

© 2013 Yen-Wei Huang

ASYMPTOTIC ANALYSIS FOR MULTI-USER CHANNELS

BY

YEN-WEI HUANG

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:

Professor Pierre Moulin, Chair
Professor Tamer Başar
Professor Bruce Hajek
Professor Venugopal V. Veeravalli
Associate Professor Olgica Milenkovic

ABSTRACT

This dissertation studies the asymptotics of two multi-user channel problems. The fingerprinting channel is associated with digital fingerprinting, which is an emerging technology to protect multimedia from unauthorized redistribution. The encoder embeds fingerprints into a host sequence and provides the decoder with the capability to trace back pirated copies to the colluders. The multiple access channel (MAC) is a classical problem in the field of network information theory. Multiple senders cooperate with one another to maximize their rates of communication to a single receiver. We address the problem of asymptotic analysis when the size of the problem goes to infinity.

The fundamental metric of measuring the detection capability of a fingerprinting system is capacity. It has recently been derived as the limit value of a sequence of maximin games with mutual information as their payoff functions. However, these games generally do not admit saddle-point solutions and are very hard to solve. Here under a modified version of the combined digit model proposed by Škorić et al., we reformulate the capacity as the value of a single two-person zero-sum game, and show that it is achieved by a saddle-point solution.

For fingerprinting capacity games with k pirates, we provide capacities along with optimal strategies for both players of the game when k is small. For large k , we show that capacity is asymptotic to A/k^2 where the constant A is specified as the maximin value of a continuous functional game. Saddle-point solutions to the game are obtained using methods of variational calculus.

For multiple access channels we study the maximum achievable rate region for a given blocklength n and a desired error probability ϵ . The inner region for the discrete memoryless MAC is approximated by a single-lettered expression $\mathbf{I} - \frac{1}{\sqrt{n}} \mathcal{Q}_{\text{inv}}(\mathbf{V}, \epsilon)$ where $\mathbf{I}$ is associated with the capacity pentagon bounds by Ahlswede and Liao, $\mathbf{V}$ is the MAC dispersion matrix, and $\mathcal{Q}_{\text{inv}}$ is

the inverse complementary multivariate Gaussian cumulative distribution region. For outer regions, we provide general converse bounds for both average error probability and maximal error probability criteria.

To my family

ACKNOWLEDGMENTS

It has been a long journey in quest of my Ph.D. degree, and this dissertation marks the end of it. I would like to express my hearty gratitude to all the people who have helped me during my Ph.D. study.

I would like to express my deepest gratitude to my adviser Prof. Pierre Moulin, for believing in me and providing me with the freedom to explore without any bounds. My graduate education experience has allowed me to grow in many different aspects. Without his support and trust in me, none of this would have been possible. I am very fortunate to have such a great adviser and partner to work with during the past six years.

I thank Professors Tamer Başar, Bruce Hajek, Venugopal V. Veeravalli, and Olgica Milenkovic for serving on my doctoral committee, taking time to read versions of the dissertation, and for their helpful suggestions and advice. A special mention is appropriate for Dr. Luddy Harrison for the guidance he provided during my first year in graduate school. I wish to thank Professors Alexander Barg, J. Nicholas Laneman, Boris Škorić, Vincent Tan, Doctors Ehsan Amiri, Teddy Furon, and Ebrahim MolavianJazi, for illuminating discussion and helpful comments.

I gratefully acknowledge the funding provided by the National Science Foundation under grants CCF 06-35137 and CCF 07-29061, and by DARPA under the ITMANET program.

Thanks are due to my research colleagues Maha El Choubassi, Shankar Sadasivam, Jean-Francois Jourdas, Ibtissam Ezzeddine, Scott Deeann Chen, Rohit Naini, Guillaume Gigaud, Honghai Yu, Ben Chidester, Patrick Johnstone, and Igor Fedorov, for enlightening discussion on various research problems and the good times we spent at the Beckman Institute.

Finally, I would like to thank my family, from whom I am incredibly fortunate to receive such ongoing love and support. My parents Ming-Lang and Ying-Jen, my sister Ching-Ying, my brother-in-law Wei-Chia, and my

nephew Chen-Hao have always been my safe haven under any circumstances. My wife Chieh-Chun provided me the inspiration and motivation as we experienced the thick and thin of graduate student life together. Even in times of crisis, I could always count on her to bring a smile to my face.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER 1 INTRODUCTION	1
1.1 Focus and Perspective	2
1.2 Dissertation Outline	4
1.3 Notation	5
CHAPTER 2 PRELIMINARIES ON FINGERPRINTING	7
2.1 Fingerprinting System	7
2.2 Randomized Fingerprinting Codes	8
2.3 Collusion Channel	10
2.4 Error Probabilities and Capacity	14
CHAPTER 3 MUTUAL INFORMATION GAMES OF FINGER- PRINTING CAPACITY	16
3.1 Mutual Information Game for Joint Decoder	16
3.2 Mutual Information Game for Simple Decoder	18
3.3 Two-Person Zero-Sum Games of Fingerprinting Capacity	19
CHAPTER 4 BINARY ALPHABET FINGERPRINTING CA- PACITY UNDER THE MARKING ASSUMPTION	23
4.1 Game Definition	23
4.2 Analysis of the Convex Games	27
4.3 Capacity Bounds	28
4.4 Asymptotics for Large Coalitions	33
4.5 Numerical Results for Beta Distributions	38
4.6 Why Are the Arcsine Distribution and the Interleaving At- tack Optimal for Large Coalitions?	39
CHAPTER 5 ARBITRARY ALPHABET FINGERPRINTING CA- PACITY UNDER THE COMBINED DIGIT MODEL	42
5.1 Fighting Small Coalitions	42
5.2 Asymptotics for Large Coalitions	44

CHAPTER 6	SECOND-ORDER CODING RATES FOR MUL-	
	TIPLE ACCESS CHANNELS	59
6.1	Definitions	59
6.2	Main Results	61
6.3	Inner Region	64
6.4	Outer Region	67
CHAPTER 7	SUMMARY AND FUTURE WORK	73
7.1	Digital Fingerprinting	73
7.2	Multiple Access Channel	74
APPENDIX A	PROOFS FOR FINGERPRINTING CAPACITY . .	75
A.1	Derivation of $C_{k,2}^{\text{simple}}(\mathcal{P}_W, \mathbf{p}^*)$	75
A.2	Proof of Theorem 5.6	76
A.3	Useful Lemmas for Change of Variable in Sec. 5.2.2	82
APPENDIX B	PROOFS FOR MULTIPLE ACCESS CHANNELS . .	88
REFERENCES		90

LIST OF TABLES

2.1	CDM Parameters for Variants of the Marking Assumption . .	14
5.1	Saddle-Point Solutions for the Joint Capacity Game: Coalition Size $k = 2$ and Alphabet Size $q = 2, 3$	45

LIST OF FIGURES

2.1	Fingerprinting system model	8
4.1	Capacities $C_{k,2}(\mathcal{P}_W, \mathcal{P}_{\text{RDM}})$, $C_{k,2}(\mathcal{P}_W, \mathbf{p}^*)$, $C_{k,2}(f_W^*, \mathcal{P}_{\text{RDM}})$, and upper and lower bounds of the (a) joint and (b) simple fingerprinting games	29
4.2	Cumulative distribution function of the arcsine distribution P_W^* and the optimal distributions $P_W^{(k,2)}$ for the (a) joint and (b) simple fingerprinting games under the marking assumption	29
4.3	The interleaving attack $\mathbf{p}^*$ and the optimal attacks $\mathbf{p}^{(k,2)}$ for the (a) joint and (b) simple fingerprinting games under the marking assumption	30
4.4	$g_{\text{opt}}(z/k)$ and minimizing collusion channels $\mathbf{p}^{(k,2)}$ for $k =$ 10, 20, and 30 and $P_W = \text{Beta}(\beta, \beta)$	38
4.5	$J_{\text{opt}}(w)$ and normalized payoff function $\tilde{I}_k(w, \mathbf{p}^{(k,2)})$ for $k =$ 10, 30, and 50 and $P_W = \text{Beta}(\beta, \beta)$	38
5.1	Fingerprinting capacities $C_{1,q}(\mathcal{P}_{\text{CDM}})$	43
6.1	The region $\mathcal{Q}_{\text{inv}}(\Sigma, \epsilon)$ for $\Sigma = \begin{bmatrix} 1 & .3 \\ .3 & 1 \end{bmatrix}$ and $\epsilon = .3$	62

CHAPTER 1

INTRODUCTION

The domain of network information theory studies the communication problem where multiple senders and receivers transmit signals in a system. The interactions among users in the system include interference, cooperation, feedback, etc. The general problem of what the fundamental limits of such transmission are has not yet been solved and it is one of the most active areas of research in the past decade in information theory. In this dissertation we consider the following two special problems:

1. **Digital Fingerprinting:** Distributors of digital contents (e.g. images, videos, audios, programs, etc.) embed a uniquely identified *fingerprint* into each distributed copy. Upon receiving a *forgery* generated by a coalition of users combining their copies, a fingerprint decoder analyzes traces of fingerprint left in the forgery and identifies the coalition.
2. **Multiple Access Channel:** This is a common model of multiple senders sending information to a single receiver. Examples include ground stations transmitting signals to a satellite receiver or cell phones communicating with a base station. The receiver decodes the received signal to messages from each sender.

Despite the apparent differences between the two problems, the digital fingerprinting channel is indeed a special case of the multiple access channel: the users in the coalition are the multiple senders, the fingerprint decoder is the single receiver, and the individual identities of the coalition are the message the decoder tries to decipher. The main difference between the two problems is that while in the general setup of multiple access channels the senders try to maximize the communication rates to the receiver, the senders (i.e. the coalition of users) in a fingerprinting system try to minimize the communication rates to the receiver to avoid detection.

We also focus on different senses of asymptotics for the two problems we study. For the digital fingerprinting problem, the major challenges are the uncertainty of the collusion channel, especially when the coalition size k is large. Therefore the asymptotic analysis focuses on characterizing fingerprinting capacity when k goes to infinity. For the multiple access channel problem, we assume the channel is known and the number of senders is fixed, and we attack the problem of characterizing the coding rate region as the blocklength n goes to infinity. In the rest of the chapter, we briefly review both problems and provide an outline of this dissertation.

1.1 Focus and Perspective

1.1.1 Fingerprinting Capacity

Before distribution of a copyrighted content, a fingerprinting system embeds a unique identification fingerprint into each copy, which can be extracted and helps trace unauthorized redistribution. A collusion attack is a powerful attack by a group of malicious users called *pirates* who combine their copies to generate a new version that contains only weak traces of their fingerprints. The goal of a fingerprinting decoder is to identify the pirates even under severe collusion attacks.

The challenges of designing an effective fingerprinting system are manifold. Firstly, it is difficult to construct a suitable model that describes the colluders' activities for different applications. Secondly, theoretical studies have shown that in some models, the maximum achievable rate, or *capacity*, decays quadratically with the size of the coalition k [1]. Thirdly, recent constructions on capacity-achieving fingerprinting codes require exponential computational complexity [2, 3] while efficient designs fall short in achieving capacity [4, 3].

The Boneh-Shaw marking assumption [5] is one of the most popular fingerprinting models in the literature. It assumes that pirates can only “cut-and-paste” their copies. This assumption, however, omits the possibility of pirates performing signal processing attacks which occasionally remove or replace a fingerprint symbol. The *combined digit model* (CDM) recently proposed by Škorić et al. [6] is a significant generalization of the Boneh-Shaw

marking assumption which successfully accounts for both types of attack. In attempting to bring the theory of fingerprinting one step closer to practical scenarios, we study fingerprinting capacity games under CDM in this work.

We show that fingerprinting capacity is the value of a two-person zero-sum game which admits a saddle-point solution. In the binary alphabet case under the marking assumption, new capacity bounds are provided in closed-form expressions. The ratio between the upper and lower bounds of the joint decoding scheme is $\pi^2/2$, while that of the simple decoding scheme is only $\pi^2/4$ (for large k). These bounds not only show that the binary fingerprinting capacity is in $\Theta(1/k^2)$, but they also provide secure strategies for both players of the game. Numerical solutions for small k are also presented in comparison with the bounds.

In Chap. 5 we show that the pirates can perform a stronger attack under CDM and capacity under CDM is about 20-30% less than that under the marking assumption. This suggests that the marking assumption is indeed too optimistic in addressing the colluders' capability of manipulating their fingerprint copies. The results also shed light on the structure of capacity-achieving codes, and on how the colluders can maximally utilize their copies to avoid detection. We then analyze the asymptotics of the game when the coalition size k is large, generalizing the asymptotic results for binary alphabets under the marking assumption [7]. Assuming a certain regularity condition, we show that capacity for q -ary alphabet against size- k coalition still decays quadratically with k under CDM and the constant in front is the value of a q -dimensional continuous-kernel game. Using the methods of variational calculus, a set of Euler-Lagrange differential equations, Lagrange multipliers, and boundary conditions can be explicitly specified as necessary and sufficient conditions. We derive the following results:

1. Binary alphabet with marking assumption: The asymptotic capacity is $(k^2 2 \ln 2)^{-1}$, the optimal embedding distribution is the arcsine distribution, and the optimal attack is the interleaving attack.
2. Binary alphabet with CDM: The asymptotic capacity is strictly smaller than $(k^2 2 \ln 2)^{-1}$ and the gap depends on the parameters of CDM. Somewhat surprisingly the optimal embedding is still the arcsine distribution, but the optimal attack is no longer the interleaving attack.

3. Arbitrary alphabet with marking assumption: The asymptotic capacity is $(q - 1)/(2k^2 \ln q)$, the optimal embedding distribution belongs to the family of Dirichlet distributions, and the optimal attack is the interleaving attack.

1.1.2 Multiple Access Channel

Shannon’s single-user channel coding problem and other classical channel problems give us an elegant formulation of capacities or capacity regions. The theorems state that the maximum rate of transmission approaches capacity or capacity region as the blocklength tends to infinity. However, such theorems do not provide convergence rates. This issue is addressed by asymptotic channel coding analysis conducted by Strassen [8] and others in the early 1960s and recently revisited by Polyanskiy et al. [9] and Hayashi [10]. Extension of the techniques therein to multiuser settings is quite challenging [11, 12, 7, 13, 14].

In this work, we study the second-order coding rates for MAC. The characterization of the average error capacity region of the discrete memoryless MAC (DM-MAC) is given independently by Ahlswede [15] and Liao [16]. Unlike the single-user channel, the maximal error capacity region is generally smaller than the average error capacity region [17] and its characterization remains an open problem. The strong converse results in [18, 19] further strengthened the notion of capacity region in the sense that no rate pairs outside the capacity region can be ϵ -achievable for any $\epsilon < 1$. Recent development of this line of work includes Tan and Kosut’s [11], MolavianJazi and Laneman’s [12], Scarlett et al.’s [20], and our [7] analyses on the finite blocklength capacity inner region for DM-MAC, MolavianJazi and Laneman’s [13] further extension of their results to Gaussian MAC, and Verdú’s [14] and our [7] general converses.

1.2 Dissertation Outline

The outline of this dissertation is as follows: In Chap. 2, we introduce our fingerprinting model and formally defined fingerprinting capacity. The capacity formulas derived in [3] are reviewed and reformulated in Chap. 3.

Chap. 4 and Chap. 5 are devoted to the binary alphabet case and the arbitrary alphabet case, respectively. The multiple access channel problem is studied in Chap. 6, and a brief summary is provided in Chap. 7.

1.3 Notation

We use capital letters to represent random variables, and lowercase letters to represent their realizations. Boldface denotes vectors, and calligraphic letters denote finite sets. For example, $\mathbf{X} \in \mathcal{X}^n$ denotes a random vector $(X_1, \dots, X_n)$, with each X_i taking values in $\mathcal{X}$. The probability distribution of $\mathbf{X}$ is characterized by its distribution function $P_{\mathbf{X}}(\mathbf{x}) \triangleq \Pr(X_1 \leq x_1, \dots, X_n \leq x_n)$. If the distribution is discrete, we also describe it by its probability mass function (pmf) $p_{\mathbf{X}}(\mathbf{x}) \triangleq \Pr(\mathbf{X} = \mathbf{x})$. Otherwise if $P_{\mathbf{X}}$ has the form

$$P_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\mathbf{X}}(\mathbf{x}) dx_1 \dots dx_n$$

then we characterize the distribution by its probability density function (pdf) $f_{\mathbf{X}}$. Mathematical expectation of a function $g(\mathbf{X})$ with respect to $P_{\mathbf{X}}$ is defined by

$$\mathbb{E}_{P_{\mathbf{X}}} [(g(\mathbf{X}))] \triangleq \int g(\mathbf{x}) dP(\mathbf{x}).$$

The mutual information of X and Y is denoted by $I(X; Y) = H(X) - H(X|Y)$. Should the dependency on the underlying pmf's be explicit, we write the pmf's as subscripts, e.g. $H_{p_X}(X)$ and $I_{p_X p_{Y|X}}(X; Y)$. Given a pair of sequences $(\mathbf{x}, \mathbf{y})$, we denote by $I(\mathbf{x}; \mathbf{y})$ the empirical mutual information of the joint pmf $p_{\mathbf{xy}}$. We also denote the binary entropy function by $h_q(p) = -p \log_q p - (1-p) \log_q (1-p)$ and $h(\mathbf{p}) = (h(p_1), \dots, h(p_n))'$. The base- q Kullback-Leibler divergence between two probability vectors $\mathbf{p}$ and $\mathbf{q}$ is denoted by $D_q(\mathbf{p} \parallel \mathbf{q})$, and the base-2 Kullback-Leibler divergence between two Bernoulli random variables with respective expectations p and q is denoted by $d_2(p \parallel q) \triangleq p \log_2 \frac{p}{q} + (1-p) \log_2 \frac{1-p}{1-q}$, where $\log_q$ denotes base q logarithm and $\ln$ denotes natural logarithm throughout the dissertation.

Sequences are denoted by $\langle \cdot \rangle$. The size or cardinality of a finite set $\mathcal{A}$ is denoted by $|\mathcal{A}|$. The indicator function of a subset $\mathcal{A}$ of a set $\mathcal{X}$ is a function

$\mathbb{1}_{\mathcal{A}} : \mathcal{X} \rightarrow \{0, 1\}$ defined as

$$\mathbb{1}_{\mathcal{A}}(x) = \begin{cases} 1 & \text{if } x \in \mathcal{A} \\ 0 & \text{if } x \notin \mathcal{A} \end{cases}.$$

The power set of a finite set $\mathcal{X}$, denoted by $2^{\mathcal{X}}$, is the set of all subsets of $\mathcal{X}$, including the empty set and $\mathcal{X}$ itself. The support of a probability distribution P , denoted by $\text{supp}(P)$, is the smallest set whose complement has probability zero. The support of a family $\mathcal{P}$ of probability distributions, denoted by $\text{supp}(\mathcal{P})$, is the union of the support of each probability distribution in the family, i.e., $\bigcup_{P \in \mathcal{P}} \text{supp}(P)$.

Asymptotic notations are defined as follows: Suppose $f(k)$ and $g(k)$ are two functions defined on positive real numbers. We say $f(k) = O(g(k))$ if $\exists c_1 > 0, k_1 > 0$ such that $f(k) \leq c_1 g(k), \forall k \geq k_1$. Also, $f(k) = \Omega(g(k))$ if $\exists c_2 > 0, k_2 > 0$ such that $f(k) \geq c_2 g(k), \forall k \geq k_2$. We write $f(k) = \Theta(g(k))$ if $f(k) = O(g(k))$ and $f(k) = \Omega(g(k))$. The expression $f(k) = o(g(k))$ or $f(k) = \omega(g(k))$ means that $f(k)/g(k)$ tends to 0 or ∞ respectively. The shorthand $f \sim g$, $f \gtrsim g$, and $f \lesssim g$ denote the asymptotic relations $\lim_{k \rightarrow \infty} \frac{f(k)}{g(k)} = 1$, $\liminf_{k \rightarrow \infty} \frac{f(k)}{g(k)} \geq 1$, and $\limsup_{k \rightarrow \infty} \frac{f(k)}{g(k)} \leq 1$ respectively.

CHAPTER 2

PRELIMINARIES ON FINGERPRINTING

2.1 Fingerprinting System

Let $\mathcal{X} = \mathcal{Q} = \{0, 1, \dots, q-1\}$ denote a size- q alphabet, $\mathcal{M} = \{1, \dots, M\}$ the user index set, and $\mathcal{W}_n$ a set of secret key values. An (n, M) randomized fingerprinting code $\mathcal{C}$ over $\mathcal{X}$, or *fingerprinting ensemble*, consists of an encoder-decoder pair (e_n, d_n) . The encoder

$$e_n : \mathcal{M} \times \mathcal{W}_n \rightarrow \mathcal{X}^n$$

assigns user m a length- n fingerprint $\mathbf{X}_m = e_n(m, W_n)$ where W_n is the random key. The mapping e_n is known to the public but the realization of W_n is kept secret from the pirates.

A *coalition* is formed by a subset of k users who are also called colluders. Without loss of generality, we denote the set of colluder indices by $\mathcal{K} = \{1, \dots, k\}$. The colluders observe fingerprint sequences $\mathbf{X}_{\mathcal{K}} = \{\mathbf{X}_1, \dots, \mathbf{X}_k\}$ and use them to produce a *forgery* $\mathbf{Y} \in \mathcal{Y}^n$. The process is modeled by passing $\mathbf{X}_{\mathcal{K}}$ through a *collusion channel* $p_{\mathbf{Y}|\mathbf{X}_{\mathcal{K}}}$, which is a conditional probability distribution on $\mathbf{Y}$ given $\mathbf{X}_{\mathcal{K}}$.

Based upon the received forgery $\mathbf{Y}$ and the secret key W_n , the decoder

$$d_n : \mathcal{Y}^n \times \mathcal{W}_n \rightarrow 2^{\mathcal{M}}, \quad \hat{\mathcal{K}} = d_n(\mathbf{Y}, W_n)$$

accuses a subset of users $\hat{\mathcal{K}}$. Two error criteria are often used: Under the detect-all criterion an error occurs when $\hat{\mathcal{K}}$ is not equal to $\mathcal{K}$, and under the detect-one criterion an error occurs when $\hat{\mathcal{K}}$ is empty or contains some innocent user $m \notin \mathcal{K}$. Our complete model for the fingerprinting system is shown in Fig. 2.1.

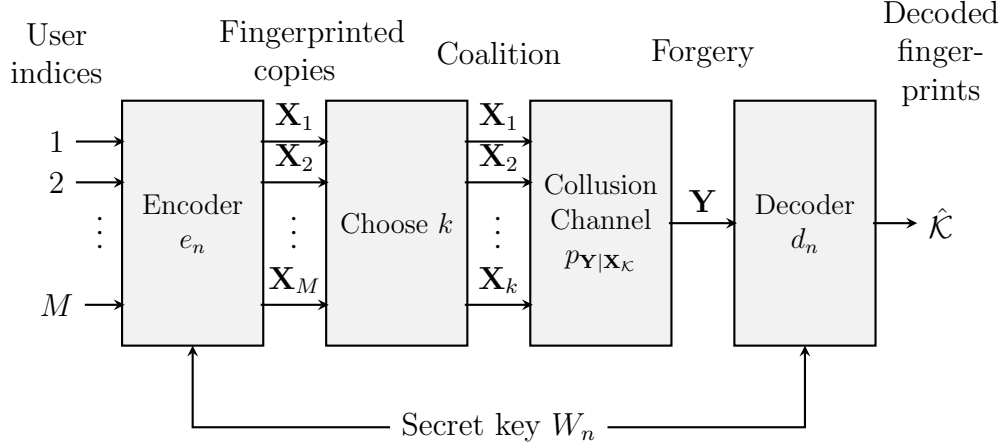


Figure 2.1: Fingerprinting system model

2.2 Randomized Fingerprinting Codes

The formal definition of an ensemble of fingerprinting codes is as follows.

Definition 2.1. A fingerprinting ensemble (E_n, D_n) is formed by the fingerprinting embedder randomly choosing from a family $\{e_n(\cdot, w_n), d_n(\cdot, w_n), w_n \in \mathcal{W}_n\}$ of (n, M) fingerprinting codes according to some probability distribution on the set $\mathcal{W}_n$ of keys.

We assume that the family of fingerprinting codes and the probability distribution on $\mathcal{W}_n$ are known to the public, but the realization w_n is only known to the encoder and the decoder.

As shown in [3], it suffices to consider the following two-phase fingerprinting construction and joint/simple decoding scheme in studying capacity. The secret key W_n shared by the encoder and the decoder in this scheme is the set of random variables $\{\mathbf{W}_i\}_{i=1}^n \cup (X_{m,i})_{M \times n}$.

2.2.1 Encoding Scheme

Let $P_{\mathbf{W}}$ be a probability distribution on the $(q-1)$ -dimensional simplex $\mathcal{W} \triangleq \{\mathbf{w} \in \mathbb{R}^q : \|\mathbf{w}\|_1 = 1, \mathbf{w} \geq \mathbf{0}\}$. A sequence of auxiliary “time-sharing” random variables $\{\mathbf{W}_i\}_{i=1}^n$ is drawn independent and identically from the distribution $P_{\mathbf{W}}$. For each $j \in \{1, \dots, n\}$, $\{X_{i,j}\}_{i=1}^m$ are m independent and

identically distributed random variables constructed from a categorical distribution¹ with parameter $\mathbf{W}_i$, i.e.,

$$Pr(X_{1,i} = x_1, \dots, X_{M,i} = x_M | \mathbf{W}_i = \mathbf{w}) = \prod_{m=1}^M w_{x_m} \quad (2.1)$$

for $x_1, \dots, x_m \in \mathcal{X}$. In general, there is no constraint on the choice of the embedding distribution $P_{\mathbf{W}}$, which means that we can choose it from the class of all probability distributions on $\mathcal{W}$, denoted by $\mathcal{P}_{\mathbf{W}}$. However, we may want to limit $P_{\mathbf{W}}$ to a subclass $\mathcal{P}^e$ of $\mathcal{P}_{\mathbf{W}}$ in some applications. For instance, Nuida et al. [21] limited $P_{\mathbf{W}}$ to be discrete with a finite spectrum. Furon and Perez-Freire [22] studied the case when $P_{\mathbf{W}}$ is the arcsine distribution (defined in (4.3) or the uniform distribution over the unit interval for binary fingerprinting codes, in which $\mathcal{P}^e$ is just a singleton.

In most of our results we require $\mathcal{P}^e$ to be compact. In some results we also require the following condition. Note that $\mathcal{P}^e$ satisfying (2.2) is compact.

Condition 2.2. *$\mathcal{P}^e$ coincides with the class of all probability distributions on $\text{supp}(\mathcal{P}^e)$, i.e.,*

$$\mathcal{P}^e = \{P_{\mathbf{W}} \in \mathcal{P}_{\mathbf{W}} : \text{supp}(P_{\mathbf{W}}) \subseteq \text{supp}(\mathcal{P}^e)\}. \quad (2.2)$$

Analogous to the symbol-symmetric fingerprinting codes proposed by Škorić et al. [23], it is intuitively reasonable to adopt a probability distribution $P_{\mathbf{W}}$ that is invariant to permutations of the symbols. Formally, let π be a permutation of $\mathcal{X}$ and define

$$P_{\mathbf{W}}^{\pi}(w_0, \dots, w_{q-1}) \triangleq P_{\mathbf{W}}(w_{\pi(0)}, \dots, w_{\pi(q-1)}). \quad (2.3)$$

Then we have

Definition 2.3. *An embedding distribution $P_{\mathbf{W}}$ is symbol-symmetric if*

$$P_{\mathbf{W}}^{\pi} = P_{\mathbf{W}}, \quad \forall \pi. \quad (2.4)$$

¹The categorical distribution is a special case of the multinomial distribution with the number of trials set to 1.

Definition 2.4. A subset $\mathcal{P}^e$ of $\mathcal{P}_{\mathbf{W}}$ is said to be symbol-symmetric if

$$P_{\mathbf{W}} \in \mathcal{P}^e \Rightarrow P_{\mathbf{W}}^{\pi} \in \mathcal{P}^e, \quad \forall \pi. \quad (2.5)$$

Definition 2.5. The symbol-symmetric subclass of an embedding class $\mathcal{P}^e$ is defined by

$$\mathcal{P}_{\text{sym}}^e = \{P_{\mathbf{W}} \in \mathcal{P}^e : P_{\mathbf{W}} \text{ is symbol-symmetric}\}. \quad (2.6)$$

The optimality of limiting $P_{\mathbf{W}}$ to $\mathcal{P}_{\text{sym}}^e$ will be discussed in the next section.

2.2.2 Decoding Scheme

We briefly review Moulin's two decoding schemes proposed in [3]: the simple decoder tests candidate fingerprints one by one, while the joint decoder utilizes a joint decoding rule. The simple decoder evaluates the empirical mutual information $I(\mathbf{x}_m; \mathbf{y}|\mathbf{w})$ for each user m . A threshold η^{simple} is chosen and user m is accused if and only if $I(\mathbf{x}_m; \mathbf{y}|\mathbf{w}) > \eta^{\text{simple}}$. If $I(\mathbf{x}_m; \mathbf{y}|\mathbf{w}) \leq \eta^{\text{simple}}$ for all $m \in \mathcal{M}$, then $\hat{\mathcal{K}} = \emptyset$. The joint decoder evaluates the following score for each coalition $\mathcal{A} \subseteq \mathcal{M}$:

$$S(\mathcal{A}) = \begin{cases} 0, & \text{if } \mathcal{A} = \emptyset \\ I(\mathbf{x}_{\mathcal{A}}; \mathbf{y}|\mathbf{w}) - |\mathcal{A}|\eta^{\text{joint}}, & \text{otherwise} \end{cases} \quad (2.7)$$

where η^{joint} is a threshold. The set $\mathcal{A}$ that has the largest score is then accused. With the parameters η^{simple} and η^{joint} , both decoders allow to tune the trade-offs between false positive and false negative error probabilities.

It is shown in [3] that the joint decoding scheme achieves capacity while the simple decoding scheme has a smaller maximum achievable rate. However, the computational complexity of a joint decoder is generally vastly greater than that of a simple decoder.

2.3 Collusion Channel

The set of admissible collusion channels determines the coalition's capability of manipulating the fingerprint sequences available to them. The Boneh-Shaw marking assumption and its nonbinary alphabet variants [5, 23] are the

most commonly adopted model in the literature. However, these models fail to completely capture plausible colluders' actions. Škorić et al. [6] recently proposed a more general setup called the combined digit model. However, CDM combines the collusion and the decoding process and is very hard to analyze. Here we introduce a modified version of CDM which is tractable and retains the spirit of the original model. We show how all variants of the marking assumption fit into our new model and how parameters can be determined based on applications.

We first introduce a number of properties of collusion channels. We consider a *memoryless* collusion channel $p_{\mathbf{Y}|\mathbf{X}_{\mathcal{K}}}$, i.e.,

$$p_{\mathbf{Y}|\mathbf{X}_{\mathcal{K}}}(\mathbf{y}|\mathbf{x}_{\mathcal{K}}) = \prod_{i=1}^n p_{Y|X_{\mathcal{K}}}(y_i|x_{\mathcal{K},i}),$$

which implies that for each coordinate i the forgery symbol y_i depends only on $x_{1,i}, \dots, x_{k,i}$. As exploited in [3], the memoryless restriction can be relaxed without changing the fingerprinting capacity. For simplicity we impose this constraint so the colluders' strategy can be characterized by the single-lettered channel $p_{Y|X_{\mathcal{K}}}$ and we will drop the coordinate index i in the following analysis.

Given a single-lettered collusion channel $p_{Y|X_{\mathcal{K}}}$, consider the user-permuted collusion channel

$$p_{Y|X_{\pi(\mathcal{K})}}(y|x_1, \dots, x_k) \triangleq p_{Y|X_{\mathcal{K}}}(y|x_{\pi(1)}, \dots, x_{\pi(k)}) \quad (2.8)$$

where π is a permutation of $\mathcal{K}$. We say that $p_{Y|X_{\mathcal{K}}}$ is *user-symmetric* if

$$p_{Y|X_{\pi(\mathcal{K})}} = p_{Y|X_{\mathcal{K}}}, \quad \forall \pi. \quad (2.9)$$

A class $\mathcal{P}^c$ of collusion channels is said to be user-symmetric if

$$p_{Y|X_{\mathcal{K}}} \in \mathcal{P}^c \Rightarrow p_{Y|X_{\pi(\mathcal{K})}} \in \mathcal{P}^c, \quad \forall \pi. \quad (2.10)$$

Note that in general not all elements of such $\mathcal{P}^c$ are user-symmetric. The user-symmetric subclass of a collusion class $\mathcal{P}^c$ that consists of user-

symmetric collusion channels is defined by

$$\mathcal{P}_{\text{fair}}^c = \{p_{Y|X_K} \in \mathcal{P}^c : p_{Y|X_K} \text{ is user-symmetric}\}. \quad (2.11)$$

Symbol-symmetry can also be defined for collusion channels. Let π be a permutation of $\mathcal{X}$ and define

$$p_{Y|X_K}^\pi(y|x_1, \dots, x_k) \triangleq p_{Y|X_K}(\pi(y)|\pi(x_1), \dots, \pi(x_k)). \quad (2.12)$$

Then we have

Definition 2.6. A collusion channel $p_{Y|X_K}$ is **symbol-symmetric** if

$$p_{Y|X_K}^\pi = p_{Y|X_K}, \quad \forall \pi. \quad (2.13)$$

Definition 2.7. A class $\mathcal{P}^c$ of collusion channels is said to be symbol-symmetric if

$$p_{Y|X_K} \in \mathcal{P}^c \Rightarrow p_{Y|X_K}^\pi \in \mathcal{P}^c, \quad \forall \pi. \quad (2.14)$$

Definition 2.8. The symbol-symmetric subclass of a collusion class $\mathcal{P}^c$ is defined by

$$\mathcal{P}_{\text{sym}}^c = \{p_{Y|X_K} \in \mathcal{P}^c : p_{Y|X_K} \text{ is symbol-symmetric}\}. \quad (2.15)$$

With the assumption of a memoryless collusion channel and user-symmetry², we let $\mathbf{Z} = [Z_0, Z_1, \dots, Z_{q-1}]' \in \{0, \dots, k\}^q$ with

$$Z_x = \sum_{m=1}^k \delta_{X_m x}, \quad x \in \mathcal{X}$$

which indicates how many times the k colluders see symbol x at a given coordinate. Let $\mathcal{Z} \triangleq \{\mathbf{z} \in \{0, \dots, k\}^q : \sum_{x=0}^{q-1} z_x = k\}$. We let $\mathcal{Y} = \mathcal{Q} \cup \{\text{e}\}$ where ‘e’ denotes an erasure. Based on the user-symmetry assumption, $\mathbf{Z}$ is a sufficient statistic for generating Y , and the colluders’ strategy can be modeled by the $(q+1) \times |\mathcal{Z}|$ matrix $\Theta \triangleq \{\theta(\mathbf{z})\}_{\mathbf{z} \in \mathcal{Z}}$ where

$$\theta_y(\mathbf{z}) \triangleq p_{Y|\mathbf{Z}}(y|\mathbf{z}), \quad y \in \mathcal{Y}, \mathbf{z} \in \mathcal{Z} \quad (2.16)$$

²We do not need to impose symbol-symmetry in our definition of collusion channels here.

is the conditional probability of selecting forgery symbol y given the count vector $\mathbf{z}$. Our modified CDM admits two parameter vectors: $\mathbf{r}_c = [r_{c,1}, \dots, r_{c,q-1}]'$ and $\mathbf{r}_e = [r_{e,1}, \dots, r_{e,q}]'$. The value $r_{c,N}$ specifies an upper bound on the probability of pirates generating an unseen symbol given they have seen N distinct symbols, and the value $r_{e,N}$ specifies an upper bound on the probability of pirates generating the erasure given they have seen N distinct symbols. Mathematically, a collusion channel Θ is admissible if and only if it satisfies the linear constraints

$$\sum_{y \in \mathcal{Q}, y \notin X_{\mathcal{K}}} \theta_y(\mathbf{z}) \leq r_{c,N(\mathbf{z})} \quad (2.17a)$$

$$\theta_e(\mathbf{z}) \leq r_{e,N(\mathbf{z})} \quad (2.17b)$$

for $\mathbf{z} \in \mathcal{Z}$ where $N(\mathbf{z}) \leq k$ denotes the number of distinct symbols seen by the pirates.

This is indeed a fairly general model as it reduces to the following four variants of the marking assumption in the literature with different choices of $\mathbf{r}_c$ and $\mathbf{r}_e$:

- **Restricted digit model (RDM):** RDM only allows the colluders to “mix and match” their copies of the content, i.e., Y can only be one of the symbols in $X_{\mathcal{K}}$.
- **Unreadable digit model (UDM):** UDM allows slightly stronger attacks. Besides the symbols they have, the colluders can also generate an erasure when $N(\mathbf{z}) > 1$.
- **Arbitrary digit model (ADM):** ADM does not allow erasures, but Y can be any symbol in $\mathcal{Q}$ when $N(\mathbf{z}) > 1$. In Sec. 5.1 we show that ADM allows stronger attacks than UDM.
- **General digit model (GDM):** GDM allows Y to be any symbol in $\mathcal{Q}$ or an erasure when $N(\mathbf{z}) > 1$. This allows the strongest attacks.

Table 2.1 summarizes the parameters corresponding to different models. It also shows realistic values of parameters based on simulations in [6]. Collusion classes are denoted by $\mathcal{P}_{\text{RDM}}$, $\mathcal{P}_{\text{UDM}}$, etc. for specific models or $\mathcal{P}^c$ in general. Collusion channels in our following analysis are represented by the $(q+1) \times t$ matrix $\Theta \triangleq [\theta(\mathbf{z}_1), \dots, \theta(\mathbf{z}_t)]$.

Table 2.1: CDM Parameters for Variants of the Marking Assumption

Model	Forgery alphabet	Parameters	
		$\mathbf{r}'_c$	$\mathbf{r}'_e$
RDM	$\mathcal{Q}$	$[0, 0, \dots, 0]$	$[0, 0, \dots, 0]$
UDM	$\mathcal{Q} \cup \{e\}$	$[0, 0, \dots, 0]$	$[0, 1, \dots, 1]$
ADM	$\mathcal{Q}$	$[0, 1, \dots, 1]$	$[0, 0, \dots, 0]$
GDM	$\mathcal{Q} \cup \{e\}$	$[0, 1, \dots, 1]$	$[0, 1, \dots, 1]$
CDM ₁	$\mathcal{Q} \cup \{e\}$	$[\text{.01}, \text{.01}, \dots, \text{.01}]$	$[\text{.20}, \text{.65}, \text{.84}, \text{.92}, \text{.92}, \dots]$
CDM ₂		$[\text{.05}, \text{.05}, \dots, \text{.05}]$	$[\text{.06}, \text{.40}, \text{.57}, \text{.72}, \text{.75}, \dots]$
CDM ₃		$[\text{.10}, \text{.10}, \dots, \text{.10}]$	$[\text{.03}, \text{.25}, \text{.44}, \text{.54}, \text{.63}, \dots]$

Note: our CDM is not the same as the original definition in [6], but one can still see the tradeoff between $\mathbf{r}_c$ and $\mathbf{r}_e$

2.4 Error Probabilities and Capacity

Under fingerprinting ensemble (E_n, D_n) , coalition size k , and collusion channel $p_{Y|X_K}$, we consider the following error probabilities:

- The probability of false positives (accusing an innocent user):

$$P_e^{\text{FP}}(E_n, D_n, p_{Y|X_K}) = \Pr(\hat{\mathcal{K}} \setminus \mathcal{K} \neq \emptyset). \quad (2.18)$$

- The probability of failing to catch any single pirate:

$$P_e^{\text{one}}(E_n, D_n, p_{Y|X_K}) = \Pr(\hat{\mathcal{K}} \cap \mathcal{K} = \emptyset). \quad (2.19)$$

- The probability of failing to catch the full coalition:

$$P_e^{\text{all}}(E_n, D_n, p_{Y|X_K}) = \Pr(\mathcal{K} \not\subseteq \hat{\mathcal{K}}). \quad (2.20)$$

The error probabilities above can be written explicitly as

$$P_e(E_n, D_n, p_{Y|X_K}) = \sum_{e_n, d_n} \Pr(e_n, d_n) \sum_{\mathbf{x}_{\mathcal{M}}, \mathbf{y}} \prod_{i=1}^n \int dP_{\mathbf{W}}(\mathbf{w}_i) \left(\prod_{m=1}^M w_{x_{m,i}} \right) p_{Y|X_K}(y_i | x_{\mathcal{K},i}) \mathbb{1}_{\mathcal{E}} \quad (2.21)$$

where the error event $\mathcal{E}$ is given by $\mathcal{E}^{\text{FP}} = \{d_n(\mathbf{y}, w_n) \setminus \mathcal{K} \neq \emptyset\}$, $\mathcal{E}^{\text{one}} = \{d_n(\mathbf{y}, w_n) \cap \mathcal{K} = \emptyset\}$, and $\mathcal{E}^{\text{all}} = \{\mathcal{K} \not\subseteq d_n(\mathbf{y}, w_n)\}$, when P_e is given by (2.18),

(2.19), and (2.20) respectively. The worst-case error probability for a collusion class is given by

$$P_{e,k}(E_n, D_n, \mathcal{P}^c) = \max_{\substack{\mathcal{K} \subseteq \mathcal{M} \\ |\mathcal{K}| \leq k}} \max_{p_{Y|X_{\mathcal{K}}} \in \mathcal{P}^c} P_e(E_n, D_n, p_{Y|X_{\mathcal{K}}}). \quad (2.22)$$

Having defined the error probabilities of the randomized fingerprinting scheme, we now define the notion of capacity.

Definition 2.9. *A rate R is achievable for embedding class $\mathcal{P}^e$, collusion channel $\mathcal{P}^c$, and size- k coalitions under the **detect-one** criterion if there exists a sequence of fingerprinting ensembles (E_n, D_n) generated by $P_{\mathbf{W}} \in \mathcal{P}^e$ for $M = \lceil 2^{nR} \rceil$ users such that both $P_{e,k}^{\text{FP}}(E_n, D_n, \mathcal{P}^c)$ and $P_{e,k}^{\text{one}}(E_n, D_n, \mathcal{P}^c)$ vanish as n tends to infinity.*

Definition 2.10. *A rate R is achievable for embedding class $\mathcal{P}^e$, collusion channel $\mathcal{P}^c$, and size- k coalitions under the **detect-all** criterion if there exists a sequence of fingerprinting ensembles (E_n, D_n) generated by $P_{\mathbf{W}} \in \mathcal{P}^e$ for $M = \lceil 2^{nR} \rceil$ users such that both $P_{e,k}^{\text{FP}}(E_n, D_n, \mathcal{P}^c)$ and $P_{e,k}^{\text{all}}(E_n, D_n, \mathcal{P}^c)$ vanish as n goes to infinity.*

Definition 2.11. *Fingerprinting capacities $C_{k,q}^{\text{one}}(\mathcal{P}^e, \mathcal{P}^c)$ and $C_{k,q}^{\text{all}}(\mathcal{P}^e, \mathcal{P}^c)$ are the suprema of all achievable rates with respect to the detect-one and detect-all criteria, respectively.*

Remark 2.1. *When the embedding class $\mathcal{P}^e$ is a singleton $\{P_{\mathbf{W}}\}$ or the collusion class $\mathcal{P}^c$ is a singleton $\{p_{Y|X_{\mathcal{K}}}\}$, we denote the corresponding capacities as $C_{k,q}(P_{\mathbf{W}}, \cdot)$ and $C_{k,q}(\cdot, p_{Y|X_{\mathcal{K}}})$ respectively, which is a slight abuse of notation.*

CHAPTER 3

MUTUAL INFORMATION GAMES OF FINGERPRINTING CAPACITY

In this chapter we first review the mutual information games associated with both the joint and the simple decoding schemes in [3]. We show how these games can be simplified under the combined digit model, and we show the existence of saddle-point solutions.

3.1 Mutual Information Game for Joint Decoder

To present the capacity formula, we first introduce the following setup: for a fixed embedding class $\mathcal{P}^e$, let

$$\mathcal{P}_l^e \triangleq \{p_{\mathbf{W}} \in \mathcal{P}^e : |\text{supp}(p_{\mathbf{W}})| \leq l\} \quad (3.1)$$

be the class of probability distributions with finite spectrum composed of no more than l points of the $(q-1)$ -dimensional simplex $\mathcal{W}$. A random variable $\mathbf{W}$ is drawn from some $p_{\mathbf{W}} \in \mathcal{P}_l^e$, and $\{X_m\}_{m=1}^k$ are independent and identically distributed with categorical distribution with parameter $\mathbf{W}$, i.e.,

$$p_{X_{\mathcal{K}}|\mathbf{W}}(x_{\mathcal{K}}|\mathbf{W}) = \prod_{m=1}^k p_{X_m|\mathbf{W}}(x_m|\mathbf{W}) \quad (3.2)$$

where

$$p_{X|\mathbf{W}}(x|\mathbf{W}) = w_x, \quad x \in \mathcal{X}.$$

The collusion class $\mathcal{P}^c$ is the set of all feasible channels $p_{Y|X_{\mathcal{K}}}$. Let

$$C_{k,q}^{\text{joint},l}(\mathcal{P}^e, \mathcal{P}^c) = \max_{p_{\mathbf{W}} \in \mathcal{P}_l^e} \min_{p_{Y|X_{\mathcal{K}}} \in \mathcal{P}^c} \frac{1}{k} I(X_{\mathcal{K}}; Y|\mathbf{W}). \quad (3.3)$$

The following theorem summarizes the main results of fingerprinting capacity proposed in [3] under the marking assumption.

Theorem 3.1. *Assume that Condition 2.2 is satisfied and $\mathcal{P}^c$ is compact. Then*

1. *The detect-all capacity is upper bounded by the detect-one capacity, i.e.,*

$$C_{k,q}^{\text{all}}(\mathcal{P}^e, \mathcal{P}^c) \leq C_{k,q}^{\text{one}}(\mathcal{P}^e, \mathcal{P}^c).$$

2. *Suppose further that $\mathcal{P}^c$ is user-symmetric, then*

$$C_{k,q}^{\text{one}}(\mathcal{P}^e, \mathcal{P}^c) = C_{k,q}^{\text{one}}(\mathcal{P}^e, \mathcal{P}_{\text{fair}}^c) = C_{k,q}^{\text{all}}(\mathcal{P}^e, \mathcal{P}_{\text{fair}}^c) = \lim_{l \rightarrow \infty} C_{k,q}^{\text{joint},l}(\mathcal{P}^e, \mathcal{P}^c). \quad (3.4)$$

3. *All rates below $C_{k,q}^{\text{one}}(\mathcal{P}^e, \mathcal{P}^c)$ are achievable by the joint decoding scheme in Sec. 2.2.2.*

Theorem 3.1 states that the detect-all capacity can never exceed the detect-one capacity, which is no surprise since it can only be harder for the decoder to detect all the pirates than to detect only one of the pirates. However, if the collusion channel is user-symmetric, which we can intuitively think of as the case when each colluder “contributes” the same number of samples to the forgery (hence the term “fair”), then the detect-one and the detect-all capacities are the same.

To distinguish capacity from the highest achievable rate achievable by the simple decoding scheme (discussed in the next section), we will in the rest of the dissertation refer to the detect-one capacity $C_{k,q}^{\text{one}}(\mathcal{P}^e, \mathcal{P}^c)$, denoted by $C_{k,q}^{\text{joint}}(\mathcal{P}^e, \mathcal{P}^c)$, as the joint fingerprinting capacity for embedding class $\mathcal{P}^e$ and collusion channel $\mathcal{P}^c$.

In the game-theoretic point of view, $C_{k,q}^{\text{joint},l}$ is the maximin value of a two-person zero-sum game for each l . Observe that the sequence $\langle C_{k,q}^{\text{joint},l} \rangle_{l=1}^{\infty}$ is nondecreasing since $\langle \mathcal{P}_l^e \rangle_{l=1}^{\infty}$ is nondecreasing (i.e. $\mathcal{P}_1^e \subseteq \mathcal{P}_2^e \subseteq \dots$). Thus the game can be interpreted as the following: the maximizer, the fingerprint embedder, picks $p_{\mathbf{w}}$ with an increasing flexibility in the support size, while the minimizer, the coalition, counters the embedder’s choice for each l by minimizing the mutual information payoff function. Fingerprinting capacity is the limit value of the sequence of maximin games.

However, the maximin game of (3.3) is in general very difficult to solve even for small values of l since a saddle-point solution cannot be guaranteed.

For the binary alphabet ($q = 2$) and $l = 1$ under RDM, we can derive the maximin value as

$$C_k^{\text{joint},1}(\mathcal{P}_{\mathbf{W}}, \mathcal{P}_{\text{RDM}}) = \frac{1}{k} 2^{-(k-1)} \quad (3.5)$$

which is not achieved by a saddle-point solution when $k > 2$. Also, this is a very loose lower bound on $C_{k,q}^{\text{joint}}(\mathcal{P}_{\mathbf{W}}, \mathcal{P}_{\text{RDM}})$ for large k comparing to the $\Theta(k^{-2})$ bound we will show in Sec. 4.3.

3.2 Mutual Information Game for Simple Decoder

As mentioned in Sec. 2.2, computationally joint decoding is too complex. Thus it is also interesting to study the maximum achievable rate of the simple decoding scheme.

Theorem 3.2. *[3] Assume that Condition 2.2 is satisfied and $\mathcal{P}^c$ is compact and user-symmetric. Let*

$$C_{k,q}^{\text{simple},l}(\mathcal{P}^e, \mathcal{P}^c) = \max_{\mathbf{P}_{\mathbf{W}} \in \mathcal{P}_l^e} \min_{p_{Y|X_{\mathcal{K}}} \in \mathcal{P}_{\text{fair}}^c} I(X_1; Y | \mathbf{W}) \quad (3.6)$$

for $l \geq 1$ and let

$$C_{k,q}^{\text{simple}}(\mathcal{P}^e, \mathcal{P}^c) = \lim_{l \rightarrow \infty} C_{k,q}^{\text{simple},l}(\mathcal{P}^e, \mathcal{P}^c). \quad (3.7)$$

Then all rates below $C_{k,q}^{\text{simple}}(\mathcal{P}^e, \mathcal{P}^c)$ are achievable by the simple decoding scheme for embedding class $\mathcal{P}^e$, collusion channel $\mathcal{P}^c$, and size- k coalitions under the detect-one criterion.

Corollary 3.3. *[3] For $\mathcal{P}^e$ satisfying Condition 2.2 and compact and user-symmetric $\mathcal{P}^c$, we have*

$$C_{k,q}^{\text{simple}}(\mathcal{P}^e, \mathcal{P}^c) \leq C_{k,q}^{\text{joint}}(\mathcal{P}^e, \mathcal{P}^c). \quad (3.8)$$

Although we do not have a notion of capacity for the quantity $C_{k,q}^{\text{simple}}$, it will be referred to as the “simple” fingerprinting capacity as opposed to the joint fingerprinting discussed in the previous subsection.

3.3 Two-Person Zero-Sum Games of Fingerprinting Capacity

To establish the desired saddle-point property, we first reformulate both the joint and the simple fingerprinting capacities as the respective values of the following two fingerprinting maximin games.

Theorem 3.4. *Assume that Condition 2.2 is satisfied and $\mathcal{P}^c$ is compact and user-symmetric. Then*

$$C_{k,q}^{\text{joint}}(\mathcal{P}^e, \mathcal{P}^c) = \max_{P_{\mathbf{W}} \in \mathcal{P}^e} \min_{p_{Y|X_{\mathcal{K}}} \in \mathcal{P}^c} \frac{1}{k} I(X_{\mathcal{K}}; Y | \mathbf{W}) \quad (3.9)$$

and

$$C_{k,q}^{\text{simple}}(\mathcal{P}^e, \mathcal{P}^c) = \max_{P_{\mathbf{W}} \in \mathcal{P}^e} \min_{p_{Y|X_{\mathcal{K}}} \in \mathcal{P}_{\text{fair}}^c} I(X_1; Y | \mathbf{W}). \quad (3.10)$$

Proof. Let

$$I_{k,q}^{\text{joint}}(\mathbf{w}, p_{Y|X_{\mathcal{K}}}) = \frac{1}{k} I_{p_{Y|X_{\mathcal{K}}}}(X_{\mathcal{K}}; Y | \mathbf{W} = \mathbf{w}) \quad (3.11)$$

and let

$$I_{k,q}^{\text{simple}}(\mathbf{w}, p_{Y|X_{\mathcal{K}}}) = I_{p_{Y|X_{\mathcal{K}}}}(X_1; Y | \mathbf{W} = \mathbf{w}). \quad (3.12)$$

Then the payoff functions of (3.9) and (3.10) become $\mathbb{E}_{P_{\mathbf{W}}} [I_k^{\text{joint}}(\mathbf{W}, p_{Y|X_{\mathcal{K}}})]$ and $\mathbb{E}_{P_{\mathbf{W}}} [I_{k,q}^{\text{simple}}(\mathbf{W}, p_{Y|X_{\mathcal{K}}})]$ respectively. Denote also the right-hand sides of (3.9) and (3.10) by $\tilde{C}_{k,q}^{\text{joint}}(\mathcal{P}^e, \mathcal{P}^c)$ and $\tilde{C}_{k,q}^{\text{simple}}(\mathcal{P}^e, \mathcal{P}^c)$ respectively. The inequality

$$C_{k,q}(\mathcal{P}^e, \mathcal{P}^c) \leq \tilde{C}_{k,q}(\mathcal{P}^e, \mathcal{P}^c) \quad (3.13)$$

follows directly from the fact that $\mathcal{P}_l^e \subseteq \mathcal{P}^e$ for each l . Now let the optimal achieving distributions for (3.9) or (3.10) be $P_{\mathbf{W}}^{(k,q)}$ and $p_{Y|X_{\mathcal{K}}}^{(k,q)}$. Then by completeness of $\mathcal{P}^e$, there exists a sequence of distributions $\langle p_{\mathbf{W}}^l \rangle_{l=1}^{\infty}$ with $p_{\mathbf{W}}^l \in \mathcal{P}_l^e$ that converges in distribution to $P_{\mathbf{W}}^{(k,q)}$. Both the functions $I_{k,q}^{\text{joint}}$ and $I_{k,q}^{\text{simple}}$ are bounded and continuous with respect to $\mathbf{w}$. By [24, p.249 Theorem 1] we have

$$\lim_{l \rightarrow \infty} \mathbb{E}_{p_{\mathbf{W}}^l} [I_{k,q}(\mathbf{W}, p_{Y|X_{\mathcal{K}}}^{(k,q)})] = \mathbb{E}_{P_{\mathbf{W}}^{(k,q)}} [I_{k,q}(\mathbf{W}, p_{Y|X_{\mathcal{K}}}^{(k,q)})] \quad (3.14)$$

and thus

$$C_{k,q}(\mathcal{P}^e, \mathcal{P}^c) \geq \lim_{l \rightarrow \infty} \mathbb{E}_{p_{\mathbf{W}}^l} \left[I_{k,q}(\mathbf{W}, p_{Y|X_{\mathcal{K}}}^{(k,q)}) \right] = \tilde{C}_{k,q}(\mathcal{P}^e, \mathcal{P}^c). \quad (3.15)$$

Combining (3.13) and (3.15) yields (3.9) and (3.10). $\square$

Theorem 3.4 shows that the joint and simple capacities are the maximin values of two single two-person zero-sum games. Note that the theorem only specifies the capacities when the embedding class $\mathcal{P}^e$ satisfies Condition 2.2. With slight modification of the proofs in [3], it can be shown that (3.9) and (3.10) still hold for any compact $\mathcal{P}^e$. Furthermore, we can show that the maximin and minimax values of the games are equal in general, and there are always saddle-point strategies for both players of the games.

We define the minimax values associated with the above games:

Definition 3.1. *The minimax value of the joint fingerprinting game is defined by*

$$\bar{C}_{k,q}^{\text{joint}}(\mathcal{P}^e, \mathcal{P}^c) = \min_{p_{Y|X_{\mathcal{K}}} \in \mathcal{P}^c} \max_{P_{\mathbf{W}} \in \mathcal{P}^e} \mathbb{E}_{P_{\mathbf{W}}} \left[I_{k,q}^{\text{joint}}(\mathbf{W}, p_{Y|X_{\mathcal{K}}}) \right]. \quad (3.16)$$

Definition 3.2. *The minimax value of the simple fingerprinting game is defined by*

$$\bar{C}_{k,q}^{\text{simple}}(\mathcal{P}^e, \mathcal{P}^c) = \min_{p_{Y|X_{\mathcal{K}}} \in \mathcal{P}_{\text{fair}}^c} \max_{P_{\mathbf{W}} \in \mathcal{P}^e} \mathbb{E}_{P_{\mathbf{W}}} \left[I_{k,q}^{\text{simple}}(\mathbf{W}, p_{Y|X_{\mathcal{K}}}) \right]. \quad (3.17)$$

When Condition 2.2 is satisfied (for example, when $\mathcal{P}^e = \mathcal{P}_{\mathbf{W}}$), the minimax games can be simplified as the following:

Lemma 3.5. *Assume that Condition 2.2 is satisfied. Then the minimax values of the joint and simple fingerprinting games can be respectively written as*

$$\bar{C}_{k,q}^{\text{joint}}(\mathcal{P}^e, \mathcal{P}^c) = \min_{p_{Y|X_{\mathcal{K}}} \in \mathcal{P}^c} \max_{\mathbf{w} \in \text{supp}(\mathcal{P}^e)} I_{k,q}^{\text{joint}}(\mathbf{w}, p_{Y|X_{\mathcal{K}}}) \quad (3.18)$$

and

$$\bar{C}_{k,q}^{\text{simple}}(\mathcal{P}^e, \mathcal{P}^c) = \min_{p_{Y|X_{\mathcal{K}}} \in \mathcal{P}_{\text{fair}}^c} \max_{\mathbf{w} \in \text{supp}(\mathcal{P}^e)} I_{k,q}^{\text{simple}}(\mathbf{w}, p_{Y|X_{\mathcal{K}}}). \quad (3.19)$$

Proof. Note that randomization is no longer necessary for the minimax games

of (3.16) and (3.17), so they have the same respective values as (3.18) and (3.19). $\square$

We now present the main saddle-point property of the fingerprinting games. The results owe to the convexity of the payoff function with respect to the minimizer's strategy. Such games are generally called convex games [25, §2.5].

Theorem 3.6. *For compact $\mathcal{P}^e$, compact and user-symmetric $\mathcal{P}^c$, and for both the joint and simple games, $\bar{C}_{k,q}(\mathcal{P}^e, \mathcal{P}^c) = C_{k,q}(\mathcal{P}^e, \mathcal{P}^c)$. Suppose further that $\mathcal{P}^e$ and $\mathcal{P}^c$ are symbol-symmetric, then the first argument $\mathcal{P}^e$ can be replaced by $\mathcal{P}_{\text{sym}}^e$ and/or the second argument $\mathcal{P}^c$ can be replaced by $\mathcal{P}_{\text{fair}}^c$, $\mathcal{P}_{\text{sym}}^c$, or $\mathcal{P}_{\text{fair,sym}}^c$ without changing the minimax or the maximin value. For all these games, the minimizer has an optimal strategy $p_{Y|X_K}^{(k,q)} \in \mathcal{P}_{\text{fair,sym}}^c$ while the maximizer has an optimal strategy $P_{\mathbf{W}}^{(k,q)} \in \mathcal{P}_{\text{sym}}^e$. In particular, when Condition 2.2 is satisfied, the maximizing strategy $p_{\mathbf{W}}^{(k,q)} \in \mathcal{P}_{\text{sym}}^e$ has a finite spectrum. The values of all these games equal the (joint or simple) fingerprinting capacity $C_{k,q}(\mathcal{P}^e, \mathcal{P}^c)$.*

Proof. We show that the functions $I_{k,q}^{\text{joint}}$ and $I_{k,q}^{\text{simple}}$ are convex functions of $p_{Y|X_K}$ for fixed $\mathbf{w}$. The convexity of $I_{k,q}^{\text{joint}}$ is shown in [26, Theorem 2.7.4]. To show the convexity of $I_{k,q}^{\text{simple}}(\mathbf{w}, p_{Y|X_K})$, we fix $\mathbf{w}$ and consider two different conditional distributions $p_{Y|X_K}^1$ and $p_{Y|X_K}^2$. Note that

$$\begin{aligned} I_{k,q}^{\text{simple}}(\mathbf{w}, p_{Y|X_K}) &= I(X_1; Y | \mathbf{W} = \mathbf{w}) \\ &= \sum_{x,y} p_{X_1|\mathbf{w}}(x|\mathbf{w}) p_{Y|X_1\mathbf{w}}(y|x, \mathbf{w}) \log \frac{p_{Y|X_1\mathbf{w}}(y|x, \mathbf{w})}{p_{Y|\mathbf{w}}(y|\mathbf{w})} \\ &= \sum_x p_{X_1|\mathbf{w}}(x|\mathbf{w}) D(p_{Y|X_1\mathbf{w}} \| p_{Y|\mathbf{w}} | \mathbf{W} = \mathbf{w}). \end{aligned} \quad (3.20)$$

For any $0 \leq \lambda \leq 1$, we have

$$\begin{aligned}
& \lambda I^{\text{simple}}(\mathbf{w}, p_{Y|X_K}^1) + (1 - \lambda) I^{\text{simple}}(\mathbf{w}, p_{Y|X_K}^2) \\
&= \sum_x p_{X_1|\mathbf{w}}(x|\mathbf{w}) [\lambda D(p_{Y|X_1}^1 \mathbf{w} \parallel p_{Y|\mathbf{w}}^1 | \mathbf{W} = \mathbf{w}) \\
&\quad + (1 - \lambda) D(p_{Y|X_1}^2 \mathbf{w} \parallel p_{Y|\mathbf{w}}^2 | \mathbf{W} = \mathbf{w})] \\
&\geq \sum_x p_{X_1|\mathbf{w}}(x|\mathbf{w}) D(p_{Y|X_1}^\lambda \mathbf{w} \parallel p_{Y|\mathbf{w}}^\lambda | \mathbf{W} = \mathbf{w}) \\
&= I^{\text{simple}}(\mathbf{w}, p_{Y|X_K}^\lambda)
\end{aligned} \tag{3.21}$$

where $p_{Y|X_K}^\lambda = \lambda p_{Y|X_K}^1 + (1 - \lambda) p_{Y|X_K}^2 \in \mathcal{P}^c$ by compactness. The inequality follows from the convexity of relative entropy [26, Theorem 2.7.2]. Hence $I_{k,q}^{\text{simple}}$ is convex in $p_{Y|X_K}$.

Now since $I_{k,q}(\mathbf{w}, p_{Y|X_K})$ is a convex function of $p_{Y|X_K}$ for fixed $\mathbf{w} \in \mathcal{W}^q$, $\mathbb{E}_{P_{\mathbf{W}}}[I_{k,q}(\mathbf{W}, p_{Y|X_K})]$ is also a convex function of $p_{Y|X_K}$ for fixed $P_{\mathbf{W}} \in \mathcal{P}_{\mathbf{W}}$. On the other hand, $\mathbb{E}_{P_{\mathbf{W}}}[I_{k,q}(\mathbf{W}, p_{Y|X_K})]$ is a linear function of $P_{\mathbf{W}}$ for fixed $p_{Y|X_K}$. By the minimax theorem [27], the game admits a saddle-point solution.

If $\mathcal{P}^e$ is symbol-symmetric and let $P_{\mathbf{W}}$ be a minimizing saddle-point strategy, then by symbol-symmetry each $P_{\mathbf{W}}^\pi \in \mathcal{P}^e$ is a minimizing saddle-point strategy for any permutation π of $\mathcal{X}$. The symbol-permutation averaged distribution

$$\bar{P}_{\mathbf{W}} = \frac{1}{q!} \sum_{\pi} P_{\mathbf{W}}^\pi \tag{3.22}$$

is also a minimizing saddle-point strategy and is symbol-symmetric by construction. Similarly if $\mathcal{P}^c$ is user-symmetric and symbol-symmetric, we can construct a maximizing saddle-point strategy that is both user-symmetric and symbol-symmetric.

Finally if $\mathcal{P}^e$ is the class of all probability distributions on $\text{supp}(\mathcal{P}^e)$ (Condition 2.2), the game becomes a so-called convex game whose minimizing strategy has a finite spectrum (see [25, §2.5]). $\square$

CHAPTER 4

BINARY ALPHABET FINGERPRINTING CAPACITY UNDER THE MARKING ASSUMPTION

In this chapter we study intensively the joint and simple fingerprinting games for the binary alphabet under the marking assumption. In this case the four variations RDM, UDM, ADM, and GDM of the marking assumption discussed in [23] are equivalent in terms of capacity. Hence for simplicity we assume $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and denote the collusion channel by $\mathcal{P}_{\text{RDM}}$. Tight upper and lower bounds on capacities are provided under several different setups.

4.1 Game Definition

The mutual information games for joint and simple decoder in the binary case can be simplified as follows:

1. Fingerprinting Codes

The auxiliary random vector $\mathbf{W}$ now has only one degree of freedom, and we redefine it as $W = W_1 \in [0, 1]$. P_W denotes its distribution and $p_{X|W} \sim \text{Bernoulli}(W)$.

Suppose $\mathcal{P}^e$ is compact and symbol-symmetric. Then by Theorem 3.6, it suffices to consider symbol-symmetric P_W , which in the binary case means that the distribution of W is symmetric about $1/2$, i.e.,

$$\Pr(W \leq w) = \Pr(W \geq 1 - w), \quad w \in [0, 1]. \quad (4.1)$$

In the numerical results in Sec. 4.5, we will consider a subset of the family of beta distributions, which is a family of continuous probability distributions defined on $(0, 1)$:

$$dP_W^\beta(w) = \frac{1}{B(\beta, \beta)} [w(1 - w)]^{\beta-1} \quad (4.2)$$

where the beta function, $B(\beta, \beta) \triangleq \int_0^1 [t(1-t)]^{\beta-1} dt$, appears as a normalization constant, and the parameter $\beta > 0$. The arcsine distribution, which is a special case of the beta distribution with $\beta = 1/2$, has pdf

$$dP_W^*(w) = \frac{1}{\pi \sqrt{w(1-w)}} \quad (4.3)$$

on $(0, 1)$. The arcsine distribution was first used in generating randomized fingerprinting codes by Tardos [4] and is sometimes referred to as the “Tardos distribution” in the literature.

2. Collusion Channel

Suppose $\mathcal{P}^c$ is compact and user- and symbol-symmetric. Then by Theorem 3.6 it suffices to consider user-symmetric attacks. Let $Z \triangleq \sum_{i=1}^k X_i \in \{0, 1, \dots, k\}$, which is the number of 1’s in $X_{\mathcal{K}}$. User-symmetry makes Z a sufficient statistic in producing Y . If we let $\mathbf{p} = (p_0, \dots, p_k)'$ where $p_z \triangleq p_{Y|Z}(1|z)$, $z = 0, \dots, k$, then the collusion channel can be completely characterized by $\mathbf{p}$. The marking assumption enforces that

$$p_0 = 0 \text{ and } p_k = 1. \quad (4.4)$$

On the other hand, symbol-symmetry allows us to consider $\mathbf{p}$ with

$$p_z = 1 - p_{k-z}, \quad z = 0, \dots, k. \quad (4.5)$$

The *interleaving attack* $\mathbf{p}^*$ (a.k.a. “uniform channel” in [28] and “blind colluders” in [29]) defined by

$$p_z^* = \frac{z}{k}, \quad z = 0, \dots, k \quad (4.6)$$

is frequently adopted to model the coalition’s strategy and can be easily implemented by drawing each y_j randomly from $x_{1,j}, \dots, x_{k,j}$ at each position j . One can verify that it satisfies the marking assumption (4.4) and is both user- and symbol-symmetric (4.1). We will further discuss the performance of this attack in Sec. 4.3 and Sec. 4.4.

3. Payoff Functions

Let $\boldsymbol{\alpha}(w) = (\alpha_0(w), \dots, \alpha_k(w))'$ and similarly for $\boldsymbol{\alpha}^1(w)$ and $\boldsymbol{\alpha}^0(w)$

where

$$\alpha_z(w) \triangleq p_{Z|W}(z|w) = \binom{k}{z} w^z (1-w)^{k-z} \quad (4.7)$$

is the binomial law with parameter w and k trials, and

$$\alpha_z^1(w) \triangleq p_{Z|X_1 W}(z|1, w) = \begin{cases} \binom{k-1}{z-1} w^{z-1} (1-w)^{k-z}, & 1 \leq z \leq k \\ 0, & z = 0 \end{cases} \quad (4.8)$$

and

$$\alpha_z^0(w) \triangleq p_{Z|X_1 W}(z|0, w) = \begin{cases} \binom{k-1}{z} w^z (1-w)^{k-z-1}, & 0 \leq z \leq k-1 \\ 0, & z = k \end{cases} \quad (4.9)$$

are the (shifted for $\alpha^1(w)$) binomial laws with parameter w and $k-1$ trials.

Recall that $W \rightarrow X_{\mathcal{K}} \rightarrow Z \rightarrow Y$ forms a Markov chain. We have

$$\begin{aligned} p_{Y|X_1 W}(1|x, w) &= \sum_{z=0}^k p_{Z|X_1 W}(z|x, w) p_{Y|Z}(1|z) \\ &= \sum_{z=0}^k \alpha_z^x(w) p_z = \boldsymbol{\alpha}^{x'} \mathbf{p} \end{aligned} \quad (4.10)$$

for $x = 0, 1$. The payoff function for the joint fingerprinting game is then

$$\begin{aligned} I_{k,2}^{\text{joint}}(w, \mathbf{p}) &= \frac{1}{k} I(X_{\mathcal{K}}; Y | W = w) \\ &= \frac{1}{k} I(Z; Y | W = w) \\ &= \frac{1}{k} [H(Y | W = w) - H(Y | Z, W = w)] \\ &= \frac{1}{k} \left[h_2 \left(\sum_{z=0}^k \alpha_z(w) p_z \right) - \sum_{z=0}^k \alpha_z(w) h_2(p_z) \right] \\ &= \frac{1}{k} [h_2(\boldsymbol{\alpha}' \mathbf{p}) - \boldsymbol{\alpha}' h_2(\mathbf{p})]. \end{aligned} \quad (4.11)$$

Another representation of $I_{k,2}^{\text{joint}}$ is

$$\begin{aligned}
I_{k,2}^{\text{joint}}(w, \mathbf{p}) &= \frac{1}{k} D_2(p_{ZY|W} \parallel p_{Z|W} p_{Y|W} | W = w) \\
&= \frac{1}{k} \sum_{z=0}^k \sum_{y=0}^1 p_{Z|W}(z|w) p_{Y|Z}(y|z) \log_2 \frac{p_{Y|Z}(y|z)}{p_{Y|W}(y|w)} \\
&= \frac{1}{k} \sum_{z=0}^k \alpha_z(w) \left[p_z \log_2 \frac{p_z}{\alpha' \mathbf{p}} + (1 - p_z) \log_2 \frac{1 - p_z}{1 - \alpha' \mathbf{p}} \right] \\
&= \frac{1}{k} \sum_{z=0}^k \alpha_z(w) d_2(p_z \parallel \alpha' \mathbf{p}). \tag{4.12}
\end{aligned}$$

For the simple fingerprinting game, we have

$$\begin{aligned}
I_{k,2}^{\text{simple}}(w, \mathbf{p}) &= I(X_1; Y | W = w) \\
&= D_2(p_{X_1 Y | W} \parallel p_{X_1 | W} p_{Y | W} | W = w) \\
&= \sum_{x=0}^1 \sum_{y=0}^1 p_{X_1 | W}(x|w) p_{Y | X_1 W}(y|x, w) \log_2 \frac{p_{Y | X_1 W}(y|x, w)}{p_{Y | W}(y|w)} \\
&= w D_2(p_{Y | X_1=1, W=w} \parallel p_{Y | W=w}) \\
&\quad + (1 - w) D_2(p_{Y | X_1=0, W=w} \parallel p_{Y | W=w}) \\
&\stackrel{(a)}{=} w d_2(\alpha^{1'} \mathbf{p} \parallel \alpha' \mathbf{p}) + (1 - w) d_2(\alpha^{0'} \mathbf{p} \parallel \alpha' \mathbf{p}) \tag{4.13}
\end{aligned}$$

where (a) follows from (4.10).

4. Fingerprinting Games

The fingerprinting games for the binary alphabet can now be written as

$$C_{k,2}(\mathcal{P}^e, \mathcal{P}^c) = \max_{P_W} \min_{\mathbf{p}} \mathbb{E}_{P_W} [I_{k,2}(W, \mathbf{p})] \tag{4.14}$$

$$= \min_{\mathbf{p}} \max_{P_W} \mathbb{E}_{P_W} [I_{k,2}(W, \mathbf{p})] \tag{4.15}$$

where the maximization is subject to $P_W \in \mathcal{P}^e$ and the symbol-symmetry condition (4.1) while the minimization is subject to $\mathbf{p} \in \mathcal{P}^c$ and the symbol-symmetry condition (4.5). The maximizing and minimizing strategies are denoted by $P_W^{(k,2)}$ and $\mathbf{p}^{(k,2)}$ respectively. If $\mathcal{P}^e$ satisfies

Condition 2.2, then by Lemma 3.5 we have

$$C_{k,2}(\mathcal{P}^e, \mathcal{P}^c) = \min_{\mathbf{p}} \max_w I_{k,2}(w, \mathbf{p}) \quad (4.16)$$

where the maximization is subject to $w \in \text{supp}(\mathcal{P}^e)$.

4.2 Analysis of the Convex Games

We consider the following three cases:

1. Colluders' Strategy is Fixed

$\mathcal{P}^c = \{\mathbf{p}\}$ in this case. For general $\mathcal{P}^e$ the game is still an infinite-dimensional maximization problem. However when Condition 2.2 is satisfied, it reduces to one-dimensional by (4.16) and a simple line search gives us the capacity under collusion channel $\mathbf{p}$. Note that for any $\mathbf{p} \in \mathcal{P}_{\text{RDM}}$, $C_{k,2}(\mathcal{P}^e, \mathbf{p})$ is an upper bound on $C_{k,2}(\mathcal{P}^e, \mathcal{P}_{\text{RDM}})$.

2. Fingerprinting Embedder's Strategy is Fixed

$\mathcal{P}^e = \{P_W\}$ in this case. The game reduces to a $\Theta(k)$ -dimensional minimization problem. Since the payoff function $\mathbb{E}_{P_W} [I_{k,2}(W, \mathbf{p})]$ is convex in $\mathbf{p}$, we use the conditional gradient method to solve the constrained convex optimization problem (4.14) (see [30]). For the joint fingerprinting game, Furon and Perez-Freire [22] proposed a Blahut-Arimoto algorithm which, however, cannot be applied to the simple fingerprinting game. Note that for any $P_W \in \mathcal{P}_W$, $C_{k,2}(P_W, \mathcal{P}^e)$ is a lower bound on $C_{k,2}(\mathcal{P}_W, \mathcal{P}^e)$.

3. Fingerprinting Capacities Under the Marking Assumption

We consider specifically $\mathcal{P}^e = \mathcal{P}_W$ and $\mathcal{P}^c = \mathcal{P}_{\text{RDM}}$. Solving the maximin game of (4.14) or the minimax game of (4.16) is much more difficult than solving the above maximization or minimization problems. In particular, the alternating maximization and minimization algorithm generally diverges.

Owing to the existence of a saddle-point solution, $\mathbf{p}^{(k,2)}$ and $p_W^{(k,2)}$ (note that it is a pmf by Theorem 3.6) must satisfy the following:

- (a) When $\mathbf{p} = \mathbf{p}^{(k,2)}$ is fixed, $I_{k,2}(w, \mathbf{p}^{(k,2)})$ is a differentiable function over the unit interval. The support $\text{supp}(p_W^{(k,2)})$ of $p_W^{(k,2)}$ can only take values at the maximizers of $I_{k,2}(w, \mathbf{p}^{(k,2)})$ [25, §2.5]. Hence we have

$$\begin{cases} I_{k,2}(w, \mathbf{p}^{(k,2)}) = C_{k,2}(\mathcal{P}_W, \mathcal{P}_{\text{RDM}}) \\ \frac{\partial}{\partial w} I_{k,2}(w, \mathbf{p}^{(k,2)}) = 0 \\ \frac{\partial^2}{\partial w^2} I_{k,2}(w, \mathbf{p}^{(k,2)}) < 0 \end{cases} \quad (4.17)$$

for any $w \in \text{supp}(p_W^{(k,2)})$.

- (b) When $p_W = p_W^{(k,2)}$ is fixed and the constraint (4.5) is imposed, we have

$$\mathbb{E}_{p_W^{(k,2)}} \left[\frac{\partial}{\partial p_z} I_{k,2}(W, \mathbf{p}^{(k,2)}) \right] = 0, \quad z = 1, \dots, k-1. \quad (4.18)$$

By the convexity in $\mathbf{p}$ of the payoff function, we have $|\text{supp}(p_W^{(k,2)})| \leq k$ (see [25, §2.5]). With a fixed spectrum cardinality, we can obtain candidate capacity-achieving strategies $p_W^{(k,2)}$ and $\mathbf{p}^{(k,2)}$ by solving (4.17) and (4.18), and then verify whether those candidate distributions are optimal by examining the second partial derivatives. Once $\mathbf{p}^{(k,2)}$ and $p_W^{(k,2)}$ are found, we can evaluate $C_{k,2}(\mathcal{P}_W, \mathcal{P}_{\text{RDM}})$ by substituting them into (4.14).

Numerical solutions to the joint and simple fingerprinting games are shown in Fig. 4.1-4.3. Observe that capacities for both games (Fig. 4.1(a)-(b)), the optimal distributions $P_W^{(k,2)}$ for both games (Fig. 4.2(a)-(b)), and the optimal attacks $\mathbf{p}^{(k,2)}$ for the joint fingerprinting game (Fig. 4.3(a)) all seem to converge as k grows. The optimal attacks $\mathbf{p}^{(k,2)}$ for the simple fingerprinting game (Fig. 4.3(b)), however, exhibit some wild oscillations in both amplitude and frequency as k grows. We will study the asymptotics of the joint fingerprinting game for large k in the next section.

4.3 Capacity Bounds

The analysis of Sec. 4.2 allows us to solve the fingerprinting game numerically for small k . However, evaluating or even approximating the capacity value

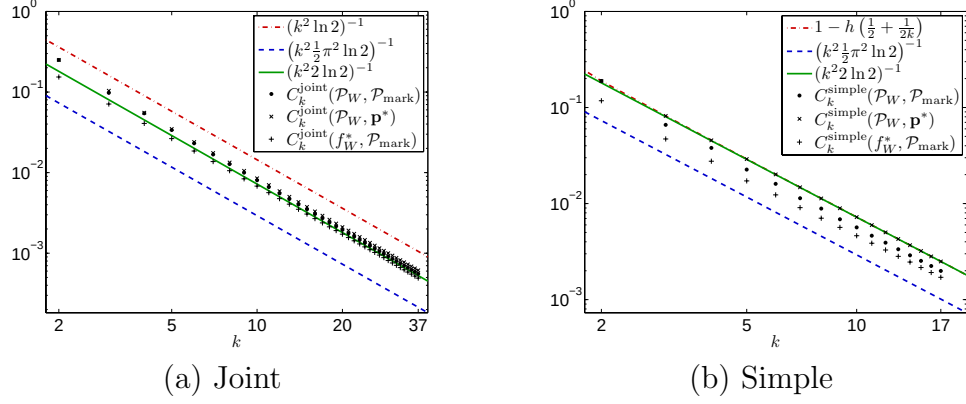


Figure 4.1: Capacities $C_{k,2}(\mathcal{P}_W, \mathcal{P}_{\text{RDM}})$, $C_{k,2}(\mathcal{P}_W, \mathbf{p}^*)$, $C_{k,2}(f_W^*, \mathcal{P}_{\text{RDM}})$, and upper and lower bounds of the (a) joint and (b) simple fingerprinting games

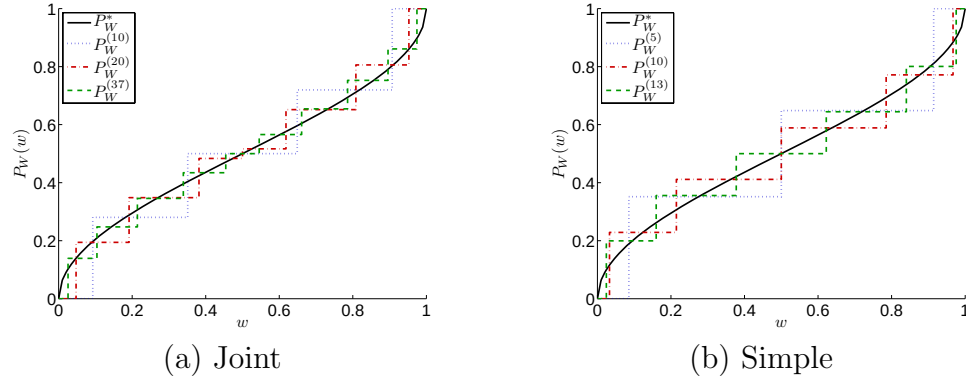


Figure 4.2: Cumulative distribution function of the arcsine distribution P_W^* and the optimal distributions $P_W^{(k,2)}$ for the (a) joint and (b) simple fingerprinting games under the marking assumption

for large k is still a difficult task. In this subsection, we provide tight upper and lower bounds on capacity.

For simplicity of notation, we let

$$g_k(w) \triangleq p_{Y|W}(1|w) = \sum_{z=0}^k \alpha_z(w) p_z = \boldsymbol{\alpha}(w)' \mathbf{p} \quad (4.19)$$

which by the definition of $\alpha_z(w)$ in (4.7) is a polynomial in w of degree $\leq k$. Note that $g_k(0) = p_0 = 0$ and $g_k(1) = p_k = 1$ by the marking assumption.

The following lemmas will be useful for the proofs:

Lemma 4.1 (Pinsker's inequality). [26, Lemma 11.6.1]

$$d_2(r \parallel s) \geq \frac{2}{\ln 2} (r - s)^2. \quad (4.20)$$

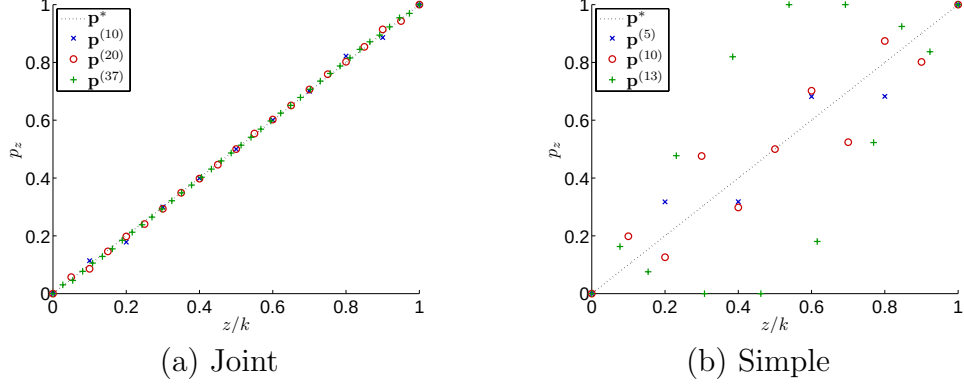


Figure 4.3: The interleaving attack $\mathbf{p}^*$ and the optimal attacks $\mathbf{p}^{(k,2)}$ for the (a) joint and (b) simple fingerprinting games under the marking assumption

Lemma 4.2. *Equalities*

$$\alpha^{1'} \mathbf{p} - \alpha' \mathbf{p} = \frac{1-w}{k} g'_k(w)$$

and

$$\alpha^{0'} \mathbf{p} - \alpha' \mathbf{p} = -\frac{w}{k} g'_k(w)$$

hold for $z = 0, \dots, k$ and $w \in [0, 1]$.

Proof. The equalities follow directly from (4.7), (4.8), (4.9), and (4.19). $\square$

Lemma 4.3. *Let dP_W be a pdf on $[0, 1]$. Then*

$$\int_0^1 \frac{dw}{dP_W(w)w(1-w)} \geq \pi^2 \quad (4.21)$$

with equality if and only if dP_W is the arcsine distribution of (4.3).

Proof. By the Cauchy-Schwarz inequality, we have

$$\int_0^1 \frac{dw}{dP_W(w)w(1-w)} \geq \frac{\left[\int_0^1 \frac{dw}{\sqrt{w(1-w)}} \right]^2}{\int_0^1 dP_W(w)dw} = \pi^2.$$

Equality holds if and only if $dP_W(w) \propto \frac{1}{\sqrt{w(1-w)}}$, which leads us to the arcsine distribution. $\square$

4.3.1 Upper Bounds

The following two theorems bound from above capacities under the interleaving attack of (4.6).

Theorem 4.4. [1, Theorem 4.2]

$$C_{k,2}^{\text{joint}}(\mathcal{P}_W, \mathbf{p}^*) \leq \frac{1}{k^2 \ln 2}. \quad (4.22)$$

Proof.

$$\begin{aligned} C_{k,2}^{\text{joint}}(\mathcal{P}_W, \mathbf{p}^*) &= \max_{w \in [0,1]} I_{k,2}^{\text{joint}}(w, \mathbf{p}^*) \\ &\stackrel{(a)}{=} \frac{1}{k} \max_{w \in [0,1]} \left\{ h_2(w) - \sum_{z=0}^k \alpha_z(w) h_2\left(\frac{z}{k}\right) \right\} \\ &\stackrel{(b)}{\leq} \frac{1}{k^2 \ln 2} \end{aligned}$$

where (a) follows from (4.11) and (b) results from [28, Theorem 4.3]. $\square$

Theorem 4.5. [31, Proposition 4.2]

$$C_{k,2}^{\text{simple}}(\mathcal{P}_W, \mathbf{p}^*) = 1 - h_2\left(\frac{1}{2} + \frac{1}{2k}\right) = \frac{1}{k^2 2 \ln 2} + O\left(\frac{1}{k^4}\right). \quad (4.23)$$

Proof. It can be shown that $I_{k,2}^{\text{simple}}(w, \mathbf{p}^*)$ takes its maximum at $w = 1/2$ (See Appendix A.1). Hence

$$\begin{aligned} C_{k,2}^{\text{simple}}(\mathcal{P}_W, \mathbf{p}^*) &= \max_{w \in [0,1]} I_{k,2}^{\text{simple}}(w, \mathbf{p}^*) \\ &= 1 - h_2\left(\frac{1}{2} + \frac{1}{2k}\right) = \frac{1}{k^2 2 \ln 2} + O\left(\frac{1}{k^4}\right). \quad \square \end{aligned}$$

4.3.2 Lower Bounds

The following theorem provides a lower bound on both the joint and simple capacities under a continuous probability distribution dP_W .

Theorem 4.6. *Let P_W has a pdf dP_W on $[0, 1]$. Then*

$$C_{k,2}^{\text{joint}}(P_W, \mathcal{P}_{\text{RDM}}) \geq C_{k,2}^{\text{simple}}(P_W, \mathcal{P}_{\text{RDM}}) \geq \frac{2}{k^2 \ln 2} \left[\int_0^1 \frac{dw}{dP_W(w)w(1-w)} \right]^{-1}. \quad (4.24)$$

The lower bound is maximized when $P_W = P_W^*$ where it takes the value $\frac{2}{k^2 \pi^2 \ln 2}$.

Proof. For any $\mathbf{p} \in \mathcal{P}_{\text{RDM}}$, we have

$$\begin{aligned} & \mathbb{E}_{P_W} \left[I_{k,2}^{\text{simple}}(W, \mathbf{p}) \right] \\ & \stackrel{(a)}{=} \int_0^1 \left[w d_2(\boldsymbol{\alpha}^{1'} \mathbf{p} \parallel \boldsymbol{\alpha}' \mathbf{p}) + (1-w) d_2(\boldsymbol{\alpha}^{0'} \mathbf{p} \parallel \boldsymbol{\alpha}' \mathbf{p}) \right] dP_W(w) dw \\ & \stackrel{(b)}{\geq} \frac{2}{\ln 2} \int_0^1 \left[w (\boldsymbol{\alpha}^{1'} \mathbf{p} - \boldsymbol{\alpha}' \mathbf{p})^2 + (1-w) (\boldsymbol{\alpha}^{0'} \mathbf{p} - \boldsymbol{\alpha}' \mathbf{p})^2 \right] dP_W(w) dw \\ & \stackrel{(c)}{=} \frac{2}{k^2 \ln 2} \int_0^1 [g'_k(w)]^2 w(1-w) dP_W(w) dw \\ & \stackrel{(d)}{\geq} \frac{2}{k^2 \ln 2} \frac{\left[\int_0^1 g'_k(w) dw \right]^2}{\int_0^1 \frac{dw}{dP_W(w)w(1-w)}} \\ & \stackrel{(e)}{=} \frac{2}{k^2 \ln 2} \left[\int_0^1 \frac{dw}{dP_W(w)w(1-w)} \right]^{-1}. \end{aligned}$$

(a) follows from (4.13). (b) follows from Pinsker's inequality (Lemma 4.1).

(c) follows from Lemma 4.2. (d) follows from the Cauchy-Schwarz inequality.

Finally, (e) follows from the marking assumption. Hence

$$\begin{aligned} C_{k,2}^{\text{simple}}(P_W, \mathcal{P}_{\text{RDM}}) &= \min_{\mathbf{p} \in \mathcal{P}_{\text{RDM}}} \mathbb{E}_{P_W} \left[I_{k,2}^{\text{simple}}(W, \mathbf{p}) \right] \\ &\geq \frac{2}{k^2 \ln 2} \left[\int_0^1 \frac{dw}{dP_W(w)w(1-w)} \right]^{-1}. \end{aligned}$$

Following Lemma 4.3, the lower bound is maximized when $P_W = P_W^*$, which coincides with the lower bound given in [31]. $\square$

The following corollaries summarizes the upper and lower bounds on capacities under the marking assumption:

Corollary 4.7.

$$\frac{2}{k^2 \pi^2 \ln 2} \leq C_{k,2}^{\text{joint}}(\mathcal{P}_W, \mathcal{P}_{\text{RDM}}) \leq \frac{1}{k^2 \ln 2}. \quad (4.25)$$

Corollary 4.8.

$$\frac{2}{k^2 \pi^2 \ln 2} \leq C_{k,2}^{\text{simple}}(\mathcal{P}_W, \mathcal{P}_{\text{RDM}}) \leq \frac{1}{k^2 2 \ln 2} + O\left(\frac{1}{k^4}\right). \quad (4.26)$$

4.4 Asymptotics for Large Coalitions

The upper and lower bounds on $C_{k,2}^{\text{joint}}(\mathcal{P}_W, \mathcal{P}_{\text{RDM}})$ provided in the previous section are within a factor of about five. As can be seen in Fig. 4.1, the numerical results suggest that $C_{k,2}^{\text{joint}}(\mathcal{P}_W, \mathcal{P}_{\text{RDM}})$ approximates $(k^2 2 \ln 2)^{-1}$ even for small values of k . Amiri and Tardos [2] claimed the same asymptotic rate but only provided the proof for the lower bound in [32, Theorem 15]. Here we analyze not only this rate but the complete asymptotics of the joint fingerprinting game.

We consider the sequence of mutual information games for joint decoding. To study the asymptotics when $k \rightarrow \infty$, we first assume that the collusion channel $\mathbf{p}$ satisfies the following regularity condition:

Condition 4.1. *There exists a bounded and twice differentiable function $g(w)$ on $[0, 1]$ with $g(0) = 0$ and $g(1) = 1$ such that*

$$p_z = g\left(\frac{z}{k}\right), \quad \forall z \in \{0, \dots, k\}. \quad (4.27)$$

Certainly the condition restricts the colluders' strategy to a smaller space. We however claim that this is a very mild limitation on their power for the following reasons:

1. For each k , the collusion channels take values of g at only the lattice points in $[0, 1]$, hence intuitively the class of collusion channels satisfying Condition 4.1 remains large.
2. Fig. 4.4 shows the minimizing collusion channels $\mathbf{p}^{(k,2)}$ for several different embedding distributions. For each case it seems the continuous interpolation of $\mathbf{p}$ does converge to some g on $[0, 1]$. Indeed, our following analysis still holds if we relax the restriction of (5.12) to

$$p_z = g\left(\frac{z}{k}\right) + o\left(\frac{1}{k}\right), \quad \forall z \in \{0, \dots, k\}. \quad (4.28)$$

The following reparameterization of the class (5.12) of collusion channels will simplify our analysis:

Definition 4.2. Let G and J be functions on $[0, 1]$ defined as

$$G(w) \triangleq \cos^{-1}[1 - 2g(w)] \quad (4.29)$$

and

$$J(w) \triangleq w(1 - w)[G'(w)]^2 \quad (4.30)$$

where $g(w)$ satisfies Condition 4.1.

The outline of our asymptotic analysis is as follows: we fix $w \in (0, 1)$ and we study the asymptotics of $I_{k,2}^{\text{joint}}(w, \mathbf{p})$. The binomial distribution of Z can be approximated by the Gaussian distribution with mean kw and variance $kw(1 - w)$, and by which we can approximate the dominating terms of $I_{k,2}^{\text{joint}}(w, \mathbf{p})$. Theorem 4.11 yields $I_{k,2}^{\text{joint}}(w, \mathbf{p}) \sim J(w)/(k^2 2 \ln 2)$, where J is the transformation of g defined in (4.30). The maximin game with J as the payoff function can be solved explicitly and hence the asymptotics of the fingerprinting game can be obtained.

The following lemmas will be useful for our analysis:

Lemma 4.9. [33, Sec. 2.5]

$$d_2(r \parallel s) = \frac{(r - s)^2}{s(1 - s)2 \ln 2} + O(|r - s|^3). \quad (4.31)$$

Lemma 4.10. For $Z \sim \text{Binomial}(k, w)$, we have

$$\Pr[|Z - kw| \geq \sqrt{k \ln k}] \leq 1/k^2. \quad (4.32)$$

Proof. This is special case of Hoeffding's inequality (see [34]). $\square$

Recall that the expectation of Y given $W = w$, which we denote by $g_k(w)$ in (4.19), can be written as

$$g_k(w) \triangleq p_{Y|W}(1|w) = \sum_{z=0}^k \alpha_z(w) p_z = \sum_{z=0}^k \alpha_z(w) g\left(\frac{z}{k}\right)$$

where

$$\alpha_z(w) \triangleq p_{Z|W}(z|w) = \binom{k}{z} w^z (1 - w)^{k-z}$$

is the binomial pmf which concentrates around its mean kw as $k \rightarrow \infty$. g_k is a polynomial in w of degree $\leq k$ and is known as the Bernstein polynomial of order k of the function g [35]. By Condition 4.1 g is bounded and the second derivative $g''(w)$ exists; therefore, from Bernstein [35, §1.6] we have

$$g_k(w) = g(w) + \frac{w(1-w)}{2k} g''(w) + o\left(\frac{1}{k}\right). \quad (4.33)$$

On the other hand, for $z = k(w + \epsilon)$, we have

$$\begin{aligned} p_z &= g\left(\frac{z}{k}\right) = g(w + \epsilon) \\ &= g(w) + \epsilon g'(w) + O(\epsilon^2). \end{aligned} \quad (4.34)$$

We now write the asymptotic approximation of $I_{k,2}^{\text{joint}}$ in terms of g .

Firstly by the bounds presented in Sec. 4.3, we can focus on w and g such that $I_{k,2}^{\text{joint}}(w, \mathbf{p}) = \Omega(1/k^2)$. Let $\delta = \sqrt{\ln k/k}$. We have

$$\begin{aligned} I_{k,2}^{\text{joint}}(w, \mathbf{p}) &\stackrel{(a)}{=} \frac{1}{k} \sum_{z=0}^k \alpha_z(w) d_2(p_z \parallel g_k(w)) \\ &\stackrel{(b)}{\sim} \frac{1}{k} \sum_{z: |z-kw| < k\delta} \alpha_z(w) d_2(p_z \parallel g_k(w)) \end{aligned} \quad (4.35)$$

where (a) follows from (4.12) and (4.19) and (b) from Lemma 4.10.

Now if we let $z = kw + \eta$, where $\eta = O(\sqrt{k \ln k})$, then by (4.34) we have

$$p_z = g(w) + \frac{\eta}{k} g'(w) + O\left(\frac{\eta^2}{k^2}\right) \quad (4.36)$$

and combining with (4.33) we have

$$p_z - g_k(w) = \frac{\eta}{k} g'(w) + O\left(\frac{\ln k}{k}\right). \quad (4.37)$$

for $\eta = \omega(1)$. The contribution to (4.35) for $\eta = O(1)$ decays exponentially with k and thus can be neglected.

By (4.37) and Lemma A.2, we have

$$d_2(p_z \parallel g_k(w)) = \frac{[p_z - g_k(w)]^2}{g_k(w)(1 - g_k(w))2 \ln 2} + o\left(\frac{1}{k}\right) \quad (4.38)$$

and hence (4.35) yields

$$\begin{aligned}
I_{k,2}^{\text{joint}}(w, \mathbf{p}) &\stackrel{(a)}{\sim} \frac{1}{k^2 \ln 2} \sum_{z: |z-kw| < k\delta} \alpha_z(w) \frac{[p_z - g_k(w)]^2}{g_k(w)(1 - g_k(w))} \\
&\stackrel{(b)}{\sim} \frac{[g'(w)]^2}{k^3 g(w)(1 - g(w)) 2 \ln 2} \sum_{z: |z-kw| < k\delta} \alpha_z(w) (z - kw)^2 \\
&\stackrel{(c)}{\sim} \frac{[g'(w)]^2 w(1 - w)}{k^2 g(w)(1 - g(w)) 2 \ln 2} \\
&\stackrel{(d)}{=} \frac{1}{k^2 2 \ln 2} J(w)
\end{aligned} \tag{4.39}$$

where (a) follows from (4.38), (b) from (4.33) and (4.37), (c) from Lemma 4.10, and (d) directly from the definitions in (4.29) and (4.30). The following theorem concludes what we have proved thus far:

Theorem 4.11. *Assume that Condition 4.1 is satisfied, then*

$$I_{k,2}^{\text{joint}}(w, \mathbf{p}) \sim \frac{1}{k^2 2 \ln 2} J(w), \quad \forall w \in (0, 1). \tag{4.40}$$

The joint fingerprinting game of (4.16) can now be approximated by the game with J as its payoff function. We consider continuous probability distributions P_W satisfying the following condition:

Condition 4.3. *The probability distribution P_W has a continuous pdf dP_W on $[0, 1]$ with*

$$\int_0^1 \frac{dw}{dP_W(w)w(1-w)} < \infty. \tag{4.41}$$

The following lemma shows the solution to the minimization problem with J as its payoff function.

Lemma 4.12. *Let $g(w)$ satisfy Condition 4.1 and fix P_W satisfying Condition 4.3. Then*

$$\begin{aligned}
&\min_g \mathbb{E}_{P_W} [J(W)] \\
&= \min_g \int_0^1 J(w) dP_W(w) dw = \pi^2 \left[\int_0^1 \frac{dw}{dP_W(w)w(1-w)} \right]^{-1}
\end{aligned} \tag{4.42}$$

where J is defined in (4.30). The minimum is achieved by

$$g_{\text{opt}}(w) = \frac{1}{2} \left[1 - \cos \left(\frac{\pi \int_0^w \frac{dv}{dP_W(v)v(1-v)}}{\int_0^1 \frac{dv}{dP_W(v)v(1-v)}} \right) \right]. \quad (4.43)$$

Proof. We have

$$\begin{aligned} \int J(w) dP_W(w) dw &= \int_0^1 w(1-w) [G'(w)]^2 dP_W(w) dw \\ &\stackrel{(a)}{\geq} \frac{\left[\int_0^1 G'(w) dw \right]^2}{\int_0^1 \frac{dw}{w(1-w) dP_W(w)}} \\ &\stackrel{(b)}{=} \pi^2 \left[\int_0^1 \frac{dw}{dP_W(w) w(1-w)} \right]^{-1} \end{aligned} \quad (4.44)$$

where (a) follows from the Cauchy-Schwarz inequality and (b) follows from the boundary conditions $G(0) = 0$ and $G(1) = \pi$ following directly from Condition 4.1 and the definition of (4.29). Equality holds in (a) when

$$G'_{\text{opt}}(w) = \frac{\pi}{\int_0^1 \frac{dv}{dP_W(v)v(1-v)}} \cdot \frac{1}{dP_W(w)w(1-w)}, \quad (4.45)$$

which leads us to (4.43) by (4.29). $\square$

Corollary 4.13. *For P_W satisfying Condition 4.3, we have*

$$C_{k,2}^{\text{joint}}(P_W, \mathcal{P}_{\text{RDM}}) \sim \frac{\pi^2}{k^2 2 \ln 2} \left[\int_0^1 \frac{dw}{dP_W(w)w(1-w)} \right]^{-1}. \quad (4.46)$$

Proof. This follows directly from Theorem 4.11 and Lemma 4.12. $\square$

Corollary 4.14.

$$C_{k,2}^{\text{joint}}(P_W^*, \mathcal{P}_{\text{RDM}}) \sim \frac{1}{k^2 2 \ln 2}. \quad (4.47)$$

Proof. The right-hand side of (4.46) is maximized when $P_W = P_W^*$ by Lemma 4.3. Also note that by (4.43) we have $g_{\text{opt}}(w) = w$, which leads us to the interleaving attack. $\square$

Corollary 4.15. *The interleaving attack is an “equalizing strategy” for the colluders that makes the payoff function $J(w)$ asymptotically independent of w :*

$$I_{k,2}^{\text{joint}}(w, \mathbf{p}^*) \sim \frac{1}{k^2 2 \ln 2}, \quad \forall w \in (0, 1). \quad (4.48)$$

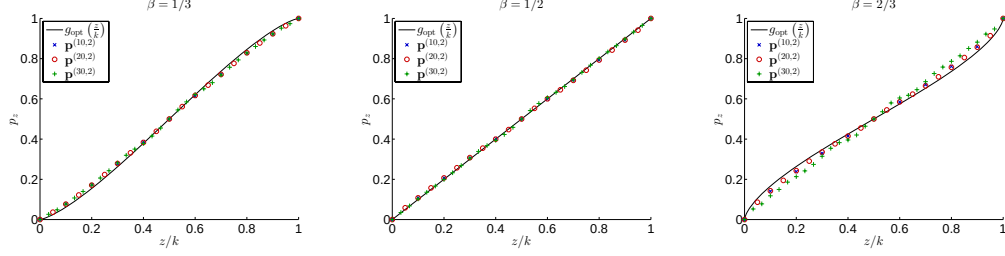


Figure 4.4: $g_{\text{opt}}(z/k)$ and minimizing collision channels $\mathbf{p}^{(k,2)}$ for $k = 10, 20$, and 30 and $P_W = \text{Beta}(\beta, \beta)$

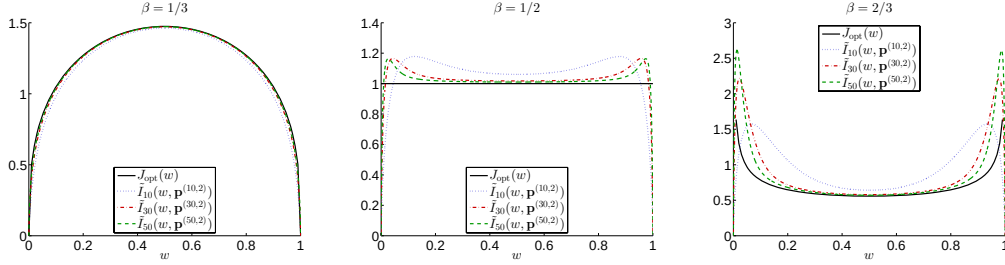


Figure 4.5: $J_{\text{opt}}(w)$ and normalized payoff function $\tilde{I}_k(w, \mathbf{p}^{(k,2)})$ for $k = 10, 30$, and 50 and $P_W = \text{Beta}(\beta, \beta)$

Proof. Let $g(w) = w$. Then $\mathbf{p}$ becomes the interleaving attack by (5.12) and $J(w) \equiv 1$ by Definition 4.2. $\square$

Corollary 4.16. *The fingerprinting capacity under the marking assumption satisfies*

$$C_{k,2}^{\text{joint}}(\mathcal{P}_W, \mathcal{P}_{\text{RDM}}) \sim \frac{1}{k^2 2 \ln 2}. \quad (4.49)$$

Furthermore, the arcsine distribution P_W^* and the interleaving attack $\mathbf{p}^*$ are the respective maximizing and minimizing strategies that achieve the asymptotic capacity value.

Proof. The asymptotic relations (4.47) and (4.48) establish matching asymptotic lower and upper bounds on $C_{k,2}^{\text{joint}}(\mathcal{P}_W, \mathcal{P}_{\text{RDM}})$ respectively. $\square$

4.5 Numerical Results for Beta Distributions

We now use the family of Beta distributions to illustrate the asymptotics of the joint fingerprinting game. Let P_W^β be the probability distribution defined in (4.2). Condition 4.3 is satisfied for any $\beta \in (0, 1)$. For $\beta = 1/3, 1/2$, and $2/3$, we find the minimizing collision channels $\mathbf{p}^{(k,2)}$ for P_W^β and compare

them with $g_{\text{opt}}(z/k)$ obtained by (4.43). Fig. 4.4 shows that $p_z^{(k,2)}$ does converge to $g_{\text{opt}}(z/k)$ as $k \rightarrow \infty$ as expected, which also rationalizes our assumption of Condition 4.1.

Consider the normalized payoff function $\tilde{I}_{k,2}^{\text{joint}}(w, \mathbf{p}) \triangleq k^2 2 \ln 2 \cdot I_{k,2}^{\text{joint}}(w, \mathbf{p})$, which by Theorem 4.11 is asymptotically close to $J(w)$. Suppose J_{opt} is obtained by substituting g_{opt} of (4.43) into (4.29) and (4.30). Again for $\beta = 1/3, 1/2$, and $2/3$, we compare $\tilde{I}_{k,2}^{\text{joint}}(w, \mathbf{p}^{(k,2)})$ with $J_{\text{opt}}(w)$ in Fig. 4.5. As shown in the figure, $\tilde{I}_{k,2}^{\text{joint}}(w, \mathbf{p}^{(k,2)})$ is asymptotically flat over $(0, 1)$ when $\beta = 1/2$, which is the case when P_W^β is chosen properly. If $\beta < 1/2$, which means P_W^β has too much weight around 0 and 1, then the colluders' choice of g_{opt} makes J_{opt} peak at $w = 1/2$. If on the contrary $\beta > 1/2$, then too much weight around $1/2$ is put on P_W^β , and g_{opt} makes J_{opt} peaks at $w = 0, 1$.

4.6 Why Are the Arcsine Distribution and the Interleaving Attack Optimal for Large Coalitions?

The analysis and the numerical results above show not only that the asymptotic capacity is $(k^2 2 \ln 2)^{-1}$, but also that both the arcsine distribution for the maximizer and the interleaving attack for the minimizer achieve the same asymptotic value. Such results are very interesting, and at the same time raise some issues for further investigation. One topic concerns the regularity constraint (Condition 4.1) upon which the asymptotic analysis in Sec. 4.4 is based. However, it is reasonable to conjecture that the same asymptotics hold without the regularity condition. Our numerical results indeed suggest this condition may not be necessary. Moreover, it is important to mention that both the asymptotic lower bound on capacity (see [32, Theorem 15]) and Corollary 4.15 (which contributes to the asymptotic upper bound) hold without the regularity constraint.

Both the arcsine distribution and the interleaving attack have been extensively studied in the literature. In 2003, Tardos applied the arcsine distribution to fingerprinting [4]. How he fine-tuned his codes, however, had been a mystery until Škorić et al. [36] and Furon et al. [29] rationalized Tardos' choices based on Gaussian approximations. On the other hand, the interleaving attack has been frequently adopted to model the collusion channel in the literature [28, 29, 22], but no conceptual reasoning has been proposed

on why it should be the coalition's optimal choice. Fortunately, owing to the discovery of the capacity formulas (Theorem 3.4), we can now study fingerprinting games from the information-theoretic point of view. In the previous subsection, we established the optimality of these two strategies based on asymptotic methods. Here we provide a statistical interpretation.

We may think of the (joint) fingerprinting capacity game as follows: the coalition is given k independent *observations* $X_1, \dots, X_k$ distributed according to an unknown distribution $\text{Bernoulli}(W)$ chosen at random by the fingerprinting embedder from the family $\{\text{Bernoulli}(W) : W \in [0, 1]\}$ according to a known *prior distribution* P_W . Upon generating Y according to the conditional distribution $p_{Y|Z}$ given the sufficient statistic $Z = \sum_{i=1}^k X_i$, the coalition suffers a loss $I(Z; Y|W = w)$. The *risk* of the game $I(Z; Y|W)$ is the average loss under P_W .

As emphasized in previous works [3, 22], the choice of the embedding distribution P_W , or *prior selection* in statistician's language, is crucial to the fingerprinting game. If no randomization takes place [37], or equivalently, if the realization w is revealed to the pirates [22], then the maximin game value decays exponentially with coalition size k (see (3.5)). Loosely speaking, the loss the pirates suffer is mainly due to their error in estimating W . If they have a good estimation of the time-sharing random variable W , then the loss they suffer is small.

Jeffreys' prior [38] is a "non-informative" prior that plays an important role in Bayesian statistics. Given a family of distributions with an unknown parameter, Jeffreys' prior is proportional to the square root of the Fisher information. Conceptually speaking, Jeffreys' prior is the "least-favorable" distribution in regard to estimating that parameter. For the Bernoulli trial with the probability of success w as parameter, the Fisher information is $I(w) = [w(1 - w)]^{-1}$ and thus Jeffreys' prior is

$$f(w) \propto \frac{1}{\sqrt{w(1 - w)}} \quad (4.50)$$

which is exactly the arcsine distribution!

The optimality of the interleaving attack is closely related to the concept of "equalizer rule" in game theory. From Corollary 4.15, interleaving is the asymptotic equalizing strategy, which is the desirable attribute we expect for a saddle-point solution. The optimal collusion channel depends on the prior

by (4.43), and from the proof of Corollary 4.14, the interleaving attack is optimal under the arcsine distribution. Also observe that the interleaving attack is the strategy where the colluders generate Y according to $\text{Bernoulli}(\widehat{W})$, where $\widehat{W} = Z/k$ is the maximum likelihood estimator of W , which is asymptotically unbiased (as $k \rightarrow \infty$) and has minimum asymptotic variance (equal to $(kI(w))^{-1} = w(1-w)/k$).

Another interesting question is what are the asymptotics of the simple fingerprinting game. In Corollary 4.8, we established upper and lower bounds on $C_{k,2}^{\text{simple}}$. Note that the upper bound is obtained by assuming the interleaving attack for the coalition and it coincides with the asymptotic rate $(k^2 2 \ln 2)^{-1}$ of $C_{k,2}^{\text{joint}}$. On the other hand, Fig. 4.3(b) indicates that the optimal attack is actually quite different from the interleaving attack. This suggests that the pirates can exploit the suboptimality of the single-user decoder and perform a stronger attack. The study of the exact asymptotics of the simple fingerprinting game is discussed in the next chapter.

CHAPTER 5

ARBITRARY ALPHABET FINGERPRINTING CAPACITY UNDER THE COMBINED DIGIT MODEL

In this chapter we study the joint and simple fingerprinting games with k pirates for arbitrary alphabet under the combined digit model. For small k , capacities along with optimal strategies for both players of the game are obtained explicitly. For large k , we extend our previous asymptotic analysis for the binary alphabet with the marking assumption to this general model and show that capacity is asymptotic to A/k^2 where the constant A is specified as the maximin value of a continuous functional game. Saddle-point solutions to the game are obtained using methods of variational calculus.

5.1 Fighting Small Coalitions

Owing to the existence of the saddle-point solutions, the joint and simple fingerprinting capacity games can be solved either analytically or numerically when k is small. In the previous chapter we showed solutions for the binary alphabet case. Our algorithm can be generalized to the q -ary alphabet with the combined digit model. Also we impose both the user-symmetry and the symbol-symmetry constraints discussed in Chap. 3.

5.1.1 One Pirate

For $k = 1$, the joint and simple capacity games are the same. The maximin games of (3.9) and (3.10) reduce to compound channel capacity games [39], or to watermarking games [40]. For the four variants of the marking assumption, capacity is trivially one. For CDM, the optimal embedding distribution is $W_x^{(k,q)} \equiv 1/q$ for all $x \in \mathcal{X}$. Upon receiving $\mathbf{z} = [1, 0, \dots, 0]'$, the colluder utilizes strategy $\boldsymbol{\theta}(\mathbf{z}) = [1 - p_c - p_e, p_c/(q-1), \dots, p_c/(q-1), p_e]'$ where p_c is the crossover probability and p_e is the erasure probability. The optimal

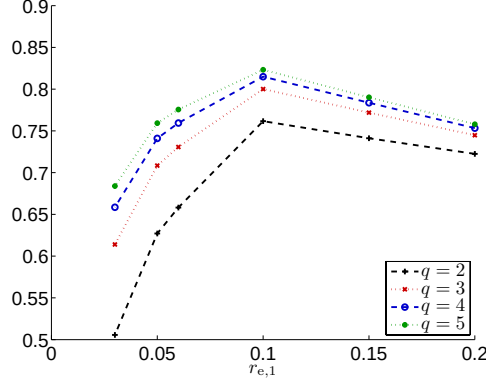


Figure 5.1: Fingerprinting capacities $C_{1,q}(\mathcal{P}_{\text{CDM}})$

strategy is $p_e^{(k,q)} = r_{e,1}$ and $p_c^{(k,q)} = \min \left\{ \frac{q-1}{q}(1 - r_{e,1}), r_{c,1} \right\}$. Capacity is given by

$$\begin{aligned}
C_{1,q}^{\text{joint}}(\mathcal{P}_{\text{CDM}}) &= C_{1,q}^{\text{simple}}(\mathcal{P}_{\text{CDM}}) \\
&= \left\{ (1 - r_{e,1}) \left[1 - h_q \left(\frac{r_{e,1}}{1 - r_{e,1}} \right) \right] - r_{c,1} \log_q(q - 1) \right\}_+
\end{aligned}$$

where $(x)_+ \triangleq \max(x, 0)$. Capacities using parameters in [6] are shown in Fig. 5.1. We can see that the loss in capacity from the ideal case is about 20% if parameters are chosen properly.

5.1.2 Two Pirates

By studying the simplest case involving collusion attacks, we can see how different models allow different strengths of attacks. By symbol-symmetry there are only two types of patterns in $\mathcal{Z}$, namely $\mathbf{z}_1 = [2, 0, \dots, 0]'$ and $\mathbf{z}_2 = [1, 1, 0, \dots, 0]'$. Similar to the analysis in 5.1.1, the crossover probabilities $\mathbf{p}_c$ and the erasure probabilities $\mathbf{p}_e$ determine the collusion attack. More specifically, we let $\boldsymbol{\theta}(\mathbf{z}_1) = \left[1 - p_{c,1} - p_{e,1}, \frac{p_{c,1}}{q-1}, \dots, \frac{p_{c,1}}{q-1}, p_{e,1} \right]'$ and $\boldsymbol{\theta}(\mathbf{z}_2) = \left[\frac{1-p_{c,2}-p_{e,2}}{2}, \frac{1-p_{c,2}-p_{e,2}}{2}, \frac{p_{c,2}}{q-2}, \dots, \frac{p_{c,2}}{q-2}, p_{e,2} \right]'$ due to symbol-symmetry. The optimal collusion channel can then be characterized by $\mathbf{p}_c^{(k,q)}$ and $\mathbf{p}_e^{(k,q)}$.

- **Restricted Digit Model:** Under RDM, $\mathbf{p}_c = \mathbf{p}_e = \mathbf{0}$ and the optimal embedding distribution is $W_x^{(k,q)} \equiv 1/q$ for all $x \in \mathcal{X}$ for both the joint

and simple capacity games. Capacities are given by

$$C_{2,q}^{\text{joint}}(\mathcal{P}_{\mathbf{W}}, \mathcal{P}_{\text{RDM}}) = \frac{1}{2} \left(1 - \frac{q-1}{q} \log_q 2 \right) \quad (5.1)$$

$$C_{2,q}^{\text{simple}}(\mathcal{P}_{\mathbf{W}}, \mathcal{P}_{\text{RDM}}) = \frac{1}{2} \left(1 + \frac{1}{q} \right) \log_q (q+1) - \log_q 2. \quad (5.2)$$

- **Unreadable Digit Model:** UDM allows the coalition to generate erasures but not unseen symbols. For $q = 2$ and 3 , the colluders' optimal attack does not utilize erasures even they are allowed to do so. For $q \geq 4$, $p_{e,2}^{(2,q)}$ is nonzero and randomization of $\mathbf{W}$ is necessary.
- **Arbitrary/General/Combined Digit Model:** In these models $p_{c,2}^{(2,q)}$ and/or $p_{e,2}^{(2,q)}$ are nonzero and randomization of $\mathbf{W}$ is necessary for some cases for $q = 3$.

Numerical solutions for the joint capacity game for binary and ternary alphabets are shown in Table 5.1. The simple capacity game has similar results. We see that ADM does allow stronger attacks than UDM, which essentially says that the coalition would prefer generating unseen symbols over erasures. A simple way to explain this is that when observing an erasure symbol, the decoder learns that (for sure or with high probability) the colluders have more than one symbol in that coordinate. On the other hand, generating an unseen symbol can divert the decoder to accuse some innocent users. Such preference can also be seen by observing the solutions to CDM with different parameters.

5.2 Asymptotics for Large Coalitions

The analysis in Sec. 5.1 allows us to solve the fingerprinting games for small coalitions. However, the solution itself has cardinality in the order of k^{q-1} . For the binary case this grows linearly with k and we showed the solutions for up to tens of pirates [7]. For the nonbinary case, the curse of dimensionality makes the task difficult even for small k . Therefore good approximations of the games' solutions for large-coalition asymptotics ($k \rightarrow \infty$) are important. Independently of our work, Boesten and Škorić [41] extended our asymptotic results in [7] to nonbinary alphabets under RDM. The analysis showed that

Table 5.1: Saddle-Point Solutions for the Joint Capacity Game: Coalition Size $k = 2$ and Alphabet Size $q = 2, 3$

q	Model	$C_{2,q}^{\text{joint}}$	Embed distribution		Collusion channel			
			$\mathbf{w}^{(2,q)'} $	$p_{\mathbf{w}}^{(2,q)}$	$p_{c,1}^{(2,q)}$	$p_{c,2}^{(2,q)}$	$p_{e,1}^{(2,q)}$	$p_{e,2}^{(2,q)}$
2	(R,U,A,G)DM	.25	[.5, .5]	1	0	0	0	0
	CDM ₁	.18	[.5, .5]	1	.01	.01	.20	.20
	CDM ₂	.16	[.5, .5]	1	.05	.05	.06	.06
	CDM ₃	.13	[.5, .5]	1	.10	.10	.03	.03
3	(R,U)DM	.29	[.33, .33, .33]	1	0	0	0	0
	*(A,G)DM	.19	[.06, .47, .47]	.33	0	.17	0	0
	CDM ₁	.19	[.33, .33, .33]	1	.01	.01	.20	.48
	CDM ₂	.18	[.33, .33, .33]	1	.05	.05	.06	.22
	CDM ₃	.14	[.33, .33, .33]	1	.10	.10	.03	.10

***Note:** The optimal distribution takes probability mass $1/3$ at permutations of $\mathbf{w}^{(2,q)}$ shown. All other rows the optimal distributions take $\mathbf{w}^{(2,q)}$ with probability one.

the asymptotic joint capacity is $(q - 1)/(2k^2 \ln q)$ but it did not reveal the optimal strategies. Here we show how to extend our analysis of [7] to the more general combined digit model for both the joint and simple capacity games and derive optimal strategies. Our main results are the following three theorems and two corollaries:

Theorem 5.1. *Assume that Condition 5.1 (defined below) is satisfied, then we have*

$$C_{k,q}^{\text{joint}}(\mathcal{P}^e, \mathcal{P}^c) \sim C_{k,q}^{\text{simple}}(\mathcal{P}^e, \mathcal{P}^c) \sim \frac{A(\mathcal{P}^e, \mathcal{P}^c)}{2k^2 \ln q} \quad (5.3)$$

when $A(\mathcal{P}^e, \mathcal{P}^c)$ (specified in (5.16)) is nonzero.¹ Note that A depends on the embedding distribution class $\mathcal{P}^e$ and the collusion class $\mathcal{P}^c$ and is independent of the coalition size k .

Theorem 5.2. *The binary ($q = 2$) fingerprinting capacities under CDM with $0 \leq r_{e,1} = r_{e,2} \leq 1$ and $0 \leq r_{c,1} < (1 - r_{e,1})/2$ satisfy*

$$C_{k,2}^{\text{joint}}(\mathcal{P}_{\text{CDM}}) \sim C_{k,2}^{\text{simple}}(\mathcal{P}_{\text{CDM}}) \sim \frac{(1 - r_{e,1}) \left(1 - \frac{4}{\pi} \sin^{-1} \sqrt{\frac{r_{c,1}}{1 - r_{e,1}}}\right)^2}{2k^2 \ln 2}. \quad (5.4)$$

¹With a slight abuse of notation, we use the same notation $\mathcal{P}^c$ as in Sec. 2 where it is defined for a fixed coalition size k . The collusion class $\mathcal{P}^c$ here represents a sequence of collusion channels as $k \rightarrow \infty$ and is in general non-compact. A detailed explanation of the compactness of $\mathcal{P}^c$ is provided in Sec. 5.2.3

Furthermore, the maximizing and minimizing strategies that achieve the asymptotic capacity value are respectively the arcsine distribution $dP_{\mathbf{W}}^*(w_0, w_1) \propto (w_0 w_1)^{-1/2} d\mathbf{w}$ and $\boldsymbol{\theta}(\mathbf{z})$ defined by

$$\theta_0(\mathbf{z}) = (1 - r_{e,1}) \cos^2 \gamma_1 \quad (5.5a)$$

$$\theta_1(\mathbf{z}) = (1 - r_{e,1}) \sin^2 \gamma_1 \quad (5.5b)$$

$$\theta_e(\mathbf{z}) = r_{e,1} \quad (5.5c)$$

where $\gamma_1 = (1 - \frac{4}{\pi} \gamma_{\max}) \phi_1 + \gamma_{\max}$, $\phi_1 = \sin^{-1} \sqrt{\frac{z_1}{k}}$, and $\gamma_{\max} = \sin^{-1} \sqrt{\frac{r_{c,1}}{1-r_{e,1}}}$. For $0 \leq r_{e,1} = r_{e,2} \leq 1$ but $r_{c,1} \geq (1 - r_{e,1})/2$, we have $C_{k,2}^{\text{joint}}(\mathcal{P}_{\text{CDM}}) = C_{k,2}^{\text{simple}}(\mathcal{P}_{\text{CDM}}) = 0$ and the minimizing strategy is $\boldsymbol{\theta}(\mathbf{z}) \equiv [(1 - r_{e,1})/2, (1 - r_{e,1})/2, r_{e,1}]'$ for all $\mathbf{z}$.

Corollary 5.3. For $q = 2$, RDM is the special case of CDM with $r_{e,1} = r_{e,2} = r_{c,1} = 0$, under which we have

$$C_{k,2}^{\text{joint}}(\mathcal{P}_{\text{RDM}}) \sim C_{k,2}^{\text{simple}}(\mathcal{P}_{\text{RDM}}) \sim \frac{1}{2k^2 \ln 2}. \quad (5.6)$$

The arcsine distribution $dP_{\mathbf{W}}^*(w_0, w_1) \propto (w_0 w_1)^{-1/2} d\mathbf{w}$ and the interleaving attack

$$\boldsymbol{\theta}^*(\mathbf{z}) = \begin{bmatrix} \frac{1}{k} \mathbf{z} \\ 0 \end{bmatrix} \quad (5.7)$$

are the respective maximizing and minimizing strategies that achieve the asymptotic capacity value.

Theorem 5.4. The fingerprinting capacities under CDM with $0 \leq r_{e,1} = \dots = r_{e,q} \leq 1$ and $\mathbf{r}_c = \mathbf{0}$ satisfies

$$C_{k,q}^{\text{joint}}(\mathcal{P}_{\text{CDM}}) \sim C_{k,q}^{\text{simple}}(\mathcal{P}_{\text{CDM}}) \sim \frac{(1 - r_{e,1})(q - 1)}{2k^2 \ln q}. \quad (5.8)$$

Furthermore, $dP_{\mathbf{W}}^*(\mathbf{w}) \propto \prod_x w_x^{-1/2} d\mathbf{w}$ and

$$\boldsymbol{\theta}(\mathbf{z}) = \begin{bmatrix} \frac{1-r_{e,1}}{k} \mathbf{z} \\ r_{e,1} \end{bmatrix} \quad (5.9)$$

are the respective maximizing and minimizing strategies that achieve the asymptotic capacity value.

Corollary 5.5. *RDM is the special case of CDM with $\mathbf{r}_e = \mathbf{0}$ and $\mathbf{r}_c = \mathbf{0}$, under which we have*

$$C_{k,q}^{\text{joint}}(\mathcal{P}_{\text{RDM}}) \sim C_{k,q}^{\text{simple}}(\mathcal{P}_{\text{RDM}}) \sim \frac{q-1}{2k^2 \ln q}. \quad (5.10)$$

The distribution $dP_{\mathbf{w}}^*(\mathbf{w}) \propto \prod_x w_x^{-1/2} d\mathbf{w}$ and the interleaving attack

$$\boldsymbol{\theta}^*(\mathbf{z}) = \begin{bmatrix} \frac{1}{k}\mathbf{z} \\ 0 \end{bmatrix} \quad (5.11)$$

are the respective maximizing and minimizing strategies that achieve the asymptotic capacity value.

5.2.1 Continuous Functional Game Approximation

We consider the sequence of joint/simple fingerprinting games. Similarly to [7], we impose a regularity condition on the collusion channel Θ . The q -dimensional simplex is denoted by $\mathcal{G} \triangleq \{\mathbf{g} \in \mathbb{R}^{q+1} : \|\mathbf{g}\|_1 = 1, \mathbf{g} \geq 0\}$.

Condition 5.1. *There exists a bounded and twice differentiable function $\mathbf{g} : \mathcal{W} \rightarrow \mathcal{G}$ such that²*

$$\boldsymbol{\theta}(\mathbf{z}) = \mathbf{g}\left(\frac{1}{k}\mathbf{z}\right) + o\left(\frac{1}{k}\right) \quad (5.12)$$

as $k \rightarrow \infty$.

The condition assumes a smooth limiting function of $\mathbf{z}$ for the sequence of collusion channels. For the binary case, the assumption seems valid for the joint capacity game but not for the simple capacity game [7]. However, as the simple decoding scheme is suboptimal, this assumption does not seem overly restrictive.

Under the regularity constraint, the following theorem is instrumental to prove Theorem 5.1. It reduces the payoff functions for the joint/simple capacity games to a functional $T[\mathbf{w}, \mathbf{g}, \nabla \mathbf{g}]$ where $\nabla \mathbf{g}$ is the $(q+1) \times q$ Jacobian matrix of $\mathbf{g}$ versus $\mathbf{w}$. The proof is given in Appendix A.2. Denote by $\nabla g_y(\mathbf{w})$ the gradient row vector of g_y with respect to $\mathbf{w}$ and by $\text{diag}(\mathbf{w})$ the diagonal matrix whose diagonal entries are the components of $\mathbf{w}$.

²The little-o notation is defined uniformly over all $\mathbf{z} \in \mathcal{Z}$.

Theorem 5.6. *Assume that Condition 5.1 is satisfied, then*

$$I_{k,q}^{\text{joint}}(\mathbf{w}, \boldsymbol{\Theta}) \sim I_{k,q}^{\text{simple}}(\mathbf{w}, \boldsymbol{\Theta}) \sim \frac{T[\mathbf{w}, \mathbf{g}, \nabla \mathbf{g}]}{2k^2 \ln q} \quad (5.13)$$

as $k \rightarrow \infty$ where

$$T[\mathbf{w}, \mathbf{g}, \nabla \mathbf{g}] \triangleq \sum_{y \in \mathcal{Y}} \frac{\nabla g_y(\mathbf{w}) [\text{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}'] \nabla' g_y(\mathbf{w})}{g_y(\mathbf{w})} \quad (5.14)$$

$$= \sum_{y \in \mathcal{Y}} \frac{1}{g_y(\mathbf{w})} \left\{ \sum_{x \in \mathcal{Q}} w_x \left(\frac{\partial g_y(\mathbf{w})}{\partial w_x} \right)^2 - \left(\sum_{x \in \mathcal{Q}} w_x \frac{\partial g_y(\mathbf{w})}{\partial w_x} \right)^2 \right\}. \quad (5.15)$$

5.2.2 Change of Variables

By Theorem 5.6 it is straightforward to prove Theorem 5.1 if we let $A(\mathcal{P}^e, \mathcal{P}^c)$ be the value of the following functional maximin game:³

$$A(\mathcal{P}^e, \mathcal{P}^c) \triangleq \sup_{P_{\mathbf{W}}} \inf_{\mathbf{g}} \int_{\mathcal{W}} T[\mathbf{w}, \mathbf{g}, \nabla \mathbf{g}] dP_{\mathbf{W}}(\mathbf{w}) \quad (5.16)$$

where the maximization is subject to $P_{\mathbf{W}} \in \mathcal{P}^e$ and the minimization is subject to constraints of $\mathcal{P}^c$. Before solving the functional maximin game of (5.16), we first simplify the payoff functional T by making the following two changes of variables. The first transforms $(\mathbf{w}, \mathbf{g})$ on the product of simplices $\mathcal{W} \times \mathcal{G}$ to $(\mathbf{v}, \mathbf{h})$ on the product of hyperspheres $\mathcal{V} \times \mathcal{H}$ defined by $\mathcal{V} \triangleq \{\mathbf{v} \in \mathbb{R}^q : \|\mathbf{v}\|_2 = 1, \mathbf{v} \geq \mathbf{0}\}$ and $\mathcal{H} \triangleq \{\mathbf{h} \in \mathbb{R}^{q+1} : \|\mathbf{h}\|_2 = 1, \mathbf{h} \geq \mathbf{0}\}$, which is similar to that of [41, Sec. 3.4]. The second further transforms $(\mathbf{v}, \mathbf{h})$ into hyperspherical coordinates $(\boldsymbol{\phi}, \boldsymbol{\gamma})$, which reduces T to the square of the translated Frobenius norm of the Jacobian matrix $\nabla \boldsymbol{\gamma}$ of $\boldsymbol{\gamma}$ versus $\boldsymbol{\phi}$.

Let $v_x = \sqrt{w_x}$ and $h_y = \sqrt{g_y}$ for all $x \in \mathcal{Q}$ and $y \in \mathcal{Y}$ and we have

$$\frac{\partial g_y}{\partial w_x} = \frac{\partial g_y}{\partial h_y} \cdot \frac{\partial h_y}{\partial v_x} \cdot \frac{\partial v_x}{\partial w_x} = \frac{h_y}{v_x} \cdot \frac{\partial h_y}{\partial v_x}, \quad x \in \mathcal{Q}, y \in \mathcal{Y}.$$

³Note that we change “max-min” to “sup-inf” since the collusion class $\mathcal{P}^c$ is in general non-compact in the asymptotic game. Therefore saddle-point solutions may not exist for these cases.

Therefore

$$\begin{aligned}
T[\mathbf{w}, \mathbf{g}, \nabla \mathbf{g}] &= \sum_{y \in \mathcal{Y}} \frac{1}{g_y(\mathbf{w})} \left\{ \sum_{x \in \mathcal{Q}} w_x \left(\frac{\partial g_y}{\partial w_x} \right)^2 - \left(\sum_{x \in \mathcal{Q}} w_x \frac{\partial g_y}{\partial w_x} \right)^2 \right\} \\
&= \sum_{y \in \mathcal{Y}} \frac{1}{h_y^2(\mathbf{v})} \left\{ \sum_{x \in \mathcal{Q}} v_x^2 \left(\frac{h_y}{v_x} \cdot \frac{\partial h_y}{\partial v_x} \right)^2 - \left(\sum_{x \in \mathcal{Q}} v_x \cdot \frac{h_y}{v_x} \cdot \frac{\partial h_y}{\partial v_x} \right)^2 \right\} \\
&= \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{Q}} \left(\frac{\partial h_y}{\partial v_x} \right)^2 - \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{Q}} v_x \frac{\partial h_y}{\partial v_x} \right)^2 \\
&= \|\nabla \mathbf{h}(\mathbf{v})\|_F^2 - \|\nabla(\mathbf{h}(\mathbf{v}))\mathbf{v}\|_2^2 \triangleq T[\mathbf{v}, \mathbf{h}, \nabla \mathbf{h}]
\end{aligned} \tag{5.17}$$

where $\mathbf{v} \triangleq [v_0, \dots, v_{q-1}]' \in \mathbb{R}^q$, $\mathbf{h} \triangleq [h_0, \dots, h_{q-1}, h_q = h_e]' \in \mathbb{R}^{q+1}$, $\nabla \mathbf{h}$ is the $(q+1) \times q$ Jacobian matrix of $\mathbf{h}$ versus $\mathbf{v}$, $\|\cdot\|_F$ is the Frobenius norm, and $\|\cdot\|_2$ is the Euclidean norm. The constraints $\mathbf{w} \in \mathcal{W}$ and $\mathbf{g} \in \mathcal{G}$ translate to $\mathbf{v} \in \mathcal{V}$ and $\mathbf{h} \in \mathcal{H}$ where $\mathcal{V}$ and $\mathcal{H}$ are respectively the $(q-1)$ -dimensional hypersphere $\mathcal{V} \triangleq \{\mathbf{v} \in \mathbb{R}^q : \|\mathbf{v}\|_2 = 1, \mathbf{v} \geq \mathbf{0}\}$ and the q -dimensional hypersphere $\mathcal{H} \triangleq \{\mathbf{h} \in \mathbb{R}^{q+1} : \|\mathbf{h}\|_2 = 1, \mathbf{h} \geq \mathbf{0}\}$.

We then transform $(\mathbf{v}, \mathbf{h})$ to hyperspherical coordinates $(\boldsymbol{\phi}, \boldsymbol{\gamma})$. Let $\mathbf{v}_i^j, 0 \leq i \leq j \leq q-1$ be the vector $[v_i, \dots, v_j]' \in \mathbb{R}^{j-i+1}$, $\boldsymbol{\phi} \triangleq [r = \phi_0, \phi_1, \dots, \phi_{q-1}]' \in \mathbb{R}^q$ be defined as

$$r = \|\mathbf{v}\|_2 \tag{5.18a}$$

$$\phi_i = \cot^{-1} \frac{v_{i-1}}{\|\mathbf{v}_i^{q-1}\|_2}, \quad 1 \leq i \leq q-1 \tag{5.18b}$$

which has the inverse transformation

$$v_0 = r \cos \phi_1 \tag{5.19a}$$

$$v_1 = r \sin \phi_1 \cos \phi_2 \tag{5.19b}$$

$$\vdots$$

$$v_{q-2} = r \sin \phi_1 \cdots \sin \phi_{q-2} \cos \phi_{q-1} \tag{5.19c}$$

$$v_{q-1} = r \sin \phi_1 \cdots \sin \phi_{q-2} \sin \phi_{q-1}, \tag{5.19d}$$

and $\boldsymbol{\gamma} \triangleq [R = \gamma_0, \gamma_1, \dots, \gamma_{q-1}, \gamma_q]' \in \mathbb{R}^{q+1}$ be defined as

$$R = \|\mathbf{h}_0^{q-1}\|_2 \quad (5.20a)$$

$$\gamma_j = \cot^{-1} \frac{h_{j-1}}{\|\mathbf{h}_j^{q-1}\|_2}, \quad 1 \leq j \leq q-1 \quad (5.20b)$$

$$\gamma_q = h_e \quad (5.20c)$$

which has the inverse transformation

$$h_0 = R \cos \gamma_1 \quad (5.21a)$$

$$h_1 = R \sin \gamma_1 \cos \gamma_2 \quad (5.21b)$$

$$\vdots$$

$$h_{q-2} = R \sin \gamma_1 \cdots \sin \gamma_{q-2} \cos \gamma_{q-1} \quad (5.21c)$$

$$h_{q-1} = R \sin \gamma_1 \cdots \sin \gamma_{q-2} \sin \gamma_{q-1} \quad (5.21d)$$

$$h_e = \gamma_q. \quad (5.21e)$$

Now since $\nabla \mathbf{h} = \mathbf{J}_\gamma^{\mathbf{h}} \nabla \boldsymbol{\gamma} \mathbf{J}_\mathbf{v}^\phi$ where $\mathbf{J}_\gamma^{\mathbf{h}}$ is the $(q+1) \times (q+1)$ Jacobian matrix of $\mathbf{h}$ versus $\boldsymbol{\gamma}$, $\mathbf{J}_\mathbf{v}^\phi$ is the $q \times q$ Jacobian matrix of $\boldsymbol{\phi}$ versus $\mathbf{v}$, and $\nabla \boldsymbol{\gamma}$ is the $(q+1) \times q$ Jacobian matrix of $\boldsymbol{\gamma}$ versus $\boldsymbol{\phi}$, we have

$$\begin{aligned} & T[\mathbf{v}, \mathbf{h}, \nabla \mathbf{h}] \\ &= \|\nabla \mathbf{h}\|_F^2 - \|(\nabla \mathbf{h})\mathbf{v}\|_2^2 \\ &= \text{tr}(\nabla \mathbf{h}' \nabla \mathbf{h}) - \mathbf{v}' (\nabla \mathbf{h})' (\nabla \mathbf{h}) \mathbf{v} \\ &= \text{tr} \left[\left(\mathbf{J}_\mathbf{v}^\phi \right)' (\nabla \boldsymbol{\gamma})' \left(\mathbf{J}_\gamma^{\mathbf{h}} \right)' \mathbf{J}_\gamma^{\mathbf{h}} \nabla \boldsymbol{\gamma} \mathbf{J}_\mathbf{v}^\phi \right] - \mathbf{v}' \left(\mathbf{J}_\mathbf{v}^\phi \right)' (\nabla \boldsymbol{\gamma})' \left(\mathbf{J}_\gamma^{\mathbf{h}} \right)' \mathbf{J}_\gamma^{\mathbf{h}} \nabla \boldsymbol{\gamma} \mathbf{J}_\mathbf{v}^\phi \mathbf{v} \\ &= \text{tr} \left[(\nabla \boldsymbol{\gamma})' \left(\mathbf{J}_\gamma^{\mathbf{h}} \right)' \mathbf{J}_\gamma^{\mathbf{h}} \nabla \boldsymbol{\gamma} \mathbf{J}_\mathbf{v}^\phi \left(\mathbf{J}_\mathbf{v}^\phi \right)' \right] - \text{tr} \left[(\nabla \boldsymbol{\gamma})' \left(\mathbf{J}_\gamma^{\mathbf{h}} \right)' \mathbf{J}_\gamma^{\mathbf{h}} \nabla \boldsymbol{\gamma} \mathbf{J}_\mathbf{v}^\phi \mathbf{v} \mathbf{v}' \left(\mathbf{J}_\mathbf{v}^\phi \right)' \right] \\ &= \text{tr} \left\{ (\nabla \boldsymbol{\gamma})' \left(\mathbf{J}_\gamma^{\mathbf{h}} \right)' \mathbf{J}_\gamma^{\mathbf{h}} \nabla \boldsymbol{\gamma} \left[\mathbf{J}_\mathbf{v}^\phi \left(\mathbf{J}_\mathbf{v}^\phi \right)' - \mathbf{J}_\mathbf{v}^\phi \mathbf{v} \mathbf{v}' \left(\mathbf{J}_\mathbf{v}^\phi \right)' \right] \right\} \triangleq T[\boldsymbol{\phi}, \boldsymbol{\gamma}, \nabla \boldsymbol{\gamma}] \end{aligned}$$

where $\text{tr}(\mathbf{A})$ denotes the trace of square matrix $\mathbf{A}$. By Lemma A.5 we have

$$\begin{aligned} & (\mathbf{J}_\gamma^{\mathbf{h}})' \mathbf{J}_\gamma^{\mathbf{h}} \\ &= \text{diag} \left(1, R^2, R^2 \sin^2 \gamma_1, R^2 \sin^2 \gamma_1 \sin^2 \gamma_2, \dots, R^2 \prod_{l=1}^{q-2} \sin^2 \gamma_l, 1 \right) \end{aligned} \quad (5.22)$$

$$\begin{aligned} & \mathbf{J}_v^\phi (\mathbf{J}_v^\phi)' - \mathbf{J}_v^\phi \mathbf{v} \mathbf{v}' \\ &= \text{diag} \left(1 - r^2, r^{-2}, r^{-2} \sin^{-2} \phi_1, r^{-2} \sin^{-2} \phi_1 \sin^{-2} \phi_2, \dots, r^{-2} \prod_{l=1}^{q-2} \sin^{-2} \phi_l \right). \end{aligned} \quad (5.23)$$

By incorporating the constraint $r = \|\mathbf{v}\|_2 \equiv 1$ we have

$$\begin{aligned} T[\phi, \gamma, \nabla \gamma] &= \text{tr} \left[(\nabla \gamma)' \tilde{\Gamma}^2 \nabla \gamma \tilde{\Phi}^{-2} \right] \\ &= \left\| \tilde{\Gamma} \nabla \gamma \tilde{\Phi}^{-1} \right\|_F^2 \end{aligned} \quad (5.24a)$$

$$= \sum_{\substack{1 \leq i \leq q-1 \\ 0 \leq j \leq q}} \left(\frac{\partial \gamma_j}{\partial \phi_i} \cdot \frac{\tilde{\gamma}_j}{\tilde{\phi}_i} \right)^2 \quad (5.24b)$$

where

$$\begin{aligned} \tilde{\Gamma} &\triangleq \text{diag}(\tilde{\gamma}_0, \dots, \tilde{\gamma}_q) \\ &= \text{diag} \left(1, R, R \sin \gamma_1, R \sin \gamma_1 \sin \gamma_2, \dots, R \prod_{j=1}^{q-2} \sin \gamma_j, 1 \right) \end{aligned} \quad (5.25)$$

$$\begin{aligned} \tilde{\Phi}^{-1} &\triangleq \text{diag}(0, \tilde{\phi}_1^{-1}, \dots, \widetilde{\phi_{q-1}}^{-1}) \\ &= \text{diag} \left(0, 1, \sin^{-1} \phi_1, \sin^{-1} \phi_1 \sin^{-1} \phi_2, \dots, \prod_{i=1}^{q-2} \sin^{-1} \phi_i \right). \end{aligned} \quad (5.26)$$

The constraints $\mathbf{v} \in \mathcal{V}$ and $\mathbf{h} \in \mathcal{H}$ translate to $\phi \in \Phi$ and $\gamma \in \Gamma$ where Φ and Γ are respectively the $(q-1)$ -dimensional space $\Phi \triangleq \{\phi \in \mathbb{R}^q : r = 1, 0 \leq \phi_i \leq \pi/2, 1 \leq i \leq q-1\}$ and the q -dimensional space $\Gamma \triangleq \{\gamma \in \mathbb{R}^{q+1} : R^2 + \gamma_q^2 = 1, 0 \leq \gamma_j \leq \pi/2, 1 \leq j \leq q-1\}$.

5.2.3 Solving the Continuous Functional Game

Sec. 5.2.2 reduces the continuous functional game of (5.16) to

$$A(\mathcal{P}^e, \mathcal{P}^c) = \sup_{P_{\Phi}} \inf_{\gamma} \int_{\Phi} T[\phi, \gamma, \nabla \gamma] dP_{\Phi}(\phi) \quad (5.27)$$

where T is defined in (5.24) and maximization and minimization are subject to constraints on P_{Φ} and γ corresponding to $P_{\mathbf{W}} \in \mathcal{P}^e$ and $\mathbf{g} \in \mathcal{P}^c$.

In Chap. 4 we solve the game (5.16) for the binary case ($q = 2$) under the marking assumption (which is equivalent to RDM, UDM, ADM, and GDM) without the change of variables trick in Sec. 5.2.2. The minimization problem of (5.16) is solved explicitly for fixed $P_{\mathbf{W}}$ using the Cauchy-Schwarz inequality, and the maximization problem can then be solved using another Cauchy-Schwarz inequality. Indeed since the game of (5.16) has a convex payoff functional with respect to $\mathbf{g}$ (analogously to Theorem 3.6), there exists a saddle-point solution pair $(\mathbf{P}_{\mathbf{W}}^*, \mathbf{g}^*)$, which are the arcsine distribution and the interleaving attack for the binary case. They are the maximizing and minimizing strategies for the capacity game.

For the binary and nonbinary case under the general combined digit model, the Cauchy-Schwarz trick does not yield the analytic solution in general. However, convexity of the payoff functional and the saddle-point property still hold. Thus if we can find a minimizer $\gamma(\phi)$ for a fixed P_{Φ}^* such that T is a constant for all ϕ (such $\gamma(\phi)$ is also called an “equalizing strategy”), then $(P_{\Phi}^*, \gamma(\phi))$ is the maximizing-minimizing saddle-point strategy pair, and the value $A(\mathcal{P}^e, \mathcal{P}^c)$ of the game equals the constant value of T . We use variational calculus to solve the minimization problem where a set of Euler-Lagrange differential equations, Lagrange multipliers, and boundary conditions are specified as necessary and sufficient conditions. Saddle-point solutions can be obtained analytically for various cases. For details of the method of variational calculus see [42, 43].

Binary Alphabet with Marking Assumption

We start with the simplest case where $q = 2$, $\mathbf{r}_e = \mathbf{0}$, and $\mathbf{r}_c = \mathbf{0}$. Instead of solving the game (5.16) as in [7, Sec. V-A] we solve the much simpler transformed game (5.27). We fix the uniform distribution $dP_{\Phi}^*(\phi) \propto d\phi$

which is equivalent to $dP_{\mathbf{W}}^*(w_0, w_1) \propto (w_0 w_1)^{-1/2} d\mathbf{w}$. The constraints of the marking assumption translate to $R \equiv 0$ and $\gamma_2 \equiv 0$ and therefore we have the payoff function

$$T[\phi_1, \gamma, \nabla \gamma] = \left(\frac{\partial R}{\partial \phi_1} \right)^2 + R^2 \left(\frac{\partial \gamma_1}{\partial \phi_1} \right)^2 + \left(\frac{\partial \gamma_2}{\partial \phi_1} \right)^2 = \left(\frac{\partial \gamma_1}{\partial \phi_1} \right)^2.$$

The boundary conditions $g_0(w_0 = 0) = 0$ and $g_1(w_1 = 0) = 0$ translate to $\gamma_1(\phi_1 = 0) = 0$ and $\gamma_1(\phi_1 = \pi/2) = \pi/2$. Thus

$$\begin{aligned} \int_{\Phi} T[\phi_1, \gamma, \nabla \gamma] dP_{\Phi}^*(\phi) &\propto \int_0^{\pi/2} \left(\frac{d\gamma_1}{d\phi_1} \right)^2 d\phi_1 \\ &\stackrel{(a)}{\geq} \frac{\left(\int_0^{\pi/2} \frac{d\gamma_1}{d\phi_1} d\phi_1 \right)^2}{\int_0^{\pi/2} 1 \cdot d\phi_1} \\ &\stackrel{(b)}{=} \frac{\pi}{2} \end{aligned}$$

where (a) follows from the Cauchy-Schwarz inequality and (b) follows from the boundary conditions. Equality holds in (a) when $\frac{d\gamma_1^*}{d\phi_1} = 1$ or $\gamma_1^* = \phi_1$, which translates to the interleaving attack of (5.3). The maximin value of the game $A(\mathcal{P}_{\text{RDM}}) = T|\gamma^* \equiv \left(\frac{d\gamma_1^*}{d\phi_1} \right)^2 = 1$. By Theorem 5.1 we establish Corollary 5.3.

Binary Alphabet with Combined Digit Model

We first examine the setup $0 \leq r_{e,1} = r_{e,2} \leq 1$ and $0 \leq r_{c,1} < (1 - r_{e,1})/2$. Again we fix the uniform distribution $dP_{\Phi}^*(\phi) \propto d\phi$. The payoff function can be written as

$$\begin{aligned} T[\phi_1, \gamma, \nabla \gamma] &= \left(\frac{\partial R}{\partial \phi_1} \right)^2 + R^2 \left(\frac{\partial \gamma_1}{\partial \phi_1} \right)^2 + \left(\frac{\partial \gamma_2}{\partial \phi_1} \right)^2 \\ &= (J_1^0)^2 + R^2 (J_1^1)^2 + (J_1^2)^2 \end{aligned}$$

where we use the notation $J_i^j \triangleq \frac{\partial \gamma_j}{\partial \phi_i}$ interchangeably. The constraints of (2.17) can be simplified as $g_e \leq r_{e,1}$, $g_0(w_0 = 0) \leq r_{c,1}$ and $g_1(w_1 = 0) \leq r_{c,1}$, which translate to $\gamma_2 \leq \sqrt{r_{e,1}}$, $\gamma_1(\phi_1 = 0) \leq \gamma_{\max}$ and $\gamma_1(\phi_1 = \pi/2) \geq \pi/2 - \gamma_{\max}$ where $\gamma_{\max} \triangleq \sin^{-1} \sqrt{\frac{r_{c,1}}{1-r_{e,1}}}$. Therefore we are interested in the following

minimization problem:

$$\text{Minimize} \quad \int_0^{\pi/2} T[\phi_1, \gamma, \nabla \gamma] d\phi_1 \quad (5.28a)$$

$$\text{subject to} \quad G_1[\gamma] \triangleq R^2 + \gamma_2^2 - 1 = 0 \text{ and } G_2[\gamma] \triangleq \gamma_2 - \sqrt{r_{e,1}} \leq 0 \quad (5.28b)$$

(global constraints)

$$\gamma_1(\phi_1 = 0) \leq \gamma_{\max} \text{ and } \gamma_1(\phi_1 = \pi/2) \geq \pi/2 - \gamma_{\max} \quad (5.28c)$$

(boundary conditions)

The Cauchy-Schwarz inequality cannot help us here so we turn to variational calculus:

- (i) **Euler-Lagrange Equations:** By [42, Sec. 12], if a functional $\gamma(\phi_1)$ minimizes (5.28a), then there exists $\lambda_1(\phi_1)$ and $\lambda_2(\phi_1)$ satisfying

$$\frac{\partial T}{\partial \gamma_j} + \lambda_1 \frac{\partial G_1}{\partial \gamma_j} + \lambda_2 \frac{\partial G_2}{\partial \gamma_j} = \frac{\partial}{\partial \phi_1} \left(\frac{\partial T}{\partial J_1^j} \right), \quad 0 \leq j \leq 2. \quad (5.29)$$

We verify the functional $\gamma(\phi_1)$ defined by $R \equiv \sqrt{1 - r_{e,1}}$, $\gamma_1 = (1 - \frac{4}{\pi} \gamma_{\max})\phi_1 + \gamma_{\max}$, and $\gamma_2 \equiv \sqrt{r_{e,1}}$ satisfies (5.29) with $\lambda_1 = -(1 - \frac{4}{\pi} \gamma_{\max})^2$ and $\lambda_2 = -\lambda_1 \sqrt{r_{e,1}}$:⁴

$$j = 0: \quad \text{LHS} = 2R (J_1^1)^2 + 2\lambda_1 R = 0 = \frac{\partial}{\partial \phi_1} (2J_1^0) = \text{RHS}$$

$$j = 1: \quad \text{LHS} = 0 = \frac{\partial}{\partial \phi_1} (2R^2 J_1^1) = \text{RHS}$$

$$j = 2: \quad \text{LHS} = 2\lambda_1 \gamma_2 + 2\lambda_2 = 0 = \frac{\partial}{\partial \phi_1} = \text{RHS}$$

- (ii) **Global Constraints:** The inequality constraint $G_2 \leq 0$ is active for the solution so we need $\lambda_2 \geq 0$, which is true since $\lambda_1 < 0$.

- (iii) **Boundary Conditions:** Both boundary conditions are active for the solution so we also need $\frac{\partial T}{\partial J_1^1} \Big|_{\phi_1=0} \geq 0$ and $\frac{\partial T}{\partial J_1^1} \Big|_{\phi_1=\pi/2} \geq 0$, which can be verified by $\frac{\partial T}{\partial J_1^1} = 2R^2 J_1^1 > 0$ ($\because \gamma_{\max} < \pi/4$).

The minimizer $\gamma(\phi_1)$ translates to (5.5) and the maximin value of the game is $A(\mathcal{P}_{\text{CDM}}) = T|_{\gamma} \equiv (1 - r_{e,1}) \left(1 - \frac{4}{\pi} \sin^{-1} \gamma_{\max}\right)^2$. Therefore by Theorem

⁴LHS stands for left-hand side and RHS stands for right-hand side.

5.1 we establish the first part of Theorem 5.2.⁵

If $0 \leq r_{e,1} = r_{e,2} \leq 1$ but $r_{c,1} \geq (1 - r_{e,1})/2$, we can let γ be defined as $R \equiv \sqrt{1 - r_{e,1}}$, $\gamma_1 \equiv \pi/4$, and $\gamma_2 = \sqrt{r_{e,1}}$ which yields $\theta(\mathbf{z}) \equiv [(1 - r_{e,1})/2, (1 - r_{e,1})/2, r_{e,1}]'$ for all $\mathbf{z}$ and $T \equiv 0$, i.e., this is the case where the coalition can generate the same distribution of forgery symbols regardless of what they receive, hence capacity is zero.

One may try to solve the game for the general setup where $0 \leq r_{e,1} < r_{e,2} \leq 1$. However, this results in a non-compact collusion class $\mathcal{P}^c$. More specifically, the constraints (2.17b) alone can be translated to

$$\gamma_2(\phi_1 = 0) \leq \sqrt{r_{e,1}} \quad (5.30a)$$

$$\gamma_2(\phi_1 = \pi/2) \leq \sqrt{r_{e,1}} \quad (5.30b)$$

$$\gamma_2(0 < \phi_1 < \pi/2) \leq \sqrt{r_{e,2}}, \quad (5.30c)$$

which is not a compact space. The minimizer $\gamma(\phi_1)$ above fails to satisfy condition (b) for global constraints, while the conjectured minimizer with γ_2 defined by

$$\gamma_2(\phi_1) = \begin{cases} \sqrt{r_{e,1}}, & \phi_1 = 0, \pi/2 \\ \sqrt{r_{e,2}}, & \phi_1 \in (0, \pi/2) \end{cases}$$

does not satisfy the regularity constraint (Condition 5.1).

General Alphabet with Combined Digit Model

Finally we examine the general alphabet case where $q \geq 2$. We start with the case where $0 \leq r_{e,1} = \dots = r_{e,q} \leq 1$ and $\mathbf{r}_c = \mathbf{0}$. Let $\tilde{\phi}_i = \prod_{l=1}^{i-1} \sin \phi_l$ for $1 \leq i \leq q-1$. We fix the distribution

$$dP_{\Phi}^*(\phi) \propto \prod_{i=1}^{q-1} \tilde{\phi}_i d\phi = \prod_i^{q-1} (\sin \phi_i)^{q-i-1} d\phi \quad (5.31)$$

⁵We note that by [43, Chap. 2] we indeed need all (i-iii) to verify the minimizer $\gamma(\phi)$. As a counter example, the functional $\gamma(\phi_1)$ defined by $R \equiv \sqrt{1 - r_{e,1}}$, $\gamma_1 = \phi_1$, and $\gamma_2 \equiv \sqrt{r_{e,1}}$ satisfy (a) Euler-Lagrange Equations with $\lambda_1 = -1$ and $\lambda_2 = \sqrt{r_{e,1}}$ and (b) $\lambda_2 > 0$, but we also need (c) $\left. \frac{\partial T}{\partial J_1^1} \right|_{\phi_1=0} = 0$ for inactive boundary condition $\gamma_1(\phi_1 = 0) < \gamma_{\max}$, which isn't true.

which is equivalent to $dP_{\mathbf{W}}^*(\mathbf{w}) \propto \prod_x w_x^{-1/2} d\mathbf{w}$. The payoff function can be written as

$$T[\phi, \gamma, \nabla \gamma] = \sum_{\substack{1 \leq i \leq q-1 \\ 0 \leq j \leq q}} \left(\frac{\tilde{\gamma}_j}{\tilde{\phi}_i} \cdot J_i^j \right)^2$$

where

$$\tilde{\gamma}_j = \begin{cases} R \prod_{l=1}^{j-1} \sin \gamma_l, & 1 \leq j \leq q-1 \\ 1, & j = 0, q. \end{cases} \quad (5.32)$$

The constraints of (2.17) can be simplified as $g_e \leq r_{e,1}$ and $g_i(w_i = 0) = 0$ for $0 \leq i \leq q-1$, which translates to $\gamma_q \leq \sqrt{r_{e,1}}$ and $\gamma_i(\phi_i = 0) = 0$, $\gamma_i(\phi_i = \pi/2) = \pi/2$ for $1 \leq i \leq q-1$. Therefore we are interested in the following minimization problem:

$$\text{Minimize} \quad \int_0^{\pi/2} T[\phi, \gamma, \nabla \gamma] dP_{\Phi}^*(\phi) \quad (5.33a)$$

$$\text{subject to} \quad G_1[\gamma] \triangleq R^2 + \gamma_q^2 - 1 = 0 \text{ and } G_2[\gamma] \triangleq \gamma_q - \sqrt{r_{e,1}} \leq 0 \quad (5.33b)$$

(global constraints)

$$\gamma_i(\phi_i = 0) = 0 \text{ and } \gamma_i(\phi_i = \pi/2) = \pi/2 \text{ for } 0 \leq i \leq q-1 \quad (5.33c)$$

(boundary conditions)

The minimizer $\gamma(\phi)$ is given by $R \equiv \sqrt{1 - r_{e,1}}$, $\gamma_q \equiv \sqrt{r_{e,1}}$, and $\gamma_i = \phi_i$ for $1 \leq i \leq q-1$. Again we verify this solution using variational calculus:

(i) **Euler-Lagrange Equations:** By [42, Sec. 12] we verify that the minimizer $\gamma(\phi)$ satisfies

$$\begin{aligned} & \frac{\partial}{\partial \gamma_j} \left(T dP_{\Phi}^*(\phi) \right) + \lambda_1 \frac{\partial G_1}{\partial \gamma_j} dP_{\Phi}^*(\phi) + \lambda_2 \frac{\partial G_2}{\partial \gamma_j} dP_{\Phi}^*(\phi) \\ &= \sum_{i=1}^{q-1} \frac{\partial}{\partial \phi_i} \left[\frac{\partial}{\partial J_i^j} \left(T dP_{\Phi}^*(\phi) \right) \right], \quad 0 \leq j \leq q. \end{aligned} \quad (5.34)$$

with $\lambda_1 = -(q-1)$ and $\lambda_2 = -2\lambda_1 \sqrt{r_{e,1}}$:

- $j = 0$ ($\gamma_0 = R$): we have

$$\begin{aligned}
\text{LHS} &\stackrel{(a)}{=} \left[\sum_{1 \leq i, l \leq q-1} 2 (J_i^l)^2 \cdot \frac{\tilde{\gamma}_l}{\tilde{\phi}_i^2} \cdot \frac{\partial \tilde{\gamma}_l}{\partial R} \right] \cdot dP_{\Phi}^*(\phi) + 2\lambda_1 R \cdot dP_{\Phi}^*(\phi) \\
&\stackrel{(b)}{=} \left(\sum_{i=1}^{q-1} 2 \cdot \frac{\tilde{\gamma}_i}{\tilde{\phi}_i^2} \cdot \frac{\tilde{\gamma}_i}{R} \right) \cdot dP_{\Phi}^*(\phi) + 2\lambda_1 R \cdot dP_{\Phi}^*(\phi) \\
&\stackrel{(c)}{=} 2R(q-1)dP_{\Phi}^*(\phi) + 2\lambda_1 R \cdot dP_{\Phi}^*(\phi) \stackrel{(d)}{=} 0 \\
\text{RHS} &= \sum_{i=1}^{q-1} \frac{\partial}{\partial \phi_i} \left[\frac{2J_i^0}{\tilde{\phi}_i^2} \cdot dP_{\Phi}^*(\phi) \right] \stackrel{(e)}{=} 0
\end{aligned}$$

where (a) follows from (5.32), (b) and (e) follows from $J_i^l = 1$ for $1 \leq i = j \leq q-1$ and $J_i^l = 0$ otherwise, (c) follows from $\tilde{\gamma}_i = R\tilde{\phi}_i$ for $1 \leq i \leq q-1$, and (d) follows from $R = \sqrt{1 - r_{e,1}}$.

- $1 \leq j \leq q-1$: we have

$$\begin{aligned}
\text{LHS} &\stackrel{(a)}{=} \left[\sum_{i=1}^{q-1} \sum_{l=j+1}^{q-1} 2 (J_i^l)^2 \cdot \frac{\tilde{\gamma}_l}{\tilde{\phi}_i^2} \cdot \frac{\partial \tilde{\gamma}_l}{\partial \gamma_j} \right] \cdot dP_{\Phi}^*(\phi) \\
&\stackrel{(b)}{=} \left[\sum_{i=j+1}^{q-1} 2 \cdot \frac{\tilde{\gamma}_i}{\tilde{\phi}_i^2} \cdot \frac{\partial \tilde{\gamma}_i}{\partial \gamma_j} \right] \cdot dP_{\Phi}^*(\phi) \\
&\stackrel{(c)}{=} \left[\sum_{i=j+1}^{q-1} 2 \cdot \frac{\tilde{\gamma}_i^2}{\tilde{\phi}_i^2} \cdot \cot \gamma_j \right] \cdot dP_{\Phi}^*(\phi) \\
&\stackrel{(d)}{=} 2(q-j-1)R^2 \cot \phi_j \cdot dP_{\Phi}^*(\phi) \\
\text{RHS} &= \sum_{i=1}^{q-1} \frac{\partial}{\partial \phi_i} \left[2J_i^j \cdot \frac{\tilde{\gamma}_j^2}{\tilde{\phi}_i^2} \cdot dP_{\Phi}^*(\phi) \right] \\
&\stackrel{(e)}{=} \frac{\partial}{\partial \phi_j} \left(2 \cdot \frac{\tilde{\gamma}_j^2}{\tilde{\phi}_j^2} \cdot dP_{\Phi}^*(\phi) \right) \\
&\stackrel{(f)}{=} \frac{\partial}{\partial \phi_j} \left(2R^2 \cdot dP_{\Phi}^*(\phi) \right) \\
&\stackrel{(g)}{=} 2R^2 \cdot dP_{\Phi}^*(\phi) \cdot (q-j-1) \cot \phi_j
\end{aligned}$$

where (a) follows from (5.32), (b) and (e) follows from the values of J_i^l , (c) follows from $\frac{\partial \tilde{\gamma}_i}{\partial \gamma_j} = \tilde{\gamma}_i \cot \gamma_j$ for $i \geq j+1$, and (d) and (f) follows from $\tilde{\gamma}_i = R\tilde{\phi}_i$ for $1 \leq i \leq q-1$, and (g) follows from

(5.31).

- $j = q$: we have

$$\begin{aligned} \text{LHS} &\stackrel{(a)}{=} 2\lambda_1\gamma_q dP_{\Phi}^*(\phi) + \lambda_2 dP_{\Phi}^*(\phi) = 0 \\ \text{RHS} &= \sum_{i=1}^{q-1} \frac{\partial}{\partial \phi_i} \left[\frac{2J_i^q}{\tilde{\gamma}^2} \cdot dP_{\Phi}^*(\phi) \right] \stackrel{(b)}{=} 0 \end{aligned}$$

where (a) follows from (5.32) and (b) follows from the values of J_i^q .

- (ii) **Global Constraints:** The inequality constraint $G_2 \leq 0$ is active for the solution so we need $\lambda_2 \geq 0$, which is true since $\lambda_1 < 0$.
- (iii) **Boundary Conditions:** All boundary conditions are equalities so we do not need to verify any other conditions.

The minimizer $\gamma(\phi)$ translates to (5.9) and the maximin value of the game is $A(\mathcal{P}_{\text{CDM}}) = T|\gamma| \equiv (q-1)(1-r_{e,1})$. Therefore by Theorem 5.1 we establish Theorem 5.4.

For the case where $\mathbf{r}_c$ is nonzero, unfortunately, a linear function of ϕ does not yield a minimizer as in the binary case. We leave this part as future work. Also similar to the binary case, a general setup where $0 \leq r_{e,1} < \dots < r_{e,q} \leq 1$ yields a non-compact collusion class $\mathcal{P}^c$, which prevents us from finding the minimizer using variational calculus.

CHAPTER 6

SECOND-ORDER CODING RATES FOR MULTIPLE ACCESS CHANNELS

In this chapter we study the second-order coding rates of multiple access channels. In order to follow the conventional notation of MAC, all notation in this chapter is redefined and is independent of that of digital fingerprinting in the previous chapters.

6.1 Definitions

Formally, a two-user MAC¹ is characterized by a transition distribution $P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}$ where $\mathbf{X}_1 \in \mathcal{X}_1^n$, $\mathbf{X}_2 \in \mathcal{X}_2^n$, and $\mathbf{Y} \in \mathcal{Y}^n$, and is called discrete memoryless if $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{Y}$ are finite and $P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) = \prod_{i=1}^n W(y_i|x_{1i}, x_{2i})$. Hence, a DM-MAC is determined by the single-lettered conditional distribution $W(y|x_1, x_2)$. An (n, M_1, M_2) MAC code $\mathcal{C}$ consists of two codebooks $\{\mathbf{x}_1(m_1)\}$ and $\{\mathbf{x}_2(m_2)\}$ where $m_j \in \{1, \dots, M_j\} \triangleq \mathcal{M}_j$ for $j = 1, 2$ and a possibly randomized decoder $\phi(\mathbf{y}) = (\widehat{m}_1, \widehat{m}_2)$. Code rates are defined by $R_j \triangleq n^{-1} \log M_j$ for $j = 1, 2$.

The error probability for passing a message pair (m_1, m_2) using MAC code $\mathcal{C}$ through channel $P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}$ is denoted by

$$P_e(m_1, m_2, P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}, \mathcal{C}) \triangleq \sum_{\mathbf{y} \in \mathcal{Y}^n} P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}(\mathbf{y}|\mathbf{x}_1(m_1), \mathbf{x}_2(m_2)) \Pr\{\phi(\mathbf{y}) \neq (m_1, m_2)\}.$$

The average error probability is defined by

$$P_e^{\text{avg}}(P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}, \mathcal{C}) = \frac{1}{M_1 M_2} \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} P_e(m_1, m_2, P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}, \mathcal{C})$$

¹Generalizations to k -user MAC for $k \geq 3$ are straightforward and thus omitted.

and the maximal error probability is defined by

$$P_e^{\max}(P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}, \mathcal{C}) = \max_{(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} P_e(m_1, m_2, P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}, \mathcal{C}).$$

An (n, M_1, M_2, ϵ) MAC code $\mathcal{C}$ under average (resp. maximal) error probability satisfies $P_e^{\text{avg}}(P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}, \mathcal{C}) \leq \epsilon$ (resp. $P_e^{\max}(P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}, \mathcal{C}) \leq \epsilon$). A rate pair (R_1, R_2) is (n, ϵ) -achievable under average (resp. maximal) error probability if there exists an (n, M_1, M_2, ϵ) MAC code such that $\log M_1 \geq R_1$ and $\log M_2 \geq R_2$ under average (resp. maximal) error probability.

The capacity region of a DM-MAC W is closely related to statistics of the *information density vector* $\mathbf{i}(U, X_1, X_2, W)$ abbreviated by $\mathbf{i}$ (should there be no risk of confusion) and defined by

$$\mathbf{i}(U, X_1, X_2, W) = \begin{bmatrix} i(X_1; Y|X_2, U) \\ i(X_2; Y|X_1, U) \\ i(X_1, X_2; Y|U) \end{bmatrix} = \begin{bmatrix} \log \frac{W(Y|X_1, X_2)}{P(Y|X_2, U)} \\ \log \frac{W(Y|X_1, X_2)}{P(Y|X_1, U)} \\ \log \frac{W(Y|X_1, X_2)}{P(Y|U)} \end{bmatrix}$$

where U is an auxiliary “time sharing” random variable and $U \leftrightarrow (X_1, X_2) \leftrightarrow Y$ forms a Markov chain, i.e.,

$$P_{UX_1X_2Y} = P_U P_{X_1X_2|U} W.$$

The marginal distributions $P(Y|X_1, U)$, $P(Y|X_2, U)$ and $P(Y|U)$ are obtained by marginalizing out X_2 , X_1 , and (X_1, X_2) , respectively. The *mutual information vector* $\mathbf{I}(P_{UX_1X_2}, W)$ or $\mathbf{I}$ is defined by the expectation of the information density vector under $P_{UX_1X_2}W$:

$$\mathbf{I}(P_{UX_1X_2}, W) \triangleq \mathbb{E}[\mathbf{i}(U, X_1, X_2, W)] = \begin{bmatrix} I(X_1; Y|X_2, U) \\ I(X_2; Y|X_1, U) \\ I(X_1, X_2; Y|U) \end{bmatrix}. \quad (6.1)$$

The (n, ϵ) -capacity region $\mathcal{R}_{n, \epsilon}^{\text{avg}}$ (resp. $\mathcal{R}_{n, \epsilon}^{\max}$) under average (resp. maximal) error probability is defined by the set of all (n, ϵ) -achievable rate pairs. For the DM-MAC, as $\epsilon \searrow 0$ and $n \rightarrow \infty$, $\mathcal{R}_{n, \epsilon}^{\text{avg}}$ approaches the MAC capacity

region by Ahlswede [15] and Liao [16]:

$$\mathcal{C}(W) = \bigcup_{P_U P_{X_1|U} P_{X_2|U}} \mathcal{R}(P_U P_{X_1|U} P_{X_2|U}) \quad (6.2)$$

where $\mathcal{R}(P_U P_{X_1|U} P_{X_2|U})$ is the set of all rate pairs $[R_1, R_2]'$ satisfying

$$\mathbf{R} \leq \mathbf{I}(P_{UX_1X_2}, W). \quad (6.3)$$

The rate vector is defined by $\mathbf{R} = [R_1, R_2, R_1 + R_2]'$ and the inequality denotes element-wise inequality. The strong converse by Dueck [18], later sharpened by Ahlswede [19], stated that every sequence of (n, M_1, M_2, ϵ) MAC codes (average error probability) satisfies

$$[\log M_1, \log M_2]' \in n\mathcal{C}(W) + O(\sqrt{n} \log n). \quad (6.4)$$

It should be noted that while $\mathcal{R}_{n,\epsilon}^{\max}$ is in general smaller than $\mathcal{R}_{n,\epsilon}^{\text{avg}}$ even in the first-order term [17], they do coincide if encoder randomization is allowed.²

The following notation will be used throughout the chapter: given a sequence of vectors $\mathbf{a}_n$, a sequence of regions $\mathcal{F}_n$, and a sequence of functions g_n , the asymptotic notation $\mathbf{a}_n \in \mathcal{F}_n + O(g_n)$ means there exists a constant c such that $\mathbf{a}_n \in \mathcal{F}_n + cg_n \mathbf{1}$ for sufficiently large n . Here $\mathbf{1}$ denotes the vector comprised of all ones, and ‘+’ denotes Minkowski set addition.

6.2 Main Results

The main goal of this work is to provide sharp inner and outer bounds on DM-MAC second-order coding rates. Our bounds are associated with the following quantities:

Under $P_{UX_1X_2}W$, we define the *dispersion matrix* $\mathbf{V}(P_{UX_1X_2}, W)$ as

$$\mathbf{V}(P_{UX_1X_2}, W) \triangleq \mathbb{E}_{P_U} \left\{ \text{Cov}_{P_{X_1X_2|U}W} [\mathbf{i}(U, X_1, X_2, W) | U] \right\}. \quad (6.5)$$

Let $\mathbf{X}$ be a d -dimensional Gaussian vector with mean $\mathbf{0}$ and covariance

²In particular, randomized permutation of codeword assignments equalizes the decoding error probability for all possible message pairs.

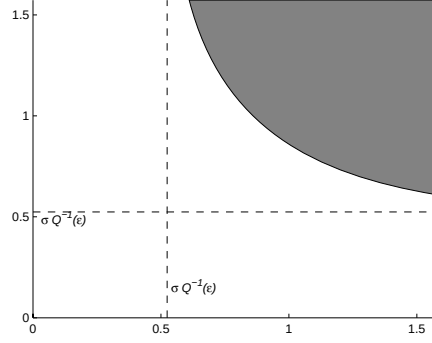


Figure 6.1: The region $\mathcal{Q}_{\text{inv}}(\Sigma, \epsilon)$ for $\Sigma = \begin{bmatrix} 1 & .3 \\ .3 & 1 \end{bmatrix}$ and $\epsilon = .3$.

matrix Σ . For any $\epsilon \in (0, 1)$ let

$$\mathcal{Q}_{\text{inv}}(\Sigma, \epsilon) \triangleq \{\mathbf{x} \in \mathbb{R}^d : \Pr(\mathbf{X} \leq \mathbf{x}) \geq 1 - \epsilon\} \quad (6.6a)$$

$$= \{\mathbf{x} \in \mathbb{R}^d : \Pr(\mathbf{X} \geq -\mathbf{x}) \geq 1 - \epsilon\}. \quad (6.6b)$$

Note that for $d = 1$, we have $\mathcal{Q}_{\text{inv}}(\sigma^2, \epsilon) = \{z \geq \sigma Q^{-1}(\epsilon)\}$ where Q^{-1} is the complementary Gaussian cumulative distribution function. An example of a two-dimensional $\mathcal{Q}_{\text{inv}}$ region is shown in Fig. 6.1.

Fix distributions P and $\{Q^j\}_{j=1}^3$ we consider the ternary composite hypothesis testing (HT) problem:

$$H_0 : Y \sim Q^j, \quad j = 1, 2, 3$$

$$H_1 : Y \sim P.$$

A randomized hypothesis test is denoted by $\delta(y) = \Pr(\text{Say } H_1 | Y = y), y \in \mathcal{Y}$.

Definition 6.1.³ *The ternary hypothesis testing performance region $\mathcal{E}_\epsilon$ is defined as the set of achievable false positive vectors for power $(1 - \epsilon)$ tests:*

$$\begin{aligned} \mathcal{E}_\epsilon \left(P, \{Q^j\}_{j=1}^3 \right) &\triangleq \{[e_1, e_2, e_3]^\top : \exists \text{ test } \delta \text{ s.t.} \\ &\quad \mathbb{E}_P[\delta(Y)] \geq 1 - \epsilon \text{ and } \mathbb{E}_{Q^j}[\delta(Y)] \leq e_j, 1 \leq j \leq 3\}. \end{aligned}$$

This is a generalization of the binary hypothesis testing performance func-

³Note that the definition is different from that of our ISIT paper [7].

tion β_α of [9, (100)], which can also be written as:

$$\beta_\alpha(P, Q) = \min \{ \epsilon' : \exists \text{ test } \delta \text{ s.t. } \mathbb{E}_P[\delta(Y)] \geq \alpha \text{ and } \mathbb{E}_Q[\delta(Y)] \leq \epsilon' \}.$$

Our inner bound characterizes the asymptotics of the (n, ϵ) -capacity region $\mathcal{R}_{n, \epsilon}^{\text{avg}}$ under the average error probability criterion of DM-MAC up to the $O\left(\frac{\log n}{n}\right)$ term. The characterization utilizes the mutual information vector of (6.1), the second-order dispersion matrix $\mathbf{V}$ of (6.5), and the $\mathcal{Q}_{\text{inv}}$ region of (6.6).

Theorem 6.1. *For any fixed distribution $P_{UX_1X_2}$ let*

$$\mathcal{R}_{n, \epsilon}^{\text{in}}(P_{UX_1X_2}, W) \triangleq \mathbf{I}(P_{UX_1X_2}, W) - \frac{1}{\sqrt{n}} \mathcal{Q}_{\text{inv}}(\mathbf{V}(P_{UX_1X_2}, W), \epsilon)$$

and

$$\mathcal{R}_{n, \epsilon}^{\text{in}}(W) \triangleq \cup_{P_U P_{X_1|U} P_{X_2|U}} \mathcal{R}_{n, \epsilon}^{\text{in}}(P_{UX_1X_2}, W). \quad (6.7)$$

Then for sufficiently large n any rate vector $\mathbf{R} \in \mathcal{R}_{n, \epsilon}^{\text{in}}(W) + O\left(\frac{\log n}{n}\right)$ is (n, ϵ) -achievable (average error probability).

Our single-lettered outer bounds are based on the following general outer regions for discrete MAC (not necessarily memoryless). We say $\mathcal{C}'$ an (n, Ω, ϵ) MAC subcode of $\mathcal{C}$ under maximal error probability if $\forall (m_1, m_2) \in \Omega \subseteq \mathcal{M}_1 \times \mathcal{M}_2$ we have $P_e(m_1, m_2, P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}, \mathcal{C}) \leq \epsilon$.

Theorem 6.2. *Let $\mathcal{C}$ be an (n, M_1, M_2, ϵ) MAC code (maximal error probability) and let Ω be a subset of the set of message pairs $\mathcal{M}_1 \times \mathcal{M}_2$. Then*

$$|\Omega| \leq \inf_{Q_{\mathbf{Y}}} \max_{(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{C}(\Omega)} [\beta_{1-\epsilon}(P_{\mathbf{Y}|\mathbf{X}_1=\mathbf{x}_1, \mathbf{X}_2=\mathbf{x}_2}, Q_{\mathbf{Y}})]^{-1} \quad (6.8)$$

where $Q_{\mathbf{Y}}$ ranges over all distributions on $\mathcal{Y}^n$.

Theorem 6.3. *Given an $(n, \bar{\Omega}, \epsilon)$ MAC subcode $\mathcal{C}'$ (maximal error probability) of $\mathcal{C}$ for channel $P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}$, a subset Ω of $\bar{\Omega}$ such that $|\bar{\Omega}| \geq |\Omega| \geq$*

$3 \max \{M_1, M_2\}$. Then we have

$$\left[\frac{3M_2}{|\Omega|}, \frac{3M_1}{|\Omega|}, \frac{3}{|\Omega|} \right] \\ \in \bigcap_{Q_{\mathbf{Y}|\mathbf{X}_2}^1, Q_{\mathbf{Y}|\mathbf{X}_1}^2, Q_{\mathbf{Y}}^3} \bigcup_{(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{C}(\Omega)} \mathcal{E}_\epsilon (P_{\mathbf{Y}|\mathbf{X}_1=\mathbf{x}_1, \mathbf{X}_2=\mathbf{x}_2}, \{Q_{\mathbf{Y}|\mathbf{X}_2=\mathbf{x}_2}^1, Q_{\mathbf{Y}|\mathbf{X}_1=\mathbf{x}_1}^2, Q_{\mathbf{Y}}^3\})$$

where $Q_{\mathbf{Y}|\mathbf{X}_2}^1$, $Q_{\mathbf{Y}|\mathbf{X}_1}^2$, and $Q_{\mathbf{Y}}^3$ range over all distributions on $\mathcal{Y}^n$.

6.3 Inner Region

In this section, we prove the characterization of the asymptotic inner bound of Theorem 6.1 for the (n, ϵ) -capacity region $\mathcal{R}_{n, \epsilon}^{\text{avg}}$ under the average error probability criterion of DM-MAC. For any fixed distribution $P_{U X_1 X_2} = P_U P_{X_1|U} P_{X_2|U}$, we consider the following random codes:

Encoder: Let $\{u_i\}_{i=1}^n, n \geq 1$ be deterministic sequences such that for large enough n the empirical distribution $\hat{P}_U(u) \triangleq \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{u_i=u\}}$ approaches $P_U(u)$ in $O(1/n)$, i.e.,

$$\|\hat{P}_U - P_U\|_1 = O(1/n) \quad (6.9)$$

where $\|\cdot\|_1$ denotes the L_1 norm. For each message pair $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2$, draw X_{ji} independently from $P_{X_j|U}(\cdot|u_i)$ for $1 \leq i \leq n, j = 1, 2$.

Decoder: Define the (variable-size) list $\mathcal{L}$ of all messages (m_1, m_2) whose score exceeds a prescribed threshold vector $\boldsymbol{\tau}_n = [\tau_{1,n}, \tau_{2,n}, \tau_{3,n}]'$:

$$\mathcal{L} \triangleq \{(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2 : \mathbf{Z}(m_1, m_2) \geq \boldsymbol{\tau}_n\}$$

where $\mathbf{Z}(m_1, m_2) \triangleq \sum_{i=1}^n \mathbf{Z}_i(m_1, m_2)$ and

$$\mathbf{Z}_i(m_1, m_2) = \mathbf{i}(u_i, X_{1i}(m_1), X_{2i}(m_2), W), \quad 1 \leq i \leq n. \quad (6.10)$$

Analysis of the threshold decoding scheme is based on a generalized multi-dimensional Berry-Esseen theorem [44], which approximates the distribution of the sum of independent random vectors to a multivariate Gaussian distribution, and large deviations analysis.

Without loss of generality, we assume $m_1 = m_2 = 1$. The analysis below

shows that there exists a threshold vector $\boldsymbol{\tau}_n$ such that

$$\Pr(\mathcal{L} = \emptyset) \leq \epsilon - \alpha_n \quad (6.11)$$

$$\Pr(\exists [(m'_1, m'_2) \neq (1, 1)] \in \mathcal{L}) \leq \alpha_n \quad (6.12)$$

for some $\alpha_n = O(n^{-1/2})$. Now $\mathbf{Z}(1, 1)$ of (6.10) is a sum of independent (but not identically distributed) random vectors. For $1 \leq i \leq n$ we let

$$\mathbf{C}_i = \mathbb{E}[\mathbf{i}(u_i, X_{1i}(1), X_{2i}(1), W)]$$

and

$$\boldsymbol{\Sigma}_i = \text{Cov}(\mathbf{i}(u_i, X_{1i}(1), X_{2i}(1), W))$$

and therefore

$$\begin{aligned} \overline{\mathbf{C}}_n &\triangleq \frac{1}{n} \mathbb{E}[\mathbf{Z}(1, 1)] = \frac{1}{n} \sum_{i=1}^n \mathbf{C}_i \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\mathbf{i}(u_i, X_{1i}(1), X_{2i}(1), W)] \\ &= \sum_u \hat{P}_U(u) \mathbb{E}[\mathbf{i}(u, X_1, X_2, W)] \\ &= \mathbf{I}(\hat{P}_U P_{X_1|U} P_{X_2|U}, W) \end{aligned} \quad (6.13)$$

and

$$\begin{aligned} \overline{\boldsymbol{\Sigma}}_n &\triangleq \frac{1}{n} \text{Cov}(\mathbf{Z}(1, 1)) = \frac{1}{n} \sum_{i=1}^n \boldsymbol{\Sigma}_i \\ &= \frac{1}{n} \sum_{i=1}^n \text{Cov}(\mathbf{i}(u_i, X_{1i}(1), X_{2i}(1), W)) \\ &= \sum_u \hat{P}_U(u) \text{Cov}(\mathbf{i}(u, X_1, X_2, W)) \\ &= \mathbf{V}(\hat{P}_U P_{X_1|U} P_{X_2|U}, W) \end{aligned} \quad (6.14)$$

(both linear functions of $\hat{P}_U$). We apply Theorem B.2 by letting $\mathbf{S}_i = \mathbf{Z}_i - \mathbf{C}_i$. By (6.13) and (6.14) we have $\mathbb{E}[\mathbf{S}_i] = \mathbf{0}$ and $\mathbf{M} = \text{Cov}(\mathbf{S}) = n \overline{\boldsymbol{\Sigma}}_n$. Here we assume that $\mathbf{V}(P_U P_{X_1|U} P_{X_2|U}, W) > 0$. By Lemma B.1 we have $\lambda_{\min}(\overline{\boldsymbol{\Sigma}}_n) > 0$

for sufficiently large n and thus we have

$$\begin{aligned}\beta_i &= \mathbb{E}[\|\mathbf{M}^{-1/2}\mathbf{S}_i\|_2^3] \leq \mathbb{E}\left[\left|\lambda_{\max}(\mathbf{M}^{-1/2})\right|^3 \|\mathbf{S}_i\|_2^3\right] \\ &= n^{-3/2} [\lambda_{\min}(\overline{\Sigma}_n)]^{-3/2} \mathbb{E}[\|\mathbf{S}_i\|_2^3]\end{aligned}$$

and therefore

$$\beta = \sum_{i=1}^n \beta_i \leq \frac{1}{\sqrt{n}} [\lambda_{\min}(\overline{\Sigma}_n)]^{-3/2} \overline{T}_n \quad (6.15)$$

where $\overline{T}_n \triangleq \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\|\mathbf{Z}_i - \mathbf{C}_i\|_2^3]$ which is bounded for DM-MAC. If we let

$$\gamma_n \triangleq \frac{c [\lambda_{\min}(\overline{\Sigma}_n)]^{-3/2} \overline{T}_n}{\sqrt{n}} = O\left(\frac{1}{\sqrt{n}}\right) \quad (6.16)$$

(where c is the constant defined in Theorem B.2) and

$$A = \{\mathbf{S} \in \mathbb{R}^3 : \mathbf{S} \geq \boldsymbol{\tau}_n - n\overline{\mathbf{C}}_n\},$$

then we have

$$\begin{aligned}\Pr\{\mathbf{S} \in A\} &= \mathbb{P}\left\{\sum_{i=1}^n \mathbf{S}_i \geq \boldsymbol{\tau}_n - n\overline{\mathbf{C}}_n\right\} = \Pr\{\mathbf{Z} \geq \boldsymbol{\tau}_n\} \\ \Pr\{\mathbf{X} \in A\} &= \Pr\{\mathbf{X} \geq \boldsymbol{\tau}_n - n\overline{\mathbf{C}}_n\}\end{aligned}$$

and thus by Theorem B.2, (6.15), and (6.16) we have

$$|\mathbb{P}\{\mathbf{Z} \geq \boldsymbol{\tau}_n\} - \Pr\{\mathbf{X} \geq \boldsymbol{\tau}_n - n\overline{\mathbf{C}}_n\}| \leq \Delta(\mathcal{S}) \leq c\beta \leq \gamma_n. \quad (6.17)$$

We now pick $\boldsymbol{\tau}_n$ arbitrarily from

$$n\overline{\mathbf{C}}_n - \sqrt{n}\mathcal{Q}_{\text{inv}}(\overline{\Sigma}_n, \epsilon - \alpha_n - \gamma_n).$$

Then by the definition (6.6) of $\mathcal{Q}_{\text{inv}}$ we have

$$\Pr(\mathbf{Z} \geq \boldsymbol{\tau}_n) \geq \underbrace{\Pr\{\mathbf{X} \geq \boldsymbol{\tau}_n - n\overline{\mathbf{C}}_n\}}_{\geq 1 - \epsilon + \alpha_n + \gamma_n} - \gamma_n \geq 1 - \epsilon + \alpha_n$$

and therefore (6.11) is satisfied. On the other hand, using large deviations

analysis we obtain

$$\begin{cases} \Pr(\exists((m'_1 \neq 1), 1) \in \mathcal{L}) \leq M_1 \exp\{-\tau_{1,n} + O(1)\} \\ \Pr(\exists(1, (m'_2 \neq 1)) \in \mathcal{L}) \leq M_2 \exp\{-\tau_{2,n} + O(1)\} \\ \Pr(\exists((m'_1 \neq 1), (m'_2 \neq 2)) \in \mathcal{L}) \\ \leq M_1 M_2 \exp\{-\tau_{3,n} + O(1)\}. \end{cases}$$

Therefore (6.12) is satisfied if

$$[R_1, R_2, R_1 + R_2]' \in \frac{\boldsymbol{\tau}_n}{n} - O\left(\frac{\log n}{n}\right).$$

Finally, by combining $\|\bar{\mathbf{C}}_n - \mathbf{I}(P_{UX_1X_2}, W)\|_1 = O(1/n)$ and $\|\bar{\boldsymbol{\Sigma}}_n - \mathbf{V}(P_{UX_1X_2}, W)\|_1 = O(1/n)$ (which can be shown using (6.9)) and the fact that $P_{UX_1X_2}$ and $\boldsymbol{\tau}_n$ are arbitrary, we prove the theorem for the case $\mathbf{V}(P_U P_{X_1|U} P_{X_2|U}, W) > 0$.

Now we consider the case where $\mathbf{V}(P_U P_{X_1|U} P_{X_2|U}, W)$ is singular but $p = \text{rank}(\mathbf{V}(P_U P_{X_1|U} P_{X_2|U}, W)) \geq 1$. In this case we can find a transformation matrix $\mathbf{T} \in \mathbb{R}^{p \times 3}$ and analyze the statistics of $\tilde{\mathbf{Z}} = \mathbf{T}\mathbf{Z}$. The covariance matrix of $\tilde{\mathbf{Z}}$ is non-singular and we can follow the same lines of proof. If $\mathbf{V}(P_U P_{X_1|U} P_{X_2|U}, W) = \mathbf{0}$, then $\bar{\boldsymbol{\Sigma}}_n = O\left(\frac{1}{n}\right)$ so the above analysis still holds.

6.4 Outer Region

We provide general outer regions for discrete multiple access channels (not necessarily memoryless), for both average and maximal error probability criteria. The results can be regarded as generalizations of the general converses for single-user channels in [9].

6.4.1 Average Error Probability

Theorem 6.4. *For any four channels $P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}$ and $\{Q_{\mathbf{X}_1\mathbf{X}_2\mathbf{Y}}^j\}_{j=1}^3$ and a MAC code $\mathcal{C}$, let*

$$\begin{aligned}\epsilon &= \text{avg. error prob. with } P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2} \\ \epsilon'_j &= \text{avg. error prob. with } Q_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}^j, j = 1, 2, 3 \\ P_{\mathbf{X}_1\mathbf{X}_2} &= P_{\mathbf{X}_1}P_{\mathbf{X}_2} = Q_{\mathbf{X}_1\mathbf{X}_2}^j = \text{encoder output} \\ &\text{distribution with indep. equiprobable codewords.}\end{aligned}$$

Then we have

$$[1 - \epsilon'_1, 1 - \epsilon'_2, 1 - \epsilon'_3] \in \mathcal{E}_\epsilon \left(P_{\mathbf{X}_1\mathbf{X}_2\mathbf{Y}}, \{Q_{\mathbf{X}_1\mathbf{X}_2\mathbf{Y}}^j\}_{j=1}^3 \right).$$

Proof. This is a generalization of [9, Theorem 26]. Consider the following test δ for deciding between P and $\{Q^j\}$: denote the observed pair by $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y})$; $\mathbf{y}$ is fed into the decoder and the test declares P if the message is decoded correctly. The probability that the test is correct if P is the actual distribution is $\mathbb{E}_P[\delta(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})] = 1 - \epsilon$, and the probability that the test is incorrect if Q^j is the actual distribution is $\mathbb{E}_{Q^j}[\delta(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})] = 1 - \epsilon'_j$, $j = 1, 2, 3$. Therefore the error vector $[1 - \epsilon'_1, 1 - \epsilon'_2, 1 - \epsilon'_3]'$ is achievable. $\square$

Theorem 6.5. *Every (n, M_1, M_2, ϵ) MAC code (average error probability) satisfies*

$$\begin{aligned}[1/M_1, 1/M_2, 1/M_1M_2]' &\in \bigcup_{P_{\mathbf{X}_1\mathbf{X}_2}} \bigcap_{Q_{\mathbf{Y}|\mathbf{X}_2}^1, Q_{\mathbf{Y}|\mathbf{X}_1}^2, Q_{\mathbf{Y}}^3} \mathcal{E}_\epsilon \left(P_{\mathbf{X}_1\mathbf{X}_2} \times P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}, \right. \\ &\left. \{P_{\mathbf{X}_1\mathbf{X}_2} \times Q_{\mathbf{Y}|\mathbf{X}_2}^1, P_{\mathbf{X}_1\mathbf{X}_2} \times Q_{\mathbf{Y}|\mathbf{X}_1}^2, P_{\mathbf{X}_1\mathbf{X}_2} \times Q_{\mathbf{Y}}^3\} \right) \end{aligned} \quad (6.18)$$

where $Q_{\mathbf{Y}|\mathbf{X}_2}^1$, $Q_{\mathbf{Y}|\mathbf{X}_1}^2$, and $Q_{\mathbf{Y}}^3$ range over all distributions on $\mathcal{Y}^n$.

Proof. By particularizing $Q_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}^1$ in Theorem 6.4 to be independent of $\mathbf{X}_1$, the error probability ϵ'_1 is at least $1 - 1/M_1$. Similarly we have $\epsilon'_2 \geq 1 - 1/M_2$ and $\epsilon'_3 \geq 1 - 1/M_1M_2$. Hence the theorem. $\square$

6.4.2 Maximal Error Probability

Theorem 6.6. *For any two channels $P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}$ and $Q_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}$, a MAC code $\mathcal{C}$, and a subset Ω of the set of message pairs $\mathcal{M}_1 \times \mathcal{M}_2$, let*

$$\begin{aligned}\epsilon_\Omega &= \max_{(m_1, m_2) \in \Omega} P_e(m_1, m_2, P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}) \\ \epsilon'_\Omega &= \max_{(m_1, m_2) \in \Omega} P_e(m_1, m_2, Q_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}).\end{aligned}$$

Then

$$\min_{(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{C}(\Omega)} \beta_{1-\epsilon_\Omega}(P_{\mathbf{Y}|\mathbf{X}_1=\mathbf{x}_1, \mathbf{X}_2=\mathbf{x}_2}, Q_{\mathbf{Y}|\mathbf{X}_1=\mathbf{x}_1, \mathbf{X}_2=\mathbf{x}_2}) \leq 1 - \epsilon'_\Omega.$$

Proof. Let $(m_1^*, m_2^*) \in \Omega$ be the message pair with $P_e(m_1^*, m_2^*, Q_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}) = \epsilon'_\Omega$. Consider the test δ for deciding between $P_{\mathbf{Y}|\mathbf{X}_1=\mathbf{x}_1(m_1^*), \mathbf{X}_2=\mathbf{x}_2(m_2^*)}$ and $Q_{\mathbf{Y}|\mathbf{X}_1=\mathbf{x}_1(m_1^*), \mathbf{X}_2=\mathbf{x}_2(m_2^*)}$ that declares P if the message is decoded correctly. The probability that the test is correct if P is the actual distribution is $\mathbb{E}_P[\delta(\mathbf{x}_1(m_1^*), \mathbf{x}_2(m_2^*), \mathbf{Y})] \geq 1 - \epsilon_\Omega$, and the probability that the test is incorrect if Q is the actual distribution is $\mathbb{E}_Q[\delta(\mathbf{x}_1(m_1^*), \mathbf{x}_2(m_2^*), \mathbf{Y})] = 1 - \epsilon'_\Omega$. Therefore

$$\begin{aligned}\beta_{1-\epsilon_\Omega}(P_{\mathbf{Y}|\mathbf{X}_1=\mathbf{x}_1(m_1^*), \mathbf{X}_2=\mathbf{x}_2(m_2^*)}, \\ Q_{\mathbf{Y}|\mathbf{X}_1=\mathbf{x}_1(m_1^*), \mathbf{X}_2=\mathbf{x}_2(m_2^*)}) \leq 1 - \epsilon'_\Omega. \quad \square\end{aligned}$$

By particularizing $Q_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}$ in Theorem 6.6 independent of $\mathbf{X}_1$ and $\mathbf{X}_2$, the maximal error probability ϵ'_Ω is at least $1 - 1/|\Omega|$. Also ϵ_Ω is trivially upper-bounded by ϵ . Therefore Theorem 6.2 follows.

Theorem 6.7. *Given an $(n, \bar{\Omega}, \epsilon)$ MAC subcode (maximal error probability) of $\mathcal{C}$ for channel $P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}$, a subset Ω of $\bar{\Omega}$, and three arbitrary MAC channels $Q_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}^j, j = 1, 2, 3$. Then for any $(m_1, m_2) \in \Omega$, we have*

$$\begin{aligned}& \begin{bmatrix} 1 - P_e(m_1, m_2, Q_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}^1, \mathcal{C}) \\ 1 - P_e(m_1, m_2, Q_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}^2, \mathcal{C}) \\ 1 - P_e(m_1, m_2, Q_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}^3, \mathcal{C}) \end{bmatrix} \\ & \in \bigcup_{(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{C}(\Omega)} \mathcal{E}_\epsilon \left(P_{\mathbf{Y}|\mathbf{X}_1=\mathbf{x}_1, \mathbf{X}_2=\mathbf{x}_2}, \left\{ Q_{\mathbf{Y}|\mathbf{X}_1=\mathbf{x}_1, \mathbf{X}_2=\mathbf{x}_2}^j \right\}_{j=1}^3 \right). \quad (6.19)\end{aligned}$$

Proof. Consider the test δ for deciding between $P_{\mathbf{Y}|\mathbf{X}_1=\mathbf{x}_1(m_1),\mathbf{X}_2=\mathbf{x}_2(m_2)}$ and $\left\{Q_{\mathbf{Y}|\mathbf{X}_1=\mathbf{x}_1(m_1),\mathbf{X}_2=\mathbf{x}_2(m_2)}^j\right\}_{j=1}^3$ that declares P if the message pair is decoded correctly. The probability that the test is correct if P is the actual distribution is

$$\mathbb{E}_P[\delta(\mathbf{x}_1(m_1), \mathbf{x}_2(m_2), \mathbf{Y})] = 1 - P_e(m_1, m_2, P_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}, \mathcal{C}) \geq 1 - \epsilon,$$

and the probability that the test is incorrect if Q^j for $j = 1, 2, 3$ is the actual distribution is

$$\mathbb{E}_{Q^j}[\delta(\mathbf{x}_1(m_1), \mathbf{x}_2(m_2), \mathbf{Y})] = 1 - P_e(m_1, m_2, Q_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}^j, \mathcal{C}).$$

Thus by the definition of $\mathcal{E}_\epsilon$ we have

$$\begin{aligned} & \begin{bmatrix} 1 - P_e(m_1, m_2, Q_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}^1, \mathcal{C}) \\ 1 - P_e(m_1, m_2, Q_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}^2, \mathcal{C}) \\ 1 - P_e(m_1, m_2, Q_{\mathbf{Y}|\mathbf{X}_1\mathbf{X}_2}^3, \mathcal{C}) \end{bmatrix} \\ & \in \mathcal{E}_\epsilon \left(P_{\mathbf{Y}|\mathbf{X}_1=\mathbf{x}_1(m_1),\mathbf{X}_2=\mathbf{x}_2(m_2)}, \left\{ Q_{\mathbf{Y}|\mathbf{X}_1=\mathbf{x}_1(m_1),\mathbf{X}_2=\mathbf{x}_2(m_2)}^j \right\}_{j=1}^3 \right) \end{aligned}$$

and therefore (6.19) is proved. $\square$

Theorem 6.3 can be shown by applying Theorem 6.7. Using the fact that $Q_{\mathbf{Y}}^3$ is independent of $\mathbf{X}_1$ and $\mathbf{X}_2$, we have

$$\begin{aligned} \sum_{(m_1, m_2) \in \Omega} P_e(m_1, m_2, Q_{\mathbf{Y}}^3, \mathcal{C}) &= \sum_{(m_1, m_2) \in \Omega} \sum_{\mathbf{y} \in \mathcal{Y}^n} Q_{\mathbf{Y}}^3(\mathbf{y}) \Pr\{\phi(\mathbf{y}) \neq (m_1, m_2)\} \\ &= \sum_{\mathbf{y} \in \mathcal{Y}^n} Q_{\mathbf{Y}}^3(\mathbf{y}) \underbrace{\sum_{(m_1, m_2) \in \Omega} \Pr\{\phi(\mathbf{y}) \neq (m_1, m_2)\}}_{\geq |\Omega| - 1} \\ &\geq |\Omega| - 1. \end{aligned}$$

Similarly we have

$$\begin{aligned}
& \sum_{(m_1, m_2) \in \Omega} P_e(m_1, m_2, Q_{\mathbf{Y}|\mathbf{X}_2}^1, \mathcal{C}) \\
&= \sum_{(m_1, m_2) \in \Omega} \sum_{\mathbf{y} \in \mathcal{Y}^n} Q_{\mathbf{Y}|\mathbf{X}_2}^1(\mathbf{y}|\mathbf{x}_2(m_2)) \Pr\{\phi(\mathbf{y}) \neq (m_1, m_2)\} \\
&= \sum_{m_2 \in \mathcal{M}_2} \sum_{\mathbf{y} \in \mathcal{Y}^n} Q_{\mathbf{Y}|\mathbf{X}_2}^1(\mathbf{y}|\mathbf{x}_2(m_2)) \underbrace{\sum_{m_1: (m_1, m_2) \in \Omega} \Pr\{\phi(\mathbf{y}) \neq (m_1, m_2)\}}_{\geq N_1(m_2) - 1} \\
&\geq \sum_{m_2 \in \mathcal{M}_2} [N_1(m_2) - 1] \sum_{\mathbf{y} \in \mathcal{Y}^n} Q_{\mathbf{Y}|\mathbf{X}_2}^1(\mathbf{y}|\mathbf{x}_2(m_2)) \\
&= |\Omega| - M_2
\end{aligned}$$

where $N_1(m_2) \triangleq |\{m_1 : (m_1, m_2) \in \Omega\}|$ is the number of message pairs in Ω whose second message is m_2 . In sum, we have

$$|\Omega| - M_2 \leq \sum_{(m_1, m_2) \in \Omega} P_e(m_1, m_2, Q_{\mathbf{Y}|\mathbf{X}_2}^1, \mathcal{C}) \leq |\Omega| \quad (6.20)$$

$$|\Omega| - M_1 \leq \sum_{(m_1, m_2) \in \Omega} P_e(m_1, m_2, Q_{\mathbf{Y}|\mathbf{X}_1}^2, \mathcal{C}) \leq |\Omega| \quad (6.21)$$

$$|\Omega| - 1 \leq \sum_{(m_1, m_2) \in \Omega} P_e(m_1, m_2, Q_{\mathbf{Y}}^3, \mathcal{C}) \leq |\Omega|. \quad (6.22)$$

Now if we can find a message pair $(m_1, m_2) \in \Omega$ such that

$$P_e(m_1, m_2, Q_{\mathbf{Y}|\mathbf{X}_2}^1, \mathcal{C}) \geq 1 - \frac{3M_2}{|\Omega|} \triangleq \epsilon'_1 \quad (6.23)$$

$$P_e(m_1, m_2, Q_{\mathbf{Y}|\mathbf{X}_1}^2, \mathcal{C}) \geq 1 - \frac{3M_1}{|\Omega|} \triangleq \epsilon'_2 \quad (6.24)$$

$$P_e(m_1, m_2, Q_{\mathbf{Y}}^3, \mathcal{C}) \geq 1 - \frac{3}{|\Omega|} \triangleq \epsilon'_3, \quad (6.25)$$

where $\epsilon'_1, \epsilon'_2, \epsilon'_3 \in (0, 1)$ then by substituting (6.23)-(6.25) into (6.19) we can prove the theorem. $|\bar{\Omega}| \geq |\Omega| \geq 3 \max\{M_1, M_2\}$

Suppose not, i.e., for any $(m_1, m_2) \in \Omega$ at least one of the three inequalities among (6.23)-(6.25) into (6.19) is not satisfied. Let

$$\mathcal{S}_j \triangleq \{(m_1, m_2) : P_e(m_1, m_2, Q^j, \mathcal{C}) < \epsilon'_j\}, \quad j = 1, 2, 3.$$

Since every message pair $(m_1, m_2) \in \Omega$ belongs to at least one of the $\mathcal{S}_j$'s,

there exists $j^* \in \{1, 2, 3\}$ such that $|\mathcal{S}_{j^*}| \geq \frac{|\Omega|}{3}$. Then we have

$$\begin{aligned}
& \sum_{(m_1, m_2) \in \Omega} P_e(m_1, m_2, Q^{j^*}, \mathcal{C}) \\
&= \sum_{(m_1, m_2) \in \Omega \cap \mathcal{S}_{j^*}} P_e(m_1, m_2, Q^{j^*}, \mathcal{C}) + \sum_{(m_1, m_2) \in \Omega \setminus \mathcal{S}_{j^*}} P_e(m_1, m_2, Q^{j^*}, \mathcal{C}) \\
&< |\mathcal{S}_{j^*}| \cdot \epsilon'_{j^*} + (|\Omega| - |\mathcal{S}_{j^*}|) \cdot 1 \\
&= |\Omega| - |\mathcal{S}_{j^*}| (1 - \epsilon'_{j^*}) \\
&\leq |\Omega| - \frac{|\Omega|}{3} (1 - \epsilon'_{j^*}) \\
&= |\Omega| \left[1 - \frac{1}{3} (1 - \epsilon'_{j^*}) \right] \\
&= \begin{cases} |\Omega| - M_2 & j^* = 1 \\ |\Omega| - M_1 & j^* = 2 \\ |\Omega| - 1 & j^* = 3, \end{cases}
\end{aligned}$$

which contradicts with (6.20), (6.21), or (6.22). Finally, $|\Omega| \geq 3 \max \{M_1, M_2\}$ ensures that $\epsilon'_1, \epsilon'_2, \epsilon'_3 \in (0, 1)$ and therefore Theorem 6.3 follows.

CHAPTER 7

SUMMARY AND FUTURE WORK

In this work we focus on characterizing the asymptotics of two multi-user channels. The fingerprinting channel is associated with the technology of digital fingerprinting, which is an active topic in the area of information forensics and security. The multiple access channel is one of the most fundamental building blocks of communication systems. The asymptotic analyses study the fundamental limits as the sizes of the problems become intractable.

7.1 Digital Fingerprinting

The difficulty of analyzing fingerprinting capacity games stems from the uncertainty of pirates' attack. As the number k of pirates grows, their attack can be more powerful. Moreover, as shown in Chap. 4 and Chap. 5, the game is already unmanageable even when k is in single digits. Therefore it is crucial to study the large-coalition fingerprinting games. In Chap. 4 we study the asymptotics of fingerprinting games for the binary alphabets under the marking assumption, and the results are generalized to arbitrary alphabets under the combined digit model in Chap. 5. We show that capacity is asymptotic to A/k^2 where the constant A is explicitly specified. The asymptotic optimal strategies of the fingerprinting capacity game are also obtained.

The capacity game results are mostly negative. The quadratic decay of capacity suggests that it is fundamentally impossible to construct an efficient fingerprinting code that is robust to large-coalition attacks. Furthermore, the capacity values under the marking assumption derived in previous work are too optimistic and are about 20-30% less under CDM. However, the fact that the simple capacity game is asymptotically the same as the joint capacity game suggests a positive result: One can essentially achieve the maximum

achievable rate with computationally feasible codes. Along this line of work one interesting problem to explore is how to design an effective fingerprinting code that approaches capacity. Tardos’ construction [4] and its q -ary alphabet generalization [23] already achieved the same $1/k^2$ rate as capacity, but how to achieve or approach the same constant factor A remains an open problem. A recent work by Oosterwijk et al. [45] shed light on how our theoretical capacity results are indeed in align with practical fingerprinting code construction: They improved upon their earlier construction [23] by changing the distribution of the time-sharing variable $\mathbf{W}$ from $dP_{\mathbf{W}}(\mathbf{w}) \propto \prod_x w_x^{-1/q} d\mathbf{w}$ to $dP_{\mathbf{W}}^*(\mathbf{w}) \propto \prod_x w_x^{-1/2} d\mathbf{w}$, and by tailoring the so-called “suspicion function” against the interleaving attack. This matches exactly our maximizing and minimizing strategies for the asymptotic fingerprinting game under RDM (Corollary 5.5)! Therefore, we believe that, by studying the asymptotic capacity game for the general model of CDM, one can construct an efficient fingerprinting code against practical collusion attacks.

7.2 Multiple Access Channel

The study of second-order channel coding rate, initiated by Strassen [8] in 1962 and recently revisited by Polyanskiy et al. [9] and Hayashi [10], has become one of the most active areas of research in information theory. The characterization of second-order coding rates for multiple access channels is studied in many recent literatures [11, 12, 7, 13, 14, 20], but it remains an open problem.

Our characterization of inner region for the discrete memoryless MAC has been the tightest bound in the literature until Scarlett et al.’s very recent work [20]. On the other hand, there has not been any significant progress on the outer region for discrete memoryless MAC other than some trivial bounds that can be derived from single-user channels [11]. We believe our general outer region bounds can be extended to tight single-lettered outer bound. Such extension, however, requires strong large deviation analysis for composite hypothesis problems.

APPENDIX A

PROOFS FOR FINGERPRINTING CAPACITY

A.1 Derivation of $C_{k,2}^{\text{simple}}(\mathcal{P}_W, \mathbf{p}^*)$

The function $I_{k,2}^{\text{simple}}(w, \mathbf{p}^*)$ is indeed symmetric around $w = 1/2$ and has a global maximum at $w = 1/2$ as suggested by numerical experiments in [22]. We first prove the following lemma:

Lemma A.1. *For $r \geq s \geq 0$ and $r + s \leq 1$, we have*

$$d_2(r \parallel s) \geq d_2(s \parallel r). \quad (\text{A.1})$$

Proof. The difference $\delta(r, s)$ between the two sides is

$$\begin{aligned} \delta(r, s) &= d_2(r \parallel s) - d_2(s \parallel r) \\ &= r \log_2 \frac{r}{s} + (1-r) \log_2 \frac{1-r}{1-s} - s \log_2 \frac{s}{r} - (1-s) \log_2 \frac{1-s}{1-r} \\ &= (r+s) \log_2 \frac{r}{s} + (2-r-s) \log_2 \frac{1-r}{1-s}. \end{aligned} \quad (\text{A.2})$$

Then

$$\frac{\partial}{\partial s} \delta(r, s) = \frac{s-r}{s(1-s) \ln 2} + \log_2 \frac{r(1-s)}{s(1-r)} \quad (\text{A.3})$$

and

$$\frac{\partial^2}{\partial s^2} \delta(r, s) = \frac{(r-s)(1-2s)}{s^2(1-s)^2 \ln 2} \geq 0 \quad (\text{A.4})$$

for all $r \geq s$ and $r + s \leq 1$. Now when $s \leq r \leq 1/2$, we have $\frac{\partial}{\partial s} \delta(r, s)|_{s=r} = \delta(r, r) = 0$. Thus $\frac{\partial}{\partial s} \delta(r, s) \leq 0$ and thus $\delta(r, s) \geq 0$. When $1/2 \leq r \leq 1-s$, we have $\frac{\partial}{\partial s} \delta(r, s)|_{s=1-r} \leq 0$ and $\delta(r, 1-r) = 0$. Hence similarly $\frac{\partial}{\partial s} \delta(r, s) \leq 0$ and hence $\delta(r, s) \geq 0$. To prove the inequality $\frac{\partial}{\partial s} \delta(r, s)|_{s=1-r} \leq 0$, let

$$\lambda(r) \triangleq \frac{\partial}{\partial s} \delta(r, s) \Big|_{s=1-r} = \frac{1-2r}{r(1-r) \ln 2} + 2 \log_2 \frac{r}{1-r}$$

and since $\lambda(1/2) = 0$ and $\lambda'(r) = -\frac{(1-2r)^2}{r^2(1-r)^2 \ln 2} \leq 0$ it follows that $\lambda(r) \leq 0$ for all $r \in [1/2, 1]$. $\square$

Now the payoff function can be written as

$$\begin{aligned} I_{k,2}^{\text{simple}}(w, \mathbf{p}^*) &\stackrel{(a)}{=} wd_2(\boldsymbol{\alpha}^{1'} \mathbf{p}^* \parallel \boldsymbol{\alpha}' \mathbf{p}^*) + (1-w)d_2(\boldsymbol{\alpha}^{0'} \mathbf{p}^* \parallel \boldsymbol{\alpha}' \mathbf{p}^*) \\ &\stackrel{(b)}{=} wd_2(w + \frac{1-w}{k} \parallel w) + (1-w)d_2(w - \frac{w}{k} \parallel w) \end{aligned} \quad (\text{A.5})$$

where (a) follows from (4.13) and (b) follows from (4.6) and Lemma 4.2, and by which we can easily verify the symmetry property $I_{k,2}^{\text{simple}}(w, \mathbf{p}^*) = I_{k,2}^{\text{simple}}(1-w, \mathbf{p}^*)$. Hence it suffices to show that $I_{k,2}^{\text{simple}}(w, \mathbf{p}^*)$ is nondecreasing for $0 \leq w \leq 1/2$. Taking the derivative of (A.5) with respect to w and after some simplifications, we obtain

$$\frac{\partial}{\partial w} I_{k,2}^{\text{simple}}(w, \mathbf{p}^*) = d_2(w^+ \parallel w^-) - d_2(w^- \parallel w^+) \quad (\text{A.6})$$

where $w^+ \triangleq w + \frac{1-w}{k}$ and $w^- \triangleq w - \frac{w}{k}$. By Lemma A.1 it follows that $\frac{\partial}{\partial w} I_{k,2}^{\text{simple}}(w, \mathbf{p}^*) \geq 0$ for all $w \in [0, 1/2]$ and hence $I_{k,2}^{\text{simple}}(w, \mathbf{p}^*)$ achieves its maximum at $w = 1/2$.

A.2 Proof of Theorem 5.6

By the definitions in Chap. 2, the conditional distribution of $\mathbf{Z}$ given $\mathbf{w}$ follows the multinomial law with parameter $\mathbf{w}$ and k trials. We denote the probability mass functions by

$$\alpha_{\mathbf{z}}(\mathbf{w}) \triangleq p_{\mathbf{Z}|\mathbf{W}}(\mathbf{z}|\mathbf{w}) = \binom{k}{\mathbf{z}} \prod_{x=0}^{q-1} w_x^{z_x} \quad (\text{A.7})$$

for $\mathbf{z} \in \mathcal{Z}$ and $\mathbf{w} \in \mathcal{W}^q$ where $\binom{k}{\mathbf{z}}$ is the multinomial coefficient defined by $\binom{k}{\mathbf{z}} = \frac{k!}{z_0! z_1! \dots z_{q-1}!}$. Let $\boldsymbol{\alpha}(\mathbf{w}) = [\alpha_{\mathbf{z}_1}(\mathbf{w}), \dots, \alpha_{\mathbf{z}_t}(\mathbf{w})]'$ be a length- t probability vector. Similarly

$$\alpha_{\mathbf{z}}^x(\mathbf{w}) \triangleq p_{\mathbf{Z}|X_1 \mathbf{W}}(\mathbf{z}|x, \mathbf{w}) = \binom{k-1}{\mathbf{z} - \mathbf{e}_x} \prod_{l=0}^{q-1} w_l^{z_l - \delta_{xl}} \quad (\text{A.8})$$

for $x \in \mathcal{X}$, $\mathbf{z} \in \mathcal{Z}$, and $\mathbf{w} \in \mathcal{W}^q$ where $\mathbf{e}_x$ is the unit vector with the x -th coordinate equals to one. We have the following moments of $\mathbf{Z}$:

$$\mathbb{E}[\mathbf{Z}|\mathbf{W} = \mathbf{w}] = k\mathbf{w} \quad (\text{A.9})$$

$$\text{Cov}(\mathbf{Z}|\mathbf{W} = \mathbf{w}) = k(\text{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}') \quad (\text{A.10})$$

$$\mathbb{E} [\|\mathbf{Z} - k\mathbf{W}\|_2^3 | \mathbf{W} = \mathbf{w}] = O(k) \quad (\text{A.11})$$

$$\mathbb{E}[\mathbf{Z}|\mathbf{W} = \mathbf{w}, X_1 = x] = \mathbf{e}_x + (k-1)\mathbf{w} \quad (\text{A.12})$$

$$\text{Cov}(\mathbf{Z}|\mathbf{W} = \mathbf{w}, X_1 = x) = (k-1)(\text{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}') \quad (\text{A.13})$$

$$\mathbb{E} [\|\mathbf{Z} - \mathbf{e}_x - (k-1)\mathbf{w}\|_2^3 | \mathbf{W} = \mathbf{w}, X_1 = x] = O(k) \quad (\text{A.14})$$

The following lemmas will be useful for our analysis:

Lemma A.2. [33, Sec. 2.5]

$$D_q(\mathbf{f} \parallel \mathbf{g}) = \frac{1}{2 \ln q} \sum_i \frac{(f_i - g_i)^2}{g_i} + O(\|\mathbf{f} - \mathbf{g}\|_2^3) \quad (\text{A.15})$$

as $\|\mathbf{f} - \mathbf{g}\|_2 \rightarrow 0$, where $\mathbf{f}$ and $\mathbf{g}$ are probability vectors.

Lemma A.3. For $\mathbf{Z}$ following the multinomial law with parameters $\mathbf{w}$ and k trials, we have

$$\Pr \left[\|\mathbf{Z} - k\mathbf{w}\|_2 \geq q\sqrt{k \ln k} \right] = O\left(\frac{1}{k^2}\right). \quad (\text{A.16})$$

Proof. We have

$$\begin{aligned} \Pr \left[\|\mathbf{Z} - k\mathbf{w}\|_2 \geq q\sqrt{k \ln k} \right] &\leq \Pr \left[\exists x \in \mathcal{Q} : |Z_x - kw_x| \geq \sqrt{k \ln k} \right] \\ &\stackrel{(a)}{\leq} \sum_{x \in \mathcal{Q}} \Pr \left[|Z_x - kw_x| \geq \sqrt{k \ln k} \right] \stackrel{(b)}{\leq} \frac{2q}{k^2} \end{aligned}$$

where (a) follows from the union bound and (b) follows from Hoeffding's inequality (see [34]). $\square$

To prove Theorem 5.6, we first rewrite the mutual information of the payoff

functions as

$$\begin{aligned}
I_{k,q}^{\text{joint}}(\mathbf{w}, \boldsymbol{\Theta}) &\triangleq \frac{1}{k} I(\mathbf{Z}; Y | \mathbf{W} = \mathbf{w}) \\
&= \frac{1}{k} \sum_{\mathbf{z}} p_{\mathbf{Z}|\mathbf{W}}(\mathbf{z}|\mathbf{w}) \sum_y p_{Y|\mathbf{Z},\mathbf{W}}(y|\mathbf{z}, \mathbf{w}) \log_q \frac{p_{Y|\mathbf{Z},\mathbf{W}}(y|\mathbf{z}, \mathbf{w})}{p_{Y|\mathbf{W}}(y|\mathbf{w})} \\
&\stackrel{(a)}{=} \frac{1}{k} \sum_{\mathbf{z}} \alpha_{\mathbf{z}}(\mathbf{w}) \sum_y \theta_y(\mathbf{z}) \log_q \frac{\theta_y(\mathbf{z})}{(\boldsymbol{\Theta}\boldsymbol{\alpha})_y} \\
&= \frac{1}{k} \sum_{\mathbf{z}} \alpha_{\mathbf{z}}(\mathbf{w}) D_q(\boldsymbol{\theta}(\mathbf{z}) \parallel \boldsymbol{\Theta}\boldsymbol{\alpha})
\end{aligned} \tag{A.17}$$

where (a) follows from (2.16) and (A.7), and

$$\begin{aligned}
I_{k,q}^{\text{simple}}(\mathbf{w}, \boldsymbol{\Theta}) &\triangleq I(X_1; Y | \mathbf{W} = \mathbf{w}) \\
&= \sum_{x \in \mathcal{Q}} p_{X_1|\mathbf{W}}(x|\mathbf{w}) \sum_y p_{Y|X_1,\mathbf{W}}(y|x, \mathbf{w}) \log_q \frac{p_{Y|X_1,\mathbf{W}}(y|x, \mathbf{w})}{p_{Y|\mathbf{W}}(y|\mathbf{w})} \\
&\stackrel{(a)}{=} \sum_{x \in \mathcal{Q}} w_x \sum_y (\boldsymbol{\Theta}\boldsymbol{\alpha}^x)_y \log_q \frac{(\boldsymbol{\Theta}\boldsymbol{\alpha}^x)_y}{(\boldsymbol{\Theta}\boldsymbol{\alpha})_y} \\
&= \sum_x w_x D_q(\boldsymbol{\Theta}\boldsymbol{\alpha}^x \parallel \boldsymbol{\Theta}\boldsymbol{\alpha}).
\end{aligned} \tag{A.18}$$

where (a) follows from (2.1) and (A.8). By Condition 5.1 and Taylor's Theorem, we have for each $y \in \mathcal{Y}$,

$$\begin{aligned}
\theta_y(\mathbf{z}) &= g_y \left(\frac{1}{k} \mathbf{z} \right) + o \left(\frac{1}{k} \right) \\
&= g_y(\mathbf{w}) + \frac{1}{k} \nabla g_y(\mathbf{w}) (\mathbf{z} - k\mathbf{w}) + \frac{1}{2k^2} (\mathbf{z} - k\mathbf{w})' \nabla^2 g_y(\mathbf{w}) (\mathbf{z} - k\mathbf{w}) \\
&\quad + O \left(\frac{1}{k^3} \|\mathbf{z} - k\mathbf{w}\|_2^3 \right) + o \left(\frac{1}{k} \right)
\end{aligned} \tag{A.19}$$

as $k \rightarrow \infty$, where $\nabla^2 g_y$ is the q -by- q Hessian matrix. Therefore

$$\begin{aligned}
(\Theta\alpha)_y &= \mathbb{E}[\theta_y(\mathbf{Z})|\mathbf{W} = \mathbf{w}] \\
&= g_y(\mathbf{w}) + \frac{1}{k} \nabla g_y(\mathbf{w}) \underbrace{\mathbb{E}[\mathbf{Z} - k\mathbf{W}|\mathbf{W} = \mathbf{w}]}_{\stackrel{(a)}{=} \mathbf{0}} \\
&\quad + \frac{1}{2k^2} \mathbb{E}[(\mathbf{Z} - k\mathbf{W})' \nabla^2 g_y(\mathbf{w}) (\mathbf{Z} - k\mathbf{W}) | \mathbf{W} = \mathbf{w}] \\
&\quad + \underbrace{O\left(\frac{1}{k^3} \mathbb{E}[\|\mathbf{Z} - k\mathbf{W}\|_2^3 | \mathbf{W} = \mathbf{w}]\right)}_{\stackrel{(b)}{=} O(1/k^2)} + o\left(\frac{1}{k}\right) \\
&= g_y(\mathbf{w}) + \frac{1}{2k^2} \text{tr}[\nabla^2 g_y(\mathbf{w}) \text{Cov}(\mathbf{Z}|\mathbf{W} = \mathbf{w})] + o\left(\frac{1}{k}\right) \\
&\stackrel{(c)}{=} g_y(\mathbf{w}) + \frac{1}{2k} \text{tr}[\nabla^2 g_y(\mathbf{w}) (\text{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}')] + o\left(\frac{1}{k}\right) \quad (\text{A.20})
\end{aligned}$$

where (a), (b), and (c) follows from (A.9), (A.11), and (A.10) respectively,

and

$$\begin{aligned}
& (\Theta \alpha^x)_y \\
&= \mathbb{E} [\theta_y(\mathbf{Z}) | \mathbf{W} = \mathbf{w}, X_1 = x] \\
&= g_y(\mathbf{w}) + \frac{1}{k} \nabla g_y(\mathbf{w}) \underbrace{\mathbb{E} [\mathbf{Z} - k\mathbf{W} | \mathbf{W} = \mathbf{w}, X_1 = x]}_{\stackrel{(a)}{=} \mathbf{e}_x - \mathbf{w}} \\
&\quad + \frac{1}{2k^2} \mathbb{E} [(\mathbf{Z} - k\mathbf{W})' \nabla^2 g_y(\mathbf{W}) (\mathbf{Z} - k\mathbf{W}) | \mathbf{W} = \mathbf{w}, X_1 = x] \\
&\quad + O \left(\underbrace{\frac{1}{k^3} \mathbb{E} [\|(\mathbf{Z} - \mathbf{e}_x - (k-1)\mathbf{w}) + \mathbf{e}_x - \mathbf{w}\|_2^3 | \mathbf{W} = \mathbf{w}, X_1 = x]}_{\stackrel{(b)}{=} O(1/k^2)} \right) + o \left(\frac{1}{k} \right) \\
&= g_y(\mathbf{w}) + \frac{1}{k} \nabla g_y(\mathbf{w}) (\mathbf{e}_x - \mathbf{w}) \\
&\quad + \frac{1}{2k^2} \text{tr} \{ \nabla^2 g_y(\mathbf{w}) \mathbb{E} [(\mathbf{Z} - k\mathbf{w}) (\mathbf{Z} - k\mathbf{w})' | \mathbf{W} = \mathbf{w}, X_1 = x] \} + o \left(\frac{1}{k} \right) \\
&= g_y(\mathbf{w}) + \frac{1}{k} \nabla g_y(\mathbf{w}) (\mathbf{e}_x - \mathbf{w}) \\
&\quad + \frac{1}{2k^2} \text{tr} \{ \nabla^2 g_y(\mathbf{w}) [\text{Cov}(\mathbf{Z} | \mathbf{W} = \mathbf{w}, X_1 = x) + (\mathbf{e}_x - \mathbf{w})(\mathbf{e}_x - \mathbf{w})'] \} + o \left(\frac{1}{k} \right) \\
&\stackrel{(c)}{=} g_y(\mathbf{w}) + \frac{1}{k} \nabla g_y(\mathbf{w}) (\mathbf{e}_x - \mathbf{w}) + \frac{k-1}{2k^2} \text{tr} [\nabla^2 g_y(\mathbf{w}) (\text{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}')] + o \left(\frac{1}{k} \right) \\
&= g_y(\mathbf{w}) + \frac{1}{k} \nabla g_y(\mathbf{w}) (\mathbf{e}_x - \mathbf{w}) + \frac{1}{2k} \text{tr} [\nabla^2 g_y(\mathbf{w}) (\text{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}')] + o \left(\frac{1}{k} \right) \\
&\tag{A.21}
\end{aligned}$$

where (a), (b), and (c) follow from (A.12), (A.14), and (A.13) respectively.

Now we have

$$\begin{aligned}
I_{k,q}^{\text{joint}}(\mathbf{w}, \Theta) &= \frac{1}{k} \sum_{\mathbf{z}} \alpha_{\mathbf{z}}(\mathbf{w}) D_q(\theta(\mathbf{z}) \parallel \Theta \alpha) \\
&\stackrel{(a)}{\sim} \frac{1}{k} \sum_{\mathbf{z}: \|\mathbf{z} - k\mathbf{w}\|_2 \leq q\sqrt{k \ln k}} \alpha_{\mathbf{z}}(\mathbf{w}) D_q(\theta(\mathbf{z}) \parallel \Theta \alpha) \tag{A.22}
\end{aligned}$$

where (a) follows from Lemma A.3. If we let $\mathbf{z} = k\mathbf{w} + \boldsymbol{\eta}$ where $\boldsymbol{\eta} = O(\sqrt{k \ln k})$, then combining (A.19) and (A.20) yields

$$\theta_y(\mathbf{z}) - (\Theta \alpha)_y = \frac{1}{k} \nabla g_y(\mathbf{w}) \boldsymbol{\eta} + O \left(\frac{\ln k}{k} \right)$$

for $\boldsymbol{\eta} = \omega(1)$. The contributions to (A.22) for $\boldsymbol{\eta} = O(1)$ decays exponentially with k and thus can be neglected. Hence we have

$$\begin{aligned}
I_{k,q}^{\text{joint}}(\mathbf{w}, \boldsymbol{\Theta}) &\sim \frac{1}{k} \sum_{\mathbf{z}: \|\mathbf{z} - k\mathbf{w}\|_2 \leq q\sqrt{k \ln k}} \alpha_{\mathbf{z}}(\mathbf{w}) D_q(\boldsymbol{\theta}(\mathbf{z}) \parallel \boldsymbol{\Theta}\boldsymbol{\alpha}) \\
&\stackrel{(a)}{\sim} \frac{1}{k} \sum_{\mathbf{z}: \|\mathbf{z} - k\mathbf{w}\|_2 \leq q\sqrt{k \ln k}} \alpha_{\mathbf{z}}(\mathbf{w}) \cdot \frac{1}{2 \ln q} \sum_y \frac{\left[\frac{1}{k} \nabla g_y(\mathbf{w})(\mathbf{z} - k\mathbf{w})\right]^2}{g_y(\mathbf{w})} \\
&= \frac{1}{2k^3 \ln q} \sum_y \frac{1}{g_y(\mathbf{w})} \sum_{\mathbf{z}: \|\mathbf{z} - k\mathbf{w}\|_2 \leq q\sqrt{k \ln k}} \alpha_{\mathbf{z}}(\mathbf{w}) [\nabla g_y(\mathbf{w})(\mathbf{z} - k\mathbf{w})]^2 \\
&\stackrel{(b)}{\sim} \frac{1}{2k^3 \ln q} \sum_y \frac{1}{g_y(\mathbf{w})} \sum_{\mathbf{z}} \alpha_{\mathbf{z}}(\mathbf{w}) [\nabla g_y(\mathbf{w})(\mathbf{z} - k\mathbf{w})]^2 \\
&= \frac{1}{2k^3 \ln q} \sum_y \frac{\nabla g_y(\mathbf{w}) \text{Cov}(\mathbf{Z} | \mathbf{W} = \mathbf{w}) \nabla' g_y(\mathbf{w})}{g_y(\mathbf{w})} \\
&\stackrel{(c)}{=} \frac{1}{2k^2 \ln q} \sum_y \frac{\nabla g_y(\mathbf{w}) [\text{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}'] \nabla' g_y(\mathbf{w})}{g_y(\mathbf{w})} \tag{A.23}
\end{aligned}$$

where (a) follows from Lemma A.2, (b) follows from Lemma A.3, and (c) follows from (A.10).

The derivation of $I_{k,q}^{\text{simple}}(\mathbf{w}, \boldsymbol{\Theta})$ is simpler:

$$\begin{aligned}
I_{k,q}^{\text{simple}}(\mathbf{w}, \boldsymbol{\Theta}) &= \sum_x w_x D_q(\boldsymbol{\Theta}\boldsymbol{\alpha}^x \parallel \boldsymbol{\Theta}\boldsymbol{\alpha}) \\
&\stackrel{(a)}{\sim} \frac{1}{2 \ln q} \sum_x w_x \sum_y \frac{[(\boldsymbol{\Theta}\boldsymbol{\alpha}^x)_y - (\boldsymbol{\Theta}\boldsymbol{\alpha})_y]^2}{(\boldsymbol{\Theta}\boldsymbol{\alpha})_y} \\
&\stackrel{(b)}{\sim} \frac{1}{2k^2 \ln q} \sum_x w_x \sum_y \frac{[\nabla g_y(\mathbf{w})(\mathbf{e}_x - \mathbf{w})]^2}{g_y(\mathbf{w})} \\
&= \frac{1}{2k^2 \ln q} \sum_y \frac{1}{g_y(\mathbf{w})} \nabla g_y(\mathbf{w}) \left[\sum_x w_x (\mathbf{e}_x - \mathbf{w})(\mathbf{e}_x - \mathbf{w})' \right] \nabla' g_y(\mathbf{w}) \\
&= \frac{1}{2k^2 \ln q} \sum_y \frac{\nabla g_y(\mathbf{w}) [\text{diag}(\mathbf{w}) - \mathbf{w}\mathbf{w}'] \nabla' g_y(\mathbf{w})}{g_y(\mathbf{w})} \tag{A.24}
\end{aligned}$$

as $k \rightarrow \infty$, where (a) follows from Lemma A.2 and (b) follows from (A.21) and (A.20).

A.3 Useful Lemmas for Change of Variable in Sec. 5.2.2

The following lemma follows directly from the definitions of (5.18), (5.19), (5.20), and (5.21):

Lemma A.4. *We have*

$$\|\mathbf{v}_i^{q-1}\|_2 = \begin{cases} r, & i = 0 \\ v_{i-1} \tan \phi_i, & 1 \leq i \leq q-1 \end{cases} \quad (\text{A.25})$$

$$= r \prod_{l=1}^i \sin \phi_l, \quad 0 \leq i \leq q-1 \quad (\text{A.26})$$

$$\|\mathbf{h}_j^{q-1}\|_2 = \begin{cases} R, & j = 0 \\ h_{j-1} \tan \gamma_j, & 1 \leq j \leq q-1 \end{cases} \quad (\text{A.27})$$

$$= R \prod_{l=1}^j \sin \gamma_l, \quad 0 \leq j \leq q-1. \quad (\text{A.28})$$

Lemma A.5. *By the definition of (5.18), (5.19), (5.20), and (5.21) we have*

$$\begin{aligned} & (\mathbf{J}_\gamma^h)' \mathbf{J}_\gamma^h \\ &= \text{diag} \left(1, R^2, R^2 \sin^2 \gamma_1, R^2 \sin^2 \gamma_1 \sin^2 \gamma_2, \dots, R^2 \prod_{l=1}^{q-2} \sin^2 \gamma_l, 1 \right) \end{aligned} \quad (\text{A.29a})$$

$$\begin{aligned} & \mathbf{J}_v^\phi (\mathbf{J}_v^\phi)' \\ &= \text{diag} \left(1, r^{-2}, r^{-2} \sin^{-2} \phi_1, r^{-2} \sin^{-2} \phi_1 \sin^{-2} \phi_2, \dots, r^{-2} \prod_{l=1}^{q-2} \sin^{-2} \phi_l \right) \end{aligned} \quad (\text{A.29b})$$

$$\begin{aligned} & \mathbf{J}_v^\phi \mathbf{v} \mathbf{v}' (\mathbf{J}_v^\phi)' \\ &= \text{diag}(r^2, 0, \dots, 0). \end{aligned} \quad (\text{A.29c})$$

Proof. By (5.21) we have

$$\mathbf{J}_{\boldsymbol{\gamma}}^{\mathbf{h}} = \begin{bmatrix} h_0/R & -h_0 \tan \gamma_1 & 0 & \cdots & 0 & 0 \\ h_1/R & h_1 \cot \gamma_1 & -h_1 \tan \gamma_2 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ h_{q-2}/R & h_{q-2} \cot \gamma_1 & h_{q-2} \cot \gamma_2 & \ddots & -h_{q-2} \tan \gamma_{q-1} & 0 \\ h_{q-1}/R & h_{q-1} \cot \gamma_1 & h_{q-1} \cot \gamma_2 & \cdots & h_{q-1} \cot \gamma_{q-1} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

or

$$\begin{aligned} \left(\mathbf{J}_{\boldsymbol{\gamma}}^{\mathbf{h}}\right)_{ij} &= \frac{\partial h_i}{\partial \gamma_j} \\ &= \begin{cases} h_i/R, & j = 0, 0 \leq i \leq q-1 \\ \delta_{ij}, & \max(i, j) = q \\ 0, & 0 \leq i < j-1 < q-2 \\ -h_i \tan \gamma_j, & 0 \leq i = j-1 \leq q-2 \\ h_i \cot \gamma_j, & 1 \leq j \leq i \leq q-1. \end{cases} \end{aligned}$$

Now by

$$M_{ij} \triangleq \left[\left(\mathbf{J}_{\boldsymbol{\gamma}}^{\mathbf{h}}\right)' \mathbf{J}_{\boldsymbol{\gamma}}^{\mathbf{h}} \right]_{ij} = \left(\frac{\partial \mathbf{h}}{\partial \gamma_i} \right)' \left(\frac{\partial \mathbf{h}}{\partial \gamma_j} \right) = \sum_{l=0}^q \left(\frac{\partial h_l}{\partial \gamma_i} \right) \left(\frac{\partial h_l}{\partial \gamma_j} \right), \quad 0 \leq i \leq q$$

we can find $\{M_{ij}\}$ as follows (by symmetry it suffices to consider $0 \leq j \leq i \leq q$):

(i) $i = j = 0$:

$$M_{00} = \sum_{l=0}^{q-1} \frac{h_l^2}{R^2} = \frac{1}{R^2} \|\mathbf{h}_0^{q-1}\|_2^2 = 1.$$

(ii) $j = 0, 1 \leq i \leq q-1$:

$$\begin{aligned} M_{ij} &= \frac{h_{i-1}}{R} (-h_{i-1} \tan \gamma_i) + \sum_{l=i}^{q-1} \frac{h_l}{R} \cdot h_l \cot \gamma_i \\ &= \frac{1}{R} (-h_{i-1}^2 \tan \gamma_i + \|\mathbf{h}_i^{q-1}\|_2^2 \cot \gamma_i) \\ &\stackrel{(a)}{=} \frac{1}{R} (-h_{i-1}^2 \tan \gamma_i + h_{i-1}^2 \tan^2 \gamma_i \cot \gamma_i) = 0 \end{aligned}$$

where (a) follows from (A.27).

(iii) $1 \leq j = i \leq q - 1$:

$$\begin{aligned}
M_{ij} &= \sum_{l=i-1}^{q-1} \left(\frac{\partial h_l}{\partial \gamma_i} \right)^2 = h_{i-1}^2 \tan^2 \gamma_i + \sum_{l=i}^{q-1} h_l^2 \cot^2 \gamma_i \\
&\stackrel{(a)}{=} h_{i-1}^2 \tan^2 \gamma_i + h_{i-1}^2 \tan^2 \gamma_i \cot^2 \gamma_i \\
&\stackrel{(b)}{=} h_{i-1}^2 \sec^2 \gamma_i = R^2 \prod_{l=1}^{i-1} \sin^2 \gamma_l
\end{aligned}$$

where (a) follows from (A.27) and (b) follows from the Pythagorean identity.

(iv) $1 \leq j < i \leq q - 1$:

$$\begin{aligned}
M_{ij} &= \sum_{l=i-1}^{q-1} \left(\frac{\partial h_l}{\partial \gamma_i} \right) \left(\frac{\partial h_l}{\partial \gamma_j} \right) \\
&= (-h_{i-1} \tan \gamma_i)(h_{i-1} \cot \gamma_j) + \sum_{l=i}^{q-1} (h_l \cot \gamma_i)(h_l \cot \gamma_j) \\
&= -h_{i-1}^2 \tan \gamma_i \cot \gamma_j + \|\mathbf{h}_i^{q-1}\|_2^2 \cot \gamma_i \cot \gamma_j \\
&\stackrel{(a)}{=} -h_{i-1}^2 \tan \gamma_i \cot \gamma_j + h_{i-1}^2 \tan^2 \gamma_i \cot \gamma_i \cot \gamma_j \\
&= h_{i-1}^2 (-\tan \gamma_i \cot \gamma_j + \tan \gamma_i \cot \gamma_j) = 0
\end{aligned}$$

where (a) follows from (A.27).

(v) $i = q, 0 \leq j \leq q$: $M_{qj} = \delta_{qj}$ by inspection.

Therefore

$$\mathbf{M} \triangleq (\mathbf{J}_{\boldsymbol{\gamma}}^{\mathbf{h}})' \mathbf{J}_{\boldsymbol{\gamma}}^{\mathbf{h}} = \text{diag} \left(1, R^2, R^2 \sin^2 \gamma_1, R^2 \sin^2 \gamma_1 \sin^2 \gamma_2, \dots, R^2 \prod_{l=1}^{q-2} \sin^2 \gamma_l, 1 \right).$$

By (5.18) we have (all norms are Euclidean norm in the sequel)

$$\mathbf{J}_{\mathbf{v}}^{\phi} = \begin{bmatrix} v_0/r & v_1/r & \cdots & v_{q-2}/r & v_{q-1}/r \\ -\frac{\|\mathbf{v}_1^{q-1}\|}{\|\mathbf{v}_0^{q-1}\|} & \frac{v_0 v_1}{\|\mathbf{v}_1^{q-1}\| \|\mathbf{v}_0^{q-1}\|^2} & \cdots & \frac{v_0 v_{q-2}}{\|\mathbf{v}_1^{q-1}\| \|\mathbf{v}_0^{q-1}\|^2} & \frac{v_0 v_{q-1}}{\|\mathbf{v}_1^{q-1}\| \|\mathbf{v}_0^{q-1}\|^2} \\ 0 & -\frac{\|\mathbf{v}_2^{q-1}\|}{\|\mathbf{v}_1^{q-1}\|^2} & \ddots & \frac{v_1 v_{q-2}}{\|\mathbf{v}_2^{q-1}\| \|\mathbf{v}_1^{q-1}\|^2} & \frac{v_1 v_{q-1}}{\|\mathbf{v}_2^{q-1}\| \|\mathbf{v}_1^{q-1}\|^2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -\frac{\|\mathbf{v}_{q-1}^{q-1}\|}{\|\mathbf{v}_{q-2}^{q-1}\|^2} & \frac{v_{q-2} v_{q-1}}{\|\mathbf{v}_{q-1}^{q-1}\| \|\mathbf{v}_{q-2}^{q-1}\|^2} \end{bmatrix}$$

or

$$\begin{aligned} (\mathbf{J}_{\mathbf{v}}^{\phi})_{ij} &= \frac{\partial \phi_i}{\partial v_j} \\ &= \begin{cases} v_j/r, & i = 0, 0 \leq j \leq q-1 \\ 0, & 0 \leq j < i-1 \leq q-2 \\ -\frac{\|\mathbf{v}_i^{q-1}\|}{\|\mathbf{v}_{i-1}^{q-1}\|^2}, & 0 \leq j = i-1 \leq q-2 \\ \frac{v_{i-1} v_j}{\|\mathbf{v}_i^{q-1}\| \|\mathbf{v}_{i-1}^{q-1}\|^2}, & 1 \leq i \leq j \leq q-1. \end{cases} \end{aligned}$$

Now by

$$N_{ij} \triangleq \left[\mathbf{J}_{\mathbf{v}}^{\phi} (\mathbf{J}_{\mathbf{v}}^{\phi})' \right]_{ij} = \left(\frac{\partial \phi_i}{\partial \mathbf{v}} \right) \left(\frac{\partial \phi_j}{\partial \mathbf{v}} \right)' = \sum_{l=0}^{q-1} \left(\frac{\partial \phi_i}{\partial v_l} \right) \left(\frac{\partial \phi_j}{\partial v_l} \right), \quad 0 \leq i \leq q-1$$

we can find $\{N_{ij}\}$ as follows (by symmetry it suffices to consider $0 \leq j \leq i \leq q-1$):

(i) $i = j = 0$:

$$N_{00} = \sum_{l=0}^{q-1} \frac{v_l^2}{r^2} = \frac{1}{r^2} \|\mathbf{v}\|_2^2 = 1.$$

(ii) $j = 0, 1 \leq i \leq q-1$:

$$\begin{aligned} N_{ij} &= \frac{v_{i-1}}{r} \left(-\frac{\|\mathbf{v}_i^{q-1}\|}{\|\mathbf{v}_{i-1}^{q-1}\|^2} \right) + \sum_{l=i}^{q-1} \frac{v_l}{r} \cdot \frac{v_{i-1} v_l}{\|\mathbf{v}_i^{q-1}\| \|\mathbf{v}_{i-1}^{q-1}\|^2} \\ &= \frac{v_{i-1}}{r \|\mathbf{v}_{i-1}^{q-1}\|^2} \left(-\|\mathbf{v}_i^{q-1}\| + \frac{\|\mathbf{v}_i^{q-1}\|^2}{\|\mathbf{v}_i^{q-1}\|} \right) = 0. \end{aligned}$$

(iii) $1 \leq j = i \leq q - 1$:

$$\begin{aligned}
N_{ij} &= \sum_{l=i-1}^{q-1} \left(\frac{\partial \phi_i}{\partial v_l} \right)^2 = \frac{\|\mathbf{v}_i^{q-1}\|^2}{\|\mathbf{v}_{i-1}^{q-1}\|^4} + \sum_{l=i}^{q-1} \frac{v_{i-1}^2 v_l^2}{\|\mathbf{v}_i^{q-1}\|^2 \|\mathbf{v}_{i-1}^{q-1}\|^4} \\
&= \frac{1}{\|\mathbf{v}_{i-1}^{q-1}\|^4} \left(\|\mathbf{v}_i^{q-1}\|^2 + \frac{v_{i-1}^2}{\|\mathbf{v}_i^{q-1}\|^2} \cdot \|\mathbf{v}_i^{q-1}\|^2 \right) \\
&= \|\mathbf{v}_{i-1}^{q-1}\|^{-2} \stackrel{(a)}{=} r^{-2} \prod_{l=1}^{i-1} \sin^{-2} \phi_l
\end{aligned}$$

where (a) follows from (A.26).

(iv) $1 \leq j < i \leq q - 1$:

$$\begin{aligned}
N_{ij} &= \sum_{l=i-1}^{q-1} \left(\frac{\partial \phi_i}{\partial v_l} \right) \left(\frac{\partial \phi_j}{\partial v_l} \right) \\
&= \left(-\frac{\|\mathbf{v}_i^{q-1}\|}{\|\mathbf{v}_{i-1}^{q-1}\|^2} \right) \left(\frac{v_{j-1} v_{i-1}}{\|\mathbf{v}_j^{q-1}\| \|\mathbf{v}_{j-1}^{q-1}\|^2} \right) + \sum_{l=i}^{q-1} \left(\frac{v_{i-1} v_l}{\|\mathbf{v}_i^{q-1}\| \|\mathbf{v}_{i-1}^{q-1}\|^2} \right) \left(\frac{v_{j-1} v_l}{\|\mathbf{v}_j^{q-1}\| \|\mathbf{v}_{j-1}^{q-1}\|^2} \right) \\
&= \frac{v_{i-1} v_{j-1}}{\|\mathbf{v}_{i-1}^{q-1}\|^2 \|\mathbf{v}_j^{q-1}\| \|\mathbf{v}_{j-1}^{q-1}\|^2} \left(-\|\mathbf{v}_i^{q-1}\| + \frac{\|\mathbf{v}_i^{q-1}\|^2}{\|\mathbf{v}_i^{q-1}\|} \right) = 0.
\end{aligned}$$

Therefore

$$\mathbf{N} \triangleq \mathbf{J}_{\mathbf{v}} \boldsymbol{\phi} \left(\mathbf{J}_{\mathbf{v}} \boldsymbol{\phi} \right)' = \text{diag} \left(1, r^{-2}, r^{-2} \sin^{-2} \phi_1, r^{-2} \sin^{-2} \phi_1 \sin^{-2} \phi_2, \dots, r^{-2} \prod_{l=1}^{q-2} \sin^{-2} \phi_l \right).$$

Finally by

$$\tilde{v}_i \triangleq \left(\mathbf{J}_{\mathbf{v}} \boldsymbol{\phi} \right)_i = \sum_{l=0}^{q-1} \frac{\partial \phi_i}{\partial v_l} v_l, \quad 0 \leq i \leq q - 1$$

we can find $\{\tilde{v}_i\}$ as follows:

(i) $i = 0$:

$$\tilde{v}_0 = \sum_{l=0}^{q-1} \frac{v_l}{r} \cdot v_l = \frac{\|\mathbf{v}\|}{r} = r.$$

(ii) $1 \leq i \leq q-1$:

$$\begin{aligned}\tilde{v}_i &= \left(-\frac{\|\mathbf{v}_i^{q-1}\|}{\|\mathbf{v}_{i-1}^{q-1}\|^2} \right) \cdot v_{i-1} + \sum_{l=i}^{q-1} \frac{v_{i-1}v_l}{\|\mathbf{v}_i^{q-1}\|\|\mathbf{v}_{i-1}^{q-1}\|^2} \cdot v_l \\ &= \frac{v_{i-1}}{\|\mathbf{v}_{i-1}^{q-1}\|^2} \left(-\|\mathbf{v}_i^{q-1}\| + \frac{\|\mathbf{v}_i^{q-1}\|^2}{\|\mathbf{v}_{i-1}^{q-1}\|} \right) = 0.\end{aligned}$$

Therefore $\mathbf{J}_\mathbf{v}^\phi \mathbf{v} = r\mathbf{e}_0$ and therefore

$$\mathbf{J}_\mathbf{v}^\phi \mathbf{v} \mathbf{v}' \left(\mathbf{J}_\mathbf{v}^\phi \right)' = r^2 \mathbf{e}_0 \mathbf{e}_0' = \text{diag}(r^2, 0, \dots, 0). \quad \square$$

APPENDIX B

PROOFS FOR MULTIPLE ACCESS CHANNELS

Lemma B.1. *Let $\mathbf{A}_n, \mathbf{B}_n \in \mathbb{R}^{3 \times 3}$ be two sequences of positive semi-definite symmetric matrices satisfying $\|\mathbf{A}_n - \mathbf{B}_n\|_1 = O\left(\frac{1}{n}\right)$. Then*

$$\lambda_{\min}(\mathbf{A}_n) \geq \lambda_{\min}(\mathbf{B}_n) - O\left(\frac{1}{n}\right).$$

Proof. Let $\mathbf{A}_n = \mathbf{B}_n + \mathbf{\Delta}_n$ and let $\mathbf{u}_n$ be the unit vector corresponding to $\lambda_{\min}(\mathbf{A}_n)$. Then

$$\begin{aligned} \lambda_{\min}(\mathbf{A}_n) &= \mathbf{u}_n' \mathbf{A}_n \mathbf{u}_n = \mathbf{u}_n' \mathbf{B}_n \mathbf{u}_n + \mathbf{u}_n' \mathbf{\Delta}_n \mathbf{u}_n \\ &\geq \lambda_{\min}(\mathbf{B}_n) - \lambda_{\max}(\mathbf{\Delta}_n) = \lambda_{\min}(\mathbf{B}_n) - O\left(\frac{1}{n}\right). \quad \square \end{aligned}$$

Theorem B.2. [44, Theorem 1.1 with dimension $d = 3$] *Let $\mathbf{S}_1, \dots, \mathbf{S}_n$ be independent (but not necessarily identically distributed) random vectors in $\mathbb{R}^3$ with $\mathbf{E}\mathbf{S}_i = \mathbf{0}$ for all i . Let $\mathbf{S} = \sum_{i=1}^n \mathbf{S}_i$. Assume $\mathbf{M} = \text{Cov}(\mathbf{S}) > 0$ and let $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{M})$. Let*

$$\beta = \sum_{i=1}^n \beta_i, \quad \text{where } \beta_i \triangleq \mathbf{E}[\|\mathbf{M}^{-1/2} \mathbf{S}_i\|_2^3]$$

and let

$$\Delta(\mathcal{S}) = \sup_{A \in \mathcal{S}} |\Pr\{\mathbf{S} \in A\} - \Pr\{\mathbf{X} \in A\}|$$

where $\mathcal{S}$ is the class of all convex subsets of $\mathbb{R}^3$. Then there exists $c > 0$ independent n such that $\Delta(\mathcal{S}) \leq c\beta$.

Lemma B.3. *For fixed positive definite matrix $\mathbf{\Sigma} \in \mathbb{R}^d$, $\epsilon \in (0, 1)$, and $\delta_n = o(1)$, we have*

$$\mathcal{Q}_{\text{inv}}(\mathbf{\Sigma}, \epsilon + \delta_n) = \mathcal{Q}_{\text{inv}}(\mathbf{\Sigma}, \epsilon) + O(\delta_n) \tag{B.1}$$

Proof. Let $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \Sigma)$. Then for fixed $\mathbf{x} \in \mathbb{R}^d$ the mapping $\delta \rightarrow \Pr\{\mathbf{X} \leq \mathbf{x} + \delta \mathbf{1}\}$ is continuously differentiable and strictly increasing. Therefore we have

$$\Pr\{\mathbf{X} \leq \mathbf{x} + \delta_n \mathbf{1}\} = \Pr(\mathbf{X} \leq \mathbf{x}) + O(\delta_n)$$

and by the definition (6.6) of $\mathcal{Q}_{\text{inv}}$ we have (B.1). □

REFERENCES

- [1] Y.-W. Huang and P. Moulin, “Saddle-point solution of the fingerprinting capacity game under the marking assumption,” in *Proc. IEEE Intl. Symposium on Information Theory (ISIT 2009)*, 2009, pp. 2256–2260.
- [2] E. Amiri and G. Tardos, “High rate fingerprinting codes and the fingerprinting capacity,” in *Proc. 20th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2009)*, 2009, pp. 336–345.
- [3] P. Moulin, “Universal fingerprinting: Capacity and random-coding exponents,” May 2011. [Online]. Available: <http://arxiv.org/abs/0801.3837v3>
- [4] G. Tardos, “Optimal probabilistic fingerprint codes,” in *Proc. 35th ACM Symposium on Theory of Computing (STOC 2003)*, 2003, pp. 116–125.
- [5] D. Boneh and J. Shaw, “Collusion-secure fingerprinting for digital data,” *IEEE Trans. Inf. Theory*, vol. 44, no. 5, pp. 1897–1905, Sept. 1998.
- [6] B. Škorić, S. Katzenbeisser, H. G. Schaathun, and M. U. Celik, “Tardos fingerprinting codes in the combined digit model,” *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 906–919, Sept. 2011.
- [7] Y.-W. Huang and P. Moulin, “On the saddle-point solution and the large-coalition asymptotics of fingerprinting games,” *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 1, pp. 160–175, Feb. 2012.
- [8] V. Strassen, “Asymptotische abschätzungen in Shannons informationstheorie,” *Trans. Third. Prague Conf. Inf. Th.*, pp. 689–723, 1962.
- [9] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Channel coding rate in the finite blocklength regime,” *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [10] M. Hayashi, “Information spectrum approach to second-order coding rate in channel coding,” *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.

- [11] V. Y. F. Tan and O. Kosut, "On the dispersions of three network information theory problems," Feb. 2012. [Online]. Available: <http://arxiv.org/abs/1201.3901v2>
- [12] E. MolavianJazi and J. N. Laneman, "Multiaccess communication in the finite blocklength regime," in *Information Theory and Applications Workshop (ITA), 2012*, Feb. 2012.
- [13] E. MolavianJazi and J. N. Laneman, "A random coding approach to Gaussian multiple access channels with finite blocklength," in *50th Annual Allerton Conference on Communication, Control, and Computing*, 2012.
- [14] S. Verdú, "Non-asymptotic achievability bounds in multiuser information theory," in *50th Annual Allerton Conference on Communication, Control, and Computing*, 2012.
- [15] R. Ahlswede, "Multi-way communication channels," in *Proceedings of 2nd International Symposium on Information Theory*, 1971, pp. 23–52.
- [16] H. H. J. Liao, "Multiple access channels," Ph.D. dissertation, University of Hawaii, Honolulu, 1972.
- [17] G. Dueck, "Maximal error capacity regions are smaller than average error capacity regions for multi-user channels," *Problems of Control and Information Theory*, vol. 7, no. 1, pp. 11–19, 1978.
- [18] G. Dueck, "The strong converse to the coding theorem for the multiple-access channel," *J. Comb. Inform. Syst. Sci.*, vol. 6, pp. 187–196, 1981.
- [19] R. Ahlswede, "An elementary proof of the strong converse theorem for the multiple-access channel," *J. Comb. Inform. Syst. Sci.*, vol. 7, no. 3, pp. 216–230, 1982.
- [20] J. Scarlett, A. Martinez, and A. G. i Fabregas, "Second-order rate region of constant-composition codes for the multiple-access channel," Jul 2013. [Online]. Available: arxiv.org/abs/1303.6167v2
- [21] K. Nuida, M. Hagiwara, H. Watanabe, and H. Imai, "Optimization of Tardos's fingerprinting codes in a viewpoint of memory amount," in *Information Hiding: 9th Intl. Workshop, IH 2007, Saint Malo, France, Jun. 11-13, 2007, Revised Selected Papers*, ser. LNCS, vol. 4567. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 279–293.
- [22] T. Furon and L. Perez-Freire, "Worst case attacks against binary probabilistic traitor tracing codes," Aug. 2009, preprint, available at <http://arxiv.org/abs/0903.3480v2>.

- [23] B. Škorić, S. Katzenbeisser, and M. U. Celik, “Symmetric Tardos fingerprinting codes for arbitrary alphabet sizes,” *Designs, Codes and Cryptography*, vol. 46, no. 2, pp. 137–166, Feb. 2008.
- [24] W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd ed. New York, NY: John Wiley & Sons, 1968, vol. II.
- [25] L. A. Petrosjan and N. A. Zenkevich, *Game Theory*. River Edge, NJ: World Scientific, 1996.
- [26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: John Wiley & Sons, 2006.
- [27] M. Sion, “On general minimax theorems,” *Pacific J. Math.*, vol. 8, no. 1, pp. 171–176, 1958.
- [28] N. P. Anthapadmanabhan, A. Barg, and I. Dumer, “On the fingerprinting capacity under the marking assumption,” *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2678–2689, Jun. 2008.
- [29] T. Furon, A. Guyader, and F. C  rou, “On the design and optimization of Tardos probabilistic fingerprinting codes,” in *Information Hiding: 10th Intl. Workshop, IH 2008, Santa Barbara, CA, USA, May 19-21, 2008, Revised Selected Papers*, ser. LNCS, vol. 5284. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 341–356.
- [30] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1999.
- [31] Y.-W. Huang and P. Moulin, “Capacity-achieving fingerprint decoding,” in *Proc. First IEEE Intl. Workshop on Information Forensics and Security (WIFS 2009)*, 2009, pp. 51–55.
- [32] E. Amiri, “Fingerprinting codes: Higher rates, quick accusations,” Ph.D. dissertation, Simon Fraser University, Burnaby, BC, Canada, 2010.
- [33] F. Bavaud, “Information theory, relative entropy and statistics,” in *Formal Theories of Information: From Shannon to Semantic Information Theory and General Concepts of Information*, ser. LNCS, vol. 5363. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 54–78.
- [34] W. Hoeffding, “Probability inequalities for sums of bounded random variables,” *J. Amer. Statistical Assoc.*, vol. 58, no. 301, pp. 13–30, Mar. 1963.
- [35] G. G. Lorentz, *Bernstein Polynomials*, 2nd ed. New York, NY: AMS Bookstore, 1986.

- [36] B. Škorić, T. Vladimirova, M. Celik, and J. Talstra, “Tardos fingerprinting is better than we thought,” *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3663–3676, Aug. 2008.
- [37] A. Somekh-Baruch and N. Merhav, “Achievable error exponents for the private fingerprinting game,” *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1827–1838, May 2007.
- [38] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*. New York, NY: John Wiley & Sons, 2000.
- [39] D. Blackwell, L. Breiman, and A. J. Thomasian, “The capacity of a class of channels,” *The Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1229–1241, Dec. 1959.
- [40] P. Moulin and J. A. O’Sullivan, “Information-theoretic analysis of information hiding,” *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 563–593, Mar. 2003.
- [41] D. Boesten and B. Škorić, “Asymptotic fingerprinting capacity for non-binary alphabets,” in *Information Hiding: 13th Intl. Conf., IH 2011, Prague, Czech Republic, May 18-20, 2011, Revised Selected Papers*, ser. LNCS, vol. 6958. Springer-Verlag, 2011, pp. 1–13.
- [42] S. V. Gelfand, I. M.; Fomin, *Calculus of Variations*. Prentice-Hall, 1963, vol. I.
- [43] B. Dacorogna, *Introduction to the Calculus of Variations*. Imperial College Press, 2004.
- [44] V. Bentkus, “A Lyapunov-type bound in R^d ,” *Theory of Probability & Its Applications*, vol. 49, no. 2, pp. 311–323, 2005.
- [45] J.-J. Oosterwijk, B. Škorić, and J. Doumen, “Optimal suspicion functions for Tardos traitor tracing schemes,” in *Proc. First ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec 2013)*, 2013, pp. 19–28.