On fundamental limits and design of explicit schemes for multi-user networks

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the

Graduate School of The Ohio State University

By

Mohammad Shahmohammadi,

Graduate Program in Electrical and Computer Engineering

The Ohio State University

2011

Dissertation Committee:

Prof. Hesham El Gamal, Adviser

Prof. C. Emre Koksal

Prof. Atilla Eryilmaz

ABSTRACT

Multi-user settings present new challenges to the problems of characterizing the fundamental limits and design of explicit schemes for a network. In the first direction of the dissertation we consider the problem of explicit scheme design for digital fingerprinting, which is a multi-user extension of watermarking. First, minimum distance decoding is studied and an explicit scheme based on graph codes under Belief Propagation decoding is proposed which achieves vanishing misidentification probability for rates as high as 0.11. This is a marked improvement over the best available designs in the literature. A specific coalition attack nicknamed "zero capacity attack" is also identified for which the minimum distance decoder fails to extend to larger coalition sizes than 2.

Next, we study the application of tree codes and sequential decoding to the problem and establish the existence of a good code which can achieve vanishing error probability with finite average complexity for rates below a cut-off rate. Using random Convolutional codes we provide numerical results indicating vanishing error probability for rates as high as 0.16. Our numerical results are extended to coalition size of 3 under the zero capacity collusion attack using a novel family of non-linear tree codes based on concatenation of random Convolutional codes and tree codes under bidirectional sequential decoding. We achieve vanishing error probability for rates as high as 0.05.

In the second direction of the thesis, we study the fundamental limits of multi-user communications. We begin by considering the X channel which contains most of the well studied multi-terminal networks (e.g., multiple-access, broadcast and interference channels). We propose a signaling scheme and derive the achievable rate region. The achievable region includes the best known rate regions for the special cases and outperforms the best available one in the literature. We conclude this direction of the thesis by studying cognitive cooperation in downlink communication for cellular systems. A singling scheme based on joint interference alignment and dirty paper coding is proposed and the achievable sum degrees of freedom region is characterized. Moreover, an outer bound on this region is derived and it is established that our proposed scheme is optimal for some special cases.

to Maman and Baba

ACKNOWLEDGMENTS

First of all, I would like to express my gratitude towards my parents, my brother and sister for their support throughout my PhD work.

I would like to thank my thesis advisor, Prof. Hesham El Gamal, for all his advice, support and guidance during my graduate studies. His support and critical feedback on my work have been immensely fruitful.

I also extend my gratitude to Prof. C. Emre Koksal and Prof. Atilla Eryilmaz for serving on my candidacy and dissertation committees. Their comments and suggestions were very helpful to me. I would like to thank Prof. Tamer Khattab of Qatar University for the helpful discussions as well as the financial support for my work. I thank Prof. Gulsah Akar for serving on my dissertation committee as well. I also thank Jeri McMichael for her help throughout my stay at OSU.

I would also like to thank my fellow IPS students, for the valuable and constructive discussions, creating a constructive environment for research. Here is a partial list in no particular order: Aditi Kothiyal, Lifeng Lai, Kambiz Azarian, Praveen Gopala, Arun Kannu, Arul Murugan, Ozan Koyluoglu, Young-Han Nam, Liza Toher, Sung-Jun Hwang, Sibasish Das, Naveen Ramakrishnan, Subhojit Som, Sugumar Murugesan, Justin Ziniel, Onur Gungor, Derya Gol, Arun Sridharan, Rohit Aggrawal and Rahul Srivatsava.

PUBLICATIONS

Research Publications

M. Shahmohammadi, O. Koyluoglu, T. Khattab and H. El Gamal, "On the Degrees of Freedom of The Cognitive Broadcast Channel," *to be Submitted to IEEE Tran. Comm.*

M. Shahmohammadi, O. Koyluoglu, T. Khattab and H. El Gamal, "On the Degrees of Freedom of The Cognitive Broadcast Channel," *Submitted to IEEE Int. Symp. on Info. Theory*

M. Shahmohammadi, O. Koyluoglu, T. Khattab and H. El Gamal, "Joint Interference Cancellation and Dirty Paper Coding for Cognitive Cellular Networks," *To appear in Proc. of IEEE Wireless Comm. and Netw. Conference.*

O. Koyluoglu, M. Shahmohammadi and H. El Gamal, "An Achievable Rate Region for the Discrete Memoryless X Channel," *in Proc. IEEE Int. Symp. on Info. Theory (ISIT) 2009, pp. 2427-2431.*

M. Shahmohammadi and H. El Gamal, "Tree Search for Digital Fingerprinting," *in Proc. Allerton Conf. on Comm., Cont. and Comp. 2008, pp. 842-849.*

S. C. Lin, M. Shahmohammadi, and H. El Gamal, "Fingerprinting With Minimum Distance Decoding," *IEEE Trans. on Info. For. and Security*, Vol.4, pp.59-69, Mar. 2009.

S. C. Lin, M. Shahmohammadi, and H. El Gamal, "Fingerprinting With Minimum Distance Decoding," *Proc. Allerton Conf. on Comm., Cont. and Comp. 2007.*

FIELDS OF STUDY

Major Field: Electrical and Computer Engineering

TABLE OF CONTENTS

Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Shannon laid down the mathematical foundation of communications also known as *information theory* in his seminal work [1]. Among fundamental ideas of [1] are: Typical sequences, random coding argument and typical set decoding. Because the main objective in information theory is to characterize the fundamental limits of information transmission, apart from fundamental constraints (e.g., transmission power limit) no practical limitation (e.g., decoding complexity) is assumed for either the transmitter(s) or receiver(s). Achieving the reliable rates promised by the information theoretic results using real-world systems poses many interesting problems which arise from shortcomings of practical systems. For example, the encoder-decoder pair used in Shannon's work are based on typical sequences. This means that the encoder needs to maintain a look up table for messages the size of which grows *exponentially* with the code length (i.e., the number of channel uses). Also, the decoder needs to search the entire codebook for a typical pair which again means exponential decoding complexity.

Achieving near capacity results, with tolerable and not the prohibitive exponential complexity fueled a lot of research effort over the past decades. An interesting problem posed after derivation of capacity bounds, is the design of schemes that can

achieve or approach those bounds with a tolerable complexity and will be loosely referred to as *explicit schemes*. In general, information theoretic results not only characterize the limits but also shed light on the structure of optimal codes and inspire design of explicit schemes. The main focus of the original work of [1] is point to point (i.e, single user) communication. Multi user scenarios present new dimensions to the problem both in terms of characterizing capacity bounds and designing efficient coding schemes as well.

In the first direction of the thesis we study the problem of explicit design for digital fingerprinting. With the development of file sharing applications and accessibility of high speed Internet, the protection of copy righted data such as image, music, video and software has become increasingly important. Watermarking is a powerful tool employed for the protection of copyrighted data from illegal distribution. Loosely speaking, under a watermarking strategy a short redundant data is inserted in the copyrighted data before distribution to the users. In other words, before the distributor distribute its copy righted data among a set of licensed users, it embeds a unique watermark in each licensed copy.

Fingerprinting can be thought of as a *multi-user* extension of watermarking schemes. The goal of the users that copy the data in an illegal manner is to minimize the probability of being traced back by the distributer. In a multi-user scenario, when the user can collaborate to produce a forgery of their copy-righted files, identifying the guilty users becomes significantly more challenging for the distributer. This difficulty can be compounded by the fact that usually the distributer is ignorant about the strategy with which the forged copy is produced. The fundamental limits of a fingerprinting system in terms of the length of the inserted mark, the number of colluding

users (coalition size) and a reliability measure is studied in [2–4]. As it is typical in information theoretic analysis the aforementioned contributions are not constrained in terms of encoding/decoding complexity.

The first direction of this thesis aims at designing explicit schemes for fingerprinting. Two different approaches based on *graph codes* and *tree codes* are devised. We achieve vanishing error probability with tolerable complexity of encoding and decoding using the proposed schemes. Our contribution in this direction of the thesis is summarized below:

First, an achievable rate for binary linear codes in a fingerprinting system under minimum distance decoding is derived. Random graph codes and belief propagation decoding emerge as a natural fit to approximate the minimum distance decoding. Next, we provide simulation results for a recently proposed class of low rate graph codes under coalition attack with belief propagation decoding. A specific attack for coalition sizes $t > 2$ is identified for which the minimum distance decoding fails.

In the second part of this direction of the thesis, we consider the application of tree codes and tree search decoding to digital fingerprinting. It is a well known fact that the complexity of tree search decoding is a random variable. Unlike other powerful available decoding algorithms i.e., belief propagation and Viterbi decoding for which the complexity of both is constant in the code-length. The complexity of tree search decoding under the considered fingerprinting system is studied in Chapter 3 and the existence a good tree code with finite complexity under the proposed tree search decoding is established when the coalition size is equal to 2. This result improves over the achievable rate under the minimum distance decoding. For the coalition attack

under which the minimum distance decoder fails, we propose a novel non-linear construction for tree codes and study its performance numerically. Numerical results indicate vanishing error probability with finite complexity of decoding for this attack. Our results in the first direction of the thesis is reported in [5–7]

In the second direction of the thesis we consider the fundamental limits of multi-user networks. Shannon's original work on point to point channels have been extended to multi-user cases in many different directions. The following significant examples are worth mentioning:

- The many to one communication or *multiple access channel* for which the capacity region is fully known and can be achieved by typical set encoding and decoding [8]

- One to many communication or *broadcast channel*. The capacity region is known for the special case of *degraded* broadcast channel using superposition encoding and typical set decoding [9].

- Interference channel: which is a communication scenario with two transmitters and two receivers. Each receiver, receives an intended signal from its corresponding transmitter and an unintended one which is *interference* from the other transmitter.

In the first part we consider a general model of two user communication which will be referred to as the "X" channel. The X channel represents a communication model with two transmitters and two receivers. Unlike the conventional interference channel, in the X channel each transmitter has a message for each receiver. The X channel was

first considered in [10] and the achievable rate region and degrees of freedom region was studied using dirty paper coding and interference alignment. Our contribution in this part is that by borrowing ideas from Marton's binning scheme for broadcast channel [11] , Han and Kobayashi's message splitting technique [12] and superposition codebook [9] we propose a new achievable rate region which improves over and contains the previously proposed scheme of [10] and it is the best available rate region in literature. Our result in this part is reported in [13].

In the second part of this direction of the thesis we consider the recently proposed idea of cognitive cooperation. Cognitive cooperation has been recently proposed to improve the performance of the existing (single frequency, single protocol) wireless systems [14, 15]. In this work we consider the idea of cognitive cooperation in a cellular network. More specifically we assume a primary base station and a secondary one. The secondary base station is assumed to know the messages of the primary base station *non-causally* and thus is referred to as cognitive. The validity of this assumption hinges on the fact that in many cellular applications the bases stations can be assumed to be connected by high capacity links.

One of the improvements offered by the cognitive technology is better efficiency in exploiting the available resources. Traditionally it is assumed that a cognitive device is capable of sensing if a dimension (e.g., a frequency band, a time sharing slot) is idle and in that case transfer its transmission to the unused band allowing for an increase in the overall efficiency of the system. In our work, we do not confine the cognitive node to only transmitting in the unused parts of the spectrum domain. We consider the general case where the bases stations can transmit over the same dimension at the same time.

Transmission over the same dimension could cause *interference* on unintended receivers. An important component in our contribution is the manner in which the interference is handled and/or mitigated. In our work we apply the following approaches to manage different types of interference :

- Dirty paper coding: When the channel interference is non-causally known at the transmitter, its capacity is derived by Gelfand and Pinsker [16]. The coding scheme proposed by [16] involves a binning structure and the obtained formula contains the channel's probability distribution, a distribution on its input alphabet as well as an auxiliary random variable.

  For the special case of an additive Gaussian channel the capacity of the corresponding interference channel serves as an outer bound for the channel that experiences interference. When the known interference at the transmitter is distributed according to a Gaussian distribution, Costa [17] determined the optimal distribution of the auxiliary random variable in the Gelfand Pinsker scheme which achieves the outer bound. The scheme with the given distribution is coined as *dirty paper coding*.

  In our proposed scheme interference mitigation in the secondary cell is achieved by dirty paper coding in conjunction with the common zero forcing beamforming approach in a multiple input multiple output (MIMO) broadcast channel.

- Interference alignment: In general, determining the exact capacity region is often an intractable problem for a multi-user network. To mover forward, the notion of *"degrees of freedom"* has been proposed.

The degrees of freedom (or the multiplexing gain) can be thought of as the scaling rate of the network's capacity with $\log(SNR)$ as $SNR \to \infty$ where SNR denotes the signal to noise ratio. Basically, in the high SNR region the degrees of freedom is an approximation of the capacity within a constant independent of the SNR.

Broadly speaking, we can generate degrees of freedom by expansions in time (e.g., through multiple fading blocks), in frequency (e.g., by using multiple carriers) and in space if the communication nodes are equipped with multiple antennas. Several approaches have been proposed to efficiently exploit the multiplexing gains offered by spatial dimensions [18,19]. One of the significant ideas in this arena is *interference alignment* which was first proposed to achieve the full multiplexing gains of a K user interference channel [20].

Intuitively speaking, by a certain design of beam-forming vectors the signals at the unindented receivers are aligned to some dimensions using the proposed scheme of [20].

By doing so, we can zero force them together which means by discarding the dimensions dedicated to interference, we will be able to remove the interference completely. Next, the intended data can be recovered on the rest of dimensions available at each receiver. Interference alignment is a fundamental component along with zero forcing in achieving the full degrees of freedom region for a two user multiple input multiple output X channel [21] too.

The idea of zero forcing the unintended data altogether, by aligning them can also be extended to different types of interference as well. In [22], a cellular system is considered and at each mobile user the inter-cell and intra-cell interferences are aligned

to the same linear space and thus can be mitigated together.

Upper and lower bounds on the capacity region of an interference channel with a cognitive transmitter has been studied in [23, 24]. The achievable degrees of freedom region under cognition for interference and X channels is considered in [20,21,25]. The cognitive paradigm has been studied for a cellular setting (for example see [26–28]) however many fundamental problems regarding a cognitive cellular network remains open.

Our contribution in this part of the thesis is summarized in the following. We propose a general model for the downlink communication in a cognitive cellular network. Next borrowing ideas from dirty paper coding, interference alignment and zero forcing we propose a signaling scheme and study the achievable sum degrees of freedom region in the cognitive and non-cognitive cells.

We also present an outer bound on the performance of a cognitive system in the downlink and determine some special cases for which our proposed signaling scheme is optimal.

In addition, as mentioned earlier achieving the gains promised by interference alignment hinges on the careful design of beam-forming vectors. To determine the correct directions all of the transmitters are required to perfectly know the channel state information (CSI) of each other. In our work, we consider the case when full CSI is not available at the non-cognitive base station. The results of this part of the thesis are reported in [29, 30].

CHAPTER 2

FINGERPRINTING WITH MINIMUM DISTANCE DECODING

In this and the next chapter the problem of designing explicit schemes for digital fingerprinting is considered. We adopt an information theoretic framework for the design of collusion-resistant coding/decoding schemes for digital fingerprinting. More specifically, in this chapter the minimum distance decision rule is used to identify one of the colluding users. Achievable rates, under this detection rule, are characterized under any collusion attack that satisfies the marking assumption. For $t = 2$ pirates, we show that rate $1 - H(0.25) \approx 0.188$ is achievable using an ensemble of random linear codes. Inspired by our theoretical analysis, we then construct coding/decoding schemes for fingerprinting based on the celebrated Belief-Propagation framework. Using an explicit accumulate repeat accumulate code we obtain a vanishingly small probability of misidentification at rate $1/9$ with $t = 2$. These results represent a marked improvement over the best available designs in the literature.

2.1 Introduction

Digital fingerprinting is a paradigm for protecting copyrighted data against illegal distribution [31] and have lately received a lot of attention particularly from the information theory community. In a nutshell, a *distributor* wishes to distribute its data $\mathbb{D}$ among a number of licensed *users*. To identify each user, a set of redundant digits referred to as a *marks* or *fingerprints* are embedded inside the copyrighted data. The locations of the marks are kept *hidden* from the users and are only known to the distributor. Their positions, however, remain the same for all users.

If any user inadvertently re-distributes its assigned copy, it will be caught by its fingerprint. However, several users may collude to form a *coalition* enabling them to produce an unauthorized copy (also referred to as *forged copy* or *forgery*) which is difficult to trace back. In the literature, the colluding members are typically referred to as *pirates* or *colluders*. Hence, the need arises for the design of collusion-resistant digital fingerprinting techniques. The focus of the first direction of the thesis is to develop an information theoretic framework for the design of low complexity *pirate-identification* schemes which are loosely referred to as *explicit schemes*. To this end, in this chapter we consider graph codes and minimum distance decoding and the next chapter is dedicated to tree codes and tree search decoding.

Based on the constraints faced by the coalition to devise a collusion attack, different settings have been considered for the fingerprinting problem. There are two main settings in the literature:

- The *distortion* setting, in which the "difference" between forged copy and each of the licensed copies cannot exceed a certain threshold

- The *marking assumption* setting which is a popular model for software fingerprinting and roughly states that the pirates can only change the coordinates where they see a difference in their assigned fingerprint.

In this thesis, we consider the *marking assumption* setting first proposed in [31]. In this framework, the pirates attempt to identify the positions occupied by the fingerprinting digits by comparing their copies. Afterwards, they can *only* modify the identified coordinates, in any desired way, to minimize the probability of traceability. The validity of the marking assumption hinges on the assumption that any modification to the data content $\mathbb{D}$ could damage it permanently. This prevents the users from modifying any location in which they do not identify as a fingerprinting digit since it *may be* a data symbol. For more on fingerprinting, the validity of the marking assumption and motivations behind it we refer the readers to [2,31,32] and references therein.

Boneh and Shaw [31] were the first to construct fingerprinting codes that are resistant to attacks that satisfy the marking assumption. This approach was later extended in [32] using the idea of separating codes [33]. To the best of our knowledge, the best available explicit binary fingerprinting codes are the *low rate* codes presented in [32]. For example, for $t = 2$, the best available code has a rate≈ 0.0092.

Upper and lower bounds on fingerprinting capacity under the marking assumption was first considered in [2]. In a recent publication [3] Moulin unified the different settings and formulations of fingerprinting and derived upper and lower bounds on capacity under various settings. The decoders presented in [2,3], however, is based on exhaustive search, and hence, would suffer from an exponentially growing complexity in the code-length. This prohibitive complexity motivates our proposed approaches.

In this chapter, we show that using linear fingerprinting codes and minimum distance (MD) decoding, one can achieve rates less than 0.188 when the coalition size is $t = 2$.

Since the complexity of the *exact* MD decoder can be prohibitive when the code-length is long, we develop a low complexity belief-propagation identification approach [34]. This detector only requires a linear complexity in the code-length n, and offers remarkable performance gain over the best known explicit construction [34]. For the marking assumption set-up, we achieve a vanishingly small misidentification probability for rates up to 1/9 using the a recently proposed class of low rate accumulate repeat accumulate (ARA) codes [35]. It is worth noting that these results represent a marked improvement over the state of the art in the literature. Furthermore, one would expect additional performance enhancement by optimizing the degree sequences of the codes (which is beyond the scope of this work).

The rest of this chapter is organized as follows: In Section 2.2, we introduce the mathematical notations and formally define our problem setup. Then we explore the theoretical limits of fingerprinting using the MD decoder in Section 2.3. The simulation results based on the BP framework are presented in Section 2.4. Finally, Section 2.5 offers some concluding remarks.

2.2 Notations and Problem statement

Throughout the thesis the entropy function is denoted by $H(\cdot)$, with the argument being the probability mass function as in [8]. For simplicity, we denote $H(p, 1-p)$ as $H(p)$, where $0 \leq p \leq 1$.

Moreover, for two functions of n, we write $a(n) \doteq b(n)$ if:

$$\lim_{n\to\infty} \frac{1}{n} \frac{a(n)}{b(n)} = 0, \tag{2.1}$$

for example, $\binom{n}{d} \doteq 2^{nH(\frac{d}{n})}$.

Also random variables are denoted by capital letters and their realization with smaller case letter. Vectors of length n are denoted by bold-face letters. For example, $\mathbf{x}$ denotes the vector $\{x_1, x_2, \ldots, x_n\}$. $d_H(\mathbf{x}_1, \mathbf{x}_2)$ denotes the Hamming distance between $\mathbf{x}_1, \mathbf{x}_2$.

Assume there are M users denoted by $\{1, 2, \ldots, M\}$ in the fingerprinting system. In general, a coalition of size t is a subset U of $\{1, 2, \ldots, M\}$ where $|U| = t$. We will concentrate on case $t = 2$, i.e. $U = \{u_1, u_2\}$.

In general we can not confine the coalition to employ a specific type of attack. We consider a formulation that can capture all attacks under the marking assumption. It should be noted that, we use a similar mathematical formulation presented in [2], for completeness however, we repeat it here. As mentioned in [31], deterministic fingerprinting under the marking assumption is not possible in general. Therefore, the distributor needs to employ some kind of randomization.

A collection of binary codes (F, G) is composed of K pairs of encoding and decoding functions as:

$$f_k : \{1, 2, \ldots, M\} \to \{0, 1\}^n$$

$$g_k : \{0, 1\}^n \to \{1, 2, \ldots, M\}$$

$$k = 1, 2, \ldots, K.$$

The code rate R is equal to $\frac{\log_2 M}{n}$

The secret key, k is a random variable employed to randomize the codebook family. By applying randomization, the very codebook utilized for fingerprinting is also kept hidden from the users. It should be noted that, adhering to the common conventions in cryptography, the encoding and decoding functions as well as the probability distribution of the secret key, $p(k)$, are known to all users.

The fingerprints corresponding to the coalition of users (also referred to as pirates or colluders), u_1, u_2 are denoted by $f_k(u_1, u_2) = \{\mathbf{x}_1, \mathbf{x}_2\}$. A position i, in the coalition's fingerprints is called *undetectable* [31] if $x_{1i} = x_{2i}$, otherwise it is called *detectable*. Define $E(U)$ the set of feasible forged copies for a coalition U by:

$$E(U) = \{\mathbf{y} \in \{0,1\}^n \mid y_i = x_{1i}, \forall i \quad \textit{undetectable}\}. \tag{2.2}$$

In general, coalition U utilizes a random strategy that satisfies the marking assumption to produce $\mathbf{y}$. That is, let $V(\mathbf{y} \mid \mathbf{x}_1, \mathbf{x}_2)$ be the probability that $\mathbf{y}$ is created, given coalition is $\mathbf{x}_1, \mathbf{x}_2$, then we will have:

$$V(\mathbf{y} \mid \mathbf{x}_1, \mathbf{x}_2) = 0 \qquad \text{for all } \mathbf{y} \notin E(U). \tag{2.3}$$

Denote the set of all strategies that satisfies (2.3) by $\mathcal{V}$. The probability of misidentification for a given coalition U, in the codebook family (F, G) which utilizes the collusion strategy V, is defined by:

$$P_m(U, F, G, V) \;:=\; \mathbb{E}_K\Big(\sum_{\mathbf{y} \in \mathbf{E(U)}, \mathbf{g_k(y)} \notin \mathbf{U}} V(\mathbf{y} \mid f_K(U)) \Big) \tag{2.4}$$

Maximum probability of misidentification under any attack that satisfies the marking assumption, averaged over all possible coalitions is considered. The maximum

probability of misidentification is more convenient to analyze and is defined as:

$$\overline{P}_m(F,G) := \frac{1}{\binom{M}{t}} \sum_{U} \max_{V \in \mathcal{V}} P_m(U,F,G,V). \tag{2.5}$$

2.3 General Collusion Attack with MD Decoder

In this section we study the performance of minimum distance decoder when the coalition employs *any* strategy as long as the marking assumption is satisfied. The problem is studied from an information theoretic perspective.

First achievable rates are derived using random coding arguments, next we present our numerical results to show that the minimum distance decoder can be well approximated by belief propagation.

Before proceeding with our main results, the concept of *close* pairs of binary vectors is defined and a lemma is proved about the probability of two randomly picked vectors from a random coding ensemble being close.

For a small ε, two binary vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ of length n are said to be *close* if:

$$d_H(\mathbf{x}_1, \mathbf{x}_2) \leq n(\frac{1}{2} + \varepsilon).$$

A closed pair $(\mathbf{x}_1, \mathbf{x}_2)$ is denoted by:

$$\mathbf{x}_1 \overset{C}{\leftrightarrow} \mathbf{x}_2.$$

The pair $(\mathbf{x}_1, \mathbf{x}_2)$ that is not close is referred to as *non-close*. A non-close pair is denoted by:

$$\mathbf{x}_1 \overset{N}{\leftrightarrow} \mathbf{x}_2$$

The following lemma establishes the probability of a randomly picked pair from an ensemble of random binary and linear random binary codes being close.

Lemma 2.1. *For both ensembles of codes the probability of a randomly selected pair of codes being non-close is vanishingly small.*

Proof. An ensemble of binary random codes is composed of random binary codewords each coordinate of the codeword drawn identically and independently with the same probability of being 0 and 1. An ensemble of linear random binary code is defined as a linear code whose parity check matrix is a random matrix all elements of which are drawn identically and independently with the same probability of being 0 and 1.

Let us consider the i.i.d random ensemble first. For a code C in the i.i.d ensemble and $1 \leq d \leq n$ define:

$$S_c(d) = \sum_{i=1}^{M} \sum_{j=1}^{i-1} \Phi\{d_H(\mathbf{x}_i, \mathbf{x}_j) = d\}. \tag{2.6}$$

where $\Phi(\cdot)$ is the indicator function. In [36], it is established that with probability going to one as $n \to \infty$:

$$S_c(d) = \begin{cases} 2^{n(2R+H(\frac{d}{n})-1)} & n\delta_{GV}(2R) < d < n(1-\delta_{GV}(2R)) \\ 0 & \text{otherwise} \end{cases} \tag{2.7}$$

where $\delta_{GV}(\cdot)$ is the Gilbert-Varshamov distance. The Gilbert-Varshamov distance is defined as

$$\delta_{GV}(R) = \begin{cases} \delta & H(\delta) = 1 - R \text{ such that, } \delta < 0.5 \text{ and } R < 1 \\ 0 & R \geq 1 \end{cases} \tag{2.8}$$

Using (2.7), The probability that a randomly picked pair from an i.i.d random codebook being non-close can be written as:

$$\frac{\sum_{d>n(1/2+\epsilon)}^{n(1-\delta_{GV}(2R))} 2^{n(2R+H(d/n)-1)}}{2^{2nR}} < \frac{n2^{n(2R-1+H(\frac{1}{2}+\epsilon))}}{2^{2nR}}, \tag{2.9}$$

which goes exponentially to zero as $n \to \infty$.

Now let us turn to the random linear ensemble. For a code C in the linear ensemble

and $1 \leq d \leq n$ by the symmetry of the linear ensemble and considering the hamming distance of the code, it can be stated that

$$S_c(d) = \sum_{i=1}^{M} \sum_{j=1}^{i-1} \Phi\{d_H(\mathbf{x}_i, \mathbf{x}_j) = d\} = \frac{1}{2} \sum_{i=1}^{M} \sum_{j \neq i} \Phi\{d_H(\mathbf{x}_i, \mathbf{x}_j) = d\} = \frac{M}{2} N_c(d) \doteq 2^{nR} N_c(d), \tag{2.10}$$

where $N_c(d) \triangleq \sum_{j \neq i} \Phi\{d_H(\mathbf{x}_i, \mathbf{x}_j) = d\}$.

In [36], it is shown that with probability going to one as $n \to \infty$

$$N_c(d) \doteq \begin{cases} 2^{n(R+H(d/n)-1)}, & n\delta_{GV}(R) < d < n(1-\delta_{GV}(R)) \\ 0, & \text{otherwise.} \end{cases} \tag{2.11}$$

Therefore, the average probability of a pair being non-close can be written as

$$\frac{\sum_{d>n(1/2+\epsilon)}^{n(1-\delta_{GV}(R))} 2^{n(2R+H(d/n)-1)}}{2^{2nR}} < \frac{n 2^{n(2R-1+H(\frac{1}{2}+\epsilon))}}{2^{2nR}}, \tag{2.12}$$

which again goes exponentially to zero as $n \to \infty$. □

The first part of our contribution is presented in the following theorems:

Theorem 2.1. *For all rates less than* $1 - H(0.25)$ *there is a fingerprinting code for which the probability of misidentification goes to zero as the code-length goes to infinity, when* $t = 2$.

Proof. We use a random coding argument to establish the above theorem. Consider the ensemble of binary random codes as the following: A binary random vectors (fingerprints) of length n is assigned to the $M = 2^{nR}$ users where each coordinate is chosen independently at random with equal probability of being $0, 1$.

We can upper bound the misidentification probability by considering two events: First, if the assigned fingerprints to users u_1 and u_2 compose a non-close, an error is declared. Even if the minimum distance decoder can catch a pirate we consider the

aforementioned event as an error. Next, we try to upper bound the misidentification probability given the forged copy is produced by a close pair.

That is, given a forged fingerprint $\mathbf{y}$, the probability of misidentification can be upper bounded by:

$$P_m \leq P_m(\mathbf{y}|\mathbf{x}_1 \overset{C}{\leftrightarrow} \mathbf{x}_2) + P(\mathbf{x}_1 \overset{N}{\leftrightarrow} \mathbf{x}_2),$$

where $P_m(\mathbf{y}|\mathbf{x}_1 \overset{C}{\leftrightarrow} \mathbf{x}_2)$ is the misidentification probability when $\mathbf{y}$ is produced by a close pair $(\mathbf{x}_1, \mathbf{x}_2)$ and $P(\mathbf{x}_1 \overset{N}{\leftrightarrow} \mathbf{x}_2)$ is the probability that the pirates did not constitute a close pair.

Through the following argument, we show that the average of those probabilities over the random coding ensemble goes exponentially to zero as n goes to infinity hence the proof.

In Lemma 2.1 have proved that $P(\mathbf{x}_1 \overset{N}{\leftrightarrow} \mathbf{x}_2)$ over the ensemble goes to zero as n goes to infinity.

Now we turn to $P_m(\mathbf{y}|\mathbf{x}_1 \overset{C}{\leftrightarrow} \mathbf{x}_2)$. Since $d_H(\mathbf{x}_1, \mathbf{x}_2) < n(\frac{1}{2} + \varepsilon)$, the Hamming distance of the forged copy $\mathbf{y}$ with at least one of the pirates must be less than

$$h(n) = n(\frac{1}{4} + \frac{\varepsilon}{2}),$$

due to the marking assumption. Without loss of generality we assume that pirate to be $\mathbf{x}_1$.

Using minimum distance decoding, misidentification occurs if there is another binary vector $\mathbf{z}$ of length n in the codebook such that $d_H(\mathbf{y}, \mathbf{z}) \leq d_H(\mathbf{y}, \mathbf{x}_1)$. Let the hamming distance between $\mathbf{y}$ and $\mathbf{z}$ be equal to l i.e, $d_H(\mathbf{y}, \mathbf{z}) = l$. Because, there are a total $\binom{n}{l}$ number of positions that $\mathbf{z}$ and $\mathbf{y}$ can disagree, given any $\mathbf{y}$ the probability of

such $\mathbf{z}$ be in the codebook is equal to:

$$\frac{\binom{n}{l}}{2^n} \tag{2.13}$$

Therefore, the total misidentification probability given the original fingerprints are close can be written as:

$$P_m(\mathbf{y}|\mathbf{x}_1 \overset{C}{\leftrightarrow} \mathbf{x}_2) = \frac{\sum_{i=1}^{h(n)} \binom{n}{i}}{2^n} \doteq 2^{-n(1-H(0.25))}$$

Using the union bound the total probability of misidentification in a random code of size M is at most:

$$P_e \leq M 2^{-n(1-H(0.25))} = 2^{-n(1-H(0.25)-R)}$$

The above probability goes exponentially to zero as $n \to \infty$ for all rates $R < 1 - H(0.25)$ □

We observed that with the probability going to one the forged copy will be produced by a close pair, and due to the marking assumption the pirate with the closer distance, $\mathbf{x}_1$ to $\mathbf{y}$ is approximately $n/4$. The key observation is that, we can treat the "channel" between $\mathbf{y}$ and $\mathbf{x}_1$ with a BSC with crossover probability 1/4.

The above theorem establishes the existence of a *good* fingerprinting code. However, Theorem 2.1 does not provide any information about the *structure* of the code. In order for any code to be encoded and decoded with tolerable complexity, it needs to have certain algebraic structures which will help reduce the complexity of encoding/decoding.

The next theorem establishes the existence of a good *linear* fingerprinting code. Linearity of the code helps reduce its *encoding* complexity from exponential (the complexity of the look up table required for the random codes) to polynomial. Our result is extended to binary linear codes in the following theorem:

Theorem 2.2. *For all rates less than* $1-H(0.25)$*, there exists a* linear *MD achievable fingerprinting code, when* $t = 2$*.*

Proof. Consider the ensemble of binary linear codes with binary parity check matrix H where elements of H are chosen equally and independently from $\{0, 1\}$. Here we only consider linear codes with full rank parity check matrices and discard the exponentially small number of codes which violate this assumption.

The size of matrix H is $l \times n$, with rate $R = 1 - l/n$ and the codeword length n. It should also be noted that in the following all matrix multiplications and additions are done in module-2 unless otherwise is stated. Similar to the proof of Theorem 2.1, we can re-write the probability of misidentification given a forged copy as

$$P_m \leq P_m(\mathbf{y}|\mathbf{x}_1 \overset{C}{\leftrightarrow} \mathbf{x}_2) + P(\mathbf{x}_1 \overset{N}{\leftrightarrow} \mathbf{x}_2). \tag{2.14}$$

In Lemma 2.1 we have established that over the ensemble of linear random codes described above, $P(\mathbf{x}_1 \overset{N}{\leftrightarrow} \mathbf{x}_2)$ also goes to zero as the code length goes to infinity. Now let us consider $P_m(\mathbf{y}|\mathbf{x}_1 \overset{C}{\leftrightarrow} \mathbf{x}_2)$. In order to randomize the codebook, the distributor employs the following strategy:

The secret keys, $\mathbf{k}$'s are independent binary random vectors of length n, whose coordinates are chosen to be $0, 1$ independently with the same probability. After encoding, the secret key is added in the binary domain to the codeword i.e. $\mathbf{x}_1 \oplus \mathbf{k}$, $\mathbf{x}_2 \oplus \mathbf{k}$ and the resulting vectors are assigned to the users.

As we mentioned earlier, the secret key is unknown to the users and is only known to the distributor. Upon receiving $\mathbf{y}$, the decoder removes $\mathbf{k}$, and performs the minimum distance decoding. Similar to the proof of Theorem 2.1, again due to the marking assumption at least one of the fingerprints will have a hamming distance less

than or equal to $h(n) = n(\frac{1}{4} + \frac{\epsilon}{2})$ with one of the fingerprints which without loss of generality will be assumed to be $\mathbf{x}_1$. That is:

$$d_H(\mathbf{y}, \mathbf{x}_1) < h(n) \tag{2.15}$$

To prove the theorem we need the following properties of the considered linear ensemble which is proved in [37]:

- For any binary vector $\mathbf{z}$ the probability of $\mathbf{z}$ being in a code C of the ensemble is equal to 2^{-n}
- For binary vectors $\mathbf{z}$, $\mathbf{z}'$ such that $\mathbf{z} \oplus \mathbf{z}' \neq \mathbf{0}$ then:

$$\Pr(\mathbf{z}, \mathbf{z}') \in C = 2^{-2n},$$

that is the event that $\mathbf{z}$ and $\mathbf{z}'$ be in the same codebook C, are independent.

Now in the proof of Theorem 2.1 we only user *pairwise* independence. Basically, the probability of any $\mathbf{z}$ with the Hamming distance of l from $\mathbf{y}$ being in the linear codebook is still equal to:

$$\frac{\binom{n}{l}}{2^n}$$

Therefore, the probability of misidentification for the random linear ensemble can still be upper bounded by:

$$M2^{-n(1-H(0.25))} = 2^{-n(1-H(0.25)-R)},$$

which goes exponentially to zero as $n \to \infty$ for all rates $R < 1 - H(0.25)$. □

2.4 Belief Propagation for Fingerprinting

The application of an explicit linear code defined over a random graph with Belief Propagation (BP) decoding is studied in this section. As mentioned earlier BP iterative decoding can be employed as an approximation to the exact minimum distance decoding. The complexity of full MD decoding is still exponential whereas BP decoding has linear complexity in the code-length.

2.4.1 Accumulate Repeat Accumulate codes for Fingerprinting

The performance of accumulate repeat accumulate codes with different rates under the 2-pirate memoryless attack is studied by computer simulation. In this attack, when the pirates encounter a detectable position, they choose $0, 1$ independently and with equal probability to form the forged copy.

We use rates 1/8, 1/9 and 1/10 ARA codes based on the low rate protographs presented in [35]. The protographs of the codes are depicted in Fig 2.1. For a formal description on ARA codes, we refer the interested readers to [35,38,39], and references therein.

The decoding is done iteratively using the sum product algorithm with maximum number of iterations equal to 60. The decoder treats the forged fingerprint as the output of a binary symmetric channel with crossover probability equal to 0.25. In Fig 2.2, the probability of misidentification P_m is depicted versus different code lengths for different rates.

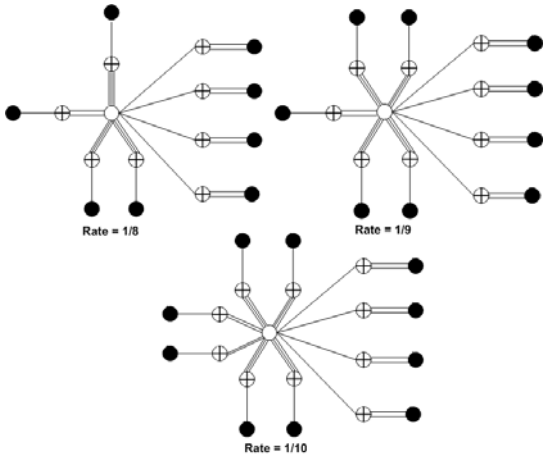

Fig. 2.1: Protographs of rate 1/8, 1/9, 1/10 ARA codes

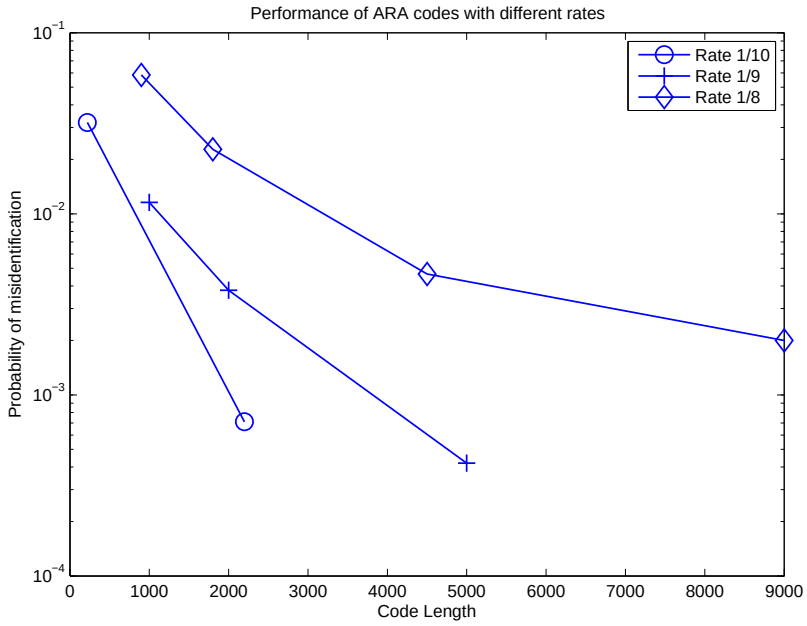

Fig. 2.2: Probability of misidentification for ARA code with different rates under iterative decoding

2.5 Conclusion

In this chapter we studied the code design and performance of the MD decoder for digital fingerprinting. We established that for rates up to $1 - H(0.25) \approx 0.188$, there is a linear fingerprinting code whose probability of misidentification goes to zero exponentially fast in the code length under the minimum distance decoding for any general collusion attack.

It is worth mentioning that the fingerprinting capacity in the considered setting is derived [3] to be 0.25. However, the decoder employed in [3] is far more complex than a belief propagation decoder. We also provided numerical results using an explicit ARA code with iterative decoding to validate our theory. To the best of our knowledge, the best proposed code [32] has a rate approximately equal to 0.0092 for the similar case of $t = 2$ and is decoded with polynomial complexity in the code-length whereas the iterative decoder has linear complexity in n. It should be noted that when the coalition size, t is larger than 2 the minimum distance decoding will fail due to the following argument:

Let $t = 3$, and assume the forged copy is produced by setting

$$y_i = (x_{1i} \oplus x_{2i} \oplus x_{3i}) \tag{2.16}$$

For all detectable positions i. For $t > 3$ the coalition can consider only three of the pirates, ignore the rest and apply the strategy. We will refer to this strategy as the *zero capacity attack*.

It is straightforward to see for the studied random coding ensemble, when the pirates apply the aforementioned simple strategy the mutual information between the forged copy and each of the pirates will be zero, and thus decoding for only one

pirate is impossible. Therefore we will need to perform some type of joint decoding. Joint decoding is also a favorable candidate for closing the gap between 0.188 and the capacity 0.25. We will treat a joint decoder and the zero capacity attack in Chapter 3.

CHAPTER 3

TREE CODES FOR DIGITAL FINGERPRINTING

In this chapter we propose the use of *tree codes* and *sequential decoding* for designing low complexity fingerprinting schemes. We constrict our focus to fair and memoryless coalition attacks both of which will be motivated and explained in the following. The attack can then be characterized by a probability law governing the input alphabets (which belong to the pirates) and the output alphabet (belonging to the forged copy). It should be noted that the concept of fair attacks have also been considered in [3], but for memoryless attacks our definition is more general than the one proposed in [3].

Sequential decoding is a decoding algorithm for tree codes, originally proposed for point to point channels. It can be extended to any code with a tree structure as well. The main issue with sequential decoding is its search effort which is characterized by the number of nodes visited by the decoder also referred to as *complexity*. Using a random coding argument, we show all rates below a cut-off rate are achievable for fair memoryless fingerprinting attacks with finite average complexity. Inspired by this result, we study the performance of random Convolutional codes and a special random tree code for fingerprinting. Our *numerical* results show that for rates as high as 1/6 and vanishing misidentification error is achieved with finite average complexity.

Using a novel non-linear construction based on concatenation of random tree and Convolutional codes rates as high as $1/20 \approx 0.05$ are achievable when $t = 3$ under the zero capacity attack.

3.1 Introduction

In Chapter 2, we studied the applications of minimum distance decoding to digital fingerprinting. In this chapter we look at the problem of from another perspective and study tree search decoding in designing collusion attack-resistant digital fingerprinting schemes.

Our setting is the same Boneh-Shaw's marking assumption introduced in the previous chapter. We studied minimum distance decoding for the problem and proposed a low complexity scheme capable of catching one pirate for a coalition size equal to 2. A coalition attack named the zero capacity attack was identified for larger coalition sizes for which the MD decoder fails.

A main ingredient of our work in this chapter is the concept of *fair* attacks. The intuition behind the assumption of fair attack is that all pirates would want to undergo the same risk of getting caught by the distributor. In order to understand the concept of fair attacks better, consider the following non-fair strategy as an example: In a coalition the colluders ignore one pirate and produce the unauthorized copy based on the $t - 1$ remaining ones. In this case it will be impossible for the distributor to catch the ignored pirate (since it did not have any effect on the forged copy) and the ignored pirate will escape being traced. However, because the attack is mounted by less number of pirates than t it will be easier for the distributer to catch one of the guilty colluders. Then the other members of the coalition will be at a higher risk of

being caught.

Moreover, to facilitate the mathematical analysis we assume that the attack is also *memoryless* which is a common assumption in communication theory.

Our development in this chapter is based on *tree* codes and *tree search decoding*. Tree codes constitute a set of powerful coding schemes. Convolutional codes can be both represented by their *trellis* and *tree* structure. Considering Convolutional codes as a special case of tree codes and its application to forward error correcting have been vastly studied for point to point channels. Arikan in [40] and Balakirsky in [41] have also considered the application of tree codes to the multiple access channel which is a multi user setting.

In this chapter, we propose another explicit fingerprinting scheme based on tree codes. We show the existence of a tree code that satisfies the following properties when the code rate is below a cut-off rate:

- It can be decoded sequentially with *finite* average complexity per decoded symbol.

- Achieves vanishingly small probability of error.

Sequential decoding (SD) which is a procedure for decoding codes on trees was introduced by Wozencraft [42] and further extended and studied in [43–45].

Intuitively SD can be thought of as searching over the code's tree to minimize a cost function in which the search is guided by a *metric*. It was first proposed for decoding of Convolutional codes but it can be applied to any code with tree structure. The Viterbi decoder achieves ML performance for Convolutional codes but its complexity grows exponentially in the constraint length.

Thus, when the constraint length of the code is long, the Viterbi algorithm's complexity becomes prohibitive. The complexity of sequential decoding, however, is almost independent of the code's constraint length. Along with Belief Propagation decoding and Viterbi decoding, sequential decoding constitute the most powerful decoding methods available today.

Unlike Viterbi and BP decoding the complexities of which grow *linearly* in the code-length and thus when the code-length is constant their complexities will always be constant, SD has *variable complexity* of decoding. That is, the complexity of SD is a *random variable* depending on the channel probability distribution, the code, and the information sequence. Studying the complexity issues of various sequential decoding procedures, and problems surrounding them has fueled a considerable amount of research since it was proposed. For an additive noise Gaussian channel, this phenomenon can be intuitively explained as follows:

If the channel noise is high the decoder will need to backtrack too many times to decode an information bit. The sequential decoder succeeds in decoding a block if the code rate is relatively low enough for the channel noise.

Characterizing the complexity is not mathematically tractable for a specific code, but it can be analyzed for an ensemble of codes using random coding arguments. It can be shown that the average complexity (i.e., the expectation of this random variable, where the expectation is also taken over the ensemble) per node is bounded if the code rate is lower than the channel's *computational complexity cut-off rate* [45].

In our work the decoding is carried out by searching for a coalition which satisfies the marking assumption *sequentially* on the Cartesian product of the tree code with its replicas.

Using a random coding argument, the decoder's cut-off rate is characterized for a coalition size of $t = 2$. Average error probability for the random ensemble is also proved to be vanishing for rates less than the cut-off rate. Finally, using an ex-purgation argument we establish the existence of a tree code with vanishing error probability and finite complexity. The derived cutoff rate improves on the achievable rate under the minimum distance decoding.

Numerical results are also provided to support our mathematical analysis. First using random Convolutional codes we obtain vanishing error probability -which can be caused by either complexity overflow or misidentification- at rate $1/6$ when $t = 2$. This corresponds to almost 30% improvement under the same attack considered in Chapter 2.

Second, we propose a novel explicit construction using concatenation of random tree codes and Convolutional codes and obtain vanishing probabilities of misidentification with linear complexity at $1/20$ with $t = 3$ under the zero capacity attack . It should be noted again that the MD decoder fails under this attack. Bidirectional tree search decoding is also studied which provides a trade-off between the decoding complexity and misidentification error probability.

The rest of the chapter is organized as follows. In Section 3.2 we present the system model and mathematical notations. Section 3.3 contains our main results and we present the numerical results in Section 3.4.

3.2 Notations and Problem Statement

The same model of a fingerprinting system under the marking assumption and notations of Chapter 2 is applied in this chapter. As denoted before the probability

law $V(\mathbf{y} \mid \mathbf{x}_1, \cdots, \mathbf{x}_t)$ under which the forged copy $\mathbf{y}$ is created determines the fingerprinting strategy.

A fingerprinting strategy V is *memoryless* if at all coordinates the distribution of the forged copy depends only on the corresponding coordinates of the coalition marks and is independent of the other coordinates.

That is, for a memoryless attack we will have:

$$V(\mathbf{y} \mid \mathbf{x}_1, \cdots, \mathbf{x}_t) = \prod_{i=1}^{n} V(\mathbf{y}_i \mid \mathbf{x}_{1i}, \cdots, \mathbf{x}_{ti}) \tag{3.1}$$

Obviously a strategy satisfies the marking assumption if:

$$V(Y = 1 \mid X_1 = 1, \cdots, X_t = 1) = V(Y = 0 \mid X_1 = 0, \cdots, X_t = 0) = 1 \tag{3.2}$$

The memoryless fingerprinting strategy V is said to be *fair* if:

$$\Pr(Y_i = X_{1i}) = \Pr(Y_i = X_{2i}) = \cdots = \Pr(Y_i = X_{ti}), \qquad 1 \leq i \leq n. \tag{3.3}$$

Heuristically speaking, the fair assumption ensures some uniformity over the forged copy with respect to the pirates.

Denote the set of all fair memoryless strategies i.e., those that satisfy (3.2) ,(3.3) by $\hat{\mathcal{V}}$.

We reiterate the fact that deterministic fingerprinting under the marking assumption is not possible and we use a randomized family of codes. The probability of error for a given coalition U, in the codebook family (E, G) which utilizes the collusion strategy V, is defined as:

$$P_m(U, F, G, V) = \mathbb{E}_K\Big(\sum_{\mathbf{y} \in \mathbf{E(U)}, \mathbf{g_k(y)} \neq \mathbf{U}} V(\mathbf{y} \mid f_K(U)) \Big) \tag{3.4}$$

Similar to Chapter 2, we consider the maximum probability of error but the average is taken over all possible coalitions under the memoryless and fair attacks:

$$\overline{P}_m(E, G) := \frac{1}{\binom{M}{t}} \sum_{U} \max_{V \in \hat{\mathcal{V}}} P_m(U, E, G, V). \tag{3.5}$$

The rest of the section is dedicated to binary tree codes.

Consider a general binary information sequence $\mathbf{u}$. The first i digits of $\mathbf{u}$ will be denoted by $\mathbf{u}(\cdots i)$, i.e. $\mathbf{u}(\cdots i) = [\mathbf{u}(1), \mathbf{u}(2), \cdots, \mathbf{u}(i)]$.
Also, for r-tuples $\mathbf{a} = (a_1, a_2, \cdots, a_r), \mathbf{b} = (b_1, b_2, \cdots, b_r), \cdots, \mathbf{c} = (c_1, c_2, \cdots, c_r)$, we define $\mathbf{a} \times \mathbf{b} \times \cdots \times \mathbf{c}$ as:

$$\Big((a_1, b_1, \cdots, c_1), \cdots, (a_r, b_r, \cdots, c_r)\Big)$$

The encoder is denoted by $\mathbf{e}$ and its function is to generate a binary block of length N with respect to each letter of sequence $\mathbf{u}$. The ith block will be denoted by $\mathbf{eu}(i)$ and the first i blocks by $\mathbf{eu}(\cdots i)$. The code rate will be equal to $\frac{1}{N}$.

A code for which $\mathbf{eu}(i)$ only depends on $\mathbf{u}(\cdots i)$ will be referred to as a *tree code*. We will also refer to $\mathbf{u}(\cdots i)$ as a *node* and the block $\mathbf{eu}(i)$ which connects node $\mathbf{u}(\cdots i-1)$ to $\mathbf{u}(\cdots i)$ as a *branch*. Clearly a tree code can be represented by a tree diagram using the above notions of node and branch.

For a tree code the decoding can be performed *sequentially*. Define the *t-tuple tree* for a tree code $\mathbf{e}$, as the function which assigns $\big(\mathbf{eu}_1 \times \mathbf{eu}_2 \times \cdots \times \mathbf{eu}_t\big)$ to the t-tuple of information sequences $\big(\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_t\big)$. The t-tuple tree is essentially the Cartesian product of the code with its replicas. Similar to a tree's node and branch we will refer to $(\mathbf{u}_1(\cdots i) \times \mathbf{u}_2(\cdots i) \cdots \times \mathbf{u}_t(\cdots i))$ as a node and to $\mathbf{eu}_1(i) \times \mathbf{eu}_2(i) \cdots \times \mathbf{eu}_t(i)$ as a branch.

A node $(\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_t)$ at depth i is called the *right* node of $(\mathbf{u}'_1, \mathbf{u}'_2, \cdots, \mathbf{u}'_t)$ if: $\mathbf{u}_1(\cdots i-1) = \mathbf{u}'_1, \mathbf{u}_2(\cdots i-1) = \mathbf{u}'_2, \cdots, \mathbf{u}_t(\cdots i-1) = \mathbf{u}'_t(\cdots i-1)$. Also $(\mathbf{u}_1(\cdots i-1) \times \mathbf{u}_2(\cdots i-1) \cdots \times \mathbf{u}_t(\cdots i-1))$ is called the *left* node of $(\mathbf{u}_1(\cdots i) \times \mathbf{u}_2(\cdots i) \cdots \times \mathbf{u}_t(\cdots i))$.

For vectors $\mathbf{a}, \mathbf{b} \cdots, \mathbf{y}$ of the same length r, we say $(\mathbf{a} \times \cdots \times \mathbf{b}, \mathbf{y})$ satisfies the

marking assumption, if for all $1 \leq i \leq r$ for which $\mathbf{a}(i) = \cdots = \mathbf{b}(i)$, we have:

$$\mathbf{y}(i) = \mathbf{a}(i)$$

We say a node in the t-tuple tree satisfies the marking assumption if $(\mathbf{eu}_1(\cdots i) \times \mathbf{eu}_2(\cdots i) \cdots \times \mathbf{eu}_t(\cdots i), \mathbf{y}(\cdots iN))$ does.

3.3 Main Results

Having rigorously defined the basic concepts and the tree codes, we are prepared to study the proposed decoding scheme. The idea is to search over the t-tuple tree for a combination that satisfies the marking assumption by hypothesizing different nodes on it. Each step of the search consists of updating a node on the t-tuple tree that satisfies the marking assumption with the forged copy and will be referred to as the *active node*. In each step we move from one node to a neighboring one. Only two different types of moves are allowed in the tree search:

1. Forward move : for which we replace the active node with one of its right nodes that satisfies the marking assumption. There could be more than one right node with that property, in which case we do the replacement according to some arbitrary ordering. If it is the first time that a node is hypothesized as the active node, we choose the first right node. Otherwise we choose the next right node that has not been hypothesized yet.

If we can not make any forward move i.e., all of the right nodes that satisfy the marking assumption have already been visited, the next move will be a backward move (explained below).

A backward move means that all the descending nodes from the active node on

the tree contradict the marking assumption and therefore the node will be discarded from further consideration. Also note that by this procedure every node will be the destination of a forward move at most once. Decoding is finished when we reach the end of the tree.

2. Backward move : for which we replace the active node with its left node. A backward move is only made when it is not possible to make a forward move. After making a backward move, if it is possible to make a forward move on the new active node the next move will be a forward move otherwise the next move will again be a backward move. If we are at the origin and the next move is a backward move we declare failure.

The number of nodes visited on the t-tuple tree to decode for the forged fingerprint is a random variable depending on the code, the coalition and the forged copy. Let the *correct path* be the path corresponding to the fingerprints belonging to the coalition. It is helpful to look at the tree as the correct path and the rest which are incorrect paths originated from a node in the correct path.

Define W_n as the number of *forward move* on the incorrect tree originated from the n-th node on the correct path. To study the complexity we only count the number of forward moves made on the incorrect paths. The reason for only considering forward moves on the incorrect paths are:

- For each backward move on a node, there is a forward move to its descendant node from which the backward move were made. Thus, the total number of backward moves at depth $l-1$ is at most equal to the number of forward moves at depth l and if the average number of forward moves are bounded the total number of backward moves will also be bounded.

- Because all of the nodes on the correct path satisfy the marking assumption, no backward move will be made on the correct path.

  Thus, all the moves on the correct path are forward and each node is visited at most once on the correct path. Therefore the number of moves on the correct path is always bounded and to bound the average complexity it suffices to only consider the forward moves on the incorrect paths.

The average complexity per node as is defined as:

$$C = \lim_{n\to\infty} \frac{1}{n} \sum_{n=0}^{\infty} W_n \tag{3.6}$$

Analyzing the distribution of the above random variable for a particular code is intractable in general. Instead we adopt the common approach in the sequential decoding literature and treat it for an ensemble of codes. The framework is similar to that of [45] which was later extended to multiple access channels in [40].

Our goal is to find an upper bound on $\bar{C}$, the average complexity over the coalition, the forged copy and the ensemble and show that if the code rate is less than a threshold it is a finite constant independent of the code length. If this property is satisfied for the ensemble then there should exist at least one good code with finite complexity of decoding.

To this aim, we first argue that $\bar{W}_0$ i.e., the average of number of incorrect forward moves on the wrong subtree originated from the origin over the random ensemble is finite. Next we note that since the strategy is memoryless and all codewords are chosen independently the statistical description of $W_0, W_1, ..., W_n, ...$ will be the same and therefore $\bar{C}$ will be finite.

Theorem 3.1. *Under fair and memoryless fingerprinting attacks, all rates below* 0.2 *are achievable under the decoder explained above, when* $t = 2$.

Proof. The proof consists of three parts: First we establish that the average complexity of tree search over the random ensemble on the 2-tuple tree is finite for all rates less than 0.2. Next we establish the error probability is vanishing and finally by employing an expurgation argument we show the existence of a code satisfying both properties.

Assign a binary information sequence of length $\log_2(M)$ to all users in the system. Consider the following ensemble of tree codes: For any information sequence $\mathbf{u}$, $\mathbf{eu}(i)$ is a random binary sequence of length N where $0, 1$ are chosen independently and randomly with distribution $B(0.5)$, where $B(p)$ denotes the Bernoulli distribution with argument p. It should be noted that that for this random ensemble of tree codes if two different information sequences are the same up to depth h and different after that i.e., $u_1(\cdots h) = u_2(\cdots h)$ and $u_1(h+1) \neq u_2(h+1)$ their corresponding codewords will be the same for the first h blocks and and statistically independent beyond it.

Let $k(l)$ be the k-th node at depth l according to the same ordering with which the 2-tuple tree is searched. Define the binary random variable $w[k(l)]$ as:

$$w[k(l)] = \begin{cases} 1 & \text{if a forward move arrives} \\ & \text{in the } k\text{-th node of depth } l \\ 0 & \text{otherwise} \end{cases}$$

As we mentioned earlier every node can be at most once the destination of one forward move so the total number of forward moves on the incorrect tree can be upper bounded by:

$$W_0 \leq \sum_{l=0}^{\infty} \sum_{k(l)} w[k(l)] \tag{3.7}$$

We are interested in the expectation of W_0 over the random ensemble. Because the expectation of the sum is equal to the sum of expectation we have:

$$\overline{W_0} \leq \sum_{l=0}^{\infty} \sum_{k(l)} \overline{w[k(l)]} \tag{3.8}$$

The k-th node of depth l in the wrong tree can be destination of a forward move only if it satisfies the marking assumption with $\mathbf{y}(\cdots lN)$.

There are two kinds of branches on the wrong 2-tuple tree:

- Type 0: for which none of the components are equal to either of the original fingerprints.
- Type 1: for which one the components is equal to one of the of the original fingerprints.

To distinguish those two we re-write (3.8) with indicator functions ϕ_0, ϕ_1 corresponding to errors of type 0,1 respectively as:

$$\overline{W_0} \leq \sum_{l=0}^{\infty} \sum_{m(l)} \overline{\phi_0[m(l)]} + \sum_{l=0}^{\infty} \sum_{n(l)} \overline{\phi_1[n(l)]} \tag{3.9}$$

Let us study $\overline{\phi_1[n(l)]}$ first:

The wrong branch consists of an original fingerprint to depth l, $\mathbf{x}_1(\cdots lN)$ and a wrong codeword $\mathbf{z}_1(\cdots lN)$ of the same length. Assume they are equal for the first h blocks where $0 \leq h \leq l$. As mentioned earlier, $\mathbf{x}_1(\cdots lN)$, $\mathbf{z}_1(\cdots lN)$ will be the same for the first h blocks and independent in the next $l - h$ ones.

$\overline{\phi_1[n(l)]}$ can then be broken into:

$$\overline{\phi_1[n(l)]} = \overline{\phi_1[\cdots hN]} \times \overline{\phi_1[hN + 1 \cdots lN]}$$

Over the random ensemble $\overline{\phi_1[\cdots hN]}$ is equal to the probability that $(\mathbf{x}_1(\cdots hN) \times \mathbf{x}_1(\cdots hN), \mathbf{y}(\cdots hN))$ satisfies the marking assumption.

The possibility of this event to hold is when $\mathbf{x}_1(\cdots hN) = \mathbf{y}(\cdots hN))$. In Lemma 3.1 we will show that for equiprobable inputs and fair memoryless attack the following holds:

$$\Pr(X_1 = 0, Y = 0) = \frac{(1+w)}{4}$$
$$\Pr(X_1 = 1, Y = 1) = \frac{(2-w)}{4},$$

for some w with $0 \leq w \leq 1$ and thus $\Pr(X_1 = Y) = \frac{(1+w)}{4} + \frac{(2-w)}{4} = \frac{3}{4}$.

Then, due to memoryless property of the attack we have:

$$\Pr\left(\mathbf{x}_1(\cdots hN) = \mathbf{y}(\cdots hN)\right) = \left(\frac{3}{4}\right)^{hN} \tag{3.10}$$

For the second part, $\overline{\phi_1[hN+1\cdots lN]}$ is equal to the probability that in the random ensemble

$$\Big(\mathbf{x}_1(hN+1\cdots lN) \times \mathbf{z}_1(hN+1\cdots lN), \mathbf{y}(hN+1\cdots lN)\Big)$$

satisfies the marking assumption which as will be proved in Lemma 3.1 is equal to $(\frac{7}{8})^{(l-h)N}$.

Thus,

$$\overline{\phi_1[n(l)]} = (\frac{3}{4})^{(hN)} \times (\frac{7}{8})^{(l-h)N} < (\frac{7}{8})^{lN} \tag{3.11}$$

Because at depth l, the total number of branches with a component equal to one of the pirates is at most 2^l, $\sum_{l=0}^{\infty} \sum_{n(l)} \overline{\phi_1[n(l)]}$ can be upper bounded by:

$$\sum_{l=0}^{\infty} 2^l (7/8)^{lN} \tag{3.12}$$

which converges if $2(7/8)^N < 1$, that is $2^R(7/8) < 1$ which holds for:

$$R < \log_2(\frac{8}{7}) \approx 0.2 \tag{3.13}$$

Let us consider $\overline{\phi_0[m(l)]}$ now. This is the probability that two codewords $\mathbf{z}_1(\cdots lN)$, $\mathbf{z}_2(\cdots lN)$ different from the original fingerprints and belonging to the pirates, satisfy the marking assumption with the forged copy. If those are equal to depth $\acute{h}$ and different after it, we can similarly break down $\overline{\phi_0[m(l)]}$ to

$$\overline{\phi_0[\cdots \acute{h}N]} \times \overline{\phi_0[\acute{h}N+1 \cdots lN]}$$

For all the positions $i, 1 \leq i \leq \acute{h}N$, out of two options of being 0 or 1 $\mathbf{z}_1(i)$ can only be equal to $\mathbf{y}(i)$ for the marking assumption to be satisfied. Thus, $\overline{\phi_0[\cdots \acute{h}N]}$ can be upper-bounded by $(\frac{1}{2})^{(\acute{h}N)}$.

For positions $i, \acute{h}N+1 \leq i \leq lN$, out of the four different combinations $\mathbf{z}_1(i), \mathbf{z}_2(i)$ can be all the 3 options except for being simultaneously equal to $\mathbf{y}(i) \oplus 1$ and we have:

$$\overline{\phi_0[\acute{h}N+1 \cdots lN]} \leq (\frac{3}{4})^{(l-\acute{h})N}$$

Thus, $\overline{\phi_0[m(l)]} \leq (\frac{3}{4})^{lN}$.

Noting that the total number of branches of the wrong tree for which none of the components are equal to the original fingerprints at depth l is at most 2^{2l}, we arrive at:

$$\sum_{l=0}^{\infty} \sum_{m(l)} \overline{\phi_0[m(l)]} \leq \sum_{l=0}^{\infty} 2^{2l}(3/4)^{lN} \tag{3.14}$$

which converges for $R < 0.2$.

As mentioned earlier the statistical description for all W_n's are the same as that of W_0, therefore the average complexity of decoding per node over the random ensemble is bounded for rates less than 0.2.

Now let us consider the average error probability: Decoding error occurs if one of the paths on the wrong trees reaches the end. Let the probability of the wrong tree from node i reaches the end of tree be P_i, the average error probability can be upper bounded by:

$$\overline{P_e} \leq \overline{P_0} + \overline{P_1} + \cdots + \overline{P_i} + \cdots$$

In the above proof for complexity, the probabilities of being on the wrong paths over the random ensemble of tree codes were upper bounded by:

$$\overline{P_i} \doteq 2^{2R(n-i)}(3/4)^{n-i} + 2^{R(n-i)}(7/8)^{n-i}$$

which goes exponentially to zero as n goes to infinity for $R < 0.2$. And finally consider the following expurgation argument to complete the proof: Let L be the number of codes in the ensemble of binary tree codes explained above. Since we are assigning $0, 1$ with the same probability of $1/2$ each of the codes will have the same equal probability of $1/L$. Let us sort the codes in the descending order of error probability. Since the probability of error is positive, none of the codes in the upper half - denoted by family Q- can have an error probability larger than the average error probability in the whole ensemble. Thus, the error probability for *all* the codes in this family is vanishing in n. Now, we are going to prove the average complexity over this family is also finite.

Assume the ensemble to be split in two sets with the same number of codes in each one. Showing that the average complexity in either of the sets cannot exceed $2W$ will complete the proof. Let us sort the codes in the ensemble in ascending order of complexity. Denote the family of the first half (with larger complexity) by A and the other half by B. Obviously the average complexity of Q cannot be larger than A.

Now, the average complexity in A can be written as:

$$\frac{2}{L}(W_{A1} + W_{A2} + ... + W_{A(L/2)}) \leq$$
$$\frac{2}{L}\{(W_{A1} + W_{B1}) + (W_{A2} + W_{B2}) + ... + (W_{A(L/2)} + W_{B(L/2)})\} = 2W$$

where W is the average complexity over the ensemble and was proved to be finite. Thus, the average complexity in the family Q is bounded by $2W$ which is still finite and therefore there must be at least one code in the family with a complexity less than $2W$. Because all codes in Q have vanishingly small error probabilities we established the existence of a code with vanishing error probability that can be decoded with finite complexity. □

Lemma 3.1. *For randomly chosen codewords of length m under a fair memoryless fingerprinting attack, the average error probabilities of type 0 and 1 can be upper bounded by $(3/4)^m$ and $(7/8)^m$ respectively.*

Proof. Let $\mathbf{y}$ be the forged fingerprint and $\mathbf{x}_1$ and $\mathbf{x}_2$ be random codewords with length m the coordinates of which are chosen i.i.d with distribution $B(0.5)$ that correspond to the pirates. The considered error of type 0 happens when two random codewords $\mathbf{z}_1$ and $\mathbf{z}_2$ constructed in the same way as $\mathbf{x}_1$ and $\mathbf{x}_2$ but different from them satisfy the marking assumption with $\mathbf{y}$.

At each position $1 \leq i \leq m$, for the incorrect codewords $\mathbf{z}_1(i)$ and $\mathbf{z}_2(i)$ can be any of the 4 possible combinations except for being simultaneously equal to $\mathbf{y}_1(i) \oplus 1$ and therefor there are 3 choices. Thus, the total probability in the random ensemble can be upper bounded by:

$$(\frac{3}{4})^m \tag{3.15}$$

Now let us focus on type 1 errors that are more complicated to analyze. Without loss of generality we can assume the codeword $\mathbf{x}_1$ and another codeword $\mathbf{z}_1$ satisfy the marking assumption with $\mathbf{y}_1$. We can write the probability of error averaged over the random ensemble as:

$$\sum_{(\mathbf{x}_1,\mathbf{x}_2,\mathbf{z}_1)} \Pr(\mathbf{x}_1,\mathbf{x}_2,\mathbf{z}_1)\Pr(err \mid (\mathbf{x}_1,\mathbf{x}_2)) = \tag{3.16}$$
$$\sum_{(\mathbf{x}_1,\mathbf{x}_2,\mathbf{z}_1)} \Pr(\mathbf{x}_1,\mathbf{x}_2,\mathbf{z}_1)\sum_{\mathbf{y}} V^m\Big(\mathbf{y} \mid \mathbf{x}_1,\mathbf{x}_2\Big)\Phi_M(\mathbf{y},\mathbf{x}_1,\mathbf{z}_1),$$

where $\Phi_M(\mathbf{y},\mathbf{x}_1,\mathbf{z}_1)$, is 1 when its arguments satisfy the marking assumption and is 0 otherwise and V^m denotes m use of the fair, memoryless fingerprinting "channel".

Noting that the codewords are chosen independently, equation (3.16) can be written as:

$$\sum_{\mathbf{y}}\sum_{\mathbf{x}_1,\mathbf{x}_2} V^n(\mathbf{y} \mid \mathbf{x}_1,\mathbf{x}_2)\Pr(\mathbf{x}_1)\Pr(\mathbf{x}_2)\sum_{\mathbf{z}_1}\Pr(\mathbf{z}_1)\Phi_M(\mathbf{y},\mathbf{x}_1,\mathbf{z}_1) \tag{3.17}$$

In all the positions that $\mathbf{y}$, and $\mathbf{x}_1$ are the same, $\mathbf{z}_1$ can be either 0 or 1. But in the positions where $\mathbf{y}$, and $\mathbf{x}_1$ disagree for the marking assumption to be satisfied $\mathbf{z}_1(i)$ has only one choice which is equal to $\mathbf{y}(i)$. Thus, equation (3.17) can be written as:

$$\sum_{\mathbf{y},\mathbf{x}_1} V^n(\mathbf{y},\mathbf{x}_1)2^{-d_H(\mathbf{y},\mathbf{x}_1)} \tag{3.18}$$

Because the fingerprinting strategy is assumed to be memoryless, equation (3.17) can be written as:

$$\prod_{i=1}^{n}\sum_{\mathbf{y}_i,\mathbf{x}_{1i}} V(\mathbf{y}_i,\mathbf{x}_{1i})2^{-d_H(\mathbf{y}_i,\mathbf{x}_{1i})}. \tag{3.19}$$

It is straightforward to verify that under equi-probable input alphabets the memoryless attack is fair if:

$$V(Y=0 \mid X_1=0,X_2=1) = V(Y=0 \mid X_1=1,X_2=0)$$

Let us assume $V(Y = 0 \mid X_1 = 0, X_2 = 1) = w$. Under equiprobable inputs with which the random coding ensemble is constructed we will have:

$$V(Y = 0, X_1 = 0) =$$
$$\frac{1}{4}\Big(V(Y = 0 \mid X_1 = 0, X_2 = 0) + V(Y = 0 \mid X_1 = 0, X_2 = 1)\Big) = \frac{(1 + w)}{4}.$$

Similarly we can calculate:

$$V(Y = 0, X_1 = 1) = \frac{w}{4}$$
$$V(Y = 1, X_1 = 0) = \frac{(1 - w)}{4}$$
$$V(Y = 1, X_1 = 1) = \frac{(2 - w)}{4}$$

Then, we arrive at:

$$\sum_{\mathbf{y}_i, \mathbf{x}_{1i}} V(\mathbf{y}_i, \mathbf{x}_{1i}) 2^{-d_H(\mathbf{y}_i, \mathbf{x}_{1i})} = \frac{1 + w}{4} + \frac{(2 - w)}{4} + \frac{(1 - w)}{8} + \frac{w}{8} = \frac{7}{8}, \tag{3.20}$$

and equation (3.18) can be written as:

$$\Big(\frac{7}{8}\Big)^m \tag{3.21}$$

Which completes the proof of the lemma. □

3.4 Numerical results

3.4.1 The coalition size $t = 2$

the following numerical experiments were performed to validate our theoretical results. For $t = 2$ we used a binary Convolutional code with a random encoder with the same constraint length of 30 for the two different rates of 1/6,1/8 to perform the simulation. The encoding matrix is composed of binary numbers picked independently and randomly with distribution $B(0.5)$. The collusion attack is the same explained

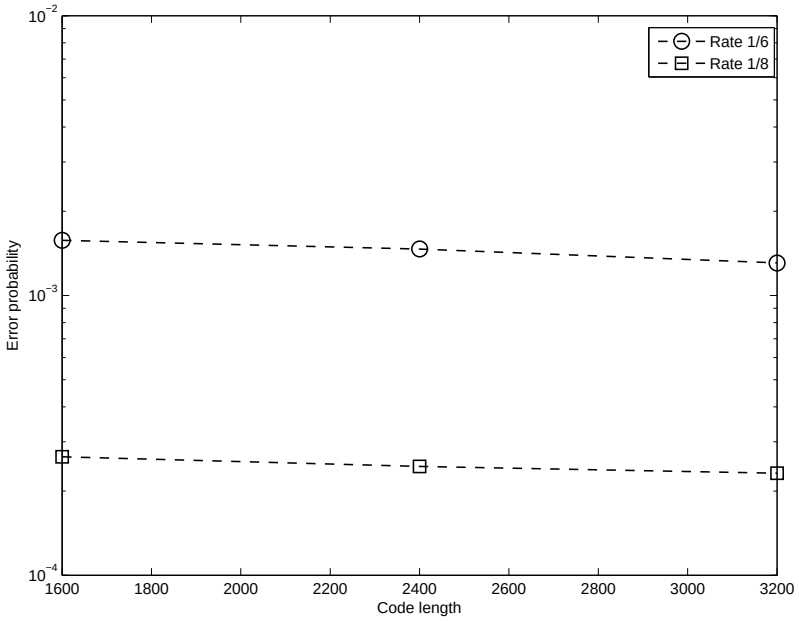

Fig. 3.1: Probability of decoding failure for $t = 2$

in Section 2.4.1.

We declare decoding failure if either the decoded t-tuple is not the same as the original coalition or the number of visited nodes exceeds a threshold. The latter event will be referred to as *complexity overflow*. It should be noted that to make a fair comparison for different code lengths, the threshold is scaled such that the ratio of threshold over the length of information sequence is kept constant. The results are depicted in Figs 3.1 and 3.2.

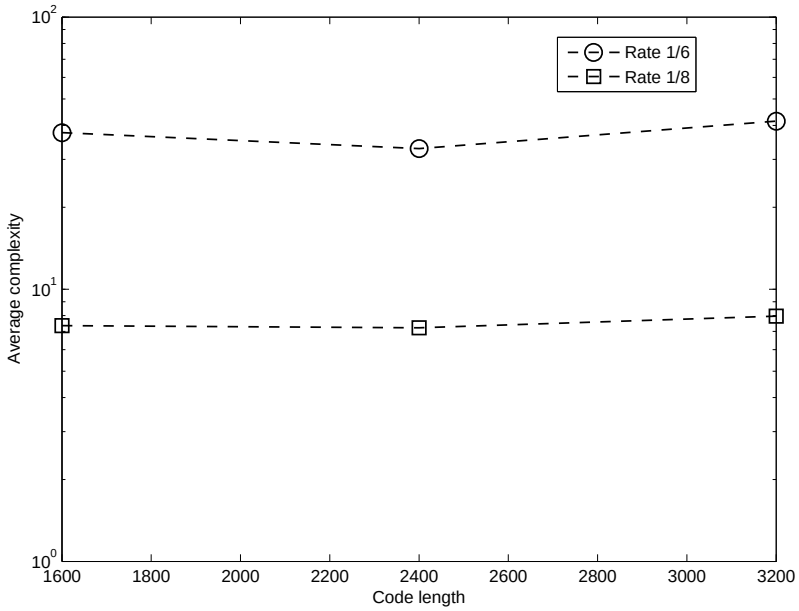

Fig. 3.2: Average decoding complexity per node for $t = 2$

3.4.2 The zero capacity attack with coalition size $t = 3$

The proof presented for the main theorem does not extend to a coalition size more than 2. However, we tried the proposed decoder under the zero capacity. This heuristic approach only relies on numerical results.

Consider three fingerprints $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ and the attack where the forged copy $\mathbf{y}$ is produced by setting

$$\mathbf{y} = \mathbf{x}_1 \oplus \mathbf{x}_2 \oplus \mathbf{x}_3.$$

It is straightforward to verify that the zero capacity attack is fair, memoryless and satisfies the marking assumption. As it was explained in Chapter 2, the minimum distance decoder will fail for this type of coalition attack. In addition, for a linear code, the forged copy $\mathbf{y}$ will be in the codebook and for any fingerprint $\mathbf{x}_i$ and the

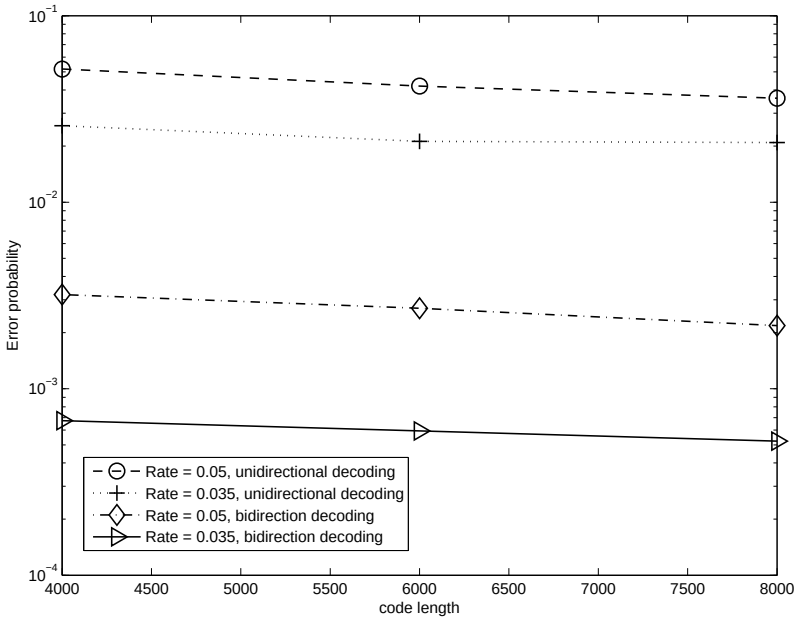

Fig. 3.3: Error probability under zero capacity attack, $t = 3$

triplet composed of $(\mathbf{x}_1, \mathbf{x}_i, \mathbf{y} \oplus \mathbf{x}_1 \oplus \mathbf{x}_i)$ all of which in the codebook will satisfy the marking assumption with $\mathbf{y}$ and it is not possible to user linear codes.

To overcome this problem, we devised a construction of non linear tree codes using the concatenation of an inner random Convolutional code with a long constraint length with an outer binary random tree code (which is non-linear). The encoding for the random tree code is performed by a look up table the constraint length of which in our experiment is 17. The results in terms of average complexity per node and decoding failure probability for rates 1/20, 1/28 is depicted in Figs. 3.3, 3.4.

Bidirectional search: As it can be observed from Fig. 3.3 the proposed scheme suffers from an error floor. To reduce the error floor of our proposed scheme, we adopt bidirectional search i.e., the tree is searched from both its end and beginning.

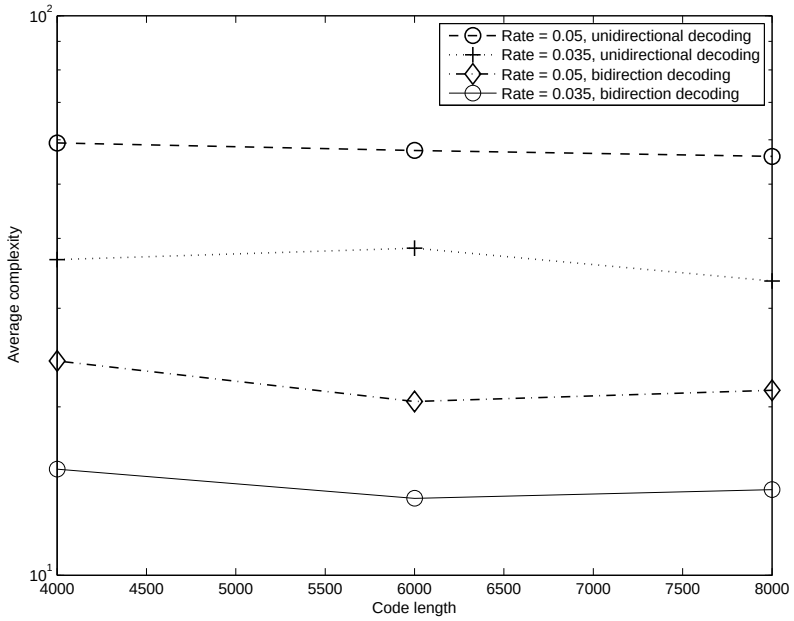

Fig. 3.4: Average decoding complexity per node under zero capacity attack, $t = 3$

Once we reach the other end the decoding is terminated. For example if decoding starts from the end of tree, we stop once we reach its beginning.

Most of the complexity overflows in our experiment are due to the following catastrophic pattern:

Let us denote the information sequences by $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ if two of those vectors (without loss of generality $\mathbf{u}_1$,$\mathbf{u}_2$) agree in the first p bits the corresponding codewords will also agree in the first p blocks i.e., the first pN bits.

In that case it is straightforward to verify that all the nodes composed of an arbitrary binary vector of length pN and the codewords of $\mathbf{u}_1(...p), \mathbf{u}_2(...p)$ will satisfy the marking assumption. Even for a small p this event will be catastrophic as the decoder will have to visit 2^p i.e., an exponential number in p of nodes before finally

finding the correct path.

Using the bidirectional search the probability of complexity overflow due to this catastrophic bit pattern will be almost squared compared with the forward search alone.

CHAPTER 4

AN ACHIEVABLE RATE REGION FOR X CHANNEL

This chapter is the first part in the second direction of the thesis where fundamental limits of a multi-user networks are studied. The considered channel model in this chapter is the discrete memoryless X channel, the most general 2 by 2 channel model in which for the two transmitters and two receivers every transmitter has a message for every receiver. An achievable scheme based on message splitting and binning codebooks is proposed. The achievable rate region under joint decoding is derived and it is established that the region contains the best known rate region for the special cases of interference channel, broadcast channel and multiple-access channel. To the best of our knowledge, the proposed scheme is the best inner bound on the capacity region of the X channel available in the literature.

4.1 Introduction

The X channel refers to a communication scenario in which each transmitter has a message for every receiver. This model involves most of the multi-user channels studied in information theory; such as multiple access channel, broadcast channel and interference channel.

The degrees of freedom region for the MIMO X channel is studied in [10, 21, 46]. It is established that for the MIMO X channel with $M > 1$ number of antennas at all nodes and with non-degenerate channel matrices, the degrees of freedom is equal to $\frac{4}{3}M$. The MIMO X channel is the first known example that has non-integer degrees of freedom and has received much attention lately. In this chapter, we propose a signaling scheme for the X channel using message splitting and binning. The proposed scheme uses the message splitting technique to split messages in two parts (common and private) and utilizes the common messages in the construction of the cloud centers ([9, 11, 47]), which are used to design superposition codes. A binning technique is used for the private messages at the transmitters to allow coding in the sense of [16] (see also [48]). In the special case of broadcast channel, our proposed rate region reduces to the best known achievable region for the two user broadcast channel [11], in the case of interference channel it reduces to that of [12] (a simpler description of this region is recently given in [49], see also [50]) and in the case of multiple access channel it achieves the capacity region (see, e.g,. [8]).

The proposed region outperforms that of [10], which to the best of our knowledge was the best proposed achievable region for the X channel before our contribution. For example, in the special case of degraded broadcast channel the proposed scheme is capacity achieving whereas [10] is not. In addition, in the case of interference channel [10] reduces to the scheme of considering interference as noise, whereas the proposed region allows for interference cancellation and achieves the Han and Kobahayashi region ([12]).

The remaining sections of this chapter are organized as follows. In Section 4.2, we provide the system model. Section 4.3 is devoted to the main result of the paper and some special cases are discussed in Section 4.4.

4.2 System Model

We consider a two-user discrete memoryless X channel (XC), composed of two transmitter-receiver pairs (see Fig. 4.1), and is denoted by

$$(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1, y_2|x_1, x_2), \mathcal{Y}_1 \times \mathcal{Y}_2),$$

for some finite sets $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}_1, \mathcal{Y}_2$. Here the symbols $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$ are the channel inputs and the symbols $(y_1, y_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2$ are the channel outputs observed at the decoder 1 and decoder 2, respectively. The channel is memoryless and time-invariant:

$$p(y_1(t), y_2(t)|\mathbf{x}_1^t, \mathbf{x}_2^t, \mathbf{y}_1^{t-1}, \mathbf{y}_2^{t-1}) = p(y_1(t), y_2(t)|x_1(t), x_2(t)).$$

We assume that each transmitter $k \in \{1, 2\}$ has messages W_{k1} and W_{k2} which is to be transmitted to receiver 1 and receiver 2, respectively, in n channel uses. In this setting, we define $(n, M_{11}, M_{12}, M_{21}, M_{22}, P_{e,1}^{(n)}, P_{e,2}^{(n)})$ codebook with the following components:

- The message sets $\mathcal{W}_{k1} = \{1, ..., M_{k1}\}$ and $\mathcal{W}_{k2} = \{1, ..., M_{k2}\}$ for transmitter $k = 1, 2$.

- An encoding function $f_k(.)$ at transmitter k which maps the messages to the transmitted symbols, $f_k : (w_{k1}, w_{k2}) \to \mathbf{X}_k$ for each $(w_{k1}, w_{k2}) \in \mathcal{W}_{k1} \times \mathcal{W}_{k2}$ for $k = 1, 2$.

- Decoding function $\phi_k(.)$ at receiver k which maps the received symbols to an estimate of the message: $\phi_k(\mathbf{Y}_k) = (\hat{w}_{1k}, \hat{w}_{2k})$ for $k = 1, 2$.

- Reliability of the transmission for receiver k is measured by $P_{e,k}^{(n)}$, where

$$lP_{e,k}^{(n)} = \frac{1}{M_{1k}M_{2k}} \sum_{(w_{1k},w_{2k})\in\mathcal{W}_{1k}\times\mathcal{W}_{2k}} Pr\big\{\phi_k(\mathbf{Y}_k) \neq (w_{1k}, w_{2k}) \\ |(w_{11}, w_{12}, w_{21}, w_{22}) \text{ is sent}\big\},$$

  for $k = 1, 2$.

The rate tuple $(R_{11}, R_{12}, R_{21}, R_{22})$ is said to be achievable for the X channel, if, for any given $\epsilon > 0$, there exists an $(n, M_{11}, M_{12}, M_{21}, M_{22}, P_{e,1}^{(n)}, P_{e,2}^{(n)})$ codebook such that,

$$\begin{aligned} \frac{1}{n}\log(M_{11}) &= R_{11}, \\ \frac{1}{n}\log(M_{12}) &= R_{12}, \\ \frac{1}{n}\log(M_{21}) &= R_{21}, \\ \frac{1}{n}\log(M_{22}) &= R_{22}, \\ \max\{P_{e,1}^{(n)}, P_{e,2}^{(n)}\} &\leq \epsilon, \end{aligned}$$

for sufficiently large n. The capacity region is the closure of the set of all achievable rate pairs (R_1, R_2) and is denoted by $\mathbb{C}^{\text{XC}}$.

4.3 Main Result

The proposed scheme is based on message splitting and binning. First we split the message of transmitter i to receiver j, W_{ij}, into the following: 1) A common message W_{ijc}, and 2) A private message W_{ijp}. The receivers are required to decode all common messages of the transmitters and the intended private messages.

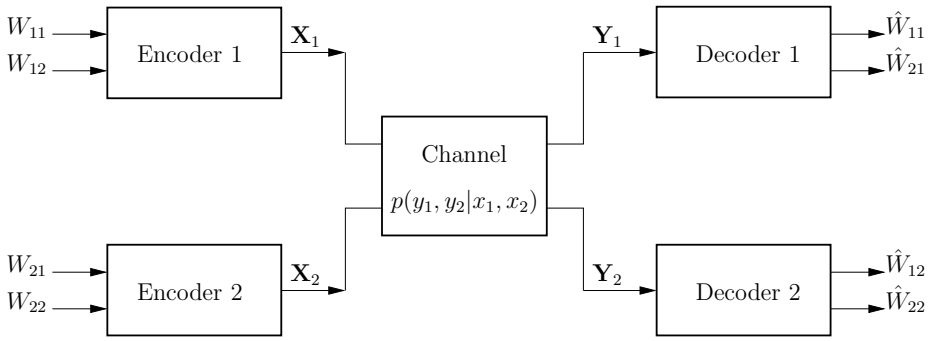

Fig. 4.1: The two user discrete memoryless X channel.

The encoding procedure is explained in the following: The common messages are used for superposition coding, and the codeword serves as "cloud centers" [9] (see also [8, 47]) for the rest of random variables, however, a general binning codebook is considered (see, e.g., [11]). The common message allows for partial interference cancellation in the sense of [12] as we require joint decoding at the transmitters. Finally, we use the binning technique of [48] to jointly encode the private messages. This allows to design the private messages, part of which can be considered as non-causally known interference in the sense of [16]. The general inner bound to the capacity region of an "X" channel is given below.

Theorem 4.1. *Let $\mathcal{P}$ be the set of probability distributions $p(.)$ that factor as*

$$\begin{aligned} & lp(q, v_{1c}, v_{11p}, v_{12p}, v_{2c}, v_{21p}, v_{22p}, x_1, x_2) \\ &= p(q)p(v_{1c}, v_{11p}, v_{12p}|q)p(v_{2c}, v_{21p}, v_{22p}|q) \\ & p(x_1|v_{1c}, v_{11p}, v_{12p}, q)p(x_2|v_{2c}, v_{21p}, v_{22p}, q). \end{aligned} \tag{4.1}$$

For any $p \in \mathcal{P}$, $\mathcal{R}_I(p)$ is the set of non-negative rate tuples $(R_{11c}, R_{11p},\ R_{12c}, R_{12p},$ $R_{21c}, R_{21p},\ R_{22c}, R_{22p})$ satisfying

$$
\begin{aligned}
R_{1c} + R_{11p^*} &< I(V_{1c}, V_{11p}; Y_1 | V_{2c}, V_{21p}, Q) \\
R_{11p^*} &< I(V_{11p}; Y_1 | V_{1c}, V_{2c}, V_{21p}, Q) \\
R_{21p^*} &< I(V_{21p}; Y_1 | V_{1c}, V_{2c}, V_{11p}, Q) \\
R_{2c} + R_{21p^*} &< I(V_{2c}, V_{21p}; Y_1 | V_{1c}, V_{11p}, Q) \\
R_{1c} + R_{11p^*} + R_{21p^*} &< I(V_{1c}, V_{11p}, V_{21p}; Y_1 | V_{2c}, Q) \\
R_{11p^*} + R_{21p^*} &< I(V_{11p}, V_{21p}; Y_1 | V_{1c}, V_{2c}, Q) \\
R_{1c} + R_{2c} + R_{11p^*} + R_{21p^*} &< I(V_{1c}, V_{2c}, V_{11p}, V_{21p}; Y_1 | Q) \\
R_{2c} + R_{11p^*} + R_{21p^*} &< I(V_{2c}, V_{11p}, V_{21p}; Y_1 | V_{1c}, Q) \\
R_{11p} + R_{12p} &< R_{11p^*} + R_{12p^*} - I(V_{11p}; V_{12p} | V_{1c}, Q) \\
R_{21p} + R_{22p} &< R_{21p^*} + R_{22p^*} - I(V_{21p}; V_{22p} | V_{2c}, Q) \\
R_{1c} &= R_{11c} + R_{12c} \\
R_{2c} &= R_{21c} + R_{22c},
\end{aligned}
$$

and

$$
\begin{aligned}
R_{11p} \leq R_{11p^*} \quad &, \quad R_{12p} \leq R_{12p^*}, \\
R_{21p} \leq R_{21p^*} \quad &, \quad R_{22p} \leq R_{22p^*}.
\end{aligned} \tag{4.2}
$$

Similarly we define $\mathcal{R}_{II}(p)$*, which is the set of non-negative tuples* $(R_{11c}, R_{11p}, R_{12c}, R_{12p}, R_{21c}, R_{21p}, R_{22c}, R_{22p})$ *satisfying equations (4.2) with the indices swapped everywhere. For a set* $\mathcal{S}$ *of tuples* $(R_{11c}, R_{11p}, R_{12c}, R_{12p}, R_{21c}, R_{21p}, R_{22c}, R_{22p})$*, we define* $\prod(\mathcal{S})$ *as the set of tuples* $(R_{11}, R_{12}, R_{21}, R_{22})$ *such that* $R_{11} = R_{11c} + R_{11p}$*,* $R_{12} = R_{12c} + R_{12p}$*,* $R_{21} = R_{21c} + R_{21p}$*, and* $R_{22} = R_{22c} + R_{22p}$*.*

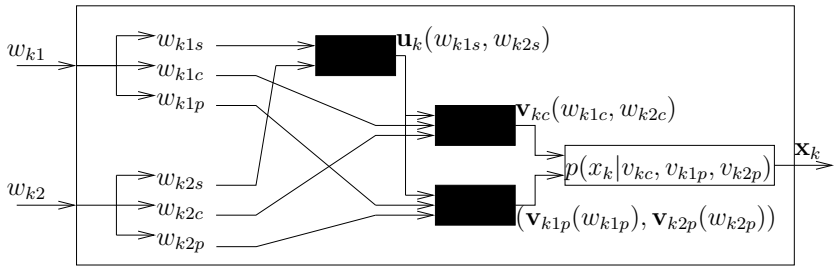

Fig. 4.2: The proposed encoder structure for transmitter k.

The set

$$\mathcal{R} = \prod \left(\bigcup_{p \in \mathcal{P}} \mathcal{R}_I(p) \cap \mathcal{R}_{II}(p) \right) \tag{4.3}$$

is an achievable region for the discrete memoryless X channel.

Proof. First we fix $p(q)$, $p(v_{1c}, v_{11p}, v_{12p}|q)$, $p(v_{2c}, v_{21p}, v_{22p}|q)$, $p(x_1|v_{1c}, v_{11p}, v_{12p}, q)$, $p(x_2|v_{2c}, v_{21p}, v_{22p}, q)$, and the channel is given by $p(y_1, y_2|x_1, x_2)$.

We then generate a random typical sequence $\mathbf{q}$, where $p(\mathbf{q}) = \prod_{i=1}^{n} p(q^{(i)})$ and each entry is chosen i.i.d. according to $p(q)$. Every node knows the sequence $\mathbf{q}^n$. Below we describe the codebook generation and encoding for transmitter 1. We follow a similar procedure at transmitter 2. Please refer to Fig. 2 for a depiction of the encoder structure.

Codebook Generation:

Each codebook in the ensemble is constructed as follows. We first split the message W_{11}, which is to be decoded at the receiver 1, as $W_{11} = \{W_{11c}, W_{11p}\}$, where W_{11c} and W_{11p} are the common and the private messages of transmitter 1 destined to receiver 1, and split message W_{12}, which is to be decoded at the receiver 2, as $W_{12} = \{W_{12c}, W_{12p}\}$, where W_{12c} and W_{12p} are the common and the private messages of

transmitter 1 intended for receiver 2.

$$W_{11c} = [1, 2, \cdots, 2^{nR_{11c}}]$$

$$W_{11p} = [1, 2, \cdots, 2^{nR_{11p}}]$$

$$W_{12c} = [1, 2, \cdots, 2^{nR_{12c}}]$$

$$W_{12p} = [1, 2, \cdots, 2^{nR_{12p}}]$$

Similarly we have W_{21c}, W_{21p}, W_{22c}, and W_{22p} for transmitter 2. We generate $2^{nR_{1c}}$ i.i.d. sequences $\mathbf{v}_{1c}(w_{11c}, w_{12c})$, where

$$R_{1c} = R_{11c} + R_{12c}, \tag{4.4}$$

according to the distribution $\prod_{t=1}^{n} p(v_{1c}(t)|q(t))$, where the tuple (w_{11c}, w_{12c}) gives the codeword index denoted by $w_{1c} \in \{1, \cdots, 2^{n(R_{11c}+R_{12c})}\}$. In the sequel, we also denote these codewords with $\mathbf{v}_{1c}(w_{1c})$. For each $\mathbf{v}_1(w_{1c})$, we generate $2^{nR_{11p^*}}$ i.i.d. sequences $\mathbf{v}_{11p}(w_{1c}, w_{11p}, w_{11p'})$ according to the distribution $\prod_{t=1}^{n} p(v_{11p}(t)|v_{1c}(t), q(t))$ and randomly throw them into $2^{nR_{11p}}$ bins, where we choose

$$R_{11p} \leq R_{11p^*} \tag{4.5}$$

Here, $w_{11p} \in \{1, 2, \cdots, 2^{nR_{11p}}\}$ denotes the bin index and $w_{11p'}$ denotes the codeword index within a particular bin. Combining these two we also enumerate the codewords with $\mathbf{v}_{11p}(w_{1c}, w^*_{11p})$, where $w^*_{11p} \in \{1, 2, \cdots, 2^{nR_{11p^*}}\}$. Similarly, we generate $2^{nR_{12p^*}}$

i.i.d. sequences $\mathbf{v}_{12p}(w_{1c}, w_{12p}, w_{12p'})$ according to the distribution $\prod_{t=1}^{n} p(v_{12p}(t)|v_{1c}(t), q(t))$, and randomly throw them into $2^{nR_{12p}}$ bins, where we choose

$$R_{12p} \leq R_{12p^*} \tag{4.6}$$

Here, $w_{12p} \in \{1, 2, \cdots, 2^{nR_{12p}}\}$ denotes the bin index and $w_{12p'}$ denotes the codeword index of a particular bin. Combining these two we also enumerate the codewords with $\mathbf{v}_{12p}(w_{1c}, w^*_{12p})$, where $w^*_{12p} \in \{1, 2, \cdots, 2^{nR_{12p^*}}\}$. Second transmitter uses a similar strategy to generate the following sequences: $\mathbf{v}_{2c}(w_{21c}, w_{22c})$, $\mathbf{v}_{21p}(w_{2c}, w_{21p}, w_{21p'})$, and $\mathbf{v}_{22p}(w_{2c}, w_{22p}, w_{22p'})$, where we require

$$\begin{aligned} R_{21p} &\leq R_{21p^*} \\ R_{22p} &\leq R_{22p^*} \\ R_{2c} &= R_{21c} + R_{22c}. \end{aligned} \tag{4.7}$$

Encoding: To transmit message tuple (w_{11}, w_{12}), transmitter 1 first splits them into $(w_{11c}, w_{11p}, w_{12c}, w_{12p})$. Then, it looks for codewords $\mathbf{v}_{11p}$ in the bin w_{11p} and codewords $\mathbf{v}_{12p}$ in the bin w_{12p}, respectively, satisfying

$$\begin{aligned} &(\mathbf{q}, \mathbf{v}_{1c}(w_{1c}), \mathbf{v}_{11p}(w_{1c}, w_{11p}, j), \mathbf{v}_{12p}(w_{1c}, w_{12p}, k)) \\ &\qquad \in \mathcal{A}^{(n)}_{\epsilon}(Q, V_{1c}, V_{11p}, V_{12p}). \end{aligned} \tag{4.8}$$

Here indices j and k denote codeword indices within the given bins. If there is no such pair of codewords, then an encoding error will be declared. If there is more than one pair, then one is randomly chosen. After finding such a tuple, the encoder generates the channel input $\mathbf{x}_1$ according to $p(\mathbf{x}_1) = \prod_{t=1}^{n} p(x_1(t)|v_{1c}(t), v_{11p}(t), v_{12p}(t), q(t))$.

Similarly, transmitter 2 generates its channel input $\mathbf{x}_2$.

Decoding:

In the following we describe the decoding strategy for receiver 1. Similar steps are taken at receiver 2. Receiver 1 tries to obtain the estimates $(\hat{w}_{11c}, \hat{w}_{11p}, \hat{w}_{21c}, \hat{w}_{21p})$ to construct the message estimates $(\hat{w}_{11}, \hat{w}_{21})$. Accordingly, it looks for tuples $(w_{1c}, w_{11p^*}, w_{2c}, w_{21p^*})$ satisfying $(\mathbf{q}, \mathbf{v}_{1c}(w_{1c}), \mathbf{v}_{2c}(w_{2c}), \mathbf{v}_{11p}(w_{1c}, w_{11p^*}), \mathbf{v}_{21p}(w_{2c}, w_{21p^*}), \mathbf{y}_1)$

$$l \in \mathcal{A}_\epsilon^{(n)}(Q, V_{1c}, V_{2c}, V_{11p}, V_{21p}, Y_1) \tag{4.9}$$

If such tuple exists with unique indices, it will first obtain w_{11c} from w_{1c}, w_{21c} from w_{2c}, w_{11p} from w_{11p^*}, and w_{21p} from w_{21p^*}, then it will set $\hat{w}_{11c} = w_{11c}$, $\hat{w}_{21c} = w_{21c}$, $\hat{w}_{11p} = w_{11p}$ and $\hat{w}_{21p} = w_{21p}$; otherwise it will declare an error. After estimating $(\hat{w}_{11c}, \hat{w}_{21c}, \hat{w}_{11p}, \hat{w}_{21p})$ the receiver will obtain the corresponding message estimates $\hat{w}_{11}$ and $\hat{w}_{21}$.

Error Probability Analysis:

We first focus on error probability $P_{e,1}^{(n)}$. Without loss of generality and by the symmetrical property of the ensemble it suffices to consider $w_{11} = w_{12} = w_{21} = w_{22} = 1$ is transmitted. We also assume that, if there is no encoding error, the first codewords in the bins are chosen at the encoders (for example, $j = k = 1$ in (4.8).

We consider the following events.

$$
\begin{aligned}
lE_1 : \text{There is no pair } &(\mathbf{v}_{11p}, \mathbf{v}_{12p}) \text{ such that} \\
&(\mathbf{q}, \mathbf{v}_{1c}(1), \mathbf{v}_{11p}(1,1,k_1), \mathbf{v}_{12p}(1,1,j_1)) \\
&\in \mathcal{A}_\epsilon^{(n)}(Q, V_{1c}, V_{11p}, V_{12p}) \\
E_2 : \text{There is no pair } &(\mathbf{v}_{21p}, \mathbf{v}_{22p}) \text{ such that} \\
&(\mathbf{q}, \mathbf{v}_{2c}(1), \mathbf{v}_{21p}(1,1,j_2), \mathbf{v}_{22p}(1,1,k_2)) \\
&\in \mathcal{A}_\epsilon^{(n)}(Q, V_{2c}, V_{21p}, V_{22p}) \\
E_3 : &(\mathbf{q}, \mathbf{v}_{1c}(1), \mathbf{v}_{2c}(1), \mathbf{v}_{11p}(1,1,1), \mathbf{v}_{21p}(1,1,1), \mathbf{y}_1) \\
&\text{does not satisfy (4.9)} \\
E_4 : &(\mathbf{q}, \mathbf{v}_{1c}(i_1), \mathbf{v}_{2c}(i_2), \mathbf{v}_{11p}(i_1,k_1^*), \mathbf{v}_{21p}(i_2,k_2^*), \mathbf{y}_1) \\
&\text{satisfies (4.9) with } (i_1,i_2,k_1^*,k_2^*) \neq (1,1,1,1)
\end{aligned}
$$

From the analysis of encoding error probability in [47,48], $Pr\{E_1\} \leq \epsilon$ as $n \to \infty$, if

$$R_{11p} + R_{12p} < R_{11p^*} + R_{12p^*} - I(V_{11p}; V_{12p}|V_{1c}, Q). \tag{4.10}$$

Similarly $Pr\{E_2\} \leq \epsilon$ as $n \to \infty$, if

$$R_{21p} + R_{22p} < R_{21p^*} + R_{22p^*} - I(V_{21p}; V_{22p}|V_{2c}, Q). \tag{4.11}$$

Asymptotic equipartition property (see, e.g., [8]) assures that $Pr\{E_3\} \leq \epsilon$ for sufficiently large n.

It remains to show the conditions for which $Pr\{E_4|E_3^c\} \leq \epsilon$ for sufficiently large n, as $P_{e,1}^{(n)} \leq Pr\{E_1\} + Pr\{E_2\} + Pr\{E_3\} + Pr\{E_4|E_3^c\}$. We first define the following

event

$$E_4(\mathbf{i}) = \big\{(\mathbf{q}, \mathbf{v}_{1c}(i_1), \mathbf{v}_{2c}(i_2), \mathbf{v}_{11p}(i_1, k_1^*), \mathbf{v}_{21p}(i_2, k_2^*), \mathbf{y}_1) \\ \in \mathcal{A}_\epsilon^{(n)}(Q, V_{1c}, V_{2c}, V_{11p}, V_{21p}, Y_1) \big| E_3^c\big\}$$

where the index vector is given by $\mathbf{i} = \{i_1, i_2, k_1^*, k_2^*\}$. Then, using the Boole's inequality (a.k.a, the union bound), we write

$$\begin{aligned}
Pr\{E_4|E_3^c\} &= Pr\left\{\bigcup_{(i_1,i_2,k_1^*,k_2^*)\neq(1,1,1,1)} E_4(\mathbf{i})\right\} \\
&\leq \sum_{\substack{i_1\neq 1, i_2=1\\ k_1^*=1, k_2^*=1}} Pr\{E_4(\mathbf{i})\} \\
&+ \sum_{\substack{i_1\neq 1, i_2\neq 1\\ k_1^*=1, k_2^*=1}} Pr\{E_4(\mathbf{i})\} + \sum_{\substack{i_1= 1, i_2\neq 1\\ k_1^*=1, k_2^*=1}} Pr\{E_4(\mathbf{i})\} \\
&+ \sum_{\substack{i_1\neq 1, i_2= 1\\ k_1^*\neq 1, k_2^*=1}} Pr\{E_4(\mathbf{i})\} + \sum_{\substack{i_1= 1, i_2= 1\\ k_1^*\neq 1, k_2^*=1}} Pr\{E_4(\mathbf{i})\} \\
&+ \sum_{\substack{i_1\neq 1, i_2\neq 1\\ k_1^*\neq 1, k_2^*=1}} Pr\{E_4(\mathbf{i})\} + \sum_{\substack{i_1= 1, i_2\neq 1\\ k_1^*\neq 1, k_2^*=1}} Pr\{E_4(\mathbf{i})\} \\
&+ \sum_{\substack{i_1\neq 1, i_2= 1\\ k_1^*= 1, k_2^*\neq 1}} Pr\{E_4(\mathbf{i})\} + \sum_{\substack{i_1= 1, i_2= 1\\ k_1^*= 1, k_2^*\neq 1}} Pr\{E_4(\mathbf{i})\} \\
&+ \sum_{\substack{i_1\neq 1, i_2\neq 1\\ k_1^*= 1, k_2^*\neq 1}} Pr\{E_4(\mathbf{i})\} + \sum_{\substack{i_1= 1, i_2\neq 1\\ k_1^*= 1, k_2^*\neq 1}} Pr\{E_4(\mathbf{i})\} \\
&+ \sum_{\substack{i_1\neq 1, i_2= 1\\ k_1^*\neq 1, k_2^*\neq 1}} Pr\{E_4(\mathbf{i})\} + \sum_{\substack{i_1= 1, i_2= 1\\ k_1^*\neq 1, k_2^*\neq 1}} Pr\{E_4(\mathbf{i})\} \\
&+ \sum_{\substack{i_1\neq 1, i_2\neq 1\\ k_1^*\neq 1, k_2^*\neq 1}} Pr\{E_4(\mathbf{i})\} + \sum_{\substack{i_1= 1, i_2\neq 1\\ k_1^*\neq 1, k_2^*\neq 1}} Pr\{E_4(\mathbf{i})\}
\end{aligned}$$

From joint typicality results (see, e.g., [8]), we can show that $Pr\{E_4|E_3^c\}$ vanishes for sufficiently large n, once the rates satisfy the following equations.

$$R_{1c} < I(V_{1c}, V_{11p}; Y_1|V_{2c}, V_{21p}, Q) \quad (4.12)$$

$$R_{1c} + R_{2c} < I(V_{1c}, V_{11p}, V_{2c}, V_{21p}; Y_1|Q) \quad (4.13)$$

$$R_{2c} < I(V_{2c}, V_{21p}; Y_1|V_{1c}, V_{11p}, Q) \quad (4.14)$$

$$R_{1c} + R_{11p^*} < I(V_{1c}, V_{11p}; Y_1|V_{2c}, V_{21p}, Q) \quad (4.15)$$

$$R_{11p^*} < I(V_{11p}; Y_1|V_{1c}, V_{2c}, V_{21p}, Q) \quad (4.16)$$

$$R_{1c} + R_{11p^*} + R_{2c} < I(V_{1c}, V_{11p}, V_{2c}, V_{21p}; Y_1|Q) \quad (4.17)$$

$$R_{11p^*} + R_{2c} < I(V_{11p}, V_{2c}, V_{21p}; Y_1|V_{1c}, Q) \quad (4.18)$$

$$R_{1c} + R_{21p^*} < I(V_{1c}, V_{11p}, V_{21p}; Y_1|V_{2c}, Q) \quad (4.19)$$

$$R_{21p^*} < I(V_{21p}; Y_1|V_{1c}, V_{2c}, V_{21p}, Q) \quad (4.20)$$

$$R_{1c} + R_{2c} + R_{21p^*} < I(V_{1c}, V_{11p}, V_{2c}, V_{21p}; Y_1|Q) \quad (4.21)$$

$$R_{2c} + R_{21p^*} < I(V_{2c}, V_{21p}; Y_1|V_{1c}, V_{11p}, Q) \quad (4.22)$$

$$R_{1c} + R_{11p^*} + R_{21p^*} < I(V_{1c}, V_{11p}, V_{21p}; Y_1|V_{2c}, Q) \quad (4.23)$$

$$R_{11p^*} + R_{21p^*} < I(V_{11p}, V_{21p}; Y_1|V_{1c}, V_{2c}, Q) \quad (4.24)$$

$$R_{1c} + R_{2c} + R_{11p^*} + R_{21p^*} < I(V_{1c}, V_{2c}, V_{11p}, V_{21p}; Y_1|Q) \quad (4.25)$$

$$R_{2c} + R_{11p^*} + R_{21p^*} < I(V_{2c}, V_{11p}, V_{21p}; Y_1|V_{1c}, Q) \quad (4.26)$$

Noting that (4.12), (4.13), (4.14), (4.17), (4.18), (4.19), and (4.21) are redundant, we obtain that $P_{e,1}^{(n)}$ vanishes as n increases if (4.4), (4.5), (4.6), (4.7), (4.10), (4.11), (4.15), (4.16), (4.20), (4.22), (4.23), (4.24), (4.25), (4.26) are satisfied, which gives the rate region defined by $\mathcal{R}_I(p)$.

Similarly $P_{e,2}^{(n)}$ vanishes for sufficiently large n, if the rates belong to the region $\mathcal{R}_{II}(p)$. Finally, it can be readily observed that, for a given p, any rate tuple inside the region $\mathcal{R}_I(p) \bigcap \mathcal{R}_{II}(p)$ is achievable, which concludes the proof of the theorem. □

4.4 Special Cases

In the following subsections we discuss the special cases of the rate region given above.

4.4.1 The Broadcast Channel

For a given X channel specified by $p(y_1, y_2|x_1, x_2) = p(y_1, y_2|x_1)$, if the second transmitter is silenced (that is x_2 is removed) the setting is reduced to a broadcast channel with $p(y_1, y_2|x_1)$. For this special case, the region $\mathcal{R}_I(p) \cap \mathcal{R}_{II}(p)$ reduces to the following

$$
\begin{aligned}
R_{1c} + R_{11p^*} &< I(V_{1c}, V_{11p}; Y_1|Q) && (4.27)\\
R_{11p^*} &< I(V_{11p}; Y_1|V_{1c}, Q) && (4.28)\\
R_{1c} + R_{12p^*} &< I(V_{1c}, V_{12p}; Y_2|Q) && (4.29)\\
R_{12p^*} &< I(V_{12p}; Y_2|V_{1c}, Q) && (4.30)\\
R_{11p} + R_{12p} &< R_{11p^*} + R_{12p^*} - I(V_{11p}; V_{12p}|V_{1c}, Q) && (4.31)\\
R_{1c} &= R_{11c} + R_{12c} && (4.32)\\
R_{11p} &\leq R_{11p^*} && (4.33)\\
R_{12p} &\leq R_{12p^*} && (4.34)
\end{aligned}
$$

for a given $p \in \mathcal{P}$, where we set V_{2c}, V_{21p}, V_{22p} to be deterministic (as channel input of transmitter 2 does not affect the received signals, this does not reduce the achievable

rate region). We further set Q to be deterministic. Let $R_{11} = R_{11c} + R_{11p}$, $R_{12} = R_{12c} + R_{12p}$, by applying the Fourier-Motzkin elimination we obtain the following.

Any non-negative rate pair (R_{11}, R_{12}) satisfying

$$\begin{aligned} R_{11} &< I(V_{1c}, V_{11p}; Y_1) && (4.35)\\ R_{12} &< I(V_{1c}, V_{12p}; Y_2) && (4.36)\\ R_{11} + R_{12} &< \min\{I(V_{1c}; Y_1), I(V_{1c}; Y_2)\} + I(V_{11p}; Y_1|V_{1c}) + I(V_{12p}; Y_2|V_{2c}) \\ &\quad - I(V_{11p}, V_{12p}|V_{1c}) && (4.37) \end{aligned}$$

for some $p(v_{1c}, v_{11p}, v_{12p})p(x_1|v_{1c}, v_{11p}, v_{12p})$ is achievable for the broadcast channel given by $p(y_1, y_2|x_1)$. This is the Marton's rate region [11] which is the best known inner bound to the capacity region of the two-user broadcast channels.

4.4.2 The Interference Channel

For the interference channel, the cross messages do not exist and hence we set random variables V_{12p}, V_{21p} to be deterministic. We also choose the common and private auxiliary random variables in the region $\mathcal{R}_I(p) \cap \mathcal{R}_{II}(p)$, to be independent. Next, we set $R_{11} = R_{11c}+R_{11p}$, $R_{22} = R_{22c}+R_{22p}$, $R_{12c} = R_{12p} = R_{21c} = R_{21p} = 0$, and apply Fourier-Motzkin elimination to the obtained region and choose $p(x_1|v_{1c}, v_{11p}, q)$, $p(x_2|v_{2c}, v_{22p}, q)$ to be deterministic functions. We obtain the following region.

Any non-negative rate pair (R_{11}, R_{22}) satisfying

$$\begin{aligned}
R_{11} &< I(X_1; Y_1|V_{2c}, Q) \\
R_{11} &< I(X_1; Y_1|V_{1c}, V_{2c}, Q) + I(V_{1c}; Y_2|X_2, Q) \\
R_{22} &< I(X_2; Y_2|V_{1c}, Q) \\
R_{22} &< I(X_2; Y_2|V_{1c}, V_{2c}, Q) + I(V_{2c}; Y_1|X_1, Q) \\
R_{11} + R_{22} &< I(X_2; Y_2|V_{1c}, V_{2c}, Q) + I(X_1, V_{2c}; Y_1|Q) \\
R_{11} + R_{22} &< I(X_1; Y_1|V_{1c}, V_{2c}, Q) + I(X_2, V_{1c}; Y_2|Q) \\
R_{11} + R_{22} &< I(X_1, V_{2c}; Y_1|V_{1c}, Q) + I(X_2, V_{1c}; Y_2|V_{2c}, Q)
\end{aligned}$$

$$\begin{aligned}
2R_{11} + R_{22} &< I(X_1; Y_1|V_{1c}, V_{2c}, Q) + I(X_1, V_{2c}; Y_1|Q) + I(X_2, V_{1c}; Y_2|V_{2c}, Q) \\
R_{11} + 2R_{22} &< I(X_2; Y_2|V_{1c}, V_{2c}, Q) + I(X_2, V_{1c}; Y_2|Q) + I(X_1, V_{2c}; Y_1|V_{1c}, Q)
\end{aligned}$$

for some $p(q)p(v_{1c}|q)p(v_{11p}|q)p(x_1|v_{1c}, v_{11p}, q)p(v_{2c}|q)\ p(v_{22p}|q)p(x_2|v_{2c}, v_{22p}, q)$, is achievable for the interference channel given by $p(y_1, y_2|x_1, x_2)$.

This region is the region given in Lemma 1 of [50], which is shown to be equal to the compact form of Han and Kobahayashi rate region (see also [49]) by utilizing Lemma 2 of [50].

It should be noted that the scheme of [50] does not consider the event of not correctly decoding unintended common messages at the receivers as an error event. However, in deriving our rate region for the X channel, we consider that event as a decoding error.

Interestingly, applying the Fourier-Motzkin elimination to the proposed rate region, for the special case of interference channel, results in the compact form of the

Han and Kobayashi rate region. Therefore, the exclusion of the aforementioned error event does not affect the result obtained in [50].

CHAPTER 5

DOWNLINK COMMUNICATION IN A COGNITIVE CELLULAR NETWORK

Cognitive broadcast channel, where two multi-antenna transmitters communicate with their respective receivers, is considered. One of the transmitters is said to be cognitive (secondary) as it is assumed to know the messages of the other (primary) transmitter prior to their transmission. The goal is to design cooperative schemes between the two transmitters, which impose only minimal changes to the primary broadcast channel (compared to the non-cognitive scenario). Towards this end, an achievable scheme is provided under which both intra cell and inter cell interferences at the primary receivers are aligned. The interference at the secondary receivers, on the other hand, is canceled by dirty paper coding. The corresponding achievable region and an outer bound region are provided in terms of the degrees of freedom (DoF) metric. Special cases shows the optimality of the proposed scheme in the high SNR regime for those cases. We also illustrate the advantage of cognitive cooperation, over the non-cognitive system by proving that the achieved sum DoF is strictly larger than the non-cognitive case.

5.1 Introduction

A significant factor in limiting the performance of cellular systems is the interference from other cells also known as *inter cell interference*. Interference from other users also degrades the achievable throughput in a K-user MIMO interference channel. An important technique proposed to mitigate these effects is interference alignment (IA) [20], [10]. Roughly speaking, IA *aligns* all the unwanted (interfering) signals to certain dimensions allowing the intended messages to be communicated over the remaining interference free ones.

To achieve the gains promised by IA, the users need to have perfect and global knowledge of each others channel state information (CSI). That is, all the transmitters need to know *all* the channel realizations before forming their signals. Because the CSI needs to be obtained through training sequences and feedback, this introduces a serious overhead to the system. In [51], [52] the authors have considered the DoF region of MIMO networks in the absence of CSI at the transmitters. They have established the negative result that in most cases the degrees of freedom region can be achieved by simple time sharing which means nothing can be gained beyond the simple time division access.

In this chapter, we consider the problem of interference management for a cellular system in the downlink when the CSI is not fully available at all of the transmitters. Specifically, the secondary BS (to be defined below) is assumed to have full channel CSI, while the primary BS is assumed to only have CSI knowledge of its own cell as well the inter-cell CSI.

One of the base stations is also assumed to be *cognitive*. The cognitive message

sharing means that the messages of one cell's (henceforth the non-cognitive or primary cell) are made available non-causally to the other one (henceforth the cognitive or secondary cell). In developing achievable strategies for a cognitive system, we want to make sure that few changes are made to the communication scheme of the primary cell. However, the cognitive cell can adapt its signaling strategy to handle different cases.

One factor that makes it challenging to apply the idea of IA in a cellular system is the fact that if we align the interference signals on one of the users they may not be aligned at the rest of the users in the same cell. In a cellular system, each mobile user experiences two kinds of interference: 1) The interference caused by the other cell or the *inter-cell* interference, 2) The interference caused by the message intended for the other users within its own cell or the *intra-cell* interference. Given this interference setting we can *align* the intra-cell interference in the primary cell for each user to the linear space spanned by the inter-cell interference caused by the secondary BS. By doing so, they can be canceled at the same time. As the cognitive BS has full CSI knowledge and knows the the primary messages, the inter-cell interference from the primary BS to cognitive users are canceled using dirty paper coding (DPC) [17]. Finally, using the techniques proposed in [53] and [25], we derive an outer bound on the sum DoF region and show that our proposed schemes under some special cases to be optimal.

The rest of the chapter is organized as follows: Section 5.2 introduces the system model and problem formulation. In Section 5.3, we explain our proposed signaling scheme in detail. The outer bounds and special cases are studied in Section 5.4.

5.2 System model

A cellular system with one primary and one secondary base stations (denoted by $\mathcal{P}$ and $\mathcal{S}$ respectively) is considered. The primary and secondary base stations serve K_P (denoted by $\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_{K_P}$) and K_S (denoted by $\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_{K_S}$) mobile users in their cells respectively.

Let us denote the messages intended for primary users by $W_{\mathcal{P}_1}, W_{\mathcal{P}_2}, \cdots, W_{\mathcal{P}_{K_P}}$, and for the secondary users by $W_{\mathcal{S}_1}, W_{\mathcal{S}_2}, \cdots, W_{\mathcal{S}_{K_S}}$. The total power available at the base stations is denoted by ρ. Rates of $R_{\mathcal{P}_i}(\rho)$, $R_{\mathcal{S}_j}(\rho)$ are said to be achievable with power ρ, if there exists a coding scheme to reliably communicate messages of sizes $|W_{\mathcal{P}_i(\rho)}| = 2^{nR_{\mathcal{P}_i}(\rho)}$ and $|W_{\mathcal{S}_j}(\rho)| = 2^{nR_{\mathcal{S}_j}(\rho)}$ to mobile users $\mathcal{P}_i$ and $\mathcal{S}_j$, where n is the number of channel uses. The set of all achievable rate tuples at power ρ is denoted by $\mathcal{C}(\rho)$. Following the notation introduced in [21], the sum DoF in the secondary and primary cells d_S and d_P are respectively defined by:

$$d_S = \limsup_{\rho\to\infty} \Big[\sup_{\mathcal{R}(\rho)\in\mathcal{C}(\rho)} \big[\sum_{j=1}^{K_S} R_{\mathcal{S}_j}(\rho) \big] \frac{1}{\log(\rho)} \Big] \tag{5.1}$$

and

$$d_P = \limsup_{\rho\to\infty} \Big[\sup_{\mathcal{R}(\rho)\in\mathcal{C}(\rho)} \big[\sum_{i=1}^{K_P} R_{\mathcal{P}_i}(\rho) \big] \frac{1}{\log(\rho)} \Big] \tag{5.2}$$

The set of all achievable pairs of (d_S, d_P) is denoted by $\mathcal{D}_{\text{sum}}$, and is referred to as sum DoF region. We also refer to $d_S + d_P$ as the *total sum DoF*.

Throughout the chapter, it is assumed that the secondary base station non-causally knows the messages intended for the primary users i.e., at $\mathcal{S}$, the messages $\{W_{\mathcal{P}_i}(\rho)\}_{i=1}^{K_P}$ are known prior to transmission. The users are also assumed to be equipped with multiple antennas. Let m_P and m_S denote the number of antennas

at the primary and secondary base stations and n_P and n_S be the number of antennas at the primary and secondary users. The explained model is referred to as a $\{m_P, m_S, n_P, n_S, K_P, K_S\}$ cognitive system.

In general, using time and/or frequency expansions (by multiple fading blocks and/or multiple OFDM subcarriers) we can generate L extra dimensions on each user. In that case, the number of available dimensions on each node are equal to: $M_P = L \times m_P$, $M_S = L \times m_S$, $N_P = L \times n_P$ and $N_S = L \times n_S$ respectively.
The received signal at user $\mathcal{P}_i$ is equal to:

$$\mathbf{y}_{\mathcal{P}_i} = \mathbf{H}_{\mathcal{P}_i}\mathbf{x}_P + \mathbf{H}'_{\mathcal{P}_i}\mathbf{x}_S + \mathbf{z}_{\mathcal{P}_i}, \tag{5.3}$$

where $\mathbf{H}$'s represents the *extended* MIMO channel coefficients to the users and $\mathbf{H}'$'s are the extended MIMO channel from the cognitive base station. For both, $\mathbf{H}$ and $\mathbf{H}'$ the subscript denotes the receiver.

For example, $\mathbf{H}_{P_i}$ which is a matrix of size $N_P \times M_P$ and $\mathbf{H}'_{P_i}$ which is $N_P \times M_S$ are the channels from the primary and secondary base stations to the i-th primary user respectively. $(\mathbf{x}_P)_{M_P \times 1}$ and $(\mathbf{x}_S)_{M_S \times 1}$ denote the signals transmitted from the primary and secondary base stations and $\mathbf{z}$'s are the zero mean unit variance i.i.d additive white Gaussian noise.

Similarly, the signal received at the j-th secondary mobile user will be:

$$\mathbf{y}_{\mathcal{S}_j} = \mathbf{H}_{\mathcal{S}_j}\mathbf{x}_S + \mathbf{H}'_{\mathcal{S}_j}\mathbf{x}_P + \mathbf{z}_{\mathcal{S}_j} \tag{5.4}$$

In this chapter, it is assumed that the channel coefficients are drawn independently from a continuous distribution and thus the channel matrices are full rank almost surely. Also, we assume the cognitive base station to have full CSI knowledge. The

primary base station is assumed to know its own cell's CSI as well as the inter cell CSI. That is, all $\mathbf{H}_{\mathcal{P}_i}$'s and $\mathbf{H}'_{\mathcal{P}_i}$'s are known at the primary BS.

5.3 Main Result

In this section, we present an achievable region for the sum DoF in the primary and secondary cells under the explained system model.

Theorem 5.1. *For a $\{m_P, m_S, n_P, n_S, K_P, K_S\}$ cognitive system, denote the set of all pairs (d_S, d_P) for which*

$$\begin{cases} 0 \leq d_S \leq \min\left\{K_S n_S, m_S\right\} \\ 0 \leq d_P \leq \min\left\{K_P(n_P - d_S)^+, (m_P + m_S - d_S)\right\}, \end{cases}$$

by $\mathcal{D}^{in}_{\text{sum}}$. Then, $\mathcal{D}^{in}_{\text{sum}} \subseteq \mathcal{D}_{\text{sum}}$

Proof. First let us assume d_S to be a rational number and denote its irreducible form by:

$$d_S = \frac{S}{L}$$

$L \in \mathbb{Z}^+$ extra dimensions in time or frequency (through multiple fading blocks or OFDM subcarriers) are generated. The achievable sum DoF in the secondary cell cannot exceed m_S, and we always have: $S \leq M_S$.

Because the cognitive BS has full CSI knowledge including the CSI from itself to the primary users and also knows all the primary messages it can *lend* $(M_S - S)^+$ of its available dimensions which are not used for data transmission of the secondary users to the primary BS. More specifically, for each primary user $\mathcal{P}_i$ we "append" the channels from the first $(M_S - S)$ dimensions of the cognitive BS (corresponding to the first $(M_S - S)$ columns of $\mathbf{H}'_{\mathcal{P}_i}$) to $\mathbf{H}_{\mathcal{P}_i}$ to form a new matrix $\bar{\mathbf{H}}_{\mathcal{P}_i}$ of size $N_P \times (M_P + M_S - S)$. Those columns are deleted from $\mathbf{H}'_{\mathcal{P}_i}$ and the $N_P \times S$ matrix

$\bar{\mathbf{H}}'_{\mathcal{P}_i}$ is formed.

Similarly, by allocating first $(M_S - S)$ dimensions of the secondary BS to data transmission of the primary users, the data transmitted from those can be thought of as inter-cell interference on each secondary user. In other words, we can form the matrices $\bar{\mathbf{H}}_{\mathcal{S}_j}$ and $\bar{\mathbf{H}}'_{\mathcal{S}_j}$ of sizes $N_S \times S$ and $N_S \times (M_P + M_S - S)$ capturing the matrices which carry intra-cell and inter-cell data to user $\mathcal{S}_j$.

Let us specify the transmission scheme to the primary users first. For each $(\bar{\mathbf{H}}'_{\mathcal{P}_i})_{[N_P \times S]}$, the primary base station calculates r linearly independent normalized basis vectors of its null space denoted by $\mathbf{u}^1_{\mathcal{P}_i}, ..., \mathbf{u}^r_{\mathcal{P}_i}$ where $r = (N_P - S)^+$. The goal of $\mathcal{P}$, is to find zero forcing beam forming vectors such that on each primary user the intra-cell interference is *aligned* to the same linear space spanned by the inter-cell interference.

To this end, the following matrix is formed:

$$(\bar{\mathbf{U}}_{\mathcal{P}})_{\left[K_P(N_P-S)^+ \times \left(M_P+M_S-S\right)\right]} = \left(\begin{array}{c} (\mathbf{u}^1_{\mathcal{P}_1})^T \bar{\mathbf{H}}_{\mathcal{P}_1} \\ (\mathbf{u}^2_{\mathcal{P}_1})^T \bar{\mathbf{H}}_{\mathcal{P}_1} \\ \cdots \\ (\mathbf{u}^r_{\mathcal{P}_1})^T \bar{\mathbf{H}}_{\mathcal{P}_1} \\ \hline (\mathbf{u}^1_{\mathcal{P}_2})^T \bar{\mathbf{H}}_{\mathcal{P}_2} \\ (\mathbf{u}^2_{\mathcal{P}_2})^T \bar{\mathbf{H}}_{\mathcal{P}_2} \\ \cdots \\ (\mathbf{u}^r_{\mathcal{P}_2})^T \bar{\mathbf{H}}_{\mathcal{P}_2} \\ \hline \cdots \\ \hline (\mathbf{u}^1_{\mathcal{P}_{K_P}})^T \mathbf{H}_{\mathcal{P}_{K_P}} \\ (\mathbf{u}^2_{\mathcal{P}_{K_P}})^T \bar{\mathbf{H}}_{\mathcal{P}_{K_P}} \\ \cdots \\ (\mathbf{u}^r_{\mathcal{P}_{K_P}})^T \bar{\mathbf{H}}_{\mathcal{P}_{K_P}} \end{array}\right)$$

To achieve $d_P \leq \frac{1}{L}\min\{M_P + M_S - S, K_P(N_P - S)^+\}$ sum DoF in the primary cell, without loss of generality pick the first $P = L \times d_P$ rows of $\bar{\mathbf{U}}_{\mathcal{P}}$ as:

$$\bar{\mathbf{U}}_P = \begin{pmatrix} \frac{(\bar{\mathbf{u}}_1)^T}{} \\ \frac{(\bar{\mathbf{u}}_2)^T}{} \\ \frac{\cdots}{} \\ \frac{(\bar{\mathbf{u}}_P)^T}{} \end{pmatrix}$$

Denote the right pseudo-inverse of matrix $\bar{\mathbf{U}}_P$ by $\bar{\mathbf{V}}_P$:

$$\bar{\mathbf{V}}_P = \left[\begin{array}{c|c|c|c} \bar{\mathbf{v}}_{p_1} & \bar{\mathbf{v}}_{p_2} & \cdots & \bar{\mathbf{v}}_{p_P} \end{array} \right]$$

which means $\bar{\mathbf{U}}_P\bar{\mathbf{V}}_P = (\mathbf{I})_{P\times P}$.

For each $\bar{\mathbf{v}}_{p_i}$ denote its first M_P elements by $\mathbf{v}_{p_i}$ and the next $(M_S - S)$ ones by $\mathbf{v}'_{p_i}$. The signal transmitted from the primary BS is formed as:

$$\mathbf{x}_{\mathcal{P}} = \mathbf{x}_{p_1}\mathbf{v}_{p_1} + \mathbf{x}_{p_2}\mathbf{v}_{p_2} + \cdots + \mathbf{x}_{p_P}\mathbf{v}_{p_P},$$

where P streams of data, $\mathbf{x}_{p_1}, \mathbf{x}_{p_2} \cdots, \mathbf{x}_{p_P}$ are encoded Using a Gaussian codebook.

It remains to specify the transmission scheme from the cognitive BS. Let us define:

$$(\bar{\mathbf{H}}_{\mathcal{S}})_{[K_S N_S \times S]} = \begin{pmatrix} \frac{\bar{\mathbf{H}}_{\mathcal{S}_1}}{} \\ \frac{\mathbf{H}_{\mathcal{S}_2}}{} \\ \frac{\cdots}{} \\ \frac{\bar{\mathbf{H}}_{\mathcal{S}_{K_S}}}{} \end{pmatrix}$$

In order to achieve a sum DoF of d_S in the secondary cell, S streams of data are required to be reliably transmitted to the secondary users. To this end, let us denote the S rows of $\bar{\mathbf{H}}_{\mathcal{S}}$ by

$$[(\bar{\mathbf{h}}_{s_1})^T; ...; (\bar{\mathbf{h}}_{s_S})^T],$$

and form the zero forcing beam-forming vectors to the secondary users similar to a MIMO broadcast channel. That is, the beamforming vector $\mathbf{v}'_{s_i}$ is picked as the orthonormal basis of the null space of the vector space spanned by:

$$[(\bar{\mathbf{h}}_{s_1})^T; ...; (\bar{\mathbf{h}}_{s_{i-1}})^T; (\bar{\mathbf{h}}_{s_{i+1}})^T; ...; (\bar{\mathbf{h}}_{s_S})^T]$$

This is possible for $S \leq K_S \times N_S$. Next, the data stream $\mathbf{x}_{s_i}$ is encoded using dirty paper coding in $\hat{\mathbf{x}}_{s_i}$ considering $(\bar{\mathbf{h}}'_{s_i})^T \bar{\mathbf{x}}_{\mathcal{P}}$ to be the *known* interference where $(\bar{\mathbf{h}}'_{s_i})^T$ is the i-th row of $(\bar{\mathbf{H}}'_{\mathcal{S}})$ and $\bar{\mathbf{x}}_{\mathcal{P}}$ is defined as:

$$\bar{\mathbf{x}}_{\mathcal{P}} = \mathbf{x}_{p_1}\bar{\mathbf{v}}_{p_1} + \mathbf{x}_{p_2}\bar{\mathbf{v}}_{p_2} + \cdots + \mathbf{x}_{p_P}\bar{\mathbf{v}}_{p_P},$$

It should be again noted that because we assumed the secondary BS to be cognitive and have full CSI knowledge thus, $(\bar{\mathbf{h}}'_{s_i})^T \bar{\mathbf{x}}_{\mathcal{P}}$ is fully known at $\mathcal{S}$ and the secondary BS can employ dirty paper coding to cancel the interference caused by the data intended for primary mobiles on its user.

Define $\tilde{\mathbf{v}}_{p_i}$ and $\mathbf{v}_{s_j}$ as:

$$\tilde{\mathbf{v}}_{p_i} = \big[(\mathbf{v}'_{p_i})^T | \underbrace{0, \cdots, 0}_{(S)}\big]^T \quad \mathbf{v}_{s_j} = \big[\underbrace{0, \cdots, 0}_{(M_S - S)} | (\mathbf{v}'_{s_j})^T\big]^T.$$

Then, the signal transmitted by the secondary BS is equal to:

$$\mathbf{x}_{\mathcal{S}} = \sum_{i=1}^{P} \tilde{\mathbf{v}}_{p_i}\mathbf{x}_{p_i} + \sum_{j=1}^{S} \mathbf{v}_{s_j}\hat{\mathbf{x}}_{s_j}$$

Decoding: Without loss of generality, we explain the decoding of the first data streams intended for the first primary and secondary users i.e., $\mathbf{x}_{s_1}$ and $\mathbf{x}_{p_1}$. Because we are concerned with degrees of freedom which is studied asymptotically as SNR goes to infinity, following [21] for simplicity the noise vectors are ignored. In that case, the signal received at user $\mathcal{S}_1$ is equal to:

$$\mathbf{y}_{\mathcal{S}_1} = \mathbf{H}'_{\mathcal{S}_1}\mathbf{x}_{\mathcal{P}} + \mathbf{H}_{\mathcal{S}_1}\mathbf{x}_{\mathcal{S}} = \bar{\mathbf{H}}'_{\mathcal{S}_1}\bar{\mathbf{x}}_{\mathcal{P}} + \bar{\mathbf{H}}_{\mathcal{S}_1}(\hat{\mathbf{x}}_{s_1}\mathbf{v}'_{s_1} + \cdots \hat{\mathbf{x}}_{s_S}\mathbf{v}'_{s_S})$$

Due to the choice of zero-forcing vectors, the signal received on the first dimension (first antenna on the first OFDM carrier) of user $\mathcal{S}_1$ which carries $\mathbf{x}_{s_1}$ is equal to:

$$\hat{\mathbf{x}}_{s_1} + (\bar{\mathbf{h}}'_{s_1})^T \bar{\mathbf{x}}_{\mathcal{P}},$$

and $\mathbf{x}_{s_1}$ is recovered by dirty paper decoding considering $(\bar{\mathbf{h}}'_{s_1})^T\bar{\mathbf{x}}_{\mathcal{P}}$ as the known interference.

The signal received at $\mathcal{P}_1$ is equal to:

$$\mathbf{y}_{\mathcal{P}_1} = \mathbf{H}_{\mathcal{P}_1}\mathbf{x}_{\mathcal{P}} + \mathbf{H}'_{\mathcal{P}_1}\mathbf{x}_{\mathcal{S}} =$$
$$\bar{\mathbf{H}}_{\mathcal{P}_1}\left(\mathbf{x}_{p_1}\bar{\mathbf{v}}_{p_1} + \cdots + \mathbf{x}_{p_P}\bar{\mathbf{v}}_{p_P}\right) +$$
$$\bar{\mathbf{H}}'_{\mathcal{P}_1}\left(\hat{\mathbf{x}}_{s_1}\mathbf{v}'_{s_1} + \cdots + \hat{\mathbf{x}}_{s_S}\mathbf{v}'_{s_S}\right)$$

Next, to decode $\mathbf{x}_{p_1}$, user $\mathcal{P}_1$ multiplies its received signal in $\mathbf{u}^1_{\mathcal{P}_1}$, i.e.,

$$(\mathbf{u}_1)^T\mathbf{y}_{\mathcal{P}_1}$$

Because $\mathbf{u}^1_{\mathcal{P}_1}$ is in the null space of $\bar{\mathbf{H}}'_{\mathcal{P}_1}$, the inter-cell interference is canceled. By construction of $\bar{\mathbf{v}}_{p_i}$'s we have:

$$(\mathbf{u}^1_{\mathcal{P}_1})^T\bar{\mathbf{H}}_{\mathcal{P}_1}\bar{\mathbf{v}}_{p_1} = 1,$$
$$(\mathbf{u}^1_{\mathcal{P}_1})^T\bar{\mathbf{H}}_{\mathcal{P}_1}\bar{\mathbf{v}}_{p_2} = \cdots = (\mathbf{u}^1_{\mathcal{P}_1})^T\bar{\mathbf{H}}_{\mathcal{P}_1}\bar{\mathbf{v}}_{p_P} = 0$$

Thus by multiplying $(\mathbf{u}^1_{\mathcal{P}_1})^T$ in the received signal at $\mathcal{P}_1$, the inter-cell interference carried by $\bar{\mathbf{H}}'_{\mathcal{P}_1}$ and the intra-cell interference caused by $\bar{\mathbf{v}}_{p_2}, \cdots, \bar{\mathbf{v}}_{p_P}$ are both zero forced at the same time through the utilized interference alignment technique and $\mathbf{x}_{p_1}$ is decoded at $\mathcal{P}_1$.

Therefore, the proposed joint interference alignment and dirty paper coding (IA+DPC) achieves the sum DoF of $d_P \leq \min\left\{K_P(n_P - d_S)^+, (m_P + m_S - d_S)\right\}$ in the primary cell with a secondary sum DoF of $d_S \leq \min\left\{K_S n_S, m_S\right\}$.

Now for an irrational d_S there is a sequence of rational numbers converging to it. For each one the claimed d_P is achievable which means in the limit it will go to the same d_P which completes the proof of the theorem. □

5.4 Special cases and discussion

We begin the study of special cases by presenting an outer bound on the achievable sum DoF. Next, using the derived outer bounds we establish, for a special case, that the corner points of the optimal sum DoF region are achieved by the proposed signaling scheme. The benefit of the cognitive cooperation is also established for the considered special case using the outer bound.

Theorem 5.2. *Let $d_{\mathcal{P}_i}$ and $d_{\mathcal{S}_j}$ be an achievable DoF for mobile users $\mathcal{P}_i$ and $\mathcal{S}_j$. Then the DoF region of a cognitive cellular system satisfies the following bounds:*

$$\begin{aligned} L_1 &: \ d_{\mathcal{P}_i} \leq n_P, for\ 1 \leq i \leq K_P,\ d_P + d_S \leq (m_P + m_S) \\ L_2 &: \ d_S \leq \min\{m_S, K_S n_S\}. \\ L_3 &: \ d_{\mathcal{P}_i} + d_S \leq \max\{m_S, n_P\}, \quad 1 \leq i \leq K_P. \end{aligned}$$

Proof. The bound L_1 follows from the outer bounds on the point to point MIMO channel and the fact the degrees of freedom cannot exceed the number of receive or transmit antennas.

L_2 follows by assuming full cooperation between the mobile users of the secondary base station and assuming they perfectly know the interference caused by the primary base station. This cannot reduce the DoF region and L_2 follows from the outer-bounds on the DoF of the point to point MIMO channel as well.

To establish L_3, we first let $d_{\mathcal{P}_l} = 0$ for $l \neq i$ to get a bound on $d_{\mathcal{P}_i}$ and assume full cooperation at the secondary users. This reduces the problem to a MIMO interference channel (IC) with a cognitive transmitter. After the problem is reduced to a cognitive MIMO IC, we apply the sum DoF outer bound of [25] which is based on the genie aided method of [53]. It should be noted [25] assumes full and global CSI at all

nodes which is not the case in our problem. However, we can assume that the extra CSI information is also provided to all nodes which does not reduce the DoF region. Using [25] we can directly show that:

$$d_{\mathcal{P}_i} + d_S \leq \max\{m_S, n_P\}$$

For completeness however we also present a full proof in the Appendix. □

5.4.1 Special Case I

Let us consider the system with $m_P = m_S = n_P = K + 1$,$n_S = 1$ and $K_P = K + 2, K_S = K$ for an integer $K > 1$. That is a $\{K + 1, K + 1, K + 1, 1, K + 2, K\}$ cognitive system. Henceforth, we refer to this system as *example channel of type* K. Using the achievable strategy, and applying Theorem 5.1 the achievable region is the line connecting the following points:

$$\big(d_S = K; d_P = K + 2\big)$$

$$\big(d_S = 0; d_P = 2(K + 1)\big)$$

Noting that the total sum DoF of the system cannot exceed the number of transmit antennas, and the fact that $d_S = 0$ and $d_S = K$ corresponds to the corner point of the sum DoF region we can conclude that for $\{K+1, K+1, K+1, 1, K+2, K\}$ cognitive system the proposed signaling scheme is optimal in terms of DoF (i.e., in the high SNR regime). Now, let us consider the case when cognitive message sharing is not possible. If we assume all the users in the primary and secondary cell fully cooperate with all the users in their own cells, in the absence of cognition this system reduces

to a

$$\{M_1 = K+1, M_2 = K+1, N_1 = (K+2)(K+1), N_2 = K\}$$

MIMO interference channel where M_1, M_2 denote the number of antennas on the first and second transmitter and N_1, N_2 are the number of antennas on the first and second receiver respectively.

This full cooperation cannot reduce the DoF region of the cellular system without cognition. Using the bound derived in [53], the maximum total sum DoF of this MIMO interference channel is equal to $K+1$ whereas our proposed scheme with cognition achieves $d_P + d_S = 2(K+1)$ which shows cognitive cooperation under our proposed scheme *strictly* outperforms the case when cognitive message sharing of the primary messages is not available to the secondary base station.

The achievable sum DoF region for the example channels of type 2 and 3 are depicted in Fig. 5.1.

Corollary 5.1. *For a* $\{K+1, K+1, K+1, 1, \kappa, K\}$ *cognitive system with* $K > 1$ *achieves the optimal corner points of the sum DoF region for* $1 \leq \kappa \leq K+1$.

Proof. Applying Theorem 5.1 at $d_S = K$, the following point is achieved by the proposed scheme:

$$(d_S = K; d_P = \kappa)$$

In other words using the proposed scheme the DoF of $d_{\mathcal{P}_i} = d_{\mathcal{S}_j} = 1$ per mobile user is achievable for this system. Note that for this system the maximum sum DoF of the cognitive cell is equal to K, which is also achieved by the proposed signaling scheme. If we assume the cognitive cell does not loose any of its DoF by helping the primary cell i.e., transmitting at $d_S = K$, and applying the bound L_3, we arrive at $d_{\mathcal{P}_i} \leq 1$.

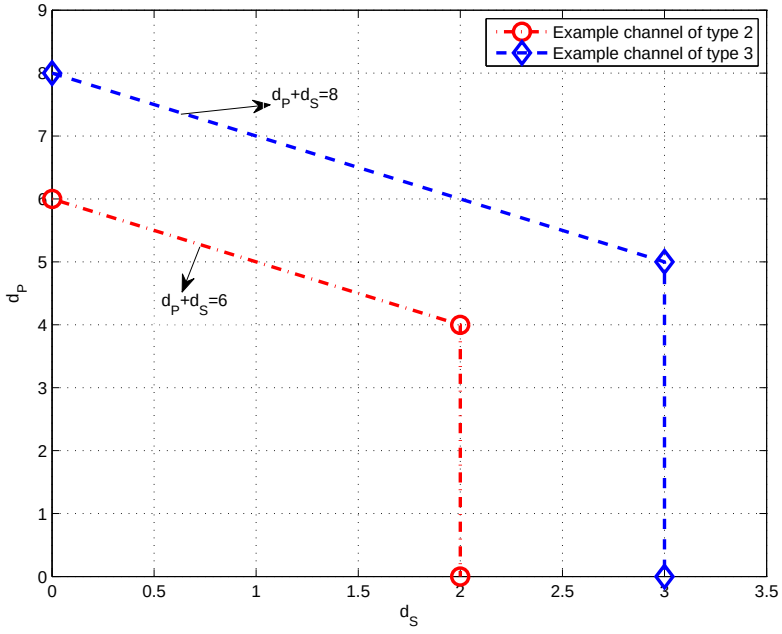

Fig. 5.1: Achievable sum DoF for the example channels of type 2 and 3

By the proposed method, $d_{\mathcal{P}_i} = 1$ is attainable. Basically, $(d_S = K; d_P = \kappa)$ is a corner point of the sum DoF region, $\mathcal{D}_{\text{sum}}$.

On the other extreme at $d_S = 0$, and applying Theorem 5.1 sum DoF in the primary cell is equal to $d_P = 2(K+1)$ (see also Theorem 5.2 for the converse) , i.e., $\mathcal{D}_{\text{sum}}^{\text{in}}$ in this case includes all the corner points of $\mathcal{D}_{\text{sum}}$. □

5.4.2 Special Case II: One antenna at all nodes

In this subsection, we apply the result obtained in Theorem 5.1 to a system with $m_P = m_S = n_P = n_S = 1$ and equal number of users in each cell i.e., $K_P = K_S = K \geq 2$. This system without a cognitive base station is considered in [22] and the normalized DoF of $\frac{1}{K+1}$ per mobile user which translates to normalized sum DoF of

$\frac{K}{K+1}$ per cell is achieved. That is, the point

$$(d_S = \frac{K}{K+1}, d_P = \frac{K}{K+1}), \tag{5.5}$$

is achieved by [22]. Our goal in this subsection is to show that point (5.5) is included in the sum DoF region achieved by our proposed scheme. Moreover, in [22] both of the base stations need to adapt their signaling scheme to handle their interference on the users of the other cell. However, in our proposed scheme the primary BS does not modify its transmission scheme to handle its interference on the secondary users and those are canceled by DPC.

Applying the Theorem 1 the following sum DoF region is achievable

$$\begin{aligned} 0 &\leq d_S \leq 1 \\ 0 &\leq d_P \leq \min\{(2-d_S), K(1-d_S)\} \end{aligned}$$

Using the above, at $d_S = \frac{K}{K+1}$ the sum DoF in primary cell is $d_P = \frac{K}{K+1}$ which means the point (5.5) is included in our region. Fig. 5.2 compares the achievable sum DoF region of our proposed scheme for the cognitive system and that of [22] for the non-cognitive counterpart.

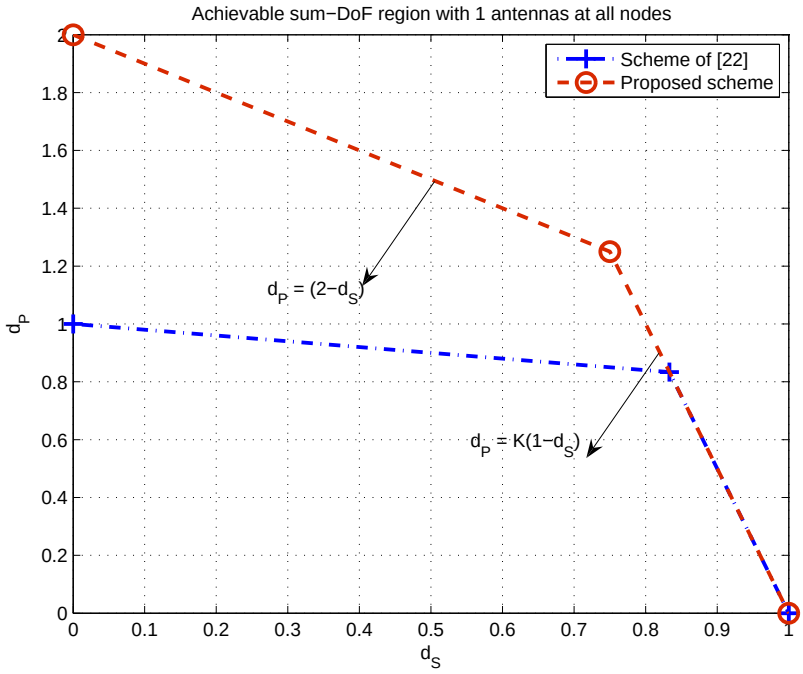

Fig. 5.2: Achievable sum DoF region with 1 antennas at all nodes, for $K = 5$

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

6.1 Summary

In the first part of the thesis we focused on the design of explicit schemes for digital fingerprinting. First, we established that for all rates less than 0.18 there is a good linear code under MD decoder. Our numerical results using ARA codes under BP decoder established it to be a close approximation of the exponentially complex MD decoder. Zero capacity coalition attack for $t = 3$ was identified and it was shown that the MD decoder fails under this attack.

In Chapter 3 We considered the application of tree codes and sequential decoding for designing low complexity fingerprinting schemes against collusion attacks that are fair and memoryless. A cut-off rate was derived for $t = 2$, and the existence of a good code that can be efficiently decoded by sequential decoding was established. To validate our proposed method we presented numerical results using random Convolutional codes for $t = 2$.

A heuristic approach for $t = 3$ and the zero capacity attack was presented using a novel non-linear construction. Bidirectional sequential decoding was also presented to reach a small error probability and overcome the error floor.

In chapter 4 a new achievable rate region for the two-user X channel was established. The proposed scheme is based on a combination of the binning technique for broadcast channels and the message splitting for interference channels with joint decoding. The proposed method generalizes the scheme in [10] which was the best available rate region priori to our contribution and, to the best of our knowledge, achieves the largest region for the two user discrete memoryless X channel.

In Chapter 5 downlink communication for a cognitive cellular system was studied and a novel signaling scheme based on interference alignment, zero forcing and dirty paper coding was proposed. We also presented an outer bound and showed that our proposed scheme is optimal for some special cases. The benefits of the cognitive paradigm was also illustrated using the outer bound by proving the total sum DoF of the system is strictly larger than the case where cognitive message sharing is not available.

6.2 Future Directions

A natural extension of our work in designing low complexity collusion resistant fingerprinting schemes, would be to extend our results to coalition sizes of $t > 3$ under *general* coalition attacks. Extending our work on sequential decoding and tree codes using more sophisticated metrics to guide the search could be a promising approach in this direction. Moreover, coming up with new methods for analyzing the complexity of the sequential decoder instead of the common approach of dividing the tree into one correct path and considering the rest as incorrect paths is worth attention.

Our work on the X channel can be extended in several ways, including simplification of the achievable rate region, as well as considering networks involving more than 2 users. In addition, finding outer bounds on the capacity region and identifying special cases where the capacity region is achieved are of research interest as well.

Finally studying different cases of CSI availability at the base stations as well as extending the work to more than 2 cells with different scenarios of cognitive cooperation are possible extensions for our contribution in cognitive cellular networks.

APPENDIX A

PROOFS OF THE OUTER BOUND IN CHAPTER 5

In order to derive the bound L_3, first we assume that the secondary users fully cooperate. In this case we can "lump" the users $\mathcal{S}_j$'s into one user $\mathbb{S}$ with $K_S n_S$ antennas. Second, we also let the primary messages be given to $\mathcal{S}$ by a genie. The above assumptions cannot reduce the achievable DoF region. Consider the system with transmitter $\mathcal{P}, \mathcal{S}$ and receivers $\mathcal{P}_1, \mathbb{S}$. The signals received at $\mathcal{P}_1$ ans $\mathbb{S}$ are:

$$\begin{aligned} \mathbf{y}_{\mathcal{P}_1} &= \mathbf{H}_{\mathcal{P}_1}\mathbf{x}_{\mathcal{P}} + \mathbf{H}'_{\mathcal{P}_1}\mathbf{x}_{\mathcal{S}} + \mathbf{z}_{\mathcal{P}_1}, \\ \mathbf{y}_{\mathbb{S}} &= \mathbf{H}_{\mathbb{S}}\mathbf{x}_{\mathcal{S}} + \mathbf{H}'_{\mathbb{S}}\mathbf{x}_{P} + \mathbf{z}_{\mathbb{S}}, \end{aligned} \tag{A.1}$$

To establish the bound we follow the steps of [53]:

1) The noise at $\mathcal{P}_1$ is reduced by changing its covariance matrix to:

$$\mathbf{K}' = \mathbf{I}_{n_P} - \mathbf{H}'_{\mathcal{P}_1}(\mathbf{H}'^{T}_{\mathcal{P}_1}\mathbf{H}'_{\mathcal{P}_1})^{-1}\mathbf{H}'^{T}_{\mathcal{P}_1} + \alpha\mathbf{H}'_{\mathcal{P}_1}\mathbf{H}'^{T}_{\mathcal{P}_1} \tag{A.2}$$

where $\alpha = \min(\frac{1}{\sigma^2(\mathbf{H}'_{\mathcal{P}_1})}, \frac{1}{\sigma^2(\mathbf{H}_{\mathbb{S}})})$ in which $\sigma^2(.)$ denotes the maximum singular value of a matrix.

2) A genie provides $\mathbb{S}$ with $\mathbf{x}_{\mathcal{P}}$, and since $\mathbf{H}'_{\mathbb{S}}$ is known at $\mathbb{S}$, it can subtract $\mathbf{H}'_{\mathbb{S}}\mathbf{x}_{\mathcal{P}_1}$ from its signal and arrive at $\mathbf{y}'_{\mathbb{S}} = \mathbf{H}_{\mathbb{S}}\mathbf{x}_{\mathcal{S}} + \mathbf{z}_{\mathbb{S}}$.

3) $\mathcal{P}_1$ is assumed to be able to decode its message reliably thus it can also subtract

$\mathbf{H}_{\mathcal{P}_1}\mathbf{x}_{\mathcal{P}}$ and arrive at $\mathbf{y}'_{\mathcal{P}_1} = \mathbf{H}'_{\mathcal{P}_1}\mathbf{x}_{\mathcal{S}} + \mathbf{z}'_{\mathcal{P}_1}$.

4) Having reached the following equations:

$$\begin{aligned} \mathbf{y}'_{\mathcal{P}_1} &= \mathbf{H}'_{\mathcal{P}_1}\mathbf{x}_{\mathcal{S}} + \mathbf{z}'_{\mathcal{P}_1} \\ \mathbf{y}'_{\mathbb{S}} &= \mathbf{H}_{\mathbb{S}}\mathbf{x}_{\mathbb{S}} + \mathbf{z}_{\mathcal{S}} \end{aligned} \tag{A.3}$$

our aim is to show that if $\mathbb{S}$ can decode $\mathbf{x}_{\mathbb{S}}$; P_1 will be able to decode it as well and thus the sum DoF will be less than n_P. To see this consider the singular value decomposition $\mathbf{H}_{\mathbb{S}} = \mathbf{U}_{\mathbb{S}}\mathbf{\Lambda}_{\mathbb{S}}\mathbf{V}_{\mathbb{S}}$, by multiplying $\mathbf{y}'_{\mathbb{S}}$ in $\mathbf{V}_{\mathbb{S}}^T\mathbf{\Lambda}_{\mathbb{S}}^{-1}\mathbf{U}_{\mathbb{S}}^T$ we obtain a channel with input $\mathbf{x}_{\mathcal{S}}$ and a noise vector with variances $\frac{1}{\sigma^2(\mathbf{H}_{\mathbb{S}})}$.

5) $\mathbf{y}'_{\mathcal{P}_1}$ is multiplied in $\mathbf{T} = (\mathbf{H}_{\mathcal{P}_1}^T\mathbf{H}_{\mathcal{P}_1})^{-1}\mathbf{H}_{\mathcal{P}_1}$. The noise variance matrix with this operation will be $\mathbf{T}\mathbf{K}'\mathbf{T}^T$, which is straightforward to check is equal to a diagonal matrix with diagonal elements equal to α. Thus, the receiver $\mathcal{P}_1$ can be made less noisy than $\mathbb{S}$ and if $\mathbf{x}_{\mathcal{S}}$ is decode-able at $\mathbb{S}$ it must be decode-able, at $\mathcal{P}_1$ as well.

6) Using the bound on the degrees of freedom on the MIMO channels, we can conclude that still $d_{\mathcal{P}_1} + d_S \leq n_P$.

7) For the case where $n_P < m_S$ and the matrix ${\mathbf{H}'_{\mathcal{P}_1}}^T\mathbf{H}'_{\mathcal{P}_1}$ is not invertible, similar to the argument given in [53] we can add more antennas at $\mathcal{P}_1$ without reducing the DoF region and follow the steps above. In that case the sum DoF will be less than m_S.

8) Following the steps 1-7 we arrive at:

$$d_S + d_{\mathcal{P}_1} \leq \max(m_S, n_P)$$

BIBLIOGRAPHY

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, 1948.

[2] N. P. Anthapadmanabhan, A. Barg, and I. Dumer, "On the fingerprinting capacity under the marking assumption," 2008.

[3] P. Moulin, "Universal fingerprinting: capacity and random-coding exponents," 2008.

[4] A. Somekh-Baruch and N. Merhav, "On the capacity game of private fingerprinting systems under collusion attacks," *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 884 – 899, 2005.

[5] S. C. Lin, M. Shahmohammadi, and H. El Gamal, "Fingerprinting under minimum distance decoding," in *Proc. Allerton Conf. on Comm., Cont. and Comp.*, 2007.

[6] S. C. Lin, M. Shahmohammadi, and H. E. Gamal, "Fingerprinting under minimum distance decoding," *IEEE Trans. on Info. Foren. and Security*, vol. 4, no. 1, pp. 59 – 69, 2009.

[7] M. Shahmohammadi and H. E. Gamal, "Tree search for digital fingerprinting," in *Proc. Allerton Conf. on Comm., Cont. and Comp.*, 2008.

[8] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley, 1991.

[9] T. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, 1972.

[10] M. A. Maddah-Ali, A. S. Motahari, and A. K. Khandani, "Signaling over mimo multi-base systems: Combination of multi-access and broadcast schemes," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, pp. 2104 – 2108, 2006.

[11] M. K., "A coding theorem for the discrete memoryless broadcast channels," *IEEE Trans. Inf. Theory*, vol. 25, no. 3, pp. 306 – 311, 1979.

[12] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, pp. 49 – 60, January 1981.

[13] O. Koyluoglu, M. Shahmohammadi, and H. E. Gamal, "An achievable rate region for the discrete memoryless x channel," in *Proc. IEEE Int. Symp. on Info. Theory (ISIT)*, 2009.

[14] J. Mitola and G. Q. Maguire, "Cognitive radio: making software radios more personal," *IEEE personal communications*, vol. 6, no. 4, pp. 13 – 18, 1999.

[15] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 1813 – 1827, 2006.

[16] S. Gel'fand and M. Pinsker, "Coding for channels with random parameters," *Probl. Contr. and Inform. Theory*, 1980.

[17] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, pp. 439 – 441, May 1983.

[18] Q. Spencer, A. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser mimo channels," *IEEE Transactions on Signal Processing,*, vol. 52, no. 2, pp. 461 – 471, 2004.

[19] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Trans. on Telecomm*, 1999.

[20] V. R. Cadambe and S. Jafar, "Interference alignment, and the degrees of freedom of the k user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, pp. 3425 – 3441, August 2009.

[21] S. A. Jafar and S. Shamai, "Degrees of freedom region for the mimo x channel," *IEEE Trans. Inf. Theory*, vol. 54, pp. 151 – 170, January 2008.

[22] C. Suh, M. Ho, and D. Tse, "Downlink interference alignment," 2010.

[23] A. Jovicic and P. Viswanath, "Cognitive radio: An information-theoretic perspective," *IEEE Trans. on Info. Theory*, vol. 55, no. 9, pp. 3945 – 3958, 2009.

[24] W. Wu, S. Vishwanath, and A. Arapostathis, "Capacity of a class of cognitive radio channels: Interference channels with degraded message sets," *IEEE Trans. Info. Theory*, no. 11, pp. 4391 – 4399, 2007.

[25] C. Huang and S. A. Jafar, "Degrees of freedom of the mimo interference channel with cooperation and cognition," *IEEE Trans. Info. Theory*, vol. 55, no. 9, pp. 4211 – 4220, 2009.

[26] J. Sachs, I. Maric, and A. Goldsmith, "Cognitive cellular systems within the tv spectrum," in *IEEE Int. Symp. on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, pp. 1 – 12, 2010.

[27] M. D. R. M. Mueck and M. Debbah, "Opportunistic relaying for cognitive radio enhanced cellular networks: Infrastructure and initial results," in *IEEE ISWPC*, pp. 556 – 561, May 2010.

[28] H. Cheng and Y. Yao, "Cognitive-relay-based intercell interference cancellation in cellular systems," *IEEE Trans. on Vehicular Tech.*, vol. 59, pp. 1901 – 1909, May 2010.

[29] M. Shahmohammadi, O. Koyluoglu, T. Khattab, and H. E. Gamal, "Joint interference cancellation and dirty paper coding for cognitive cellular networks," in *Proc. of IEEE Wireless Communications and Networking Conference*, 2011.

[30] M. Shahmohammadi, O. Koyluoglu, T. Khattab, and H. E. Gamal, "On the degrees of freedom of the cognitive broadcast channel," *in Submission*, 2011.

[31] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1897 – 1905, September 1998.

[32] A. Barg, G. Blakley, and G. Kabatiansky, "Digital fingerprinting codes: Problem statements, constructions, identification of traitors," *IEEE Trans. Inform. Theory*, vol. 49, pp. 852 – 865, April 2003.

[33] A. D. Friedman, R. L. Graham, and J. D. Ullman, "Universal single transition time asynchronous state assignments," *IEEE Trans. Comput.*, vol. 18, June 1969.

[34] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 599 – 618, 2001.

[35] D. Divsalar, S. Dolinar, and C. Jones, "Low-rate ldpc codes with simple protograph structure," in *Proc. International Symp. on Info. Theory*, pp. 1622 – 1626, 2005.

[36] A. Barg and G. D. Forney, "Random codes: minimum distances and error exponents," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2568 – 2573, 2003.

[37] R. Gallager, *Low-Density Parity-Check Codes*. MIT Press, 1963.

[38] D. Divsalar, H. Jin, and R. McEliece, "Coding theorems for turbo-like codes," in *Proc. 36th Allerton Conf. on Communication, Control, and Computing*, 1998.

[39] H. D. Pfister and I. Sason, "Accumulate-repeat-accumulate codes: Capacity-achieving ensembles of systematic codes for the erasure channel with bounded complexity," *IEEE Trans. Inform. Theory*, vol. 53, no. 6, pp. 2088 – 2115, 2007.

[40] E. Arikan, "Sequential decoding for multiple access channels," *IEEE Trans. Inform. Theory*, vol. 34, pp. 246 – 259, March 1988.

[41] V. Balakirsky, "An upper bound on the distribution of computation of a sequential decoder for multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 42, March 1996.

[42] J. M. Wozencraft, "Sequential decoding for reliable communication," *1957 National IRE Convention Record*, 1957.

[43] R. M. Fano, "A heuristic discussion of probabilistic decoding," *IEEE Trans. Inf. Theory*, vol. 9, no. 2, pp. 64 – 74, 1963.

[44] F. Jelinek, "Fast sequential decoding algorithm using a stack," *IBM Journal of Research and Development*, 1969.

[45] R. G. Gallager, *Information Theory and Reliable Communication*. Wiley, New York, 1968.

[46] M. A. Maddah-Ali, A. S. Motahari, and A. K. Khandani, "Communication over mimo x channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3457 – 3470, 2008.

[47] T. M. Cover, "Comments on broadcast channels,," *IEEE Trans. Inf. Theory*, 1998.

[48] A. E. Gamal and E. van der Meulen, "A proof of marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 27, January 1981.

[49] G. Kramer, "Review of rate regions for interference channels," in *Proc. Int. Zurich Seminar,*, pp. 162 – 165, 2006.

[50] H.-F. Chong, M. Motani, H. K. Garg, and H. E. Gamal, "On the han-kobayashi region for the interference channel,," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3188–3195, 2008.

[51] C. Huang, S. A. Jafar, S. S. (Shitz), and S. Vishwanath, "On degrees of freedom region of mimo networks without csit," 2011.

[52] C. S. Vaze and M. K. Varanasi, "The degrees of freedom regions of mimo broadcast, interference, and cognitive radio channels with no csit," 2009.

[53] S. A. Jafar and M. Fakhereddin, "Degrees of freedom for the mimo interference channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2637 – 2642, 2007.