \left[1 + \frac{1}{1 - 2^{-n[I(U;M,S)-\gamma/4]}} \right] \\
&\overset{(5)}{\leq} 2^{-n\gamma/2} \left[1 + \frac{1}{1 - 2^{-n[I(U;M,S)-\gamma/4]}} \right] \\
&\leq \epsilon/2.
\end{aligned}
$$

In the above derivation, (3) is due to the definition of j, (4) follows from the fact that, there exists n_6 such that for all $n > n_6$,

$$\Pr\{(u^n, m^n, s^n) \text{ is jointly typical} | m^n, s^n\} < 2^{-n[I(U;M,S)-\gamma/4]},$$

and finally, (5) is attributable to the fact that $\gamma < I(U;Y) - I(U;M,S)$. Putting all inequalities above together, we get that for all $n > \max\{n_5, n_6\}$,

$$\Pr\{E' \cap \bar{E} | m^n, s^n, u^n(i,j)\} < \epsilon. \tag{2.21}$$

- Employing a similar approach, we now upper bound the probability $\Pr\{E'' \cap \bar{E} | m^n, s^n, u^n(i,j)\}$. To this end, first define $r = r(m^n)$ to be the index of m^n in the bin $C(i)$. Note that all vectors before the rth vector in $C(i)$ are not equal to m^n. Then one has

$$\begin{aligned}
&\Pr\{E'' \cap \bar{E} | m^n, s^n, u^n(i,j)\} \leq \Pr\{(\hat{m}^n, u^n(i,j), y^n) \text{ is jointly } \epsilon\text{-typical, } \hat{m}^n \in C(i), \\
&\qquad \text{but } \hat{m}^n \neq m^n, (u^n(i,j), y^n) \in A_\epsilon^{(n)}(U,Y) | m^n, s^n, u^n(i,j)\} \\
&= \sum_{y^n : (u^n(i,j), y^n) \in A_\epsilon^{(n)}(U,Y)} p(y^n | m^n, s^n, u^n(i,j)) \eta(m^n, s^n, u^n(i,j), y^n),
\end{aligned}$$

where

$$\begin{aligned}
&\eta(m^n, s^n, u^n(i,j), y^n) \stackrel{\Delta}{=} \Pr\{(\hat{m}^n, u^n(i,j), y^n) \text{ is jointly } \epsilon\text{-typical, } \hat{m}^n \in C(i), \\
&\qquad \text{but } \hat{m}^n \neq m^n | m^n, s^n, u^n(i,j), y^n\}. \\
&= \sum_{l=1}^{|C(i)|} \Pr\{(\hat{m}^n, u^n(i,j), y^n) \text{ is typical }, \hat{m}^n \in C(i) \text{ but } \hat{m}^n \neq m^n, r = l | m^n, s^n, u^n(i,j), y^n\} \\
&\leq \sum_{l=1}^{|C(i)|} \sum_{k=1, k\neq l}^{|C(i)|} \Pr\{(\hat{m}^n_{(k)}, u^n(i,j), y^n) \text{ is typical }, r = l | m^n, s^n, u^n(i,j), y^n\} \\
&= \sum_{l=1}^{|C(i)|} \Bigg[\sum_{k=l+1}^{|C(i)|} \Pr\{(\hat{m}^n_{(k)}, u^n(i,j), y^n) \text{ is typical }, r = l | m^n, s^n, u^n(i,j), y^n\} \\
&\qquad + \sum_{k=1}^{l-1} \Pr\{(\hat{m}^n_{(k)}, u^n(i,j), y^n) \text{ is typical }, r = l | m^n, s^n, u^n(i,j), y^n\} \Bigg] \\
&\leq \sum_{l=1}^{|C(i)|} \big[2^{-n[I(M;U,Y) - \gamma/16]} |C(i)| \Pr\{r = l | m^n, s^n, u^n(i,j), y^n\} \\
&\qquad + \sum_{k=1}^{l-1} \Pr\{(\hat{m}^n_{(k)}, u^n(i,j), y^n) \text{ is typical }, \hat{m}^n_{(a)} \neq m^n, a = 1, 2, ..., l-1, a \neq k, \\
&\qquad \hat{m}^n_{(l)} = m^n | m^n, s^n, u^n(i,j), y^n\} \big]
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{l=1}^{|C(i)|} 2^{-n[I(M;U,Y)-\gamma/16]} \Big[2^{n[I(M;U,Y)-\gamma/8]} \Pr\{r=l|m^n,s^n,u^n(i,j),y^n\} \\
&\quad + \sum_{k=1}^{l-1} \frac{\Pr\{\hat{m}^n_{(a)} \neq m^n, a<l, \hat{m}^n_{(l)} = m^n | m^n, s^n, u^n(i,j), y^n\}}{\Pr\{\hat{m}^n_{(k)} \neq m^n | m^n\}} \Big] \\
&\leq \sum_{l=1}^{|C(i)|} 2^{-n[I(M;U,Y)-\gamma/16]} \Big[2^{n[I(M;U,Y)-\gamma/8]} \Pr\{r=l|m^n,s^n,u^n(i,j),y^n\} \\
&\quad + \sum_{k=1}^{l-1} \frac{\Pr\{r=l|m^n,s^n,u^n(i,j),y^n\}}{1-2^{-n[H(M)-\gamma/4]}} \Big] \\
&\leq 2^{-n[I(M;U,Y)-\gamma/16]} 2^{n[I(M;U,Y)-\gamma/8]} \left[1 + \frac{1}{1-2^{-n[H(M)-\gamma/4]}}\right] \\
&\leq 2^{-n\gamma/16} \left[1 + \frac{1}{1-2^{-n[H(M)-\gamma/4]}}\right] \\
&\leq \epsilon
\end{aligned}
$$

for all large n. Thus, there exists n_7 such that for all numbers $n > n_7$,

$$\Pr\{E'' \cap \bar{E} | m^n, s^n, u^n(i,j)\} < \epsilon. \tag{2.22}$$

Finally, combining (2.10)-(2.13), (2.14),(2.15),(2.21) and (2.22) together, we get

$$\mathbf{E}_{C,W} p_e(C,W) \leq 7\epsilon, \tag{2.23}$$

for all $n > n' = \max_{i=0,1,\ldots,7}\{n_i\}$

2.6.4 Analysis of Distortion Constraint

Let x_0^n be the fixed stegotext if an encoding error is declared. By the watermark encoding scheme we have

$$
\begin{aligned}
&\mathbf{E}_{C,W}\mathbf{E}_{M^n,S^n}[d(S^n,X^n)] = \mathbf{E}[d(S^n,X^n)] \\
&= \Pr\{\cup_{i=0}^3 E_i\}\mathbf{E}[d(S^n,X^n)|\cup_{i=0}^3 E_i] + \Pr\{\cap_{i=0}^i \bar{E}_i\}\mathbf{E}[d(S^n,X^n)|\cap_{i=0}^3 \bar{E}_i] \\
&\leq \Pr\{\cup_{i=0}^3 E_i\} d_{max} + \Pr\{\cap_{i=0}^i \bar{E}_i\}\mathbf{E}[d(S^n,X^n)|\cap_{i=0}^3 \bar{E}_i] \\
&\leq 4\epsilon d_{max} + \epsilon + D
\end{aligned} \tag{2.24}
$$

where the last inequality follows from the fact that there exists a large number n_8 such that for $n > n_8$, $d(s^n, x^n) \leq D + \epsilon$ for all jointly ϵ-typical sequences (m^n, s^n, u^n, x^n) with respect to $p(m, s, u, x)$ with $\mathbf{E}d(S, X) \leq D$, and from the analysis of error probabilities of encoding in the last subsection.

2.6.5 Existence of Watermarking Encoders and Decoders

By the Markov inequality and (2.23), one has

$$\Pr\{p_e(C, W) \geq \sqrt{7\epsilon}\} \quad \leq \sqrt{7\epsilon}.$$

Let

$$\Gamma = \{(C, W) : p_e(C, W) \leq \sqrt{7\epsilon}\}. \tag{2.25}$$

Then $\Pr\{\Gamma\} \geq 1 - \sqrt{7\epsilon}$.

So, from (2.24) one has

$$\begin{aligned} &\sum_{(C,W)\in\Gamma} \Pr(C, W)\mathbf{E}_{S^n,M^n}\left(d(S^n, X^n)|C, W\right) \\ &\quad \leq \quad \mathbf{E}_{C,W}[\mathbf{E}_{S^n,M^n}\left(d(S^n, X^n)|C, W\right)] \\ &\quad \leq \quad 4\epsilon d_{max} + \epsilon + D. \end{aligned}$$

Thus,

$$\begin{aligned} &\sum_{(C,W)\in\Gamma} \frac{\Pr(C, W)}{\Pr\{\Gamma\}}\mathbf{E}_{S^n,M^n}\left(d(S^n, X^n)|C, W\right) \\ &= \quad \frac{1}{\Pr\{\Gamma\}} \sum_{(C,W)\in\Gamma} \Pr(C, W)\mathbf{E}_{S^n,M^n}\left(d(S^n, X^n)|C, W\right) \\ &\leq \quad \frac{4\epsilon d_{max} + \epsilon + D}{1 - \sqrt{7\epsilon}} = D + \epsilon' \end{aligned} \tag{2.26}$$

where $\epsilon' = \frac{(4d_{max}+1)\epsilon + D\sqrt{7\epsilon}}{1-\sqrt{7\epsilon}}$ goes to 0 as $\epsilon \to 0$.

Combining (2.25) and (2.26) yields the existence of watermarking encoder and watermarking decoder for each $n > \max_{i=0,1,..8}\{n_i\}$ such that the error probability is $\leq \sqrt{7\epsilon}$ and

the averaged distortion is $\leq D + \epsilon'$. Since $\epsilon > 0$ is arbitrary, it follows that the probability distribution $p(m, s)$ is publicly admissible with respect to D. This completes the the proof of the direct part of Theorem 2.1.

2.6.6 Proof of Corollary 2.1

Assume $R_{public}^{correlated}(D) > 0$ for $D \in [0, D_{\max})$. From Theorem 2.1, it suffices to show that $p(m, s)$ is publicly admissible with respect to D if $H(M) = R_{public}^{correlated}(D)$. Indeed, for any sufficiently small $\epsilon > 0$, one has $R_{public}^{correlated}(D + \epsilon) > R_{public}^{correlated}(D) = H(M)$ since $R_{public}^{correlated}(D)$ is strictly increasing in $[0, D_{\max})$ by Property 2.3. Thus, by Theorem 2.1, $p(m, s)$ is publicly admissible with respect to $D + \epsilon$ for any sufficiently small $\epsilon > 0$. This, together with the definition of public admissibility, implies that $p(m, s)$ is also publicly admissible with respect to D.

□

2.7 Proof of the Converse Part

To prove the converse part of Theorem 2.1, it suffices to show that $H(M) \leq R_{public}^{correlated}(D)$ if $p(m, s)$ is admissible with respect to D. Suppose $p(m, s)$ is admissible with respect to D. Then for any $\epsilon > 0$, there exists, for any sufficiently large n, a watermarking encoder and public decoder pair (f_n, g_n) with length n such that

$$\begin{aligned} \mathbf{E}d(S^n, f_n(M^n, S^n)) &\leq D + \epsilon, \\ p_e(f_n, g_n) = \Pr\{g_n(Y^n) \neq M^n\} &< \epsilon, \end{aligned}$$

where Y^n is generated by the attack channel with input $X^n = f_n(M^n, S^n)$. In the following we will show that $H(M) \leq R_{public}^{correlated}(D)$.

We first upper bound $I(M^n; Y^n) - I(M^n; S^n)$. Using an approach similar to [40], we

have

$$
\begin{aligned}
I(M^n;Y^n) - I(M^n;S^n) &= \sum_{i=1}^{n}[I(M^n;Y_i|Y_1^{i-1}) - I(M^n;S_i|S_{i+1}^n)] \\
&= \sum_{i=1}^{n}[H(Y_i|Y_1^{i-1}) - H(Y_i|M^n,Y_1^{i-1}) - H(S_i|S_{i+1}^n) + H(S_i|M^n,S_{i+1}^n)] \\
&= \sum_{i=1}^{n}[H(Y_i|Y_1^{i-1}) - H(Y_i|M^n,Y_1^{i-1},S_{i+1}^n) - I(Y_i;S_{i+1}^n|M^n,Y_1^{i-1}) \\
&\quad -H(S_i|S_{i+1}^n) + H(S_i|M^n,S_{i+1}^n)] \\
&= \sum_{i=1}^{n}[H(Y_i|Y_1^{i-1}) - H(Y_i|M^n,Y_1^{i-1},S_{i+1}^n) - H(S_i|S_{i+1}^n) + H(S_i|M^n,S_{i+1}^n)] \\
&\quad - \sum_{i=1}^{n} I(Y_i;S_{i+1}^n|M^n,Y_1^{i-1}) \\
&\overset{(a)}{=} \sum_{i=1}^{n}[H(Y_i|Y_1^{i-1}) - H(Y_i|M^n,Y_1^{i-1},S_{i+1}^n) - H(S_i|S_{i+1}^n) + H(S_i|M^n,S_{i+1}^n)] \\
&\quad - \sum_{i=1}^{n} I(Y_1^{i-1};S_i|M^n,S_{i+1}^n) \\
&\overset{(b)}{=} \sum_{i=1}^{n}[H(Y_i|Y_1^{i-1}) - H(Y_i|M^n,Y_1^{i-1},S_{i+1}^n) - H(S_i) + H(S_i|M^n,Y_1^{i-1},S_{i+1}^n)] \\
&\overset{(c)}{\leq} \sum_{i=1}^{n}[H(Y_i) - H(Y_i|M^n,Y_1^{i-1},S_{i+1}^n) - H(S_i) + H(S_i|M^n,Y_1^{i-1},S_{i+1}^n)] \\
&= \sum_{i=1}^{n}[I(V_i;Y_i) - I(V_i;S_i)],
\end{aligned}
$$

where $V_i = (M^n, Y_1^{i-1}, S_{i+1}^n)$. In the above, the first three equalities follow from the definition and the chain rule of mutual information, (b) is attributable to the memorylessness of the source S, (c) is due to the fact that conditions cannot increase entropy, and finally

(a) is derived from the following fact

$$\begin{aligned}\sum_{i=1}^{n} I(Y_i; S_{i+1}^n | M^n, Y_1^{i-1}) &= \sum_{i=1}^{n}\sum_{j=i+1}^{n} I(Y_i; S_j | M^n, Y_1^{i-1}, S_{j+1}^n) \\ &= \sum_{j=1}^{n}\sum_{i=1}^{j-1} I(Y_i; S_j | M^n, Y_1^{i-1}, S_{j+1}^n) \\ &= \sum_{j=1}^{n} I(Y_1^{j-1}; S_j | M^n, S_{j+1}^n) \\ &= \sum_{i=1}^{n} I(Y_1^{i-1}; S_i | M^n, S_{i+1}^n).\end{aligned}$$

Now let $T \in \{1, 2, ..., n\}$ be a time-sharing random variable, uniformly distributed and independent of all other random variables. Define $S = S_i$, $M = M_i$, $X = X_i$, $Y = Y_i$, and $V = V_i$ when $T = i$, and let $U = (V, T)$. It is easy to see that the joint distribution of M and S is exactly given by $p(m, s)$, and $(M, S, U) \to X \to Y$ forms a Markov chain with the transition probability from X to Y given by $p(y|x)$. Furthermore, since $d(s^n, x^n) = \frac{1}{n}\sum_{i=1}^{n} d(s_i, x_i)$, it follows that

$$\begin{aligned}\mathbf{E}d(S, X) &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}d(S_i, X_i) \\ &= \mathbf{E}d(S^n, X^n) \le D + \epsilon.\end{aligned}$$

Since $I(T; S) = 0$, it follows that

$$\begin{aligned}I(M^n; Y^n) - I(M^n; S^n) &\le n[I(V; Y|T) - I(V; S|T)] \\ &\le n[I(V, T; Y) - I(V, T; S)] \\ &= n[I(U; Y) - I(U; S)]. \qquad (2.27)\end{aligned}$$

Therefore, we have

$$\begin{aligned}
nH(M|S) &= H(M^n|S^n) \\
&= I(M^n;Y^n|S^n) + H(M^n|Y^n,S^n) \\
&= I(M^n;Y^n|S^n) + H(M^n|Y^n) - I(M^n;S^n|Y^n) \\
&= I(M^n;S^n,Y^n) - I(M^n;S^n) + H(M^n|Y^n) - I(M^n;S^n|Y^n) \\
&= I(M^n;Y^n) - I(M^n;S^n) + H(M^n|Y^n) \\
&\overset{(d)}{\leq} n[I(U;Y) - I(U;S)] + H(M^n|Y^n) \\
&\overset{(e)}{\leq} n[I(U;Y) - I(U;S)] + 1 + np_e(f_n,g_n)\log|\mathcal{M}| \qquad (2.28)
\end{aligned}$$

where (d) follows from inequality (2.27) and (e) is due to the Fano inequality,

$$H(M^n|Y^n) \leq 1 + np_e(f_n,g_n)\log|\mathcal{M}|.$$

On the other hand, one has

$$\begin{aligned}
nI(M;U,Y) &= nI(M;V,T,Y) \\
&\overset{(f)}{=} nI(M;V,Y|T) \\
&= \sum_{i=1}^{n} I(M_i;V_i,Y_i) \\
&= \sum_{i=1}^{n} H(M_i) - \sum_{i=1}^{n} H(M_i|V_i,Y_i) \\
&\overset{(g)}{=} \sum_{i=1}^{n} H(M_i) - \sum_{i=1}^{n} H(M_i|V_i,S_i) \\
&= \sum_{i=1}^{n} I(M_i;V_i,S_i) \\
&= nI(M;V,S|T) \\
&\overset{(h)}{=} nI(M;V,T,S) \\
&= nI(M;U,S) \\
&= n[I(M;S) + I(M;U|S)] \qquad (2.29)
\end{aligned}$$

where (f) and (h) follow from the independence of M and T, and (g) is attributable to the fact that M_i is uniquely determined by V_i for each i by the construction of V_i. Thus,

$$\begin{aligned} H(M) &= H(M|S) + I(M;S) \\ &\overset{(i)}{\leq} I(U;Y) - I(U;S) + I(M;U,Y) - I(M;U|S) + \frac{1}{n} + p_e(f_n, g_n)\log|\mathcal{M}| \\ &= I(U;Y) - I(U;M,S) + I(M;U,Y) + \frac{1}{n} + p_e(f_n, g_n)\log|\mathcal{M}|, \\ &\leq I(U;Y) - I(U;M,S) + I(M;U,Y) + \frac{1}{n} + \epsilon\log|\mathcal{M}| \\ &\leq R_{public}^{correlated}(D+\epsilon) + \frac{1}{n} + \epsilon\log|\mathcal{M}| \end{aligned}$$

where (i) follows from inequalities (2.28) and (2.29). Since $R_{public}^{correlated}(D)$ as a function of D is continuous, letting $n \to \infty$ and then $\epsilon \to 0$ yields

$$H(M) \leq R_{public}^{correlated}(D).$$

This completes the proof of the converse part.

2.8 Summary

A new digital watermarking scenario has been studied, where the watermark source and the covertext source are correlated. A necessary and sufficient condition has been derived under which the watermark source can be recovered with high probability at the end of a public watermarking decoder after the watermarked signal is disturbed by a fixed memoryless attack channel. It has been demonstrated that there exists some public watermarking system with a correlated watermark and covertext for which reliably transmitting the watermark to the watermark receiver is still possible even when the entropy of the watermark source is strictly greater than the standard public watermarking capacity. Moreover, by using similar techniques, a combined source coding and Gel'fand-Pinsker channel coding theorem has also been established, and an open problem proposed recently by Cox et al has been solved.

Chapter 3

Joint Compression and Information Embedding When Watermarks and Covertexts Are Correlated

The problem of joint compression and watermarking is addressed for public watermarking systems with correlated watermark and covertext sources. Sufficient and necessary conditions are determined under which watermarks can be recovered with high probability at the end of public watermark decoding after the compression rate-constrained watermarked signal is disturbed by a fixed memoryless attack channel.

3.1 Introduction

In real applications, watermarked signals are likely to be stored and/or transmitted in compressed format. Obviously, the simplest way of watermarking is to embed watermarks first via a standard watermarking encoder and then compress the watermarked signals via a standard compression encoder with a given compression rate to get compressed watermarked signals. But, the drawback of this approach employing separated watermarking

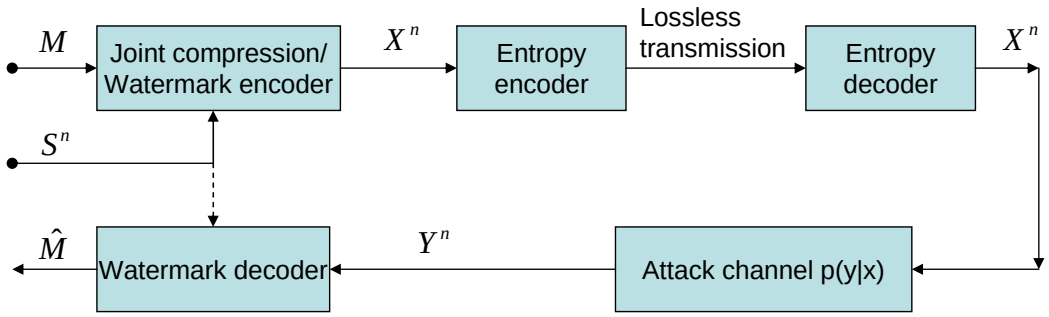

Figure 3.1: Model of joint compression and watermarking system

and compression is obvious, since the compression in such a way could remove certain watermarks from the watermarked signals, and degrade or damage the robustness of watermarked signals. Therefore, instead of treating watermarking and compression separately, it is interesting and beneficial to look at the joint design of watermarking and compression system, as introduced in the following. All notations in Chapter 2 are adopted here.

The communication model of joint compression and watermarking system is designated in Figure 3.1. Here, as before, a watermark M is assumed to be a random variable uniformly taking values over $\mathcal{M} = \{1, 2, ..., |\mathcal{M}|\}$, and a covertext S^n is a sequence of independent and identical drawings of a random variable S with probability distribution $p_S(s)$ taking values over a finite alphabet $\mathcal{S}$, that is, $p(s^n) = \prod_{i=1}^{n} p_S(s_i)$.

Definition 3.1 *A **joint compression and watermarking encoder** of length n with distortion level D with respect to a distortion measure d and watermarking rate R_w and compression rate R_c is a mapping f_n from $\mathcal{M} \times \mathcal{S}^n$ to $\mathcal{X}^n$, $x^n = f_n(m, s^n)$ such that*

$$\begin{aligned} \mathbf{E}d(S^n, X^n) &\leq D, \\ R_w &= \frac{1}{n}\log|\mathcal{M}| \\ \frac{H(X^n)}{n} &\leq R_c. \end{aligned}$$

Note, since $\frac{H(X^n)}{n} \leq R_c$, the stegotext X^n can be entropy-encoded with rate at most R_c.

Definition 3.2 *A mapping $g_n : \mathcal{Y}^n \to \mathcal{M}$, $\hat{m} = g_n(y^n)$ is called a **public watermarking decoder** of length n. Here, the forgery y^n is generated by the attacker according to the attack channel $p(y^n|x^n)$ with input stegotext x^n.*

Given a joint compression and watermarking encoder f_n and a public watermarking decoder g_n, the error probability of watermarking averaged over all watermarks and covertexts is defined by

$$p_e(f_n, g_n) = \Pr\{\hat{M} \neq M\}.$$

Definition 3.3 *A pair (R_w, R_c) is called **publicly achievable** with respect to distortion level D if for any $\epsilon > 0$, there exists, for any n sufficiently large, an n-length joint compression and watermarking encoder f_n with distortion level $D+\epsilon$ and watermarking rate $R_w - \epsilon$ and compression rate R_c, and a public watermarking decoder g_n such that $p_e(f_n, g_n) < \epsilon$.*

Definitions for private case can be given in the same manner.

In this scenario of joint compression and watermarking, the main problem studied in information theory is to describe tradeoffs between watermarking rate, compression rate, distortion between covertexts and watermarked signals and robustness of watermarked signals. Karakos and Papamarcou [19,18] determined best tradeoffs for joint compression and private watermarking systems with finite alphabets and with Gaussian covertext sources, respectively. Maor and Merhav [20,21] gave best tradeoffs for joint compression and public watermarking systems with finite alphabets and no attack or under a fixed attack channel, respectively. The best tradeoffs for the private case were extended to the case of abstract alphabets in [45].

In all these mentioned works on joint compression and digital watermarking, the watermark to be embedded is assumed independent of the covertext. This chapter, as Chapter

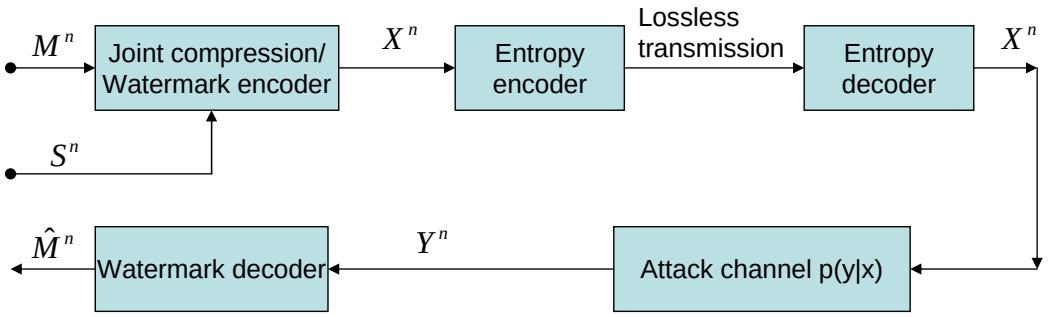

Figure 3.2: Model of joint compression and watermarking system with correlated watermarks and covertexts

2, assuming that watermarks and covertexts (M^n, S^n) are generated identically and independently according to a joint probability distribution $p(m, s)$, investigates the framework in Chapter 2 but with an additional constraint for compression rate of watermarked signals. To be specific, let D be a given distortion level between the covertext S^n and the watermarked signal X^n with respect to the distortion measure d, R_c a given compression rate for watermarked signal X^n, we determine a necessary and sufficient condition for the case of public watermarking, under which the watermark M^n can be fully recovered with high probability at the end of watermark decoding after the compression rate-constrained watermarked signal is disturbed by a fixed memoryless attack channel $p(y|x)$.

3.2 Problem Formulation and Result Statement

3.2.1 Problem Formulation

The joint compression and public watermarking model studied in this chapter is designated in Figure 3.2, in which watermark $M^n \in \mathcal{M}^n$ and covertext $S^n \in \mathcal{S}^n$ are generated

independently and identically according to a joint probability distribution $p(m, s)$, that is,

$$p(m^n, s^n) = \prod_{i=1}^{n} p(m_i, s_i).$$

for any n and $m^n \times s^n \in \mathcal{M}^n \times \mathcal{S}^n$. Let $p(y^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i)$ be a fixed memoryless attack channel with input x^n and output y^n, known to both watermark encoder and watermark decoder.

Definition 3.4 *A **joint compression and watermarking encoder** of length n with distortion level D with respect to the distortion measure d and compression rate R_c is a mapping f_n from $\mathcal{M}^n \times \mathcal{S}^n$ to $\mathcal{X}^n$, $x^n = f_n(m^n, s^n)$ such that*

$$\begin{aligned} \mathbf{E}d(S^n, X^n) &\leq D, \\ \frac{H(X^n)}{n} &\leq R_c. \end{aligned}$$

A public watermarking decoder is defined in Definition 3.2.

Given a joint compression and watermarking encoder f_n and a public watermarking decoder pair g_n, the error probability of watermarking averaged over all watermarks and covertexts is defined by

$$p_e(f_n, g_n) = \Pr\{\hat{M}^n \neq M^n\}.$$

Definition 3.5 *The joint probability distribution $p(m, s)$ of a correlated watermark and covertext source (M, S) is called **publicly admissible** with respect to distortion level D and compression rate R_c if for any $\epsilon > 0$, there exists for any n sufficiently large, an n-length joint compression and watermarking encoder f_n with distortion level $D + \epsilon$ and compression rate R_c and a public watermarking decoder g_n such that $p_e(f_n, g_n) < \epsilon$.*

The aim of this chapter is to determine sufficient and necessary conditions for a joint probability distribution $p(m, s)$ under which $p(m, s)$ is publicly admissible with respect to a distortion level D and a compression rate R_c.

3.2.2 Main Result and Discussion

Let $p(m,s)$, $D, R_c \geq 0$ be given as before. Define

$$R_w^{correlated}(D,R_c) \stackrel{def}{=} \sup_{p(u,x|m,s):\mathbf{E}d(S,X)\leq D} \min\{R_c - I(M,S;U,X) + I(M;U,Y), \\ I(U;Y) - I(U;M,S) + I(M;U,Y)\} \tag{3.1}$$

where the sup is taken over all random variables U, X taking values from finite alphabets $\mathcal{U}$, $\mathcal{X}$ respectively, jointly distributed with M, S, Y with the joint probability distribution $p(m,s,u,x,y) = p(m,s)p(u,x|m,s)p(y|x)$.

The following theorem is the main result of this chapter, which describes the sufficient and necessary conditions for public admissibility of a joint probability $p(m,s)$.

Theorem 3.1 *Let $p(m,s)$ be the joint probability distribution of a joint watermark and covertext source (M,S). For any $D \geq 0, R_c \geq 0$, if $R_w^{correlated}(D,R_c) > 0$, then, $p(m,s)$ is publicly admissible with respect to D and R_c if*

$$H(M) < R_w^{correlated}(D,R_c); \tag{3.2}$$

$p(m,s)$ is not publicly admissible with respect to D and R_c if

$$H(M) > R_w^{correlated}(D,R_c). \tag{3.3}$$

Moreover, for any fixed $R_c > 0$, if $R_w^{correlated}(D,R_c) > 0$ for $D \in [0, D_{\max}(R_c))$, then $p(m,s)$ is publicly admissible with respect to D and R_c if and only if

$$H(M) \quad \leq \quad R_w^{correlated}(D,R_c), \tag{3.4}$$

where $D_{\max}(R_c)$ is the least distortion that achieves the maximum of $R_w^{correlated}(D,R_c)$ over D.

Discussion and Comments:

- The model studied in this chapter can be regarded as a combination of the model in Chapter 2 and the models of [19, 18, 20, 21], in which watermarks are assumed independent of covertexts.

- If the watermark source and the covertext source are independent, that is, $p(m,s) = p(m)p(s)$, then $R_w^{correlated}(D,R_c)$ is equal to the joint compression and public watermarking capacity $C_w(D,R_c)$, as given in [21] by

$$C_w(D,R_c) = \max_{p(u,x|s):\mathbf{E}d(S,X)\le D} \min\{R_c - I(S;U,X), I(U;Y) - I(U;S)\}. \quad (3.5)$$

  Thus, Theorem 3.1 is degraded to the joint compression and public watermarking coding theorem in [21].

- If no compression rate for stegotext is constrained in the model of this chapter, then Theorem 3.1 is equivalent to Theorem 2.1 and Corollary 2.1.

3.3 Properties of $R_w^{correlated}(D,R_c)$

In this section we shall study $R_w^{correlated}(D,R_c)$ by determining some properties, which will be used in the following proofs.

Property 3.1 *Let $p(m,s)$ be a fixed joint probability distribution of (M,S). Then, the supremum in $R_w^{correlated}(D,R_c)$ can be replaced by maximum since the cardinality of $\mathcal{U}$ can be upper bounded by $|\mathcal{M}||\mathcal{S}||\mathcal{X}|+2$, that is,*

$$\begin{aligned} R_w^{correlated}(D,R_c) = \max_{p(u,x|m,s):\mathbf{E}d(S,X)\le D} \min\{&R_c - I(M,S;U,X) + I(M;U,Y), \\ &I(U;Y) - I(U;M,S) + I(M;U,Y)\} \end{aligned} \quad (3.6)$$

where the maximum is taken over all auxiliary random variables (U,X) taking values over $\mathcal{U}\times\mathcal{X}$ with $|\mathcal{U}| \le |\mathcal{M}||\mathcal{S}||\mathcal{X}|+2$, jointly distributed with (M,S,Y) with the joint probability distribution of (M,S,U,X,Y) given by $p(m,s,u,x,y) = p(m,s)p(u,x|m,s)p(y|x)$, and satisfying $\mathbf{E}d(S,X) \le D$.

The proof is very similar to that of Property 2.1, and omitted here.

Property 3.2 *(1) $R_w^{correlated}(D, R_c)$ is concave with respect to (D, R_c);*

(2) For any $R_c > 0$, $R_w^{correlated}(D, R_c)$ as a function of D is increasing and continuous in $[0, \infty)$, and strictly increasing in $[0, D_{max}(R_c))$.

Proof: Since (2) easily follows from (1), we only need to prove the claim in (1).

First, for any random variables (M, S, U, X, Y), we can write

$$\begin{aligned} I(U;Y) - I(U;M,S) + I(M;U,Y) &= H(Y) - H(M,S) + H(M,S|U) + H(M) - H(M,Y|U) \\ I(M,S;U,X) - I(M;U,Y) &= I(M,S;U) + I(M,S;X|U) - I(M;U) - I(M;Y|U). \end{aligned}$$

Now, let $\lambda_1 \geq 0, \lambda_2 \geq 0$ with $\lambda_1 + \lambda_2 = 1$, and $(D_1, R_c^{(1)}), (D_2, R_c^{(2)})$ be any two points. Let (M, S, U_i, X_i, Y_i), $i = 1, 2$ be the random variables achieving $R_w^{correlated}(D_i, R_c^{(i)})$, $T \in \{1, 2\}$ be a random variable independent of all other random variables with $\lambda_i = \Pr\{T = i\}$. Define new random variables

$$U = (U_T, T), X = X_T, Y = Y_T.$$

Then, it is easy to check that for the constructed random vector (M, S, U, X, Y), $\mathbf{E}d(S, X) \leq \lambda_1 D_1 + \lambda_2 D_2$. Moreover, one has

$$\begin{aligned}
&\lambda_1 R_c^{(1)} + \lambda_2 R_c^{(2)} - I(M, S; U, X) + I(M; U, Y) \\
&= \lambda_1 R_c^{(1)} + \lambda_2 R_c^{(2)} - I(M, S; U) - I(M, S; X|U) + I(M; U) + I(M; Y|U) \\
&= \lambda_1 R_c^{(1)} + \lambda_2 R_c^{(2)} - I(M, S; U_T, T) - I(M, S; X|U_T, T) + I(M; U_T, T) + I(M; Y|U_T, T) \\
&= \lambda_1 R_c^{(1)} + \lambda_2 R_c^{(2)} - I(M, S; T) - I(M, S; U_T|T) - I(M, S; X|U_T, T) \\
&\quad + I(M; T) + I(M; U_T|T) + I(M; Y|U_T, T) \\
&= \lambda_1 R_c^{(1)} + \lambda_2 R_c^{(2)} - I(M, S; U_T|T) - I(M, S; X|U_T, T) + I(M; U_T|T) + I(M; Y|U_T, T) \\
&= \lambda_1 \left(R_c^{(1)} - I(M, S; U_1) - I(M, S; X_1|U_1) + I(M; U_1) + I(M; Y_1|U_1) \right) \\
&\quad + \lambda_2 \left(R_c^{(2)} - I(M, S; U_2) - I(M, S; X_2|U_2) + I(M; U_2) + I(M; Y_2|U_2) \right) \\
&= \lambda_1 \left(R_c^{(1)} - I(M, S; U_1, X_1) + I(M; U_1, Y_1) \right) + \lambda_2 \left(R_c^{(2)} - I(M, S; U_2, X_2) + I(M; U_2, Y_2) \right) \\
&\geq \lambda_1 R_w^{correlated}(D_1, R_c^{(1)}) + \lambda_2 R_w^{correlated}(D_2, R_c^{(2)}),
\end{aligned}$$

and similarly,

$$I(U; Y) - I(U; M, S) + I(M; U, Y) \geq \lambda_1 R_w^{correlated}(D_1, R_c^{(1)}) + \lambda_2 R_w^{correlated}(D_2, R_c^{(2)}).$$

So, by the definition of $R_w^{correlated}(D, R_c)$, one has

$$\begin{aligned}
&R_w^{correlated}(\lambda_1 D_1 + \lambda_2 D_2, \lambda_1 R_c^{(1)} + \lambda_2 R_c^{(2)}) \\
&\geq \min\{\lambda_1 R_c^{(1)} + \lambda_2 R_c^{(2)} - I(M, S; U, X) + I(M; U, Y), I(U; Y) - I(U; M, S) + I(M; U, Y)\} \\
&\geq \lambda_1 R_w^{correlated}(D_1, R_c^{(1)}) + \lambda_2 R_w^{correlated}(D_2, R_c^{(2)}).
\end{aligned}$$

The concavity of $R_w^{correlated}(D, R_c)$ with respect to D, R_c is proved. □

3.4 Proof of the Direct Part

Suppose $R_w^{correlated}(D, R_c) > 0$ for $D \geq 0, R_c > 0$. By employing random coding argument we first shall show $p(m, s)$ is publicly admissible with respect to D and R_c if $H(M) <$

$R_w^{correlated}(D, R_c)$, then show that $p(m, s)$ is publicly admissible with respect to $R_c > 0, D \in [0, D_{max}(R_c))$ if $H(M) = R_w^{correlated}(D, R_c)$, by exploiting properties of $R_w^{correlated}(D, R_c)$.

Now, assume that (U, X) be random variables over finite alphabets $\mathcal{U} \times \mathcal{X}$ achieving the $R_w^{correlated}(D, R_c)$, jointly distributed with M, S, Y with the probability distribution $p(m, s, x, u, y)$ of (M, S, U, X, Y) given by $p(m, s, x, u, y) = p(m, s)p(x, u|m, s)p(y|x)$. Thus, $\mathbf{E}d(S, X) \leq D$ and

$$R_w^{correlated}(D, R_c) = \min\{R_c - I(M, S; U, X) + I(M; U, Y), I(U; Y) - I(U; M, S) + I(M; U, Y)\}.$$

Define $\gamma \stackrel{\triangle}{=} \min\{R_c - I(M, S; U, X) + I(M; U, Y), I(U; Y) - I(U; M, S) + I(M; U, Y)\} - H(M)$. We want to show that, for any $\epsilon > 0$w there exist joint compression and watermarking encoder and public watermarking decoder (f_n, g_n) for all sufficiently large n such that $\mathbf{E}d(S^n, f_n(M^n, S^n)) < D + \epsilon$, $H(f_n(M^n, S^n))/n \leq R_c$ and $p_e(f_n, g_n) < \epsilon$.

3.4.1 Random Joint Compression and Watermarking Coding

Random Codes Generation: Three random codebooks C,W and G will be generated as follows.

- First, generate identically distributed and independently $\exp(n[H(M) - I(M; U, Y) + \gamma/8])$ vectors $m^n \in \mathcal{M}^n$ according to the probability $p(m)$, and uniformly distribute all these vectors into $t \stackrel{\triangle}{=} \exp(n[H(M) - I(M; U, Y) + \gamma/4])$ bins, each bin $C(i)$, $i = 1, ..., t$, containing $\exp(n[I(M; U, Y) - \gamma/8])$ vectors. Denote the random codebook by $C = \{C(i)\}_{i=1}^t$.

- Second, for each index $i = 1, ..., t$ generate a bin of vectors $W(i) = \{u^n(i, j) : j = 1, 2, ..., \exp(n[I(U; M, S) + \gamma/4])\}$, each $u^n(i, j) \in \mathcal{U}^n$ is generated identically and independently according to the probability $p(u)$ derived from the joint probability $p(m, s, x, u, y)$. Denote the random codebook by $W = \{W(i)\}_{i=1}^t$.

- Third, for each vector $u^n(i, j) \in W(i)$, $i = 1, ..., t$, $j = 1, 2, ..., \exp(n[I(U; M, S) + \gamma/4])$ generate a bin of vectors $G(i, j) = \{x^n(i, j, l) : l = 1, 2, ..., \exp(n[I(M, S; X|U) +$

$\gamma/4])\}$, each $x^n(i,j,l) \in \mathcal{X}^n$ is generated identically and independently according to the probability $p(x^n|u^n(i,j))$. Denote the random codebook by $G = \{G(i,j), i = 1,...,t, j = 1,2,...,\exp(n[I(U;M,S)+\gamma/4])\}$.

- Finally, the two codebooks C and W are distributed to the watermarking decoder, and the codebook G is sent to the lossless decompressor of stegotexts.

Watermarking encoding: Fix codebooks C, W, G. Given a watermark m^n and a covertext s^n.

- If (m^n, s^n) is not jointly ϵ-typical, then an encoding error is declared;
- If (m^n, s^n) is jointly ϵ-typical, but no $C(i)$ contains m^n, $i = 1,2,...t$, then an encoding error is declared;
- If (m^n, s^n) is jointly ϵ-typical and $C(i)$ is the first bin in C containing m^n, but no vector $u^n \in W(i)$ such that (m^n, s^n, u^n) is jointly ϵ-typical, then an encoding error is declared;
- If (m^n, s^n) is jointly ϵ-typical, $C(i)$ is the first bin in C containing m^n, and $u^n(i,j)$ is the first vector in $W(i)$ such that $(m^n, s^n, u^n(i,j))$ is jointly ϵ-typical, but no vector x^n in the bin $G(i,j)$ such that $(m^n, s^n, u^n(i,j), x^n)$ is jointly ϵ-typical, then an encoding error is declared;
- If (m^n, s^n) is jointly ϵ-typical, $C(i)$ is the first bin in C containing m^n, and $u^n(i,j)$ is the first vector in $W(i)$ such that $(m^n, s^n, u^n(i,j))$ is jointly ϵ-typical, then choose the first vector $x^n(i,j,l)$ as the stegotext in the bin $G(i,j)$ such that $(m^n, s^n, u^n(i,j), x^n(i,j,l))$ is jointly ϵ-typical;
- If an encoding error is declared, then define a fixed x_0^n as the stegotext.

Watermarking decoding: Fix codebooks C, W. The decoding scheme is exactly the same as that of Chapter 2. To be specific, let y^n be a forgery received by the watermarking decoder when $m^n \in C(i)$ is transmitted using s^n, $u^n(i,j) \in W(i)$ and $x^n(i,j,l) \in G(i,j)$.

- The decoder finds the first vector, say $u^n(i_0, j_0)$, in the codebook W such that $(u^n(i_0, j_0), y^n)$ is jointly ϵ-typical with respect to $p(u, y)$;

- If no or more than one u^n are found in W such that (u^n, y^n) is jointly ϵ-typical, then a decoding error is declared;

- If only one $u^n(i_0, j_0)$ is found in W such that $(u^n(i_0, j_0), y^n)$ is jointly ϵ-typical, then the decoder finds the unique vector $\hat{m}^n \in C(i_0)$ such that $(\hat{m}^n, u^n(i_0, j_0), y^n)$ is jointly ϵ-typical with respect to $p(m, u, y)$, and decodes $\hat{m}^n$ to be the watermark;

- If only one $u^n(i_0, j_0)$ is found in W such that $(u^n(i_0, j_0), y^n)$ is jointly ϵ-typical, but no or more than one $\hat{m}^n$ are found in the bin $C(i_0)$ such that $(\hat{m}^n, u^n(i_0, j_0), y^n)$ is jointly ϵ-typical, then a decoding error is also declared.

3.4.2 Averaged Error Probability

From the random watermarking encoding and decoding scheme, there are the following encoding error events:

- E_0: (m^n, s^n) is not jointly ϵ-typical;

- E_1: $(m^n, s^n) \in \bar{E}_0$, but $m^n \notin C$, where $\bar{E}_0$ denotes the complement of E_0;

- E_2: $(m^n, s^n) \notin E_0 \cup E_1$, but no $u^n \in W(i)$ such that (m^n, s^n, u^n) is ϵ-typical, where $i = i(m^n)$ is the smallest i such that $C(i)$ contains m^n; and

- E_3: $(m^n, s^n) \notin E_0 \cup E_1 \cup E_2$, $u^n(i, j)$ is the first vector in $W(i)$ such that $(m^n, s^n, u^n(i, j))$ is ϵ-typical, but no $x^n \in G(i, j)$ is found such that $(m^n, s^n, u^n(i, j), x^n)$ is ϵ-typical.

Suppose that encoding (m^n, s^n) is successful via C, W, G with the stegotext x^n. Let y^n be a forgery generated by the attacker. Then, there are the following decoding error events:

- E: $(m^n, u^n(i,j), y^n)$ is not jointly ϵ-typical;
- E': more than one $u^n \in W$ such that (u^n, y^n) is ϵ-jointly typical; and
- E'': more that one $\hat{m}^n \in C(i)$ such that $(\hat{m}^n, u^n(i,j), y^n)$ is jointly ϵ-typical.

Based on the random coding argument in this section and the analysis of error probabilities in Subsection 2.6.3 of Chapter 2, we can obtain for sufficiently large n

$$\mathbf{E}_{C,W,G} p_e(C,W,G) \leq 8\epsilon \tag{3.7}$$

if we could show that

$$\Pr\{E_3\} \leq \epsilon \tag{3.8}$$

To reach this, we note that $x^n \in G(i,j)$ is generated identically and independently according to $p(x^n|u^n(i,j))$, so for large n the probability

$$\Pr\{(m^n, s^n, u^n(i,j), x^n) \text{ is jointly } \epsilon\text{-typical}\} > 2^{-n[I(X;M,S|U)+\gamma/8]}.$$

Thus, there exists a large number n_1 such that for all $n > n_1$

$$\begin{aligned}
\Pr\{E_3\} &= \Pr\{\text{no } x^n \in G(i,j) \text{ such that } (m^n, s^n, u^n(i,j), x^n) \text{ is jointly } \epsilon\text{-typical}\} \\
&\leq (1 - 2^{-n[I(X;M,S|U)+\gamma/8]})^{2^{n[I(X;M,S|U)+\gamma/4]}} \\
&\leq 2^{-2^{n\gamma/8}} \leq \epsilon.
\end{aligned} \tag{3.9}$$

3.4.3 Distortion Constraint and Compression Rate Constraint

Let x_0^n be the fixed stegotext if an encoding error is declared. By the watermark encoding scheme we have

$$\begin{aligned}
&\mathbf{E}_{C,W,G}\mathbf{E}_{M^n,S^n}[d(S^n, X^n)] = \mathbf{E}[d(S^n, X^n)] \\
&= \Pr\{\cup_{i=0}^3 E_i\}\mathbf{E}[d(S^n, X^n)|\cup_{i=0}^3 E_i] + \Pr\{\cap_{i=0}^i \bar{E}_i\}\mathbf{E}[d(S^n, X^n)|\cap_{i=0}^3 \bar{E}_i] \\
&\leq \Pr\{\cup_{i=0}^3 E_i\} d_{max} + \Pr\{\cap_{i=0}^i \bar{E}_i\}\mathbf{E}[d(S^n, X^n)|\cap_{i=0}^3 \bar{E}_i] \\
&\leq 4\epsilon d_{max} + \epsilon + D
\end{aligned} \tag{3.10}$$

where the last inequality follows from the fact that for large n, $d(s^n, x^n) \leq D + \epsilon$ since $(m^n, s^n, u^n(i,j), x^n(i,j,l))$ is jointly ϵ-typical with respect to $p(m,s,u,x)$ with $\mathbf{E}d(S,X) \leq D$

Finally, by the construction of the codebook G,

$$\begin{aligned}
\frac{H(X^n)}{n} &\leq \frac{1}{n}\log|G| \\
&\leq [H(M) - I(M;U,Y) + \gamma/4] + [I(M,S;X|U) + \gamma/4] + [I(M,S;U) + \gamma/4] \\
&\leq R_c - \gamma/4 \\
&\leq R_c
\end{aligned} \tag{3.11}$$

since

$$\begin{aligned}
&H(M) - I(M;U,Y) + I(M,S;X|U) + I(M,S;U) \\
&= H(M) - I(M;U,Y) + I(M,S;X,U) \\
&\leq R_c - \gamma
\end{aligned}$$

by the definition of γ.

3.4.4 Existence of Watermarking Encoders and Decoders

By Markov inequality and (3.7), one has

$$\begin{aligned}
\Pr\{p_e(C,W,G) \geq \sqrt{8\epsilon}\} &\leq \frac{\mathbf{E}_{C,W,G}p_e(C,W,G)}{\sqrt{8\epsilon}} \\
&= \sqrt{8\epsilon}.
\end{aligned}$$

Let

$$\Gamma = \{(C,W,G) : p_e(C,W,G) \leq \sqrt{8\epsilon}\}, \tag{3.12}$$

then $\Pr\{\Gamma\} \geq 1 - \sqrt{8\epsilon}$.

So, from (3.10) one has

$$\begin{aligned}
&\sum_{(C,W,G)\in\Gamma} \Pr(C,W,G)\mathbf{E}_{S^n,M^n}\left(d(S^n,X^n)|C,W,G\right)\\
&\quad\leq\ \mathbf{E}_{C,W,G}[\mathbf{E}_{S^n,M^n}\left(d(S^n,X^n)|C,W,G\right)]\\
&\quad\leq\ 4\epsilon d_{max}+\epsilon+D.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\sum_{(C,W,G)\in\Gamma} \frac{\Pr(C,W,G)}{\Pr\{\Gamma\}}\mathbf{E}_{S^n,M^n}\left(d(S^n,X^n)|C,W,G\right)\\
&=\ \frac{1}{\Pr\{\Gamma\}}\sum_{(C,W,G)\in\Gamma} \Pr(C,W,G)\mathbf{E}_{S^n,M^n}\left(d(S^n,X^n)|C,W,G\right)\\
&\leq\ \frac{4\epsilon d_{max}+\epsilon+D}{1-\sqrt{8\epsilon}}=D+\epsilon'
\end{aligned}\tag{3.13}$$

for some small number $\epsilon'>0$ and $\epsilon'\to 0$ as $\epsilon\to 0$.

Combination of (3.11),(3.12) and (3.13) guarantees the existence of joint compression and watermarking encoder and public watermarking decoder for all large n such that the error probability is less than $\sqrt{8\epsilon}$, the averaged distortion $\mathbf{E}d(S^n,X^n)$ is less than $D+\epsilon'$ and $H(X^n)/n\leq R_c$, that is, the probability $p(m,s)$ is publicly admissible with respect to D and compression rate R_c if $R_w^{correlated}(D,R_c)>H(M)$.

To finish the proof of the direct part, in the following we shall prove that the probability $p(m,s)$ is publicly admissible with respect to $D\in[0,D_{\max}(R_c))$ and compression rate R_c if $H(M)=R_w^{correlated}(D,R_c)$. Indeed, for any small $\epsilon>0$, one has

$$R_w^{correlated}(D+\epsilon,R_c)>R_w^{correlated}(D,R_c)=H(M)$$

since $R_w^{correlated}(D,R_c)$ is strictly increasing in $[0,D_{\max}(R_c))$ by Property 3.2. Thus, $p(m,s)$ is publicly admissible with respect to $D+\epsilon$ and compression rate R_c by the proof of the first step. Because $\epsilon>0$ can be arbitrarily small, $p(m,s)$ is publicly admissible with respect to D and compression rate R_c. The proof of the direct part is finished.

3.5 Proof of the Converse Part

In this section we shall prove the converse part, that is, for any arbitrary but fixed number $\epsilon > 0$, if there exists for any sufficiently large n, a joint compression and watermarking encoder and public decoder pair (f_n, g_n) with length n such that

$$\begin{aligned} \mathbf{E}d(S^n, f_n(M^n, S^n)) &\leq D + \epsilon, \\ \frac{1}{n}H(f_n(M^n, S^n)) &\leq R_c, \\ p_e = \Pr\{g_n(Y^n) \neq M^n\} &< \epsilon, \end{aligned}$$

where Y^n is generated by the attack channel with input $X^n = f_n(M^n, S^n)$, then $H(M) \leq R_w^{correlated}(D, R_c)$.

The proof is very long and will be finished in four steps.

Step one. Since $X^n = f_n(M^n, S^n)$ is a function of (M^n, S^n), $H(X^n|S^n, M^n) = 0$. Thus

$$\begin{aligned} H(M^n|S^n) &= H(M^n|S^n) - H(X^n|M^n, S^n) \\ &= H(M^n, X^n|S^n) \\ &= H(M^n, X^n) - I(S^n; M^n, X^n) \\ &= H(M^n|X^n) + H(X^n) - I(S^n; M^n, X^n). \end{aligned}$$

Following from the basic properties of mutual information, the Markov chain $(M^n, S^n) \to X^n \to Y^n$ and from the fact that the covertext source $\{S_i\}_{i=1}^{\infty}$ is memoryless, it is not hard

to get

$$\begin{aligned} I(S^n; M^n, X^n) &= I(S^n; M^n, X^n, Y^n) - I(S^n; Y^n | M^n, X^n) \\ &= I(S^n; M^n, X^n, Y^n) \\ &= \sum_{i=1}^{n} I(S_i; M^n, X^n, Y^n | S_{i+1}^n) \\ &= \sum_{i=1}^{n} \left[I(S_i; M^n, X^n, Y^n, S_{i+1}^n) - I(S_i; S_{i+1}^n) \right] \\ &= \sum_{i=1}^{n} I(S_i; M^n, S_{i+1}^n, Y_1^{i-1}, X_i, X_1^{i-1}, X_{i+1}^n, Y_i^n). \end{aligned}$$

Let $V_i = (M^n, S_{i+1}^n, Y_1^{i-1})$. Obviously,

$$\sum_{i=1}^{n} I(S_i; V_i, X_i, X_1^{i-1}, X_{i+1}^n, Y_i^n) \geq \sum_{i=1}^{n} I(S_i; V_i, X_i); \tag{3.14}$$

and

$$\sum_{i=1}^{n} I(M_i; V_i, Y_i) = \sum_{i=1}^{n} I(M_i; V_i, X_i, S_i). \tag{3.15}$$

So, combination of (3.14) and (3.15) yields

$$\begin{aligned} \sum_{i=1}^{n} I(S_i; V_i, X_i, X_1^{i-1}, X_{i+1}^n, Y_i^n) &\geq \sum_{i=1}^{n} I(S_i; V_i, X_i) - \sum_{i=1}^{n} I(M_i; V_i, Y_i) + \sum_{i=1}^{n} I(M_i; V_i, X_i, S_i) \\ &= \sum_{i=1}^{n} I(S_i; V_i, X_i) - \sum_{i=1}^{n} I(M_i; V_i, Y_i) \\ &\quad + \sum_{i=1}^{n} I(M_i; S_i) + \sum_{i=1}^{n} I(M_i; V_i, X_i | S_i) \\ &= \sum_{i=1}^{n} I(M_i; S_i) + \sum_{i=1}^{n} I(M_i, S_i; V_i, X_i) - \sum_{i=1}^{n} I(M_i; V_i, Y_i). \end{aligned}$$

Therefore,

$$\begin{aligned} H(M^n|S^n) &= H(M^n|X^n) + H(X^n) - \sum_{i=1}^{n} I(S_i; M^n, S_{i+1}^n, Y_1^{i-1}, X_i, X_1^{i-1}, X_{i+1}^n, Y_i^n) \\ &= H(M^n|X^n) + H(X^n) - \sum_{i=1}^{n} I(S_i; V_i, X_i, X_1^{i-1}, X_{i+1}^n, Y_i^n) \\ &\leq H(M^n|X^n) + H(X^n) - \sum_{i=1}^{n} I(M_i; S_i) \\ &\quad - \sum_{i=1}^{n} I(M_i, S_i; V_i, X_i) + \sum_{i=1}^{n} I(M_i; V_i, Y_i). \end{aligned} \tag{3.16}$$

Step two. By using the same approach as that in Section 2.7 of Chapter 2, we have

$$I(M^n; Y^n) - I(M^n; S^n) \leq \sum_{i=1}^{n} [I(V_i; Y_i) - I(V_i; S_i)]. \tag{3.17}$$

Step three. Let $T \in \{1, 2, ..., n\}$ be a time-sharing random variable, uniformly distributed and independent of all other random variables. Define $S = S_i$, $M = M_i$, $X = X_i$, $Y = Y_i$, $V = V_i$ when $T = i$, and $U = (V, T)$. Define the joint probability distribution of (M, S, U, X, Y) as

$$p(m, s, u, x, y) = \frac{1}{n} \sum_{i=1}^{n} \Pr\{(M_i, S_i, X_i, U_i, Y_i) = (m, s, x, u, y)\}.$$

By the additivity of $d(s^n, x^n) = \frac{1}{n} \sum_{i=1}^{n} d(s_i, x_i)$ and the definition of (M, S, U, X, Y), it is obvious that $\mathbf{E}d(S, X) \leq D + \epsilon$ since $\mathbf{E}d(S^n, X^n) \leq D + \epsilon$. Moreover, from the construction of (M, S, U, X, Y), $(M, S, U) \to X \to Y$ forms a Markov chain.

By Fano's inequality, one has

$$\begin{aligned} \frac{1}{n} H(M^n|X^n) &\leq \frac{1}{n} H(M^n|Y^n) \\ &\leq \frac{1}{n} + p_e \log |\mathcal{M}|. \end{aligned}$$

In light of $I(M,S;T)=0, I(M;T)=0$, and (3.16), one can obtain

$$\begin{aligned}
H(M|S) &= \frac{1}{n}H(M^n|S^n) \\
&\leq \frac{1}{n}H(M^n|X^n)+\frac{1}{n}H(X^n)-\frac{1}{n}\sum_{i=1}^{n}I(M_i;S_i) \\
&\quad -\frac{1}{n}\sum_{i=1}^{n}I(M_i,S_i;V_i,X_i)+\frac{1}{n}\sum_{i=1}^{n}I(M_i;V_i,Y_i) \\
&\leq \frac{1}{n}+p_e\log|\mathcal{M}|+R_c \\
&\quad -I(M,S;V,X|T)-I(M;S|T)+I(M;V,Y|T) \\
&= \frac{1}{n}+p_e\log|\mathcal{M}|+R_c \\
&\quad -I(M,S;V,T,X)+I(M,S;T)-I(M;S)+I(M;V,T,Y)-I(M;T) \\
&= \frac{1}{n}+p_e\log|\mathcal{M}|+R_c \\
&\quad -I(M,S;U,X)-I(M;S)+I(M;U,Y) \\
&\leq R_c-I(M,S;U,X)-I(M;S)+I(M;U,Y)+\frac{1}{n}+\epsilon\log|\mathcal{M}|.
\end{aligned}$$

So,

$$\begin{aligned}
H(M)-I(M;U,Y) &= H(M|S)+I(M;S)-I(M;U,Y) \\
&\leq R_c-I(M,S;U,X)+\frac{1}{n}+\epsilon\log|\mathcal{M}|. \qquad (3.18)
\end{aligned}$$

On the other hand, from (3.17) and the construction of U, we can show

$$\frac{1}{n}I(M^n;Y^n)-\frac{1}{n}I(M^n;S^n) \leq I(U;Y)-I(U;S),$$

which, combined with (2.29), yields

$$\begin{aligned}
H(M)-I(M;U,Y) &= H(M|S)+I(M;S)-I(M;U,Y)\\
&= \frac{1}{n}H(M^n|S^n)+I(M;S)-I(M;U,Y)\\
&= \frac{1}{n}I(M^n;Y^n)-\frac{1}{n}I(M^n;S^n)+\frac{1}{n}H(M^n|Y^n)+I(M;S)-I(M;U,Y)\\
&\leq I(U;Y)-I(U;S)+I(M;S)-I(M;U,Y)+\frac{1}{n}+p_e\log|\mathcal{M}|\\
&= I(U;Y)-I(U;S)-I(M;U|S)+\frac{1}{n}+p_e\log|\mathcal{M}|\\
&= I(U;Y)-I(U;M,S)+\frac{1}{n}+p_e\log|\mathcal{M}|\\
&\leq I(U;Y)-I(U;M,S)+\frac{1}{n}+\epsilon\log|\mathcal{M}| \qquad (3.19)
\end{aligned}$$

Step four. From (3.18) and (3.19), we have

$$\begin{aligned}
H(M) \leq\ & \min\{R_c-I(M,S;U,X)+I(M;U,Y), I(U;Y)-I(U;M,S)+I(M;U,Y)\}\\
&+\frac{1}{n}+\epsilon\log|\mathcal{M}|.
\end{aligned}$$

Thus,

$$H(M) \leq R_w^{correlated}(D+\epsilon,R_c)+\frac{1}{n}+\epsilon\log|\mathcal{M}.$$

Since $\epsilon\log|\mathcal{M}|$ can be arbitrarily small, $1/n\to 0$ and $R_w^{correlated}(D,R_c)$ is continuous with respect to $D\geq 0$, one has

$$H(M) \leq R_w^{correlated}(D,R_c).$$

The proof of the converse part is completed.

3.6 Summary

In this chapter, we investigate a joint compression and public watermarking scenario with correlated watermarks and covertexts, which can be regarded as a generalization of existing

joint compression and watermarking models. For given distortion level between covertexts and stegotexts and compression rate for stegotexts, sufficient and necessary conditions are determined under which reliably transmitting watermarks via correlated covertexts is successful with high probability even after the compressed stegotexts are disturbed by a fixed attacker.

Chapter 4

Information Embedding with Fidelity Criterion for Watermarks

In this chapter, the models of digital watermarking of Chapter 2 and Chapter 3 will be investigated continuously respectively, but with a relaxed and more reasonable assumption on recovery of watermarks from a viewpoint of real applications. More specifically, in this chapter it is assumed that the decoded watermark has tolerant distortion with respect to the transmitted watermark instead of fully recovering the transmitted watermark as in Chapter 2 and Chapter 3. With this assumption in mind, sufficient conditions are determined for the case without compression of stegotexts and the case with compression of stegotexts, under which transmitting watermarks to public watermark receivers is reliable in the presence of a fixed attack channel.

4.1 Problem Formulation and Main Results

All notations in previous chapters are kept here. Specifically, (M^n, S^n) are watermarks and covertexts generated identically and independently by a random vector $(M, S) \in \mathcal{M} \times \mathcal{S}$ with a joint probability distribution $p(m, s)$. Let $p(y|x)$ be a fixed attack channel with

input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ known to watermark transmitter and watermark decoder, and d the distortion measure between $\mathcal{S}$ and $\mathcal{X}$. Furthermore, let $\hat{\mathcal{M}}$ be a reproduction alphabet of decoded watermarks, and define a distortion measure $d_1 : M \times \hat{\mathcal{M}} \to [0, \infty)$ with $d'_{max} = \max_{s \in \mathcal{S}, \hat{s} \in \hat{\mathcal{S}}} d_1(s, \hat{s})$. Without loss of generality, assume that $\max_{s \in \mathcal{S}} \min_{\hat{s} \in \hat{\mathcal{S}}} d_1(s, \hat{s}) = 0$.

The definitions of watermarking encoder and joint compression and watermarking encoder are the same as those in the previous chapters, however, they are re-stated here for completeness.

Definition 4.1 *A **watermarking encoder** of length n with distortion level D with respect to the distortion measure d is a mapping f_n from $\mathcal{M}^n \times \mathcal{S}^n$ to $\mathcal{X}^n$ with $x^n = f_n(m^n, s^n)$ such that $\mathbf{E}d(S^n, X^n) \leq D$.*

*A **joint compression and watermarking encoder** of length n with distortion level D with respect to the distortion measure d and compression rate R_c is a mapping f_n from $\mathcal{M}^n \times \mathcal{S}^n$ to $\mathcal{X}^n$ with $x^n = f_n(m^n, s^n)$ such that $\mathbf{E}d(S^n, X^n) \leq D$ and $H(X^n)/n \leq R_c$.*

Definition 4.2 *A mapping $g_n : \mathcal{Y}^n \to \hat{\mathcal{M}}^n$ with $\hat{m}^n = g_n(y^n)$ is called a **public watermarking decoder** with length n and distortion level D_1 with respect to d_1 if $\mathbf{E}d_1(M^n, \hat{M}^n) \leq D_1$.*

Definition 4.3 *The joint probability distribution $p(m, s)$ of a correlated watermark and covertext source (M, S) is called **publicly admissible with respect to distortion level** D, D_1 (**publicly admissible with respect to distortion level** D, D_1 **and compression rate** R_c) if for arbitrary $\epsilon > 0$, there exists, for any sufficiently large n, a watermarking encoder f_n with length n and distortion level $D + \epsilon$ (a joint compression and watermarking encoder f_n with distortion level $D + \epsilon$ and compression rate R_c) and a public watermarking decoder g_n with distortion level $D_1 + \epsilon$.*

This chapter will address the problem on the best tradeoffs between the public admissibility of $p(m, s)$, distortion levels D, D_1, and compression rate R_c. In other words, under

what conditions is $p(m,s)$ publicly admissible for given distortion level D, D_1 and compression rate R_c? Obviously, if $D_1 = 0$, then the problem coincides with the previous ones in Chapter 2 and Chapter 3. In the following, only sufficient conditions are determined for the case without compressing stegotexts and the case with compressing stegotexts under which $p(m,s)$ is publicly admissible, and no necessary conditions are obtained here.

To state the main results of this chapter, some notations are defined as follows. Given distortion levels D, D_1 with respect to distortion measures d, d_1, let $F(D, D_1)$ be the set of all random vectors $(V, U, X) \in \mathcal{V} \times \mathcal{U} \times \mathcal{X}$, $\mathcal{V}$,$\mathcal{U}$ be any finite alphabets, with the joint probability distribution of (M, S, V, U, X, Y) given by

$$p(m,s,v,u,x,y) = p(m,s)p(v,u,x|m,s)p(y|x)$$

such that

- $\mathbf{E}d(S,X) \leq D$, and
- there exists a function $g : \mathcal{V} \times \mathcal{U} \times \mathcal{Y} \to \hat{\mathcal{M}}$ such that

$$\mathbf{E}d_1(M, g(V,U,Y)) \quad \leq \quad D_1.$$

Now, define

$$\begin{aligned} R_{fidelity}(D, D_1) \quad = \quad & \sup_{(V,U,X)\in F(D,D_1)} [I(U;Y) - I(U;M,S,V) + I(V;U,Y) \\ & + H(M,S|V) - H(S|M)], \end{aligned}$$

and

$$\begin{aligned} R_{fidelity}(D, D_1, R_c) \quad = \quad & \sup_{(V,U,X)\in F(D,D_1)} \min\{R_c - I(M,S,V;U,X) + I(V;U,Y) \\ & + H(M,S|V) - H(S|M), I(U;Y) - I(U;M,S,V) \\ & + I(V;U,Y) + H(M,S|V) - H(S|M)\}. \end{aligned}$$

It is ready to state the main results of this chapter in the following, and Theorem 4.1 applies for the case without compression of stegotexts while Theorem 4.2 applies for the case with compression of stegotexts.

Theorem 4.1 *Let $p(m,s)$ be the fixed joint probability distribution of a joint watermark and covertext source (M,S). For any $D \geq 0, D_1 \geq 0$, if $R_{fidelity}(D,D_1) > 0$, then, $p(m,s)$ is publicly admissible with respect to D, D_1 if*

$$H(M) \;<\; R_{fidelity}(D,D_1).$$

Theorem 4.2 *Let $p(m,s)$ be the fixed joint probability distribution of a joint watermark and covertext source (M,S). For any $D \geq 0, D_1 \geq 0, R_c \geq 0$, if $R_{fidelity}(D,D_1,R_c) > 0$, then, $p(m,s)$ is publicly admissible with respect to D, D_1 and R_c if*

$$H(M) \;<\; R_{fidelity}(D,D_1,R_c).$$

4.2 Proof of Theorem 4.1

In this section we shall prove Theorem 4.1. Now suppose that $(V,U,X) \in F(D,D_1)$ satisfies

$$H(M) < I(U;Y) - I(U;M,S,V) + I(V;U,Y) + H(M,S|V) - H(s|M),$$

or equivalently,

$$\begin{aligned} 0 \;&<\; I(U;Y) - I(U;M,S,V) + I(V;U,Y) - I(V;M,S) \\ &=\; I(U;Y) - I(U;V) - I(U;M,S|V) + I(V;U) + I(V;Y|U) - I(V;M,S) \\ &=\; I(U,V;Y) - I(U,V;M,S) \end{aligned}$$

and g is a function from $\mathcal{V} \times \mathcal{U} \times \mathcal{Y}$ to $\hat{\mathcal{M}}$ such that $\mathbf{E}d_1(M, g(V,U,Y)) \leq D_1$.

Denote $\gamma \stackrel{\triangle}{=} I(U;Y) - I(U;M,S) + I(V;U,Y) - I(V;M,S) > 0$. Let $\epsilon > 0$ be an arbitrarily small but fixed number. We will show the existence of watermarking encoder and public watermarking decoder pairs (f_n, g_n) for all sufficiently large n such that $\mathbf{E}d(S^n, f_n(M^n,S^n)) < D + \epsilon$, and $\mathbf{E}d_1(M^n, g_n(Y^n)) < D_1 + \epsilon$. To reach this, the following random coding argument is adopted.

4.2.1 Watermarking Coding Scheme

Random Codes Generation: Two random codebooks C and W will be generated as follows.

- Generate identically and independently $\exp(n[I(V;M,S)+\gamma/8])$ vectors $v^n \in \mathcal{V}^n$ according to the probability distribution $p(v)$ of the random V derived from the joint probability distribution $p(m,s,v,u,x,y)$ of (M,S,V,U,X,Y), then uniformly distribute them into $t \triangleq \exp(n[I(V;M,S)-I(V;U,Y)+\gamma/4])$ bins $C(i)$, $i=1,2,...,t$, each bin containing $\exp(n[I(V;U,Y)-\gamma/8])$ vectors v^n. Denote the random codebook by $C=\{C(i)\}_{i=1}^{t}$.

- Generate identically and independently $\exp(n[I(V;M,S)-I(V;U,Y)+I(U;M,S,V)+\gamma/2])$ vectors $u^n \in \mathcal{U}^n$ according to the probability distribution $p(u)$ of the random U derived from the joint probability distribution $p(m,s,v,u,x,y)$ of (M,S,V,U,X,Y), then uniformly distribute them into t bins $W(i)$, $i=1,2,...,t$, each bin containing $\exp(n[I(U;M,S,V)+\gamma/4])$ vectors u^n. Denote the random codebook by $W=\{W(i)\}_{i=1}^{t}$.

- The two codebooks C and W are then distributed to the watermarking decoder.

Watermarking encoding: Fix codebooks C, W. Given a watermark m^n and a covertext s^n.

- If (m^n, s^n) is not jointly ϵ-typical, then an encoding error is declared;

- If (m^n, s^n) is jointly ϵ-typical, but no $v^n \in \bigcup_{i=1}^{t} C(i)$ such that (m^n, s^n, v^n) is jointly ϵ-typical, then an encoding error is declared;

- Assume (m^n, s^n) is jointly ϵ-typical, $C(i)$ is the first bin of C containing a vector v^n such that (m^n, s^n, v^n) is jointly ϵ-typical. Let $v^n(i,h)$ denote the first such a vector in $C(i)$. If no $u^n \in W(i)$ such that $(m^n, s^n, v^n(i,h), u^n)$ is jointly ϵ-typical, then an encoding error is declared;

- If $u^n(i,j) \in W(i)$ be the first vector such that $(m^n, s^n, v^n(i,h), u^n(i,j))$ is jointly ϵ-typical, then the encoder randomly generates a stegotext x^n according to $p(x^n|m^n, s^n, v^n(i,h), u^n(i,j))$.

- If an encoding error is declared, then define a fixed x_0^n as the stegotext.

Watermarking decoding: Fix codebooks C, W. Let y^n be an output of the attack channel with the input x^n when m^n is transmitted using s^n, $v^n(i,h) \in C(i)$ and $u^n(i,j) \in W(i)$.

- The decoder finds the first vector u^n in the codebook W, say $u^n(i_0, j_0) \in W(i_0)$, such that $(u^n(i_0, j_0), y^n)$ is jointly ϵ-typical with respect to $p(u, y)$;

- If no or more than one $u^n \in W$ are found such that (u^n, y^n) is jointly ϵ-typical, then a decoding error is declared;

- The decoder finds a vector $v^n(i_0, h_0) \in C(i_0)$ such that $(v^n(i_0, h_0), u^n(i_0, j_0), y^n)$ is jointly ϵ-typical with respect to $p(v, u, y)$;

- If no or more than one v^n are found in the bin $C(i_0)$ such that $(v^n, u^n(i_0, j_0), y^n)$ is jointly ϵ-typical, then a decoding error is declared.

- The decoder decodes

$$\hat{m}^n = (g(v_1(i_0,h_0), u_1(i_0,j_0), y_1), g(v_2(i_0,h_0), u_2(i_0,j_0), y_2), ..., g(v_n(i_0,h_0), u_n(i_0,j_0), y_n)),$$

  where $v_t(i,h), t = 1, 2..., n$ is the tth component of $v^n(i,j)$.

- If a decoding error is declared, then decode watermarks as a fixed vector $\hat{m}_0^n$.

4.2.2 Distortion Constraint for Watermarking Encoders

Let C, W be fixed codebooks generated as above. In this subsection, we shall analyze the distortion constraint for watermarking encoders averaged over watermarks M^n and

covertexts S^n, that is, we shall prove the distortion constraint for watermarking encoders is satisfied with high probability. To begin with, an event is defined by

$$B(C,W) = \{(m^n, s^n) : \text{ an encoding error is declared when embedding } m^n \text{ into } s^n \text{ via } C, W\}.$$

Then, one has

$$\begin{aligned}
\mathbf{E}_{M^n,S^n}[d(S^n,X^n)|C,W] &= \sum_{(m^n,s^n)\in B(C,W)} p(m^n,s^n) d(s^n,x_0^n) \\
&\quad + \sum_{(m^n,s^n)\in \bar{B}(C,W)} p(m^n,s^n)\mathbf{E}_{X^n} d(s^n,X^n) \\
&\leq \Pr\{B(C,W)\} d_{max} + \sum_{(m^n,s^n)\in \bar{B}(C,W)} p(m^n,s^n) \Pr\{A(m^n,s^n,v^n(i,h),u^n(i,j))\} d_{max} \\
&\quad + \sum_{(m^n,s^n)\in \bar{B}(C,W)} p(m^n,s^n) \sum_{x^n\in \bar{A}(m^n,s^n,v^n(i,h),u^n(i,j))} p(x^n|m^n,s^n,v^n(i,h),u^n(i,j)) d(s^n,x^n) \\
&\overset{(1)}{\leq} \sum_{(m^n,s^n)\in \bar{B}(C,W)} p(m^n,s^n) \sum_{x^n\in \bar{A}(m^n,s^n,v^n(i,h),u^n(i,j))} p(x^n|m^n,s^n,v^n(i,h),u^n(i,j)) d(s^n,x^n) \\
&\quad + \Pr\{B(C,W)\} d_{max} + \epsilon d_{max} \\
&\overset{(2)}{\leq} \Pr\{B(C,W)\} d_{max} + D + 2\epsilon d_{max},
\end{aligned}$$

where (1) follows from the fact that

$$\Pr\{A(m^n,s^n,v^n(i,h),u^n(i,j))\} \leq \epsilon$$

for sufficiently large n by the generation of x^n, here

$$A(m^n,s^n,v^n(i,h),u^n(i,j)) \overset{def}{=} \{x^n : (m^n,s^n,v^n(i,h),u^n(i,j),x^n) \text{ is not } \epsilon\text{-typical}\},$$

and (2) is due to

$$\begin{aligned}
d(s^n,x^n) &= \frac{1}{n}\sum_{i=1}^n d(s_i,x_i) \\
&\leq \sum_{(s,x)\in\mathcal{S}\times\mathcal{X}} \left[p(s,x) + \frac{\epsilon}{|\mathcal{S}||\mathcal{X}|}\right] d(s,x) \\
&\leq \mathbf{E}d(S,X) + \epsilon d_{max} \\
&\leq D + \epsilon d_{max}.
\end{aligned}$$

If we can show that $\Pr\{B(C,W)\} \le \epsilon$ with high probability of C and W, that is,

$$\Pr\{\Pr\{B(C,W)\} \le \epsilon\} \ge 1-\epsilon,$$

then one has

$$\begin{aligned}\mathbf{E}_{M^n,S^n}[d(S^n,X^n)|C,W] &\le \Pr\{B(C,W)\}d_{max} + D + 2\epsilon d_{max} \\ &\le D + 3\epsilon d_{max}\end{aligned}$$

with high probability of C, W.

In the following, we will estimate the probability $\Pr\{\Pr\{B(C,W)\} \le \epsilon\}$. Obviously, if we can show that

$$\mathbf{E}_{C,W} \Pr\{B(C,W)\} \le \epsilon^2, \tag{4.1}$$

then, by the Markov inequality,

$$\begin{aligned}\Pr\{\Pr\{B(C,W)\} \le \epsilon\} &= 1 - \Pr\{\Pr\{B(C,W)\} \ge \epsilon\} \\ &\ge 1 - \frac{\mathbf{E}_{C,W} \Pr\{B(C,W)\}}{\epsilon} \\ &\ge 1-\epsilon.\end{aligned}$$

So, next we shall show the inequality (4.1). Let E_0 be the set of all non ϵ-typical sequences (m^n, s^n). For each $(m^n, s^n) \notin E_0$, we define the following events:

- $E_1(m^n, s^n)$: no $v^n \notin C$ such that (v^n, m^n, s^n) is ϵ-typical;
- $E_2(m^n, s^n)$: no $u^n \in W(i)$ such that $(m^n, s^n, v^n(i,h), u^n)$ is ϵ-typical.

First, for $n > n_0$ sufficiently large,

$$\Pr\{E_0\} \le \frac{\epsilon^2}{3}. \tag{4.2}$$

For sufficiently large $n > n_1$, we have

$$\begin{aligned}\Pr\{E_1(m^n,s^n)|\bar{E}_0\} &\le (1-2^{-n[I(V;M,S)+\gamma/16]})^{2^{n[I(V;M,S)+\gamma/8]}} &(4.3)\\ &\le 2^{-2^{n\gamma/16}} &(4.4)\end{aligned}$$

double exponentially decreasing, since there exists n_1 such that for all $n > n_1$,

$$\Pr\{(V^n, m^n, s^n) \text{ is jointly } \epsilon\text{-typical } \} \geq 2^{-n[I(V;M,S)+\gamma/16]}.$$

Similarly, if $m^n, s^n, v^n(i,h)$ are jointly typical, then there exists n_2 such that for sufficiently large $n > n_2$, one has

$$\Pr\{(m^n, s^n, v^n(i,h), U^n) \text{ is jointly } \epsilon\text{-typical}\} > 2^{-n[I(U;M,S)+\gamma/8]}.$$

Thus, for all $n > n_2$

$$\begin{aligned} \Pr\{E_2(m^n, s^n)|\bar{E}_0 \cap \bar{E}_1(m^n, s^n)\} &\leq (1 - 2^{-n[I(U;M,S,V)+\gamma/8]})^{2^{n[I(U;M,S,V)+\gamma/4]}} \\ &\leq 2^{-2^{n\gamma/8}}, \end{aligned} \tag{4.5}$$

double exponentially decreasing.

Thus, for sufficiently large $n > \{n_0, n_1, n_2\}$

$$\begin{aligned} E_{C,W} \Pr\{B(C,W)\} &\leq \Pr\{E_0\} + \sum_{(m^n,s^n)\in\bar{E}_0} \Pr\{E_1(m^n, s^n) \cup E_2(m^n, s^n)\} \\ &\leq \frac{\epsilon^2}{3} + |\mathcal{M}^n||\mathcal{S}^n|2^{-2^{n\gamma/16}} + |\mathcal{M}^n||\mathcal{S}^n|2^{-2^{n\gamma/8}} \\ &\leq \frac{\epsilon^2}{3} + \frac{\epsilon^2}{3} + \frac{\epsilon^2}{3} \\ &= \epsilon^2. \end{aligned} \tag{4.6}$$

Therefore, we have shown that for sufficiently large n

$$\mathbf{E}_{M^n,S^n}[d(S^n, X^n)|C, W] \leq D + 3\epsilon d_{max} \tag{4.7}$$

with high probability of C, W.

4.2.3 Distortion Constraint for Watermark Decoders

In this subsection we shall analyze the averaged distortion constraint between transmitted watermarks and reproduced watermarks.

First, assume that encoding m^n into s^n via C, W successfully generates a stegotext x^n with jointly typical $(m^n, s^n, v^n(i,h), u^n(i,j), x^n)$. Let y^n be a forgery generated by the attacker, and define the following events:

- E: $(v^n(i,h), u^n(i,j), y^n)$ is not jointly ϵ-typical;
- E': more than one $u^n \in W$ such that (u^n, y^n) is ϵ-jointly typical; and
- E'': more that one $v^n \in C(i)$ such that $(v^n, u^n(i,j), y^n)$ is jointly ϵ-typical.

We shall upper bound the probability

$$\begin{aligned}
&\Pr\{E \cup E' \cup E'' | m^n, s^n, v^n(i,j), u^n(i,j)\} \\
&\leq \Pr\{E|m^n, s^n, v^n(i,j), u^n(i,j)\} + \Pr\{E' \cap \bar{E}|m^n, s^n, v^n(i,j), u^n(i,j)\} \\
&\quad + \Pr\{E'' \cap \bar{E}|m^n, s^n, v^n(i,j), u^n(i,j)\}. \qquad (4.8)
\end{aligned}$$

By the Markov Lemma, there exists a large number n_3 such that for all $n > n_3$,

$$\Pr\{E|m^n, s^n, v^n(i,j), u^n(i,j)\} < \epsilon. \qquad (4.9)$$

For the second term in right side of (4.8), one has

$$\begin{aligned}
&\Pr\{E' \cap \bar{E}|m^n, s^n, v^n(i,j), u^n(i,j)\} \\
&\leq \Pr\{(u^n, y^n) \text{ is } \epsilon\text{-typical for some } u^n \neq u^n(i,j), y^n \in A_\epsilon^{(n)}(Y)|m^n, s^n, v^n(i,h), u^n(i,j)\} \\
&= \sum_{y^n \in A_\epsilon^{(n)}(Y)} p(y^n|m^n, s^n, v^n(i,h), u^n(i,j))\theta(m^n, s^n, v^n(i,h), u^n(i,j), y^n), \qquad (4.10)
\end{aligned}$$

where

$$\begin{aligned}
&\theta(m^n, s^n, v^n(i,h), u^n(i,j), y^n) \\
&= \Pr\{(u^n, y^n) \text{ is jointly } \epsilon\text{-typical for some } u^n \neq u^n(i,j)|m^n, s^n, v^n(i,h), u^n(i,j), y^n\} \\
&\leq \sum_{u^n \in W(i'), i'=1,2,\ldots t, i' \neq i} \Pr\{(u^n, y^n) \text{ is jointly } \epsilon\text{-typical}|m^n, s^n, v^n(i,h), u^n(i,j), y^n\} \qquad (4.11) \\
&\quad + \Pr\{(u^n, y^n) \text{ is jointly } \epsilon\text{-typical, } u^n \in W(i) \text{ but } u^n \neq u^n(i,j)|m^n, s^n, v^n(i,h), u^n(i,j), y^n\}.
\end{aligned}$$

By the generation of W, if $u^n \in W(i')$ and $i' \neq i$, then u^n is independent of $(m^n, s^n, v^n(i,h), u^n(i,j), y^n)$ and the probability for large n $\Pr\{(u^n, y^n)$ is jointly ϵ-typical$\} < 2^{-n[I(U;Y)-\gamma/4]}$. Therefore, there exists a large number n_4 such that for all $n > n_4$, the summation in the right side of (4.11) is less or equal to

$$\begin{aligned}
&(2^{n[H(V;M,S)-I(M;U,Y)+\gamma/4]} - 1)2^{-n[I(U;Y)-\gamma/4]}2^{n[I(U;M,S,V)+\gamma/4]} \\
&\leq \quad 2^{-n[I(U;Y)-I(U;M,S,V)+I(M;U,Y)-H(V;M,S)-3\gamma/4]} \\
&= \quad 2^{-n\gamma/4} \leq \epsilon. \qquad (4.12)
\end{aligned}$$

As to the second term in the right side of (4.11), following the exact approach used in Chapter 2 yields

$$\Pr\{(u^n, y^n) \text{ is } \epsilon\text{-typical, } u^n \in W(i) \text{ but } u^n \neq u^n(i,j) | m^n, s^n, v^n(i,h), u^n(i,j), y^n\} \quad \leq \quad \epsilon.$$

Thus, one has for all $n > n_4$,

$$\Pr\{E' \cap \bar{E} | m^n, s^n, v^n(i,h), u^n(i,j)\} < \epsilon. \qquad (4.13)$$

We now upper bound the probability $\Pr\{E'' \cap \bar{E} | m^n, s^n, v^n(i,h), u^n(i,j)\}$. One has

$$\begin{aligned}
&\Pr\{E'' \cap \bar{E} | m^n, s^n, v^n(i,h), u^n(i,j)\} \\
&\leq \quad \Pr\{(v^n, u^n(i,j), y^n) \text{ is jointly } \epsilon\text{-typical, } v^n \in C(i), \text{ but } v^n \neq v^n(i,h), \\
&\qquad (u^n(i,j), y^n) \in A_\epsilon^{(n)}(U,Y) | m^n, s^n, v^n(i,h), u^n(i,j)\} \\
&= \quad \sum_{y^n : (u^n(i,j), y^n) \in A_\epsilon^{(n)}(U,Y)} p(y^n | m^n, s^n, v^n(i,h), u^n(i,j)) \eta(m^n, s^n, v^n(i,h), u^n(i,j), y^n),
\end{aligned}$$

here

$$
\begin{aligned}
&\eta(m^n, s^n, v^n(i,h), u^n(i,j), y^n) = \Pr\{(v^n, u^n(i,j), y^n) \text{ is jointly } \epsilon\text{-typical, } v^n \in C(i), \\
&\qquad \text{but } v^n \neq v^n(i,h) | m^n, s^n, v^n(i,h), u^n(i,j), y^n\}. \\
&= \sum_{l=1}^{|C(i)|} \Pr\{(v^n, u^n(i,j), y^n) \text{ is typical }, v^n \in C(i) \text{ but } v^n \neq v^n(i,h), \\
&\qquad h = l | m^n, s^n, v^n(i,h), u^n(i,j), y^n\} \\
&\leq \sum_{l=1}^{|C(i)|} \sum_{k=1, k\neq l}^{|C(i)|} \Pr\{(v^n(i,k), u^n(i,j), y^n) \text{ is typical }, h = l | m^n, s^n, v^n(i,h), u^n(i,j), y^n\} \\
&= \sum_{l=1}^{|C(i)|} \left[\sum_{k=l+1}^{|C(i)|} \Pr\{(v^n(i,k), u^n(i,j), y^n) \text{ is typical }, h = l | m^n, s^n, v^n(i,h), u^n(i,j), y^n\} \right. \\
&\qquad \left. + \sum_{k=1}^{l-1} \Pr\{(v^n(i,k), u^n(i,j), y^n) \text{ is typical }, h = l | m^n, s^n, v^n(i,h), u^n(i,j), y^n\} \right] \\
&\leq \sum_{l=1}^{|C(i)|} \left[2^{-n[I(V;U,Y)-\gamma/16]} |C(i)| \Pr\{h = l | m^n, s^n, v^n(i,h), u^n(i,j), y^n\} \right. \\
&\qquad + \sum_{k=1}^{l-1} \Pr\{(v^n(i,k), u^n(i,j), y^n) \text{ is typical }, (m^n, s^n, v^n(i,a)) \text{ is not typical }, \\
&\qquad \left. a = 1, 2, ..., l-1, a \neq k, (m^n, s^n, v^n(i,l)) \text{ is typical } | m^n, s^n, v^n(i,h), u^n(i,j), y^n\} \right] \\
&\leq \sum_{l=1}^{|C(i)|} 2^{-n[I(V;U,Y)-\gamma/16]} \left[2^{n[I(V;U,Y)-\gamma/8]} \Pr\{h = l | m^n, s^n, v^n(i,h), u^n(i,j), y^n\} \right. \\
&\qquad \left. + \sum_{k=1}^{l-1} \frac{\Pr\{v^n(i,l) \text{ is the first typical with } (m^n, s^n) | m^n, s^n, v^n(i,h), u^n(i,j), y^n\}}{\Pr\{(m^n, s^n, v^n(i,k)) \text{ is not typical } | m^n, s^n\}} \right] \\
&\leq \sum_{l=1}^{|C(i)|} 2^{-n[I(V;U,Y)-\gamma/16]} \left[2^{n[I(V;U,Y)-\gamma/8]} \Pr\{h = l | m^n, s^n, v^n(i,h), u^n(i,j), y^n\} \right. \\
&\qquad \left. + \sum_{k=1}^{l-1} \frac{\Pr\{h = l | m^n, s^n, v^n(i,h), u^n(i,j), y^n\}}{1 - 2^{-n[I(V;M,S)-\gamma/4]}} \right] \\
&\leq 2^{-n[I(V;U,Y)-\gamma/16]} 2^{n[I(V;U,Y)-\gamma/8]} \left[1 + \frac{1}{1 - 2^{-n[I(V;M,S)-\gamma/4]}} \right] \\
&\leq 2^{-n\gamma/16} \left[1 + \frac{1}{1 - 2^{-n[I(V;M,S)-\gamma/4]}} \right] \\
&\leq \epsilon
\end{aligned}
$$

for all large n. Thus, there exists n_5 such that for all numbers $n > n_5$,

$$\Pr\{E'' \cap \bar{E} | m^n, s^n, v^n(i,j), u^n(i,j)\} < \epsilon. \tag{4.14}$$

Therefore, by the watermarking encoding and decoding scheme, for $n > \max_{i=0..5}\{n_i\}$ we have

$$\begin{aligned}
&\mathbf{E}_{C,W}\mathbf{E}_{M^n,S^n} d_1(M^n, \hat{M}^n) = \mathbf{E}_{M^n,S^n}\mathbf{E}_{C,W} d_1(M^n, \hat{M}^n) \\
&\leq \Pr\{E_0\} d'_{max} + \sum_{(m^n,s^n)\in\bar{E}_0} p(m^n, s^n) \Pr\{E_1(m^n, s^n) \cup E_2(m^n, s^n)\} d'_{max} \\
&\quad + \sum_{(m^n,s^n)\in\bar{E}_0} p(m^n, s^n) \Pr\{A(m^n, s^n, v^n(i,h), u^n(i,j)) | \bar{E}_1(m^n, s^n) \cap \bar{E}_2(m^n, s^n)\} d'_{max} \\
&\quad + \sum_{(m^n,s^n)\in\bar{E}_0} p(m^n, s^n) \Pr\{E \cup E' \cup E'' | m^n, s^n, v^n(i,h), u^n(i,j)\} d'_{max} \\
&\quad + \sum_{(m^n,s^n)\in\bar{E}_0} p(m^n, s^n) [d_1(m^n, g(v^n(i,h), u^n(i,j), y^n)) | \bar{E} \cap \bar{E}' \cap \bar{E}''] \\
&\leq D_1 + \epsilon + 8\epsilon d'_{max},
\end{aligned} \tag{4.15}$$

where the last inequality follows from the above analysis on encoding and decoding error probabilities and from the fact that $(m^n, v^n(i,h), u^n(i,j), y^n)$ are jointly ϵ-typical with respect to $p(m,v,u,y)$ and $\mathbf{E}d_1(M, g(V,U,Y)) \leq D_1$.

4.2.4 Existence of Watermarking Encoders and Decoders

Define

$$\Gamma = \{(C,W) : \mathbf{E}_{M^n,S^n}[d(S^n, X^n) | C, W] \leq D + 3\epsilon d_{max}\}, \tag{4.16}$$

then from (4.7), one has $\Pr\{\Gamma\} \geq 1 - \epsilon$.

So, from (4.15) one has

$$\begin{aligned}
&\sum_{(C,W)\in\Gamma} \Pr(C,W) \mathbf{E}_{M^n,S^n}[d_1(M^n, \hat{M}^n) | C, W] \\
&\quad \leq \mathbf{E}_{C,W}\mathbf{E}_{M^n,S^n}[d_1(M^n, \hat{M}^n) | C, W] \\
&\quad \leq D_1 + \epsilon + 8\epsilon d'_{max}.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\sum_{(C,W)\in\Gamma} \frac{\Pr(C,W)}{\Pr\{\Gamma\}} \mathbf{E}_{M^n,S^n}[d_1(M^n,\hat{M}^n)|C,W] \\
&= \frac{1}{\Pr\{\Gamma\}} \sum_{(C,W)\in\Gamma} \Pr(C,W)\mathbf{E}_{M^n,S^n}[d_1(M^n,\hat{M}^n)|C,W] \\
&\leq \frac{D_1+\epsilon+8\epsilon d'_{max}}{1-\epsilon} = D_1+\epsilon' \qquad (4.17)
\end{aligned}$$

for some small number $\epsilon' > 0$ and $\epsilon' \to 0$ as $\epsilon \to 0$.

Combination of (4.16) and (4.17) yields the existence of watermarking encoder and watermarking decoder for each sufficiently large n such that averaged distortion for watermark encoder is less than $D + 3\epsilon d_{max}$ and the averaged distortion for watermarks is less than $D_1 + \epsilon'$. Thus the probability $p(m, s)$ is publicly admissible with respect to D and D_1.

The proof of Theorem 4.1 is completed.

□

4.3 Proof of Theorem 4.2

In this section we shall prove Theorem 4.2, that is, if there exists $(V, U, X) \in F(D, D_1)$ such that

$$\begin{cases} H(M) < R_c - I(M,S,V;U,X) + I(V;U,Y) + H(M,S|V) - H(S|M) \\ H(M) < I(U;Y) - I(U;M,S,V) + I(V;U,Y) + H(M,S|V) - H(S|M), \end{cases} \qquad (4.18)$$

then, there exist watermarking encoder and public watermarking decoder pairs (f_n, g_n) for all sufficiently large n such that $\mathbf{E}d(S^n, f_n(M^n, S^n)) < D + \epsilon$, $H(f_n(M^n, S^n))/n \leq R_c$ and $\mathbf{E}d_1(M^n, g_n(Y^n)) < D_1 + \epsilon$

Obviously, (4.18) is equivalent to

$$I(V;M,S) - I(V;U,Y) < \min\{R_c - I(M,S,V;U,X), I(U;Y) - I(U;M,S,V)\}.$$

Let $\epsilon > 0$ be an arbitrarily small but fixed number, g be a function such that

$$\mathbf{E}d_1(M, g(V, U, Y)) \leq D_1,$$

and denote $\gamma \stackrel{\triangle}{=} \min\{R_c - I(M, S, V; U, X), I(U; Y) - I(U; M, S, V)\} - I(V; M, S) + I(V; U, Y)$.

4.3.1 Watermarking Coding Scheme

Random Codes Generation: Random codebooks C, W and G will be generated as follows.

- Generate identically and independently $\exp(n[I(V; M, S) + \gamma/8])$ vectors $v^n \in \mathcal{V}^n$ according to the probability distribution $p(v)$ of the random V derived from the joint probability distribution $p(m, s, v, u, x, y)$ of (M, S, V, U, X, Y), then uniformly distribute them into $t \stackrel{\triangle}{=} \exp(n[I(V; M, S) - I(V; U, Y) + \gamma/4])$ bins $C(i)$, $i = 1, 2, ..., t$, each bin containing $\exp(n[I(V; U, Y) - \gamma/8])$ vectors v^n. Denote the random codebook by $C = \{C(i)\}_{i=1}^t$.

- Generate identically and independently $\exp(n[I(V; M, S) - I(V; U, Y) + I(U; M, S, V) + \gamma/2])$ vectors $u^n \in \mathcal{U}^n$ according to the probability distribution $p(u)$ of the random U derived from the joint probability distribution $p(m, s, v, u, x, y)$ of (M, S, V, U, X, Y), then uniformly distribute them into t bins $W(i)$, $i = 1, 2, ..., t$, each bin containing $\exp(n[I(U; M, S, V) + \gamma/4])$ vectors u^n. Denote the random codebook by $W = \{W(i)\}_{i=1}^t$.

- For each vector $u^n(i, j) \in W(i)$, $i = 1, ..., t$, $j = 1, 2, ..., \exp(n[I(U; M, S, V) + \gamma/4])$, generate a bin of vectors $G(i, j) = \{x^n(i, j, l) : l = 1, 2, ..., \exp(n[I(M, S, V; X|U) + \gamma/4])\}$, each $x^n(i, j, l)$ is generated identically and independently according to the probability $p(x^n|u^n(i, j))$. Denote the random codebook by $G = \{G(i, j), i = 1, ..., t, j = 1, 2, ..., \exp(n[I(U; M, S, V) + \gamma/4])\}$.

- The two codebooks C and W are then distributed to the watermarking decoder, and the codebook G is sent to the lossless stegotext decompressor.

Watermarking encoding: Fix codebooks C, W and G. Given a watermark m^n and a covertext s^n.

- If (m^n, s^n) is not jointly ϵ-typical, then an encoding error is declared;
- If (m^n, s^n) is jointly ϵ-typical, but no $v^n \in \cup_{i=1}^{t} C(i)$ such that (m^n, s^n, v^n) is jointly ϵ-typical, then an encoding error is declared;
- If (m^n, s^n) is jointly ϵ-typical and $C(i)$ is the first bin of C containing a vector v^n such that (m^n, s^n, v^n) is jointly ϵ-typical and $v^n(i,h) \in C(i)$ is the first such a vector v^n, but no $u^n \in W(i)$ such that $(m^n, s^n, v^n(i,h), u^n)$ is jointly ϵ-typical, then an encoding error is declared;
- If (m^n, s^n) is jointly ϵ-typical and $C(i)$ is the first bin of C containing a vector v^n such that (m^n, s^n, v^n) is jointly ϵ-typical, $v^n(i,h) \in C(i)$ is the first such a vector, and $u^n(i,j) \in W(i)$ is the first vector such that $(m^n, s^n, v^n(i,h), u^n(i,j))$ is jointly ϵ-typical, but no vector $x^n(i,j,l) \in G(i,j)$ is found such that $(m^n, s^n, v^n(i,h), u^n(i,j), x^n(i,j,l))$ is jointly ϵ-typical, then an encoding error is declared;
- If (m^n, s^n) is jointly ϵ-typical, $C(i)$ is the first bin of C containing a vector v^n such that (m^n, s^n, v^n) is jointly ϵ-typical, $v^n(i,h) \in C(i)$ is the first such a vector, and $u^n(i,j) \in W(i)$ is the first vector such that $(m^n, s^n, v^n(i,h), u^n(i,j))$ is jointly ϵ-typical, then the encoder finds the first vector $x^n(i,j,l) \in G(i,j)$ such that $(m^n, s^n, v^n(i,h), u^n(i,j), x^n(i,j,l))$ is jointly ϵ-typical, and $x^n(i,j,l)$ is the stegotext;
- If an encoding error is declared, then define a fixed x_0^n as the stegotext.

Watermarking decoding: Fix codebooks C, W, G. Let y^n be a forgery received by the watermarking decoder when m^n is transmitted using s^n, $v^n(i,h) \in C(i)$, $u^n(i,j) \in W(i)$ and $x^n(i,j,l) \in G(i,j)$.

- The decoder finds the first vector u^n in the codebook W, say $u^n(i_0, j_0) \in W(i_0)$, such that $(u^n(i_0, j_0), y^n)$ is jointly ϵ-typical with respect to $p(u, y)$;

- If no $u^n \in W$ or more than one are found such that (u^n, y^n) is jointly ϵ-typical, then a decoding error is declared;

- The decoder finds the first vector $v^n(i_0, h_0) \in C(i_0)$ such that $(v^n(i_0, h_0), u^n(i_0, j_0), y^n)$ is jointly ϵ-typical with respect to $p(v, u, y)$;

- If no v^n or more than one are found in the bin $C(i_0)$ such that $(v^n, u^n(i_0, j_0), y^n)$ is jointly ϵ-typical, then a decoding error is declared;

- The decoder decodes

$$\hat{m}^n = (g(v_1(i_0, h_0), u_1(i_0, j_0), y_1), g(v_2(i_0, h_0), u_2(i_0, j_0), y_2), ..., g(v_n(i_0, h_0), u_n(i_0, j_0), y_n)).$$

- If a decoding error is declared, then decode watermarks as a fixed $\hat{m}_0^n$

4.3.2 Distortion Constraint for Watermarking Encoders

Let C, W and G be fixed codebooks generated as above, we shall analyze the distortion constraint for watermarking encoders averaged watermarks M^n and covertexts S^n. Define

$$B(C, W, G) = \{(m^n, s^n) : \text{ an encoding error is declared when encoding } m^n \text{ via } s^n, C, W, G\}.$$

Then, one has

$$\begin{aligned} \mathbf{E}_{M^n, S^n}[d(S^n, X^n)|C, W, G] &= \sum_{(m^n, s^n) \in B(C,W,G)} p(m^n, s^n) d(s^n, x_0^n) \\ &\quad + \sum_{(m^n, s^n) \in \bar{B}(C,W,G)} p(m^n, s^n) d(s^n, x^n) \\ &\leq \Pr\{B(C, W, G)\} d_{max} + D + \epsilon d_{max}, \end{aligned}$$

since $(m^n, s^n) \in \bar{B}(C, W, G)$ is jointly typical by the watermark encoding scheme, and

$$\begin{aligned} d(s^n, x^n) &= \frac{1}{n}\sum_{i=1}^{n} d(s_i, x_i) \\ &\leq \sum_{(s,x)\in\mathcal{S}\times\mathcal{X}} \left[p(s,x) + \frac{\epsilon}{|\mathcal{S}||\mathcal{X}|}\right] d(s,x) \\ &\leq \mathbf{E}d(S,X) + \epsilon d_{max} \\ &\leq D + \epsilon d_{max}, \end{aligned}$$

for sufficiently large n.

If we can show that with high probability of C, W and G, $\Pr\{B(C,W,G)\} \leq \epsilon$, that is,

$$\Pr\{\Pr\{B(C,W,G)\} \leq \epsilon\} \geq 1 - \epsilon,$$

then

$$\begin{aligned} \mathbf{E}_{M^n,S^n}[d(S^n, X^n)|C, W, G] &\leq \Pr\{B(C,W,G)\}d_{max} + D + \epsilon d_{max} \\ &\leq D + 2\epsilon d_{max} \end{aligned}$$

with high probability of C, W and G.

Therefore, we only need to estimate the probability $\Pr\{\Pr\{B(C,W,G)\} \leq \epsilon\}$. Obviously, if we can show that

$$\mathbf{E}_{C,W,G}\Pr\{B(C,W,G)\} \leq \epsilon^2, \tag{4.19}$$

then, by the Markov inequality,

$$\begin{aligned} \Pr\{\Pr\{B(C,W,G)\} \leq \epsilon\} &= 1 - \Pr\{\Pr\{B(C,W,G)\} \geq \epsilon\} \\ &\geq 1 - \frac{\mathbf{E}_{C,W,G}\Pr\{B(C,W,G)\}}{\epsilon} \\ &\geq 1 - \epsilon. \end{aligned}$$

So, in the following we shall show the inequality (4.19). Let E_0 be the set of all non ϵ-typical sequences (m^n, s^n). For each $(m^n, s^n) \notin E_0$, define the following events:

- $E_1(m^n, s^n)$: no $v^n \notin C$ such that (v^n, m^n, s^n) is ϵ-typical;

- $E_2(m^n, s^n)$: $C(i)$ is the first bin in the random codebook C containing a vector v^n such that (v^n, m^n, s^n) is ϵ-typical and $v^n(i,h)$ denotes the first such a vector, but no $u^n \in W(i)$ such that $(m^n, s^n, v^n(i,h), u^n)$ is ϵ-typical;

- $E_3(m^n, s^n)$: $v^n(i,h)$ is the first vector in $C(i)$ such that $(v^n(i,h), m^n, s^n)$ is ϵ-typical, $u^n(i,j)$ is the first vector in $W(i)$ such that $(m^n, s^n, v^n(i,h), u^n(i,j))$ is ϵ-typical, but no $x^n \in G(i,j)$ is found such that $(m^n, s^n, v^n(i,h), u^n(i,j), x^n)$ is ϵ-typical.

By employing the approach used in Subsection 4.2.2, we can show that there exists a large number n_1 such that for all $n > n_1$,

$$\begin{aligned} \Pr\{E_0\} &\leq \frac{\epsilon^2}{4} && (4.20)\\ \Pr\{E_1(m^n, s^n)|\bar{E}_0\} &\leq 2^{-2^{n\gamma/16}} && (4.21)\\ \Pr\{E_2(m^n, s^n)|\bar{E}_0, \bar{E}_1(m^n, s^n)\} &\leq 2^{-2^{n\gamma/8}}. && (4.22) \end{aligned}$$

If $m^n, s^n, v^n(i,h), u^n(i,j)$ are jointly typical, then, there exists a large number n_2 such that for all $n > n_2$,

$$\Pr\{(m^n, s^n, v^n(i,h), u^n(i,j), x^n) \text{ is jointly } \epsilon\text{-typical}\} > 2^{-n[I(X;M,S,V|U)+\gamma/8]}.$$

Thus, for all $n > n_2$

$$\begin{aligned} &\Pr\{E_3(m^n, s^n)|v^n, m^n, s^n, u^n(i,j)\} && (4.23)\\ &= \Pr\{\text{no } x^n \in G(i,j) \text{ such that } (m^n, s^n, v^n(i,h), u^n(i,j), x^n) \text{ is jointly } \epsilon\text{-typical}\}\\ &\leq (1 - 2^{-n[I(X;M,S,V|U)+\gamma/8]})^{2^{n[I(X;M,S,V|U)+\gamma/4]}} && (4.24)\\ &\leq 2^{-2^{n\gamma/8}} \end{aligned}$$

double exponentially decreasing.

Thus, for sufficiently large $n > \max\{n_1, n_2\}$,

$$\begin{aligned}
\mathbf{E}_{C,W} \Pr\{B(C,W)\} &\leq \Pr\{E_0\} + \sum_{(m^n,s^n)\in\bar{E}_0} \Pr\{E_1(m^n,s^n) \cup E_2(m^n,s^n) \cup E_3(m^n,s^n)\} \\
&\leq \frac{\epsilon^2}{4} + |\mathcal{M}^n||\mathcal{S}^n|2^{-2^{n\gamma/16}} + |\mathcal{M}^n||\mathcal{S}^n|2^{-2^{n\gamma/8}} + |\mathcal{M}^n||\mathcal{S}^n|2^{-2^{n\gamma/8}} \\
&\leq \frac{\epsilon^2}{4} + \frac{\epsilon^2}{4} + \frac{\epsilon^2}{4} + \frac{\epsilon^2}{4} \\
&= \epsilon^2. \qquad (4.25)
\end{aligned}$$

Therefore, we have shown that with high probability of C, W and G,

$$\mathbf{E}_{M^n,S^n}[d(S^n, X^n)|C, W, G] \leq D + 2\epsilon d_{max}. \qquad (4.26)$$

4.3.3 Compression Rate Constraint for Watermarking Encoders

By the construction of the codebook G, it is obvious that

$$\begin{aligned}
\frac{H(X^n)}{n} &\leq \frac{1}{n}\log|G| \\
&\leq (I(V;M,S) - I(V;U,Y) + \gamma/4) + (I(M,S,V;X|U) + \gamma/4) \\
&\quad +(I(M,S,V;U) + \gamma/4) \\
&= I(V;M,S) - I(V;U,Y) + I(M,S,V;U,X) + 3\gamma/4 \\
&\leq R_c - \gamma/4 \\
&\leq R_c \qquad (4.27)
\end{aligned}$$

since

$$I(V;M,S) - I(V;U,Y) + I(M,S,V;X,U) \leq R_c - \gamma$$

by the definition of γ. Thus, the compression rate constraint is satisfied for all watermarking encoders.

4.3.4 Averaged Distortion Constraint for Watermarks

From the random coding scheme, we can see that the watermarking decoder is exactly the same as the decoder in the case without compression of stegotexts introduced in Subsection 4.2.1. Therefore, by employing the same method of Subsection 4.2.3, we have

$$\mathbf{E}_{C,W,G}\mathbf{E}_{M^n,S^n}d_1(M^n,\hat{M}^n) \;\leq\; D_1+\epsilon+8\epsilon d'_{max}. \tag{4.28}$$

4.3.5 Existence of Watermarking Encoders and Decoders

Define

$$\Gamma=\{(C,W,G):\mathbf{E}_{M^n,S^n}[d(S^n,X^n)|C,W,G]\leq D+2\epsilon d_{max}\}, \tag{4.29}$$

then from (4.26), one has $\Pr\{\Gamma\}\geq 1-\epsilon$.

So, from (4.28) one has

$$\begin{aligned}
&\sum_{(C,W,G)\in\Gamma}\Pr(C,W,G)\mathbf{E}_{M^n,S^n}[d_1(M^n,\hat{M}^n)|C,W,G]\\
&\leq\; \mathbf{E}_{C,W,G}\mathbf{E}_{M^n,S^n}[d_1(M^n,\hat{M}^n)|C,W,G]\\
&\leq\; D_1+\epsilon+8\epsilon d'_{max}.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\sum_{(C,W,G)\in\Gamma}\frac{\Pr(C,W,G)}{\Pr\{\Gamma\}}\mathbf{E}_{M^n,S^n}[d_1(M^n,\hat{M}^n)|C,W,G]\\
&=\; \frac{1}{\Pr\{\Gamma\}}\sum_{(C,W,G)\in\Gamma}\Pr(C,W,G)\mathbf{E}_{M^n,S^n}[d_1(M^n,\hat{M}^n)|C,W,G]\\
&\leq\; \frac{D_1+\epsilon+8\epsilon d'_{max}a}{1-\epsilon}=D_1+\epsilon'
\end{aligned} \tag{4.30}$$

for some small number $\epsilon'>0$ and $\epsilon'\to 0$ as $\epsilon\to 0$.

Combination of (4.29) and (4.30) and (4.27) yields the existence of a joint compression and watermarking encoder with distortion level $D+2\epsilon d_{max}$ and compression rate R_c, and

a public watermarking decoder with distortion $D_1 + \epsilon'$ for each sufficiently large n. Thus the probability $p(m, s)$ is publicly admissible with respect to D and D_1 and compression rate R_c.

The proof of Theorem 4.2 is completed.

□

4.4 Summary

In this chapter the models of digital watermarking in Chapter 2 and in Chapter 3 but with a relaxed constraint on recovery of watermarks are investigated respectively. Under the assumption that the decoded watermark has tolerant distortion with respect to the transmitted watermark, sufficient conditions are given for the case without compression of stegotexts and the case with joint compression and watermarking, under which the transmitting watermarks to public watermark receivers is reliable after stegotexts are disturbed by a fixed attack channel.

Chapter 5

Closed-Forms of Private Watermarking Capacities for Laplacian Sources

It is well known that watermarking capacities and compression and watermarking rate regions of joint compression and watermarking systems can be expressed as optimization problems in information-theoretic quantities [20, 21, 25, 26, 32, 33]. However, this characterization does not mean that watermarking capacities and joint compression and watermarking rate regions can be calculated easily. So far, closed-form formulas for watermarking capacities are known only for watermarking systems with independent and identically distributed (iid) binary covertexts and Gaussian covertexts [26, 7], and closed-form formulas for compression and watermarking rate regions of joint compression and watermarking systems are known only for private watermarking systems with independent and identically distributed Gaussian covertexts [19]. In this chapter, private watermarking systems with iid Laplacian covertexts are investigated and nice closed-forms of watermarking capacities are determined. The motivation to study such watermarking systems is that, in most applications, source data such as transformed coefficients of image signals can be more or less

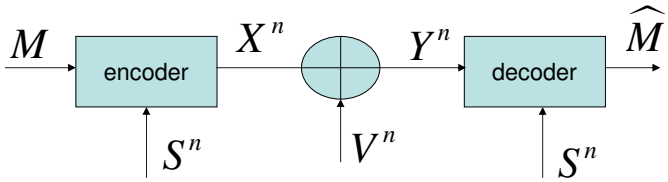

Figure 5.1: Model of private Laplacian watermarking systems

modeled as Laplacian sources and many digital watermarking schemes are implemented in frequency domain instead of space domain.

5.1 Setting of Watermarking Models and Main Results

The model of private Laplacian watermarking systems studied in this chapter is depicted in Figure 5.1, where $S^n \in \mathbb{R}^n$ is an Laplacian covertext generated independently and identically by a memoryless source S with Laplacian density function $p(s) = \frac{1}{2\alpha} e^{-|s|/\alpha}$, $\alpha > 0$, and watermark M is a random variable uniformly distributed over the set $\{1, 2, ..., e^{nR}\}$, $R \geq 0$. In this model, the magnitude-error distortion measure is employed, $d(s, x) = |x - s|$. A watermarking encoder of length n with rate R and distortion D maps (M, S^n) to $X^n = (X_1, X_2, ..., X_n) \in \mathbb{R}^n$ such that $\mathbf{E}d(S^n, X^n) = 1/n \sum_{i=1}^n d(S_i, X_i) \leq D$. An attacker uses an additive iid noise vector $V^n = (V_1, V_2, ..., V_n) \in \mathbb{R}^n$ generated by a real-valued random variable V to disturb the stegotext X^n and generates a forgery $Y^n \in \mathbb{R}^n$, that is, $Y^n = X^n + V^n$. Finally, a private watermarking decoder produces an estimate of a watermark, $\hat{M}$, from Y^n with the help of S^n.

A number $R \geq 0$ is achievable with respect to a distortion level D if for any small number $\epsilon > 0$, there exist, for sufficiently large n, a watermarking encoder of length n with rate $R - \epsilon$ and distortion $D + \epsilon$ and a private decoder such that $\Pr\{\hat{M} \neq M\} < \epsilon$. The

private watermarking capacity $C(D)$ of the private watermarking model is defined to be the maximum of all achievable embedding rates R with respect to the distortion level D.

It is well known from [5, 7, 26, 32] that the private watermarking capacity of the model in this chapter is given by

$$C(D) = \max_{\mathbf{E}d(S,X)\leq D} I(X;Y|S) \tag{5.1}$$

where $Y = X+V$, $I(X;Y|S)$ is the conditional mutual information between X and Y given S, and the maximization is taken over all random variables X such that $\mathbf{E}d(S,X) \leq D$. The aim of this chapter is to determine a closed-form of $C(D)$.

Unless otherwise specified, in this chapter all logarithms are with respect to base e and the upper and lower limits of all integrals are ∞ and $-\infty$, respectively. Now we are ready to give our main results.

Theorem 5.1 *Let V be a Laplacian random variable with the density function $g(x) = \frac{1}{2d}e^{-\frac{|x|}{d}}$, $d > 0$. Then, the private watermarking capacity $C(D)$ of the iid Laplacian watermarking system with respect to the distortion level D and under an additive Laplacian noise(ALN) V^n is given by*

$$C(D) = \log\left(1 + \frac{2d + D - \sqrt{D^2+4d^2}}{\sqrt{D^2+4d^2} - D}\right). \tag{5.2}$$

Theorem 5.2 *Let V be a real-valued random variable with the density function $g(x)$. Then, the private watermarking capacity $C(D)$ of the iid Laplacian watermarking system with respect to the distortion level D and under an additive noise(AN) V^n is given by*

$$\begin{aligned} C(D) = G(0)\sqrt{2\pi}\left[\log\int e^{-\frac{\lambda_0 l(x)}{\pi\sqrt{2\pi}}}dx - \log(G(0)\sqrt{2\pi})\right] \\ + \frac{\lambda_0 G(0)}{\pi\int e^{-\frac{\lambda_0 l(x)}{\pi\sqrt{2\pi}}}dx}\int l(x)e^{-\frac{\lambda_0 l(x)}{\pi\sqrt{2\pi}}}dx + \int g(x)\log g(x)dx, \end{aligned}$$

where $G(t)$ is the Fourier transform of $g(x)$, $l(x) = \int \frac{e^{-ixt}}{t^2 G(t)} dt$ and $\lambda_0 < 0$ satisfies

$$\pi D \int e^{-\frac{\lambda_0 l(x)}{\pi\sqrt{2\pi}}} dx = -G(0) \int l(-x) e^{-\frac{\lambda_0 l(x)}{\pi\sqrt{2\pi}}} dx.$$

Particularly, if $g(x)$ is even, then the watermarking capacity is

$$\begin{aligned} C(D) &= -\lambda_0 D + G(0)\sqrt{2\pi}\left[\log \int e^{-\frac{\lambda_0 l(x)}{\pi\sqrt{2\pi}}} dx - \log(G(0)\sqrt{2\pi})\right] \\ &\quad + \int g(x) \log g(x) dx. \end{aligned}$$

Discussion:

- Under an additive Laplacian attack, the capacity given in Theorem 5.1 has a very nice closed formula, which is independent of the parameter α of the Laplacian source, and determined only by the distortion level D and the parameter d of the Laplacian attack random variable.

- Actually, the watermarking capacity $C(D)$ in (5.2) can be simplified to be $C(D) = \log \frac{D+\sqrt{D^2+4d^2}}{2d}$. Thus, $C(D) \simeq \log \frac{D}{d}$ if $D \gg d$, and $C(D) \simeq 0$ if $D \ll d$.

- For the Laplacian attack random variable V with parameter d, the variance is $\sigma^2 = 2d^2$. So, $C(D) = \log \frac{D+\sqrt{D^2+2\sigma^2}}{\sqrt{2}\sigma}$ in terms of D and σ^2.

- It is well known that the watermarking capacity of a Gaussian watermarking system with the mean square distortion measure and under a fixed Gaussian attack with variance σ^2 is $C_G(D) = \log \frac{\sqrt{D+\sigma^2}}{\sigma}$. Therefore,

  - if $\sigma^2 \gg D$, the watermarking capacity of a Laplacian system under an additive Laplacian attack with variance σ^2 is almost equal to that of a Gaussian system under an additive Gaussian attack with the variance σ^2;
  - If $\sigma^2 \ll D$, the watermarking capacity of a Laplacian system under an additive Laplacian attack with variance σ^2 is larger than that of a Gaussian system under an additive Gaussian attack with the variance σ^2 and the difference is $\log(2D)/2$.

- For $D < 0.5$, solving $\frac{D+\sqrt{D^2+2\sigma^2}}{\sqrt{2}\sigma} = \frac{\sqrt{D+\sigma^2}}{\sigma}$ yields $\sigma^2 = 1/2 - D$. So,
  - if $\sigma^2 < 1/2 - D$, then the capacity of a Laplacian system under a Laplacian attack with variance σ^2 is bigger than that of a Gaussian system under a Gaussian attack with variance σ^2;
  - if $\sigma^2 > 1/2 - D$, then the capacity of a Laplacian system under a Laplacian attack with variance σ^2 is smaller than that of a Gaussian system under a Gaussian attack with variance σ^2.
- To determine a closed form of watermarking capacity with an arbitrary additive noise attack, one only needs to solve an equation to get the parameter λ_0.

5.2 Watermarking Capacities Under Additive Laplacian Noise Attacks

Let V be a real random variable with density function $g(x)$ and independent of all other random variables. Then, from (5.1) and the model specified in Figure 5.1, the private watermarking capacity under the additive attack V^n is given by

$$\begin{aligned} C(D) &= \max_{X:\mathbf{E}|S-X|\le D} I(X;Y|S) \\ &= \max_{X:\mathbf{E}|S-X|\le D} [H(Y|S) - H(Y|X,S)] \\ &= \max_{X:\mathbf{E}|S-X|\le D} [H(X+V|S) - H(V)] \\ &= \max_{T:\mathbf{E}|T|\le D} H(T+V|S) - H(V) \\ &= \max_{T:\mathbf{E}|T|\le D} H(T+V) - H(V) \\ &= \max_{T:\mathbf{E}|T|= D} H(T+V) - H(V). \end{aligned} \tag{5.3}$$

To obtain $C(D)$, we first compute $\max_{\mathbf{E}|T|=D} H(T+V)$ using the method of Lagrange multipliers. Let $f(\cdot)$ be the density function of a real-valued random variable T, and define

a functional for $\mu, \lambda < 0$ and $f(\cdot)$,

$$\begin{aligned}\Delta(f(\cdot), \lambda, \mu) &= \int_{x'} \left(\int_{t'} f(t')g(x'-t')dt' \right) \log \left(\int_{t'} f(t')g(x'-t')dt' \right) dx' \\ &\quad -\lambda \left(\int_{x'} |x'| f(x')dx' - D \right) - \mu \left(\int_{x'} f(x')dx' - 1 \right).\end{aligned}$$

Then

$$\begin{aligned}\frac{\partial \triangle}{\partial f(x)} &= \int_{x'} g(x'-x) \log \int_{t'} f(t')g(x'-t')dt'dx' + \int_{x'} g(x'-x)dx' - \lambda|x| - \mu \\ &= \int_{x'} g(x'-x) \log h(x')dx' - \lambda|x| - \mu + 1\end{aligned}$$

where

$$h(x) = \int_{t'} f(t')g(x-t')dt'.$$

Let $\frac{\partial \Delta}{\partial f(x)} = 0$. Then for any $x \in \mathbb{R}$

$$\int_{x'} g(x'-x) \log h(x')dx' = \lambda|x| + \mu - 1. \tag{5.4}$$

Let $G(t)$ and $H(t)$ be Fourier transforms of $g(x)$ and $\log h(x)$, respectively. Then, by the Fourier Convolution Theorem and (5.4), we have

$$\begin{aligned}\int_t G(t)H(t)e^{-ixt}dt &= \int_{x'} g(x'-x) \log h(x')dx' \\ &= \lambda|x| + \mu - 1, \forall x. \end{aligned} \tag{5.5}$$

By solving the integral equation (5.5), one has

$$\begin{aligned}G(t)H(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} (\lambda|x| + \mu - 1)e^{ixt}dx \\ &= -\frac{\lambda}{t^2\pi} + (\mu - 1)Dirac(t),\end{aligned} \tag{5.6}$$

where $Dirac(\cdot)$ is the unit impulse function, that is,

$$Dirac(t) = \begin{cases} \lim_{\epsilon \to} \frac{1}{\epsilon}, & -\epsilon/2 \le t \le \epsilon/2, \\ 0 & \text{otherwise.} \end{cases}$$

Note (5.6) holds for any additive attack V.

In the following of this section, we assume V is a Laplacian random variable with density function $g(x) = \frac{1}{2d}e^{-\frac{|x|}{d}}$, then the Fourier transform of $g(x)$ is

$$\begin{aligned} G(t) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g(x)e^{ixt}dx \\ &= \frac{1}{\sqrt{2\pi}(1+t^2d^2)}. \end{aligned}$$

Thus, from (5.6) we have

$$H(t) = \sqrt{2\pi}(1+t^2d^2)\left[-\frac{\lambda}{t^2\pi} + (\mu-1)Dirac(t)\right],$$

and applying the inverse Fourier transform to $H(t)$, it is not hard to obtain

$$\log h(x) = \pi(2\lambda x Heaviside(x) + \mu - 1 - 2d^2\lambda Dirac(x) - \lambda x),$$

where $Heaviside(x)$ is the unit step function, that is,

$$Heaviside(x) = \begin{cases} 1, & x \geq 0, \\ 0 & x \leq 0. \end{cases}$$

By the definition of $h(x)$ and the Fourier Convolution Theorem, we get

$$\begin{aligned} h(x) &= \int_{-\infty}^{\infty} G(t)F(t)e^{-ixt}dt \\ &= e^{\pi(2\lambda x Heaviside(x)+\mu-1-2d^2\lambda Dirac(x)-\lambda x)}, \end{aligned} \tag{5.7}$$

where $F(t)$ is the Fourier Transform of $f(x)$. Solving the integral equation (5.7),

$$G(t)F(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} h(x)e^{ixt}dx = -\frac{\lambda}{t^2+\pi^2\lambda^2}e^{\pi(\mu-1)},$$

so

$$F(t) = -\frac{\lambda}{t^2+\pi^2\lambda^2}\sqrt{2\pi}(1+t^2d^2)e^{\pi(\mu-1)}.$$

Using the inverse Fourier transform, we obtain

$$\begin{aligned} f(x) &= -\lambda e^{\pi(\mu-1)} \int_{-\infty}^{\infty} \frac{1+t^2d^2}{t^2+\pi^2\lambda^2} e^{-ixt} dt \\ &= -e^{\pi(\mu-1)}[e^{-\lambda\pi x}(\pi^2\lambda^2d^2-1)Heaviside(-x) \\ &\quad +e^{\lambda\pi x}(\pi^2\lambda^2d^2-1)Heaviside(x) \\ &\quad +2\pi\lambda d^2 Dirac(x)]. \end{aligned} \tag{5.8}$$

Since the density function $f(x)$ must satisfy the constraints $\int_x f(x)dx = 1$ and $\int_x |x|f(x)dx = D$, that is,

$$\begin{cases} -\frac{2}{\pi\lambda} e^{\pi(\mu-1)} &= 1 \\ -\frac{2(d^2\lambda^2\pi^2-1)}{\pi^2\lambda^2} e^{\pi(\mu-1)} &= D \end{cases}$$

one has

$$\begin{cases} \lambda &= \frac{D-\sqrt{D^2+4d^2}}{2\pi d^2}, \\ \mu &= 1+\frac{1}{\pi}\log\frac{\sqrt{D^2+4d^2}-D}{4d^2}. \end{cases} \tag{5.9}$$

Now we get the optimal real-valued random variable T with the density function $f(x)$ in (5.8) with (λ, μ) of (5.9). For this optimal $f(x)$, it is not hard to obtain the entropy of $T+V$

$$H(T+V) = 1 - \log\frac{\sqrt{D^2+4d^2}-D}{4d^2}.$$

Therefore, by (5.3) and $H(V) = 1+\log(2d)$,

$$\begin{aligned} C(D) &= 1 - \log\frac{\sqrt{D^2+4d^2}-D}{4d^2} - H(V) \\ &= -\log\frac{\sqrt{D^2+4d^2}-D}{4d^2} - \log(2d) \\ &= \log\left(1+\frac{2d+D-\sqrt{D^2+4d^2}}{\sqrt{D^2+4d^2}-D}\right). \end{aligned}$$

The proof of Theorem 5.1 is completed.

5.3 Watermarking Capacities Under Additive Noise Attacks

In this section, we assume the additive noise V^n is generated iid by a real-valued random variable V with density function $g(x)$. Then, from (5.6), we get

$$H(t) = \frac{\mu-1}{G(t)} Dirac(t) - \frac{\lambda}{\pi t^2 G(t)},$$

where $G(t)$ is the Fourier transform of $g(x)$.

Since $H(t)$ is the Fourier transform of $\log h(x)$, then by the inverse Fourier transform,

$$\begin{aligned}
\log h(x) &= \frac{1}{\sqrt{2\pi}} \int H(t) e^{-ixt} dt \\
&= \frac{\mu-1}{\sqrt{2\pi}} \int \frac{Dirac(t)}{G(t)} e^{-ixt} dt - \frac{\lambda}{\pi\sqrt{2\pi}} \int \frac{e^{-ixt}}{t^2 G(t)} dt \\
&= \frac{\mu-1}{\sqrt{2\pi} G(0)} - \frac{\lambda}{\sqrt{2\pi}\pi} l(x),
\end{aligned}$$

where

$$l(x) = \int \frac{e^{-ixt}}{t^2 G(t)} dt. \tag{5.10}$$

Let $F(t)$ be the Fourier transform of $f(x)$. Then

$$\begin{aligned}
F(t) &= \frac{1}{2\pi G(t)} \int h(x) e^{ixt} dx \\
&= \frac{1}{2\pi G(t)} e^{\frac{\mu-1}{\sqrt{2\pi} G(0)}} \int e^{-\frac{\lambda}{\sqrt{2\pi}\pi} l(x) + ixt} dx
\end{aligned}$$

by the Fourier Convolution Theorem. Applying the inverse Fourier transform to $F(t)$ yields

$$f(x) = \frac{e^{\frac{\mu-1}{\sqrt{2\pi} G(0)}}}{2\pi\sqrt{2\pi}} \int\int \frac{1}{G(t)} e^{-\frac{\lambda l(x_1)}{\pi\sqrt{2\pi}} + ix_1 t - ixt} dx_1 dt. \tag{5.11}$$

Since $\int f(x)dx = 1$ and $\int |x|f(x)dx = D$, that is,

$$\begin{aligned}\int f(x)dx &= \frac{e^{\frac{\mu-1}{\sqrt{2\pi}G(0)}}}{2\pi\sqrt{2\pi}}\int\int\left(\int e^{-ixt}dx\right)\frac{1}{G(t)}e^{-\frac{\lambda l(x_1)}{\pi\sqrt{2\pi}}+ix_1t}dx_1dt\\ &= \frac{e^{\frac{\mu-1}{\sqrt{2\pi}G(0)}}}{2\pi\sqrt{2\pi}}\int\int 2\pi Dirac(t)\frac{1}{G(t)}e^{-\frac{\lambda l(x_1)}{\pi\sqrt{2\pi}}+ix_1t}dx_1dt\\ &= \frac{e^{\frac{\mu-1}{\sqrt{2\pi}G(0)}}}{\sqrt{2\pi}G(0)}\int e^{-\frac{\lambda}{\pi\sqrt{2\pi}}l(x)}dx = 1\end{aligned}$$

and

$$\begin{aligned}\int |x|f(x)dx &= \frac{e^{\frac{\mu-1}{\sqrt{2\pi}G(0)}}}{2\pi\sqrt{2\pi}}\int\int\left(\int |x|e^{-ixt}dx\right)\frac{1}{G(t)}e^{-\frac{\lambda l(x_1)}{\pi\sqrt{2\pi}}+ix_1t}dx_1dt\\ &= \frac{e^{\frac{\mu-1}{\sqrt{2\pi}G(0)}}}{2\pi\sqrt{2\pi}}\int\int\left(\frac{-2}{t^2}\right)\frac{1}{G(t)}e^{-\frac{\lambda l(x_1)}{\pi\sqrt{2\pi}}+ix_1t}dx_1dt\\ &= -\frac{e^{\frac{\mu-1}{\sqrt{2\pi}G(0)}}}{\sqrt{2\pi}\pi}\int l(-x)e^{-\frac{\lambda}{\pi\sqrt{2\pi}}l(x)}dx = D,\end{aligned}$$

we obtain

$$\begin{cases}\lambda &= \lambda_0\\ \mu &= 1+\sqrt{2\pi}G(0)\left[\log(G(0)\sqrt{2\pi}) - \log\int e^{-\frac{\lambda_0 l(x)}{\pi\sqrt{2\pi}}}dx\right],\end{cases} \tag{5.12}$$

where $\lambda_0 < 0$ satisfies

$$\pi D\int e^{-\frac{\lambda_0 l(x)}{\pi\sqrt{2\pi}}}dx = -G(0)\int l(-x)e^{-\frac{\lambda_0 l(x)}{\pi\sqrt{2\pi}}}dx. \tag{5.13}$$

For this optimal random variable T with the density function $f(x)$ determined by (5.11), (5.12) and (5.13), we can calculate the entropy

$$\begin{aligned}H(T+V) &= G(0)\sqrt{2\pi}\left[\log\int e^{-\frac{\lambda_0 l(x)}{\pi\sqrt{2\pi}}}dx - \log(G(0)\sqrt{2\pi})\right]\\ &\quad + \frac{\lambda_0 G(0)}{\pi\int e^{-\frac{\lambda_0 l(x)}{\pi\sqrt{2\pi}}}dx}\int l(x)e^{-\frac{\lambda_0 l(x)}{\pi\sqrt{2\pi}}}dx.\end{aligned} \tag{5.14}$$

In particular, if $g(x)$ is even, then $G(t)$ is even, so $l(x)$ is. Thus, by (5.13),

$$\begin{aligned}\int l(x)e^{-\frac{\lambda_0 l(x)}{\pi\sqrt{2\pi}}}dx &= \int l(-x)e^{-\frac{\lambda_0 l(x)}{\pi\sqrt{2\pi}}}dx\\ &= -\frac{D\pi}{G(0)}\int e^{-\frac{\lambda_0 l(x)}{\pi\sqrt{2\pi}}}dx,\end{aligned}$$

and the last term in (5.14) is simplified to be $-\lambda_0 D$. In view of (5.3), the proof of Theorem 5.2 is finished.

5.4 Summary

Calculation of watermarking capacities of private Laplacian watermarking systems with the magnitude-error distortion measure under additive attacks is addressed in this chapter. First, in the case of an additive Laplacian attack, a nice closed-form formula for the capacities is derived, which involves only the distortion level and the parameter of the Laplacian attack. Second, in the case of an arbitrary additive attack, a general, but slightly more complicated formula for the capacities is given.

Chapter 6

Algorithms for Computing Joint Compression and Private Watermarking Rate Regions

Compression and watermarking rate regions of joint compression and watermarking systems can be characterized as optimization problems in information theoretic quantities. However, calculation of these optimizations problems is not straightforward. In this chapter numerical algorithms are developed for computing compression and watermarking rate regions of joint compression and private watermarking systems, and algorithms for computing joint compression and public watermarking rate regions are given in the next chapter.

6.1 Introduction

From an information-theoretic viewpoint, a major research problem on joint compression and watermarking is to determine best tradeoffs among the distortion between covertexts and stegotexts, the watermarking embedding rate, the compression rate and the robustness of stegotexts. Karakos [19, 18] determined the best tradeoffs for joint compression

and private watermarking systems with finite alphabets and with Gaussian covertexts, respectively. Maor and Merhav [20, 21] obtained the best tradeoff for joint compression and public watermarking systems with discrete alphabets. These results were extended to the case of abstract alphabets in [45]. Mathematically, for a joint compression and watermarking system the best tradeoff among distortion between covertexts and stegotexts, watermarking rate, compression rate and robustness of stegotexts can be formulated as an optimization problem. Unfortunately, the optimization problem is often difficult to solve. As a result, numerical algorithms are needed for calculating the tradeoff efficiently.

On the other hand, in the context of communication systems, in order to numerically compute channel capacities of memoryless channels and rate-distortion functions of memoryless sources, an efficient iterative algorithm was invented by Blahut [2] and Arimoto [1] independently, and its convergence was proved rigorously by Csiszar [13]. Since then, extensive studies on its generalization to other scenarios have been conducted. For instance, Chang and Davisson [4] generalized the Blahut-Arimoto algorithm to the case with continuous alphabets; Dupuis and Yu and Willems [15, 42] modified the Blahut-Arimoto algorithm to calculate capacities of Gel'fand-Pinsker channels and rate-distortion functions of Wyner-Ziv sources. Other generalizations can be found in [22, 29, 30, 37] and reference therein. However, due to the different setting up for joint compression and watermarking systems, none of these existing generalized Blahut-Arimoto algorithms is applicable for numerical calculation of joint compression and watermarking rate regions.

In this chapter, based on the Blahut-Arimoto algorithm we will present two iterative algorithms that can be combined to efficiently and numerically determine the compression and watermarking rate region of a joint compression and private watermarking system with finite alphabets. Algorithm A is used for determining numerically the private watermarking capacity, and Algorithm B is for computing the compression rate function with respect to watermarking rate and distortion.

6.2 Formulation of Joint Compression and Private Watermarking Rate Regions

For a joint compression and private watermarking system designated in Figure 3.1, we can define a privately achievable pair (R_w, R_c) with respect to D as in Definition 3.3. The set of all privately achievable rate pairs (R_w, R_c) with respect to D is called a **compression and watermarking rate region** with respect to D. To facilitate the description of the region, we define a **compression rate function** with respect to watermarking rate R_w and distortion D as

$$R_c(D, R_w) \stackrel{def}{=} \inf R_c,$$

where the inf is taken over all privately achievable pairs (R_w, R_c) with respect to D and fixed R_w. Obviously, the compression rate function with respect to R_w and D determines the best tradeoff among watermarking rate, compression rate, distortion and robustness of the watermarking system. It is well known [19, 20, 45] that for a joint compression and private watermarking system with a memoryless finite covertext source S and under a fixed attack $p(y|x)$,

$$R_c(D, R_w) = R_w + \min_{\substack{p(x|s):\mathbf{E}_{d(S,X)\le D} \\ R_w \le I(X;Y|S)}} I(S;X), \tag{6.1}$$

where $D \ge 0$, $0 \le R_w \le C(D)$, the private watermarking capacity, and

$$C(D) \stackrel{def}{=} \max_{p(x|s):\mathbf{E}_{d(S,X)\le D}} I(X;Y|S).$$

Now we analyze (6.1) in much more details. For a given $D \ge 0$, if

$$R(D) \stackrel{def}{=} \min_{p(x|s):\mathbf{E}_{d(S,X)\le D}} I(S;X),$$

the rate-distortion function of the source S, is achieved by a conditional probability $p^*(x|s)$, then, for $0 \le R_w \le I_{p^*(x|s)}(X;Y|S)$,

$$R_c(D, R_w) = R_w + R(D),$$

a linear function of R_w; in the range of $I_{p^*(x|s)}(X;Y|S) \leq R_w \leq C(D)$, $R_c(D,R_w)$ is a curve of R_w. Since the rate-distortion function $R(D)$ and $p^*(x|s)$ can be easily calculated by the Blahut-Arimoto algorithm, it is sufficient to develop algorithms for computing the private watermarking capacity $C(D)$ and $R_c(D,R_w)$ for $I_{p^*(x|s)}(X;Y|S) \leq R_w \leq C(D)$ in order to describe the compression rate function with respect to the watermarking rate R_w and distortion D, or equivalently, the joint compression and private watermarking rate region.

6.3 Algorithm A for Computing Private Watermarking Capacities

In this section, we present an algorithm (hereafter referred to as Algorithm A) for computing private watermarking capacities $C(D)$. Algorithm A is similar to the Blahut-Arimoto algorithm for computation of channel capacities with constrained inputs [2], but it has more complicated constraints and objective function. Specifically, Algorithm A is to compute $\max_{p(x|s):\mathbf{E}d(S,X)\leq D} I(X;Y|S)$, while the Blahut-Arimoto algorithm is to calculate $\max_{p(x):\mathbf{E}e(X)\leq D} I(X;Y)$ in which $e(x)$ is a function of the channel input x. Before describing Algorithm A and showing its convergence, some properties are given.

6.3.1 Properties of $C(D)$

Proposition 6.1 *The private watermarking capacity* $C(D) = \max_{p(x|s):\mathbf{E}d(S,X)\leq D} I(X;Y|S)$ *is nondecreasing, concave and continuous in* $D \geq 0$.

Proposition 6.2

$$C(D) \;=\; \lambda D + \max_{p(x|s)} \left[I(X;Y|S) - \lambda \mathbf{E}d(S,X) \right], \tag{6.2}$$

for $\lambda \geq 0$, *where* $D = \sum_{s,x} p(s)p^*(x|s)d(s,x)$ *and* $p^*(x|s)$ *achieves the maximum in (6.2).*

Proposition 6.3 *For $\lambda \geq 0$ and two probability distributions $p(x|s)$, $Q(x|s,y) > 0$, define*

$$\begin{aligned} L(p(x|s), Q(x|s,y)) &= \sum_{s,x,y} p(s)p(x|s)p(y|x) \log \frac{Q(x|s,y)}{p(x|s)} \\ &\quad -\lambda \sum_{s,x} p(s)p(x|s)d(s,x). \end{aligned}$$

Then (a).

$$C(D) = \lambda D + \max_{p(x|s)} \max_{Q(x|s,y)} L(p(x|s), Q(x|s,y)); \tag{6.3}$$

(b). For fixed $p(x|s)$, the optimal probability distributions $Q^*(x|s,y)$ to maximize $L(p(x|s), Q(x|s,y))$ is given by

$$Q^*(x|s,y) = \frac{p(x|s)p(y|x)}{\sum_{x'} p(x'|s)p(y|x')}; \tag{6.4}$$

(c). For fixed $Q(x|s,y)$, the optimal $p^*(x|s)$ to maximize $L(p(x|s), Q(x|s,y))$ is given by

$$p^*(x|s) = \frac{\exp\left(\sum_y p(y|x) \log Q(x|s,y) - \lambda d(s,x)\right)}{\sum_{x'} \exp\left(\sum_y p(y|x') \log Q(x'|s,y) - \lambda d(s,x')\right)}. \tag{6.5}$$

Moreover,

$$L(p^*(x|s), Q(x|s,y)) = \sum_s p(s) \log \sum_x \exp\left(\sum_y p(y|x) \log Q(x|s,y) - \lambda d(s,x)\right). \tag{6.6}$$

Proof: (a) For a given probability distribution $p(x|s)$, let

$$Q^*(x|s,y) = \frac{p(x|s)p(y|x)}{\sum_{x'} p(x'|s)p(y|x')}.$$

Then, for any probability distribution $Q(x|s,y)$ of X given s and y,

$$\begin{aligned} & L(p(x|s), Q^*(x|s,y)) - L(p(x|s), Q(x|s,y)) \\ &= \sum_{s,x,y} p(s)p(x|s)p(y|x) \left[\log \frac{Q^*(x|s,y)}{p(x|s)} - \log \frac{Q(x|s,y)}{p(x|s)}\right] \\ &= \sum_{s,x,y} p(s)p(x|s)p(y|x) \log \frac{Q^*(x|s,y)}{Q(x|s,y)} \\ &\geq 0 \end{aligned}$$

by the log-sum inequality. So by Proposition 6.2,

$$C(D) = \lambda D + \max_{p(x|s)} \max_{Q(x|s,y)} L(p(x|s), Q(x|s,y)).$$

(b) is obvious from (a).

(c) For fixed $Q(x|s,y)$, one has

$$\begin{aligned}
\frac{\partial L(p(x|s), Q(x|s,y))}{\partial p(x|s)} &= \sum_y \frac{\partial \left(\sum_{s',x'} p(s')p(x'|s')p(y|x') \log \frac{Q(x'|s',y)}{p(x'|s')} \right)}{\partial p(x|s)} - \lambda p(s) d(s,x) \\
&= \sum_y p(s)p(y|x) \log \frac{Q(x|s,y)}{p(x|s)} - \sum_y \sum_{s',x'} p(s')p(y|x') \frac{\partial p(x'|s')}{\partial p(x|s)} - \lambda p(s) d(s,x) \\
&= \sum_y p(s)p(y|x) \log \frac{Q(x|s,y)}{p(x|s)} - p(s) - \lambda p(s) d(s,x) \\
&= p(s) \sum_y p(y|x) \log Q(x|s,y) - p(s) \log p(x|s) - p(s) - \lambda p(s) d(s,x).
\end{aligned}$$

Assume $p(s) > 0$ for all s. By the Karush-Kuhn-Tucker(KKT) conditions and $\sum_x p(x|s) = 1$ for all s, if $p^*(x|s) > 0$, then it is not hard to get

$$p^*(x|s) = \frac{\exp(\sum_y p(y|x) \log Q(x|s,y) - \lambda d(s,x))}{\sum_{x'} \exp(\sum_y p(y|x') \log Q(x'|s,y) - \lambda d(s,x'))},$$

which is optimal to maximize $L(p(x|s), Q(x|s,y))$ for fixed $Q(x|s,y)$.

For the fixed $Q(x|s,y)$ and $p^*(x|s)$ in (6.5), it is easy to obtain

$$\begin{aligned}
L(p^*(x|s), Q(x|s,y)) &= \sum_s p(s) \left[\sum_{x,y} p^*(x|s)p(y|x) \log \frac{Q(x|s,y)}{p^*(x|s)} - \lambda \sum_x p^*(x|s) d(s,x) \right] \\
&= \sum_s p(s) \left[\sum_{x,y} p^*(x|s)p(y|x) \log Q(x|s,y) - \sum_x p^*(x|s) \log p^*(x|s) - \lambda \sum_x p^*(x|s) d(s,x) \right] \\
&= \sum_s p(s) \left\{ \sum_{x,y} p^*(x|s)p(y|x) \log Q(x|s,y) - \sum_x p^*(x|s) \left[\sum_y p(y|x) \log Q(x|s,y) - \lambda d(s,x) \right] \right. \\
&\quad \left. + \log \sum_{x'} \exp \left(\sum_y p(y|x') \log Q(x'|s,y) - \lambda d(s,x') \right) - \lambda \sum_x p^*(x|s) d(s,x) \right\} \\
&= \sum_s p(s) \log \sum_x \exp \left(\sum_y p(y|x) \log Q(x|s,y) - \lambda d(s,x) \right).
\end{aligned}$$

□

Corollary 6.1 *Probability distributions $p(x|s)$ and $Q(x|s,y)$ achieve the private watermarking capacity $C(D)$ in (6.3) if and only if they satisfy*

$$Q(x|s,y) = \frac{p(x|s)p(y|x)}{\sum_{x'} p(x'|s)p(y|x')}, \tag{6.7}$$

$$p(x|s) = \frac{\exp(\sum_y p(y|x)\log Q(x|s,y) - \lambda d(s,x))}{\sum_{x'}\exp(\sum_y p(y|x')\log Q(x'|s,y) - \lambda d(s,x'))}. \tag{6.8}$$

6.3.2 Algorithm A

Fix a small number $\epsilon > 0$.

Step 1. Choose $Q^{(1)}(x|s,y) > 0$ arbitrarily for any (s,y);

Step 2. Define

$$a^{(n)}(s,x) = \exp\left[\sum_y p(y|x)\log Q^{(n-1)}(x|s,y) - \lambda d(s,x)\right].$$

Compute

$$p^{(n)}(x|s) = \frac{a^{(n)}(s,x)}{\sum_{x'} a^{(n)}(s,x')}, \tag{6.9}$$

and

$$Q^{(n)}(x|s,y) = \frac{p^{(n)}(x|s)p(y|x)}{\sum_{x'} p^{(n)}(x'|s)p(y|x')}; \tag{6.10}$$

Step 3. If

$$\sum_s p(s)\left(\log\max_x a^{(n)}(s,x) - \log\sum_x p^{(n)}(x|s)a^{(n)}(s,x)\right) < \epsilon$$

then stop the iteration and get the optimal probability distributions $p^{(n)}(x|s)$ and $Q^{(n)}(x|s,y)$ achieving the private watermarking capacity $C(D)$; otherwise, go to Step 2 and continue the iteration.

The termination condition is similar to that of [2], so the proof is omitted here.

Remarks: After the author completed the thesis-writing, Frans Willems told him that Algorithm A developed here would be identical to that developed by Blahut for computing

channel capacity with input expense constraint if Shannon's strategies were used. To see this, let $T(\cdot)$, mapping from $\mathcal{S}$ into $\mathcal{X}$, be any strategy. Then

$$\begin{aligned} I(X;Y|S) &= I(T;S) + I(X;Y|S) \\ &= I(T;S) + I(T;Y|S) \\ &= I(T;Y,S) \end{aligned}$$

since T is independent of S. Moreover, the cost $e(t)$ of a strategy t is

$$e(t) = \sum_s p(s)d(s, x = t(s)),$$

the transition probailities

$$p(y, s|t) = p(s)p(y|x = t(s)),$$

and all joint probabilities can be realized in this way. Now, it is obvious that Algorithm A is identical to Blahut's algorithm for computing the channel capacity with input expense constraint.

6.3.3 Convergence of Algorithm A

Let $p(x|s)$ and $Q(x|s, y)$ be probability distributions achieving the private watermarking capacity $C(D)$, and $p^{(n)}(x|s)$ and $Q^{(n)}(x|s, y)$ be defined in (6.9) and (6.10). For $p(x|s)$ and $p^{(n)}(x|s)$, define

$$p(y|s) = \sum_x p(x|s)p(y|x),$$

$$p^{(n)}(y|s) = \sum_x p^{(n-1)}(x|s)p(y|x).$$

Then it is easy to obtain

$$\sum_{x'} \exp\left(\sum_{y'} p(y'|x') \log Q(x'|s, y') - \lambda d(s, x')\right) = \frac{p(y|x) \exp\left(\sum_{y'} p(y'|x) \log Q(x|s, y') - \lambda d(s, x)\right)}{p(y|s)Q(x|s, y)}$$

and

$$\sum_{x'} \exp\left(\sum_{y'} p(y'|x') \log Q^{(n-1)}(x'|s,y') - \lambda d(s,x')\right)$$
$$= \frac{p(y|x) \exp\left(\sum_{y'} p(y'|x) \log Q^{(n-1)}(x|s,y') - \lambda d(s,x)\right)}{p^{(n)}(y|s) Q^{(n)}(x|s,y)}.$$

Therefore, we can obtain

$$\begin{aligned}
& L(p(x|s), Q(x|s,y)) - L(p^{(n)}(x|s), Q^{(n-1)}(x|s,y)) \\
= & \sum_s p(s) \log \sum_{x'} \exp\left(\sum_{y'} p(y'|x') \log Q(x'|s,y') - \lambda d(s,x')\right) \\
& - \sum_s p(s) \log \sum_{x'} \exp\left(\sum_{y'} p(y'|x') \log Q^{(n-1)}(x'|s,y') - \lambda d(s,x')\right) \\
= & \sum_{s,x,y} p(s)p(x|s)p(y|x) \log \frac{Q(x|s,y)}{Q^{(n-1)}(x|s,y)} + \sum_{s,x,y} p(s)p(x|s)p(y|x) \log \frac{p^{(n)}(y|s)}{p(y|s)} \\
& + \sum_{s,x,y} p(s)p(x|s)p(y|x) \log \frac{Q^{(n)}(x|s,y)}{Q(x|s,y)} \\
= & \sum_{s,y} p(s,y) \left[D(Q(x|s,y)||Q^{(n-1)}(x|s,y)) - D(Q(x|s,y)||Q^{(n)}(x|s,y))\right] \\
& - \sum_s p(s) D(p(y|s)||p^{(n)}(y|s)),
\end{aligned}$$

where $D(p(x)||q(x))$ is the divergence between $p(x)$ and $q(x)$. Since $p(x|s), Q(x|s,y)$ achieve the private watermarking capacity, one has

$$\begin{aligned}
0 \leq & \sum_{s,y} p(s,y)[D(Q(x|s,y)||Q^{(n-1)}(x|s,y)) - D(Q(x|s,y)||Q^{(n)}(x|s,y))] \\
& - \sum_s p(s) D(p(y|s)||p^{(n)}(y|s)),
\end{aligned}$$

which implies that

$$\begin{aligned}
& \sum_{s,y} p(s,y)[D(Q(x|s,y)||Q^{(n-1)}(x|s,y)) - D(Q(x|s,y)||Q^{(n)}(x|s,y))] \\
\geq & \sum_s p(s) D(p(y|s)||p^{(n)(y|s)}) \\
\geq & 0.
\end{aligned}$$

Thus, for any $N > 1$,

$$\begin{aligned} 0 &\leq \sum_{n=2}^{N}[L(p(x|s), Q(x|s,y)) - L(p^{(n)}(x|s), Q^{(n-1)}(x|s,y))] \\ &\leq \sum_{n=2}^{N}\sum_{s,y} p(s,y)[D(Q(x|s,y)||Q^{(n-1)}(x|s,y)) - D(Q(x|s,y)||Q^{(n)}(x|s,y))] \\ &= \sum_{s,y} p(s,y)[D(Q(x|s,y)||Q^{(1)}(x|s,y)) - D(Q(x|s,y)||Q^{(N)}(x|s,y))] \\ &\leq \sum_{s,y} p(s,y)D(Q(x|s,y)||Q^{(1)}(x|s,y)) < \infty, \end{aligned}$$

which yields

$$L(p^{(n)}(x|s), Q^{(n-1)}(x|s,y)) \to L(p(x|s), Q(x|s,y))$$

as $n \to \infty$.

The proof that $p^n(x|s) \to p(x|s)$ and $Q^n(x|s,y) \to Q(x|s,y)$ is similar to that in algorithm B and omitted here.

6.4 Algorithm B for Computing Compression Rate Functions

In this section, an iterative algorithm, Algorithm B, will be developed to calculate the compression rate function with respect to watermarking rate R_w and distortion D given in (6.1), for $I_{p^*(x|s)}(X;Y|S) \leq R_w \leq C(D)$ and $D \geq 0$. As in the Section 6.3, properties of the compression rate functions will be introduced, and followed by the description of Algorithm B and the proof of its convergence.

6.4.1 Properties of Compression Rate Functions

Proposition 6.4 *$R_c(D, R_w)$ is non-increasing in $D \geq 0$ and non-decreasing in $R_w \geq 0$.*

Proposition 6.5 *$R_c(D, R_w)$ is convex in (D, R_w).*

Proof: Let $(D_1, R_w^{(1)})$ and $(D_2, R_w^{(2)})$ be two points, and $p_1(x|s)$ and $p_2(x|s)$ the probability distributions achieving $R_c(D_i, R_w^{(i)})$ for $i = 1, 2$.

Let $\lambda \in [0, 1]$, and define

$$p^*(x|s) = \lambda p_1(x|s) + (1-\lambda)p_2(x|s).$$

Then

$$\sum_{s,x} p(s)p^*(x|s)d(s,x) \leq \lambda D_1 + (1-\lambda)D_2,$$

and

$$\begin{aligned}
I_{p^*}(X;Y|S) &= \sum_{s,x,y} p(s)p^*(x|s)p(y|x)\log\frac{p(y|x)}{p^*(y|s)} \\
&= \sum_{s,x,y} p(s)\left[\lambda p_1(x|s) + (1-\lambda)p_2(x|s)\right]p(y|x)\log\frac{p(y|x)}{p^*(y|s)} \\
&= \lambda\sum_{s,x,y} p(s)p_1(x|s)p(y|x)\log\frac{p(y|x)}{p^*(y|s)} + (1-\lambda)\sum_{s,x,y} p(s)p_2(x|s)p(y|x)\log\frac{p(y|x)}{p^*(y|s)} \\
&\overset{(1)}{\geq} \lambda\sum_{s,x,y} p(s)p_1(x|s)p(y|x)\log\frac{p(y|x)}{p_1(y|s)} + (1-\lambda)\sum_{s,x,y} p(s)p_2(x|s)p(y|x)\log\frac{p(y|x)}{p_2(y|s)} \\
&= \lambda I_{p_1}(X;Y|S) + (1-\lambda)I_{p_2}(X;Y|S) \\
&\geq \lambda R_w^{(1)} + (1-\lambda)R_w^{(2)},
\end{aligned}$$

where (1) holds since

$$\sum_{s,x,y} p(s)p_i(x|s)p(y|x)\log\frac{p(y|x)}{p^*(y|s)} \geq \sum_{s,x,y} p(s)p_i(x|s)p(y|x)\log\frac{p(y|x)}{p_i(y|s)}$$

by the log-sum inequality, and

$$p^*(y|s) = \sum_x p^*(x|s)p(y|x),$$

$$p_i(y|s) = \sum_x p_i(x|s)p(y|x).$$

So

$$
\begin{aligned}
&R_c(\lambda D_1 + (1-\lambda)D_2, \lambda R_w^{(1)} + (1-\lambda)R_w^{(2)}) \le \lambda R_w^{(1)} + (1-\lambda)R_w^{(2)} + I_{p^*}(S;X) \\
&= \lambda R_w^{(1)} + (1-\lambda)R_w^{(2)} + \sum_{s,x} p(s)[\lambda p_1(x|s) + (1-\lambda)p_2(x|s)] \log \frac{p^*(s|x)}{p(s)} \\
&= \lambda R_w^{(1)} + \lambda \sum_{s,x} p(s)p_1(x|s) \log \frac{p^*(s|x)}{p(s)} + (1-\lambda)R_w^{(2)} + (1-\lambda) \sum_{s,x} p(s)p_2(x|s) \log \frac{p^*(s|x)}{p(s)} \\
&\overset{(2)}{\le} \lambda \left(R_w^{(1)} + \sum_{s,x} p(s)p_1(x|s) \log \frac{p_1(s|x)}{p(s)} \right) + (1-\lambda) \left(R_w^{(2)} + \sum_{s,x} p(s)p_2(x|s) \log \frac{p_2(s|x)}{p(s)} \right) \\
&= \lambda R_c(D_1, R_w^{(1)}) + (1-\lambda) R_c(D_2, R_w^{(2)}),
\end{aligned}
$$

where

$$
\begin{aligned}
p^*(s|x) &= \frac{p(s)p^*(x|s)}{\sum_{s_1} p(s_1)p^*(x|s)}, \\
p_i(s|x) &= \frac{p(s)p_i(x|s)}{\sum_{s_1} p(s_1)p_i(x|s)},
\end{aligned}
$$

and (2) holds since

$$
\begin{aligned}
&\sum_{s,x} p(s)p_i(x|s) \log \frac{p^*(s|x)}{p(s)} - \sum_{s,x} p(s)p_i(x|s) \log \frac{p_i(s|x)}{p(s)} \\
&= \sum_{s,x} p(s)p_i(x|s) \log \frac{p^*(s|x)}{p_i(s|x)} \\
&= \sum_{s,x} p_i(x)p_i(s|x) \log \frac{p^*(s|x)}{p_i(s|x)} \le 0
\end{aligned}
$$

by the log-sum inequality. Thus, $R_c(D, R_w)$ is convex in (D, R_w). □

Proposition 6.6 *For $\lambda \le 0$ and $\gamma \ge 0$, one has*

$$
R_c(D, R_w) = \lambda D + (1+\gamma)R_w + \min_{p(x|s)} [I(S;X) - \lambda \mathbf{E} d(S,X) - \gamma I(X;Y|S)],
$$

where

$$
D = \mathbf{E}_{p^*} d(S,X), R_w = I_{p^*}(X;Y|S)
$$

and $p^*(x|s)$ achieves the above minimum.

Proposition 6.7 *Fix $\lambda \leq 0, \gamma \geq 0$. For probability distributions $p(x|s), Q(x), Q(x|s,y)$, define*

$$J(p(x|s), Q(x), Q(x|s,y)) \stackrel{def}{=} (1+\gamma) \sum_{s,x} p(s)p(x|s) \log p(x|s)$$
$$- \sum_{s,x} p(s)p(x|s) \left(\log Q(x) + \lambda d(s,x)\right) - \gamma \sum_{s,x,y} p(s)p(x|s)p(y|x) \log Q(x|s,y).$$

Then (a).

$$R_c(D, R_w) = \lambda D + (1+\gamma)R_w + \min_{p(x|s)} \min_{\{Q(x), Q(x|s,y)\}} J(p(x|s), Q(x), Q(x|s,y)),$$

where $D = \mathbf{E}_{p^*(x|s)} d(S,X)$, $R_w = I_{p^*(x|s)}(X;Y|S)$ and $p^*(x|s)$ achieves the above minimum.

(b). For fixed $p(x|s)$, the optimal $Q^*(x)$ and $Q^*(x|s,y)$ are given by

$$\begin{aligned} Q^*(x) &= \sum_s p(s)p(x|s), \\ Q^*(x|s,y) &= \frac{p(x|s)p(y|x)}{\sum_{x'} p(x'|s)p(y|x')}. \end{aligned}$$

(c). For fixed $Q(x)$ and $Q(x|s,y)$, the optimal $p^*(x|s)$ is given by

$$p^*(x|s) = \frac{b(s,x)}{\sum_{x'} b(s,x')},$$

where

$$b(s,x) = \exp\left[\frac{1}{1+\gamma}\left(\log Q(x) + \lambda d(s,x) + \gamma \sum_y p(y|x) \log Q(x|s,y)\right)\right].$$

Moreover, the minimum of $J(p(x|s), Q(x), Q(x|s,y))$ for fixed $Q(x)$ and $Q(x|s,y)$ is equal to

$$J(p^*(x|s), Q(x), Q(x|s,y)) = -(1+\gamma) \sum_s p(s) \log \sum_{x'} b(s,x').$$

Proof: (a) and (b). For a fixed probability $p(x|s)$, let

$$Q^*(x) = \sum_s p(s)p(x|s),$$

$$Q^*(x|s,y) = \frac{p(x|s)p(y|x)}{\sum_{x'} p(x'|s)p(y|x')},$$

then it is easy to get

$$\begin{aligned}
& J(p(x|s), Q(x), Q(x|s,y)) - J(p(x|s), Q^*(x), Q^*(x|s,y)) \\
&= \sum_{s,x} p(s)p(x|s) \log \frac{Q^*(x)}{Q(x)} - \gamma \sum_{s,x,y} p(s)p(x|s)p(y|x) \log \frac{Q(x|s,y)}{Q^*(x|s,y)} \\
&= \sum_{x} Q^*(x) \log \frac{Q^*(x)}{Q(x)} - \gamma \sum_{s,x,y} p(s,y)Q^*(x|s,y) \log \frac{Q(x|s,y)}{Q^*(x|s,y)} \\
&\geq 0
\end{aligned}$$

by the log-sum inequality. So, for fixed $p(x|s)$, $\min_{Q(x),Q(x|s,y)} J(p(x|s), Q(x), Q(x|s,y))$ is achieved by $Q^*(x)$ and $Q^*(x|s,y)$. Moreover, it is easy to check that the minimum value is $I(S;X) - \lambda \mathbf{E}d(S,X) - \gamma I(X;Y|S)$. Thus, (a) and (b) are proved.

(c). One has

$$\begin{aligned}
\frac{\partial J}{\partial p(x|s)} &= p(s)(1+\gamma) \log p(x|s) + (1+\gamma)p(s) - p(s)(\lambda d(s,x) + \log Q(x)) \\
&\quad -\gamma p(s) \sum_{y} p(y|x) \log Q(x|s,y).
\end{aligned}$$

Suppose $p(s) > 0$ for all s. By the KKT conditions and $\sum_x p(x|s) = 1$, if $p^*(x|s) > 0$ then one has

$$p^*(x|s) = \frac{\exp\left[\frac{1}{1+\gamma}\left(\log Q(x) + \lambda d(s,x) + \gamma \sum_y p(y|x) \log Q(x|s,y)\right)\right]}{\sum_{x'} \exp\left[\frac{1}{1+\gamma}\left(\log Q(x') + \lambda d(s,x') + \gamma \sum_y p(y|x') \log Q(x'|s,y)\right)\right]}.$$

For this optimal $p^*(x|s)$, it is easy to get the minimum of J for fixed $Q(x)$ and $Q(x|s,y)$.

□

6.4.2 Algorithm B

Fix a small number $\epsilon > 0$.

Step 1. Choose $Q^{(1)}(x) > 0, Q^{(1)}(x|s,y) > 0$ arbitrarily for any s, y;

Step 2. Let

$$b^{(n)}(s,x) = \exp\left[\frac{1}{1+\gamma}\log Q^{(n-1)}(x) + \frac{\lambda}{1+\gamma}d(s,x) + \frac{\gamma}{1+\gamma}\sum_y p(y|x)\log Q^{(n-1)}(x|s,y)\right].$$

Compute

$$p^{(n)}(x|s) = \frac{b^{(n)}(s,x)}{\sum_{x'} b^{(n)}(s,x')}, \tag{6.11}$$

$$Q^{(n)}(x) = \sum_s p(s)p^{(n)}(x|s), \tag{6.12}$$

$$Q^{(n)}(x|s,y) = \frac{p^{(n)}(x|s)p(y|x)}{\sum_{x'} p^{(n)}(x'|s)p(y|x')}. \tag{6.13}$$

Step 3. If

$$\max_x \log Q^{(n)}(x) - \sum_x Q^{(n+1)}(x)\log Q^{(n)}(x) < \epsilon,$$

stop the iteration and get the optimal probability distributions $p^{(n)}(x|s), Q^{(n)}(x), Q^{(n)}(x|s,y)$ achieving the compression rate function $R_c(D, R_w)$; otherwise, go to Step 2.

6.4.3 Convergence of Algorithm B

For $p^{(n)}(x|s), Q^{(n-1)}(x), Q^{(n-1)}(x|s,y)$ defined in (6.11), (6.12) and (6.13), it is easy to have

$$\begin{aligned}
J(p^{(n)}(x|s), Q^{(n-1)}(x), Q^{(n-1)}(x|s,y)) &= (1+\gamma)\sum_{s,x} p(s)p^{(n)}(x|s)\log p^{(n)}(x|s) \\
&\quad - \sum_{s,x} p(s)p^{(n)}(x|s)(\log Q^{(n-1)}(x) + \lambda d(s,x)) - \gamma \sum_{s,x,y} p(s)p^{(n)}(x|s)p(y|x)\log Q^{(n-1)}(x|s,y) \\
&= (1+\gamma)\sum_{s,x} p(s)p^{(n)}(x|s)\log \frac{b^{(n)}(s,x)}{\sum_{x'} b^{(n)}(s,x')} - \sum_{s,x} p(s)p^{(n)}(x|s)(\log Q^{(n-1)}(x) \\
&\quad + \lambda d(s,x)) - \gamma \sum_{s,x,y} p(s)p^{(n)}(x|s)p(y|x)\log Q^{(n-1)}(x|s,y) \\
&= -(1+\gamma)\sum_{s} p(s)\log \sum_{x'} \exp[\frac{1}{1+\gamma}\log Q^{(n-1)}(x') + \frac{\lambda}{1+\gamma} d(s,x') \\
&\quad + \frac{\gamma}{1+\gamma}\sum_{y} p(y|x')\log Q^{(n-1)}(x'|s,y))] \\
&= -(1+\gamma)\sum_{s} p(s)\log \sum_{x'} b^{(n)}(s,x').
\end{aligned}$$

Assume $(p(x|s), Q(x), p(x|s,y))$ achieve the compression rate function $R(D, R_w)$ and define

$$p^{(n)}(s|x) = \frac{p(s)p^{(n)}(x|s)}{Q^{(n)}(x)}, p(s|x) = \frac{p(s)p(x|s)}{Q(x)},$$

then, one has

$$\sum_{x'} b^{(n)}(s,x') = \frac{p(s)b^{(n)}(s,x)}{Q^{(n)}(x)p^{(n)}(s|x)},$$

and

$$\sum_{x'} b(s,x') = \frac{p(s)b(s,x)}{Q(x)p(s|x)}.$$

So,

$$
\begin{aligned}
0 \le & J(p^{(n)}(x|s), Q^{(n-1)}(x), Q^{(n-1)}(x|s,y)) - J(p(x|s), Q(x), Q(x|s,y)) \\
= & (1+\gamma)\sum_{s,x} p(s)p(x|s)\log\frac{p(s)b(s,x)}{Q(x)p(s|x)} - (1+\gamma)\sum_{s,x} p(s)p(x|s)\log\frac{p(s)b^{(n)}(s,x)}{Q^{(n)}(x)p^{(n)}(s|x)} \\
= & \sum_{s,x} p(s)p(x|s)\log\frac{Q^{(n)}(x)}{Q^{(n-1)}(x)} - \gamma\sum_{s,x} p(s)p(x|s)\log\frac{Q(x)}{Q^{(n)}(x)} \\
& -(1+\gamma)\sum_{s,x} p(s)p(x|s)\log\frac{p(s|x)}{p^{(n)}(s|x)} + \gamma\sum_{s,x,y} p(s)p(x|s)p(y|x)\log\frac{Q(x|s,y)}{Q^{(n-1)(x|s,y)}} \\
= & \sum_{s,x} p(s)p(x|s)\log\frac{Q^{(n)}(x)}{Q^{(n-1)}(x)} - \sum_{s,x} p(s)p(x|s)\log\frac{p(s|x)}{p^{(n)}(s|x)} \\
& -\gamma\sum_{s,x,y} p(s)p(x|s)p(y|x)\log\frac{\frac{Q(x)p(s|x)}{Q(x|s,y)}}{\frac{Q^{(n)}(x)p^{(n)}(s|x)}{Q^{(n-1)(x|s,y)}}} \\
= & \sum_{s,x} p(s)p(x|s)\log\frac{Q^{(n)}(x)}{Q^{(n-1)}(x)} - \sum_{s,x} p(s)p(x|s)\log\frac{p(s|x)}{p^{(n)}(s|x)} \\
& -\gamma\sum_{s,x,y} p(s)p(x|s)p(y|x)\log\frac{p(y)}{p^{(n-1)}(y)} \\
= & \sum_{x} Q(x)\log\frac{Q^{(n)}(x)}{Q^{(n-1)}(x)} - \sum_{x} Q(x)\sum_{s} p(s|x)\log\frac{p(s|x)}{p^{(n)}(s|x)} - \gamma\sum_{y} p(y)\log\frac{p(y)}{p^{(n-1)}(y)} \\
= & D(Q(x)||Q^{(n-1)}(x)) - D(Q(x)||Q^{(n)}(x)) \\
& -\sum_{x} Q(x)D(p(s|x)||p^{(n)}(s|x)) - \gamma D(p(y)||p^{(n-1)}(y)),
\end{aligned}
$$

where

$$
p^{(n-1)}(y) = \sum_{s,x} p(s)p^{(n-1)}(x|s)p(y|x),
$$

$$
\begin{aligned}
\frac{Q(x)p(s|x)}{Q(x|s,y)} &= \frac{p(s,x)p(s)p(y|x)}{Q(x|s,y)p(s)p(y|x)} \\
&= \frac{p(s,x,y)p(s)}{p(x,s|y)p(y|x)} = \frac{p(y)p(s)}{p(y|x)}
\end{aligned}
$$

and

$$\begin{aligned}\frac{Q^{(n)}(x)p^{(n)}(s|x)}{Q^{(n-1)}(x|s,y)} &= \frac{p(s)p^{(n-1)}(x|s)p(s)p(y|x)}{Q^{(n-1)}(x|s,y)p(s)p(y|x)} \\ &= \frac{p^{(n-1)}(y)p(s)}{p(y|x)}.\end{aligned}$$

Therefore we have

(i)

$$\begin{aligned}0 \leq J(p^{(n)}(x|s), Q^{(n-1)}(x), Q^{(n-1)}(x|s,y)) - J(p(x|s), Q(x), Q(x|s,y)) \\ \leq D(Q(x)||Q^{(n-1)}(x)) - D(Q(x)||Q^{(n)}(x));\end{aligned}$$

(ii)

$$\sum_x Q(x)D(p(s|x)||p^{(n)}(s|x)) \leq D(Q(x)||Q^{(n-1)}(x)) - D(Q(x)||Q^{(n)}(x));$$

(iii)

$$D(p(y)||p^{(n-1)}(y)) \leq D(Q(x)||Q^{(n-1)}(x)) - D(Q(x)||Q^{(n)}(x)).$$

In light of (i)-(iii), for any $N > 1$, we have

(a)

$$\begin{aligned}&\sum_{n=2}^{N}[J(p^{(n)}(x|s), Q^{(n-1)}(x), Q^{(n-1)}(x|s,y)) - J(p(x|s), Q(x), Q(x|s,y))] \\ &\leq \sum_{n=2}^{N}[D(Q(x)||Q^{(n-1)}(x)) - D(Q(x)||Q^{(n)}(x))] \\ &= D(Q(x)||Q^{(1)}(x)) - D(Q(x)||Q^{(N)}(x)) \\ &\leq D(Q(x)||Q^{(1)}(x)) < \infty,\end{aligned}$$

(b)

$$\sum_{n=2}^{N}\sum_x Q(x)D(p(s|x)||p^{(n)}(s|x)) \leq D(Q(x)||Q^{(1)}(x)) < \infty,$$

(c)

$$\sum_{n=2}^{N} D(p(y)||p^{(n-1)}(y)) \leq D(Q(x)||Q^{(1)}(x)) < \infty,$$

which imply that, as $n \to \infty$,

$$J(p^{(n)}(x|s), Q^{(n-1)}(x), Q^{(n-1)}(x|s,y)) \to$$
$$J(p(x|s), Q(x), Q(x|s,y)),$$

and

$$D(p(s|x)||p^{(n)}(s|x)) \to 0,$$
$$D(p(y)||p^{(n-1)}(y)) \to 0,$$

and the last two limitations guarantee that $p^{(n)}(s|x) \to p(s|x)$ and $p^{(n)}(y) \to p(y)$ as $n \to \infty$.

It is obvious that $\{Q^{(n)}(x), Q^{(n)}(x|s,y)\}$ are bounded sequences. By the Bolzano-Weierstrass Theorem, there exists a subsequence $\{Q^{(n_i)}(x), Q^{(n_i)}(x|s,y)\}$ convergent to, say $\{Q^{**}(x), Q^{**}(x|s,y)\}$. Suppose $p^{(n_i+1)}(x|s)$ determined by $(Q^{(n_i)}(x), Q^{(n_i)}(x|s,y))$ in (6.11) approaches $p^{**}(x|s)$, then as $n \to \infty$,

$$J(p^{(n_i+1)}(x|s), Q^{(n_i)}(x), Q^{(n_i)}(x|s,y)) \to J(p^{**}(x|s), Q^{**}(x), Q^{**}(x|s,y)).$$

Since $\{J(p^{(n_i+1)}(x|s), Q^{(n_i)}(x), Q^{(n_i)}(x|s,y))\}$ is a subsequence of $\{J(p^{(n+1)}(x|s), Q^{(n)}(x), Q^{(n)}(x|s,y))\}$, we have

$$J(p^{**}(x|s), Q^{**}(x), Q^{**}(x|s,y)) = J(p(x|s), Q(x), Q(x|s,y)),$$

which means $(p^{**}(x|s), Q^{**}(x), Q^{**}(x|s,y))$ achieve the minimum. Thus we have

$$\begin{aligned} Q^{**}(x) &= \sum_s p(s)p^{**}(x|s), \\ Q^{**}(x|s,y) &= \frac{p^{**}(x|s)p(y|x)}{\sum_{x'} p^{**}(x'|s)p(y|x')}, \\ p^{**}(x|s) &= \frac{b^{**}(s,x)}{\sum_{x'} b^{**}(s,x')}, \end{aligned}$$

where

$$b^{**}(s,x) = \exp\left[\frac{1}{1+\gamma}(\log Q^{**}(x) + \lambda d(s,x) + \gamma \sum_y p(y|x) \log Q^{**}(x|s,y))\right].$$

From (i), $D(Q^{**}(x)||Q^{(n-1)}(x)) - D(Q^{**}(x)||Q^{(n)}(x)) \geq 0$, that is, $\{D(Q^{**}(x)||Q^{(n)}(x))\}$ is monotonic, so, $\{D(Q^{**}(x)||Q^{(n)}(x))\}$ has a limit. Since $\{D(Q^{**}(x)||Q^{(n_i)}(x))\}$ is its subsequence convergent to 0, so $\{D(Q^{**}(x)||Q^{(n)}(x))\}$ is convergent to 0. Thus, $Q^{(n)}(x) \to Q^{**}(x)$. Since $p^{(n)}(s|x) \to p^{**}(s|x)$ and $p^{(n)}(y) \to p^{**}(y)$ as $n \to \infty$, and

$$\frac{Q^{(n)}(x)p^{(n)}(s|x)}{Q^{(n-1)}(x|s,y)} = \frac{p^{(n-1)}(y)p(s)}{p(y|x)},$$

one has $Q^{(n)}(x|s,y) \to Q^{**}(x|s,y)$, and $p^{(n)}(x|s) \to p^{**}(x|s)$. The proof of convergence is finished.

6.5 Summary

In this chapter we develop two efficient iterative algorithms for calculating watermarking capacities and compression and watermarking rate regions of joint compression and private watermarking systems. Furthermore, the two algorithms are shown to be convergent.

Chapter 7

Algorithms for Computing Joint Compression and Public Watermarking Rate Regions

In this chapter we will develop algorithms for computing public watermarking capacities and compression and watermarking rate regions of joint compression and public watermarking systems with finite alphabets and under fixed attack channels.

7.1 Formulation of Joint Compression and Public Watermarking Rate Regions

For a joint compression and public watermarking system with a memoryless covertext source S with the probability distribution $p(s)$, depicted in Figure 3.1, a compression rate function with respect to watermarking rate R_w and distortion level D is defined as $R_c(D, R_w) \stackrel{def}{=} \inf R_c$, where the inf is taken over all publicly achievable (R_w, R_c) with respect to D and the fixed R_w. It is shown in [21] that, for any $D \geq 0$ and $0 \leq R_w \leq C(D)$, where $C(D) \stackrel{def}{=} \max_{p(u,x|s):\mathbf{E}d(S,X)\leq D}[I(U;Y)-I(U;S)]$ is the public watermarking capacity,

one has

$$R_c(D, R_w) = R_w + \min I(S; U, X), \tag{7.1}$$

where the minimum is taken over all random variables (U, X) taking values over a finite alphabet $\mathcal{U}\times\mathcal{X}$ with $|\mathcal{U}| \leq |\mathcal{S}||\mathcal{X}|+1$, jointly distributed with S, Y with the joint probability distribution $p(s, u, x, y) = p(s)p(u|s)p(x|s, u)p(y|x)$ such that $\mathbf{E}d(S, X) \leq D$ and

$$R_w \quad \leq \quad I(U; Y) - I(U; S).$$

Define

$$R_c^{(0)}(D) = \min_{p(u|s),p(x|s,u):\mathbf{E}d(S,X)\leq D} I(S; U, X).$$

If $R_c^{(0)}(D)$ is achieved at $p^*(u|s), p^*(x|s, u)$, and let

$$R_w^{(0)}(D) = I_{p^*(u|s),p^*(x|s,u)}(U; Y) - I_{p^*(u|s)}(U; S).$$

Then, it is obvious from (7.1) that

$$R_c(D, R_w) = R_w + R_w^{(0)}(D) \tag{7.2}$$

for $0 \leq R_w \leq R_w^{(0)}(D)$, that is, $R_c(D, R_w)$ is a linear function of R_w; for $R_w^{(0)}(D) \leq R_w \leq C(D)$, $R_c(D, R_w)$ is a curve of R_w given in (7.1). Thus, to determine the joint compression and watermarking rate region of the public watermarking system, or equivalently, the compression rate function $R_c(D, R_w)$ with respect to public watermarking rate R_w and distortion level D, we must calculate $R_c^{(0)}(D)$, the public watermarking capacity $C(D)$ and

$$\min_{p(u|s),p(x|u,s):\mathbf{E}d(S,X)\leq D, R_w\leq I(U;Y)-I(U;S)} I(S; U, X).$$

Actually, $R_c^{(0)}(D)$ can be computed by employing the standard Blahut-Arimoto algorithm for rate-distortion functions. To see this, we define a new alphabet $\mathcal{X}' = \mathcal{U} \times \mathcal{X}$ and a new distortion measure d' between $\mathcal{S}$ and $\mathcal{X}'$ by letting $d'(s, (u, x)) = d(s, x)$. Then, it is obvious from the definition of $R_c^{(0)}(D)$ that $R_c^{(0)}(D)$ is the standard rate-distortion function of the source S with reproduction alphabet $\mathcal{X}'$ and the distortion measure d'.

7.2 Computing Public Watermarking Capacities

For a public watermarking system under a fixed attack channel $p(y|x)$, the watermarking capacity is given in [26, 33] by

$$C(D) = \max_{p(u,x|s):\mathbf{E}d(S,X)\leq D} [I(U;Y) - I(U;S)],$$

where $|\mathcal{U}| \leq |\mathcal{S}||\mathcal{X}| + 1$, and $p(s,u,x,y) = p(s)p(u,x|s)p(y|x)$. Since $I(U;Y) - I(U;S)$ is neither convex nor concave with respect to $p(u,x|s)$, so existing algorithms for convex optimization is not applicable for computing $C(D)$. However, it is shown in [26] that $I(U;Y) - I(U;S)$ is convex with respect to $p(x|u,s)$ for fixed $p(u|s)$, and concave with respect to $p(u|s)$ for fixed $p(x|u,s)$. We shall exploit this property to develop algorithms for computing $C(D)$.

Using Lagrange multiplier, we know that for $\lambda \geq 0$

$$C(D) = \lambda D + \max_{p(u|s),p(x|u,s)} [I(U;Y) - I(U;S) - \lambda \mathbf{E}d(S,X)] \tag{7.3}$$

where $D = \sum_{s,u,x} p(s)p^*(u|s)p^*(x|u,s)d(s,x)$ and $p^*(u|s), p^*(x|u,s)$ achieve the above maximum.

Property 7.1 *For probability distributions $p(u|s), p(x|u,s), Q(u|y)$, define*

$$\begin{aligned} L(p(u|s), p(x|u,s), Q(u|y)) &= \sum_{s,u,x,y} p(s)p(u|s)p(x|u,s)p(y|x) \log \frac{Q(u|y)}{p(u|s)} \\ &\quad -\lambda \sum_{s,u,x} p(s)p(u|s)p(x|u,s)d(s,x). \end{aligned}$$

Then

$$C(D) = \lambda D + \max_{p(u|s),p(x|u,s)} \max_{Q(u|y)} L(p(u|s), p(x|u,s), Q(u|y)). \tag{7.4}$$

Proof: For fixed $p(u|s), p(x|u,s)$, define

$$Q^*(u|y) = \frac{\sum_{s,x} p(s)p(u|s)p(x|u,s)p(y|x)}{\sum_{s,u,x} p(s)p(u|s)p(x|u,s)p(y|x)}.$$

Then

$$\begin{aligned} & L(p(u|s), p(x|u,s), Q(u|y)) - L(p(u|s), p(x|u,s), Q^*(u|y)) \\ = & \sum_{s,u,x,y} p(s)p(s,u,x,y) \log \frac{Q(u|y)}{Q^*(u|y)} \\ \leq & 0 \end{aligned}$$

by the log-sum inequality. Moreover, it is easy to verify that $L(p(u|s), p(x|u,s), Q^*(u|y)) = I(U;Y) - I(U;S) - \lambda \mathbf{E}d(S,X)$. So, the proof is completed.

□

Property 7.2 *(a) For fixed $p(u|s), p(x|u,s)$, the optimal probability distribution $Q^*(u|y)$ achieving the maximum of $L(p(u|s), p(x|u,s), Q(u|y))$ is given by*

$$Q^*(u|y) = \frac{\sum_{s,x} p(s)p(u|s)p(x|u,s)p(y|x)}{\sum_{s,u,x} p(s)p(u|s)p(x|u,s)p(y|x)};$$

(b) For fixed $Q(u|y), p(x|u,s)$, the optimal probability distribution $p^*(u|s)$ achieving the maximum of $L(p(u|s), p(x|u,s), Q(u|y))$ is given by

$$p^*(u|s) = \frac{\exp\left(\sum_{x,y} p(x|u,s)p(y|x) \log Q(u|y) - \lambda \sum_x p(x|u,s)d(s,x)\right)}{\sum_{u'} \exp\left(\sum_{x,y} p(x|u',s)p(y|x) \log Q(u'|y) - \lambda \sum_x p(x|u',s)d(s,x)\right)}.$$

(c) For fixed $p(u|s), Q(u|y)$, the optimal probability distribution $p^*(x|u,s)$ achieving the maximum of $L(p(u|s), p(x|u,s), Q(u|y))$ is given by

$$p^*(x|u,s) = \begin{cases} 1, & x = \arg\max_x \sum_y p(y|x) \left[\log \frac{Q(u|y)}{p(u|s)} - \lambda d(s,x)\right], \\ 0, & \textit{otherwise.} \end{cases} \tag{7.5}$$

Note here $p^*(x|u,s)$ is a function from $\mathcal{U} \times \mathcal{S}$ to $\mathcal{X}$ and not necessarily unique.

Proof: (a) is obvious from Property 7.1.

(b) It is not hard to get

$$\begin{aligned} \frac{\partial L(p(u|s), p(x|u,s), Q(u|y))}{\partial p(u|s)} = & \sum_{x,y} p(s)p(x|u,s)p(y|x) \log \frac{Q(u|y)}{p(u|s)} \\ & -p(s) - \lambda \sum_x p(s)p(x|u,s)d(s,x). \end{aligned}$$

By the KKT conditions and $\sum_u p(u|s) = 1$ for any s, if $p^*(u|s) > 0$ one has

$$p^*(u|s) = \frac{\exp\left(\sum_{x,y} p(x|u,s)p(y|x)\log Q(u|y) - \lambda \sum_x p(x|u,s)d(s,x)\right)}{\sum_{u'} \exp\left(\sum_{x,y} p(x|u',s)p(y|x)\log Q(u'|y) - \lambda \sum_x p(x|u',s)d(s,x)\right)},$$

the optimal probability distribution achieving the maximum of $L(p(u|s), p(x|u,s), Q(u|y))$.

(c) For fixed $p(u|s), Q(u|y)$, one has

$$\begin{aligned} L(p(u|s), p(x|u,s), Q(u|y)) &= \sum_{s,u,x} p(s)p(u|s)p(x|u,s) \sum_y p(y|x) \left[\log \frac{Q(u|y)}{p(u|s)} - \lambda d(s,x)\right] \\ &\le \sum_{s,u} p(s)p(u|s) \max_x \sum_y p(y|x) \left[\log \frac{Q(u|y)}{p(u|s)} - \lambda d(s,x)\right] \\ &= \sum_{s,u} p(s)p(u|s)p^*(x|u,s) \sum_y p(y|x) \left[\log \frac{Q(u|y)}{p(u|s)} - \lambda d(s,x)\right], \end{aligned}$$

where

$$p^*(x|u,s) = \begin{cases} 1, & x = \arg\max_x \sum_y p(y|x) \left[\log \frac{Q(u|y)}{p(u|s)} - \lambda d(s,x)\right], \\ 0, & \text{otherwise} \end{cases}$$

□

Based on Property 7.2, the following algorithm for computing $C(D)$ is proposed.

Algorithm A

Fix any $\epsilon > 0$.

Step one: Initially choose probability distributions $p(x|u,s) > 0$ for all (u,s).

Step two: Choose probabilities $p^{(1)}(u|s) > 0$.

Step three: Computing

$$\begin{aligned} Q^{(n)}(u|y) &= \frac{\sum_{s,x} p(s)p^{(n)}(u|s)p(x|u,s)p(y|x)}{\sum_{s,u,x} p(s)p^{(n)}(u|s)p(x|u,s)p(y|x)}; \\ p^{(n+1)}(u|s) &= \frac{\exp\left(\sum_{x,y} p(x|u,s)p(y|x)\log Q^{(n)}(u|y) - \lambda \sum_x p(x|u,s)d(s,x)\right)}{\sum_{u'} \exp\left(\sum_{x,y} p(x|u',s)p(y|x)\log Q^{(n)}(u'|y) - \lambda \sum_x p(x|u',s)d(s,x)\right)}. \end{aligned}$$

Step four: If

$$\sum_s p(s) \max_u \sum_{x,y} \left[p(x|u,s)p(y|x) \log \frac{Q^{(n)}(u|y)}{p^{(n)}(u|s)} - \lambda p(x|u,s)d(s,x) \right] - L(p^{(n)}(u|s), p(x|u,s), Q^{(n)}(u|y)) \leq \epsilon,$$

then fix $p^{(n)}(u|s), Q^{(n)}(u|y)$, go to Step five; otherwise, go to Step three.

Step five: Compute all functions $p^*(x|u,s)$ such that

$$p^*(x|u,s) = \begin{cases} 1, & x = \arg\max_x \sum_y p(y|x) \left[\log \frac{Q^{(n)}(u|y)}{p^{(n)}(u|s)} - \lambda d(s,x) \right], \\ 0, & \text{otherwise.} \end{cases}$$

Step six: For each $p^*(x|u,s)$, compute

$$Q^*(u|y) = \frac{\sum_{s,x} p(s)p^{(n)}(u|s)p^*(x|u,s)p(y|x)}{\sum_{s,u,x} p(s)p^{(n)}(u|s)p^*(x|u,s)p(y|x)}.$$

If for all $p^*(x|u,s)$,

$$\sum_s p(s) \max_u \max_x \sum_y \left[p(y|x) \log \frac{Q^*(u|y)}{p^{(n)}(u|s)} - \lambda d(s,x) \right] - L(p^{(n)}(u|s), p^*(x|u,s), Q^*(u|y)) \leq \epsilon, \tag{7.6}$$

then we get the optimal $p^{(n)}(u|s), p^*(x|u,s), Q^*(u|y)$. Here $p^*(x|u,s)$ can be any one obtained in Step five. Otherwise, if (7.6) is not satisfied for some $p^*(x|u,s)$, then go to Step three to update $Q^{(n)}(u|y)$ and $p^{(n)}(u|s)$.

Remark: Algorithm A is similar to that in [6] used to compute channel capacities with channel side information, so proof of termination conditions and convergence are omitted here. The difference is that S is the channel side information in [6] while S is a covertext source in Algorithm A here and not involved into the attack channel. Although Algorithm A is very straightforward from the Blahut-Arimoto algorithm, the drawback is obvious in Step Five, where all $|\mathcal{X}|^{|\mathcal{U}||\mathcal{S}|}$ functions from $\mathcal{U} \times \mathcal{S}$ to $\mathcal{X}$ must be checked for finding all possible optimal $p^*(x|u,s)$, which slows down the convergence of the algorithm

significantly. Therefore, to speed up the convergence of the algorithm, we propose another algorithm based on Shannon's strategy, which is also used in [15].

Before describing the algorithm, the following theorem is given.

Theorem 7.1 *Let $\mathcal{T}$ be the set of all functions from $\mathcal{S}$ to $\mathcal{X}$. Define a new distortion measure d' between $\mathcal{S}$ and $\mathcal{T}$ by $d'(s,t) \stackrel{def}{=} d(s,t(s)), s \in \mathcal{S}, t \in \mathcal{T}$ where d is the distortion measure between $\mathcal{S}$ and $\mathcal{X}$. Then the public watermarking capacity $C(D)$ is equal to*

$$C(D) = \max_{p(t|s):\mathbf{E}d'(S,T)\leq D} [I(T;Y) - I(T;S)] \tag{7.7}$$

where T is a random variable taking values from $\mathcal{T}$, and the joint probability of $p(s,t,y)$ is given by $p(s,t,y) = p(s)p(t|s)p(y|t(s))$.

Proof: On the one hand, let (S,T) be random variables with joint probability $p(s)p^*(t|s)$ achieving the maximum in the right side of (7.7) and $\mathbf{E}d'(S,T) \leq D$. Define a new random variable X by letting $p(x|s,t) = 1$ if $x = t(s)$, and $p(x|s,t) = 0$ otherwise. Then , $\mathbf{E}d(S,X) \leq D$, and $(S,T) \to X \to Y$ forms a Markov chain. Thus

$$C(D) \geq I_{p^*}(T;Y) - I_{p^*}(T;S) = \max_{p(t|s):\mathbf{E}d'(S,T)\leq D} [I(T;Y) - I(T;S)]$$

by the definition of $C(D)$.

On the other hand, Let $p^*(u|s), p^*(x|u,s)$ achieve the public capacity $C(D)$. For the fixed $p^*(u|s)$, by employing the same approach to Property 7.2-(a) and (c), $p^*(x|u,s)$ must satisfy

$$p^*(x|u,s) = \begin{cases} 1, & x = \arg\max_x \sum_y p(y|x) \log \dfrac{\frac{\sum_{s,x} p(s)p(u|s)p^*(x|u,s)p(y|x)}{\sum_{s,u,x} p(s)p(u|s)p^*(x|u,s)p(y|x)}}{p(u|s)}, \\ 0, & \text{otherwise}, \end{cases} \tag{7.8}$$

in other words, (7.8) defines $|\mathcal{U}|$ functions from $\mathcal{S}$ to $\mathcal{X}$, and each is denoted by u. Now define a random variable T by letting

$$p(t|s) = \begin{cases} p^*(u|s), & t = u; \\ 0, & \text{otherwise}. \end{cases} \tag{7.9}$$

Then we can verify that $\mathbf{E}d'(S,T) \leq D$ and

$$\begin{aligned} C(D) &= I_{p^*(u|s),p^*(x|u,s)}(U;Y) - I_{p^*(u|s)}(U;S) \\ &= I(T;Y) - I(T;S) \\ &\leq \max_{p(t|s):\mathbf{E}d'(S,T)\leq D} [I(T;Y) - I(T;S)]. \end{aligned}$$

The proof is finished.

□

Now based on (7.7), it is easy to get

$$C(D) = \lambda D + \max_{p(t|s)} \max_{Q(t|y)} \left[\sum p(s)p(t|s)p(y|t(s)) \log \frac{Q(t|y)}{p(t|s)} - \lambda \mathbf{E}d'(S,T) \right]. \quad (7.10)$$

So, by applying the idea of the Blahut-Arimoto algorithm we have Algorithm A' stated as follows without proof.

Algorithm A'

Fix any $\epsilon > 0$.

Step one: Choose probability distributions $p^{(1)}(t|s) > 0$.

Step two: Computing

$$\begin{aligned} Q^{(n)}(t|y) &= \frac{\sum_s p(s)p^{(n)}(t|s)p(y|t(s))}{\sum_{s,t'} p(s)p^{(n)}(t'|s)p(y|t'(s))}; \\ p^{(n+1)}(t|s) &= \frac{\exp\left(\sum_y p(y|t(s)) \log Q^{(n)}(t|y) - \lambda d'(s,t)\right)}{\sum_{t'} \exp\left(\sum_y p(y|t'(s)) \log Q^{(n)}(t'|y) - \lambda d'(s,t')\right)}. \end{aligned}$$

Step four: If

$$\sum_s p(s) \left\{ \log \max_t \left[\exp\left(\sum_y p(y|t(s)) \log Q^{(n)}(t|y) - \lambda d'(s,t) \right) \right] \right.$$
$$\left. - \log \sum_t p^{(n+1)}(t|s) \left[\exp\left(\sum_y p(y|t(s)) \log Q^{(n)}(t|y) - \lambda d'(s,t) \right) \right] \right\} \leq \epsilon,$$

then stop iteration and get the optimal distributions $p^{(n+1)}(t|s), Q^{(n)}(t|y)$ achieving the watermarking capacity; otherwise, go to Step two.

Remarks

- Compared with Algorithm A, Algorithm A' is faster. However, the cost is expansion of the output alphabet of convert channels from $|\mathcal{X}||\mathcal{U}|$ to $|\mathcal{X}|^{|\mathcal{S}|}$. In general, $|\mathcal{X}||\mathcal{U}| << |\mathcal{X}|^{|\mathcal{S}|}$.
- Algorithm A' can be regarded as a generalization of the algorithm in [15], where no any constraint on $p(t|s)$ is applied.

7.3 Computing Compression Rate Functions

From the discussion in Section 7.1, we only need to calculate

$$R_c(D, R_w) = R_w + \min I(S; U, X),$$

for $R_w^{(0)} \leq R_w \leq C(D)$, where the minimum is taken over all $p(u|s), p(x|u, s)$ such that $\mathbf{E}d(S, X) \leq D$ and $R_w \leq I(U; Y) - I(U; S)$.

Using the standard Lagrange multiplier, one has

Property 7.3 *Let $\lambda \leq 0$, $\mu \leq 0$. Then*

$$R_c(D, R_w) = \lambda D - (\mu - 1)R_w + \min_{p(u|s), p(x|u,s)} [I(S; U, X) - \lambda \mathbf{E}d(S, X) + \mu(I(U; Y) - I(U; S))],$$

where

$$\begin{aligned} D &= \sum_{u,s,x} p(s)p^*(u|s)p^*(x|u, s)d(s, x), \\ R_w &= I_{p^*(u|s), p^*(x|u,s)}(U; Y) - I_{p^*(u|s)}(U; S), \end{aligned}$$

and $p^*(u|s), p^*(x|u, s)$ are optimal probability distributions achieving the above minimum.

Property 7.4 *Let $\lambda \leq 0$, $\mu \leq 0$. For any probability distributions $p(u|s), p(x|u, s)$,*

$Q(u,x), Q(u|y)$, *define*

$$L(p(u|s), p(x|u,s), Q(u,x), Q(u|y)) \stackrel{def}{=} \sum_{s,u,x} p(s)p(u|s)p(x|u,s) \log \frac{p(u|s)p(x|u,s)}{Q(u,x)}$$
$$-\lambda \sum_{s,u,x} p(s)p(u|s)p(x|u,s)d(s,x) + \mu \sum_{s,u,x,y} p(s)p(u|s)p(x|u,s)p(y|x) \log \frac{Q(u|y)}{p(u|s)}.$$

Then

$$R_c(D, R_w) = \lambda D - (\mu - 1)R_w + \min_{\{p(u|s), p(x|u,s)\}} \min_{\{Q(u,x), Q(u|y)\}} L(p(u|s), p(x|u,s), Q(u,x), Q(u|y)). \quad (7.11)$$

Proof: For fixed $p(u|s), p(x|u,s)$, define

$$Q^*(u,x) = \sum_s p(s)p(u|s)p(x|u,s) \quad (7.12)$$

$$Q^*(u|y) = \frac{\sum_{s,x} p(s)p(u|s)p(x|u,s)p(y|x)}{\sum_{s,u,x} p(s)p(u|s)p(x|u,s)p(y|x)}. \quad (7.13)$$

By the definition of mutual information and the log-sum inequality, it is not hard to verify that

$$L(p(u|s), p(x|u,s), Q^*(u,x), Q^*(u|y)) = I(S;U,X) - \lambda \mathbf{E}d(S,X) + \mu(I(U;Y) - I(U;S)),$$

and

$$L(p(u|s), p(x|u,s), Q(u,x), Q(u|y)) - L(p(u|s), p(x|u,s), Q^*(u,x), Q^*(u|y)) \geq 0.$$

Thus, the property is proved.

□

The following property can be obtained in the similar way to Property 7.2.

Property 7.5 *(a) For fixed $p(u|s), p(x|u,s)$, the optimal probability distributions $Q^*(u,x)$, $Q^*(u|y)$ achieving the minimum of $L(p(u|s), p(x|u,s), Q(u,x), Q(u|y))$ are given by (7.12) and (7.13) respectively.*

(b) For fixed $p(x|u,s)$ and $Q(u,x), Q(u|y)$, the optimal probability distributions $p^*(u|s)$ achieving the minimum of $L(p(u|s), p(x|u,s), Q(u,x), Q(u|y))$ are given by

$$p^*(u|s) = \frac{a(u,s)}{\sum_{u'} a(u',s)}$$

where

$$a(u,s) = \exp\left\{\frac{1}{\mu-1}\left[1-\mu+\sum_x p(x|u,s)\log\frac{p(x|u,s)}{Q(u,x)} + \mu\sum_{x,y} p(x|u,s)p(y|x)\log Q(u|y) \right.\right.$$
$$\left.\left. -\lambda\sum_x p(x|u,s)d(s,x)\right]\right\}.$$

(c) For fixed $p(u|s)$ and $Q(u,x), Q(u|y)$, the optimal probability distributions $p^*(x|u,s)$ achieving the minimum of $L(p(u|s), p(x|u,s), Q(u,x), Q(u|y))$ are given by

$$p^*(x|u,s) = \begin{cases} 1, & x = \arg\min_x \left(\log\frac{p(u|s)}{Q(u,x)} - \lambda d(s,x) + \mu\sum_y p(y|x)\log\frac{Q(u|y)}{p(u|s)}\right), \\ 0, & \textit{otherwise.} \end{cases}$$

Proof: (a) and (b) are omitted.

(c). It is easy to have

$$\frac{\partial L(p(u|s), p(x|u,s), Q(u,x), Q(u|y))}{\partial p(x|u,s)} = p(s)p(u|s)\log\frac{p(u|s)p(x|u,s)}{Q(u,x)}$$
$$+p(s)p(u|s) - \lambda p(s)p(u|s)d(s,x) + \mu\sum_y p(s)p(u|s)p(y|x)\log\frac{Q(u|y)}{p(u|s)}.$$

For fixed $p(u|s)$ and $Q(u,x), Q(u|y)$, obviously $L(p(u|s), p(x|u,s), Q(u,x), Q(u|y))$ is convex in $p(x|u,s)$. So by the KKT conditions, $p^*(x|u,s)$ is optimal if and only if

$$\frac{\partial L(p(u|s), p(x|u,s), Q(u,x), Q(u|y))}{\partial p(x|u,s)} = p(s)p(u|s)c_{u,s}, \quad \text{if } p^*(x|u,s) > 0,$$
$$\frac{\partial L(p(u|s), p(x|u,s), Q(u,x), Q(u|y))}{\partial p(x|u,s)} \geq p(s)p(u|s)c_{u,s}, \quad \text{if } p^*(x|u,s) = 0,$$

for $c_{u,s}$, only depending on u, s. Then, we can get the optimal probability distributions $p^*(x|u,s)$ by

$$p^*(x|u,s) = \frac{\exp\left[\lambda d(s,x) + \log\frac{Q(u,x)}{p(u|s)} + \mu\sum_y p(y|x)\log\frac{p(u|s)}{Q(u|y)}\right]}{\sum_{x'}\exp\left[\lambda d(s,x') + \log\frac{Q(u,x')}{p(u|s)} + \mu\sum_y p(y|x')\log\frac{p(u|s)}{Q(u|y)}\right]} \tag{7.14}$$

if $p^*(x|u,s) > 0$, and $L(p(u|s), p^*(x|u,s), Q(u,x), Q(u|y))$ is equal to

$$\sum_{u,s} p(s)p(u|s) \left(-\log \sum_{x'} \exp \left[\lambda d(s,x') + \log \frac{Q(u,x')}{p(u|s)} + \mu \sum_y p(y|x') \log \frac{p(u|s)}{Q(u|y)} \right] \right).$$

So, for any $p(x|u,s)$,

$$\begin{aligned} & L(p(u|s), p(x|u,s), Q(u,x), Q(u|y)) \\ \geq & \sum_{u,s} p(s)p(u|s) \left(-\log \sum_{x'} \exp \left[\lambda d(s,x') + \log \frac{Q(u,x')}{p(u|s)} + \mu \sum_y p(y|x') \log \frac{p(u|s)}{Q(u|y)} \right] \right) \\ \geq & \sum_{u,s} p(s)p(u|s) \left[-\lambda d(s,x_0) + \log \frac{p(u|s)}{Q(u,x_0)} + \mu \sum_y p(y|x_0) \log \frac{Q(u|y)}{p(u|s)} \right] \end{aligned}$$

if x_0 achieves $\min_x \left(\log \frac{p(u|s)}{Q(u,x)} - \lambda d(s,x) + \mu \sum_y p(y|x) \log \frac{Q(u|y)}{p(u|s)} \right)$, and the equalities hold if $p^*(x|u,s) = 1$ if $x = x_0$, and 0 otherwise. The proof is finishes. □

Algorithm B

Fix any $\epsilon > 0$.

Step one: Initially choose probability $p(x|u,s) > 0$.

Step two: Choose probabilities $p^{(1)}(u|s) > 0$.

Step three: Computing

$$\begin{aligned} Q^{(n)}(u,x) &= \sum_s p(s)p^{(n)}(u|s)p(x|u,s) \\ Q^{(n)}(u|y) &= \frac{\sum_{s,x} p(s)p^{(n)}(u|s)p(x|u,s)p(y|x)}{\sum_{s,u,x} p(s)p^{(n)}(u|s)p(x|u,s)p(y|x)} \\ p^{(n+1)}(u|s) &= \frac{a^{(n)}(u,s)}{\sum_{u'} a^{(n)}(u',s)} \end{aligned}$$

where

$$\begin{aligned} a^{(n)}(u,s) = \exp \Bigg\{ \frac{1}{\mu - 1} \Bigg[& 1 - \mu + \sum_x p(x|u,s) \log \frac{p(x|u,s)}{Q^{(n)}(u,x)} \\ & + \mu \sum_{x,y} p(x|u,s)p(y|x) \log Q^{(n)}(u|y) - \lambda \sum_x p(x|u,s)d(s,x) \Bigg] \Bigg\}. \end{aligned}$$

Step four: If

$$L(p^{(n)}(u|s), p(x|u,s), Q^{(n)}(u,x), Q^{(n)}(u|y)) - \sum_s p(s) \min_u \sum_x p(x|u,s)$$
$$\left[\log \frac{p^{(n)}(u|s)p(x|u,s)}{Q^{(n)}(u,x)} - \lambda d(s,x) + \mu \sum_y p(y|x) \log \frac{Q^{(n)}(u|y)}{p^{(n)}(u|s)}\right] \leq \epsilon,$$

then fix $p^{(n)}(u|s), Q^{(n)}(u,x), Q^{(n)}(u|y)$, go to Step five; otherwise, go to Step three.

Step five: Compute all functions $p^*(x|u,s)$ such that

$$p^*(x|u,s) = \begin{cases} 1, & x = \arg\min_x \left(\log \frac{p^{(n)}(u|s)}{Q^{(n)}(u,x)} - \lambda d(s,x) + \mu \sum_y p(y|x) \log \frac{Q^{(n)}(u|y)}{p^{(n)}(u|s)}\right), \\ 0, & \text{otherwise.} \end{cases}$$

Step six: For each $p^*(x|u,s)$, compute

$$\begin{aligned} Q^*(u,x) &= \sum_s p(s)p^{(n)}(u|s)p^*(x|u,s) \\ Q^*(u|y) &= \frac{\sum_{s,x} p(s)p^{(n)}(u|s)p^*(x|u,s)p(y|x)}{\sum_{s,u,x} p(s)p^{(n)}(u|s)p^*(x|u,s)p(y|x)} \end{aligned}$$

If for all $p^*(x|u,s)$,

$$L(p^{(n)}(u|s), p^*(x|u,s), Q^*(u,x), Q^*(u|y)) - \sum_s p(s)$$
$$\min_u \min_x \left[\log \frac{p^{(n)}(u|s)p^*(x|u,s)}{Q^*(u,x)} - \lambda d(s,x) + \mu \sum_y p(y|x) \log \frac{Q^*(u|y)}{p^*(u|s)}\right] \leq \epsilon, \quad (7.15)$$

then we get the optimal $p^{(n)}(u|s), p^*(x|u,s), Q^*(u,x), Q^*(u|y)$. Otherwise, if (7.15) is not satisfied for some $p^*(x|u,s)$, then go to Step three to update $Q^{(n)}(u,x), Q^{(n)}(u|y)$ and $p^{(n)}(u|s)$.

Remark: The Shannon's strategy cannot be applied in this case. To see this, similar to Theorem 7.1, we can write $\min_{p(u|s),p(x|u,s)}[I(S;U,X) - \lambda \mathbf{E}d(S,X) + \mu(I(U;Y) - I(U;S))]$ as $\min_{p(t|s)}[I(S;T,T(S)) - \lambda \mathbf{E}d'(S,T) + \mu(I(T;Y) - I(T;S))]$. However, besides $p(t|s)$,

we have to know the structure of functions t to compute $\sum_t p(t)H(t(S))$ in order to get $I(S;T,T(S)) = I(S;T) + \sum_t p(t)H(t(S))$.

The convergence of Algorithm B can be proved rigorously either by employing the same approach as that used in proving the convergence of the Algorithm B in Chapter 6, or by using directly results of [9, 14, 46], which state that a two step alternating algorithm converges to the global minimum if the optimization function is convex. Obviously, it is not hard to show that for fixed $p(x|u,s)$, $L(p(u|s), p(x|u,s), Q(u,x), Q(u|y))$ is convex over $(p(u|s), Q(u,x), Q(u|y))$. So the convergence of Algorithm B is obtained.

7.4 Summary

In this chapter by employing the idea of the Blahut-Arimoto algorithm and the Shannon strategy we developed algorithms for computing public watermarking capacities and compression rate functions, in other words, the algorithms proposed in this chapter can be combined to numerically calculate compression and watermarking rate regions for joint compression and public watermarking systems with finite memoryless covertext sources and fixed attack channels.

Chapter 8

Conclusions and Future Research

8.1 Conclusions

In digital watermarking, a watermark is embedded into a covertext resulting in a watermarked signal, which can be used for different purposes ranging from copyright protection, data authentication, fingerprinting, to information hiding. In all these cases, the watermark should be embedded in such a way that the watermarked signal is robust to certain distortion caused by either standard data processing in a friendly environment or malicious attacks in an unfriendly environment. In this thesis, we investigate digital watermarking from an information theoretic viewpoint and a numerical computation viewpoint respectively.

From the information theoretic viewpoint we study a new digital watermarking scenario, where a watermark correlated with a covertext is to be transmitted by embedding the watermark into the covertext. Assume that the watermark and the covertext are generated from a joint finite memoryless watermark and covertext source. In the case of public watermarking where the covertext is not accessible to the watermark decoder, a necessary and sufficient condition is determined under which the watermark can be fully recovered with high probability at the end of watermark decoding after the watermarked signal is

disturbed by a fixed memoryless attack channel. Interestingly, from the sufficient and necessary condition we show that watermarks still can be fully recovered with high probability even if the entropy of the watermark source is strictly above the standard public watermarking capacity. Therefore, in this sense the famous Shannon's separation theorem does not hold anymore.

The above research problem is generalized to joint compression and public watermarking scenario, where watermarks and covertexts are correlated, and the watermarked signals are been compressed further for sake of efficient transmission and/or storage. For a given distortion level between the covertext and the watermarked signal and a given compression rate of the watermarked signal, a necessary and sufficient condition is determined under which the watermark can be fully recovered with high probability at the end of watermark decoding after the watermarked signal is disturbed by a fixed memoryless attack channel and the covertexts is not available to the watermark decoder.

In some applications, it is reasonable that the reproduced watermark at the end of decoding is allowed to be within certain distortion of the original watermark. For the above joint compression and watermarking models with this less requirement, sufficient conditions are determined respectively, under which watermarks can be reproduced within a given distortion of the original watermarks at the end of watermark decoding after the watermarked signals are disturbed by a fixed memoryless attack channel and the covertexts are not available to the watermark decoder.

From the viewpoint of numerical computation, the well-known characterization of watermarking capacities and joint compression and watermarking rate regions as optimization problems does not mean that they can be calculated easily. Therefore, we first derive closed forms for watermarking capacities of private Laplacian watermarking systems with the magnitude-error distortion measure under a fixed additive Laplacian attack and under a fixed arbitrary additive attack, respectively. We then focus on algorithms development for numerically computing watermarking capacities and joint compression and watermark-

ing rate regions. Based on the idea of the Blahut-Arimoto algorithm for computing channel capacities and rate distortion functions, two iterative algorithms are proposed for calculating private watermarking capacities and joint compression and private watermarking rate regions. Similarly, based on both the Blahut-Arimoto algorithm and Shannon's strategy, iterative algorithms are proposed for calculating public watermarking capacities and joint compression and public watermarking rate regions.

8.2 Directions for Future Research

As a technique to protect copyright for digital content, digital watermarking has been recently one of the most active research fields in signal processing and information theory. From the viewpoint of information theory, there are still lots of open problems. Among them, the following questions will be studied in our future works:

- In Chapter 4, only sufficient conditions are determined for the case without compression of stegotexts and the case with compression of stegotexts respectively, under which watermarks can be reproduced within a given distortion level at the end of public decoder. But, we don't know whether the conditions are necessary or not. So, it is valuable to find sufficient and necessary conditions similar to that obtained in Chapter 2 and Chapter 3.

- How to study the problem in Chapter 4 in continuous case? In particular, it may be possible to derive a sufficient and necessary condition for Gaussian case.

- We will study reversible watermarking systems with correlated watermarks and covertexts similar to the watermarking systems investigated in [38, 39, 41].

- Investigate multi-user watermarking systems, such as multiple-access watermarking systems and broadcast watermarking systems.

Bibliography

[1] S. Arimoto,"An Algorithm for Computing the Capacity of Arbitrary Discrete Memoryless Channels", *IEEE Transactions Information Theory*, vol. 18, pp. 14–20, January 1972.

[2] R. E. Blahut,"Computation of Channel Capacity and Rate-Distortion Functions", *IEEE Transactions Information Theory*, vol. 18, pp. 460–473, July 1972.

[3] R. E. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley, Reading, MA,1987.

[4] C. Chang and L. Davisson, "On Calculating the Capacity of an Infinite-Input Finite(Infinite)-Output Channel", *IEEE Transactions Information Theory*, vol. 34, pp. 1004–1010, September 1988.

[5] B. Chen and G. W. Wornell, "Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding," *IEEE Transactions Information Theory*, vol. 47, pp. 1423–1443, May 2001.

[6] S. Cheng, V. Stankovic and Z. Xiong, "Computing the Capacity and Rate-distortion Functions with Two-Sided State Information," *IEEE Transactions Information Theory*, vol. 51, pp. 4418–4425, December 2005.

[7] A. S. Cohen and Amos Lapidoth, "The Gaussian Watermarking Game," *IEEE Transactions Information Theory*, vol. 48, pp. 1639–1667, June 2002.

[8] M. Costa, "Writing on Dirty Paper," *IEEE Transactions. Information Theory*, vol. 29, pp. 439–441, May 1983.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: John Wiley & Sons, 1991.

[10] I. Cox, M. Miller and J. Bloom, *Digital Watermarking*, Elsevier Science: Morgan Kaufmann Publishers, 2001.

[11] I. J. Cox, T. Kalker, G. Pakura and M. Scheel,"Information Transmission and Steganography", *Lecture Notes on Computer Science* 3710, pp. 15–29, 2005.

[12] I. Csiszar and J.Korner, *Information theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press New York, 1980.

[13] I. Csiszar,"On the Computation of Rate-Distortion Functions", *IEEE Transactions Information Theory*, vol. 20, pp. 122–124, January 1974.

[14] I. Csiszar, G. Tusnady, " Information Geometry and Alternating Minimization Procedures", *Statistics and Decisions*, Supplement Issue 1: 205–237, 1984.

[15] F. Dupuis, W. Yu and F. Willems,"Blahut-Arimoto Algorithms for Computing Channel Capacity and Rate-Distortion with Side Information", *Proceedings of IEEE International Symposium on Information Theory* , 27 June–2 July 2004, pp. 179.

[16] S. Gel'fand and M. Pinsker, "Coding for a Channel with Random Parameters", *Problems of Control and Information Theory*, vol. 9, pp. 19–31, 1980.

[17] J. Jahn, *Introduction to the Theory of Nonlinear Optimization*, second edition, Springer,1996.

[18] D. Karakos and A. Papamarcou, "A Relationship Between Quantization and Watermarking Rates in the Presence of Additive Gaussian Attacks", *IEEE Transactions Information Theory*, vol. 49, pp. 1970–1982, August 2003.

[19] D. Karakos, "Digital Watermarking, Fingerprinting and Compression: An Information-theoretic Perspective", PhD Dissertation, University of Maryland, 2002.

[20] A. Maor and N. Merhav, "On Joint Information Embedding and Lossy Compression," *IEEE Transactions Information Theory*, vol. 51, no. 8, pp. 2998–3008, August 2005.

[21] A. Maor and N. Merhav, "On Joint Information Embedding and Lossy Compression in the Presence of a Memoryless Attack," *IEEE Transactions Information Theory*, vol. 51, no. 9, pp. 3166–3175, September 2005.

[22] G. Matz and P. Duhamel, "Information Geometric Formulation and Interpretation of Accelerated Blahut-Arimoto-type Algorithms", *IEEE International Workshop on Information Theory*, 24–29 Oct. , 2004, pp. 66–70.

[23] N. Merhav, "On Random Coding Error Exponents of Watermarking Systems," *IEEE Transactions Information Theory*, vol. 46, no. 2, pp. 420–430, March 2000.

[24] N. Merhav and S. Shamai (Shitz), "On Joint Source-channel Coding for the Wyner-Ziv Source and the Gel'fand-Pinsker Channel," *IEEE Transactions Information Theory*, vol. 49, pp. 2844–2855, November 2003.

[25] P. Moulin and R. Koetter, "Data-Hiding Codes," *Proceedings of IEEE*, Vol. 93, No. 12, pp. 2083–2127, Dec. 2005.

[26] P. Moulin and J. A. O'Sullivan, "Information-Theoretic Analysis of Information Hiding," *IEEE Transactions Information Theory*, vol. 49, pp. 563–593, March 2003.

[27] P. Moulin and M.K. Mihcak, "The Parallel-Gaussian Watermarking Game," *IEEE Transactions Information Theory*, vol. 50, pp. 272–289, Feb. 2004.

[28] S. Pradhan, J. Chou and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Transactions Information Theory*, vol. 49, pp. 1181–1203, May 2003.

[29] M. Rezaeian and A. Grant, "A Generalization of Arimoto-Blahut Algorithm", *Proceedings of IEEE International Symposium on Information Theory*, 27 June–2 July 2004, pp. 181.

[30] M. Rezaeian and A. Grant, "Computation of Total Capacity for Discrete Memoryless Multiple-access Channels", *IEEE Transactions Information Theory*, vol. 50, pp. 2779–2784, Nov. 2004.

[31] D. Slepian and J. Wolf, "Noiseless Coding of Correlated Information Sources," *IEEE Transactions Information Theory*, Vol. 19, pp. 471–480, July 1973.

[32] A. Somekh-Baruch and N. Merhav, "On the Error Exponent and Capacity Games of Private Watermarking Systems," *IEEE Transactions Information Theory*, vol. 49, pp. 537–562, March 2003.

[33] A. Somekh-Baruch and N. Merhav, "On the Capacity Game of Public Watermarking Systems," *IEEE Transactions Information Theory*, vol. 50, no. 3, pp. 511–524, March 2004.

[34] W. Sun and E.-h. Yang, "Closed-Form Formulas for Private Watermarking Capacities of Laplacian Sources with the Magnitude-Error Distortion Measure and under Additive Attacks", *Lecture Notes on Computer Science* 3710, pp.361–371.

[35] A. Sutivong, T.M. Cover, M. Chiang, and Young-Han Kim, "Rate vs. Distortion Tradeoff for Channels with State Information," *Proc. IEEE Int. Symp. on Information Theory*, Lausanne, Switzerland, June 30–July 5, 2002, p. 226.

[36] S. Voloshynovskiy et al, "Data-hiding with Host State at the Encoder and Partial Side Information at the Decoder", Available online: http://research.microsoft.com/ kivancm/publications/sp05-slava.pdf.

[37] P. Vontobel, "A Generalized Blahut-Arimoto algorithm",*Proceedings of IEEE International Symposium on Information Theory*, 29 June–4 July, 2003, pp. 53.

[38] F. Willems and T. Kalker, "Methods for Reversible Embedding," *Proc. 40th Annual Allerton Conference on Communication, Control,and Computing*, Allerton House, Monticello, Illinois, Oct. 2–4, 2002.

[39] F. Willems and T. Kalker, "Coding Theorems for Reversible Embedding," *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 66, American Mathematical Society, pp. 61–76, 2004.

[40] F. Willems, "Information Theoretical Approach to Information Embedding," *Proc. 21st Symposium on Information Theory*, pp. 255–260, Wassenaar, The Netherlands, May 25–26, 2000.

[41] F. Willems and T. Kalker, "Semantic Compaction, Transmission, and Compression Codes," *Proc. IEEE Int. Symp. on Information Theory*, Adelaide, South Australia, Australia, 4–9 September, 2005. pp. 214–218.

[42] F. Willems, "Computation of the Wyner-Ziv Rate Distortion Function", *Eindhoven University of Technology Research Reports*, July 1983.

[43] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions Information Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[44] A. Wyner, "The rate-distortion function for source coding with side information at the decoder II: General sources," *Information and Control*, vol. 38, no. 1, pp. 60–80, July 1978.

[45] E.-h. Yang and W. Sun, "On Watermarking and Compression Rates of Joint Compression and Private Watermarking Systems with Abstract Alphabets", *Proceed-*

ings of Canadian Workshop on Information Theory, 5 June–8 June 2005, Montreal, Canada.

[46] R. W. Yeung, *A First Course in Information Theory*, New York: Kluwer, 2002.