

**Authenticating Multimedia In The Presence Of Noise**

by

**Emin Martinian**

B.S., University of California Berkeley (1997)

Submitted to the Department of Electrical Engineering and Computer Science  
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

**MASSACHUSETTS INSTITUTE OF TECHNOLOGY**

May 2000

© Massachusetts Institute of Technology 2000. All rights reserved.

Author .....  
Department of Electrical Engineering and Computer Science  
May 15, 2000

Certified by .....  
Greg Wornell  
Associate Professor Of Electrical Engineering  
Thesis Supervisor

Certified by .....  
Brian Chen  
Research Affiliate Of The Digital Signal Processing Group  
Thesis Supervisor

Accepted by .....  
Arthur C. Smith  
Chairman, Department Committee on Graduate Students



# **Authenticating Multimedia In The Presence Of Noise**

by

Emin Martinian

Submitted to the Department of Electrical Engineering and Computer Science  
on May 15, 2000, in partial fulfillment of the  
requirements for the degree of  
Master of Science in Electrical Engineering and Computer Science

## **Abstract**

The authenticity of multimedia such as photos, videos, and sound recordings is critically important in many applications. Unfortunately, due to the power and economy of modern digital processing tools, creating convincing forgeries is relatively easy. Consequently there is a need for multimedia authentication methods to restore credibility. Various researchers have proposed techniques based on digital watermarking, cryptography, and content classification. However, since authenticating multimedia is a fairly new idea, measures for evaluating different methods are not well defined.

After introducing some possible applications to clarify the problem and discussing some previously proposed methods, we present a framework for authenticating multimedia under the title “Analog Authentication”. Defining a precise problem including the performance measures, tradeoffs, and resources available provides a way to evaluate different schemes. In addition, we analyze the fundamental limits of what can and can not be done by any scheme. Our analysis yields a quantity called the fundamental distortion which serves as the dividing line between what is asymptotically achievable and what is impossible. This is analogous to how the capacity of a communications channel characterizes which transmission rates are achievable and which are impossible. In addition our analysis is constructive and yields insight into practical methods. Finally, we design some practical analog authentication schemes and evaluate their performance.

Thesis Supervisor: Greg Wornell  
Title: Associate Professor Of Electrical Engineering

Thesis Supervisor: Brian Chen  
Title: Research Affiliate Of The Digital Signal Processing Group

## Acknowledgments

I would like to thank my advisors Gregory Wornell and Brian Chen for their patience, dedication, and insight. They gave me the freedom to pursue my ideas, but provided enough guidance to prevent me from becoming lost. Also they carefully went over the drafts of this thesis and suggested many valuable improvements. Working with them was challenging, educational, and enjoyable. I could not ask for better advisors.

This material is based upon work supported under a National Science Foundation Graduate Fellowship. Consequently I would like to thank the NSF for its generous support.

In addition, I would like to thank the members of the Digital Signal Processing Group. J. Nicholas Laneman listened to my early ideas and gave me valuable advice and encouragement. I am also indebted to Nick for his clear explanation of digital communication subjects such as channel coding, multi-access techniques, and fading. Stark Draper was an excellent lunch companion and also provided a ready ear and a quick wit for discussions about information theory. Conversations with my office mate, Charles Sestok, about the stock market, our summer jobs, and life in Boston were a welcome relief from the rigors of MIT. The rest of my numerous colleagues at DSPG were a pleasure and an inspiration to work with. I am grateful to be a part of such an amazing and talented collection of people.

I also would like to thank my family and friends. My mother and father worked hard and sacrificed to bring me to this country, raise me, and make sure I received a good education. My aunt and uncle, Monica and Reuben Toumani, and my second family, the Malkins, provided much appreciated love, encouragement and support. Alex Aravanis, Jarrod Chapman, and Mike Malkin have my gratitude for their warm friendship and good company. The conversations we had about science, math and engineering at Berkeley are what inspired me to pursue graduate studies in engineering instead of seeking my fortune in another career.

Finally, I would like to thank my fiancée, Esme Baker, for her love, sense of humor, and friendship. Knowing her has made my life happier than I ever imagined. She has made my life in Boston a wonderful adventure. If not for her I would never have come to MIT.

# Contents

<b>1</b>	<b>Introduction</b>	<b>10</b>
1.1	Possible Applications . . . . .	10
1.2	Previous Research . . . . .	13
1.3	A Framework For Authenticating Multimedia . . . . .	15
1.4	Overview Of The Thesis . . . . .	16
<b>2</b>	<b>Basic Idea, Intuition, And Plan Of Analysis</b>	<b>19</b>
2.0.1	A Qualitative Example . . . . .	20
2.0.2	A Quantitative Example . . . . .	21
2.1	A General Framework For Any Analog Authentication Scheme . . . . .	25
2.1.1	Geometric View And Insights About The Decoder . . . . .	27
2.1.2	Geometric Insights About The Encoder . . . . .	28
<b>3</b>	<b>Analog Authentication: The Formal Problem Statement</b>	<b>33</b>
3.0.3	Probability Of Undetected Error . . . . .	35
3.0.4	Probability Of False Alarm Error . . . . .	38
3.0.5	Embedding Distortion . . . . .	38
3.1	Formal Problem Statement . . . . .	39
3.2	Results . . . . .	40
3.3	Analog Authentication Based On Complexity Assumptions . . . . .	43
3.3.1	Brief Summary Of Digital Signature Schemes . . . . .	43
3.3.2	Using Digital Signature Schemes For Analog Authentication . . . . .	44
<b>4</b>	<b>Calculating The Fundamental Distortion</b>	<b>47</b>
4.1	Binary Source, Hamming Distortion, BSC Channel . . . . .	48
4.2	Uniform Source, Circular Additive Noise, Squared Error Distortion . . . . .	50
4.3	Gaussian Source, Squared Error Distortion, AWGN channel . . . . .	53
4.4	Uniform Source, Squared Distortion, AWGN channel . . . . .	58

<b>5</b>	<b>Practical Implementation And Performance</b>	<b>61</b>
5.1	Algorithm Description: The Uncoded Case . . . . .	61
5.1.1	Input, Design, And Performance Parameters . . . . .	61
5.1.2	Description of Encoding . . . . .	62
5.1.3	A Small Example To Illustrate Encoding . . . . .	63
5.1.4	Description of Decoding . . . . .	65
5.1.5	A Small Example To Illustrate Successful Decoding . . . . .	67
5.1.6	A Small Example To Illustrate Forgery Detection . . . . .	67
5.1.7	A Small Example To Illustrate False Alarm Error . . . . .	68
5.1.8	Comments About The Decoding Examples . . . . .	69
5.1.9	Analysis of the Uncoded Case . . . . .	70
5.2	Algorithm Design: The Coded Case . . . . .	73
5.2.1	General Codes For Analog Authentication . . . . .	73
5.2.2	Encoding Using TCM/TCQ . . . . .	74
5.2.3	Analysis of the Trellis Coded Case . . . . .	75
<b>6</b>	<b>Conclusion</b>	<b>78</b>
<b>A</b>	<b>The Backward Compatibility Broadcast Channel</b>	<b>82</b>
A.1	Definitions . . . . .	82
A.2	Theorems For The Backward Compatibility Broadcast Channel . . . . .	84

# List of Figures

2-1	Diagram of encoding process. The original picture, $X_1^n$ , has $n$ pixels. Each pixel, $X_i$ is quantized to $Y_i = Q(X_i)$ with a uniform scalar quantizer. . . . .	21
2-2	Diagram of decoding process. The received picture, $Z_1^n$ , has $n$ pixels. Each pixel, $Z_i$ is a noisy version of $Y_i$ . To try to recover $Y_i$ , each pixel is requantized with $Q(\cdot)$ . . .	22
2-3	Channel model for the analog authentication problem. . . . .	24
2-4	The transmitter wants to send $X$ to the receiver. He processes $X$ to get $Y = Q(X)$ and sends it over the channel. The receiver gets the output of the channel $Z$ and computes $\hat{X} = D(Z)$ to estimate the original $X$ . . . . .	26
2-5	Plot of possible decoding function $D(Z)$ where $Z = (z_1, z_2)$ . The shaded area corresponds to $D(Z) = \emptyset$ and the unshaded area corresponds to an area where $D(Z)$ produces an estimate of $X$ . . . . .	27
2-6	To send, the transmitter must move $X$ from the forbidden region to the inside of an authentic cell. The transmitted signal will be $Y = Q(X)$ . . . . .	28
2-7	Two possibilities for $Q(X)$ . $Y_1 = Q_1(X)$ maps $X$ only part way to the nearest authentic cell centroid while $Y_2 = Q_2(X)$ maps $X$ directly to the nearest centroid. .	29
2-8	The shaded area is the forbidden region where $D(Z) = \emptyset$ . The original source, $X$ , is mapped to the center of the nearest sphere to get $Y = Q(X)$ . Channel noise causes the receiver to get $Z_1$ which is correctly decoded to $Y$ . A large amount of channel noise or an attempted forgery would correspond to $Z_2$ which would be rejected since it is in the forbidden region. . . . .	32
3-1	A diagram of the analog authentication problem. . . . .	33
3-2	The transmitter processes $X$ to get $Y = Q(X)$ and sends it over the channel. The sophisticated receiver gets the output of the channel $Z$ and computes $\hat{Y} = A(Z)$ to authenticate the message. . . . .	35
4-1	Channel probability distribution for additive white circular noise channel. . . . .	50
4-2	Codebook probability distribution for additive circular noise channel. . . . .	51

4-3	Minimum value of $\alpha$ for a given $\epsilon$ required to make $I(X; Y) \leq I(Y; Z)$ . . . . .	52
4-4	Comparison of distortion to noise ratio (DNR) at various signal to noise ratio (SNR)	57
4-5	Comparison of distortion to noise ratio (DNR) bounds at various signal to noise ratio (SNR). . . . .	60
5-1	Some example reconstruction points for $R = 3$ (top), and $R = 4$ (bottom). . . . .	63
5-2	Diagram of the encoding process. . . . .	64
5-3	Diagram of the decoding process. . . . .	66
5-4	Quantization points and decision regions. Any point in the region $(-2, 0]$ is mapped to the quantization point -1. Any point in the region $(0, 2]$ is mapped to the quantization point 1. . . . .	69
5-5	Uncoded probability of symbol error, $p_s$ , as a function of distortion to noise ratio (DNR). . . . .	72
5-6	Signal constellation for 8-PAM. The X's are the signal points. . . . .	75
5-7	Signal constellation for TCM. The X's are the signal points for coset 0, the O's are the signal points for coset 1, the #'s are the signal points for coset 2, and the \$ are the signal points for coset 3. . . . .	75
5-8	Coding gain achieved for TCM/TCQ. . . . .	76
A-1	A diagram of the backward compatibility broadcast channel. . . . .	83

# List of Tables

4.1	Comparison of distortion to noise ratio (DNR) at various signal to noise ratio (SNR)	56
-----	--	----

# Chapter 1

## Introduction

The authenticity of multimedia such as photos, videos, and sound recordings is critically important in many applications. Unfortunately, due to the power and economy of modern digital processing tools, creating convincing forgeries is relatively easy. Consequently multimedia authentication methods are needed to restore credibility.

We first discuss possible applications of analog authentication and then discuss some proposed techniques based on digital watermarking, cryptography, and content classification. Since authenticating multimedia is a fairly new idea, measures for evaluating different methods are not well defined. Therefore we present a general framework for analog authentication and outline the issues we plan to analyze.

### 1.1 Possible Applications

#### **Trustworthy Photographs**

Photographs are often used as evidence in courts, newspapers, magazines, TV, the Internet, etc. The relative ease with which photographs can be modified or faked presents problems. For example, imagine that Charlie wants to convince Bob about some fact using a particular photo. Bob does not trust Charlie so he is unwilling to trust the photograph that Charlie presents. In the past, photographs were trusted by the mere fact that they were photographs. Now that photos can be faked, trust must be based on the provider of the photograph. Hence it might seem that photos will be less useful as evidence.

Analog authentication schemes are designed to provide a way for creators and receiver of photographs to check authenticity. Imagine that Charlie's photo was taken by Alice who is a well-known photographer. Imagine that Bob trusts Alice. If Bob were given a photograph that he knew beyond reasonable doubt was produced by Alice, then he would trust that photo. Since Alice does not

necessarily have the time or desire to meet Bob in person, they need an authentication method. Alice should be able to encode her photo appropriately so that Bob can later check whether the photo presented by Charlie was in fact created by Alice.

If Alice had a digital camera she could use a conventional public key digital signature algorithm<sup>1</sup> to protect her photos with digital signatures. She could then distribute both the photos and their corresponding digital signatures, while posting her public key in a publicly accessible database. When Charlie wants to convince Bob that a photograph was in fact taken by Alice he gives Bob the photo and its signature so that Bob can independently verify the digital signature himself.

This technique would work if photos were always distributed in a digital format (e.g. if photos were always kept on a disk or sent via the Internet instead of being kept as a photographs on paper). There are many situations, however, where photos are distributed in an analog format. If the photograph is printed on photographic paper then verifying the digital signature becomes problematic.

When Bob wants to verify the signature he must scan the photo in to a computer. Call the original digital photo,  $P$ , and the version Bob obtains after scanning it into his computer,  $P'$ . If  $P$  and  $P'$  differ by even 1 bit due to noise in the printing and scanning process or a telltale smudge, then the digital signature created for  $P$  will not work for  $P'$ . Even if the photo is always kept in a digital format it might be modified slightly. For example, Alice could have originally stored  $P$  as a raw bitmap. Charlie might have compressed  $P$  to get  $P'$  using a lossy compression scheme such as JPEG. When Charlie presents  $P'$  to Bob, the digital signature created for  $P$  will not work for  $P'$ .

A good analog authentication technique should not classify a signal as a forgery when it is only slightly different from the original. Intuitively, if  $P$  and  $P'$  are almost the same then Bob would probably be willing to accept  $P'$  if he could have a guarantee that  $P'$  is almost the same as  $P$ . Specifically, if  $d(\cdot, \cdot)$  is a distortion metric, then we would say that  $P$  and  $P'$  are perceptually indistinguishable if  $d(P, P') \leq \epsilon$  and  $\epsilon$  is sufficiently small.

This highlights an important difference between authenticating multimedia and classic cryptographic authentication. Consider the scenario where Alice wants to send Bob a picture via the Internet. Alice can take the original picture, scan it into her computer to get,  $P$ , and sign it with a public or private key digital signature algorithm to get the signature,  $\sigma$ . She can then transmit the pair  $(P, \sigma)$  to Bob over a public network. If an enemy modifies either the picture,  $P$ , or the signature,  $\sigma$ , while the picture is in transit, then the probability that the enemy finds a valid pair  $(P_f, \sigma_f)$  is small. So when Bob checks the message–digital signature pair,  $(P_f, \sigma_f)$ , he will detect any forgery.

The analog authentication problem is different because the photo might suffer small perturbations

---

<sup>1</sup>Public key digital signature schemes are important in many authentication applications. Both [1] and [2] are classic papers on the subject. The FAQ at <http://www.rsa.com/rsalabs/faq> provides a brief tutorial on digital signatures. In addition we briefly summarize the key aspects of Digital Signature Schemes in Section 3.3.1

such as smudges, dirt, printing and scanning distortion, etc. Thus when Charlie presents the photo to Bob it will be slightly different from the original. Effectively there is a noisy channel between Alice and Bob. On the other hand, we model the Internet as a noise free bit pipe. Consequently, Alice and Bob need an analog authentication scheme which will be robust to small perturbations. Like conventional authentication methods, the scheme must be able to detect significant modifications or forgeries. Finally the encoding process must not add too much distortion to the signal being protected.

### **Trustworthy Archives**

Another application of analog authentication is trustworthy archives. Imagine that Charlie wants to convince Bob of some historical event. He takes Bob to the archives and together they search for a photograph of the event in question. If they find a photograph for the event they are interested in, how do they know if this photo is authentic? For example, how does Bob know that Charlie did not break into the archives the week before and plant a fake photograph? If the photo they find was allegedly taken by Alice, how do they verify this fact?

This scenario is almost exactly the same as the trustworthy photograph problem. One difference is that Alice might have passed away long before Bob and Charlie became interested in her pictures. In the trustworthy photograph problem it is conceivable that Bob and Charlie could ask Alice about her photograph. In the trustworthy archives problem, however, this might not be possible since Alice could have passed away. Thus an analog authentication scheme is even more essential for archives than contemporary multimedia.

### **Trustworthy Military Maps**

Accurate and authentic maps are essential in many military operations. If maps were protected using analog authentication, soldiers in the field would be able to check the authenticity of a map before using it. The authentication method would have to be robust to small perturbations such as creases or folds in the map, flecks of dirt, printing and scanning distortions, etc. In addition the scheme should be sensitive enough to detect tampering that significantly effects the map. The technical features of this scenario are the same as the trustworthy photograph problem.

### **Trustworthy Sound Bytes**

Analog authentication can be used for any kind of signal not just images. For example, imagine that Alice wants to make a short statement. She can record her statement and provide it to Charlie, a reporter. Charlie might maliciously or in the interest of brevity only report part of Alice's statement or otherwise edit the original<sup>2</sup>. If Alice protects her sound sample with an analog authentication

scheme, Charlie’s listeners can independently verify the authenticity of the sound sample.

The technical details of this scenario are the same as the previous examples. Alice needs an authentication method which is robust to acceptably small perturbations, yet still insures authenticity. These small perturbations could be introduced by re-sampling, compressing, or otherwise acceptably modifying the signal. Alternatively, if the sound sample is transmitted over AM or FM radio, then the signal would be corrupted by channel noise.

### **Trustworthy Identification Documents**

One of the main applications that motivated our work is the problem of trustworthy identification documents such as passports and drivers licenses. Many governments use ad hoc schemes such as special seals, holograms, special paper, etc. to try to add a measure of authenticity to such documents.

The problem of trustworthy identification documents is similar to the trustworthy photograph problem. For example, imagine that Alice issues a passport to Charlie. Charlie later shows his passport to a guard, Bob. Bob wants to verify that Charlie’s passport was in fact produced by Alice and not produced or modified by an unauthorized agent. The authentication scheme should be sensitive enough to detect tampering, but robust to perturbations due to lossy compression, printing and scanning distortion, smudges, etc. One possible difference between the trustworthy identification documents problem and the previous problems is that other forms of side information may be available. For example, a license number or a description of the photograph in words might be on the identification document as well.

## **1.2 Previous Research**

A number of different schemes have been proposed for the analog authentication problem. Conventional cryptographic authentication techniques by themselves are insufficient due to the small perturbations a signal might undergo. Consequently, many researchers have studied methods which incorporate cryptographic techniques in ways that are robust to small perturbations.

A popular idea involves combining fragile digital watermarking with cryptography. Techniques such as [3], [4], [5], and [6] embed secret markings called digital watermarks into the original signal. The embedded watermarks are designed to be imperceptible so that the watermarked signal does not suffer unacceptable distortion. In addition the watermarks are designed to be robust to noise

---

<sup>2</sup>Radio stations commonly edit interviews in the interest of brevity. Words like “umm”, “uh” and other such pauses are cut under the premise that these words convey no information. Since speech patterns can provide information such as whether a speaker is quick-thinking, hasty, deliberate, inarticulate, some might argue that the edited sample is a forgery.

yet fragile to tampering. The watermarks are embedded such that creating them is difficult for an enemy.

Authentication is done by first extracting the watermark. When the watermark is only slightly degraded, the signal is declared authentic and the slight degradation is ascribed to noise, lossy compression or other acceptably small perturbations. When the watermark is substantially degraded, the signal is considered to be a forgery. This requires a fine balance for the watermark. It must be robust to acceptably small perturbations, yet fragile enough that significant tampering destroys the watermark.

These techniques address three main goals of analog authentication: robustness, fidelity, and security. The robustness of digital watermarks enable these schemes to overcome acceptably small perturbations introduced by noise, lossy compression, etc. Since the process of embedding the digital watermarks is designed to induce an imperceptible amount of distortion, the protected signals are faithful to the originals. By assuming that only the transmitter and receiver know how to embed and extract valid watermark these schemes can achieve security in a shared secret key setting. Alternatively, by assuming the enemy is computationally bounded, these schemes can achieve security by using conventional cryptography in choosing, embedding, and extracting the digital watermarks. This ensures that an enemy can not create a valid watermark in either the public or secret key setting. The watermark is designed to be fragile so that significant modifications of a valid signal will destroy the watermark and prevent the modified signal from being declared authentic.

Other researchers have focused on content-based methods [7], [8], [9]. The idea behind these techniques is to extract a concise representation of the signal. The concise representation is then protected using digital signature techniques from conventional cryptography and then embedded back in the signal via digital watermarking. Methods of extracting concise representations include extracting edge maps, JPEG compression, robust bit extraction functions [10], and approximate message authentication codes [11].

The receiver performs authentication by extracting the concise representation from the watermark and checking the digital signature. If the digital signature is valid, the receiver then extracts a concise representation from the received signal and compares it to the embedded version. If the two concise representations match, then the signal is declared authentic.

Content-based methods achieve robustness by assuming the concise representation of an image is not changed much by small perturbations. If the received signal has not been significantly altered, then the concise representations of the received signal should match the concise representations of the transmitted signal. Since a concise representation of the original is embedded using a digital watermark it should be robust to noise. Therefore the receiver will be able to successfully extract the original concise representation and compare it to the concise representation of the received signal. Since the concise representation of the original is protected with a digital signature, it would be

computationally infeasible for an enemy to create a forgery.

There are some interesting parallels between the fragile watermarking methods and content-based methods. Both combine conventional cryptography with other techniques for robustness. Both are designed so that the encoding process does not introduce too much distortion. However, there is also an important contrast. The fragile digital watermarking methods could be classified as channel coding methods. They encode the signal to be robust to a particular type of noise which is considered acceptable while being fragile to another type of noise which is classified as malicious tampering. Content-based methods take a dual view analogous to source coding. These methods extract concise representations and attempt to communicate them reliably. The focus is on designing the concise representations to capture the salient features of the source yet being compact enough to be robustly embedded in a watermark. This raises the question of whether the two ideas can be combined. As we will show, a combined source-channel approach can achieve good results.

### 1.3 A Framework For Authenticating Multimedia

Most researchers have analyzed particular schemes, instead of focusing on the larger questions of what is theoretically possible and impossible. This makes it difficult to determine if the proposed schemes are performing as they are due to inherent limitations or because better schemes exist but are undiscovered. For example, is there an inherent tradeoff between the probability of false alarm and the probability of detecting forgeries? Is there an inherent tradeoff between how much encoding distortion is necessary and the other issues?

In order to compare different analog authentication schemes and study fundamental limitations we need to precisely identify the resources and goals. In discussing possible applications, we have already mentioned that some of the important parameters are the encoding distortion, security and robustness to noise. The previous research in [7], [8], [3], [4], [5], and [6] also addresses these criteria.

Most authentication techniques apply digital watermarking or other kinds of processing which introduce allegedly imperceptible distortion. To compare different schemes, we need to precisely define the amount of distortion which qualifies as acceptable. Some applications might be willing to tolerate more distortion than others. Analyzing the tradeoff between the security or robustness each scheme can achieve with a fixed amount of processing distortion is also important.

Security is an important goal for analog authentication schemes. Thus defining a quantitative measure of security is necessary. Since some perturbations are considered acceptable while others are considered forgeries, defining security precisely is not trivial. For example, most researchers assume that certain operations such as lossy compression, printing and scanning, additive noise, etc. are harmless perturbations. Conversely, intentional modifications with malicious intent constitute a forgery.

What if an enemy uses very lossy compression to blur an image in order to obscure an important detail. Should this be considered a forgery? Alternatively, consider the case where Alice prints a picture for Charlie. The printing process induces some small distortions. Bob is expecting these small distortions and is willing to tolerate them. Charlie uses image processing techniques to denoise his picture in order to reduce the printing distortions and produce a higher quality picture than Bob expects. Should this be considered a forgery?

Defining what constitutes a forgery based on the *intentions* or *tools* of a possibly enemy does not yield a satisfying definition. It seems that the concept of a forged signal should be independent of how or why it was produced. For example, assume that  $P_f$ , which was created by modifying  $P$ , is considered to be a forgery. Whether  $P_f$  was created by an active enemy or whether  $P_f$  resulted from a particularly noisy but random channel should be irrelevant. One possible definition of a forgery would be to define  $P_f$  as a forgery if the distance between  $P_f$  and  $P$  is greater than a certain threshold. We are not necessarily advocating this definition, but simply pointing out that the concept of a forgery needs to be put on a firm foundation.

## 1.4 Overview Of The Thesis

In Chapter 2 we discuss the basic ideas and intuition behind analog authentication. We compare conventional authentication via cryptography with analog authentication pointing out the similarities and differences between the two. This leads to an interpretation of analog authentication as combined error correction and error detection. The error correction provides noise robustness while the error detection uses cryptographic techniques to prevent forgeries.

We examine two detailed examples to develop some intuition for analog authentication before precisely defining the goals and available resources for analog authentication. This leads to a geometric framework which can be used to analyze and compare analog authentication schemes. We study the implications of this geometric view in regards to defining the goals and resources for analog authentication. This view illustrates that the analog authentication problem corresponds to joint source-channel coding combined with authentication.

In Chapter 3 we build upon the intuition developed in Chapter 2 to formally define the analog authentication problem. The key features of this definition are the embedding distortion allowed to the transmitter, the probability of forgery detection, and the probability of false alarm. The formal problem statement provides a way to compare different analog authentication schemes.

We analyze the asymptotic behavior of analog authentication schemes by defining a quantity called the fundamental distortion

$$D^* = \min_{p(y|x): I(X;Y) - I(Y;Z) \leq 0} E[d(X, Y)]$$

We prove a coding theorem which states that for any encoding distortion  $D > D^*$ , secure and reliable analog authentication schemes exist. In addition we prove a converse which states that for any encoding distortion  $D < D^*$ , secure and reliable analog authentication is impossible. Thus the fundamental distortion characterizes the analog authentication problem similar to how the channel capacity characterizes the classic reliable communication problem.

In Chapter 4 we examine the fundamental distortion for various source and channel models. This analysis illustrates the role of joint source-channel coding in analog authentication. We calculate bounds on the fundamental distortion for some representative models including Gaussian and uniform sources over AWGN channels. Our bounds show that the squared error distortion needed for analog authentication in these models is on the same order as the channel noise variance. For situations where the channel noise is acceptably small this indicates that analog authentication can be achieved with relatively small encoding distortions.

In chapter 5 we design and evaluate practical analog authentication schemes for a uniform source transmitted over an AWGN channel where mean square error is the distortion measure. We first design an uncoded scheme to illustrate how to apply the theoretical results and provide a baseline scheme for reference. We develop a normalized measure of the distortion to noise ratio similar to the rate normalized  $SNR_{norm}$  used in classical communications. By plotting the probability of error against the distortion to noise ratio we show that there is a significant gap to the fundamental distortion at reasonable error rates.

In order to close the gap to the fundamental distortion we discuss how to modify the uncoded scheme to introduce coding. We discuss how the duality between source and channel coding applies to codes for the analog authentication problem. As a result of this duality and the joint source-channel coding nature of analog authentication, we describe how either a source code or a channel coding can be used.

After describe how to incorporate a general coding scheme, we discuss an implementation based on trellis codes. Since trellis codes have been studied as both channel codes and source codes, they serve as a representative example. We present simulation results that show trellis codes can achieve a coding gain between 3 to 5 dB for reasonable design parameters.

In Chapter 6 we summarize our results and discuss directions for further research. Since the results we obtained for the fundamental distortion were based on information theoretic arguments, many generalizations are possible based on ideas such as side channels, universal coding, water-filling for colored source or colored distortion models, etc.

In Appnedix A we discuss a type of multi-access communication scenario called the backward compatibility broadcast channel (BCBC). The BCBC models the scenario where a transmitter broadcasts to two receivers. One receiver has a decoder while the other does not. The transmitter must choose a coding scheme which achieves reliable communication while limiting the processing distor-

tion suffered by the receiver which lacks a decoder. The results for the BCBC are almost identical to results for the analog authentication problem. Specifically, reliable communication is possible only for processing distortions  $D > D^*$ .

The BCBC is an interesting communication scenario independent of analog authentication. However, the BCBC also serves as a useful tool in understanding analog authentication. The BCBC effectively models the noisy channel aspect of analog authentication. Analog authentication can be interpreted as using coding for the BCBC combined with conventional cryptographic authentication schemes such as digital signatures.

## Chapter 2

# Basic Idea, Intuition, And Plan Of Analysis

In Section 1.1 we pointed out one way the analog authentication problem differs from conventional cryptographic authentication schemes. In analog authentication there is a noisy channel between Alice and Bob. For example, imagine that Alice issues a passport to Charlie who later shows the passport to Bob. The channel noise consists of smudges, printing and scanning distortion, and other perturbations which degrade the passport picture. We will call this the passport channel. In contrast, when using conventional authentication to send a file over the Internet, the channel between Alice and Bob is modeled as a noise free bit pipe.

In reality the transmission links that make up the Internet are also noisy channels. However, the low level protocols of the Internet use error correction to provide what looks like a noiseless channel to higher level applications. Digital signatures are then used to provide message authentication over the bit pipe. Thus the previously solved scenario of authentically transmitting a picture over the Internet seems closely related to the analog authentication scenario.

We can think of the analog authentication problem as first using coding to turn the appropriate noisy channel into a noiseless bit pipe and then applying standard digital signature techniques. One of the issues is how much processing distortion is allowed. Alice wants to turn the passport channel into a bit pipe, but she still wants the protected passport photo to look like the original.

When sending a picture over the Internet, more noise implies fewer bits can be communicated reliably. A noisier channel will require adding more redundancy in the form of error correcting codes and thus lowering the transmission rate. For example, a very noisy channel might allow  $R$  bits while a clean channel might allow  $100R$  bits. To transmit a picture through the very noisy channel we would have to compress the picture a great deal, while on the clean channel we could transmit a high fidelity picture. If the passport channel is very noisy we could use JPEG compression with a quality

of 10% while on a very clean passport channel we could use JPEG compression with a quality of 95%. Thus the noisier the channel the higher the processing distortion must be. This turns out to be the case for analog authentication as well.

### 2.0.1 A Qualitative Example

Imagine that Alice wants to issue a passport for Charlie. We model the original passport picture,  $X_1^n$ , as  $n$  real numbers. Alice quantizes  $X_1^n$  using a uniform scalar quantizer with step size  $\Delta$  to get  $Y_1^n = Q(X_1^n)$  as shown in Figure 2-1. The signal  $Y_1^n$  is represented as a sequence of bits, and Alice creates a digital signature,  $\sigma$ , by signing  $Y_1^n$ . She prints the picture corresponding to  $Y_1^n$  on the passport and attaches the digital signature to the passport.<sup>1</sup> Charlie presents the passport to Bob. We refer to the picture on the passport that Bob sees as  $Z_1^n$ . To authenticate the received picture, Bob quantizes  $Z_1^n$  as shown in Figure 2-2 and checks if the received digital signature matches  $Q(Z_1^n)$ . If so, he accepts  $Q(Z_1^n)$ , otherwise he rejects it as a forgery.

If the passport picture is perturbed slightly by smudges, or other types of channel noise, quantizing the received signal should still yield  $Q(Z_1^n) = Y_1^n$  so the digital signature will match. For example, we could model these effects as additive noise:  $Z_i = Y_i + V_i$ , where  $V_i$  is the noise value at the  $i$ th pixel. If each noise element,  $V_i$ , is small no error will be made. Specifically,  $Q(Z_1^n) = Y_1^n$  if and only if  $V_i < \frac{\Delta}{2}$  for each  $i$ .

If Charlie tries to significantly modify the passport, then quantizing will result in  $Q(Z_1^n) \neq Y_1^n$ . In this case, Bob will discover that the passport is a forgery when he tries to verify the digital signature,  $\sigma$ , against the quantized passport picture  $Q(Z_1^n)$ . The quantization step size,  $\Delta$ , will affect the noise robustness and processing distortion, but not the security to forgery. Choosing a larger quantization step size,  $\Delta$ , will result in greater noise robustness and increase the processing distortion. Regardless of the value of  $\Delta$ , strong digital signatures will make the system secure against forgery. The quantization provides the error correction to turn a noisy channel into a reliable error free bit pipe. Digital signatures are then used in conjunction with the error free bit pipe for authentication and security.

The uniform scalar quantizer example illustrates the basic principle but there is no reason to believe it is the best way to apply this principle. For example, if the source is a sequence of independent, identically, distributed Gaussian random variables, a non-uniform quantizer or vector quantizer might do better. In general we expect that the quantizer design should take into account the properties of the source. In this sense analog authentication is closely related to source coding. Alternatively, we expect that the quantizer design should take into account the statistics of the noise.

---

<sup>1</sup>For example, the digital signature could be embedded in the picture using digital watermarking. Alternatively, the digital signature could be attached to a separate part of the passport using a bar code. We explore how the signature is attached later in the thesis.

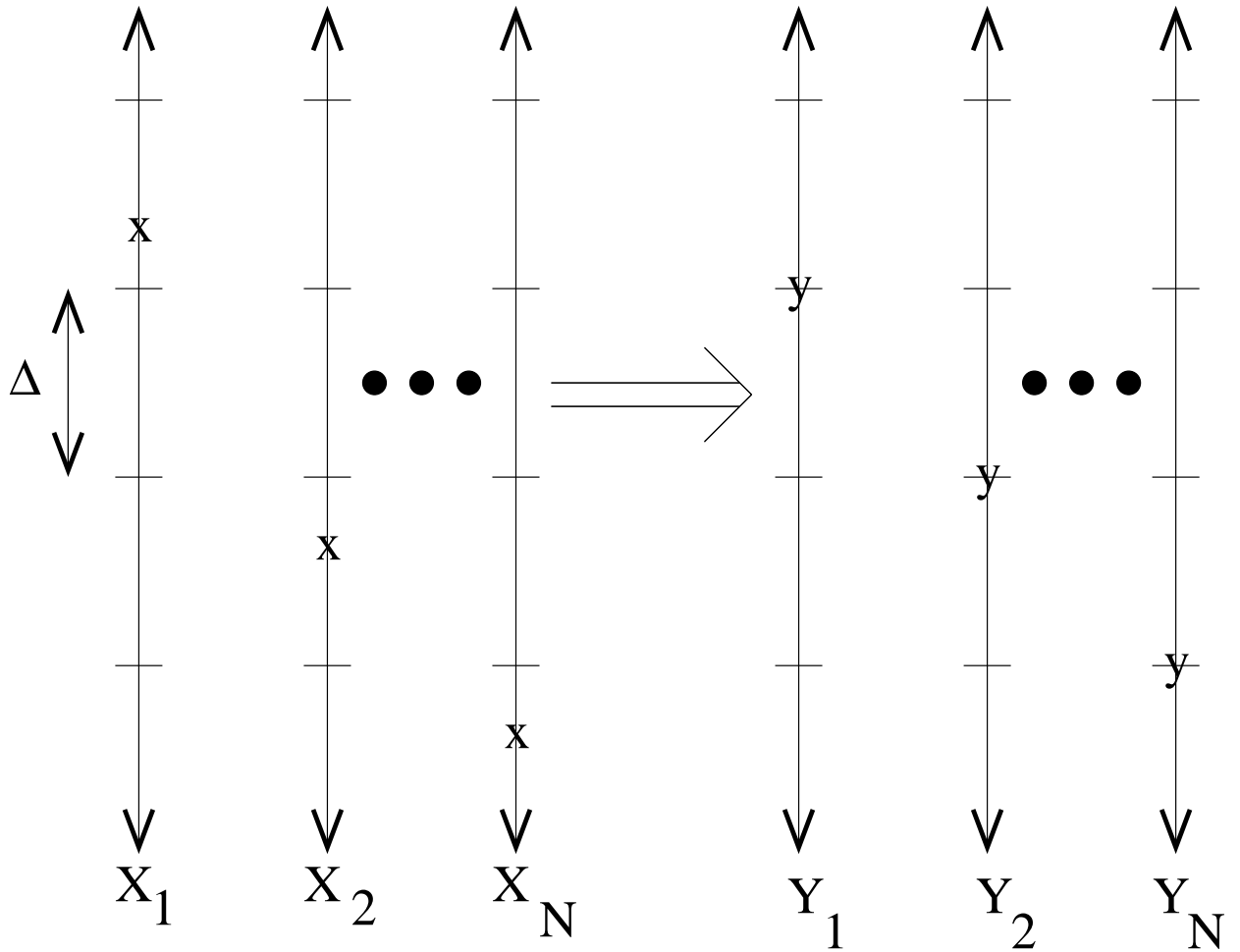


Figure 2-1: Diagram of encoding process. The original picture,  $X_1^n$ , has  $n$  pixels. Each pixel,  $X_i$  is quantized to  $Y_i = Q(X_i)$  with a uniform scalar quantizer.

In this sense analog authentication is like channel coding. As we will show, the optimal quantizer design should take into account both the source statistics and the noise statistics which corresponds to joint source-channel coding.

## 2.0.2 A Quantitative Example

Consider an i.i.d. Bernoulli 1/2 source  $X_1^n$  transmitted over a binary symmetric channel with cross over probability  $\epsilon$ . Alice is given a realization of the source which she processes to get  $Y_1^n$  and sends it through the channel. Bob receives the output  $Z_1^n$ . Distortion is measured using Hamming distance. For example, the distortion between  $X_1^n$  and an estimate,  $\hat{X}_1^n$ , is

$$D_e = E \left[ \frac{1}{n} \sum_{i=1}^n d_H(X_i, \hat{X}_i) \right]$$

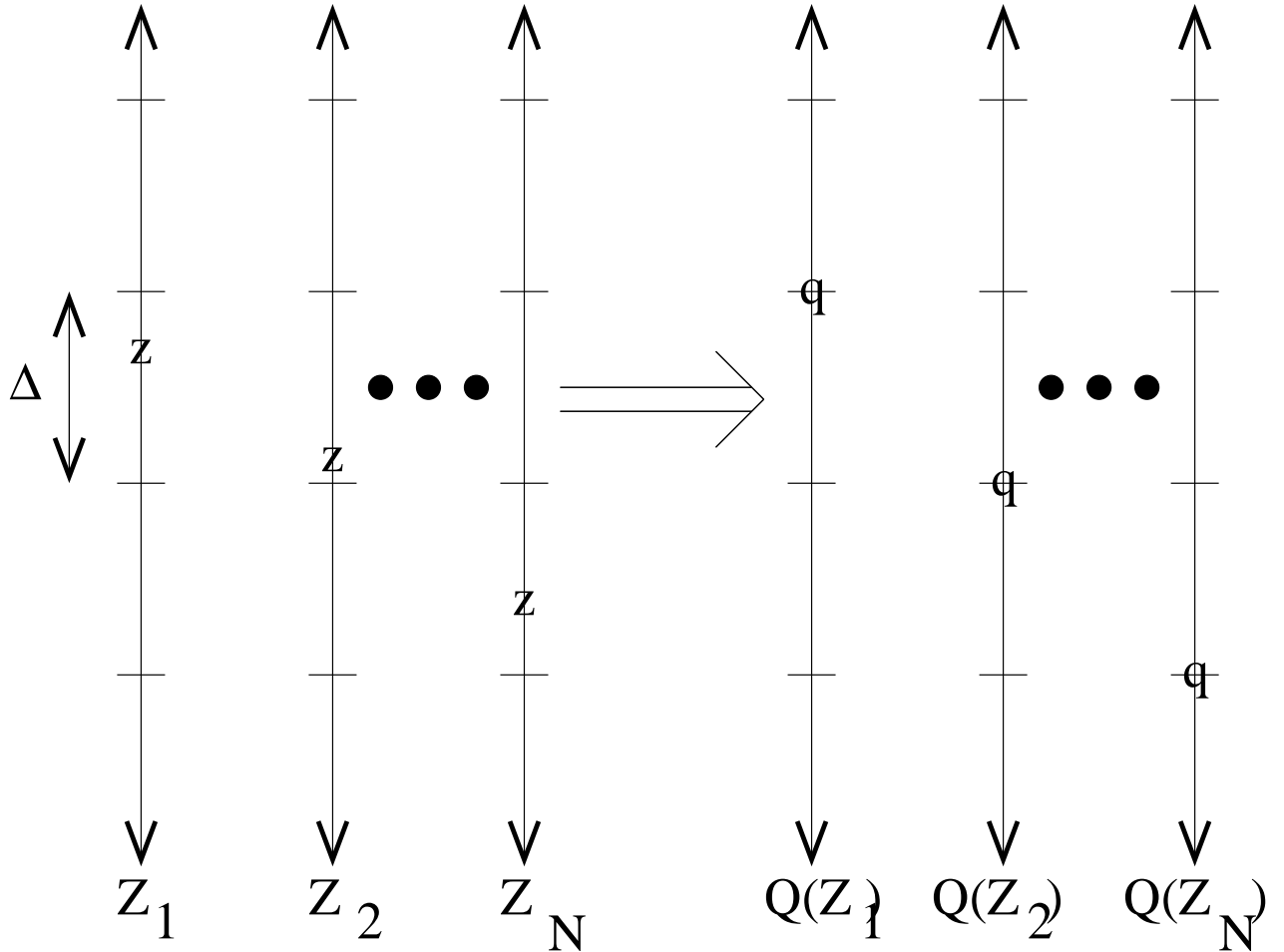


Figure 2-2: Diagram of decoding process. The received picture,  $Z_1^n$ , has  $n$  pixels. Each pixel,  $Z_i$  is a noisy version of  $Y_i$ . To try to recover  $Y_i$ , each pixel is requantized with  $Q(\cdot)$ .

where  $d_H(X_i, \hat{X}_i)$  is the Hamming distance between  $X_i$  and  $\hat{X}_i$ .

The capacity of a binary symmetric channel is well known:  $C = 1 - H_b(\epsilon)$  where  $H_b(\epsilon)$  is the binary entropy function. Similarly the rate-distortion function for an i.i.d Bernoulli 1/2 source is well known:  $R(D_e) = 1 - H_b(D_e)$  for  $D_e < 1/2$ . Without loss of optimality, source coding and channel coding can be done separately for a memoryless channel. If the goal were reliable communication over a noisy channel, all distortions which satisfy  $R(D_e) \leq C$  are achievable. Since analog authentication is more than just reliable communication, the minimum distortion any analog authentication scheme can achieve between  $X_1^n$  and  $\hat{X}_1^n$  is  $D_e \geq \epsilon$ .

The solution to the reliable communication problem corresponds to first quantizing the source using lossy source coding followed by channel coding. To adapt the approach to analog authentication in the passport example, Alice could take a passport picture and quantize it using lossy compression (e.g. JPEG). She could then use a digital signature scheme to protect the compressed file. To do channel coding, she could put this through an error correction code and print the result as a bar

code in place of the original passport picture. To decode the picture, Charlie would scan the bar code into a computer, decode the error correction code, check the digital signature and uncompress the picture.

This does not correspond to the analog authentication scenario we described. We described the problem as adding small marks or modifications to a signal which facilitate authentication. We required that the modifications be imperceptible or add a perceptually indistinguishable amount of distortion. Clearly transforming a passport picture into a bar code does not qualify as imperceptible distortion.

One of the motivations for requiring the encoding process to add an imperceptible amount of distortion is that there are two kinds of receivers for analog authentication signals. Sophisticated receivers will have the appropriate decoding equipment, while simple receivers will not. In the passport example, a guard at a high security airport might have the necessary decoding equipment to scan a bar code, decode an error correction code, verify a digital signature and reconstruct a JPEG file. However, a guard in a rural airport might not have the equipment to decode this kind of signal. If the passport picture suffers imperceptible amounts of distortion in the encoding process, the simple receiver will still be able to use the passport picture. On the other hand, if the passport picture is replaced by a bar code then the passport will not be useful to the simple receiver.

Figure 2-3 shows a diagram of the analog authentication channel. The source,  $X$ , is processed by the transmitter to obtain  $Y$  which is sent over the channel to both the simple receiver and the sophisticated receiver. The sophisticated receiver can decode the received signal. Consequently, the sophisticated receiver is interested in the distortion between the original signal,  $X$ , and the decoded version,  $\hat{X}$ . The simple receiver does not have a decoder available and is therefore interested in the distortion between the original signal,  $X$ , and the received signal  $W$ . The simple receiver views any processing done by the transmitter as a hindrance because it increases the distortion between  $X$  and  $Y$  and hence also increases the distortion between  $X$  and  $W$ . For example, if channel 1 in Figure 2-3 is a noiseless channel, then the distortion between  $X$  and  $W$  is exactly the same as the distortion between  $X$  and  $Y$ .

To make the transmitted signal acceptable to the simple receiver the transmitter must not distort the signal too much. We represent this notion with a constraint on the processing distortion:

$$E \left[ \frac{1}{n} \sum_{i=1}^n d_H(X_i, Y_i) \right] \leq D_p$$

where we use the symbol  $D_p$  to distinguish the constraint on the processing distortion from the distortion between the original source and the sophisticated receiver's estimate. Constraining the processing distortion to be small corresponds to favoring the needs of the simple receiver while allowing the processing distortion to be large corresponds to favoring the needs of the sophisticated

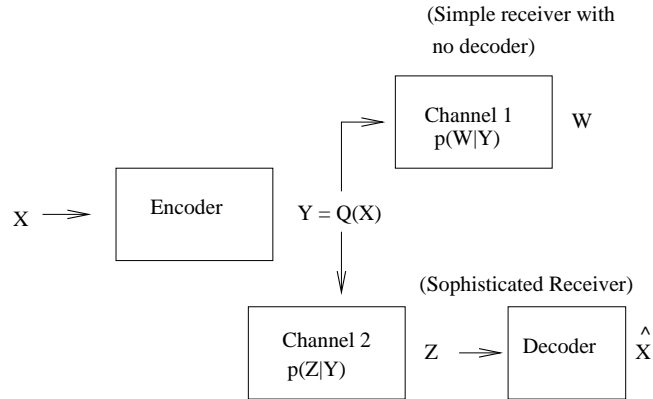


Figure 2-3: Channel model for the analog authentication problem.

receiver.

If the transmitter only catered to the sophisticated receiver, the transmitter could do lossy source coding followed by channel coding. As described previously, this would allow the transmitter to communicate the source to the sophisticated receiver with distortion  $D_e \geq \epsilon$ . Since analog authentication adds an additional constraint, this serves as a lower bound for the analog authentication problem. Later we will make these arguments precise, but for now we proceed to try to gain some intuition from this example.

Since  $D_e \geq \epsilon$  serves as a lower bound, it is natural to ask whether this lower bound is tight and to ask how  $D_e$  varies with the constraint on the processing distortion  $D_p$ . We will show that if  $D_p > \epsilon$ , then this lower bound is tight and all  $D_e \geq \epsilon$  are achievable. We will also show that if  $D_p < \epsilon$  then reliable analog authentication is not possible.

Consider the random coding argument used to prove the rate-distortion theorem in [12]. To compress  $X_1^n$  to rate  $R$ ,  $2^{nR}$  codewords with a distribution  $p(y|x)$  are chosen. As long as  $R > I(X; Y)$ , there will be a codeword which is distortion typical with the source with high probability. So the distortion will be  $E[d_H(X, Y)]$ . This is the source coding view.

Next consider the random coding argument used to prove the channel capacity theorem in [12]. To transmit reliably at rate  $R$ ,  $2^{nR}$  codewords are chosen according to a distribution  $p(y) = \sum_x p(y|x)p(x)$ . As long as  $R < I(Y; Z)$ , with high probability only the transmitted codeword will be jointly typical with the received sequence  $Z$ . This is the channel coding view.

As long as  $p(y|x)$  is chosen so that

$$I(X; Y) < R < I(Y; Z) \tag{2.1}$$

the receiver will be able to estimate  $Y$  with small probability of error and the distortion between

$X_1^n$  and  $Y_1^n$  will be  $E[d(X, Y)]$ . If

$$p(y|x) = \begin{cases} D_p, & y = 1 - x \\ 1 - D_p, & y = x \end{cases}$$

then  $I(X; Y) = 1 - H_b(D_p)$  and  $I(Y; Z) = 1 - H_b(\epsilon)$ . This implies that for  $D_p > \epsilon$ ,  $R$  can be chosen to satisfy (2.1).

Using this choice for  $p(y|x)$ , a codebook can be created that is simultaneously a good lossy source code and a good error correction code. The encoder produces a sequence,  $Y_1^n$  which will be within distortion  $D_p$  of  $X_1^n$ . This satisfies the simple receiver. The sophisticated receiver can estimate  $Y_1^n$  from  $Z_1^n$  with small probability of error. Since  $Y_1^n$  is within distortion  $D_p$  of  $X_1^n$ , the sophisticated receiver's decoded sequence,  $\hat{Y}_1^n$  is a good estimate of the source and  $D_e = D_p$ . Furthermore, the sophisticated receiver can check if the digital signature for  $Y_1^n$  matches the decoded sequence  $\hat{Y}_1^n$  to verify security.

This implies that for  $D_p > \epsilon$ , the scheme can achieve security while keeping the processing distortion acceptably small. Conversely, if  $D_p < \epsilon$  then no scheme will be able to achieve reliable communication. This is because there will exponentially many transmitted codewords which can not be distinguished by the decoder. This will preclude the use of digital signature schemes for authentication. We make these ideas precise in chapter 3.

## 2.1 A General Framework For Any Analog Authentication Scheme

In this thesis, we present a particular characterization of the analog authentication problem. While we present arguments to justify our definition of the problem, other formulations are possible. Figure 2-4 shows a diagram of an analog authentication scenario which applies to our formulation yet is general enough to include a broad range of formulations. There is an original source,  $X$ , which is processed by the transmitter to obtain the transmitted signal,  $Y$ . The transmitted signal is then sent over a noisy and possibly insecure channel. The received signal is  $Z$ . The receiver can do two things. He can try to extract an estimate of  $\hat{X} = D(Z)$  from  $Z$ , and he can try to decide if his estimate is authentic. Ideally, the receiver would always like to extract an estimate such that  $\hat{X} \approx X$ . For example, if  $Y$  undergoes a slight perturbation, the receiver would like to get a good estimate of  $X$  from  $Z$  to correct for the perturbation.

This might not be possible in some circumstances. For example, imagine that Alice wants to print a picture of Charlie onto Charlie's passport. She takes the original picture,  $X$ , processes it to get  $Y$  and puts  $Y$  onto Charlie's passport. Charlie might replace  $Y$  with  $J$ , a picture of John. John

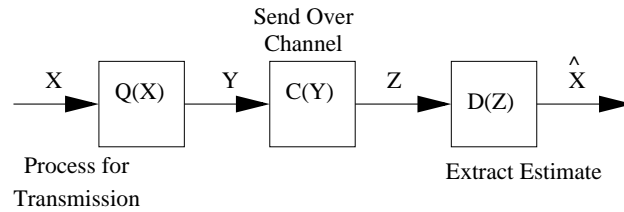


Figure 2-4: The transmitter wants to send  $X$  to the receiver. He processes  $X$  to get  $Y = Q(X)$  and sends it over the channel. The receiver gets the output of the channel  $Z$  and computes  $\hat{X} = D(Z)$  to estimate the original  $X$ .

might then try to use the forged passport. If Bob is a guard checking the passport, his received signal is  $Z = J$ . If  $J$  and  $Y$  are independent, there is no way for Bob to extract a good estimate of  $X$  from  $Z = J$ . Even though Bob would be unable to extract an estimate of the original, he would like to at least detect this and declare that he does not trust the received signal. In this case the decoder outputs a special symbol,  $D(Z) = \emptyset$ , to indicate it can not extract an authentic estimate.

Thus analog authentication can be thought of as a kind of combined error correction and error detection. When the receiver extracts an estimate of  $X$  from the received signal,  $Z$ , it is performing a kind of error correction. On the other hand, sometimes the error is so bad that error correction is not possible. In these cases we want the receiver to realize that whatever estimate the decoder makes will be an error and instead output  $D(Z) = \emptyset$ . This is a general model which captures the salient features of analog authentication.

One possible extension is to have a level of authenticity for the estimate instead of outputting either an estimate (which is implicitly assumed to be authentic) or outputting  $\emptyset$ , indicating that no authentic estimate is possible. This could be considered soft authentication as opposed to hard authentication. In practice the receiver would probably use only those estimates which had an authenticity level above a threshold and discard others. Consequently the authenticity level combined with the threshold would be equivalent to hard authentication. We only analyze hard authentication versions of the analog authentication problem, and postpone the study of soft authentication for future work.

An analog authentication scheme is completely specified by choosing the functions  $Q(X)$  and  $D(Z)$  shown in Figure 2-4. This might seem like a trivial observation, but it provides a geometric picture which can be used to reason about any kind of analog authentication scheme. Imagine that we are trying to analyze two different analog authentication schemes. Since the choice of  $Q(X)$  and  $D(Z)$  completely specify each scheme, we can ask

What are  $Q(X)$  and  $D(Z)$  for the two schemes? Why is one choice of  $Q(X)$  better than the other? Why is one choice of  $D(Z)$  better than the other?

### 2.1.1 Geometric View And Insights About The Decoder

Consider a source,  $X$ , which is a pair of real numbers,  $(x_1, x_2)$ . Recall that  $D(Z)$  either outputs an estimate of  $X$  or it outputs  $\emptyset$  indicating that it can not decode. By identifying  $Z$  with the pair  $(z_1, z_2)$  we can plot the behavior of  $D(Z)$  in a plane as shown in Figure 2-5. The shaded area corresponds to regions where  $D(Z) = \emptyset$ , and the unshaded regions correspond to regions where  $D(Z)$  produces an allegedly authentic estimate of  $X$ .

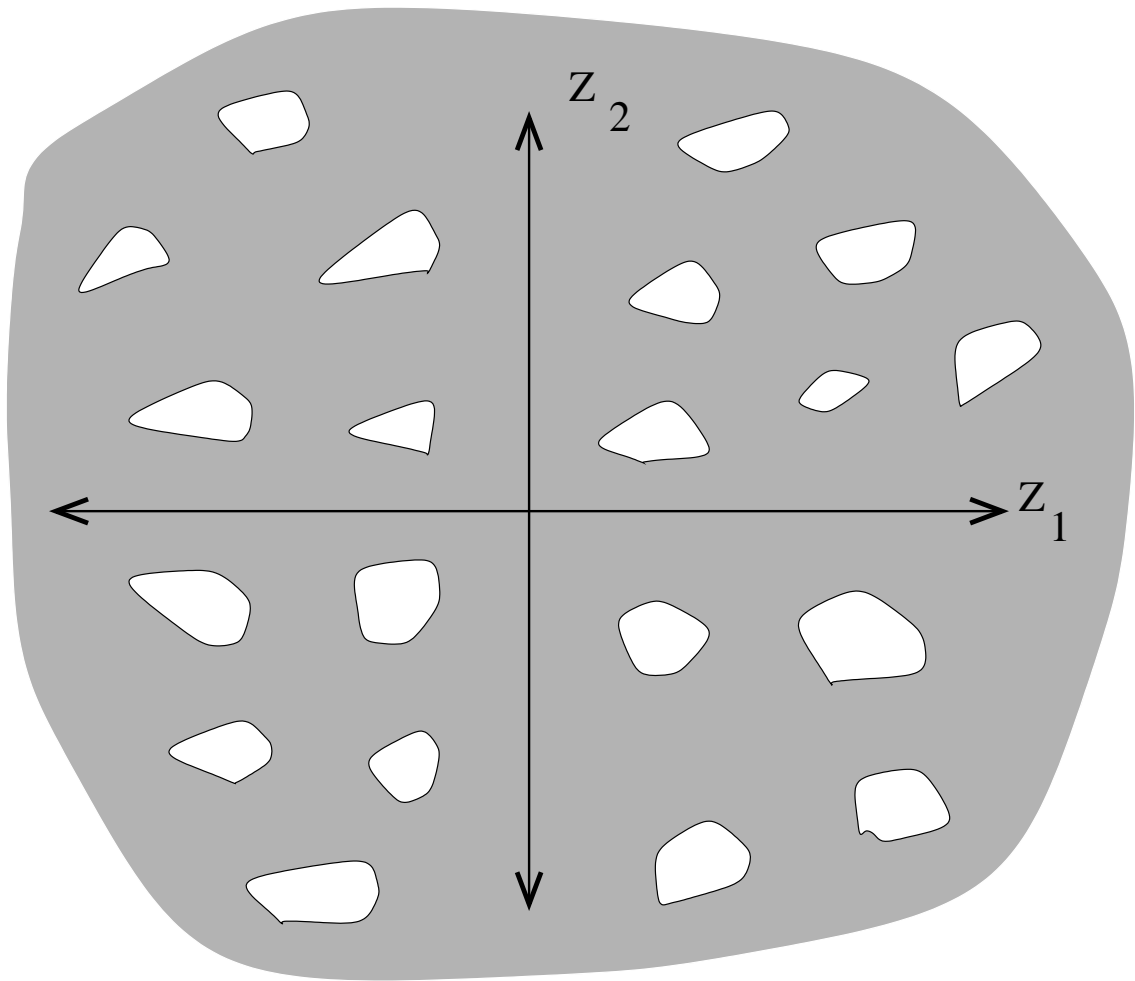


Figure 2-5: Plot of possible decoding function  $D(Z)$  where  $Z = (z_1, z_2)$ . The shaded area corresponds to  $D(Z) = \emptyset$  and the unshaded area corresponds to an area where  $D(Z)$  produces an estimate of  $X$ .

Making the ratio of the authentic region to the forbidden region small and ensuring that the enemy does not have a clear idea of where the authentic regions are will make the system secure against forgery. This is because the enemy will not be able to produce a forgery,  $Z_f$ , which is accepted as valid by the decoder. Figure 2-5 illustrates this idea since a large portion of the plane is a forbidden region where  $D(Z) = \emptyset$ . Sprinkled through the forbidden region are authentic cells where  $D(Z)$  produces an allegedly authentic estimate of  $X$ . This implies that if an enemy were to

try to produce a forgery,  $Z_f$ , without knowing  $D(\cdot)$ , then with high probability he would choose  $Z_f$  in the forbidden region so that  $D(Z_f) = \emptyset$ . The way to achieve security against forgery is to choose the forbidden region to fill most of the space.

### 2.1.2 Geometric Insights About The Encoder

The encoding process corresponds to processing  $X$  by moving it from the forbidden regions to inside an authentic cell as shown in Figure 2-6. To gain absolute protection against forgery the decoder could be designed with  $D(Z) = \emptyset$  everywhere. This would be useless because the legitimate transmitter could never send anything that the receiver would accept. Consequently, there must be some regions of  $D(Z)$  such that  $D(Z) \neq \emptyset$ . Imagine that  $D(Z)$  is fixed and the transmitter must transmit  $X = (x_1, x_2)$ . The transmitter must process  $X$  to get  $Y = Q(X)$  such that  $D(Y) \neq \emptyset$ . So  $Y$  is obtained by moving  $X$  inside an authentic cell. When  $Y$  is perturbed by the channel, as long as the perturbation is small enough that the received signal is still in the authentic cell, the decoder can use the received signal to produce an authentic estimate.

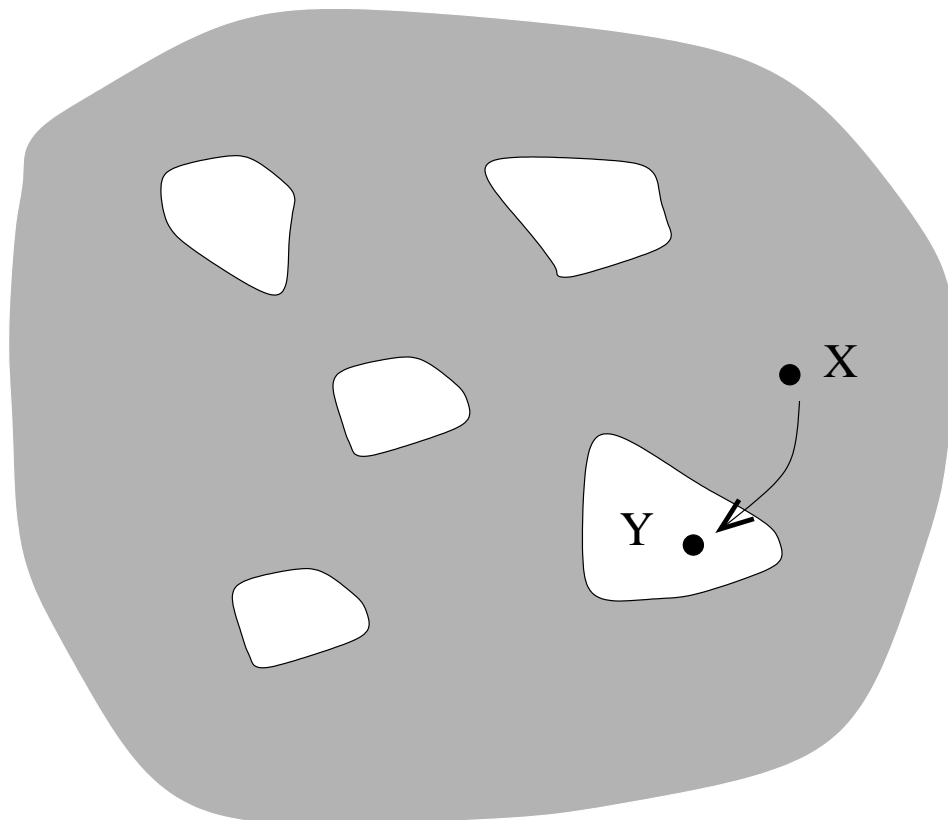


Figure 2-6: To send, the transmitter must move  $X$  from the forbidden region to the inside of an authentic cell. The transmitted signal will be  $Y = Q(X)$ .

This argument shows that  $Y$  must be in an authentic cell, but there are many ways to accomplish this. For example,  $Q(X)$  could map  $X$  to the centroid of the nearest authentic cell. In this case,

$Q(X)$  would effectively be a quantizer. Another possibility is that  $Q(X)$  moves  $X$  only part way to the nearest centroid, but not necessarily all the way.

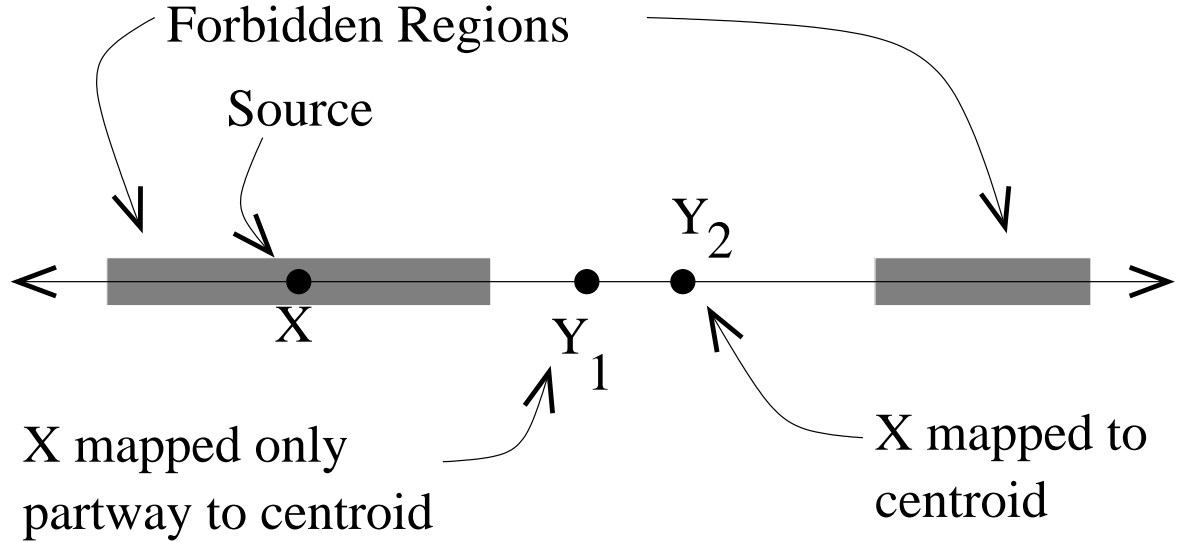


Figure 2-7: Two possibilities for  $Q(X)$ .  $Y_1 = Q_1(X)$  maps  $X$  only part way to the nearest authentic cell centroid while  $Y_2 = Q_2(X)$  maps  $X$  directly to the nearest centroid.

The quantizer  $Q_1(\cdot)$  shown in Figure 2-7 moves the source  $X$  only part way to the nearest centroid while  $Q_2(\cdot)$  moves  $X$  directly to the nearest centroid. One possible way  $Q_1(X)$  could be computed is to first compute the difference vector  $T = Q_2(X) - X$  which would move  $X$  directly to the nearest centroid. Then  $T$  could be scaled by some  $\alpha < 1$  and  $Y_1$  could be computed as  $Q_1(X) = X + \alpha T$ . This would correspond to moving  $X$  towards the centroid, but also trying to mitigate the resulting quantization distortion. This corresponds to the post-quantization processing step discussed by Chen and Wornell in analyzing digital watermarking [13].

One reason to choose different types of encoding functions  $Q(X)$  is to control the processing distortion. Since processing can be thought of as moving the original source  $X$  towards an authentic cell the amount of distortion will be related to the density of authentic cells as well as how far  $X$  is moved towards the nearest cell. However, the construction of the encoding function  $Q(X)$  must also take into account the noisy channel. Imagine that a point,  $Y$ , which is inside an authentic cell is transmitted. If the channel noise pushes  $Y$  outside of the authentic cell and into the forbidden region in producing  $Z$ , then  $D(Z) = \emptyset$ . This will correspond to a decoding failure.

Ideally, an analog authentication system would be robust to channel noise by having a very low probability of failing due to channel noise. As shown by Figure 2-7,  $Y_2$  has twice the noise margin  $Y_1$ . So  $Q_1(\cdot)$  requires less processing distortion and achieves a smaller noise margin than  $Q_2(\cdot)$ . To make a fair comparison we should compare the processing distortions,  $D_1$  and  $D_2$ , when  $Q_1(\cdot)$  and  $Q_2(\cdot)$  are chosen to have the same noise immunity. Thus one issue in constructing the encoding function,  $Q(\cdot)$  involves balancing noise immunity to processing distortion.

Another important consideration in designing the encoding function,  $Q(\cdot)$ , is the effect that it will have on the decoding function,  $D(\cdot)$ . For example, consider a quantizer that maps  $X$  to the centroid of the nearest authentic cell.<sup>2</sup> This implies that the receiver could estimate  $Y$  from  $Z$  by trying to find the centroid most likely to result in  $Z$  and declaring a decoding failure if this can not be done with reasonable confidence.

This considerably simplifies the construction and analysis of  $D(\cdot)$  and  $Q(\cdot)$ . Estimating  $X$  can be thought of as first estimating  $Y$  from  $Z$  then estimating  $X$  from  $Y$ . The decoder can be thought of as combining  $\hat{X} = D_1(Y)$  and  $\hat{Y} = D_2(Z)$  to obtain  $\hat{X} = D_1(D_2(Z))$ . If  $Y$  can be estimated perfectly, then this two-stage decoder is as good as a single stage decoder because of the Markov condition  $X \leftrightarrow Y \leftrightarrow Z$ . Constructing the decoder  $D_2(Z)$  corresponds to a multiple hypothesis testing problem. In principle, constructing  $D_2(\cdot)$  is easy since it involves a number of likelihood ratio tests. There will be a finite (or at least a countable) number of possible outputs for  $D_2(\cdot)$ . Consequently  $D_1(\cdot)$  will only have to be designed considering a finite (or countable) number of inputs and outputs. This simplifies decoder design both practically and theoretically. In addition the decoder has the possibility of estimating  $Y$  exactly. This feature has important implications for security.

Consider the example with quantizer  $Q_2(\cdot)$  in Figure 2-7. The transmitted signal is  $Y = Q_2(X)$ . We model the received signal is a noisy version of  $Y$  such that  $Z = Y + N$ . Maximum likelihood decoding corresponds to requantizing  $Z$  to the nearest centroid to estimate  $Y$  exactly with a finite probability of success. If the signal is encoded with  $Q_1(\cdot)$ , we can no longer decode  $Y$  exactly with a finite probability of success. Instead we can only estimate an approximate version of  $Y$ .

This has an important effect on the security of the system. Imagine that Alice uses the quantizer,  $Q_2(\cdot)$ . She sends  $Y = Q_2(X)$  to Bob. Charlie modifies the transmitted signal so that Bob actually receives  $Z_f$ . If Charlie makes a large modification then  $Z_f$  will probably be in a forbidden region and Bob will detect this forgery. Consider the case where Charlie makes a small modification on the same order as the noise. If the system is robust to noise then Bob would decode to  $D(Z_f) = Y$  with high probability. Thus even though Charlie has tried to pass off a forgery, Bob will either detect the forgery and reject  $Z_f$  or he will be able to decode exactly to  $Y$  so that Charlie's forgery does not affect Bob's result. This is a useful dichotomy from Bob's point of view. Either he will decode so that he knows *exactly* what Alice sent, or he will declare a decoding error. If Bob correctly accepts the received signal as authentic, Charlie's modifications will have been completely eliminated.

This property does not hold for the alternative quantizer,  $Q_1(\cdot)$ . Imagine Alice transmits  $Y = Q_1(X)$ . Again, Charlie modifies the transmitted signal so that Bob actually receives  $Z_f$ . If Charlie makes a large modification then  $Z_f$  will probably be in a forbidden region and Bob will detect this forgery. Consider the case where Charlie makes a small modification on the same order as the

---

<sup>2</sup>The crucial point is not that  $Q(X)$  maps  $X$  to the centroid of the nearest authentic cell. It is that each authentic cell must have exactly one reconstruction point and  $Q(X)$  must map  $X$  to the nearest reconstruction point. This is as opposed to a post-quantization processing system where there are a continuum of points which  $Q(X)$  maps to.

noise. If Bob tries to make a continuous estimate,  $\hat{X} = D(Z)$ , then he will be influenced by Charlie's modification. Thus even if Bob's decoder can guarantee him that  $\hat{X}$  is within some distance,  $d_0$ , from the original source  $X$  or from the transmitted signal  $Y$ , he is still affected by Charlie's modification. Bob might have a bound on the extent of Charlie's modification, but he can not eliminate Charlie's tampering. Bob has a weaker guarantee of security in this case than in the previous case. In the first case, all Charlie could do was cause Bob to declare a decoding error. However, if Bob did not declare a decoding error, then Charlie's efforts would have no effect on  $D(Z)$ . In the second case, Bob can never be sure whether his decoded result  $D(Z)$  has been affected by tampering.

### Relating Geometric View To The Thesis

We have presented a general framework for analog authentication. Specifically we view any analog authentication scheme as being completely specified by the encoder  $Q(\cdot)$  and the decoder  $D(\cdot)$ . The structure of these two determines the processing distortion, noise robustness, security and other parameters of interest. Once the goals of an analog authentication scheme are clearly specified the question

How should we design a scheme to achieve these goals?

becomes

How do we choose  $Q(\cdot)$  and  $D(\cdot)$  to optimize our figures of merit?

In the next chapter we formulate the goals of analog authentication, and analyze what encoding and decoding functions are appropriate for these goals. How does this view fit into the geometric framework we developed? There is a well known analogy between error correction/quantization and sphere packing/sphere covering (see [12], pages 243, 357). The decoder is represented by spheres in  $n$ -dimensional space. If  $Z$  is inside a sphere then  $D(Z)$  maps  $Z$  to the center of that sphere. If  $Z$  is not inside a sphere then it is in the forbidden region and  $D(Z) = \emptyset$ . The encoder maps  $X$  to the center of the nearest sphere. This process is shown in Figure 2-8.

Consider what occurs when the channel noise pushes  $Y$  away from the center but still stays inside the sphere so that the received signal is  $Z_1$ . The decoder maps  $Z_1$  to the center of the nearest sphere resulting in  $D(Z_1) = Y$  for correct decoding. If the channel noise is exceptionally large resulting in a received signal of  $Z_2$ , there will be a decoding failure. Similarly, if the enemy tries to present a forgery,  $Z_f$ , it will be in the forbidden region with high probability. Thus  $D(Z_f) = \emptyset$  and the receiver will detect the forgery. The source coding aspect corresponds to packing enough spheres in the space so that the processing distortion is small. The channel coding aspect corresponds to keeping the sphere radii large so that noise will not cause decoding failure. The cryptography aspect corresponds to making a large region of the space a forbidden region so that if the enemy tries to create a forgery it will be in a forbidden region.

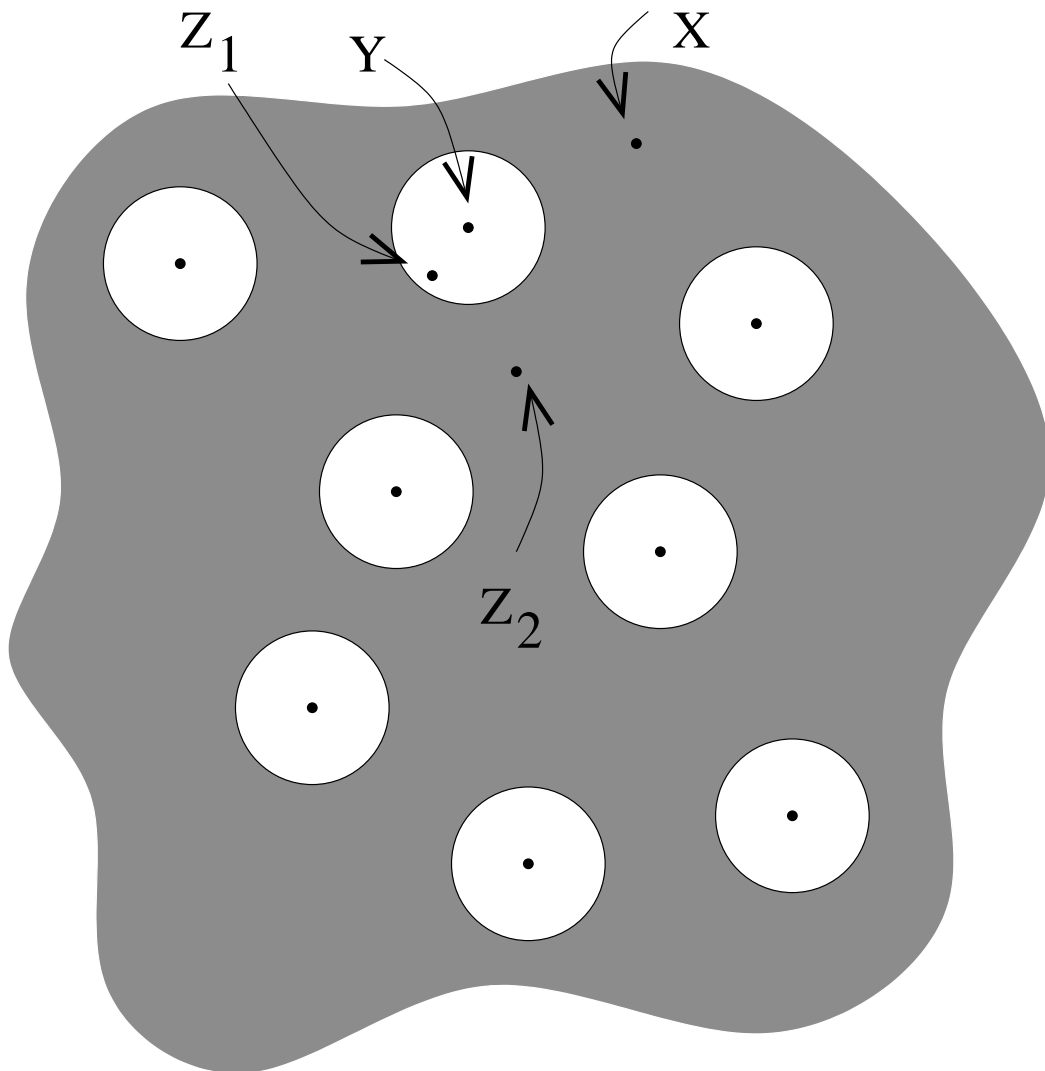


Figure 2-8: The shaded area is the forbidden region where  $D(Z) = \emptyset$ . The original source,  $X$ , is mapped to the center of the nearest sphere to get  $Y = Q(X)$ . Channel noise causes the receiver to get  $Z_1$  which is correctly decoded to  $Y$ . A large amount of channel noise or an attempted forgery would correspond to  $Z_2$  which would be rejected since it is in the forbidden region.

## Chapter 3

# Analog Authentication: The Formal Problem Statement

In the analog authentication problem, the transmitter must send a signal over a noisy and insecure channel. The transmitter's goal is to encode the signal so that a receiver with an appropriate decoder can authenticate the signal. To allow receivers without decoders to also use the signal, the encoding distortion is constrained. Figure 3-1 shows a diagram of this model. In this section we define and analyze the analog authentication problem. Although, there are many ways to formalize this problem, we define the problem as simply as possible in order to portray the fundamental ideas. In Chapter 6 we describe some generalizations.

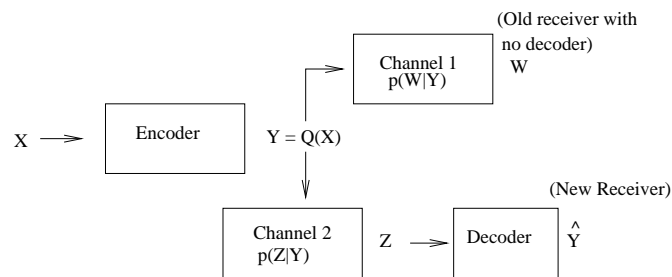


Figure 3-1: A diagram of the analog authentication problem.

The components of the analog authentication problem are as follows:

**Transmitter:** The transmitter encodes a signal to allow authentication by the appropriate receiver.

For example, a government agency that issues passports would use this device to produce passports that can be authenticated.

**Enemy:** The enemy attempts to subvert the analog authentication scheme. The enemy's goal is to produce a forged signal which appears authentic. The enemy might have access to some

signals from the transmitter or he might try to produce a forgery without first seeing a valid signal. For example, the enemy could be a spy who tries to produce a fake passport.

**Sophisticated Receiver:** The sophisticated receiver knows the details of the authentication scheme and therefore uses a decoder to authenticate received signals. For example, a customs agent who wants to check someone's passport would use a sophisticated receiver.

**Simple Receiver:** The simple receiver accepts the signal at face value and does not do any decoding. For example, a customs agent who lacks the decoding equipment to authenticate a passport would simply look at the picture on the passport without doing any decoding.

**Robustness:** The analog authentication scheme must be robust to noise and other small perturbations. For example, images might be printed and scanned, compressed, or smudged. Other types of analog signals might suffer similar perturbations. We classify these disturbances as noise and require that the analog authentication scheme be effective when they are present at certain levels.

**Security:** An important goal of analog authentication is to prevent forgery and tampering. In the passport example, an analog authentication scheme should make it difficult or impossible for unauthorized agents to produce false passports which are deemed authentic.

**Fidelity:** Analog authentication schemes must not substantially degrade the original signal. The processing distortion introduced must be perceptually indistinguishable. For example, if a government agency uses an analog authentication scheme to produce secure passports, the processed photograph on the passport should appear similar to the unprocessed photograph. The motivation for the fidelity constraint is that the signal should still be useful to those without a sophisticated receiver. In the passport example, a customs agent without the required decoding equipment should still be able to use view the picture on a passport.

The transmitter's goal is to provide the receivers with  $X^n$ . We define the source signal as a sequence of  $n$  values from the finite set  $\mathcal{X}$ . The transmitter processes the source signal and transmits  $Y^n = Q(X^n)$ .  $Y^n$  passes through possibly different channels to the sophisticated receiver and the simple receiver. The sophisticated receiver then tries to authenticate the message to produce  $\hat{Y} = A(Z)$ . If the receiver can not decode the received signal to an authentic message, it returns a special symbol,  $\emptyset$ , indicating a failure to decode. Figure 3-2 shows the path between the transmitter and the sophisticated receiver.

The message might be modified en route to the receiver. The channel could be noisy and corrupt the input. For example, the channel might be modeled as additive noise. Alternatively, the channel might be modeled as malicious modification by an enemy. The enemy might replace the input to the channel with a forgery hoping the receiver accepts the forgery as the transmitted message.

Thus the transmitter processes the signal to aid the receiver in determining whether the received message is authentic. The sophisticated receiver's goal is combined error correction and error detection. In decoding, the receiver should correct errors or perturbations due to the noise. In addition the receiver must recognize when uncorrectable errors occur and indicate this as a decoding failure. If the transmitted signal is slightly perturbed by noise, then the receiver should correct for the noise in the decoding process. On the other hand, if an enemy replaces the transmitted signal with a forgery, the receiver will not be able to correct this kind of error. Instead the receiver should indicate a decoding failure.

### 3.0.3 Probability Of Undetected Error

We define an undetected error as the event

$$\mathcal{E}_u = \{A(Z^n) \neq \emptyset\} \cap \{A(Z^n) \neq Y^n\} \quad (3.1)$$

and the probability of undetected error as

$$p_u = \Pr(\mathcal{E}_u) \quad (3.2)$$

An undetected error occurs when the receiver decodes to something other than the what transmitter sent without realizing that  $A(Z^n) \neq Y^n$ . This could happen due to noise or due to a forgery by the enemy. In either case, an undetected error is bad because the receiver will think the transmitter sent  $A(Z^n)$  when actually  $Y^n$  was sent.

For example, Charlie takes a valid passport for John and modifies the photo, of John which is  $Y^n = J$ , to look like himself,  $Z^n = C$ . If the customs agent uses the decoder to get  $A(Z^n) = \emptyset$ , then he realizes that he can not trust the passport and arrests Charlie or investigates further. If the customs agent uses the decoder and gets  $A(Z^n) = J$ , he can view the decoded passport picture,  $J$ , and realize that the passport has a photo of John instead of Charlie. In this case the customs agent does not allow Charlie to enter the country since Charlie lacks a valid passport.

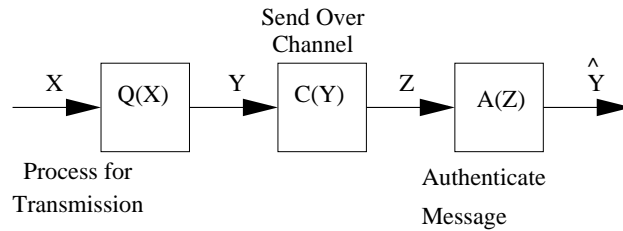


Figure 3-2: The transmitter processes  $X$  to get  $Y = Q(X)$  and sends it over the channel. The sophisticated receiver gets the output of the channel  $Z$  and computes  $\hat{Y} = A(Z)$  to authenticate the message.

Instead if the analog authentication scheme decodes to an incorrect result, then the scheme has failed due to an undetected error. In this case, the customs agent will think the passport is valid. The seemingly valid passport has a photo of Charlie, so the customs agent will let him into the country. A good analog authentication schemes should have a small probability undetected error.

### **Why The Goal Is Estimating $Y$ Not Detecting Tampering**

The receiver's goal is to either recover  $Y^n$  exactly from  $Z^n$ , or detecting a decoding failure. One might think that the receiver's goal should be to detect if the enemy has modified  $Y^n$ . This goal is impossible. If the probability distribution of the enemy's modifications have the same distribution as the channel noise, there is no way the receiver can distinguish between the cause of the perturbations. This is a fundamental point and worth stressing. In the model we have proposed, no scheme exists which could always determine whether the enemy has tampered with the signal or the signal has only been perturbed by noise.

Note that if the receiver correctly decodes  $Z^n$  to the transmitted signal  $Y^n$  when the enemy has tampered with the signal, this can be used to take action against the enemy. For example, the distance between the received signal and the transmitted signal can be compared to a threshold. If  $d(Z^n, A(Z^n)) > T$ , the bearer of the signal can be punished for attempted forgery.

### **Why The Goal Is Estimating $Y$ Not Estimating $X$**

Since  $X$  is the original signal, it is ultimately the quantity the receiver is interested in. Why not formulate the problem as estimating  $X$  instead of estimating  $Y$ ? We can always view estimating  $X$  as first estimating  $Y$  from  $Z$  and then estimating  $X$  from  $Y$ . As a result of the Markov condition  $X \leftrightarrow Y \leftrightarrow Z$ , if  $Y$  can be estimated exactly with small probability of error, this two-stage estimation is optimal. In addition, if  $E[d(X, Y)]$  is small then simply using the estimate  $\hat{X} = \hat{Y}$  will yield good results if we can decode  $Y$  exactly with small probability of error. In addition, this formulation balances the processing distortion for the sophisticated receiver and simple receiver in a reasonably way summarized by one number:  $D = E[d(X, Y)]$ . Finally, this goal is easier to analyze than a goal which involved estimating  $X$  directly.

What can be gained by formulating the problem differently? Once the encoder is fixed, formulating the goals differently does not change the decoder structure if  $Y$  can be estimated with low probability of error. This is because of the Markov condition mentioned previously. However, formulating the problem differently would lead to designing the encoder differently. Since the ultimate goal of the receiver is to obtain a reliable and authentic estimate of  $X$ , it would be reasonable to make the goal estimating  $X$  directly. This formulation might lead to lower expected distortion between  $X$  and  $\hat{X}$ . However, the alternative formulation is more difficult to analyze and has implications about security. Consequently we postpone it for future study.

The goal of analog authentication could also be defined as reliably and authentically estimating  $X$  to within some threshold,  $T$ . This would be a valid goal and some researchers seem to implicitly have this goal in mind although it has not been formally proposed. We discuss this formulation as part of the future work in Section 6.

If the receiver can either estimate  $Y$  exactly or detect that an exact estimate is not possible, then the enemy can influence the receiver in only one way. The enemy can always distort, damage, or replace  $Y$  enough so that the receiver must declare a decoding failure. However, if the receiver successfully decodes to get  $Y = A(Z) \neq \emptyset$ , then the receiver is unaffected by any modification the enemy might have made. We define this property as security via dichotomy:

**Security Via Dichotomy (SVD):** An authentication system satisfies the Security Via Dichotomy principle if tampering can only cause the receiver to reject a transmission.

A system where the receiver could always either find an estimate  $\hat{X} = A(Z)$  such that  $d(X, Z) < T$  or declare a decoding failure does not necessarily have the Security Via Dichotomy property. It is still secure in the sense that an enemy would not be able to cause the receiver to accept a signal that is very different from the transmitted signal. However, the enemy could still commit a forgery in the sense that he can make small modifications such that  $d(X, Z) < T$  but still have the modified signal accepted by the receiver.

Systems that possess the SVD property are secure in a stronger sense than systems without this property. Although systems without SVD can be secure in a weaker way, insisting on the strongest possible notions of security is a good policy. By choosing the goal as estimating  $Y$  exactly, we insure that our schemes will have the SVD property and possess a very strong notion of security.

The problem could also be defined such that the receiver attempts to estimate  $X$  reliably and authentically where  $d(\hat{X}, X) < T$  such that the system has the SVD property. Imagine that the transmitter wants to send  $X$  by processing  $X$  to get  $U$  and  $Y$ . The transmitted signal is  $Y$  as usual. The other signal,  $U$ , satisfies the condition  $d(X, U) < T$ . The decoding function is such that either  $A(Z) = U$  or  $A(Z) = \emptyset$ . The signal  $U$  corresponds to a distorted version of the original signal  $X$ .  $U$  could be a better or worse representation of  $X$  than  $Y$ , (e.g. we could have  $T > D$  or  $T < D$ ).

If the receiver recovers  $U$ , then this becomes the estimate of  $X$ . This scheme has the SVD property even though the receiver does not exactly recover the transmitted signal  $Y$ , but instead recovers  $U$ . The system has the SVD property because  $U$  was calculated by the transmitter. As long as the receiver decodes to either  $U$  or  $\emptyset$ , the enemy can not affect the signals the receiver trusts as authentic.

This formulation would have a strong notion of security. It would generalize our formulation in the sense that the processing distortion suffered by the simple receiver could be weighted differently than the processing suffered by the sophisticated receiver. We conjecture that this problem would

be characterized by a distortion region. The distortion region would be the set of pairs  $(D, T)$  which are achievable, where  $D$  is the processing distortion and  $T$  is the estimation error.

In summary, other formulations of the goals of analog authentication are possible and plausible. Specifically, choosing the goal to be estimating  $X$  instead of estimating  $Y$  would be an interesting scenario to study. However, alternative formulations are more difficult to analyze and are potentially less secure. Consequently we have chosen to focus our analysis on the goal of estimating  $Y$ .

### 3.0.4 Probability Of False Alarm Error

A well designed scheme should never fail to decode unless the enemy interferes. We define a false alarm as the event

$$\mathcal{E}_f = \{A(Z^n) = \emptyset\} | \{ \text{no tampering} \} \quad (3.3)$$

and the probability of false alarm error as

$$p_f = \Pr(\mathcal{E}_f) \quad (3.4)$$

A false alarm error occurs when the decoder fails even though no enemy has tampered with the transmission. In the passport example, this corresponds to John having a valid passport and presenting it to the customs agent. The customs agent computes  $A(Z^n) = \emptyset$  because of smudges or other noise and refuses to accept the passport. This is an undesirable event as it could result in John Doe being delayed or arrested. In practice false alarms will substantially degrade the reliability of an analog authentication scheme. A good analog authentication scheme should have a low probability of false alarm.

### 3.0.5 Embedding Distortion

We define the embedding distortion as the distortion between the original source,  $X$ , and the transmitted signal,  $Y$ . Simple receivers which lack decoders view any embedding distortion  $D$  as a hindrance. To make the transmitted signal useful to simple receivers, the embedding distortion should be acceptably small. We define an excess distortion error as the event

$$\mathcal{E}_D = \{d(X^n, Y^n) > D\} \quad (3.5)$$

and the probability of excess distortion error as

$$p_D = \Pr(\mathcal{E}_D) \quad (3.6)$$

A good analog authentication scheme should obey the distortion constraint and have a small probability of excess distortion error.

### 3.1 Formal Problem Statement

We have qualitatively discussed the pieces of the analog authentication problem. Next we precisely define the problem and analyze asymptotic results for large  $n$ . We define an analog authentication problem as consisting of the follow components:

- **Message Space:** The message is a sequence of  $n$  values from a finite set  $\mathcal{X}^n$ . Where  $\mathcal{X}^n$  is the  $n$ -fold Cartesian product of the finite set  $\mathcal{X}$ .
- **Encoding Function:** The encoding function,  $Y^n = Q(X^n)$  in Figure 3-2, is a mapping  $Q: \mathcal{X}^n \rightarrow \mathcal{X}^n$ . Furthermore we require that the output of the encoding function is reasonably close to the input:

$$\frac{1}{n} \sum_{i=1}^n d(X_i, Y_i) < D \quad (3.7)$$

where  $D$  is called the embedding distortion and  $d(\cdot, \cdot)$  is an appropriate metric (e.g.  $d(X_i, Y_i) = (X_i - Y_i)^2$  for a squared distortion metric).

- **Channel:** The noisy channel follows a probability law  $p(z|y)$  in the absence of tampering. The output space of the channel is a finite set denoted by  $\mathcal{Z}$ . When the channel is used repeatedly it is memoryless:  $p(z^n|y^n) = \prod p(z_i|y_i)$ .
- **The Enemy:** The enemy can look at the input of the channel,  $Y^n$ , and arbitrarily choose the output of the channel. Specifically, the enemy is not bound by any distortion constraint.
- **Authentication Function:** The authentication function used by the receiver either outputs an estimate of the transmitted message or outputs  $\emptyset$  to indicate that the received message could not be authenticated.
- **Probability Of Error:** We measure the effectiveness of an analog authentication scheme by the probabilities of error. We define the overall probability of error,  $\lambda$ , as

$$\lambda = \Pr[\mathcal{E}_u \cup \mathcal{E}_f \cup \mathcal{E}_D]$$

where the events  $\mathcal{E}_u, \mathcal{E}_f, \mathcal{E}_D$  are defined according to (3.1), (3.3), and (3.5) respectively.

- **Authentically Achievable Distortion:** We define an embedding distortion,  $D$ , as authentically achievable if a sequence of mappings,  $Q_n$ , and authentication functions,  $A_n$ , can be

found such that the probability of error goes to 0 asymptotically:

$$\lim_{n \rightarrow \infty} \lambda = 0$$

## 3.2 Results

We define the fundamental distortion as

$$D^* = \min_{p(y|x): I(X;Y) - I(Y;Z) \leq 0} E[d(X, Y)] \quad (3.8)$$

If no distribution satisfies the mutual information constraint above, we arbitrarily define  $D^* = \infty$ . We will give the fundamental distortion an operational meaning by proving all embedding distortions  $D > D^*$  are authentically achievable and no embedding distortions  $D < D^*$  are achievable.

The analog authentication problem consists of combined error correction and error detection. We can gain some intuition by considering these two components separately. Consider the analog authentication problem when no enemy is present. In this case, all decoding errors are false alarms. Therefore the probability of error becomes

$$\lambda = \Pr[A(Z^n) \neq Y^n]$$

This scenario corresponds closely to the well known problem of reliable communication over a noisy channel. The transmitter's goal is to encode  $X^n$  so that it can be reliably decoded by the sophisticated receiver. Instead of the usual power constraint, the transmitter is subject to a distortion constraint.

In Appendix A we analyze this communication problem which we call the backward compatibility broadcast channel (BCBC). We prove a converse and a coding theorem for the BCBC. Using the definition of the fundamental distortion in (3.8), we show that reliable communication for the BCBC is possible for encoding distortions  $D > D^*$  and impossible for encoding distortions  $D < D^*$ . The converse for the BCBC implies the converse for analog authentication.

### **Theorem 1** Converse To The Analog Authentication Coding Theorem

*If  $D^*$  is defined according to (3.8), then no embedding distortions  $D < D^*$  are authentically achievable.*

The backward compatibility broadcast channel discussed in A is a special case of the analog authentication problem. According to theorem 5, no embedding distortions  $D < D^*$  are achievable for the BCBC. Therefore, no embedding distortions  $D < D^*$  are achievable for the analog authentication problem. ♠

The proof of achievability must consider the enemy. Two common methods of dealing with enemies are by either assuming the honest parties share a secret unknown to the enemy or assuming the enemy is computationally bounded and a particular problem is computationally intractable. The former achieves information theoretic security, while the latter achieves complexity based security via public key cryptography. We first prove the coding theorem based on the shared secret key assumption and later describe how to extend it using complexity theoretic assumptions.

**Theorem 2** The Analog Authentication Coding Theorem

*If  $D^*$  is defined according to (3.8) and the transmitter and receiver share a secret then all embedding distortions  $D > D^*$  are authentically achievable.*

The proof is similar to the proof of the Backward Compatibility Coding Theorem (theorem 3 in Appendix A). Fix  $\epsilon > 0$ . Let  $p^*(y|x)$  be the distribution which achieves the minimum in the definition of  $D^*$ . By definition,  $D^* < \infty$  implies  $I(X;Y) \leq I(Y;Z)$ . Without loss of generality assume that this holds with strict inequality. If the equation holds with equality,  $p^*(y|x)$  can be changed slightly to allow a very small difference of  $\delta$  between  $I(X;Y)$  and  $I(Y;Z)$ . By continuity of the distortion function, this slight change will not significantly change  $D = E[d(X,Y)]$ . Since  $I(X;Y) < I(Y;Z)$ ,  $\gamma = [I(Y;Z) - I(X;Y)]/3 > 0$ . Choose  $R = I(X;Y) + 2\gamma$ .

The transmitter creates a random codebook,  $\mathcal{C}$ , with  $2^{nR}$  codewords where each codeword is chosen to be an i.i.d. sequence of  $n$  random variables chosen according to the distribution  $p(y)$

$$p(y) = \sum_{x \in \mathcal{X}} p^*(y|x)p(x)$$

We will refer to the sequence for the  $i$ th codeword as  $Y^n(i)$ . The codebook is revealed to transmitter, receiver, and enemy.

Next the transmitter creates the authentic set,  $\mathcal{A}$ , by choosing  $2^{n(R-\gamma)}$  integers randomly and uniformly from the set  $\{1, 2, \dots, 2^{nR}\}$ . The transmitter then secretly shares  $\mathcal{A}$  with the receiver. The authentic set is used as the shared secret key instead of making the entire codebook secret. Clearly separating the parts of the scheme which rely on a secret illustrates the role of security. Later, when we extend the scheme to complexity based assumptions, we retain the same codebook, but modify how the authentic set is designed. In addition, specifying the authentic set requires fewer bits than specifying the codebook. Therefore if shared secrets are to be used then using an authentic set is more efficient than making the entire codebook secret.

**Encoding and Decoding**

To encode a sequence,  $X^n$ , the transmitter finds a codeword, such that  $Y^n(i)$  is jointly typical with  $X^n$  and such that  $i \in \mathcal{A}$ . If more than one such codeword exists the transmitter chooses the

lexicographically least. If no such codewords exist the transmitter declares an error and sets  $i = 1$ . The transmitter sends  $Y^n(i)$ .

The receiver decodes the received sequence  $Z^n$ , by finding a jointly typical codeword  $Y^n(\hat{i})$  in  $\mathcal{C}$ . If no such codeword or more than one such codeword exists the receiver outputs  $\emptyset$  indicating a decoding failure. If exactly one jointly typical codeword,  $Y^n(\hat{i})$ , is found and  $\hat{i} \notin \mathcal{A}$  then the decoder outputs  $\emptyset$ , again indicating a decoding failure. If exactly one jointly typical codeword,  $Y^n(\hat{i})$ , is found and  $\hat{i} \in \mathcal{A}$  then the decoder outputs  $\hat{Y}^n = Y^n(\hat{i})$ .

### Analysis

When the enemy does not tamper with the transmitted signal, decoding errors can only occur due to channel noise. In this case, the probability that a decoding error occurs is upper bounded by  $2^{-n(I(Y;Z)-R)}$  so

$$\lim_{n \rightarrow \infty} \Pr[\hat{Y}^n \neq Y^n] \leq \lim_{n \rightarrow \infty} 2^{-n\gamma} = 0 \quad (3.9)$$

This follows from the noisy channel coding theorem [12], which proves that the probability of decoding error using a random codebook and typical set decoding goes to  $2^{-n(I(Y;Z)-R)}$  asymptotically for large  $n$ . The backward compatibility coding theorem (theorem 3) provides a more detailed proof of this statement which applies to the analog authentication problem in the absence of tampering.

Regardless of tampering, the probability of excess distortion error is upper bounded by  $2^{n(I(X;Y)-R+\gamma)}$  so

$$\lim_{n \rightarrow \infty} p_D \leq \lim_{n \rightarrow \infty} 2^{-n\gamma} = 0 \quad (3.10)$$

This follows from the rate-distortion theorem [12], which proves that the probability of exceeding the distortion constraint when using typical set encoding with a rate  $R'$  random codebook goes to  $2^{n(I(X;Y)-R')}$  asymptotically for large  $n$ . The backward compatibility coding theorem (theorem 3) provides a more detailed proof of this statement which applies to the analog authentication problem. Note that since the transmitter only sends codewords  $Y^n(i)$  with  $i \in \mathcal{A}$ , the rate of the encoder is effectively  $R' = R - \gamma = I(X;Y) + \gamma$ .

If the enemy modifies the transmitted signal, we are only concerned with the probability of undetected error,  $p_u$ . For an undetected error to occur,  $Y^n(\hat{i} \neq i)$  must be the only codeword jointly typical with  $Z^n$  and the codeword must be in the authentic set. Since the enemy does not know  $\mathcal{A}$ , no matter what the enemy chooses for  $Z^n$ , the probability that  $Z^n$  is jointly typical with  $Y^n(\hat{i} \neq i)$  with  $\hat{i} \in \mathcal{A}$  is upper bounded by

$$\frac{\text{number of authentic codewords}}{\text{number of codewords}} = \frac{2^{nR-\gamma}}{2^{nR}} = 2^{-n\gamma}$$

Thus the probability that the enemy can create a successful forgery is upper bounded by  $2^{-n\gamma}$ . This probability is taken over the choice of authentic sets *not* over the actions of the enemy. Therefore no matter what action the enemy takes the probability of successful forgery is bounded by  $2^{-n\gamma}$ .

Combining these results implies that the total probability of error,  $\lambda$ , goes to 0 as  $n \rightarrow \infty$  when  $\lambda$  is averaged over all random codebook and authentic sets. Consequently a deterministic codebook with these properties must exist. ♠

### 3.3 Analog Authentication Based On Complexity Assumptions

In the proof of the coding theorem, we assumed that both the transmitter and receiver secretly shared a description of the authentic set,  $\mathcal{A}$ . If we model the enemy as computationally bounded, there are other ways to design the authentic set. By assuming that certain problems are computationally intractable we can construct analog authentication schemes based on ideas from public key cryptography instead of shared secret keys. For example, the transmitter could use a random seed and a pseudo random number generator to create the authentic set. The transmitter could then encrypt the seed with the receiver's public key and reveal both the encrypted seed and the codebook to the receiver and the enemy.

There are many ways to design the authentic set based on complexity theoretic assumptions. The previous example based on public key encryption requires the assumption that public key encryption is possible. We describe an implementation based on digital signature schemes because digital signatures can be constructed from a variety of complexity assumptions including simply the assumption that one-way functions exist [14], which is the weakest possible cryptography assumption.

#### 3.3.1 Brief Summary Of Digital Signature Schemes

Many digital signature schemes have been proposed and analyzed in the cryptography literature. The main idea is to assume that some problem is computationally intractable and use this assumption to design secure schemes. The honest parties use efficient algorithms to achieve their goals, and the enemy must solve an intractable problem to achieve his goals. An important component of this type of analysis is the security parameter,  $k$ . Usually the security parameter relates to the size of the various keys involved or the amount of overhead necessary. It quantifies the tradeoff between resources and security. Efficient algorithms are those which run in time polynomial in  $k$ .

A digital signature scheme with security parameter  $k$ , is a triple of efficient algorithms,  $(\mathcal{S}, \mathcal{V}, \mathcal{G})$ , consisting of a signing algorithm, a verifying algorithm, and a key generation algorithm with the following properties:

**Key Generation:** On input  $1^k$  (by  $1^k$  we mean  $k$  ones), The key generation algorithm returns a pair of  $k$ -bit keys:  $(p_k, s_k) \leftarrow \mathcal{G}(1^k)$ .

**Signing Algorithm:** The signing algorithm takes a message and the secret key and produces a tag:  $\sigma \leftarrow \mathcal{S}(m, s_k)$ .

**Verifying Algorithm:** The verifying algorithm takes the message, tag, and public key and returns either 0 or 1. If the verifying function is given a tag which was generated using the secret key corresponding to the message  $m$  and the public key  $p_k$  it should always return 1. Thus for all  $n$ -bit binary strings,  $m$ , the sequence of operations

$$(p_k, s_k) \leftarrow \mathcal{G}(1^k); \sigma \leftarrow \mathcal{S}(m, s_k); b \leftarrow \mathcal{V}(\sigma, m, p_k)$$

should result in  $b = 1$  indicating a valid signature.

There are many definitions of security. One of the weakest notions is security against a no message attack. In this attack, the enemy is given the public key and attempts to produce a forgery consisting of a message-tag pair which will be declared valid by the verifying function. A scheme is secure against the no message attack if the probability of such an attack succeeding can be made negligible. There are progressively stronger attacks such as the  $N$  message attack where the enemy is given access to  $N$  messages and  $N$  corresponding tags before attempting to produce a forgery.

### 3.3.2 Using Digital Signature Schemes For Analog Authentication

Our proof of the Analog Authentication Coding Theorem required that the transmitter and receiver secretly share the set  $\mathcal{A}$ . This assumption provides information theoretic security. If the enemy is computationally bounded, then we can design schemes based on digital signatures. Instead of the transmitter and receiver sharing the entire set  $\mathcal{A}$ , the transmitter needs to be able to choose a codeword  $i$  such that  $i \in \mathcal{A}$ , the receiver needs to be able to check if  $i \in \mathcal{A}$ , and the enemy should be unable to produce a new<sup>1</sup>  $i' \in \mathcal{A}$  with non-negligible probability.

Consider a digital signature scheme  $(\mathcal{S}, \mathcal{V}, \mathcal{G})$  which produces  $k$ -bit keys and  $t$ -bit tags for any message  $m \in \{0, 1\}^{n(R-\gamma)}$  with  $t \leq n\gamma$ . In practice, requiring  $t \leq n\gamma$  is not a severe constraint since digital signature schemes with 1000 bit tags are considered strong enough for commercial use and schemes with 2000 bit tags are considered strong enough for military use.

We would like to use the digital signature scheme to sign a codeword. We denote the first  $n(R-\gamma)$  bits of an integer  $i$  as  $f(i)$  and the last  $n\gamma$  bits of an integer  $i$  as  $g(i)$ . If we tried to sign the index

---

<sup>1</sup>If the enemy has access to a number of valid transmissions,  $Y^n(i_1), Y^n(i_2)$ , etc., he can always replace the current transmission with previously seen valid transmissions. This type of forgery is referred to as the replay attack. By adding time stamp or other message distinguishing information, replay attacks can be made ineffective, hence we focus on showing that the enemy can not find a *new* authentic codeword.

of each codeword, there would be no way to send the tag. Therefore we interpret the first bits of a codeword as the message and the second bits as a tag. Since the codebook has  $2^{nR}$  codewords, there are  $2^{n(R-\gamma)}$  different messages each listed with all possible tags. Each message must have at least 1 valid tag so there are at least  $2^{n(R-\gamma)}$  authentic codewords such that  $\mathcal{V}(g(i), f(i), p_k) = 1$ . Thus the authentic set consists of all codewords such that  $\mathcal{V}(g(i), f(i), p_k) = 1$

Thus we use the digital signature scheme to sign messages with  $n(R - \gamma)$  bits producing tags with  $n\gamma$  bits. First we choose a random codebook,  $\mathcal{C}$  with  $2^{nR}$  codewords. The transmitter calls the key generation algorithm,  $\mathcal{G}(1^k)$  to get a public key,  $p_k$ , and a private key,  $s_k$ . All codewords such that the first part of the codeword is a valid tag and the second part of the codeword forms a valid message for that tag for  $p_k$  are considered to be in  $\mathcal{A}$ . Formally

$$\mathcal{A} = \{i \in \{0, 1\}^{nR} : \mathcal{V}(g(i), f(i), p_k) = 1\} \quad (3.11)$$

A detailed explanation which also provides insight is to view the codebook construction as a two step process. Instead of creating a random codebook with  $2^{nR}$  elements, the transmitter instead chooses a random codebook,  $\mathcal{C}_1$  with  $2^{n(R-\gamma)}$  elements. Then the transmitter chooses a blank codebook,  $\mathcal{C}$ , with space for  $2^{nR}$  codewords. The blank codebook has not yet been populated with codewords. We use  $c_1(i)$  to refer to the codeword at index  $i$  in  $\mathcal{C}_1$  and we use  $c(i)$  to refer to the codeword at index  $i$  in  $\mathcal{C}$ .

Since  $\mathcal{C}_1$  has  $2^{n(R-\gamma)}$  codewords, the encoding rate is between  $I(X; Y)$  and  $I(Y; Z)$  so the embedding distortion will be acceptably small and reliable transmission will be possible provided no tampering occurs. The transmitter populates the codewords in the blank codebook as follows. For each codeword  $c_1(i)$  in  $\mathcal{C}_1$ , the transmitter computes its index,  $j$ , in  $\mathcal{C}$  by concatenating  $i$  and  $\mathcal{S}(i, s_k)$ . The transmitter then sets the codeword at index  $j$  in  $\mathcal{C}$  to be equal to the codeword at index  $i$  in  $\mathcal{C}_1$ . This process is summarized in pseudo-code below:

```

i ← 1
while i <  $2^{n(R-\gamma)}$  do
    j ← i ◊  $\mathcal{S}(i, s_k)$ 
    ; ; comment: ◊ denotes concatenation so 00 ◊ 11 = 0011
     $c(j) = c_1(i)$ 
end while

```

Then the transmitter fills in the remaining entries in  $\mathcal{C}$  with random codewords chosen according to the distribution  $p(y)$ .

After going through these steps, the transmitter has a random codebook  $\mathcal{C}$  with  $2^{nR}$  codewords. Even though  $\mathcal{C}$  was chosen in the special way outlined above, it will have *exactly* the same distribution as a random codebook chosen in the usual way. The transmitter then reveals  $\mathcal{C}$  and the public key,

$p_k$ , to everyone while keeping  $\mathcal{C}_1$  secret. To encode, the transmitter uses the codebook  $\mathcal{C}_1$  not  $\mathcal{C}$ . The receiver decodes by first finding a jointly typical codeword at index  $i$  in  $\mathcal{C}$ . Then the receiver checks if  $\mathcal{V}(g(i), f(i), p_k) = 1$ , where  $g(i)$  is the last  $n\gamma$  bits of  $i$  and  $f(i)$  is the first  $n(R - \gamma)$  bits of  $i$ . If this is the case then he outputs codeword  $i$  as the result. Otherwise he outputs  $\emptyset$  indicating a decoding failure.

The result is that  $\mathcal{C}_1$  now plays the role of the authentic set. The transmitter only transmits a subset of the codewords in  $\mathcal{C}$ . Since the subset he is transmitting has rate  $R - \gamma > I(X; Y)$ , the encoding distortion will be small. Since the receiver uses the full codebook,  $\mathcal{C}$ , which has rate  $R < I(Y; Z)$ , the receiver will be able to reliably decode the transmitted codeword when no tampering occurs. Since the receiver has the public key,  $p_k$ , he can check if the decoded codeword  $i$  is in the authentic set by checking  $\mathcal{V}(g(i), f(i), p_k) = 1$ . Finally, if the enemy will not be able to create a forgery because to do so he would have to find a codeword  $j \neq i$  in  $\mathcal{C}$  such that  $\mathcal{V}(g(j), f(j), p_k) = 1$  which would require cracking the digital signature scheme.

Thus we have described how to extend the analog authentication coding theorem to use public key digital signature schemes instead of a shared secret key. In chapter 5 we discuss practical implementations using error correction codes instead of random coding and maximum likelihood decoding instead of typical set decoding.

## Chapter 4

# Calculating The Fundamental Distortion

In the preceding sections we have introduced the notion of the fundamental distortion,  $D^*$ , and proved that all distortions above  $D^*$  are achievable while all distortions below  $D^*$  are not achievable. In this section, we analyze  $D^*$  for a few source and channel models. Finding the value of  $D^*$  for a specific source and channel serves as a lower bound on the necessary embedding distortion. Knowing the minimum possible embedding distortion provides a way to measure the performance of practical schemes. We can measure performance by calculating how much more distortion a scheme needs than  $D^*$  at fixed probabilities of false alarm and missed detection. Measuring schemes by their gap to the fundamental distortion is analogous to measuring communication schemes using the gap to capacity. Finally, by finding the optimizing distribution  $p(y|x)$  we can discover the optimal distribution for a codebook in a practical scheme.

In some situations finding  $D^*$  is difficult and therefore we seek bounds on  $D^*$ . Furthermore, sometimes we might be more interested in computing the embedding distortion for a random codebook with a non-optimal distribution  $p(y|x)$ . For example, even though a Gaussian distribution is the capacity optimizing distribution for communication over the standard additive white Gaussian noise Channel, other distributions are used in practice. The difference between the embedding distortion for a particular distribution and the fundamental distortion for the optimal distribution can be used to measure the added distortion suffered by an encoder which uses a simpler distribution to practical issues.

## 4.1 Binary Source, Hamming Distortion, BSC Channel

Consider an i.i.d. Bernoulli(1/2) source process,  $\{X_i\}$  with distortion measured by Hamming distortion:  $d(X, Y) = 0$  if and only if  $X = Y$  and 1 otherwise. We model the channel as a Binary Symmetric Channel with crossover probability  $\epsilon < 1/2$ .

We can represent the output of the channel as  $Z = Y \oplus N$  where  $N$  is i.i.d. Bernoulli( $\epsilon$ ) and  $\oplus$  denotes modulo 2 addition. Define  $H_b(p)$  to be the binary entropy function  $H_b(p) = p \log(p) + (1 - p) \log(1 - p)$ . Then

$$\begin{aligned} I(Y; Z) &= H(Z) - H(Z|Y) \\ &= H(Z) - H(N \oplus Y|Y) \\ &= H(Z) - H(N) \\ &= H(Z) - H_b(\epsilon) \\ &\leq 1 - H_b(\epsilon) \end{aligned}$$

This provides an upper bound on  $I(Y; Z)$ .

We can use the rate-distortion theorem to find a lower bound for  $I(X; Y)$ . The rate-distortion function for a Bernoulli(1/2) source with Hamming distortion is

$$R(D) = \begin{cases} 1 - H_b(D), & D < 1/2 \\ 0, & D \geq 1/2 \end{cases}$$

Therefore any scheme which causes embedding distortion  $D$  must have  $I(X; Y) \geq 1 - H_b(D)$ .

Recall that  $D^*$  is the minimum of  $D = E[d(X, Y)]$  subject to the constraint

$$I(X; Y) \leq I(Y; Z)$$

Since  $I(X; Y) \geq R(D)$  and  $I(Y; Z) \leq 1 - H_b(\epsilon)$ ,

$$1 - H_b(D) \leq I(X; Y) < I(Y; Z) \leq 1 - H_b(\epsilon)$$

Manipulating the equation above we obtain

$$H_b(D) \geq H_b(\epsilon)$$

which implies  $D \geq \epsilon$ . Thus we have derived a lower bound for the embedding distortion:

$$D^* \geq \epsilon \tag{4.1}$$

We obtained (4.1) by upper bounding  $I(Y; Z)$  using the capacity of the BSC and lower bounding  $I(X; Y)$  using the rate-distortion function for a Bernoulli(1/2) source. The lower bound on  $I(X; Y)$  is obtained using the rate-distortion optimizing distribution

$$p_R^*(y|x) = \begin{cases} x, & 1 - D \\ 1 - x, & D \end{cases}$$

The upper bound on  $I(Y; Z)$  is obtained with the capacity optimizing distribution

$$p_C^*(y) = \begin{cases} 1/2, & y \in \{0, 1\} \\ 0, & y \notin \{0, 1\} \end{cases}$$

The rate-distortion optimizing distribution maps into the channel capacity optimizing distribution

$$\sum_x p_R^*(y|x)p(x) = p_C^*(y)$$

Therefore both bounds can be satisfied simultaneously. The optimal distribution is  $p^*(y|x) = p_R^*(y|x)$  which yields  $I(X; Y) = I(Y; Z) = 1 - H_b(\epsilon)$  and  $E[d(X, Y)] = \epsilon$ . Since  $p^*(y|x)$  achieves  $D = \epsilon$  and  $D \geq \epsilon$  from (4.1),  $D^* = \epsilon$ .

What if we were simply interested in reliable communication over this channel with no constraint on the embedding distortion? Since  $C = 1 - H_b(D)$  and  $R(D) = 1 - H_b(D)$ , we can easily compute the minimum distortion possible between the transmitter and receiver. If we write the distortion-rate function as  $D(R)$ , then the minimum possible distortion is  $D \geq D(C) = \epsilon$ .

As long as the embedding distortion is allowed to be larger than  $D^*$ , analog authentication is achievable. The sophisticated receiver can always detect tampering or decode to obtain the transmitted signal  $Y$ . For the optimal distribution  $p^*(y|x)$ ,  $E[d(X; Y)] = D = \epsilon$ . Therefore in the analog authentication problem, the sophisticated receiver can receive an estimate of  $X$  which is as good as the best possible estimate he could have expected in the standard communication problem. Consequently, the sophisticated receiver suffers no additional distortion in the analog authentication scenario than he would have in the standard communication problem.

For the receiver without a decoder, however, the expected distortion of the coded signal after going through the channel will be  $D + \epsilon - D\epsilon \geq 2\epsilon - \epsilon^2$ . If the transmitter had transmitted the signal

uncoded instead of doing analog authentication, the receiver without a decoder would see expected distortion  $\epsilon$ . Thus the receiver without a decoder must tolerate almost twice as much distortion than in the uncoded case for analog authentication to be possible.

## 4.2 Uniform Source, Circular Additive Noise, Squared Error Distortion

Consider an i.i.d source,  $X$ , which is uniformly distributed over the integers from 1 to  $N$  with  $N$  even. Consider an additive circular noise channel as shown in Figure 4-1

$$p(Z = i|Y = j) = \begin{cases} 1 - \epsilon, & i = j \\ \epsilon, & i - j \text{ modulo } N = 1 \end{cases}$$

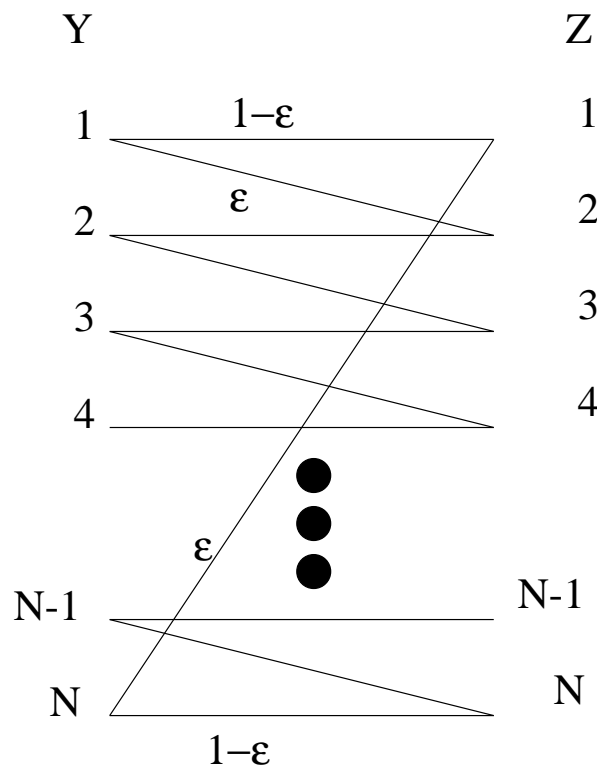


Figure 4-1: Channel probability distribution for additive white circular noise channel.

When  $\epsilon = 1/2$  this is called the noisy typewriter channel. Define the distortion as  $d(X, Y) = (X - Y)^2$ .

We do not have a closed form expression for  $D^*$  in this scenario. We could use algorithms similar to those of Arimoto [15] and Blahut [16] to solve for  $D^*$  numerically. Instead we use a simple distribution for the codebook and compute the achievable distortions for that distribution. Choose

$p(y|x)$  as follows.

$$p(Y = i|X = j) = \begin{cases} 1, & i = j \text{ for } j \text{ odd} \\ 1 - \alpha, & i = j \text{ for } j \text{ even} \\ \alpha, & i = j - 1 \text{ for } j \text{ even} \end{cases} \quad (4.2)$$

as shown in Figure 4-2

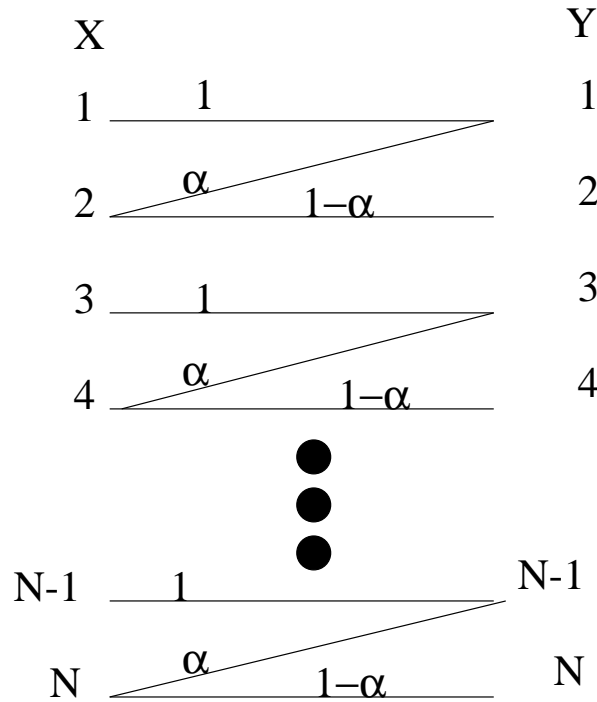


Figure 4-2: Codebook probability distribution for additive circular noise channel.

This encoding scheme corresponds to always setting  $Y = X$  when  $X$  is odd and setting  $Y = X - 1$  with probability  $\alpha$  when  $X$  is even. We calculate the interesting quantities for this scheme below:

$$\begin{aligned} E[d(X, Y)] &= \alpha/2 \\ H(Y) &= -\left(\frac{N}{2}\right) \left(\frac{1+\alpha}{N}\right) \log\left(\frac{1+\alpha}{N}\right) - \left(\frac{N}{2}\right) \left(\frac{1-\alpha}{N}\right) \log\left(\frac{1-\alpha}{N}\right) \\ H(Y|X) &= H_b(\alpha) \\ H(Z|Y) &= H_b(\epsilon) \\ H(Z) &= \log(N/2) + H_b\left(\frac{1}{2}[(1-\epsilon) + \alpha(1-\epsilon) + (1-\alpha)(\epsilon)]\right) \\ I(X; Y) - I(Y; Z) &= H_b\left(\frac{1}{2} + \frac{\alpha}{2}\right) + H_b(\epsilon) - H_b\left(\frac{1}{2}[1 + \alpha - 2\alpha\epsilon]\right) - H_b(\alpha) \end{aligned}$$

Recall that a necessary condition for analog authentication is  $I(X; Y) \leq I(Y; Z)$ . For the distribution

we have chosen, Figure (4-3) shows the minimum value of  $\alpha$  required for a given value of  $\epsilon$  to satisfy the constraint  $I(X; Y) \leq I(Y; Z)$ .

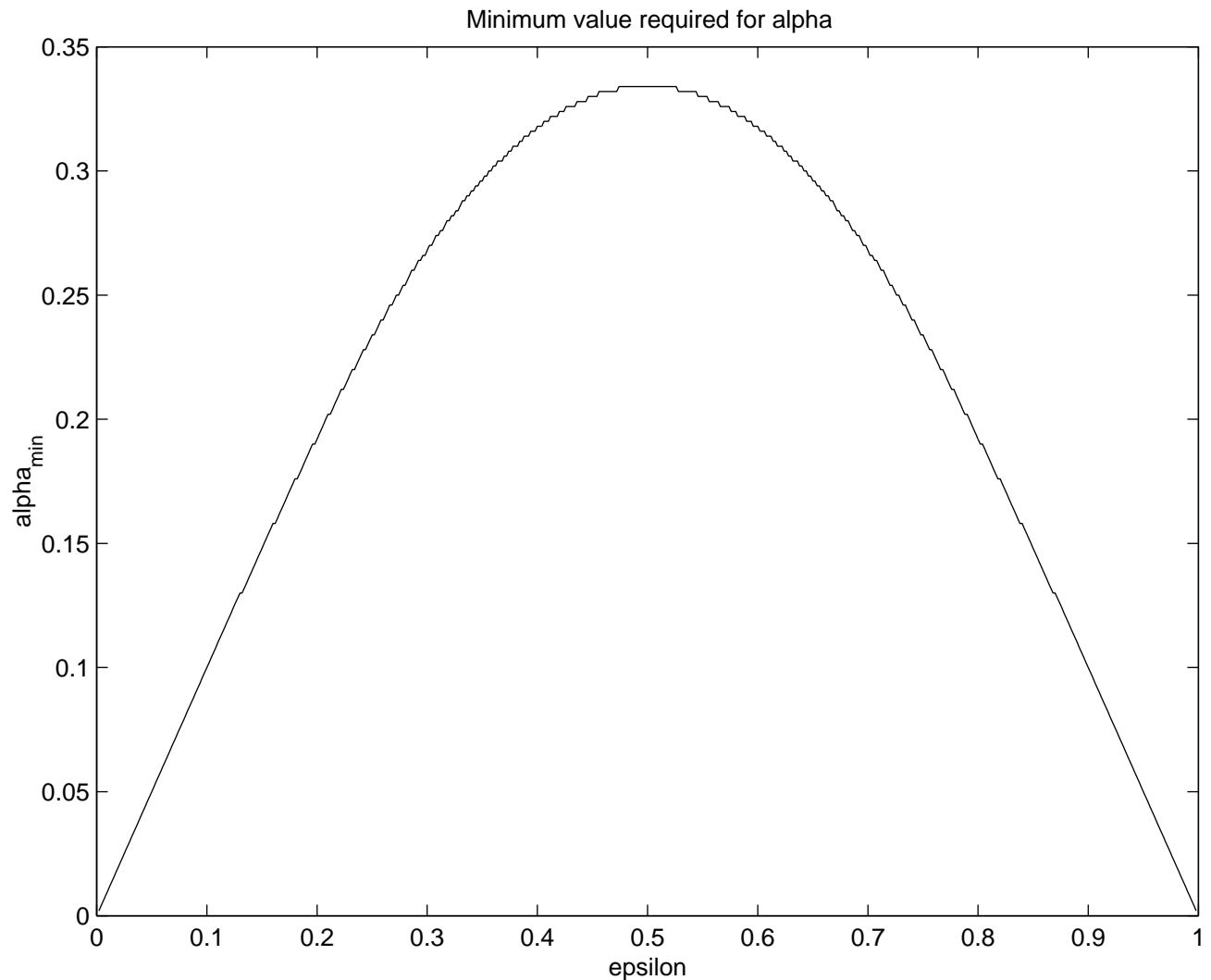


Figure 4-3: Minimum value of  $\alpha$  for a given  $\epsilon$  required to make  $I(X; Y) \leq I(Y; Z)$ .

Consider the case when  $\epsilon = 1/2$  which corresponds to the classic noisy typewriter channel. A naive coding scheme would be to only transmit odd numbers or only transmit even numbers. For example, the encoder could only transmit odd numbers by setting  $Y = X$  when  $X$  is odd and  $Y = X - 1$  when  $X$  is even which corresponds to the  $\alpha = 1$  case. Then the decoder could decode by selecting  $\hat{Y} = Z$  when  $Z$  is odd and  $\hat{Y} = Z - 1$  when  $Z$  is even. The probability of error for this scheme will be 0, and the encoding distortion would be  $D = \alpha/2 = 1/2$ . However, from Figure (4-3) we see that we can do better by using a codebook with the distribution in (4.2) and  $\alpha \approx 0.33$ . Using this distribution and appropriate coding, arbitrarily low probability of error can be achieved with embedding distortion  $D = \alpha/2 \approx 0.165$ .

Calculating  $E[d(X, Y)]$  for this simple choice of  $p(y|x)$  shows that coding can achieve a significantly lower distortion than a naive scheme. Furthermore  $E[d(X, Y)]$  shows how much coding gain we can achieve with a particular choice of  $p(y|x)$ . Once we construct a codebook for this simple  $p(Y|X)$ , which achieves a distortion close to  $E[d(X, Y)]$ , we know that there are no further gains that can be made by coding. To lower the distortion further we need to choose a better distribution for  $p(y|x)$  that achieves a shaping gain.

### 4.3 Gaussian Source, Squared Error Distortion, AWGN channel

One of the possible applications of this work includes images which can be modeled as Gaussian random variables. Similarly, a simple choice for the distortion function is mean square distortion. Finally, a reasonable model for the channel perturbations is additive white Gaussian noise. Therefore we model the source,  $\{X_k\}$ , as a sequence of i.i.d. Gaussian random variables with mean 0 and variance  $\sigma_X^2$ . We model the output of the channel as  $Z_k = Y_k + N_k$  where  $\{N_k\}$  is a sequence of i.i.d. Gaussian random variables with mean 0 and variance  $\sigma_N^2$ . The distortion function is  $d(X, Y) = (X - Y)^2$ .

Technically, the theorems we have proven do not apply to this scenario. Our proofs were for discrete alphabets with a bounded distortion function. A Gaussian source is a continuous source and mean square error is not a bounded distortion function. We could extend the theorems to continuous sources by discretizing a continuous source into  $K$  bins and letting the number of bins go to infinity. We could extend the theorems to unbounded distortion functions using the techniques in [17]. We conjecture that our results hold for this scenario and analyze the implications of this conjecture.

We would like to compute the fundamental distortion for this model. Since obtaining a closed form solution for  $D^*$  is not straight-forward we will develop upper and lower bounds for  $D^*$ . We first develop an upper bound for  $D^*$  by trying some appropriate distributions for  $p(y|x)$ .

The analog authentication problem is a joint source-channel coding problem. The source coding part for this example corresponds to quantizing a Gaussian source with mean square distortion. If we were only interested in the source coding aspect, we could find the optimum distribution from the rate-distortion theorem for the Gaussian source. The rate-distortion optimizing distribution,  $p_R(Y|X)$  is

$$Y|X \sim \mathcal{N}\left(\frac{\sigma_X^2 - D}{\sigma_X^2}x, \frac{(\sigma_X^2 - D)D}{\sigma_X^2}\right)$$

which yields  $Y \sim \mathcal{N}(0, \sigma_X^2 - D)$ . Therefore  $I(X; Y) = \frac{1}{2} \log\left(\frac{\sigma_X^2}{D}\right)$ . Furthermore since  $Z = Y + N$ ,  $Z$

will now be 0 mean with variance  $\sigma_X^2 - D + \sigma_N^2$ , so

$$\begin{aligned}
I(Y; Z) &= h(Z) - h(Z|Y) \\
&= h(Z) - h(Y + N|Y) \\
&= h(Z) - h(N) \\
&= \frac{1}{2} \log(2\pi e(\sigma_X^2 - D + \sigma_N^2)) - \frac{1}{2} \log(2\pi e\sigma_N^2) \\
I(Y; Z) &= \frac{1}{2} \log\left(\frac{\sigma_X^2 - D + \sigma_N^2}{\sigma_N^2}\right)
\end{aligned}$$

Since we require that  $I(X; Y) \leq I(Y; Z)$  we can compute the minimum possible value of  $D$  achieved by  $p_R(y|x)$  by solving for  $I(X; Y) = I(Y; Z)$  yielding

$$\frac{\sigma_X^2}{D} = \frac{\sigma_X^2 - D + \sigma_N^2}{\sigma_N^2} \quad (4.3)$$

By inspection, both  $D = \sigma_X^2$  and  $D = \sigma_N^2$  satisfy the equation above. These distortions are achievable using the distribution  $p_R(y|x)$  so

$$D^* \leq \min(\sigma_X^2, \sigma_N^2) \quad (4.4)$$

Writing this bound in terms of distortion to noise ratio and signal to noise ratio yields

$$DNR \leq \min(1, SNR)$$

We obtained the upper bound for  $D^*$  by considering the source coding aspect of the problem. We can find another upper bound by considering the channel coding aspect. We can use the results regarding channel capacity for the additive white Gaussian noise channel with a power constraint to find the maximum of  $I(Y; Z)$ . The channel capacity optimizing distribution,  $p_C(y)$ , corresponds to making  $Y$  Gaussian with mean 0 and variance  $\sigma_X^2 + D$ . This requires choosing the conditional distribution  $p_C(y|x)$  corresponding to

$$Y|X \sim \mathcal{N}(x, D)$$

This yields

$$\begin{aligned}
I(X; Y) &= \frac{1}{2} \log\left(\frac{\sigma_X^2 + D}{D}\right) \\
I(Y; Z) &= \frac{1}{2} \log\left(\frac{\sigma_X^2 + D + \sigma_N^2}{\sigma_N^2}\right)
\end{aligned}$$

We can solve for the minimum distortion for  $p_C(y|x)$  by setting  $I(X;Y) = I(Y;Z)$  to obtain

$$\frac{\sigma_X^2 + D}{D} = \frac{\sigma_X^2 + D + \sigma_N^2}{\sigma_N^2}$$

Multiplying this out and using the quadratic formula to solve for  $D$  yields

$$D = \frac{\sigma_X^2}{2} \left( -1 + \sqrt{1 + 4 \frac{\sigma_N^2}{\sigma_X^2}} \right)$$

Since this distortion is achievable it yields an upper bound for  $D^*$

$$D^* \leq \frac{\sigma_X^2}{2} \left( -1 + \sqrt{1 + 4 \frac{\sigma_N^2}{\sigma_X^2}} \right) \quad (4.5)$$

Writing this in terms of distortion to noise ratio and signal to noise ratio yields

$$DNR \leq \frac{SNR}{2} \left( -1 + \sqrt{1 + \frac{4}{SNR}} \right)$$

After developing a lower bound for  $D^*$ , we will compare the lower bound with the upper bounds in (4.5) and (4.4).

For the Bernoulli(1/2) example, viewing the problem as source coding or channel coding yielded the same results. This is because the rate-distortion optimizing distribution  $p(y|x)$  mapped into the channel capacity optimizing distribution  $p(y)$ . This does not occur for the Gaussian model.

We can derive a lower bound for  $D^*$  by lower bounding  $I(X;Y)$  and upper bounding  $I(Y;Z)$ . Since the distribution that achieves the minimum of  $I(X;Y)$  is different than the distribution that achieves the maximum of  $I(Y;Z)$ , the bound we obtain from this method will probably not be tight.

$$\begin{aligned} I(X;Y) &\geq \frac{1}{2} \log \left( \frac{\sigma_X^2}{D} \right) \\ I(Y;Z) &\leq \frac{1}{2} \log \left( \frac{\sigma_X^2 + \sigma_N^2 + D}{\sigma_N^2} \right) \\ \frac{\sigma_X^2}{D} &\leq \frac{\sigma_X^2 + \sigma_N^2 + D}{\sigma_N^2} \end{aligned}$$

Cross multiplying and gathering terms yields

$$D^2 + D(\sigma_X^2 + \sigma_N^2) - \sigma_X^2 \sigma_N^2 \geq 0$$

which implies

$$D \geq \frac{-\sigma_X^2 - \sigma_N^2 + \sqrt{\sigma_X^4 + \sigma_N^4 + 6\sigma_X^2 \sigma_N^2}}{2}$$

By dividing both sides of the equation above by  $\sigma_N^2$ , we can express the bound in terms of the distortion to noise ratio and the signal to noise ratio.

$$DNR \geq \frac{1 + SNR}{2} \left( -1 + \sqrt{1 + \frac{4SNR}{(1 + SNR)^2}} \right) \quad (4.6)$$

Table 4.1: Comparison of distortion to noise ratio (DNR) at various signal to noise ratio (SNR)

Rate-distortion optimizing distribution	$DNR = \min(1, SNR)$
Channel capacity optimizing distribution	$DNR = \frac{SNR}{2} \left( -1 + \sqrt{1 + \frac{4}{SNR}} \right)$
Lower Bound	$DNR \geq \frac{1+SNR}{2} \left( -1 + \sqrt{1 + \frac{4SNR}{(1+SNR)^2}} \right)$

Table 4.1 compares the various bounds. As shown in Figure 4-4, the channel capacity optimizing distribution is better than the rate-distortion optimizing distribution for large SNR, but the rate-distortion optimizing distribution wins for small SNR. The crossover point is at  $\frac{\sigma_X^2}{\sigma_N^2} = \frac{1}{2} = -3$  dB, as can be verified from Table (4.1). Numerical calculations show that the gap between the lower bound in equation (4.6) and the better of our two distributions is always less than 2.6 dB with the maximum gap at the crossover point. Note that the lower bound is asymptotically tight in the low SNR and high SNR limits.

How much more distortion does analog authentication cause each party over the uncoded case? In the high SNR limit, our results indicate that  $D^* = \sigma_N^2$ . Therefore the signal for the receiver without a decoder suffers distortion  $D^* + \sigma_N^2 = 2\sigma_N^2$  instead of  $\sigma_N^2$ . The receiver with a decoder suffers the same amount of distortion as in the uncoded case. Therefore the cost of analog authentication is 0 dB to the sophisticated receiver and 3 dB to the simple receiver.

When the SNR is much less than -3 dB, the optimal distribution becomes closer and closer to  $p(y|x) = \delta(y)$ . This corresponds to zeroing out the source completely. Therefore analog authentication is impossible in the low SNR limit. The rate-distortion distribution zeros out the source as soon as the SNR reaches 0 dB. Any scheme which requires  $D \geq \sigma_X^2$  is effectively ignoring the source so schemes which have  $D \geq \sigma_X^2$  do not really succeed. However, the channel capacity distribution does better than this for  $SNR > -3$  dB. Thus for  $-3 \text{ dB} < SNR < 0 \text{ dB}$ , by using the channel capacity distribution analog authentication is possible without zeroing out the source. This means that even when the channel noise is greater than the variance of the signal, analog authentication is possible.

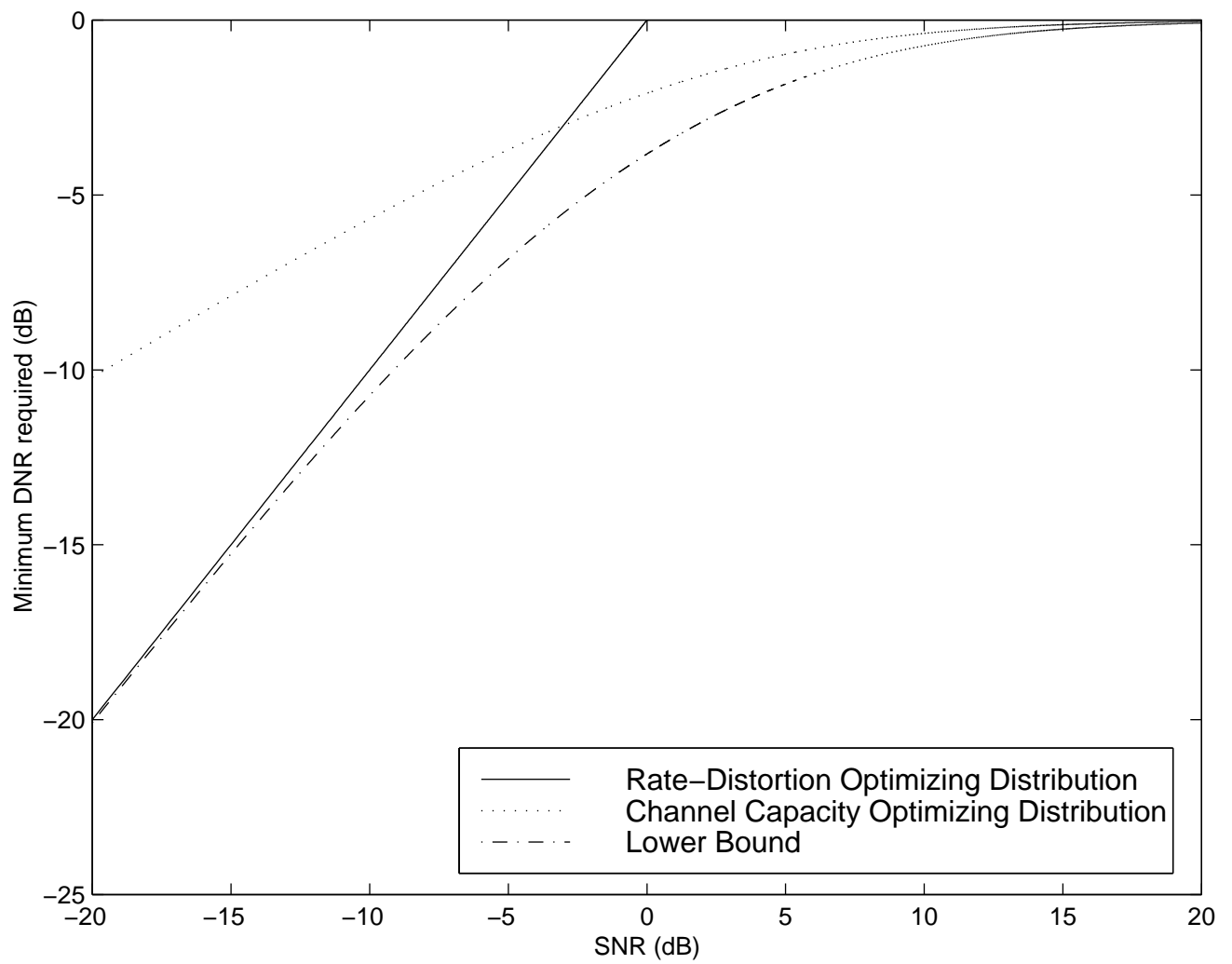


Figure 4-4: Comparison of distortion to noise ratio (DNR) at various signal to noise ratio (SNR)

## 4.4 Uniform Source, Squared Distortion, AWGN channel

Another possible model is a sequence of i.i.d. uniformly distributed random variables. We will analyze a uniform source using squared error as the distortion metric, and additive white Gaussian noise as the perturbation. In Chapter 5, we discuss practical implementations based on this model. As in the previous example, the theorems we proved did not address continuous sources. We conjecture our results hold and examine the implications.

We model the source as a sequence of i.i.d. random variables,  $\{X_k\}$ , uniformly distributed over  $[-L, L]$  and use distortion metric  $d(X, Y) = (X - Y)^2$ . We model the channel output as  $Z_k = Y_k + N_k$  where  $\{N_k\}$  is a sequence of i.i.d. Gaussian random variables with mean 0 and variance  $\sigma_N^2$ . We analyze this source by computing bounds as in the Gaussian source case. We obtain the same results with a difference of the factor  $\frac{6}{\pi e}$  which commonly occurs when comparing uniform distributions to Gaussian distributions.

We obtain a lower bound for  $I(X; Y)$  by expanding it to get  $h(X) - h(Y|X)$ . Since  $E[(X - Y)^2] \leq D$  by assumption, the continuous version of Fano's inequality implies  $h(Y|X) \leq \frac{1}{2} \log 2\pi e D$ . Therefore

$$I(X; Y) \geq \frac{1}{2} \log \frac{4L^2}{24\pi e D}$$

We obtain an expression for  $I(Y; Z)$ , by expanding mutual information in terms of entropies and noting that shifting does not change differential entropy

$$I(Y; Z) = h(Z) - h(Z|Y) \tag{4.7}$$

$$= h(Z) - h(N + Y|Y) \tag{4.8}$$

$$= h(Z) - \frac{1}{2} \log 2\pi e \sigma_N^2 \tag{4.9}$$

$$\tag{4.10}$$

We can obtain an upper bound for  $I(Y; Z)$  using the well known result that a Gaussian maximizes differential entropy subject to a variance constraint. A simple calculation shows  $Var(X) = \frac{4L^2}{12}$  and since  $E[(X - Y)^2] \leq D$ ,  $Var(Y) \leq \frac{4L^2}{12} + D$ . Therefore

$$I(Y; Z) \leq \frac{1}{2} \log \frac{\frac{4L^2}{12} + D + \sigma_N^2}{\sigma_N^2}$$

Combining the bounds on  $I(X;Y)$  and  $I(Y;Z)$  yields

$$\frac{1}{2} \log \frac{4L^2}{24\pi e D} \leq \frac{1}{2} \log \frac{\frac{4L^2}{12} + D + \sigma_N^2}{\sigma_N^2} \quad (4.11)$$

$$\frac{4L^2}{24\pi e D} \leq \frac{\frac{4L^2}{12} + D + \sigma_N^2}{\sigma_N^2} \quad (4.12)$$

$$4L^2 \sigma_N^2 \leq 8L^2 + 24\pi e D^2 + 24\pi e D \sigma_N^2 \quad (4.13)$$

from which we obtain

$$D \geq \frac{-(\sigma_X^2 + \sigma_N^2) + \sqrt{(\sigma_X^2 + \sigma_N^2)^2 + 24 \frac{\sigma_X^2 \sigma_N^2}{\pi e}}}{2}$$

Rewriting this in terms of distortion to noise ratio and signal to noise ratio yields

$$D \geq \frac{1 + SNR}{2} \left( -1 + \sqrt{1 + \frac{6}{\pi e} \frac{4SNR}{(1 + SNR)^2}} \right) \quad (4.14)$$

The lower bound above has the same form as the lower bound for the Gaussian source in (4.6) except for the factor of  $\frac{6}{\pi e}$  under the square root.

Figure 4-5 compares the lower bound derived above as well as the lower bound for a Gaussian source. The plot shows that the lower bounds for the two sources behave similarly. For the uniform source, expanding  $D/\sigma_N^2$  for large SNR, indicates  $D/\sigma_N^2$  goes to  $\frac{6}{\pi e} = -1.53$  dB. For the Gaussian source,  $D/\sigma_N^2$  goes to 0 dB in the high SNR limit. Similarly, expanding  $D/\sigma_N^2$  for small SNR indicates that  $D/\sigma_N^2$  goes to  $\frac{6}{\pi e} SNR$  for the uniform source. For the Gaussian source,  $D/\sigma_N^2$  goes to  $SNR$  in the low SNR limit. Thus the bound for the uniform source is basically the same as the bound for the Gaussian source shifted down by 1.53 dB.

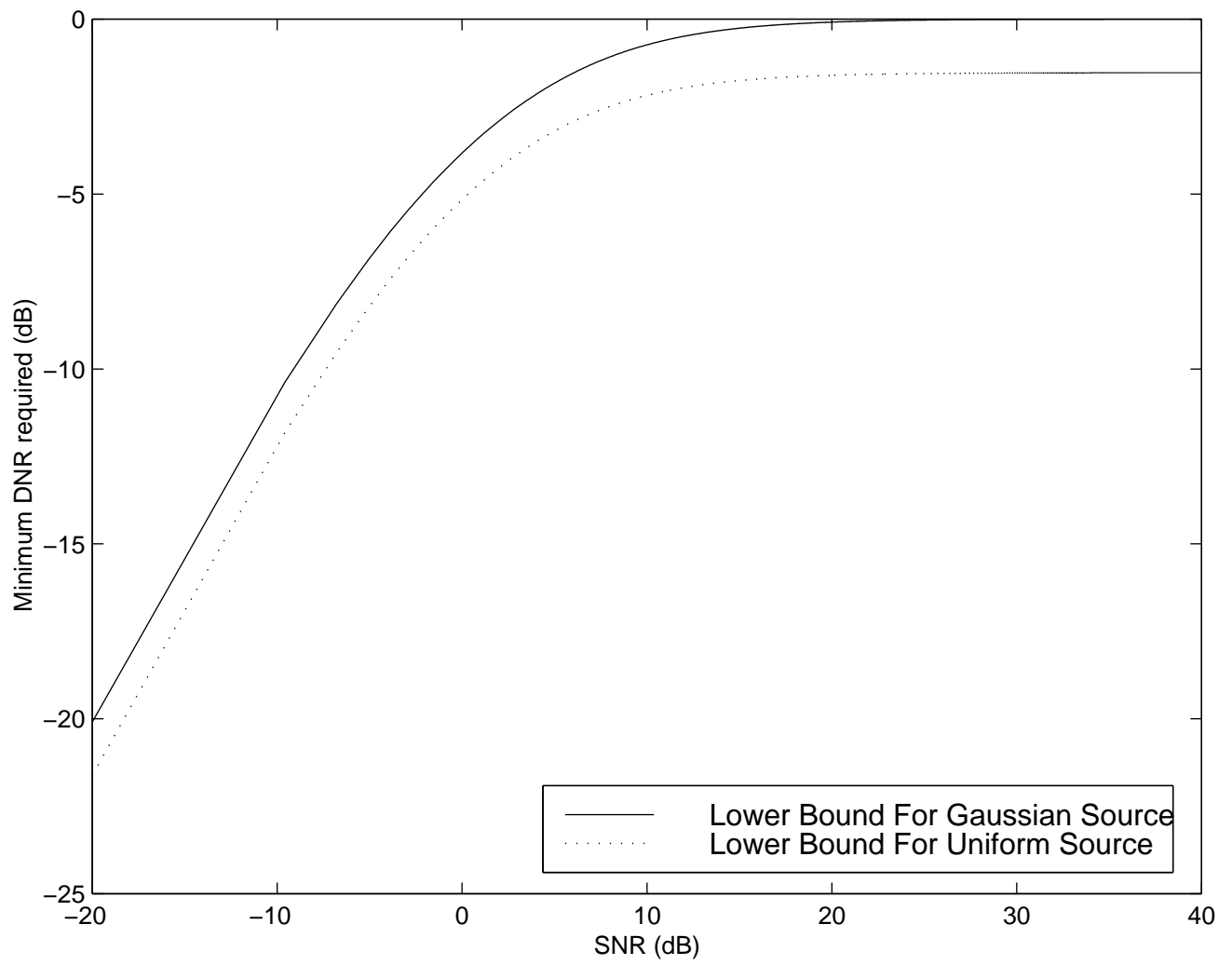


Figure 4-5: Comparison of distortion to noise ratio (DNR) bounds at various signal to noise ratio (SNR).

## Chapter 5

# Practical Implementation And Performance

In this section we develop and evaluate a detailed implementation of an analog authentication scheme for images. We model an image with  $n$  pixels, as an i.i.d sequence of random variables  $\{X_k\}_{k=1}^n$  uniformly distributed over  $[-L, L]$ . The image is processed to obtain  $Y_1^n$ . The processed image could suffer perturbations from GIF or JPEG compression, format changes, printing and scanning, smudges, etc. We model these perturbations as additive noise where the channel output is  $Z_k = Y_k + N_k$ . The noise,  $N_1^n$ , is a sequence of i.i.d. Gaussian random variables with mean zero and variance  $\sigma_N^2$ . Distortion is measured via mean square error:

$$D = \frac{1}{n} \sum_{k=1}^n E[(X_k - Y_k)^2]$$

We develop an analog authentication scheme for this scenario for a computationally bounded enemy. Since we model the enemy as computationally bounded, we can use public key digital signature schemes instead of shared secret keys. In Section 3.3.1 we summarized the important features of a digital signature scheme such as the signing algorithm,  $\mathcal{S}$ , the verifying algorithm  $\mathcal{V}$ , the security parameter,  $k$ , and the tag length,  $t$ . The sender generates his own secret key,  $s_k$ , and publishes his public key,  $p_k$ .

## 5.1 Algorithm Description: The Uncoded Case

### 5.1.1 Input, Design, And Performance Parameters

We first develop an uncoded algorithm and later describe how to add error correction coding. The problem parameters which the designer can not control are:

1. **The Source Size,  $L$ :** We model the source sequence is i.i.d. uniform over  $[-L, L]$ .
2. **The Noise Power,  $\sigma_N^2$ :** We model the channel as additive white Gaussian noise with mean 0 and variance  $\sigma_N^2$ .
3. **The Sequence Length,  $n$ :** We model the image as a sequence of  $n$  independent random variables.

The following design parameters are chosen by the engineer:

1. **The Security Parameter  $k$ :** The security parameter represents a bound on how much work the enemy must do to break the digital signature scheme.
2. **The Tag Length  $t$ :** The digital signature scheme produces  $t$  bit tags. Ideally  $t$  should be small. However most digital signature schemes have  $t$  as a function of  $k$ .
3. **The Encoding Distortion  $D$ :** The encoding distortion represents how much the original signal is disturbed in the encoding process.

The performance measures are a result of the problem parameters and the design parameters:

1. **Probability of Undetected Error,  $p_u$ :** This is the probability that the receiver accepts a signal which was not sent by the transmitter as defined in (3.2). This will largely be determined by the digital signature scheme and the security parameter  $k$ .
2. **Probability of False Alarm Error,  $p_f$ :** This is the probability that the receiver does not accept a signal which was sent by the transmitter and not tampered with as defined in (3.4).

### 5.1.2 Description of Encoding

The transmitter chooses a rate  $R$  uniform scalar quantizer to use in quantizing each sample of the source. The  $2^R$  reconstruction points are at  $\pm(2i - 1)L/2^R$  as shown in Figure 5-1. Later we specify how to choose  $R$  based on  $D$ .

If the transmitter were to simply quantize  $X$  using the specified reconstruction points, the expected distortion would be

$$D_S = \frac{4L^2}{12} 2^{-2R} = \sigma_X^2 2^{-2R}$$

Instead the transmitter chooses  $t$  distinct indices from 1 to  $n$ :  $\{I_1, I_2, \dots, I_t\}$  to store digital signature information. One good method is to choose these randomly and uniformly over all the samples, but other choices are possible. For example, the transmitter might choose  $\{28, 59, 33\}$  when  $t = 3$ . Finally, the transmitter chooses his public and private key pair  $(p_k, s_k)$  for the digital signature algorithm. The transmitter then publicly announces the digital signature chosen, the public key  $p_k$ ,

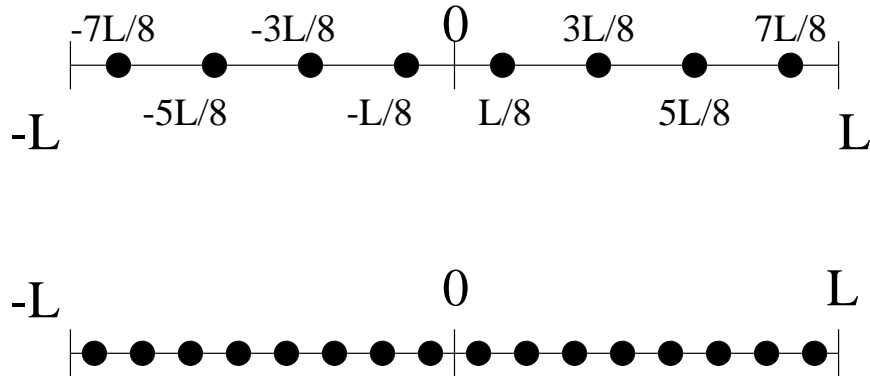


Figure 5-1: Some example reconstruction points for  $R = 3$  (top), and  $R = 4$  (bottom).

the reconstruction points and rate of the quantizer, and the indices where the digital signature tag will be embedded,  $\{I_1, I_2, \dots, I_t\}$ .

To process a source,  $X_1^n$ , the transmitter quantizes  $X_1^n$ . This maps the sequence of  $n$  source symbols to a sequence of  $n$  blocks of  $2^R$  bits each. We call the resulting sequence  $F(X_1^n)$ . The transmitter sets the least significant bit of block  $I_1$  to 0 and repeats for blocks  $I_2$  through  $I_t$ . We call the result of the second step in the quantization process  $G(F(X_1^n))$ .

Next, the transmitter computes the  $t$ -bit tag as  $\sigma = \mathcal{S}(G(F(X_1^n)), s_k)$ . The transmitter sets the least significant bit in block  $I_j$  to be the  $j$ th bit of the tag,  $\sigma$ . We call the resulting sequence of bits  $Q(G(F(X_1^n)))$ . The transmitter then constructs  $Y_1^n = P(Q(G(F(X_1^n))))$  by reconstructing the sequence of bits using the reconstruction points specified earlier. Figure 5-2 shows a diagram of the encoding process.

The process of embedding the digital signature tag in the least significant bits is based on a digital watermarking method called Low Bit Modulation (LBM) [18]. More efficient watermarking schemes such as Quantization Index Modulation (QIM) [19] could also be used. For most digital signature schemes, the tag length is small enough that the distortion difference between LBM and QIM is negligible compared to the overall processing distortion.

### 5.1.3 A Small Example To Illustrate Encoding

Consider a source with  $L = 8$  and  $n = 4$  encoded with a rate 3 quantizer. For simplicity, we model the digital signature scheme as producing 1-bit tags. The  $2^R = 8$  reconstruction points are

$$\{-7, -5, -3, -1, 1, 3, 5, 7\}$$

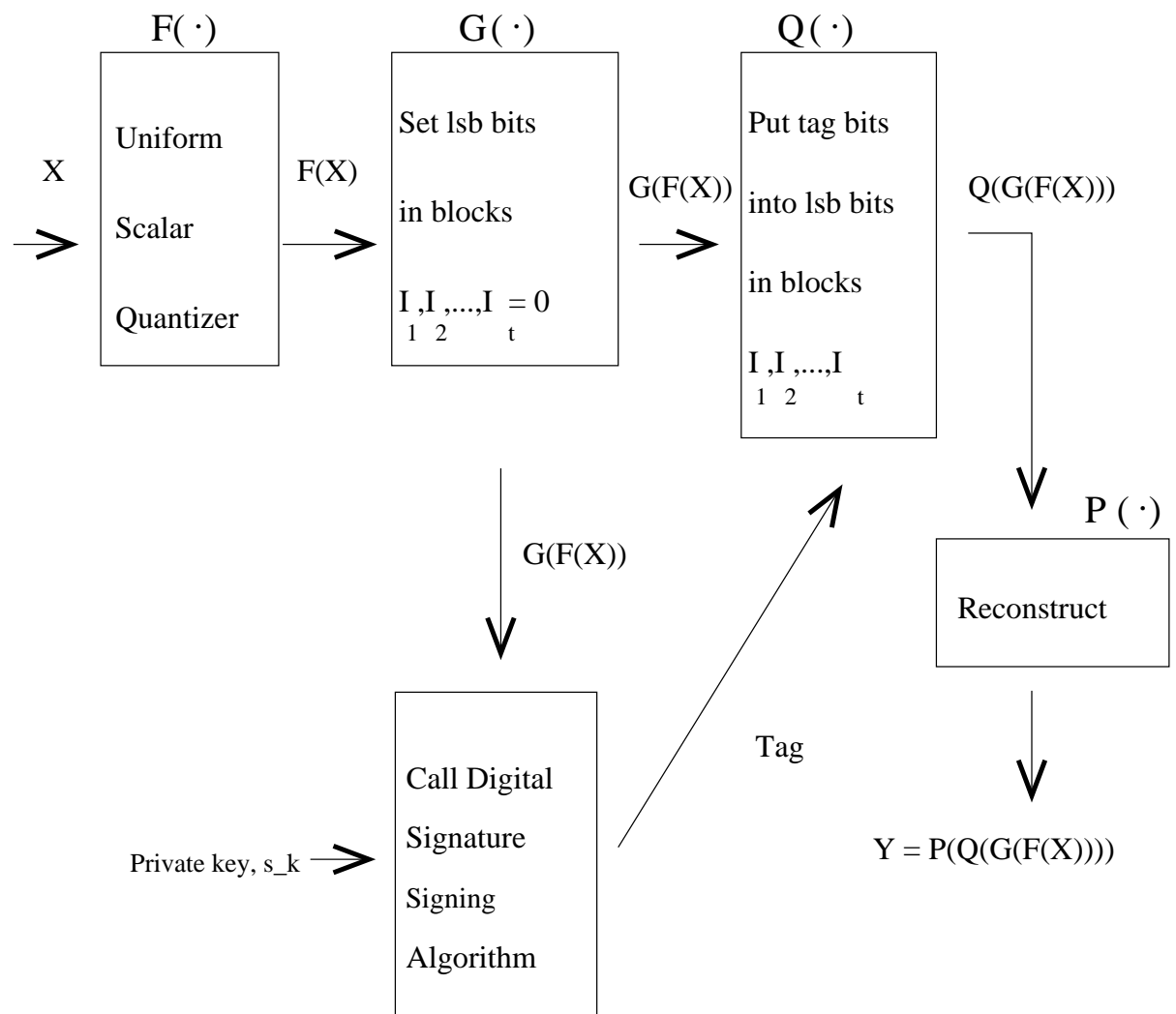


Figure 5-2: Diagram of the encoding process.

Assume that  $I_1$  is randomly chosen to be 3. To encode the source sequence  $X_1^n = \{5.6, -2.3, 7.2, 1.1\}$ , the transmitter quantizes  $X_1^n$  to get

$$F(X_1^n) = \{5, -3, 7, 1\} \rightarrow \{110, 010, 111, 100\}$$

Note that we represent points quantized to -7 as 000, points quantized to -5 as 001, and so on. Therefore -7 is the 0th quantization point, -5 is the 1st quantization point, -3 is the 2nd quantization point, etc. Next, the transmitter sets least significant bit in block  $I_1 = 3$  to be 0. Thus

$$G(F(X_1^n)) = \{110, 010, 110, 100\}$$

Assume the tag is computed to be  $\sigma = \mathcal{S}(110, 010, 110, 100, s_k) = 0$ . The transmitter sets the least significant bit in block  $I_1 = 3$  to be the tag. So

$$Q(G(F(X_1^n))) = \{110, 010, 110, 100\}$$

To get  $Y_1^n$ , the transmitter reconstructs the quantized representation with the specified reconstruction points to get

$$Y_1^n = P(Q(G(F(X_1^n)))) = \{5, -3, 5, 1\}$$

The encoding distortion is

$$\begin{aligned} D &= \frac{1}{n} \sum_{i=1}^n (X_i - Y_i) \\ D &= \frac{1}{4} [(0.6)^2 + (0.7)^2 + (2.2)^2 + (0.1)^2] \\ D &= 1.425 \end{aligned}$$

Note that there are two contributors to the encoding distortion. The first contribution comes from quantizing  $X_1^n$  with a uniform scalar quantizer. The second contribution comes from low bit modulating the low bit of  $X_3$ .

#### 5.1.4 Description of Decoding

The receiver receives  $Z_1^n$  and performs maximum likelihood decoding. The receiver first quantizes  $Z_1^n$  with the same quantizer used by the encoder to get  $F(Z_1^n)$ . Next the receiver chooses the first bit of the tag,  $\sigma$ , to be the least significant bit of block  $I_1$  and repeats this process for  $\{I_2, I_3, \dots, I_t\}$ . Then the receiver sets the least significant bit in block  $I_1$  to 0 and repeats this process for  $\{I_2, I_3, \dots, I_t\}$

to get  $G(F(Z_1^n))$ . Finally the receiver verifies the digital signature by checking if

$$\mathcal{V}(\sigma, G(F(Z_1^n)), p_k) = 1$$

If this is the case, then the receiver accepts the decoded result as  $\hat{Y}_1^n = P(Q(G(F(Z_1^n))))$ . Otherwise the receiver declares a decoding failure. Figure 5-3 shows a diagram of the decoding process.

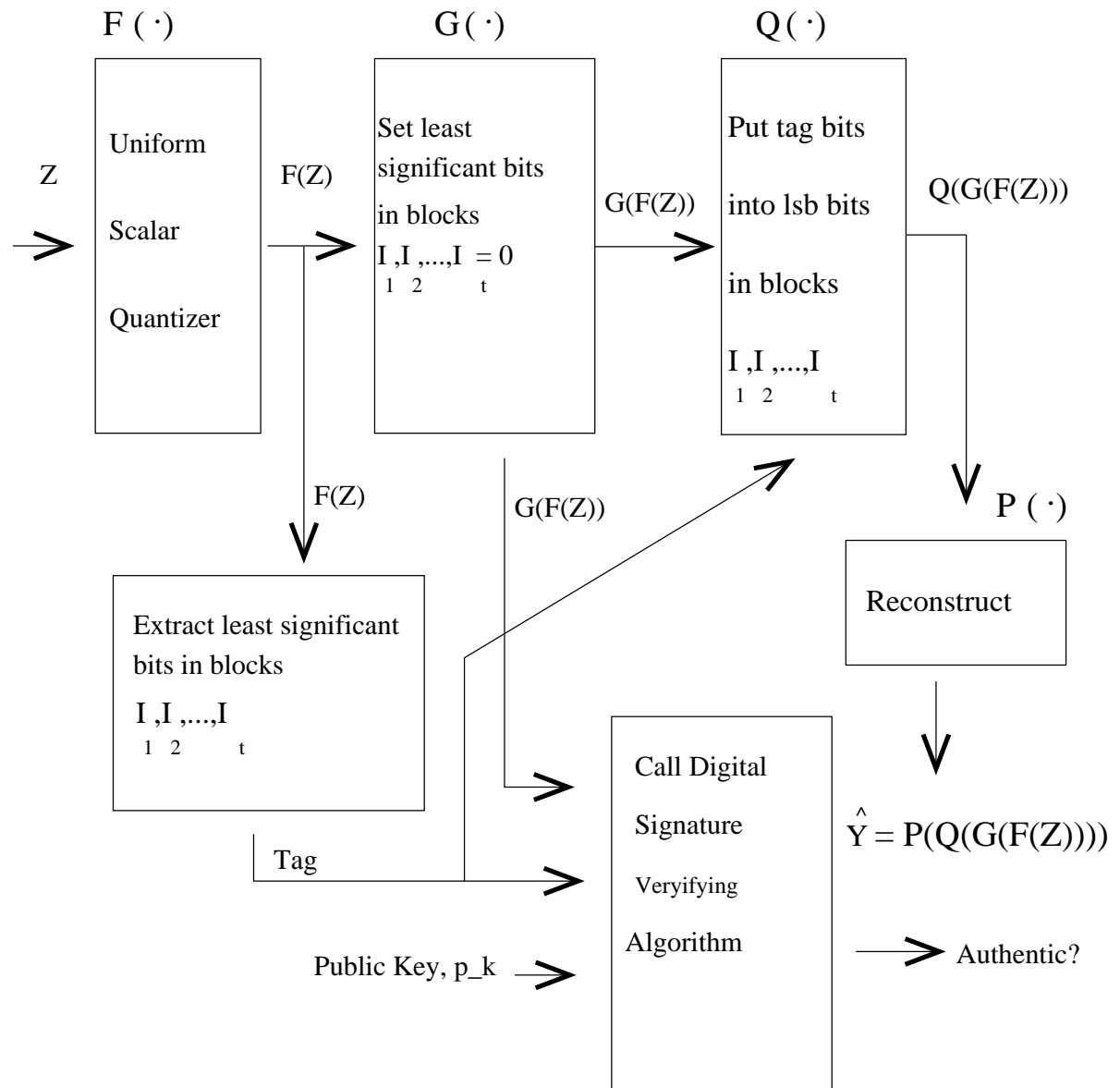


Figure 5-3: Diagram of the decoding process.

### 5.1.5 A Small Example To Illustrate Successful Decoding

In this example, the transmitter uses the design parameters outlined in Section 5.1.3 and sends  $Y_1^n = \{5, -3, 5, 1\}$ . Assume that the receiver gets  $Z_1^n = \{5.9, -3.3, 4.7, 0.1\}$ . The receiver first quantizes  $Z_1^n$  to the specified reconstruction points to get

$$F(Z_1^n) = \{5, -3, 5, 1\} \rightarrow \{110, 010, 110, 100\}$$

The receiver then chooses the first bit of the tag,  $\sigma$ , to be the least significant bit of block  $I_1 = 3$  yielding  $\sigma = 0$ . Next the receiver sets the least significant bit in block  $I_1$  to 0 to get

$$G(F(Z_1^n)) = \{110, 010, 110, 100\}$$

Finally, the receiver checks  $\mathcal{V}(\sigma, G(F(Z_1^n)), p_k)$  which must equal 1 in this case. Therefore the receiver accepts the decoded result as  $\hat{Y}_1^n = P(Q(G(F(Z_1^n)))) = \{5, -3, 5, 1\}$ . In this case the decoding is successful because the receiver decodes the received signal to exactly what the transmitter sent. Furthermore, since the signature verifies properly, the receiver is confident that the signal is not a forgery.

### 5.1.6 A Small Example To Illustrate Forgery Detection

In this example the transmitter sends  $Y_1^n = \{5, -3, 5, 1\}$  and an enemy tampers with the signal so that the receiver gets

$$Z_1^n = \{-2.2, 4.1, 5.2, 1.1\}$$

Here the enemy has modified the first two symbols to be completely different. We show that the receiver will detect this forgery.

The receiver quantizes  $Z_1^n$  to get

$$F(Z_1^n) = \{-3, 5, 5, 1\} \rightarrow \{010, 110, 110, 100\}$$

The receiver then chooses the first bit of the tag,  $\sigma$ , to be the least significant bit of block  $I_1 = 3$  yielding  $\sigma = 0$ . Next the receiver sets the least significant bit in block  $I_1$  to 0 to get

$$G(F(Z_1^n)) = \{010, 110, 110, 100\}$$

Finally, the receiver checks  $\mathcal{V}(\sigma, G(F(Z_1^n)), p_k)$ . In this case  $\sigma$  was produced as a tag for

$$Y_1^n = \{5, -3, 5, 1\} \rightarrow \{110, 010, 110, 100\}$$

Therefore a good digital signature algorithm would have  $\mathcal{V}(\sigma, G(F(Z_1^n)), p_k) = 0$ . In reality a good digital signature algorithm would require  $\sigma$  to be significantly more bits. If  $\sigma$  had  $t$ -bits, a good digital signature algorithm would have

$$Pr[\mathcal{V}(\sigma, G(F(Z_1^n)), p_k) = 1] \approx 2^{-t}$$

Therefore the receiver would declare this message to be a forgery and not accept it.

### 5.1.7 A Small Example To Illustrate False Alarm Error

Next we examine how a false alarm error can occur. In this example the transmitter sends  $Y_1^n = \{5, -3, 5, 1\}$  and due to a burst of noise in the channel the receiver gets

$$Z_1^n = \{5.1, -2.9, 4.9, -0.1\}$$

Note that the channel noise has moved the last symbol down by 1.1 units. We will show that this event will cause a false alarm error.

The receiver quantizes  $Z_1^n$  to get

$$F(Z_1^n) = \{5, -3, 5, -1\} \rightarrow \{110, 010, 110, 011\}$$

The receiver then chooses the first bit of the tag,  $\sigma$ , to be the least significant bit of block  $I_1 = 3$  yielding  $\sigma = 0$ . Next the receiver sets the least significant bit in block  $I_1$  to 0 to get

$$G(F(Z_1^n)) = \{110, 010, 110, 011\}$$

Finally, the receiver checks  $\mathcal{V}(\sigma, G(F(Z_1^n)), p_k)$ . In this case  $\sigma$  was produced as a tag for a different message. Therefore a good digital signature algorithm would result in  $\mathcal{V}(\sigma, G(F(Z_1^n)), p_k) = 0$  as discussed in the previous example. Consequently the receiver will declare the received signal a forgery. This is a false alarm error because the signal is actually valid in the sense that it has been transmitted over a noisy channel with no interference from the enemy.

### 5.1.8 Comments About The Decoding Examples

In this scheme the transmitter has chosen the quantization points such that the minimum distance between two quantization points is 2 units. Therefore the corresponding decision boundaries are at  $\pm 1$  unit from each quantization point as shown in Figure 5-4.

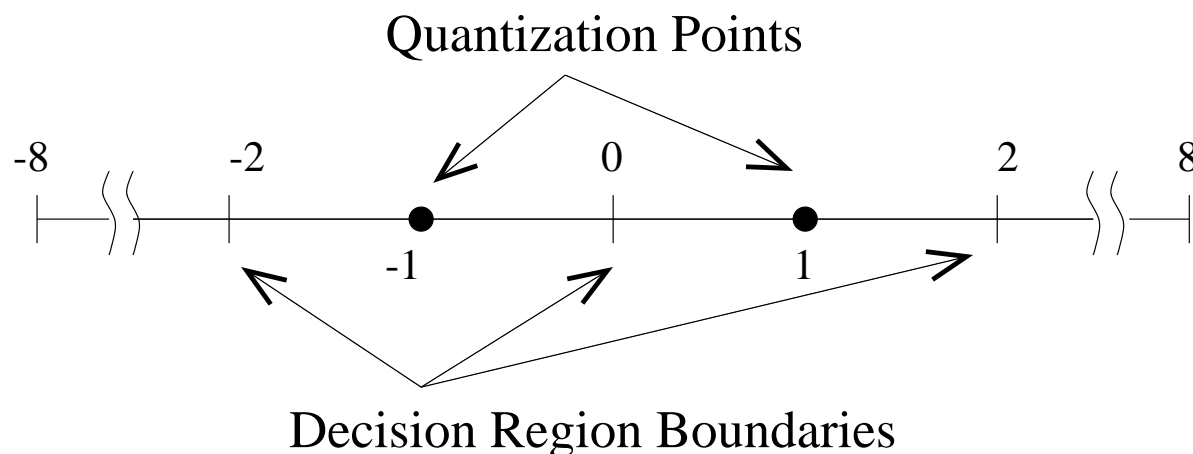


Figure 5-4: Quantization points and decision regions. Any point in the region  $(-2, 0]$  is mapped to the quantization point  $-1$ . Any point in the region  $(0, 2]$  is mapped to the quantization point  $1$ .

When the noise is less than 1 unit per symbol, it will not push a quantized point across a decision boundary. In the first example, the noise for each symbol is less than 1 unit. Therefore the decoding process can map the noisy signal points back to the original quantization points. The digital signature then verifies that the receiver has indeed received the correct signal points. Consequently the receiver recovers exactly what was transmitted and accepts the transmission as authentic.

In the second example, the enemy modifies the transmitted signal greatly. Thus the decoding process does not map the modified signal points back to the original quantization points. However the digital signature does not match the received signal points indicating that a forgery has occurred. The receiver realizes this and successfully detects the forgery.

In the third example, no tampering occurs, but the channel noise pushes the transmitted signal point at  $Y_4$  across a decision boundary. Therefore the point  $Z_4 = Y_4 + N_4$  is not mapped back to  $Y_4$ . Consequently the digital signature does not match and the receiver declares the received signal a forgery even though the enemy has not modified the signal.

If the transmitter had embedded at a rate  $R = 2$  instead of  $R = 3$ , the quantization points would have been twice as far apart. Thus the scheme would be able to tolerate twice as much noise. In this case the third example would have resulted in successful decoding instead of a false alarm error. However, decreasing the rate to  $R = 2$  would also cause a larger encoding distortion. This illustrates the tradeoff between encoding distortion and false alarm error in the uncoded case. Later we show how to use coding to lower the false alarm error without increasing the encoding distortion.

### 5.1.9 Analysis of the Uncoded Case

First we will discuss the probability of forgery. As mentioned earlier a forgery can only be successful by cracking the digital signature. A discussion of the security of digital signatures is beyond the scope of this work. However, many researches have proposed schemes which provably achieve strong levels of security given some reasonable assumptions. The level of security usually depends significantly on the tag length,  $t$ , as well as the complexity of the digital signature scheme.

The effect of large  $t$  will be to increase the encoding distortion. The encoding distortion incurred by quantizing the source to the  $2^R$  output points is

$$E[(X_i - F(X_i))^2] = \frac{4L^2}{12}2^{-2R} = \sigma_X^2 2^{-2R}$$

However, after this first quantization step, the low bits in blocks  $\{I_1, I_2, \dots, I_t\}$  are modulated with the digital signature tag,  $\sigma$ . If we model the tag bits as equally likely to be 0 or 1, the expected distortion for the modulated symbols is

$$E[(X_i - Q(X_{I_i}))^2] = \sigma_X^2 2^{-2R} + \sigma_X^2 \frac{15}{16} 2^{-2R}$$

Thus the total encoding distortion is computed by averaging these two to get

$$D = \frac{1}{n} \sum_{i=1}^n E[(X_i - Q(X_i))^2] = \sigma_X^2 2^{-2R} \left(1 + \frac{15t}{16n}\right) \quad (5.1)$$

For  $t \ll n$  we get

$$D = \sigma_X^2 2^{-2R}$$

Thus for small tag lengths, the encoding distortion is independent of  $t$ . This decouples the issues of security against forgery and encoding distortion. The designer can choose a digital signature scheme which provides enough security based on the preferred assumptions (e.g. computational resources of the enemy, symmetric key vs. asymmetric key, etc.) without regards to encoding distortion. We do not go into the details of digital signature schemes except to note that many popular schemes exist with tag lengths on the order of 500 – 1000 bits. Thus a 128 by 128 image would have  $0.029 \lesssim \frac{15t}{16n} \lesssim 0.057$ . In this example, the extra distortion caused by the factor  $(1 + \frac{15t}{16n})$  in equation (5.1) ranges from 0.1225 to 0.2417 dB. As we will show, this is a much smaller concern than the encoding distortion needed to keep the probability of false alarm error small.

Next we calculate the probability of false alarm error,  $p_f$ . If each symbol of  $Y_1^n$  is perturbed by less than half the distance between quantization points, then no errors occur. This condition is

equivalent to the event

$$\{|N_i| < L2^{-R}\} \forall i \in \{1, 2, \dots, n\}$$

Thus the probability that symbol  $i$  is in error will be  $p_s = \Pr[|N_i| > L2^{-R}]$ . For  $t \ll n$  we can write  $p_s$  in terms of  $D$  as

$$p_s = \Pr[|N_i| > \sqrt{3D}]$$

Since  $N_i$  is Gaussian with mean 0 and variance  $\sigma_N^2$  we get

$$p_s = 2Q\left(\sqrt{\frac{3D}{\sigma_N^2}}\right)$$

If any symbol is in error then the whole sequence will be in error. Since symbol errors are independent we obtain the following expression for the total probability of error:

$$p_f = 1 - (1 - p_s)^n$$

For  $p_s \ll 1$ , the total probability of false alarm error is roughly  $p_f \approx np_s$ .

We derived lower bounds on the minimum possible encoding distortion for a uniform source in an AWGN channel. We can measure performance by measuring how much more DNR the uncoded scheme requires than the lower bound at a given probability of symbol error. The lower bound for DNR depends on the signal to noise ratio. Therefore the DNR gap will also depend on the SNR. This is similar to the situation in classical communications where the gap to capacity is used as a performance measure.

Since the gap to capacity depends upon the transmission rate, the rate normalized quantity  $SNR_{norm}$  is used. This suggests that a signal to noise ratio normalized version of DNR could be useful in analog authentication. We define the normalized distortion to noise ratio as

$$DNR_{norm} = \frac{D}{D^*}$$

Since we were unable to derive an exact expression for  $D^*$ , we can not use the formula above. Instead of deriving  $D^*$  we derived a lower bound, therefore we can instead measure performance normalized to this lower bound as

$$LDNR_{norm} = \frac{D}{D_L^*}$$

where L indicates the use of a lower bound for  $D^*$ .

Every achievable scheme must have  $LDNR_{norm} \geq 1$ , however, we do not know if schemes which achieve  $LDNR_{norm} = 1$  are possible. Regardless,  $LDNR_{norm}$  is a reasonable way to measure the performance and distortion efficiency of analog authentication schemes. According to (4.14),  $D_L^*$  goes to  $\frac{6}{\pi e} \sigma_N^2$  in the high SNR limit and to  $\frac{6}{\pi e} \sigma_X^2$  in the low SNR limit. Therefore we can rewrite  $p_s$  as

$$p_s = 2Q \left( \sqrt{\frac{18}{\pi e} LDNR_{norm}} \right)$$

in the high SNR limit and

$$p_s = 2Q \left( \sqrt{\frac{18}{\pi e} \frac{LDNR_{norm}}{SNR}} \right)$$

in the low SNR limit. Figure 5-5 is a plot of  $p_s$  versus  $LDNR_{norm}$  in the high SNR limit.

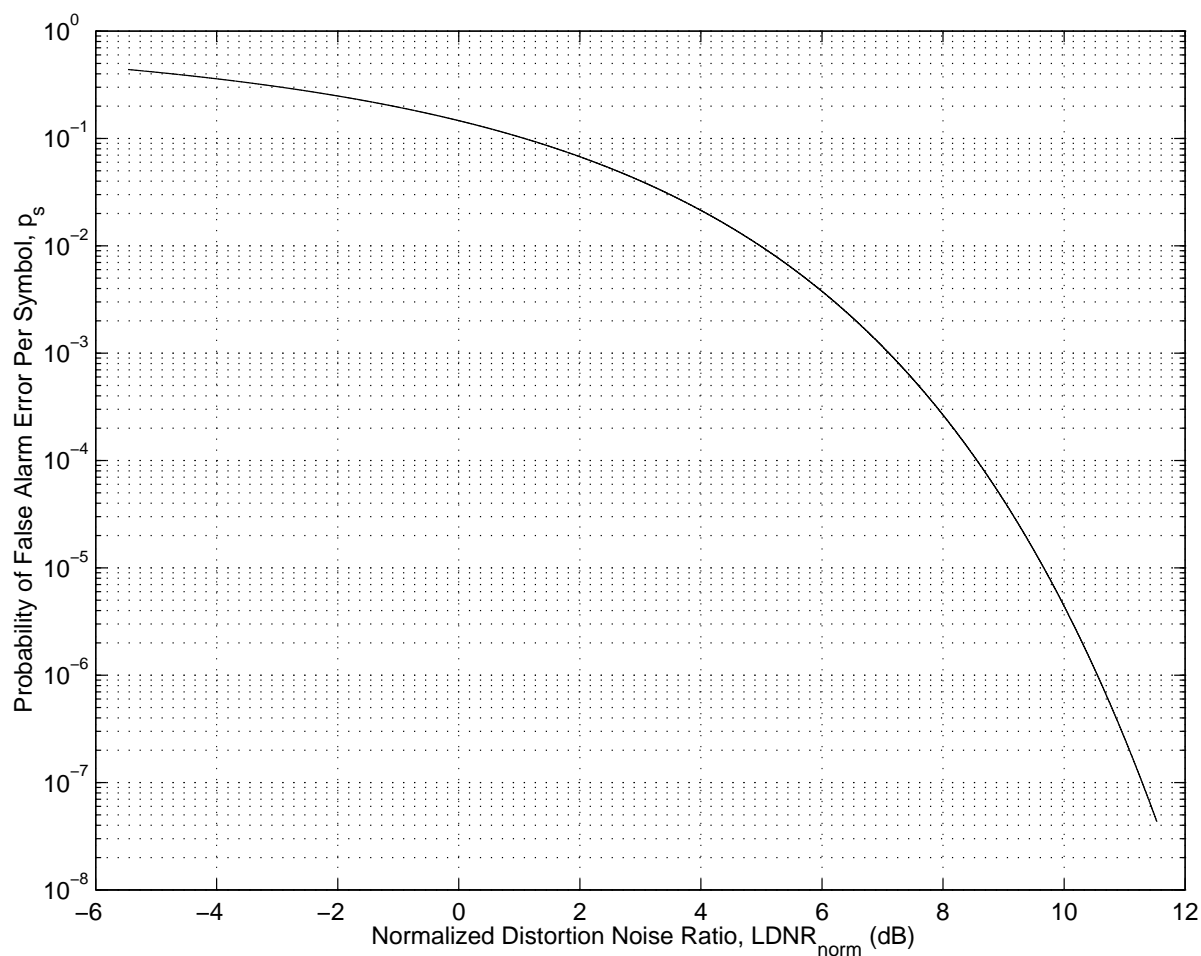


Figure 5-5: Uncoded probability of symbol error,  $p_s$ , as a function of distortion to noise ratio (DNR).

We plot the probability of symbol error,  $p_s$ , instead of the total probability of false alarm error,

$p_f$ , to separate the issue of sequence length from probability of symbol error. If a sequence with length  $n = 10^6$  is used, then

$$p_f \approx 10^{-5} \Rightarrow 10^6 p_s = 10^{-5} \Rightarrow p_s = 10^{-11}$$

Instead if we are interested in a sequence with length  $10^4$  at  $p_f \approx 10^{-5}$  we need  $p_s = 10^{-9}$ . By simply shifting the curve in Figure 5-5 by  $\log_{10} n$  we can find the required  $LDNR_{norm}$  for a given  $p_f$  with sequence length  $n$ .

The plot of  $p_s$  versus  $LDNR_{norm}$  shows the tradeoff between encoding distortion and probability of false alarm error for an uncoded system at high SNR. According to the analog authentication coding theorem arbitrarily low probability of error could be achieved as long as  $D > D^*$ . To achieve arbitrarily low  $p_f$  while bringing the encoding distortion close to  $D^*$  we need to employ coding. We analyze coded systems in the next section.

In summary, the encoding distortion, probability of forgery and probability of false alarm can be related by

$$p_f \approx nQ \left( \sqrt{\frac{3D}{\sigma_N^2} \frac{1}{1 + \frac{15t}{16n}}} \right) = nQ \left( \sqrt{\frac{3DNR}{1 + \frac{15t}{16n}}} \right)$$

The equation above includes the tradeoff between the probability of forgery via the tag length  $t$ . For many reasonable cases  $t \ll n$  and our result becomes

$$p_f \approx nQ \left( \sqrt{3DNR} \right)$$

## 5.2 Algorithm Design: The Coded Case

The distortion to noise ratio required by an uncoded scheme is roughly 10 dB from the lower bound we derived at  $p_s = 10^{-6}$ . We use coding to close the gap. We first describe a general method to apply coding and then discuss a detailed example based on Trellis Coded Modulation/Quantization (TCM/TCQ). We use TCM/TCQ as a representative example since it has reasonable complexity, provides coding gains between 3 and 6 dB, and is fairly simple to understand and implement. In Chapter 6 we mention other coding strategies which could achieve better results.

### 5.2.1 General Codes For Analog Authentication

Both the analog authentication problem and the backward compatibility broadcast channel discussed in Appendix A correspond to a joint source-channel coding problem. Therefore a coding scheme for analog authentication will have elements of both source coding and channel coding.

A source code consists of a quantizer and a reconstruction function,  $F, P$ . The quantizer maps a sequence of real numbers to bits  $F : \mathbb{R}^n \rightarrow \{0, 1\}^m$ , and the reconstruction function maps the

bits back to a sequence of real numbers  $P : \{0, 1\}^m \rightarrow \mathbb{R}^n$ . An error protection code consists of an encoder and a decoder,  $P, F$ . The encoder maps bits to a sequence of real numbers  $P : \{0, 1\}^m \rightarrow \mathbb{R}^n$ , while the decoder maps a sequence of real numbers to bits.  $F : \mathbb{R}^n \rightarrow \{0, 1\}^m$ .

Various authors have pointed out the duality between channel coding and source coding [20], [12]. This duality implies that an error correcting code and a source code can both be viewed simply as a code. This code can be used for either error correction or source coding. In analog authentication this code is used for joint error correction and source coding.

Coding can be incorporated into an analog authentication scheme by modifying the uncoded scheme shown in Figure 5-2. In the uncoded scheme  $F(\cdot)$  is a uniform scalar quantizer with the corresponding reconstruction function  $P(\cdot)$ . By instead choosing  $(F, P)$  based on a code, the performance of the scheme can be significantly enhanced. Thus using coding corresponds to choosing the appropriate code for  $(F, P)$  and also deciding which bits of  $F(X_1^n)$  are modulated to embed the digital signature tag.

Due to the duality between source coding and channel coding, a good pair  $(F, P)$  can be designed based on ideas from channel coding or source coding. If  $(F, P)$  is designed based on ideas from source coding, then  $F$  can be considered a quantizer with  $P$  as the corresponding reconstruction function. If  $(F, P)$  is designed based on ideas from channel coding, then  $P$  can be considered the encoder for an error correction code and  $F$  can be considered as the decoder for the error correction code. Ideally a code should be designed taking both viewpoints into account.

### 5.2.2 Encoding Using TCM/TCQ

As stated in the previous section, coding can be incorporated by choosing  $(F, P)$  from a code. In this section we design and evaluate  $(F, P)$  based on trellis codes. As a result of the duality between source codes and channel codes, trellis codes can be viewed either as source codes (Trellis Coded Quantization) [20], or as channel codes (Trellis Coded Modulation) [21], [22], [23].

The examples discussed for the uncoded case have the 8-PAM signal constellation shown in Figure 5-6. To apply trellis coding the constellation is doubled to yield 16 points which are divided into 4 cosets as shown in Figure 5-7. In the channel coding view,  $P$  is the 3 bit per symbol TCM encoder for the  $Z_1$  constellation discussed in [22], [23], and  $F$  is the corresponding decoder. In the source coding view,  $F$  is the 3 bit per symbol TCQ quantizer discussed in [20] and  $P$  is the corresponding reconstruction function. Both views lead to the same result for the code  $(F, P)$  but [20] analyzes the source coding gains, while [22], [23] analyzes the channel coding gains. Since both components contribute to the overall coding gain in analog authentication, each view adds a different insight.

Next we discuss where to embed the tag bits from the digital signature. One bit per symbol in  $F(X_1^n)$  corresponds to the coset choice for each symbol and the remaining bits per symbol correspond

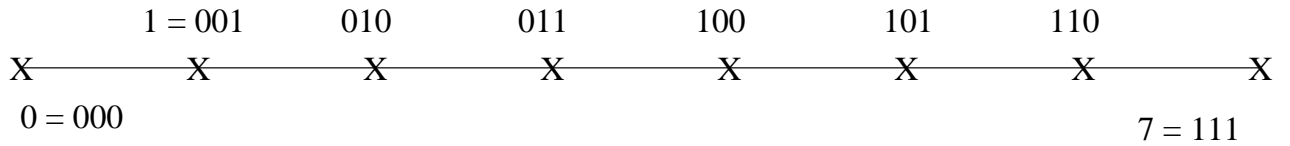


Figure 5-6: Signal constellation for 8-PAM. The X's are the signal points.

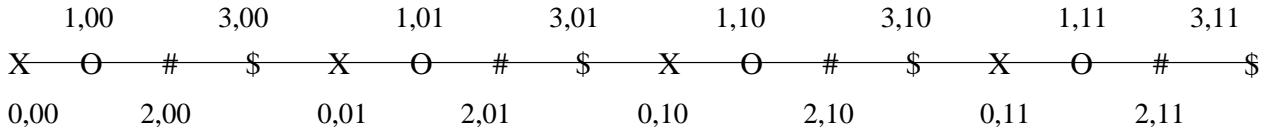


Figure 5-7: Signal constellation for TCM. The X's are the signal points for coset 0, the O's are the signal points for coset 1, the #'s are the signal points for coset 2, and the \$ are the signal points for coset 3.

to choosing a particular point in each coset. So if  $X_1^n$  is mapped to  $n(R + 1)$  bits,  $n$  bits are the coset choice bits and  $nR$  bits are the intra-coset choice bits. The tag bits are modulated into the least significant bits of the intra-coset bits. Putting the digital signature into the intra-coset bits makes calculating the tradeoff between the number of tag bits,  $t$ , and the encoding distortion straight forward.

### 5.2.3 Analysis of the Trellis Coded Case

The analysis of the probability of forgery is the same as the uncoded case. The level of security depends mostly on the tag length,  $t$ . The extra encoding distortion caused by embedding the tag will be four times greater than the corresponding amount in the uncoded case. This is because the signal points in a coset are twice as far apart as the corresponding points for a uniform scalar quantizer. Thus the processing distortion due to embedding  $t$  tag is increased by roughly  $(1 + \frac{15t}{4n})$ . For  $t \ll n$  this becomes negligible.

Rather than develop a closed form expression for the probability of false alarm error, we evaluate  $p_s$  numerically. The results of simulations for some reasonable design parameters are plotted in Figure 5-8. This plot shows the probability of false alarm error per symbol,  $p_s$ , as a function of the normalized distortion to noise ratio,  $LDNR_{norm}$ . The total probability of false alarm error,  $p_f$ , is  $n$  times  $p_s$ . Furthermore, in our plot we have assumed that the extra distortion due to embedding the tag bits is negligible.

For these experiments, the specific design parameters were based on an i.i.d source uniform over the range  $[-4, 4]$ , quantized using rate 4 TCQ. The rate 1/2 convolutional codes used for the coset selector are the 1-D codes in Table I of [23]. The results for the uncoded case were computed analytically while the results for the coded case were collected using  $10^4 - 10^6$  sequences of length 1000 (more sequences were used to collect the points with low  $p_s$ ). The number of trials were chosen so that all results are with  $\pm 10\%$  with 95% confidence. Since the source size and the rate were fixed

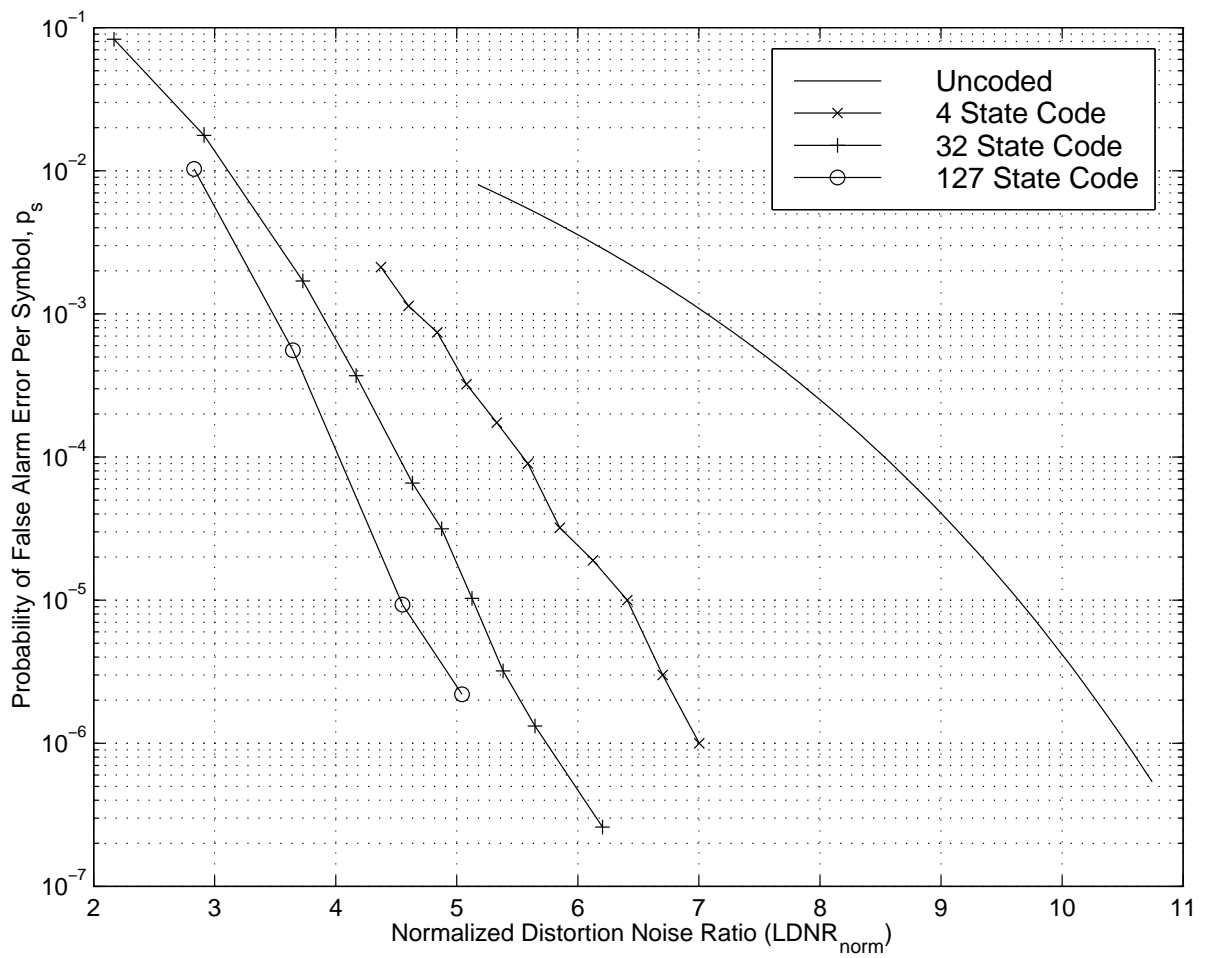


Figure 5-8: Coding gain achieved for TCM/TCQ.

for all trials, the SNR and DNR were varied by changing the noise variance.

The plot shows that TCM/TCQ achieves coding gains ranging from 3.5-5.5 dB at  $p_s = 10^{-6}$  for reasonable complexity codes. Even though the TCM/TCQ coding scheme achieves significant coding gains, a potentially large gap to the fundamental distortion still exists. In chapter 6 we discuss stronger coding schemes which can be used to reduce the gap.

## Chapter 6

# Conclusion

In this work we have formalized and studied the analog authentication. The goal of our work was to understand how images, video, speech, and multimedia can be authenticated in the presence of small perturbations. Our formulation of the problem makes it possible to compare the success of various schemes. In addition we were able to derive strong results on achievable and non-achievable schemes. In Appendix A we also analyze a type of broadcast channel which is similar to the analog authentication problem. Most of the results developed for analog authentication also apply to the backward compatibility broadcast channel.

We defined the fundamental distortion as

$$D^* = \min_{p(y|x): I(X;Y) - I(Y;Z) \leq 0} E[d(X, Y)]$$

and proved that no analog authentication scheme using processing distortion less than  $D^*$  is possible. Furthermore we showed all distortions above  $D^*$  are achievable.

In addition to proving theoretical results we designed and evaluated a practical analog authentication scheme for an i.i.d. uniform source sent over an AWGN channel using mean square error as the distortion metric. We first designed and analyzed uncoded schemes. We described how to modify this scheme to include coding. We then analyzed the benefits of using trellis coding. There are a number of ways to continue this work.

### Better Coding

We used TCM/TCQ to achieve significant coding gains. However, our scheme had a substantial gap to the fundamental distortion at  $p_s = 10^{-6}$ . We conjecture that by using more sophisticated coding techniques the efficiency of practical schemes can be substantially increased. Specifically, turbo decoding (also known as iterative decoding) and multilevel codes with multistage decoding would be excellent candidates for future research. Turbo decoding would probably provide the most

coding gain. Furthermore in most analog authentication applications decoding delay is not an issue, so the long decoding delay of turbo decoding is acceptable.

### **Prove A Strong Converse**

When we proved the converse for the coding theorem, we only showed that any schemes which have the probability of error going to 0 as  $n \rightarrow \infty$  must have distortion greater than  $D^*$ . This does not rule out the possibility of schemes having very small but finite probability of error for all  $n$ , or even schemes which have very small but finite probability as  $n \rightarrow \infty$ . We conjecture that our results can be extended to prove a strong converse which states that the probability of error has to increase exponentially with  $n$  as  $n \rightarrow \infty$  when the processing distortion is less than  $D^*$ .

### **Extend Theorems To Colored Noise, Colored Sources, or Colored Distortion Metrics**

We proved theorems for i.i.d. sources, channels with i.i.d noise, and single letter distortion metrics. Many practical settings could have colored sources or colored noise. For example, images are better represented as colored sources than as white sources. Also the effect of lossy compression is more accurately represented as colored noise than white noise. Finally, many researchers have suggested that human perception is influenced more by certain frequencies. This suggests that some kind of colored distortion metric might be more appropriate than a single letter distortion metric. We suspect that various water-filling solutions could be found in such cases.

### **Extend Theorems For Unknown Channel/Source Statistics (e.g. Universal Coding)**

One of the most fascinating areas of information theory is the study of universal schemes where the transmitter and/or the receiver does not know the source statistics or the channel law. Many exciting results have been proved for such cases in classical information theory which could be extended to the analog authentication or the backward compatibility broadcast channel.

### **Generalized Distortion Constraint**

Our problem has a constraint on  $E[d(X, Y)]$ . A more general approach would be to analyze what schemes can be achieved with the two constraints

$$E[d(X, W)] \leq D_1; \quad E[d(X, \hat{Y})] \leq D_2$$

for the scenario pictured in Figure 3-1. The first constraint is the expected distortion between the original source,  $X$ , and what the old receiver gets,  $W$ . The second constraint is between the original source,  $X$ , and what the new receiver decodes to,  $\hat{Y}$ . Note that in this case, it would be more correct to call the decoded result  $\hat{X}$  because the new user is trying to estimate  $X$  not  $Y$ . For this

scenario, the question now becomes what are the achievable pairs  $(D_1, D_2)$ ? This distortion region is analogous to the capacity region in multi-access communication. We suspect that the distortion region will be constrained by

$$D_2 \geq \min_{p(y|x): I(X;U) - I(U;Z) \leq 0} E[d(X, f(Z))] \quad (6.1)$$

$$D_1 \geq E[d(X, W)] \quad (6.2)$$

$$(6.3)$$

where  $U$  is some auxiliary random variable,  $f(\cdot)$  is some decoding function, and the expectations in (6.1) and (6.2) are taken with respect to the same conditional distribution  $p(y|x)$ . Note that in this formulation the term  $I(X;Y) - I(Y;Z)$  becomes  $I(X;U) - I(U;Z)$  which bears a strong resemblance to the capacity of the watermarking channel as discussed by Chen and Wornell [13].

Studying analog authentication in this context requires re-examining our notion of a forgery. In the new setting, the receiver would decode to  $\hat{X}$  which is an imperfect estimate of the original source,  $X$ , instead of decoding to  $\hat{Y}$  which is a perfect estimate of the transmitted signal. See Section 3.0.3 for a discussion of the advantages and disadvantages of such alternative formulations.

### Error Exponents

We obtained asymptotic results for what distortions are achievable. Studying the error exponents that could be obtained by various schemes using maximum likelihood decoding would complement these results and provide more guidance for designing practical schemes.

### Extend Theorems To Side Information

We could extend our results to the setting where a rate  $R_s$  side channel is available from the transmitter to the receiver. We conjecture that in this case the fundamental distortion would become

$$D^* = \min_{p(y|x): I(X;Y) - I(Y;Z) \leq R_s} E[d(X, Y)]$$

### Analyze Situations With Different Sources, Channels, Distortion Metrics

In our work we only analyzed a few scenarios chosen mostly for tractable computations. There might be many practical applications where the source statistics, channel law or distortion metrics are different. For example, we could examine the absolute distortion,  $d(X, Y) = |X - Y|$  instead of mean square distortion. Alternatively we could try to find distortion metrics which are more closely matched to the human perceptual system and design coding schemes based on such results.

### **Characterize Other Researchers Schemes Based On Theoretical Results**

Many other researchers have proposed analog authentication schemes. It would be interesting to examine these other schemes based on the theory we have developed. Characterizing what the best certain schemes can achieve and thereby measuring their efficiency would provide insight into the merits of these techniques.

### **Characterize Approximate Hash Functions And Robust Bit Extraction Using Similar ideas**

We suspect that an information theoretic analysis similar to the one we have provided could also be applied to the ideas of approximate hash functions and robust bit extraction. In the analog authentication scheme we described, the transmitter processes the original signal,  $X$  to get  $Y$ . This processing adds distortion. One way to view this is that an error signal,  $\mathcal{T}(X)$  is added to  $X$  to get  $Y$ :  $Y = X + \mathcal{T}(X)$ . Instead of adding this error signal, the transmitter could compute the error signal and quantize it to get a bit stream  $\mathcal{B}(X) = Q(\mathcal{T}(X))$  and transmit the original signal,  $X$ , unmodified while also sending the bit stream  $\mathcal{B}(X)$ . The receiver could then use the received signal  $Z$ , and the bit stream to check authenticity.

In this scenario, the bit stream is effectively an approximate hash function or a robust bit extraction procedure. The resources allowed would be the rate of the side bit stream,  $R_{\mathcal{B}}$  from the transmitter to the receiver. The figures of merit would noise robustness, probability of forgery and probability of false alarm. We suspect that allowing larger side rates,  $R_{\mathcal{B}}$ , would correspond to more sensitive forgery detection. Concepts from a generalized distortion constraint as well as the results on common randomness capacity by Ahlswede and Csiszar [24] could be useful in such an analysis.

# Appendix A

## The Backward Compatibility Broadcast Channel

The full analog authentication problem has quite a few pieces, one of which corresponds to reliable communication over a kind of channel we call the backward compatibility broadcast channel (BCBC). We analyze this channel separately because it is interesting in its own right. In addition, analyzing the BCBC is useful for understanding how the different pieces of the analog authentication problem fit together.

Consider a transmitter that needs to communicate to two different kinds of receivers: a new receiver and an old receiver. This situation is similar to the traditional broadcast channel. In the traditional broadcast channel, there is a transmitter and two receivers. Each receiver sees a different channel from the transmitter and uses a separate decoder. In the backward compatibility broadcast channel the new receiver has a decoder, but the old receiver does not.

This models the problem of backward compatibility. A transmitter needs to transmit to the new receiver with the newest receiving equipment, but still allow the older receiver to receive a reasonable signal. The old receiver has no decoder and must accept the signal it receives, while the new receiver can use decoding. Using the tools of information theory we can analyze a simple version of the backward compatibility broadcast channel.

### A.1 Definitions

The parameters for the BCBC consist of three finite sets  $\mathcal{X}$ ,  $\mathcal{W}$ , and  $\mathcal{Z}$ , three probability distributions  $p(x)$ ,  $p(w|y)$ ,  $p(z|y)$ , and a bounded distortion function<sup>1</sup>,  $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbf{R}^+$ . The source sequence

---

<sup>1</sup>The bound on the distortion function is a technical condition. Our results can be extended to unbounded distortion functions in the same way the classic rate-distortion theorem can be extended to unbounded distortion functions.

$X^n$  is drawn i.i.d. according to the probability law  $p(x)$ . The source sequence is a random variable and not under the control of any entity in the channel. The transmitter encodes the source to get  $Y = Q(X)$  which is broadcast over the channel. The old receiver sees output  $W$  according to the channel law  $p(w|y)$  and the new receiver sees output  $Z$  according to the channel law  $p(z|y)$ . The new receiver then decodes  $Z$  to get an estimate of the transmitted sequence,  $\hat{Y}$ . Figure A-1 shows a diagram of this model.

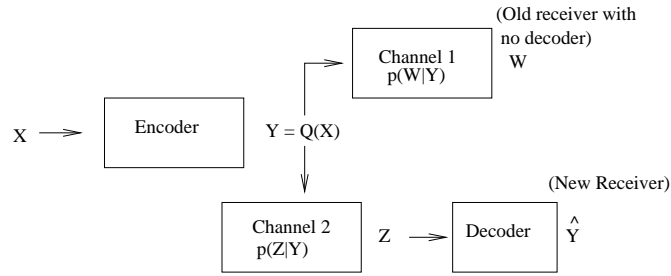


Figure A-1: A diagram of the backward compatibility broadcast channel.

A codebook,  $\mathcal{C}_n$ , for the channel consists of an encoder  $Q$  and a decoder  $g$ . The encoder  $Q : \mathcal{X}^n \rightarrow \mathcal{X}^n$  maps a source sequence,  $X^n$ , to an encoded sequence  $Y^n$ . The decoder  $g : \mathcal{Z} \rightarrow \mathcal{X}$  maps the output of the second channel to a decoded sequence  $\hat{Y}^n$ .

The probability of error,  $\lambda$ , is the probability that  $\hat{Y}^n$  is decoded incorrectly:

$$\lambda = \Pr(g(Z) \neq Y)$$

According to our definition of  $\lambda$ , the decoder tries to decode  $Y$  instead of  $X$ . We choose this measure for two reasons. First it leads to a simpler analysis. Second,  $X$  can always be estimated from  $Y$ . The Markov condition  $X \leftrightarrow Y \leftrightarrow Z$  implies that  $Y$  is a sufficient statistic for estimating  $X$ . Thus if  $Y$  can be estimated with low probability of error, estimating  $X$  by first estimating  $Y$  is optimal.

The encoding distortion,  $D$ , is the distortion between the source sequence,  $X^n$ , and the output of the encoder,  $Y^n$ , as measured according to the single letter distortion function  $d(\cdot, \cdot)$ . We extend the single letter distortion function to sequences in the natural way so that  $D = \frac{1}{n} \sum_{i=1}^n d(X_i, Y_i)$ .

An encoding distortion,  $D$ , is said to be achievable if there exists a sequence of codebooks,  $\mathcal{C}_n$ , such that the probability that the encoding distortion exceeds  $D$  or a decoding error occurs tends to zero. Formally an encoding distortion  $D$  is achievable if

$$\lim_{n \rightarrow \infty} \Pr(\{d(X^n, Y^n) > D\} \cup \{\hat{Y}^n \neq Y^n\}) = 0$$

The motivation for these definitions is that coding benefits the new receiver and hinders the old receiver. Since the old receiver has no decoder, it views the encoding as a nuisance. It receives  $W$  and views both the channel and the encoding as noise which reduce the quality of the received

signal. The goal of the new receiver is reliable transmission. It can use the codebook to overcome the noise in the channel. Thus the new receiver benefits from coding.

In the traditional broadcast channel the goals of the two receivers can be balanced differently yielding a capacity region. We could analyze the BCBC by defining a distortion region to balance the distortion for each receiver. However, at this time, we only consider the case where the new receiver achieves reliable transmission while minimizing the embedding distortion to satisfy the old receiver.

## A.2 Theorems For The Backward Compatibility Broadcast Channel

We define the fundamental distortion,  $D^*$ , as

$$D^* = \min_{p(y|x): I(X;Y) - I(Y;Z) \leq 0} E[d(X, Y)] \quad (\text{A.1})$$

We will attach an operational meaning to the fundamental distortion by showing that reliable communication can be accomplished if and only if the processing distortion is at least  $D^*$ . If the mutual information constraint in (A.1) can never be satisfied for any distribution  $p(y|x)$ , then we arbitrarily define  $D^* = \infty$  to indicate that reliable communication is impossible for every finite encoding distortion.

We will prove a coding theorem and a converse. The coding theorem states that all encoding distortions greater than  $D^*$  are achievable. The converse states that no encoding distortions below  $D^*$  are achievable. The coding theorem states that we can achieve reliable communication with finite encoding distortion. This is analogous to the classical channel coding theorem which states that reliable communication can be achieved with rate greater than 0. Thus reliable communication is possible while also satisfying the old receiver. The converse states that reliable communication comes at a price. If we want reliable communication, the old receiver must accept some minimum amount of encoding distortion.

### **Theorem 3** The Backward Compatibility Coding Theorem

*If  $D^*$  is defined according to equation (A.1), then all encoding distortions above  $D^*$  are achievable.*

The proof combines the ideas behind the classical channel coding theorem and the classical rate-distortion theorem. For any  $\epsilon > 0$ , we construct a random codebook with  $2^{nR}$  codewords chosen i.i.d. according to the distribution  $p(y) = \sum_x p(y|x)p(x)$ . For  $R > I(X;Y)$  and  $n$  large enough, the rate-distortion theorem implies the encoding distortion will be less than  $E[d(X, Y)] + \epsilon$ . Similarly, as long as  $R < I(Y;Z)$ , the classical capacity theorem implies the probability of error goes to 0 as  $n \rightarrow \infty$ . If  $I(X;Y) - I(Y;Z) < 0$ , then we can choose  $R$  such that it satisfies both requirements.

We first describe the details of our scheme and then analyze the probability of failure in detail. Choose any  $\epsilon > 0$  and choose  $p(y|x)$  to be the distortion minimizing distribution in the definition of  $D^*$ .

**Generating a Codebook:** Create a random codebook,  $\mathcal{C}_n$ , with  $2^{nR}$  elements chosen i.i.d. according to the probability law  $p(y) = \sum_x p(y|x)p(x)$ . The codebook is revealed to all interested parties.

**Encoding:** To encode a sequence  $X^n$ , find a sequence  $Y^n$  in the codebook which is strongly jointly typical with  $X^n$ . We write  $(X^n, Y^n) \in A_\epsilon^{*(n)}$  to indicate  $X^n$  and  $Y^n$  are strongly jointly  $\epsilon$ -typical. If there is more than one jointly typical sequence  $Y^n$ , pick the lexicographically least. If there is no strongly jointly typical sequence, then declare an error and set  $Y^n = X^n$ . The sequence  $Y^n$  is sent over the channel.

**Decoding:** The old receiver simply receives the output of channel 1 in Figure A-1 and does nothing further. The new receiver receives  $Z^n$  and decodes as follows. It searches the codebook for a sequence  $\hat{Y}^n$  which is strongly jointly typical with  $Z^n$ . If it finds more than one such sequence or no such sequence it declares an error. If it finds exactly one sequence  $\hat{Y}^n$  such that  $(\hat{Y}^n, Z^n) \in A_\epsilon^{*(n)}$ , then it chooses  $\hat{Y}^n$  as the decoded result.

For our scheme to succeed we need to show

$$\lim_{n \rightarrow \infty} \Pr(\{d(X^n, Y^n) > D\} \cup \{\hat{Y}^n \neq Y^n\}) = 0$$

According to the union bound it suffices to show

$$\lim_{n \rightarrow \infty} \Pr(d(X^n, Y^n) > D + \epsilon) = 0 \tag{A.2}$$

$$\lim_{n \rightarrow \infty} \Pr(\hat{Y}^n \neq Y^n) = 0 \tag{A.3}$$

We show each statement separately starting with the first.

If  $R > I(X; Y)$ , then equation (A.2) follows from the rate-distortion theorem. The encoding process for the BCBC corresponds to quantizing a source as in the rate-distortion theorem [12]. Note multiple versions of the rate-distortion theorem exist. Some versions prove that the expected distortion will be less than  $E[d(X, Y)] + \epsilon$  while others prove that the total probability that the distortion is greater than  $E[d(X, Y)] + \epsilon$  goes to 0. Equation (A.2) follows from the latter version.

If  $R < I(Y; Z)$  equation (A.3) follows from the classical channel coding theorem. To show this we rewrite the left hand side of (A.3) as

$$\lim_{n \rightarrow \infty} \sum_{y \in \mathcal{Y}} \Pr(\hat{Y}^n \neq Y^n | Y^n = y) \Pr(Y^n = y) \tag{A.4}$$

This is the probability of decoding error averaged over all the codewords. The codebook contains  $2^{nR}$  codewords chosen i.i.d. according to the distribution  $p(y) = \sum_x p(y|x)p(x)$ . By the symmetry of the encoding process, each codeword is equally likely to be transmitted. According to the classical capacity theorem, the probability of error averaged over the codebook goes to 0 as  $n \rightarrow \infty$  provided  $R < I(Y; Z)$  [12]. Therefore (A.4) goes to 0 and the statement is proved.

By definition,  $D^* < \infty$  implies  $I(X; Y) \leq I(Y; Z)$ . Without loss of generality assume that this holds with strict inequality. If the equation holds with equality,  $p(y|x)$  can be changed slightly to allow a very small difference of  $\delta$  between  $I(X; Y)$  and  $I(Y; Z)$ . By continuity of the distortion function, this slight change will not significantly change  $E[d(X, Y)]$ . Since  $I(X; Y) < I(Y; Z)$  we can choose  $R$  such that  $I(X; Y) < R < I(Y; Z)$  so that equations (A.2) and (A.3) both hold. ♠

We prove the converse in two pieces. First we prove that all achievable distortions are greater than

$$D_n^* = \lim_{n \rightarrow \infty} \min_{p(y^n|x^n): \frac{1}{n} \sum [I(X_i; Y_i) - I(Y_i; Z_i)] \leq 0} E[d(X^n, Y^n)] \quad (\text{A.5})$$

and then we single-letterize this result to prove that all achievable distortions are greater than  $D^*$ .

**Theorem 4** Converse To The Backward Compatibility Coding Theorem (part 1)

*If  $D_n^*$  is defined according to (A.5), then no encoding distortions below  $D_n^*$  are achievable.*

The intuition to the converse comes from the converse to the rate-distortion theorem and the converse to the channel coding theorem. Imagine we choose some distribution  $p(y^n|x^n)$  and try to make a random coding argument. From the converse to the rate-distortion theorem, we expect that the rate of our encoder must be above  $I(X; Y)$  to achieve distortion around  $E[d(X, Y)]$ . From the converse to the channel coding theorem we expect that to achieve reliable communication we need a rate below  $I(Y; Z)$ . This suggests that  $I(X; Y) < R < I(Y; Z)$  which implies that  $I(X; Y) < I(Y; Z)$  is a necessary condition.

Consider an encoding scheme that uses  $2^{nR}$  codewords. Define  $P_e$  as the probability that a decoding error occurs. Fano's inequality implies  $H(Y^n|Z) \leq 1 + nRP_e$ . Using this we show that if  $P_e$  goes to zero as  $n \rightarrow \infty$ , then  $R \leq \frac{1}{n} \sum_{i=1}^n I(Z_i; Y_i)$ .

$$nR = H(Y^n) = I(Y^n; Z^n) + H(Y^n|Z^n) \quad (\text{A.6})$$

$$\leq I(Y^n; Z^n) + 1 + nRP_e \quad (\text{A.7})$$

$$= H(Z^n) - H(Z^n|Y^n) + 1 + nRP_e \quad (\text{A.8})$$

$$= 1 + nRP_e + \sum_{i=1}^n [H(Z_i|Z_1^{i-1}) - H(Z_i|Y_1^n, Z_1^{i-1})] \quad (\text{A.9})$$

$$\leq 1 + nRP_e + \sum_{i=1}^n [H(Z_i) - H(Z_i|Y_i^n, Z_1, Z_2, \dots, Z_{i-1})] \quad (\text{A.10})$$

$$= 1 + nRP_e + \sum_{i=1}^n [H(Z_i) - H(Z_i|Y_i)] \quad (\text{A.11})$$

$$= 1 + nRP_e + \sum_{i=1}^n I(Z_i; Y_i) \quad (\text{A.12})$$

$$R \leq \frac{1}{n} + RP_e + \frac{1}{n} \sum_{i=1}^n I(Z_i; Y_i) \quad (\text{A.13})$$

$$R \leq \frac{1}{n} \sum_{i=1}^n I(Z_i; Y_i) \quad (\text{A.14})$$

$$(\text{A.15})$$

Reasoning:

Line	Justification
(A.6)	Expand mutual information into entropies
(A.7)	Applying Fano's Inequality
(A.8)	Expand mutual information into entropies
(A.9)	Entropy Chain Rule
(A.10)	Conditioning Reduces Entropy
(A.11)	The channel law is memoryless so $Y_{i-1} \leftrightarrow Y_i \leftrightarrow Z_i$ forms a Markov chain.
(A.12)	Expand mutual information into entropies
(A.13)	Divide both sides by $n$
(A.14)	As $n$ gets large, $1/n$ goes to 0 and $P_e$ goes to zero by assumption

Next we show that  $R \geq \frac{1}{n} \sum_{i=1}^n I(X_i; Y_i)$ .

$$nR \geq H(Y^n) \tag{A.16}$$

$$= H(Y^n) - H(Y^n|X^n) \tag{A.17}$$

$$= I(Y^n; X^n) \tag{A.18}$$

$$= H(X^n) - H(X^n|Y^n) \tag{A.19}$$

$$= \sum_{i=1}^n H(X_i) - H(X^n|Y^n) \tag{A.20}$$

$$= \sum_{i=1}^n H(X_i) - \sum_{i=1}^n H(X_i|Y^n, X_1, X_2, \dots, X_{i-1}) \tag{A.21}$$

$$\geq \sum_{i=1}^n H(X_i) - \sum_{i=1}^n H(X_i|Y_i) \tag{A.22}$$

$$= \sum_{i=1}^n I(X_i; Y_i) \tag{A.23}$$

$$R \geq \frac{1}{n} \sum_{i=1}^n I(X_i; Y_i) \tag{A.24}$$

$$\tag{A.25}$$

Reasoning:

Line	Justification
(A.16)	Because there are at most $2^{nR}$ codewords
(A.17)	Because $Y^n$ is a function of $X^n$ so $H(Y^n X^n) = 0$
(A.18)	Expand mutual information into entropies
(A.19)	Expand mutual information into entropies
(A.20)	By assumption the source is i.i.d. so $H(X^n) = \sum_{i=1}^n H(X_i)$
(A.21)	Expand $H(X^n Y^n)$ using entropy chain rule
(A.22)	Conditioning reduces entropy
(A.23)	Expand mutual information into entropies
(A.24)	Divide both sides by $n$

We established an upper bound on  $R$  based on Fano's inequality and we established a lower bound on  $R$  based on various information inequalities. Combining these results yields

$$\frac{1}{n} \sum_{i=1}^n I(X_i; Y_i) \leq R \leq \frac{1}{n} \sum_{i=1}^n I(Y_i; Z_i)$$

which implies

$$\frac{1}{n} \sum_{i=1}^n I(X_i; Y_i) - \frac{1}{n} \sum_{i=1}^n I(Y_i; Z_i) \leq 0 \quad (\text{A.26})$$

for all achievable distortions.

Next we derive a lower bound for all achievable distortions,  $D$ . Define

$$f_D(t) = \Pr[t \leq d(X^n, Y^n) \leq t + dt]$$

then

$$E[d(X^n, Y^n)] = \int_0^\infty t f_D(t) dt \quad (\text{A.27})$$

$$= \int_0^D t f_D(t) dt + \int_D^\infty t f_D(t) dt \quad (\text{A.28})$$

$$\leq \int_0^D D f_D(t) dt + \int_D^\infty d_{max} f_D(t) dt \quad (\text{A.29})$$

$$\leq D + d_{max} \Pr[d(X^n, Y^n) > D] \quad (\text{A.30})$$

$$\lim_{n \rightarrow \infty} E[d(X^n, Y^n)] \leq \lim_{n \rightarrow \infty} D + d_{max} \Pr[d(X^n, Y^n) > D] \quad (\text{A.31})$$

$$= D + 0 \cdot d_{max} \quad (\text{A.32})$$

$$D \geq \lim_{n \rightarrow \infty} E[d(X^n, Y^n)] \quad (\text{A.33})$$

$$D \geq \lim_{n \rightarrow \infty} \min_{p(y^n | x^n): \frac{1}{n} \sum [I(X_i; Y_i) - I(Y_i; Z_i)] \leq 0} E[d(X^n, Y^n)] \quad (\text{A.34})$$

Reasoning:

Line	Justification
(A.27)	By definition of expected value
(A.28)	Split the integral on the previous line
(A.29)	$t \leq D$ in the first integral and $d(\cdot, \cdot) \leq d_{max}$ by assumption
(A.30)	By definition of $\Pr[d(X^n, Y^n) > D]$
(A.31)	Take the limit of both sides as $n \rightarrow \infty$
(A.32)	Since $D$ is an achievable distortion $\lim_{n \rightarrow \infty} \Pr[d(X^n, Y^n) > D] = 0$
(A.33)	Swap the sides of the previous line
(A.34)	Minimizing over the constraint (A.26) which holds for all achievable distortions

♠

**Theorem 5** Converse To The Backward Compatibility Coding Theorem (part 2)

If  $D^*$  is defined according to (A.1), then no encoding distortions below  $D^*$  in equation (A.1) are achievable.

Using convexity arguments we single-letterize the bound in equation (A.34) obtained in theo-

rem 4. Define

$$\mathcal{F}[p(y|x)] = I(X; Y) - I(Y; Z) \quad (\text{A.35})$$

$$\mathcal{F}_n[p(y^n|x^n)] = \frac{1}{n} \sum_{i=1}^n \mathcal{F}[p_i(y_i|x_i)] = \frac{1}{n} \sum_{i=1}^n I(X_i; Y_i) - I(Y_i; Z_i) \quad (\text{A.36})$$

$$D[p(y|x)] = E[d(X, Y)] \quad (\text{A.37})$$

$$D_n[p(y^n|x^n)] = \frac{1}{n} \sum_{i=1}^n D[p_i(y_i|x_i)] = E[d(X^n, Y^n)] \quad (\text{A.38})$$

By the properties of mutual information,  $I(X_i; Y_i)$  is convex in  $p(y_i|x_i)$  and  $I(Y_i; Z_i)$  is concave in  $p(y_i) = \sum_{x_i} p(y_i|x_i)p(x_i)$  [12]. Combining these implies that  $I(X_i; Y_i) - I(Y_i; Z_i)$  is convex in  $p(y_i|x_i)$ . This shows that  $\mathcal{F}$  is a convex functional. Since  $\mathcal{F}_n$  is a sum of convex functionals it is also a convex functional. Since expectations are linear, both  $D[p(y|x)]$  and  $D_n[p(y^n|x^n)]$  are linear functionals. Note that  $D_n^*$  is the minimum of  $D_n[p(y^n|x^n)]$  subject to the constraint  $\mathcal{F}_n[p(y^n|x^n)] \leq 0$  and  $D^*$  is the minimum of  $D[p(y|x)]$  subject to the constraint  $\mathcal{F}[p(y|x)] \leq 0$ .

Assume  $p^*(y^n|x^n)$  is a distribution which minimizes  $D_n[p(y^n|x^n)]$  subject to the appropriate constraint. Note that  $D_n[\cdot]$  depends only on the marginals,  $p_i^*(y_i|x_i)$ . If we define  $q(x)$  as the average over the marginals:

$$q(y|x) = \sum_{i=1}^n p_i^*(y_i|x_i)$$

then

$$\mathcal{F}[q(y|x)] \leq \frac{1}{n} \sum_{i=1}^n \mathcal{F}[p_i^*(y_i|x_i)] \quad (\text{A.39})$$

$$\mathcal{F}[q(y|x)] \leq \mathcal{F}_n[p^*(y^n|x^n)] \quad (\text{A.40})$$

$$D[q(y|x)] = \frac{1}{n} \sum_{i=1}^n D[p_i^*(y_i|x_i)] \quad (\text{A.41})$$

$$D[q(y|x)] = D_n[p^*(y^n|x^n)] \quad (\text{A.42})$$

$$D^* \leq D[q(y|x)] \quad (\text{A.43})$$

$$D^* \leq D_n^* \quad (\text{A.44})$$

$$D^* \leq D \quad (\text{A.45})$$

Reasoning:

Line	Justification
(A.39)	$\mathcal{F}[\cdot]$ is convex
(A.40)	By definition of $\mathcal{F}_n[\cdot]$ in (A.36)
(A.41)	$D[\cdot]$ is a linear functional
(A.42)	By definition of $D_n[\cdot]$ in (A.38)
(A.43)	$D[q(y x)]$ satisfies the constraint so the minimum must be at most $D[q(y x)]$
(A.44)	Combining previous two lines
(A.45)	Combining previous line with (A.34) where $D$ is any achievable distortion



### Summary

We have presented and analyzed a new scenario called the backward compatibility broadcast channel. This scenario models a transmitter sending a signal to different kinds of receivers. The old receivers lack decoders. The new receivers have decoders and can therefore take advantage of coding done at the transmitter. The transmitters goal is to achieve reliable communication to the new receivers while minimizing the distortion to the old receivers. We have shown that no scheme can achieve this goal if the processing distortion is less than  $D^*$ , and schemes exist which achieve this goal for any  $D > D^*$ .

# Bibliography

- [1] W. Diffie; M. E. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, 67:644–654, November 1976.
- [2] R. L. Rivest; A. Shamir; L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21:120–126, February 1978.
- [3] D. Kundur; D. Hatzinakos. Digital watermarking for telltale tamper proofing and authentication. In *Proceedings of the IEEE*, volume 87, pages 1167–1180. IEEE, July 1999.
- [4] Ping Wah Wong. A public key watermark for image verification and authentication. In *ICIP 98*, volume 1, pages 445–459. International Conference On Image Processing, 1998.
- [5] Min Wu; Bede Liu. Watermarking for image authentication. In *ICIP 98*, volume 2, pages 437–441. International Conference On Image Processing, 1998.
- [6] S. Bhattacharjee; M. Kutter. Compression tolerant image authentication. In *ICIP 98*, volume 1, pages 435–439. International Conference On Image Processing, 1998.
- [7] M. P. Queluz. Towards robust, content based techniques for image authentication. In *Multimedia Signal Processing*, pages 297–302. IEEE Second Workshop on Multimedia Signal Processing, 1998.
- [8] M. Schneider; S. Chang. A robust content based digital signature for image authentication. In *ICIP 96*, volume 3, pages 227–230. International Conference On Image Processing, 1996.
- [9] Jiri Fridrich. Methods for tamper detection in digital images. *Proceedings of Multimedia and Security Workshop at ACM Multimedia*, 1999.
- [10] Jiri Fridrich. Robust bit extraction from images. *ICMCS'99*, June 1999.
- [11] Richard F. Graveman; Kevin E. Fu. Approximate message authentication codes. *Army Research Labs ATIRP Conference*, February 1999.
- [12] Cover M. Thomas; Joy A. Thomas. *Elements of Information Theory*. John Wiley and Sons, Inc., 1991.

- [13] Brian Chen; Greg W. Wornell. Preprocessed and postprocessed quantization index modulation methods for digital watermarking (preprint). In *Proc. of SPIE: Security and Watermarking of Multimedia Contents II (part of Electronic Imaging 2000)*, 2000.
- [14] L. Lamport. Constructing digital signatures from a one-way function. In *SRI Intl. CSL-98*, October 1979.
- [15] S. Arimoto. An algorithm for calculating the capacity of an arbitrary discrete memoryless channel. *IEEE Transactions on Information Theory*, IT-18:14–20, 1972.
- [16] R. Blahut. Computation of channel capacity and rate distortion functions. *IEEE Transactions on Information Theory*, IT-18:460–473, 1972.
- [17] T. Berger. *Rate Distortion Theory: A Mathematical Basis For Data Compression*. Prentice-Hall, Englewood Cliffs, NJ, 1971.
- [18] M.D. Swanson; M. Kobayashi; A.H. Tewfik. Multimedia data-embedding and watermarking technologies. In *Proceedings of the IEEE*, volume 86, pages 1064–1087, June 1998.
- [19] Brian Chen; Greg W. Wornell. Digital watermarking and information embedding using dither modulation. In *Multimedia Signal Processing*, pages 273–278. IEEE Second Workshop on Multimedia Signal Processing, 1998.
- [20] Michael W. Marcellin; Thomas R. Fischer. Trellis coded quantization of memoryless and gaussian sources. *IEEE Transactions on Communications*, 38(1):82–93, January 1990.
- [21] G. David Forney Jr. Coset codes – part 1: Introduction and geometrical classification. *IEEE Transactions on Information Theory*, 34(5):1123–1151, September 1988.
- [22] G. Ungerboeck. Trellis-coded modulation with redundant signal sets part 1: Introduction. *IEEE Communications Magazine*, 25, February 1987.
- [23] G. Ungerboeck. Trellis-coded modulation with redundant signal sets part 2: State of the art. *IEEE Communications Magazine*, 25:12–21, February 1987.
- [24] Rudolf Ahlswede; Imre Csiszar. Common randomness in information theory and cryptography– part ii: Cr capacity. *IEEE Transactions on Information Theory*, 44(1):225–240, January 1998.
- [25] G. L. Friedman. The trustworthy digital camera: Restoring credibility to the photographic image. *IEEE Transactions on Consumer Electronics*, 39:905–910, November 1993.
- [26] G. Ungerboeck. Channel coding with multilevel/phase signals. *IEEE Transactions on Information Theory*, IT-28:55–67, January 1982.

- [27] Ingemar J. Cox; Jean-Paul M. G. Linnartz. Some general methods for tampering with watermarks. *IEEE Journal On Selected Areas in Communications*, 16(4):587–593, May 1998.
- [28] John G. Proakis; Masoud Salehi. *Communication Systems Engineering*. Prentice Hall, Englewood Cliffs, New Jersey 07632, 1994.
- [29] Nasir Memon; Ping Wah Wong. Protecting digital media content. *Communications of the ACM*, 41(7):35–42, July 1998.
- [30] Fred Mintzer; Gordon W. Braudaway; Alan E. Bell. Opportunities for watermarking standards. *Communications of the ACM*, 41(7):57–64, July 1998.
- [31] Joachim Hagenauer; Elke Offer; Lutz Papke. Iterative decoding of binary block and convolutional codes. *IEEE Transactions on Information Theory*, 42(2):429–445, March 1996.