ON RESOURCE ALLOCATION FOR COMMUNICATION SYSTEMS WITH DELAY AND SECRECY CONSTRAINTS

A Dissertation

by

ANANTHARAMAN BALASUBRAMANIAN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2009

Major Subject: Electrical Engineering

ON RESOURCE ALLOCATION FOR COMMUNICATION SYSTEMS WITH DELAY AND SECRECY CONSTRAINTS

A Dissertation

by

ANANTHARAMAN BALASUBRAMANIAN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Scott L. Miller
Committee Members,	Costas N. Georghiades
	Krishna R. Narayanan
	Jim Ji
	Nancy Amato
Head of Department,	Costas N. Georghiades

December 2009

Major Subject: Electrical Engineering

ABSTRACT

On Resource Allocation for Communication Systems with
Delay and Secrecy Constraints. (December 2009)

Anantharaman Balasubramanian, B.E., College of Engineering, Chennai;
M.S., Indian Institute of Science, Bangalore
Chair of Advisory Committee: Dr. Scott L. Miller

This dissertation studies fundamental limits of modern digital communication systems in presence/absence of delay and secrecy constraints.

In the first part of this dissertation, we consider a typical time-division wireless communication system wherein the channel strengths of the wireless users vary with time with a power constraint at the base station and which is not subject to any delay constraint. The objective is to allocate resources to the wireless users in an equitable manner so as to achieve a specific throughput. This problem has been looked at in different ways by previous researchers. We address this problem by developing a systematic way of designing scheduling schemes that can achieve any point on the boundary of the rate region. This allows us to map a desired throughput to a specific scheduling scheme which can then be used to service the wireless users. We then propose a simple scheme by which users can cooperate and then show that a cooperative scheduling scheme enlarges the achievable rate region. A simple iterative algorithm is proposed to find the resource allocation parameters and the scheduling scheme for the cooperative system.

In the second part of the dissertation, a downlink time-division wireless system that is subject to a delay constraint is studied, and the rate region and optimal scheduling schemes are derived. The result of this study concludes that the achievable throughput of users decrease as the delay constraint is increased. Next, we consider a problem motivated by cognitive radio applications which has been proposed as a means to implement efficient reuse of the licensed spectrum. Previous research on this topic has focussed largely on obtaining fundamental limits on achievable throughput from a physical layer perspective. In this dissertation, we study the impact of imposing Quality of Service constraints (QoS) on the achievable throughput of users. The result of this study gives insights on how the cognitive radio system needs to be operated in the low and high QoS constraint regime.

Finally, the third part of this dissertation is motivated by the need for communicating information not only reliably, but also in a secure manner. To this end, we study a source coding problem, wherein multiple sources needs to be communicated to a receiver with the stipulation that there is no direct channel from the transmitter to the receiver. However, there are many "agents" that can help carry the information from the transmitter to the receiver. Depending on the reliability that the transmitter has on each of the agents, information is securely encoded by the transmitter and given to the agents, which will be subsequently given to the receiver. We study the overhead that the transmitter has to incur for transmitting the information to the receiver with the desired level of secrecy. The rate region for this problem is found and simple achievable schemes are proposed. The main result is that, separate secure coding of sources is optimal for achieving the sum-rate point for the general case of the problem and the rate region for simple case of this problem.

To my Parents

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my adviser, Prof. S. L. Miller, for his invaluable guidance and encouragement throughout my stay here. I will always be indebted to him for all his technical advice and ideas that have enabled me to complete this dissertation.

I also take this opportunity to thank the committee members, Prof. Costas N. Georghiades, Prof. Krishna Narayanan, Prof. Jim Ji and Prof. Nancy Amato, who have always been supportive of my work. I thank them for their constructive comments and advice on my research.

I would like to thank the professors from the Wireless Communication group for their excellent teaching. Special thanks to Prof. Tie Liu and Prof. Chamberland for the group meetings which were very helpful in my research. I owe a lot to Dr. Lingjia Liu, Prof. Tie Liu and Prof. J. F. Chamberland for their help with my research.

I would like to thank my friends, Lingjia Liu, Janath, Dilani, Makesh, Hung Ly, Jae Won, Jing Jing, Jing Jiang, Parimal, Abdallah, Wei-Yu, Kapil, Mothi, Karthik, Nitin, Mustafa, Salim, Fang and Hak, for fruitful discussions and help during my stay in College Station. Special thanks to my apartment-mate Bhanu for his great friendship.

I thank Harini, Primo, Meenu, Venky, Julia and Sriram for their love and support.

Finally, no words can suffice to thank Amma, Appa for their endless love, encouragement and for keeping me going.

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

FIGURE Page

CHAPTER I

INTRODUCTION

The study of information theory for point-to-point communication systems began with Shannon's seminal paper [1]. Since then considerable research has taken place for communication systems with one transmitter and receiver, and more or less a full understanding has been developed. The late 1990's witnessed a surge of research activities for communication systems with multiple transmit and receive antennas (i.e., MIMO systems). An important fundamental result is the characterization of the diversity-multiplexing trade off [2], indicating enhanced performance of MIMO systems over single antenna systems, especially in providing reliability/capacity. There are still some interesting ongoing research activities in understanding the theory of MIMO systems.

Not surprisingly, the gains that MIMO systems provided, captured the attention of the wireless communication industry. There have been many success in putting the theory of MIMO systems into practice, and it is important to note the contribution of [3] which is used in IEEE 802.16, WiMax wireless standards. However there has been a strong feeling that providing more than two antennas at the mobile terminals is impractical, especially given the vogue of having mobile phones as small as possible. This practical problem motivated the need of having a "close to MIMO" performance by making use of one antenna, that the mobile is typically endowed with.

The idea of user cooperation was born [4, 5]. It is based on a simple idea that the "stronger user" cooperates with a "weaker user" so that the former sacrifices some of its own throughput to help the latter, in such a manner that the system resources

The journal model is *IEEE Transactions on Automatic Control.*

are shared equitably. It is easy to see that when the inter-user channel (i.e., the channel between the cooperating users) become strong, then this reduces roughly to a MIMO system. Though it is not possible to reap all the gains of a MIMO system, cooperation does help to improve the individual throughputs (or reliability, depending on what the system is intended for) as compared to single-user systems. A part of this dissertation studies achievable gains of a downlink cooperative system. Some notable results in user cooperation include the study of diversity aspects of various cooperation protocols by Laneman [6],[7] and achievable gains of a uplink cooperative system in [8].

Since Shannon's seminal paper [1], researchers have always measured the efficiency of a communication system solely based on the throughput that it can provide. This point of view may not be always correct. For example, consider the familiar opportunistic scheme which has been proposed recently in [9] for increasing the system throughput. The primary idea in this scheme is to offer service to the user with the best channel. The problem with this approach is that users with strong channel continue to get higher throughput, while users with weak channel is deprived of any resources (i.e., opportunistic scheme maximizes the system throughput, but not the individual user's throughput). Hence a more reasonable method of measuring the system performance is by taking into account the delay associated in achieving a specific throughput. There has been an increase in the recent past in researchers embracing this point of view for measuring the system performance. One of the first results in this area was the power-delay tradeoff of Randall Berry [10]. Another tool to study the interplay of delay and throughput is the effective capacity, which is akin to the well developed concept of effective bandwidth. This tool was developed and first put to use for studying point-to-point systems in [11], MIMO systems in [12] and uplink cooperative systems in [13]. In this dissertation, we address the effect of delay constraints on the throughput of a two user downlink scheduliing system and a cognitive radio system.

Designing a good communication system not only involves techniques for communicating reliably but also securely. It should be noted that there are several algorithms that have been proposed for secure communication in the literature, most notably the Rivest, Shamir and Adleman algorithm (also called the RSA algorithm). But these algorithms work in the network layer. In this thesis we address the security issue from a physical layer perspective. The problem of physical layer security was first formulated by Shannon [14]. In particular the problem addressed in [14] was the following: What is the minimum external randomness (i.e., key bits) required in order to communicate information bits of rate R? Shannon showed that minimum rate of key bits for providing perfect secrecy should be at least the rate of the information bits. Shannon's work was followed by the remarkable work of Wyner [15] and Cheong [16]. In Wyner's model, there is assumed to be a legitimate receiver and an eavesdropper, with the eavesdropper's channel assumed to be a degraded version of the legitimate receiver's channel. It is to be noted that in Wyner's model there are *no* key bits or external randomness assumed to be present. The problem that Wyner addressed was the following: Given the equivocation rate of the wiretapper (i.e., degree of confusability), what is the maximum rate at which communication can take place securely with the legitimate receiver? Wyner showed that for perfect secrecy, the maximum secure rate that is possible is, the difference in channel capacities of the legitimate receiver and the wiretapper. There has been a burst of research activity in the area of secure communications recently. Most notable results are the secrecy capacity of the fading channels [17], multiple access channels [18], [19], broadcast channels [20], MIMO channels [21], [22] and interference channels [23].

A. Thesis Outline and Contributions

Chapter II gives the necessary background for the material discussed in the ensuing chapters. Chapter III begins by finding optimal scheduling scheme for a simple two user downlink time division wireless system. Further, it is shown that opportunistic scheduling schemes fall out as a special case of a more generalized class of optimal scheduling schemes that have been obtained. For the case when the users are allowed to cooperate, it is shown that the optimal cooperative scheduling schemes belong to the same class as that of non-cooperative (i.e., downlink) scheduling schemes. Finally, the achievable gains of a cooperative scheduling system is addressed. The primary contribution in this chapter is finding the optimal policy that one has to use for getting any point on the rate region both for the non-cooperative and the cooperative case [24], aside from establishing the rate region for the cooperative scenario. Further, for the cooperative case we also have a simple algorithm that would determine the optimal resource allocation parameters which will maximize the throughput achieved by the users [25].

In Chapter IV, we generalize the optimal scheduling scheme derived in Chapter III for the case when the wireless users have a Quality of Service (QoS) constraint. We discuss the fundamental limits on the throughput that users can achieve if a QoS constraint is imposed. The effect of coherence time on the admissible throughput of users are studied. The primary contribution in this chapter is finding the exact scheduling scheme and the rate region for a two-user problem with QoS constraints [26]. Further, for the multi-user problem, we have proposed a scheme that gives considerable gains over the opportunistic scheme for maximizing the sum-rate of effective capacity [27].

In Chapter V, we address the problem of finding the admissible rate region of a cognitive radio system with QoS constraints. This is done by modeling a cognitive radio system as a two-user priority queuing system with constant input arrival rates (which is to be determined) and variable service rates determined by the wireless channel. The primary contribution in this chapter is to model the cognitive radio system that permits analysis and compare the achievable throughput attained by users for the different models that we have proposed.

Chapter VI addresses the secrecy capacity of a source coding problem. More specifically the problem that is addressed in Chapter VI is the following: Multiple sources are to be communicated to $(L + 1)$ receivers amidst an eavesdropper. A direct channel is available from the source encoder to each of the L receivers, while the $(L + 1)^{th}$ receiver has access to the output of L channels (i.e., there is no direct channel from the source encoder to the $(L+1)^{th}$ receiver). The eavesdropper has access to a subset of the m channels $(m < L)$ and the legitimate receiver must be able to reconstruct the source from a subset of n channels, where $m < n \leq L$. The goal of the source encoder is to encode the source in such a way so as to communicate the source reliably to the $(L + 1)^{th}$ receiver with minimum overhead. The rate region of such a system is found and achievability is established using simple linear network codes. The primary contribution in this chapter is to prove that mixing a number of i.i.d sources (that is, having multiple i.i.d sources) will not help increase the randomness of the system and that one could approach the task of providing secrecy to multiple sources by providing secrecy to each source separately [28], which is shown to be optimal.

Finally, we conclude this dissertation with scope for future work in Chapter VII.

CHAPTER II

BACKGROUND

In this chapter, we provide the necessary theoretical background that is required for later chapters in this dissertation. First, we provide an overview of cooperative systems, wherein we discuss the different types of protocols in the literature. This will provide a good background for the the material in Chapter III. Next, we briefly go over the concepts of effective bandwidth and effective capacity which are required in Chapter IV and Chapter V. Finally, we discuss the main principle of information theoretic secrecy which is used in Chapter VI.

A. Overview of Cooperative Systems

Consider a system shown in Fig. 1, where there is a source 'S' which employs terminal 'R'(i.e., relay) in transmitting to the destination 'D'(i.e., destination). As in [6], we consider N consecutive uses (where N is large) of a baseband equivalent, discrete time channel model.

For direct transmission (in the absence of the relay 'R'), the channel model is:

$$y_d[n] = a_{s,d} x_s[n] + z_d[n] \tag{2.1}$$

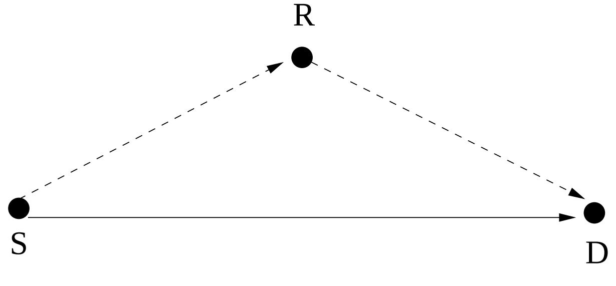


Fig. 1. Cooperative system model.

where $x_s[n]$ is the source transmitted signal, $y_d[n]$ is the destination received signal and $n = 1, 2, \ldots, N$. In the case of cooperative transmission, the channel model during the first half of the block is modeled as:

$$\begin{aligned} y_r[n] &= a_{s,r}x_s[n] + z_r[n] && (2.2) \\ y_d[n] &= a_{s,d}x_s[n] + z_d[n] && (2.3) \end{aligned}$$

for $n = 1, 2, \ldots, N/2$, where $y_r[n]$ and $y_d[n]$ are the relay and destination received signals respectively. For the second half of the block, the received signal is modeled as:

$$y_d[n] = a_{r,d}x_r[n] + z_d[n] \tag{2.4}$$

for $n = \frac{N}{2} + 1, \ldots, N$, where $x_r[n]$ is the relay transmitted signal. In (2.1)-(2.4), $a_{s,r}$, $a_{r,d}$ and $a_{s,d}$ denotes the effect of path-loss, shadowing and frequency non-selective fading and $z_r[n]$, $z_d[n]$ denotes the effects of receiver noise modeled as zero-mean mutually independent circularly symmetric, complex gaussian sequences with variance N_o.

Cooperative protocols are classified based on the types of processing employed at the relay terminals and are discussed along similar lines to the ones in [6] as follows.

1. Amplify-Forward

The system model for this protocol is given in (2.2)-(2.4). The source terminal transmits $x_s[n]$, say, for $n = 1, 2, \ldots, N/2$. During this interval, the relay processes $y_r[n]$, and relays the information by transmitting

$$x_r[n] = \beta\, y_r[n - N/2]. \tag{2.5}$$

for $n = \frac{N}{2}+1,\ldots,N$, and β is the amplifying parameter. In order that the relay remain within it power constraint, it follows that the amplifying parameter β needs to satisfy

$$\beta \leq \sqrt{\frac{P}{|a_{s,r}|^2\,P+N_o}} \tag{2.6}$$

This scheme can be thought of repetition coding being performed by the relay, except that the relay amplifies the received noise as well. The destination decodes the signal from the source and the relay by using a maximal ratio combiner. The maximum mutual information between the inputs and outputs can be shown to be [6]

$$I_{AF} = \frac{1}{2}\log\left(1+\mathtt{SNR}|a_{s,d}|^2 + f\left(\mathtt{SNR}|a_{s,r}|^2, \mathtt{SNR}|a_{r,d}|^2\right)\right) \tag{2.7}$$

where $\mathtt{SNR} := \frac{P}{N_o}$ and

$$f(x,y) = \frac{xy}{x+y+1}$$

The outage event for achieving a spectral efficiency $\mathtt{R}$ is given by the event

$$\Pr[I_{AF} < \mathtt{R}] \tag{2.8}$$

which is equivalent to the event

$$|a_{s,d}|^2 + \frac{1}{\mathtt{SNR}} f\left(\mathtt{SNR}|a_{s,r}|^2, \mathtt{SNR}|a_{r,d}|^2\right) < \frac{2^{2\mathtt{R}}-1}{\mathtt{SNR}} \tag{2.9}$$

Denoting the event in (2.8) by $P_{AF}^{out}(\mathtt{SNR},\mathtt{R})$, it has been shown in [6] that

$$P_{AF}^{out}(\mathtt{SNR},\mathtt{R}) \approx C\left(\frac{2^{2\mathtt{R}}-1}{\mathtt{SNR}}\right)^2$$

where C is as constant that depends on the average channel gain parameters. We note that this scheme achieves a full diversity gain of two.

2. Decode-and-Forward

The channel model for this scheme is given in (2.2)-(2.4). The source transmits $x_s[n]$ for $n = 1, 2, \ldots, N/2$, and the relay processes $y_r[n]$ by forming an estimate $\hat{x}_s[n]$ of the source transmitted signal, $x_s[n]$. The relay transmits the signal

$$x_r[n] = \hat{x}_s[n - N/2]$$

for $n = \frac{N}{2}+1, \ldots, N$. The destination employs a matched filter to combine the signals received from the source and the relay. The maximum mutual information for this scheme has been shown in [6] to be

$$I_{DF} = \frac{1}{2} \min \left\{ \log(1 + \mathtt{SNR}|a_{s,r}|^2), \log(1 + \mathtt{SNR}|a_{s,d}|^2 + \mathtt{SNR}|a_{r,d}|^2) \right\} \tag{2.10}$$

The first term in (2.10) represents the maximum rate at which the relay can decode the source, while the second term denotes the reliable rate at which the destination can decode the source from the source and the relay. Requiring both the relays and destination to decode the entire codeword without error results in the minimum of the two terms as in (2.10)[6]. The outage event for a spectral efficiency $\mathtt{R}$ can be obtained from (2.10), and is equivalent to the event that [6]

$$\min \left\{ |a_{s,r}|^2, |a_{s,d}|^2 + |a_{r,d}|^2 \right\} < \frac{2^{2\mathtt{R}} - 1}{\mathtt{SNR}} \tag{2.11}$$

Note that (2.11) follows from (2.10) due to high $\mathtt{SNR}$ approximation. Denoting the outage event by $P_{DF}^{out}(\mathtt{SNR}, \mathtt{R})$, it has been shown in [6] that

$$P_{DF}^{out}(\mathtt{SNR}, \mathtt{R}) \approx \tilde{C} \left(\frac{2^{2\mathtt{R}} - 1}{\mathtt{SNR}} \right)$$

where $\tilde{C}$ is a constant which depends on the average channel gain parameters. It is important to note the $\frac{1}{\mathtt{SNR}}$ decay in the above equation denotes that the decode-and-forward protocol does not offer diversity gains.

3. Selection Relaying

This scheme takes advantage of the fact that the channel coefficients can be tracked by the cooperating terminals (i.e., the relay and the destination) perfectly. More specifically, if the measured $|a_{s,r}|^2$ falls below a certain threshold, the source simply continues its transmission to the destination, in the form of repetition or more powerful codes [6]. If the measured $|a_{s,r}|^2$ lies above a threshold, the relay forwards what it received from the source, using either amplify-forward or decode-and-forward, in an attempt to achieve diversity gain [6].

Consider the performance of selection decode-and-forward scheme. The mutual information of this scheme is given by [6]:

$$I_{SDF} = \begin{cases} \frac{1}{2}\log\left(1 + 2\mathtt{SNR}|a_{s,d}|^2\right), & |a_{s,r}|^2 < \frac{2^{2\mathtt{R}}-1}{\mathtt{SNR}} \\ \frac{1}{2}\log\left(1 + \mathtt{SNR}[|a_{s,d}|^2 + |a_{r,d}|^2]\right), & |a_{s,r}|^2 \geq \frac{2^{2\mathtt{R}}-1}{\mathtt{SNR}} \end{cases} \tag{2.12}$$

The first case in (2.12) corresponds to the relay not being able to decode and the source repeating its transmission. The second case corresponds to the relay being able to decode and repeat the source transmission (i.e., in which case one can coherently combine the signals from the source and the relay). It has been shown in [6] that the outage probability decays like $\frac{1}{\mathtt{SNR}^2}$, which implies a diversity gain of two.

4. Incremental Relaying

In this scheme, the source transmits its information to the destination at spectral efficiency $\mathtt{R}$. The destination indicates success or failure by broadcasting a single bit of feedback to the source and the relay. For high enough $|a_{s,d}|^2$, there is a high probability that the destination will be able to decode the source signal in which case the relay does not re-transmit. If the channel gain, $|a_{s,d}|^2$, is not high enough (i.e., the source-destination transmission is not successful), then the relay amplifies and forwards what it received from the source, and the destination combines the two transmissions (i.e., one from the source and the other from the relay). It has been shown in [6], that the outage probability decays like $\frac{1}{\mathtt{SNR}^2}$, which suggests that this scheme reaps all the diversity gains.

B. Effective Bandwidths, Effective Capacities and Stochastic Differential Equations

1. Effective Bandwidth

Future communication networks are expected to integrate a large number of traffic sources with a wide variety of traffic characteristics while still providing some guaranteed quality of service (QoS) such as packet loss probability. Typically traffic streams transmit data at different rates where the rate may vary between zero and some peak rate. A natural question to ask for such a system is the following: Given that different data streams have different rates, how do we decide on whether to admit, or, not to admit a stream, if we need to guarantee a QoS? The effective bandwidth turns out to be a benchmark to answer the aforementioned question.

The effective bandwidth of a traffic stream is chosen as the smallest capacity (i.e., server rate) that solves the admission criterion. To view this in another way, the effective bandwidth models the stochastic behavior of a source traffic process in an asymptotic sense . In most cases, the admission criterion is taken as the allowed buffer overflow probability (as will be explained in the sequel). One of the most attractive property of using effective bandwidth in practical systems is that the sum of the effective bandwidth of two independent traffic sources is equal to the effective bandwidth of their superposition (i.e., additive), which makes calculating the admission criterion

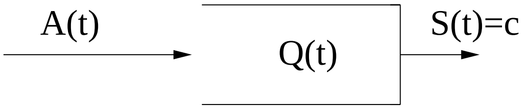


Fig. 2. Infinite length buffer with a variable arrival rate and constant service rate c.

simple.

Consider an arrival process $\{A(t), t \geq 0\}$, where $A(t)$ is the total amount of traffic generated by a source over time $[0, t]$. Consider a stochastic process $\{Z(t), t > 0\}$ that models the traffic flow. Let $r\big(Z(t)\big)$ be the rate at which the traffic flows at time t. Then

$$A(t) = \int_0^t r\big(Z(u)\big)\, du \tag{2.13}$$

Define the asymptotic log moment generating function as

$$h(v) = \lim_{t\to\infty} \frac{1}{t} \log E\left\{\exp(v\, A(t))\right\}$$

The effective bandwidth of the traffic is defined as

$$eb(v) = \frac{h(v)}{v} \tag{2.14}$$

The following properties hold

- $r^{mean} \leq eb(v) \leq r^{peak}$

- $\lim_{v\to 0} eb(v) = r^{mean}$ and $\lim_{v\to\infty} eb(v) = r^{peak}$

where $r^{mean} = E[r\big(Z(\infty)\big)]$ is the mean traffic flow rate, and, $r^{peak} = \sup_z\{r(Z)\}$ is the peak traffic flow rate.

Consider a queue of infinite buffer size served by a channel with service rate $S(t) = c$ (see Fig. 2), such as an AWGN channel. Due to the varying arrival rate $A(t)$, the queue length $Q(t)$ varies with time. Using the theory of large deviations, it can be shown that the queue length $Q(t)$ exceeding a threshold B satisfies

$$\lim_{t\to\infty} \Pr\{Q(t) \geq B\} \approx e^{-\theta B} \quad \text{as } B \to \infty \tag{2.15}$$

The exponent of the buffer overflow probability (i.e., the decay rate) is the unique solution to $eb(\theta) = c$, where $eb(\cdot)$ is defined in (2.14). For a desired exponent θ, the effective bandwidth yields the minimum service rate c, such that the behavior in (2.15) holds.

Example II.1. *Assume an on-off source with rate of transition from 'on' to 'off' denoted by λ and from 'off' to 'on' by μ. Traffic is generated at rate c, when the source is on and no traffic is generated when the source is off. Letting $e(M)$ denote the largest real eigen-value of a square matrix M. It has been shown in [29], that the asymptotic log moment generating function for the on-off process with the aforementioned parameters is*

$$h(\upsilon) = e(Q + \upsilon R)$$

where

$$Q = \begin{bmatrix} -\lambda & \lambda \\ \mu & \mu \end{bmatrix}, \qquad R = \begin{bmatrix} 0 & 0 \\ 0 & c \end{bmatrix} \tag{2.16}$$

The highest eigen-value of the matrix $Q + \upsilon R$ turns out to be

$$\frac{c\upsilon - \mu - \lambda + \sqrt{(c\upsilon - \mu - \lambda)^2 + 4\lambda c\upsilon}}{2}$$

and hence the effective bandwidth of an on-off source works out to be (from (2.14))

$$\frac{c\upsilon - \mu - \lambda + \sqrt{(c\upsilon - \mu - \lambda)^2 + 4\lambda c\upsilon}}{2\upsilon}$$

2. Effective Capacity

It is to be noted that (2.15) is derived by assuming that the service rate is fixed, while the input arrival rates are stochastic (which is why effective bandwidth characterizes the stochastic behavior of source traffic process). One natural question to ask is the following: Suppose if the input arrival rates are constant, while the output service rate is stochastic (as shown in Fig. 3), can we get a asymptotic behavior similar to (2.15). It turns out that it is possible and this is shown in [11]. The effective capacity of a traffic stream is the maximum constant input arrival rate for a given stochastic service rate that solves the admission criterion, which is why effective capacity is the dual of effective bandwidth.

Consider a standard point-to-point communication system where the source is endowed with a infinite length buffer and that the traffic flows into the buffer with constant arrival rate $A(t) = c$, with variable service rate $S(t)$, at time t. Note that the variability in service rate $S(t)$ can arise due to the effects such as shadowing, small-scale fading etc. Due the variable service rate, the queue length $Q(t)$ is also a variable. Similar to effective bandwidth, one can show that (2.15) is satisfied, where θ is the unique solution to $E_C(\theta) = c$, and

$$E_C(\upsilon) = \lim_{t\to\infty} \frac{1}{\upsilon\, t} \log \mathrm{E}\left[e^{-\upsilon\, S(t)}\right]$$

It is possible to show the following:

- $\lim_{\upsilon\to 0} E_C(\upsilon) = \mathrm{E}[S(t)]$,
- $\lim_{\upsilon\to\infty} E_C(\upsilon) = 0$.

The effective capacity can be intuitively explained as follows. Given system parameters and an exponential decay rate θ, the effective capacity is the maximum arrival

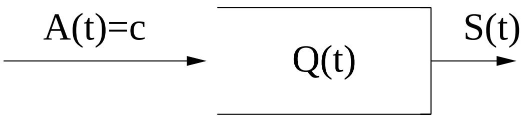


Fig. 3. Infinite length buffer with a constant arrival rate c and variable service rate.

rate for which the QoS requirement in (2.15) is fulfilled. The effective capacity of a system with constant arrival rate c, with an on-off service process is given by [11]

$$\frac{c\,\theta + \mu + \lambda - \sqrt{(c\,\theta + \mu + \lambda)^2 - 4\lambda\,c\,\theta}}{2\,\theta}$$

3. Stochastic Differential Equation

A stochastic differential equation (SDE) is one in which one or more terms is a stochastic process which results in the solution being a stochastic process as well. Here we discuss a SDE in which the random fluctuations are characterized by a jump process. The discussion presented here proceeds along similar lines to [30].

Consider a SDE excited by a jump process:

$$x(t) = x(0) + \int\limits_0^t f\big(x(\sigma), \sigma\big)\; d\sigma + \int\limits_0^t g\big(x(\sigma), \sigma\big)\; dN_\sigma \tag{2.17}$$

where N is a poisson counter. A function $x(\cdot)$ is a solution of (2.17) in the It$\hat{o}$ sense if, on an interval where N is constant, x satisfies $\dot{x} = f(x,t)$ and if, N jumps at t_1, x behaves in a neighborhood of t according to the rule [30]

$$\lim_{\substack{t\to t_1\\ t>t_1}} x(t) = g\left(\lim_{\substack{t\to t_1\\ t<t_1}} x(t), t_1\right) + \lim_{\substack{t\to t_1\\ t<t_1}} x(t) \tag{2.18}$$

It is common to write (2.17) as

$$dx = f(x,t)\; dt + g(x,t)\; dN$$

Consider a more general SDE

$$dx = f(x)\, dt + \sum_{i=1}^{m} g_i(x)\, dN_i; \quad x \in \mathbb{R}^n \tag{2.19}$$

where N_i are poisson counters of rate λ_i. Let $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ be a given function. It is possible to show that [30]

$$d\psi(t,x) = \frac{\partial \psi}{\partial t}\, dt + \langle \frac{\partial \psi}{\partial x}, f(x)\rangle dt + \sum_{i=1}^{n} [\psi\big(t, x + g_i(x)\big) - \psi(t,x)]\, dN_i \tag{2.20}$$

(2.20) is called the Itô rule for jump processes.

Example II.2. *Consider the following [30]:*

$$dx(t) = -x(t)\, dt + dN_1(t) - dN_2(t)$$

For the case when $\psi(t,x) = x^2$, one calculates $dx^2(t)$ using (2.20) to be

$$\begin{aligned} dx^2(t) &= -2x^2(t) + \big[(x(t)+1)^2 - x^2(t)\big]dN_1 + \big[(x(t)-1)^2 - x^2(t)\big]dN_2, \\ &= -2x^2(t) + \big(2x(t)+1\big)dN_1 + \big(1-2x(t)\big)dN_2. \end{aligned} \tag{2.21}$$

The expectation of (2.19) can be obtained from [30] to be:

$$\frac{d}{dt}\mathcal{E}x(t) = \mathcal{E}f\big(x(t),t\big) + \sum_{i=1}^{m} \big(\mathcal{E}g_i\big(x(t),t\big)\big) \tag{2.22}$$

Example II.3. *Consider the following [30]:*

$$dx(t) = -x(t)\, dt + dN_1 - dN_2$$

Using (2.22)

$$\frac{d}{dt}\mathcal{E}x = -\mathcal{E}x + \lambda_1 - \lambda_2$$

Further, from the Itô rule for x^2 (discussed in (2.21)) together with (2.22) yields

$$\frac{d}{dt}\mathcal{E}x^2 = -2\mathcal{E}x^2 + \mathcal{E}\big(2x(t)+1\big)\lambda_1 + \mathcal{E}\big(1-2x(t)\big)\lambda_2.$$

C. Information Theoretic Secrecy

We provide the primary idea of the Wyner's wiretap problem treated in [15] and [16]. The problem considered in shown in Fig. 4. The objective for the transmitter is to maximize the rate at which it can perform reliable communication with the legitimate receiver, with the additional constraint that the *eavesdropper be kept as ignorant as possible.* The important point to note in this model is that *the transmitter does not use any key or any external randomness to confuse the eavesdropper.* Assume that the transmitter needs to send the message W to the legitimate receiver. Hence, the transmitter does the following:

- The message W is encoded according to some mapping rule and sends out the codeword X into the channel.

- Let Y and Z be the signals received at the legitimate receiver and the eavesdropper respectively given by:

$$Y = X + n_1 \quad \text{and} \quad Z = X + n_2$$

where n_1, n_2 are the AWGN at the receivers with power N_1, N_2 respectively. Furthermore, it is assumed that the eavesdropper's channel is a degraded version of the legitimate receiver's channel (i.e., $N_2 > N_1$). The secrecy is measured by equivocation rate defined to be $H(W|Z)$ and it is desired that $H(W|Z) = H(W)$ for perfect secrecy. i.e., that is, the eavesdropper cannot make anything about the transmitted message 'W' even after it observes its received signal 'Z'.

The main result of secrecy capacity is as follows: The maximum rate at which the transmitter can send information to the legitimate receiver by keeping the *eavesdropper as ignorant as possible* is given by the difference of the capacities of the legitimate

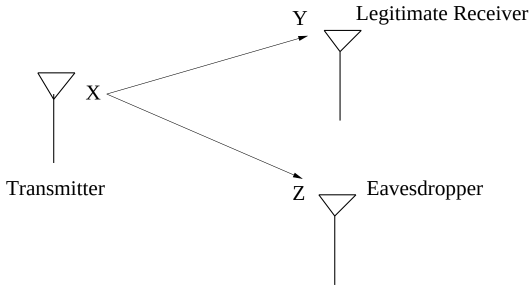


Fig. 4. A point to point communication model with an eavesdropper.

receiver's channel and the eavesdropper's channel. Denoting C_s to be the secrecy capacity, we then have:

$$C_s = \log\big(1 + \frac{P}{N_1}\big) - \log\big(1 + \frac{P}{N_2}\big) \tag{2.23}$$

We will give an intuitive explanation as to why (2.23) should hold. This will be done by using the tool of the joint typicality. Consider for the moment we do not have the eavesdropper. Then, this system becomes a point-to-point channel. We know very well for the point-to-point case that, if the codeword X is transmitted into an AWGN channel, then by joint typicality, the receiver will be able to decode to a codeword with probability one, if the size of the code book $\mathcal{C}$ has roughly $2^{nI(X;Y)}$ codewords. (that is, $|\mathcal{C}| \approx 2^{nI(X;Y)}$). Hence the rate of the message is then:

$$R = \frac{1}{n}\log_2|\mathcal{C}| = I(X;Y)$$

Let us now consider the case where there is an eavesdropper present as in Fig. 4. The main task here is to confuse the eavesdropper. In order to do this, we do the binning of codewords as shown in Fig. 5. In each bin we place $2^{I(X;Z)}$ codewords and the total number of codewords in all the bins is $2^{I(X;Y)}$. That is, in each bin there are $\frac{2^{I(X;Y)}}{2^{I(X;Z)}}$ codewords. The encoding for secure communication is done as follows:

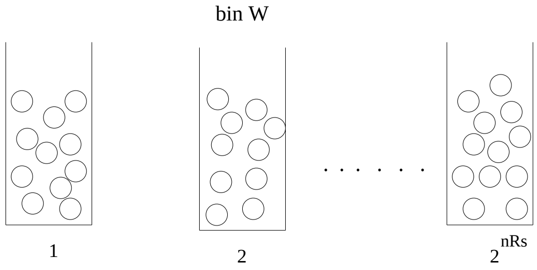


Fig. 5. The idea of binning in obtaining secure communication.

- For a given message W (for example, say '2'), the transmitter randomly chooses one of the codewords from bin W (as shown in Fig. 5) and sends it (say X) into the channel.

- The legitimate receiver receives Y and the eavesdropper receives Z.

- Both the legitimate and the eavesdropper does joint typical decoding and the *same codebooks are available with both of them.* With probability one, the eavesdropper finds one codeword in each bin, as the number of codewords in each bin is $2^{I(X;Z)}$. So the eavesdropper would be thoroughly confused as to which message was being transmitted as he finds one codeword decodable from each of the bins. But note that the legitimate receiver would decode only one codeword, as the number of codewords in all the bins is $2^{I(X;Y)}$ and with probability one he would decode to only *one* bin as the message that was sent.

- The important point to note is that secrecy does *not* imply *non-decodability.* In fact, the eavesdropper can decode codewords, but the catch is that he can decode *too many* codewords and this causes the confusion for him.

Now it is clear that that the number of messages that the transmitter can reliably send to the legitimate receiver by keeping the eavesdropper as ignorant as possible is:

$$\begin{aligned} R_s &= \frac{1}{n}\log_2\left(\frac{2^{nI(X;Y)}}{2^{nI(X;Z)}}\right) \\ &= I(X;Y) - I(X;Z) \end{aligned}$$

which yields the result in (2.23).

CHAPTER III

THE ACHIEVABLE RATE REGION OF A COOPERATIVE SCHEDULING SYSTEM

In this chapter, we derive optimal scheduling schemes that can achieve any point on the boundary of the capacity region for a wireless downlink system with a base station (BS) sending independent data to the mobile users. It is assumed that the wireless system is operated on a time division (TD) basis, wherein only one user is serviced at a time. We also show that a simple linear scheduling scheme performs very close to the optimal scheme. Furthermore, we characterize the optimal scheduling scheme for the case where the remote mobile users can perform downlink cooperation and find the achievable rate region for that case. Finally, a simple iterative algorithm is proposed for finding the resource allocation parameters and the scheduling scheme for the cooperative scheduling system.

A. Introduction

An inherent property of a wireless system is the time-variation of channel strength among the users due to factors such as shadowing and multipath fading in the received signal power. These factors can be exploited to increase the system throughput by opportunistically allocating more system resources to the user with the best channel [31], when the channel knowledge is available at the transmitter. Though opportunistic schemes yield higher overall system throughput, the individual throughput of various users can be vastly different, resulting in unfairness among the users. To clarify this further, consider a simple example. Assume, there is a base-station (BS) which serves two users (i.e., a downlink system) according to a TD model. It is also assumed that the BS can perfectly keep track of the channel gains of the wireless users. If the objective is to maximize the overall system throughput, the BS would choose to send information to the user with the best channel, which is more likely to be the user closer to the BS. Under this scheduling scheme, the user closer to the BS will have much higher probability of accessing the channel. Hence, on the average, the user which is farther away from the BS seldom gets service, which results in unfairness between the two users.

In order to answer the question of what a "fair scheduling" scheme is, one needs to define what the term "fair scheduling" scheme means. Usually, each user has its own minimum rate requirement and in the context of this chapter, a "fair scheduling" scheme is one that satisfies the minimum rate requirement for each user. The answer to the above question then requires the complete characterization of the capacity region of the system and the corresponding scheduling schemes that achieve the boundary of the capacity region. From an information theoretic point of view, the opportunistic scheduling scheme simply achieves the point on boundary of the capacity region that maximizes the sum-rate of the system. However, it is not clear as to what scheduling schemes will achieve the other rate points on the boundary of the capacity region. This chapter targets to provide an answer to this question. More specifically, a class of scheduling schemes that can achieve all points on the boundary of the capacity region are derived. It will be interesting to note that the opportunistic scheme then becomes a special case of a generalized class of scheduling schemes considered in this chapter. By looking at the boundary points on the capacity region, we can easily find those "fair" achievable rate pairs and then obtain the corresponding "fair scheduling" schemes.

There has been a lot of work related to achieving fairness in an opportunistic system. Most notable results in this area are treated in [32], [33], [34], [35], [36], [37], [38], [39], [40], [41] and [42]. One of the familiar algorithms used in the IS-856 standard (which is TDMA based) is the proportional fair scheduler (see [9], [43]). This scheduler achieves fairness by giving service to the user which has the highest ratio of requested rate at a particular time instant to the average throughput attained over a time scale of interest. More specifically, let $R_k[m], T_k[m]$ be the requested rate and the average throughput of mobile k at time slot m. Proportional fair scheduler services the user k^* which has the highest ratio of

$$\frac{R_k[m]}{T_k[m]}$$

The average throughputs, $T_k[m]$ are updated using

$$T_k[m+1] \quad = \quad \begin{cases} \big(1 - 1/t_c\big)T_k[m] + (1/t_c)R_k[m] & k = k^* \\ \big(1 - 1/t_c\big)T_k[m] & k \neq k^* \end{cases}$$

where t_c is the length of the time window, and can be interpreted as the latency time-scale of application. As the parameter t_c becomes large, it can be shown that this algorithm maximizes [43]

$$\sum_{k=1}^{K} \log T_k$$

where K is the number of users in the system. Though this algorithm is very intuitive in the sense of achieving fairness, it does not claim *optimality* of the throughputs achieved by users. On the other hand, the scheduling schemes proposed in this chapter, provides the desired long-term throughputs $(\mathbb{E}[R_1], \mathbb{E}[R_2])$ to user-1 and user-2 respectively (where $\mathbb{E}$ denotes the expectation operator), and it is claimed that there does not exist any other scheme that can provide a throughput of $(\mathbb{E}[R_1] + \epsilon_1, \mathbb{E}[R_2] + \epsilon_2)$, where $\epsilon_1, \epsilon_2 > 0$. To this end, it is shown that "fairness" can actually be achieved by carefully selecting the scheduling scheme which depends solely on the instantaneous channel strengths of users. This chapter begins with finding a class of optimal scheduling schemes for achieving all points on the boundary of the rate region for the non-cooperative case (i.e., a typical downlink scenario). This concept is extended to derive optimal scheduling schemes for the case where users are allowed to cooperate (i.e., optimal cooperative scheduling schemes are derived). Finally, the achievable rate region of a downlink cooperative system is characterized. It is to be noted that in [44], the authors have studied the achievable rate region of a variable-power and variable-time transmission scheme of a time-division downlink system. It is to be underscored that the contribution of this chapter is not in finding the achievable rate region of a TD downlink system (which is well studied in [44]), but to seek scheduling schemes that can achieve the boundary points of the capacity region.

User-cooperation has been found to increase the achievable rate region significantly in an uplink multi-access scenario. In [8], [45], the authors have shown that the rate region approaches the multi-input single-output system. Similarly, the rate region of a system with cooperation for a broadcast system has been found in [46]. In this chapter, a downlink TD system is considered wherein users are allowed to cooperate (i.e., users can help each other when the BS services a particular user). For this scenario, optimal scheduling schemes are derived and eventually the achievable rate region of this simple user-cooperative system is found.

B. Statement of the Problem

Consider the system shown in Fig. 6, where there is a base station (BS) sending independent information to two wireless users. The system is operated in a TD mode where, in each time slot only one of the wireless users can get service. The wireless channels between the BS and the users are subject to Rayleigh block fading, and there is an average transmission power constraint of P at the BS during each block. Let a_i be the channel gain of the ith user which is complex Gaussian with zero-mean

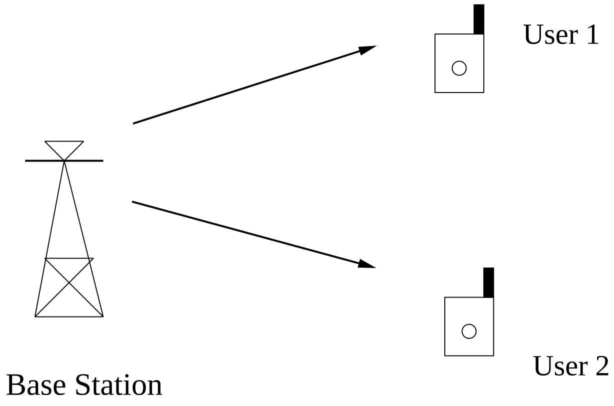


Fig. 6. Downlink system model.

and unit variance. Accordingly, the magnitude $|a_i|$ follows a Rayleigh distribution. a_i changes independently from block to block and is available both at the BS and the wireless users. Therefore, if the ith user's information is scheduled to be transmitted during a block, the received signal of the corresponding user can be written as

$$y_i = a_i\sqrt{P}S_i + n_i$$

where S_i is the transmitted signal of unit power, and, n_i is the additive white gaussian noise (AWGN) with zero mean and unit variance. Note that since $|a_i|$ follows a Rayleigh distribution, the power gain, $|a_i|^2$ of the corresponding wireless link follows an exponential distribution. Let $\gamma_i = \mathbb{E}[|a_i|^2]$ $(i = 1, 2)$ be the expected channel power gains of the links from the BS to the i^{th} user. It is emphasized that the term "scheduling scheme" refers to the manner in which the BS services the users solely based on their channel power gain $|a_i|^2$. For simplicity, henceforth let $|a_i|^2 = h_i$ $(i = 1, 2)$.

Given the fact that the BS accurately knows the channel gain of users, it can devise strategies, as to how the users are to be serviced. For example, as we have discussed in Section A, if the system throughput is to be maximized (i.e., the sum-rate

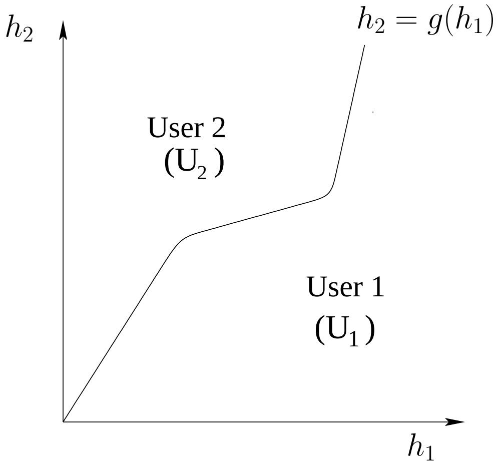


Fig. 7. Finding optimal scheduling scheme.

capacity is to be achieved), the BS needs to service the user with the best channel h_i. This chapter addresses the problem of finding a class of optimal scheduling schemes, for the downlink scenario, which essentially encompasses the opportunistic scheduling as a special case. Next, we consider the scenario where the two users are allowed to cooperate (i.e., the users can help each other when they get serviced). This can be achieved by using one user as a relay for the other user. For this case, we seek to find the optimal scheduling scheme (i.e., optimal cooperative scheduling scheme) that can achieve all points on the boundary of the capacity region. Then, based on the optimal cooperative scheduling schemes, we determine the achievable rate region of the cooperative scheduling system. It is to be noted that the power and bandwidth are kept the same in both cooperative and non-cooperative systems to make the comparison fair.

C. Optimal Non-Cooperative Scheduling Scheme

The problem of finding the optimal scheduling scheme amounts to finding a function $h_2 = g(h_1)$ in the $h_1 - h_2$ plane (see Fig. 7) such that the expected rate of user-2 is maximized, given that the expected rate of user-1 is fixed. From [47], [48], any point on the boundary of the rate region can be obtained by the following optimization problem

$$\max \quad \mu_1\,\mathbb{E}[R_1] + \mu_2\,\mathbb{E}[R_2] \quad 0 \leq \mu_1, \mu_2 \leq 1,\ \mu_1 + \mu_2 = 1 \tag{3.1}$$

where μ_i $(i = 1, 2)$ represent the weight given to the i^{th} user and $\mathbb{E}[R_i]$ is the expected rate of user-i. As μ_i ranges from 0 to 1, the weights given to i^{th} user changes which in turn gives the boundary of the rate tuple $\big(\mathbb{E}[R_1], \mathbb{E}[R_2]\big)$ corresponding to the weighted sum rate in (3.1).

Theorem III.1. *Let $K = \frac{\mu_1}{\mu_2}$ represent the ratio of the weights of the users. The optimal scheduling scheme for the two-user case has the following form*

$$\begin{aligned} &\textit{Schedule User 2, if} \\ &\qquad h_2 \;>\; \frac{(1+h_1P)^K - 1}{P}. \\ &\textit{Schedule User 1, if} \\ &\qquad h_2 \;<\; \frac{(1+h_1P)^K - 1}{P}. \end{aligned} \tag{3.2}$$

where $0 \leq K < \infty$.

Proof. Assuming without loss of generality that the noise power spectral density is $N_o = 1$, and letting $R_i = \log_2\big(1 + h_iP\big)$, $f_i(h_i) = \frac{1}{\gamma_i}\exp\big(-\frac{h_i}{\gamma_i}\big)$, the optimization problem in (3.1) becomes

$$\begin{aligned} &\max \quad \mu_1\mathbb{E}[R_1] + \mu_2\mathbb{E}[R_2] \\ = &\max\left\{ \iint\limits_{U_1} \mu_1 R_1 f_1(h_1) f_2(h_2) dh_1 dh_2 + \iint\limits_{U_2} \mu_2 R_2 f_1(h_1) f_2(h_2) dh_1 dh_2 \right\} \\ = &\max \int_0^\infty\!\!\int_0^\infty \big(\mu_1 R_1 \mathcal{I}_{U_1} + \mu_2 R_2 \mathcal{I}_{U_2}\big) f_1(h_1) f_2(h_2) dh_1 dh_2 \end{aligned} \tag{3.3}$$

where U_1 and U_2 are the regions over which the integrals are computed, and

$$\mathcal{I}_{U_i} = \begin{cases} 1 & \text{if} \quad (h_1, h_2) \in U_i, \\ 0 & \text{if} \quad (h_1, h_2) \notin U_i. \end{cases} \tag{3.4}$$

By noticing that the integrand in (3.3) is always positive, the integral in (3.3) is maximized by maximizing the terms inside the parentheses. Due to the fact that only one user gets serviced during a time slot, the integrand in (3.3) is maximized by assigning

$$\begin{aligned} \mathcal{I}_{U_1} = 1,\, \mathcal{I}_{U_2} = 0 \quad &\text{if} \quad \mu_1 R_1 > \mu_2 R_2, \\ \mathcal{I}_{U_1} = 0,\, \mathcal{I}_{U_2} = 1 \quad &\text{if} \quad \mu_1 R_1 < \mu_2 R_2. \end{aligned}$$

The optimal scheduling scheme then becomes

Schedule User-2, if

$$\begin{aligned} \mu_2 R_2 &> \mu_1 R_1, \\ \mu_2 \log_2\big(1 + h_2 P\big) &> \mu_1 \log_2\big(1 + h_1 P\big), \\ h_2 &> \frac{(1 + h_2 P)^K - 1}{P}. \end{aligned}$$

Schedule User-1, if

$$\begin{aligned} \mu_2 R_2 &< \mu_1 R_1, \\ h_2 &< \frac{(1 + h_1 P)^K - 1}{P}. \end{aligned}$$

which is the desired result. □

It is instructive to see that for the case when $K = 1$ (i.e., the sum-rate is being maximized), the scheduling scheme becomes the well known opportunistic method. The above derived result easily extends to the multi-user case which is stated below.

Theorem III.2. *For a system with n users, let $K_j = \frac{\mu_j}{\mu_i}$ $(j = 1, 2 \dots n, j \neq i)$ be the ratio of weights given to user-j with respect to user-i. Then, the optimal scheduling scheme is to schedule user-i, if $\left\{h_i > \frac{(1+h_jP)^{K_j}-1}{P}\right\}$ where $j = 1, 2 \dots n, j \neq i$.*

Proof. Using similar ideas and notations from the two user problem, the optimization problem for the n user case can be written as

$$\begin{aligned} & \max \quad \mu_1\mathbb{E}[R_1] + \mu_2\mathbb{E}[R_2] + \dots \mu_n\mathbb{E}[R_n] \qquad 0 \leq \mu_1, \mu_2 \dots \mu_n \leq 1, \sum_{i=1}^{n} \mu_i = 1 \\ = & \max \quad \int_{U_1}\!\!\cdots\int \mu_1 R_1 \prod_{i=1}^{n} f_i(h_i)\, dh_i + \dots + \int_{U_n}\!\!\cdots\int \mu_n R_n \prod_{i=1}^{n} f_i(h_i)\, dh_i \\ = & \max \quad \int\cdots\int \left(\mu_1 R_1 \mathcal{I}_{U_1} + \mu_2 R_2 \mathcal{I}_{U_2} + \dots + \mu_n R_n \mathcal{I}_{U_n}\right) \prod_{i=1}^{n} f_i(h_i)\, dh_i \end{aligned}$$

The above integral is maximized by assigning $\mathcal{I}_{U_i} = 1$ and $\mathcal{I}_{U_j} = 0$ $(j = 1, 2 \dots n, j \neq i)$ if

$$i = \arg \max_{j=1,2\dots n} \mu_j R_j$$

Letting $K_j = \frac{\mu_j}{\mu_i}$ $(j = 1, 2 \dots n, j \neq i)$ together with the above condition concludes the proof. □

1. Approximately Optimal Scheduling Scheme

Next, we answer the following: How does linear scheduling compare with optimal scheduling? The following class of linear scheduling schemes are approximately optimal (i.e., achieves rate pairs very close to the rate pairs achieved by the optimal scheduling scheme), for the two-user case:

Schedule User 1, if

$$h_1 \quad > \quad Ch_2$$

Schedule User 2, if

$$h_1 \quad < \quad Ch_2 \tag{3.5}$$

where $0 \leq C < \infty$.

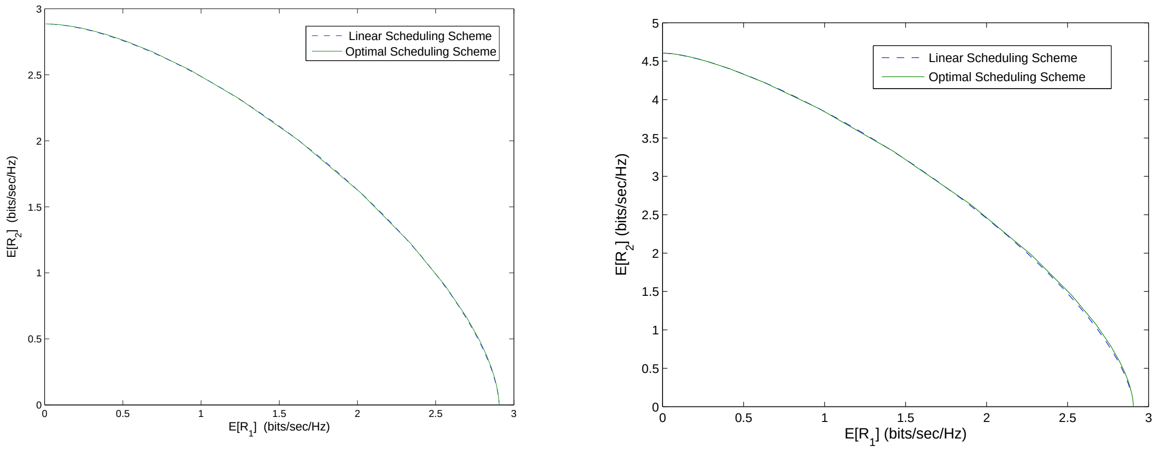
The expected rate (in bits/sec/Hz) attained by users, using the scheme in (3.5), can be worked out to be:

$$\begin{aligned}
\mathbb{E}[\mathrm{R}_1] &= \int\limits_{h_1=0}^{\infty}\int\limits_{h_2=0}^{Ch_1} \log_2\Big(1+h_1P\Big)\frac{1}{\gamma_2}\exp\Big(-\frac{h_2}{\gamma_2}\Big)\frac{1}{\gamma_1}\exp\Big(-\frac{h_1}{\gamma_1}\Big)dh_1dh_2 \\
&= \ell\Big[\int\limits_{h_1=0}^{\infty}\log\big(1+h_1P\big)\exp\big(-\frac{h_1}{\gamma_1}\big)dh_1 - \int\limits_{h_1=0}^{\infty}\log\big(1+h_1P\big)\exp\big(-(\frac{1}{\gamma_1}+\frac{C}{\gamma_2})h_1\big)dh_1\Big] \\
&= \Big[\exp\big(\frac{1}{\gamma_1P}\big)\Gamma\big(0,\frac{1}{\gamma_1P}\big) - \frac{\gamma_2}{\gamma_1C+\gamma_2}\exp\Big(\frac{\gamma_2+C\gamma_1}{\gamma_1\gamma_2P}\Big)\Gamma\Big(0,\frac{\gamma_2+C\gamma_1}{\gamma_1\gamma_2P}\Big)\Big]\log_2 e
\end{aligned}$$

where $\ell = \frac{\log_2 e}{\gamma_1}$. Similarly,

$$\begin{aligned}
\mathbb{E}[\mathrm{R}_2] &= \int\limits_{h_2=0}^{\infty}\int\limits_{h_1=0}^{\frac{h_2}{C}} \log_2\Big(1+h_2P\Big)\frac{1}{\gamma_2}\exp\Big(-\frac{h_2}{\gamma_2}\Big)\frac{1}{\gamma_1}\exp\Big(-\frac{h_1}{\gamma_1}\Big)dh_1dh_2 \\
&= \Big[\exp\big(\frac{1}{\gamma_2P}\big)\Gamma\big(0,\frac{1}{\gamma_2P}\big) - \frac{\gamma_1C}{\gamma_1C+\gamma_2}\exp\Big(\frac{\gamma_1C+\gamma_2}{\gamma_1\gamma_2CP}\Big)\Gamma\big(0,\frac{\gamma_1C+\gamma_2}{\gamma_1\gamma_2CP}\big)\Big]\log_2 e
\end{aligned}$$

where $\Gamma(a,x) = \int\limits_{x}^{\infty} t^{a-1}e^{-t}dt$ is the upper incomplete gamma function. The rate region of the optimal scheduling scheme and that of the linear scheduling scheme are shown in Fig. 8. The reason for the linear scheduling scheme to be very close to the optimal scheduling scheme can be partially explained by the fact that for small K, we have $\frac{(1+hP)^K-1}{P} \approx Kh$, and (3.2) converges to a linear scheduling scheme. It is easy to see that by switching the roles of users, $\frac{(1+hP)^{(1/K)}-1}{P} \approx (1/K)h$ for large K as well (where $K = \frac{\mu_1}{\mu_2}$). However, we see from Fig. 8, that the linear scheduling scheme approximates the optimal scheduling scheme well for all achievable rates. We thus have numerically shown that linear scheduling nearly achieves the optimal rate region.



(a) (b)

Fig. 8. Rate region of a linear and optimal scheduling scheme for P=10 and when (a) $\gamma_1 = \gamma_2 = 1$, (b) $\gamma_1 = 1$, $\gamma_2 = 4$.

D. Optimal Cooperative Scheduling Schemes and the Achievable Rate Region

In this section, a two-user scenario when the mobile users can perform downlink cooperation as shown in Fig. 9 is considered. Here again, one is interested in finding out the optimal scheduling scheme, which will then be used to study the achievable rate region. It is important to note that the optimal non-cooperative scheduling schemes derived in Section C need not be the same for the case when users are allowed to cooperate. In the cooperative scenario, the BS can choose to send all the data that belongs to the scheduled user (the user which is serviced), or, the BS can send a fraction of the other user's data as well as the data of the scheduled user. This is done by the BS with the intention that the data of the other user will subsequently be forwarded (relayed) by the user being serviced, by allocating resources such as power and bandwidth. Hence the total rate achieved by a particular user (say user-1) in a cooperative scheduling scheme will be equal to the sum of the following:

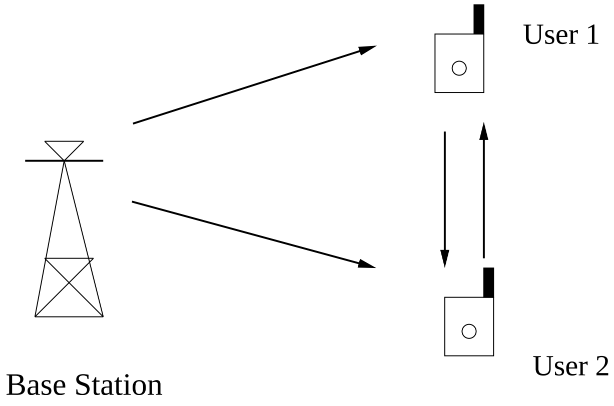


Fig. 9. Downlink cooperation model.

a) Rate achieved during the time when the BS services user-1, and

b) Rate achieved, when user-2 helps user-1, during the time the BS services user-2.

The cooperation scheme is described next.

1. System Model

Let P, P_1, P_2 be the average power constraint of the BS, user-1 and user-2 respectively. Let α_1 and $(1-\alpha_1)$ be the fraction of user-1's data and user-2's data respectively that the BS sends to user-1, when user-1 is serviced. Notice that in the non-cooperative scheme, $\alpha_1 = 1$ (i.e., the BS sends only the data of user-1). Similarly, let α_2 and $(1 - \alpha_2)$ be the fraction of user-2's data and user-1's data respectively, that the BS sends to user-2, when user-2 is serviced. Let β_1 be the bandwidth allocated for carrying user-1's data and user-2's data from the BS, whenever user-1 is serviced. The bandwidth allocated for relaying user-2's data from user-1, when user-1 is serviced, is $(1 - \beta_1)$. Here again, note that when $\beta_1 = 1$, this scheme corresponds to a typical non-cooperative scenario. Similarly, let β_2 be the bandwidth allocated for carrying user-2's data and user-1's data from the BS, whenever user-2 is serviced. The fraction of bandwidth allocated for relaying user-1's data from user-2, when user-2 is serviced, is represented by $(1 - \beta_2)$. Let h_3 be the inter-user channel power gain between the user-1 – user-2 links, which we assume for simplicity is AWGN. Furthermore, it is assumed that h_3 is constant [1], and also that the mobile users are equipped with infinite buffers. When the BS services a user (say user-2), the data intended for user-1 that is to be forwarded by user-2 need not be done instantaneously, but can be stored in the buffer until there are enough resources that can be allocated for the relaying operation. This assumption implies that the users can tolerate an arbitrary amount of delay.

2. Optimal Cooperative Scheduling Scheme

We next characterize the optimal cooperative scheduling scheme, when users are allowed to cooperate, according to the system model presented above. As done in Section C, finding the optimal scheduling scheme amounts to finding a function $h_2 = g(h_1)$, in the $h_1 - h_2$ plane such that the boundary points in the rate region are obtained, with the additional freedom that the users are allowed to cooperate. Note that the function $g(h_1)$ does not depend on the instantaneous value of the inter-user channel power gain h_3, because, by the assumption made in Section 1, the data to be relayed to the other user, by the user that is serviced, need not be done based on the instantaneous rate of the inter-user channel. However, as will be seen from the results of Theorem III.3 and Section 3 (refer (3.23)), the cooperative scheduling scheme does depend on the long term average value of the inter-user rate, through the resource allocation parameters α_i, β_i $(i = 1, 2)$.

The net rate attained by user 1 (R_1), is the sum of the direct rate attained from

[1] This assumption is made to simplify the analysis. The case when h_3 is random is discussed in Section 3

the BS when user 1 gets serviced (R_{01}) and the rate achieved by cooperating with user 2, when user 2 gets serviced (R_{21}),under a particular scheduling scheme . Notice that the rate attained by user-1 by cooperating with user-2 (R_{21}) is the rate achievable in a typical relay channel. Defining $\mathcal{I}_{U_1}$ and $\mathcal{I}_{U_2}$ as in (3.4), the expected rate achieved by user 1 (assuming noise power, $N_o = 1$) is:

$$\mathbb{E}[\mathrm{R}_1] = \mathbb{E}[\mathrm{R}_{01}] + \mathbb{E}[\mathrm{R}_{21}]$$

and,

$$\mathbb{E}[\mathrm{R}_{01}] \;=\; \alpha_1\beta_1\Big\{\mathbb{E}\big[\log_2\big(1+h_1\frac{P}{\beta_1}\big)\mathcal{I}_{U_1}\big]\Big\} \tag{3.6}$$

$$\begin{aligned}\mathbb{E}[\mathrm{R}_{21}] &= \mathbb{E}\Big\{\min\big[(1-\alpha_2)\beta_2\log_2\big(1+h_2\frac{P}{\beta_2}\big)\mathcal{I}_{U_2},(1-\beta_2)\log_2\big(1+h_3\frac{P_2}{1-\beta_2}\big)\big]\Big\}\\ &= \min\Big\{\mathbb{E}\big[(1-\alpha_2)\beta_2\log_2\big(1+h_2\frac{P}{\beta_2}\big)\mathcal{I}_{U_2}\big],(1-\beta_2)\log_2\big(1+h_3\frac{P_2}{1-\beta_2}\big)\Big\} \quad (3.7)\end{aligned}$$

where, the above equality follows from the assumption that the users are equipped with infinite buffers and h_3 being a constant. Similarly the expected rate attained by user 2, under the assumption that the users are equipped with infinite buffers is given by:

$$\mathbb{E}[\mathrm{R}_2] = \mathbb{E}[\mathrm{R}_{02}] + \mathbb{E}[\mathrm{R}_{12}] \tag{3.8}$$

where,

$$\mathbb{E}[\mathrm{R}_{02}] \;=\; \alpha_2\beta_2\Big\{\mathbb{E}\big[\log_2\big(1+h_2\frac{P}{\beta_2}\big)\mathcal{I}_{U_2}\big]\Big\} \tag{3.9}$$

$$\mathbb{E}[\mathrm{R}_{12}] \;=\; \min\Big\{\mathbb{E}\big[(1-\alpha_1)\beta_1\log_2\big(1+h_1\frac{P}{\beta_1}\big)\mathcal{I}_{U_1}\big],(1-\beta_1)\log_2\big(1+h_3\frac{P_1}{1-\beta_1}\big)\Big\} \tag{3.10}$$

Without loss of generality, one can consider the case when user-2 helps user-1 (but user-1 does not help user-2) by letting $\alpha_1 = 1, \beta_1 = 1$ and $0 \leq \alpha_2, \beta_2 \leq 1$. To see why there is no loss in generality, consider a two-user system that allows for bi-directional cooperation (i.e., both users help their partner). In particular, user-1 carries a certain data rate, Δ_1, that is forwarded to user-2 and simultaneously, user-2 carries a certain data rate, Δ_2 which is forwarded to user-1. This bi-directional scheme has a certain set of resource allocation parameters $(\alpha_1, \alpha_2, \beta_1, \beta_2)$ as well as a scheduling scheme described by K. Suppose that $\Delta_1 < \Delta_2$ (that is, user-2 provides more help to user-1 than user-1 does to user-2). Now consider replacing this system with a one-way cooperative scheme where user-2 forwards to user-1 a data rate equal to $\Delta_2 - \Delta_1$, but user-1 does not forward anything to user-2. Note that the net flow of data between user-1 and user-2 is the same in both systems. In the one-way system, user-1 will see an increase in data rate since it now can use its full bandwidth for the link between the BS and user-1 (it does not need to reserve any of the bandwidth to forward data to user-2). By the same argument, user-2 will also see an increase in its data rate as it has to forward a smaller amount of data to user-1 and hence can use more of its bandwidth for the link from the base station to user-2. Hence, for any two-way cooperative scheme, there exists a one-way cooperative scheme which performs at least as well. Hence in finding the optimal resource allocation and the scheduling scheme, it is sufficient to consider schemes with cooperation only in one direction. The scenario $\alpha_1 = 1, \beta_1 = 1$ and $0 \leq \alpha_2, \beta_2 \leq 1$ gives the maximum rates that user-1 can attain, for different rates attained by user-2. The expected rates achieved by the users are given by substituting $\alpha_1 = 1, \beta_1 = 1$ in (3.6) and (3.10). Then

$$\begin{aligned}
\mathbb{E}[\mathrm{R}_1] &= \mathbb{E}\big[\log_2\big(1 + h_1 P\big)\mathcal{I}_{U_1}\big] + \\
&\quad \min\Big\{\mathbb{E}\big[(1-\alpha_2)\beta_2 \log_2\big(1 + h_2 \frac{P}{\beta_2}\big)\mathcal{I}_{U_2}\big], (1-\beta_2)\log_2\big(1 + h_3 \frac{P_2}{1-\beta_2}\big)\Big\} \\
\mathbb{E}[\mathrm{R}_2] &= \alpha_2 \beta_2 \mathbb{E}\big[\log_2\big(1 + h_2 \frac{P}{\beta_2}\big)\mathcal{I}_{U_2}\big] \qquad (3.11)
\end{aligned}$$

Theorem III.3. *The optimal cooperative scheduling scheme belongs to the same class as that of a non-cooperative scheduling scheme given in (3.2).*

It is emphasized that Theorem III.3 does not mean that the optimal cooperative and non-cooperative scheduling schemes are the same, but that the optimal cooperative scheduling scheme belongs to the same class as that of the optimal non-cooperative scheduling schemes. This will become more clear in Section E, where an example is provided to illustrate this concept.

Proof. Let $h_2 = g(h_1)$, be the unknown scheduling scheme. By assuming that the users are equipped with infinite buffers, and also that user-2 helps user-1, the optimization problem can be formulated as (see (3.11))

$$\begin{aligned} \max \quad \mathbb{E}[\mathrm{R}_1] &= \mathbb{E}\big[\log_2\big(1+h_1P\big)\mathcal{I}_{U_1}\big] + \\ \min\Big\{\mathbb{E}\Big[(1-\alpha_2)\beta_2\log_2\Big(1+h_2\frac{P}{\beta_2}\Big)\mathcal{I}_{U_2}\Big]&, (1-\beta_2)\log_2\Big(1+h_3\frac{P_2}{1-\beta_2}\Big)\Big\} \\ \text{subject to} \quad \mathbb{E}[\mathrm{R}_2] &= \alpha_2\beta_2\mathbb{E}\Big[\log_2\Big(1+h_2\frac{P}{\beta_2}\Big)\mathcal{I}_{U_2}\Big] = C_1 \end{aligned} \tag{3.12}$$

where C_1 is a constant, $\mathcal{I}_{U_1}$ and $\mathcal{I}_{U_2}$ are the regions over which integration is performed and are defined in (3.4).

Case 1: For a given inter-user channel power gain, h_3, if [2]

$$(1-\alpha_2)\beta_2\,\mathbb{E}\Big[\log_2\Big(1+h_2\frac{P}{\beta_2}\Big)\mathcal{I}_{U_2}\Big] < (1-\beta_2)\log_2\Big(1+h_3\frac{P_2}{1-\beta_2}\Big) \tag{3.13}$$

then the optimization problem in (3.12) can be written as,

$$\max \qquad \mathbb{E}[\mathrm{R}_1] \;=\; \mathbb{E}\big[\log_2\big(1+h_1P\big)\mathcal{I}_{U_1}\big] + (1-\alpha_2)\beta_2\mathbb{E}\big[\log_2\big(1+h_2\frac{P}{\beta_2}\big)\mathcal{I}_{U_2}\big]$$

[2] For the case when h_3 is random, (3.13) is replaced by

$$(1-\alpha_2)\beta_2\,\mathbb{E}\Big[\log_2\Big(1+h_2\frac{P}{\beta_2}\Big)\mathcal{I}_{U_2}\Big] < (1-\beta_2)\,\mathbb{E}\Big[\log_2\Big(1+h_3\frac{P_2}{1-\beta_2}\Big)\Big]$$

$$\text{subject to} \qquad \mathbb{E}[\mathrm{R}_2] = \alpha_2\beta_2\mathbb{E}\big[\log_2\big(1+h_2\frac{P}{\beta_2}\big)\mathcal{I}_{U_2}\big] = C_1 \tag{3.14}$$

The above optimization further reduces to:

$$\max \quad \mathbb{E}\big[\log_2\big(1+h_1P\big)\mathcal{I}_{U_1}\big] + \big((1-\alpha_2)\beta_2+\lambda\alpha_2\beta_2\big)\mathbb{E}\big[\log_2\big(1+h_2\frac{P}{\beta_2}\big)\mathcal{I}_{U_2}\big] - \lambda.C_1 \tag{3.15}$$

where λ is the lagrangian multiplier. We can immediately recognize that the above maximization problem is of the same form as in (3.1), and it clear that the optimal co-operative scheduling scheme would belong to the same class as that of non-cooperative optimal scheduling schemes. We make this argument more precise by proving this using calculus of variations. The optimization problem in (3.15), reduces to:

$$\max \quad \Bigg(\int\limits_{h_1=0}^{\infty}\int\limits_{h_2=0}^{g(h_1)} \log_2\Big(1+h_1P\Big)f_1(h_1)f_2(h_2)dh_2dh_1\Bigg) +$$

$$\Big((1-\alpha_2)\beta_2+\lambda\alpha_2\beta_2\Big)\Bigg(\int\limits_{h_1=0}^{\infty}\int\limits_{h_2=g(h_1)}^{\infty} \log_2\Big(1+h_2\frac{P}{\beta_2}\Big)f_1(h_1)f_2(h_2)dh_2dh_1\Bigg) - \lambda C_1$$

where $f_i(h_i) = \frac{1}{\gamma_i}\exp(-\frac{h_i}{\gamma_i})$ is the pdf of a exponentially distributed random variable with mean γ_i. The above problem, further reduces to

$$\max \quad \frac{\log_2 e}{\gamma_1}\int\limits_{h_1=0}^{\infty} \mathrm{F}dh_1 - \lambda C_1 \qquad\qquad \text{where,}$$

$$\begin{aligned}\mathrm{F} = {} & \log\Big(1+h_1P\Big)\exp\Big(-\frac{h_1}{\gamma_1}\Big) + \Big((1-\alpha_2)\beta_2+\lambda\alpha_2\beta_2\Big)\times \\ & \exp\Big(\frac{\beta_2}{\gamma_2P}-\frac{h_1}{\gamma_1}\Big)\Gamma\Big(0,\frac{g(h_1)}{\gamma_2}+\frac{\beta_2}{\gamma_2P}\Big)\Bigg(\Big((1-\alpha_2)\beta_2+\lambda\alpha_2\beta_2\Big)\log\Big(1+g(h_1)\frac{P}{\beta_2}\Big) - \\ & \log\Big(1+h_1P\Big)\Bigg)\exp\Big(-\frac{g(h_1)}{\gamma_2}-\frac{h_1}{\gamma_1}\Big)\end{aligned}$$

and $\Gamma(a,x) = \int\limits_x^{\infty} t^{a-1}e^{-t}dt$ is the upper incomplete gamma function. As the integrand

F, is not dependent on the derivatives of the unknown function, $h_2 = g(h_1)$, the Euler's method of finding the unknown function ([49],Chap 17), reduces to:

$$\begin{aligned} \frac{\partial \mathrm{F}}{\partial g} &= 0 \\ \Rightarrow g(h_1) &= \frac{(1+h_1P)^{\frac{1}{(1-\alpha_2)\beta_2+\lambda\alpha_2\beta_2}} - 1}{(P/\beta_2)} \end{aligned}$$

which is exactly of the same form as derived in (3.2). For a fixed α_2, β_2, the value of λ is chosen such that the constraint in (3.14), is satisfied. It is important to note that the optimal cooperative scheduling scheme, $g(h_1)$, derived above depends on the parameters (α_2, β_2), which in turn depends on the inter-user channel power gain, h_3, through (3.13).

Case 2: For a given h_3, if [3]

$$(1-\alpha_2)\beta_2\,\mathbb{E}\Big[\log_2\Big(1+h_2\frac{P}{\beta_2}\Big)\mathcal{I}_{U_2}\Big] > (1-\beta_2)\log_2\Big(1+h_3\frac{P_2}{1-\beta_2}\Big) \qquad (3.16)$$

then the optimization problem in (3.12) can be written as

$$\begin{aligned} \max \quad \mathbb{E}[\mathrm{R}_1] &= \mathbb{E}\big[\log_2\Big(1+h_1P\Big)\mathcal{I}_{U_1}\big] + (1-\beta_2)\log_2\Big(1+h_3\frac{P_2}{1-\beta_2}\Big) \\ \text{subject to} \quad \mathbb{E}[\mathrm{R}_2] &= \alpha_2\beta_2\mathbb{E}\Big[\log_2\Big(1+h_2\frac{P}{\beta_2}\Big)\mathcal{I}_{U_2}\Big] = C_1 \qquad (3.17) \end{aligned}$$

The above optimization problem further reduces to:

$$\max \quad \mathbb{E}\big[\log_2\big(1+h_1P\big)\mathcal{I}_{U_1}\big]+(1-\beta_2)\log_2\big(1+h_3\frac{P_2}{1-\beta_2}\big)+\lambda\alpha_2\beta_2\mathbb{E}\big[\log_2\big(1+h_2\frac{P}{\beta_2}\big)\mathcal{I}_{U_2}\big]-\lambda.C_1$$

where λ is the lagrange multiplier. Here again, we recognize that this optimization

[3] For the case when h_3 is random, (3.16) is replaced by

$$(1-\alpha_2)\beta_2\,\mathbb{E}\Big[\log_2\Big(1+h_2\frac{P}{\beta_2}\Big)\mathcal{I}_{U_2}\Big] > (1-\beta_2)\,\mathbb{E}\Big[\log_2\Big(1+h_3\frac{P_2}{1-\beta_2}\Big)\Big]$$

problem is of the same form as in (3.1). By using the same technique as used for the previous case, it can be shown that the optimal scheduling scheme is given by

$$g(h_1) = \frac{(1 + h_1 P)^{\frac{1}{\lambda \alpha_2 \beta_2}} - 1}{(P/\beta_2)} \tag{3.18}$$

where λ is chosen to satisfy the constraint in (3.17). Again, it is important to note that $g(h_1)$ derived above depends on the parameters (α_2, β_2), which in turn depends on h_3 through (3.16). □

3. Achievable Rate Region of the Optimal Cooperative Scheduling Scheme

Since the optimal cooperative scheduling scheme belongs to the same class as that of optimal non-cooperative scheduling scheme, it is sufficient to consider the class of schemes in (3.2) to get the achievable rate region. The rates achievable by user-1 and user-2, under the class of optimal scheduling schemes from (3.6),(3.7), (3.9) and (3.10) are:

$$\begin{aligned} \mathbb{E}[R_1] &= \alpha_1 \beta_1 \, \mathbb{E}\left[\log_2 \left(1 + h_1 \frac{P}{\beta_1}\right) \mathcal{I}_{\left(h_2 < \frac{\left(1 + h_1 \frac{P}{\beta_1}\right)^K - 1}{(P/\beta_2)}\right)} \right] + \\ &\quad \min \left\{ \mathbb{E}\left[(1 - \alpha_2)\beta_2 \log_2 \left(1 + h_2 \frac{P}{\beta_2}\right) \mathcal{I}_{\left(h_2 > \frac{\left(1 + h_1 \frac{P}{\beta_1}\right)^K - 1}{(P/\beta_2)}\right)} \right], \right. \\ &\quad \left. (1 - \beta_2) \log_2 \left(1 + h_3 \frac{P_2}{1 - \beta_2}\right) \right\} \end{aligned} \tag{3.19}$$

$$\begin{aligned} \mathbb{E}[R_2] &= \alpha_2 \beta_2 \, \mathbb{E}\left[\log_2 \left(1 + h_2 \frac{P}{\beta_2}\right) \mathcal{I}_{\left(h_2 > \frac{\left(1 + h_1 \frac{P}{\beta_1}\right)^K - 1}{(P/\beta_2)}\right)} \right] + \\ &\quad \min \left\{ \mathbb{E}\left[(1 - \alpha_1)\beta_1 \log_2 \left(1 + h_1 \frac{P}{\beta_1}\right) \mathcal{I}_{\left(h_2 < \frac{\left(1 + h_1 \frac{P}{\beta_1}\right)^K - 1}{(P/\beta_2)}\right)} \right], \right. \\ &\quad \left. (1 - \beta_1) \log_2 \left(1 + h_3 \frac{P_1}{1 - \beta_1}\right) \right\} \end{aligned} \tag{3.20}$$

where

$$\mathcal{I}_{X>Y} = \begin{cases} 1 & \text{for} \quad X > Y \\ 0 & \text{for} \quad X \leq Y \end{cases}$$

For a fixed $(K, \alpha_1, \alpha_2, \beta_1, \beta_2)$, let $\mathcal{R}(K, \alpha_1, \alpha_2, \beta_1, \beta_2)$, denote the achievable rate region. Hence the achievable rate region of this cooperation scheme is given by

$$\mathcal{R}_c \triangleq \left(\bigcup_{\substack{0 \leq \alpha_1, \alpha_2, \beta_1, \beta_2 \leq 1 \\ 0 \leq K < \infty}} \mathcal{R}(K, \alpha_1, \alpha_2, \beta_1, \beta_2) \right)$$

To characterize $\mathcal{R}_c$, it suffices to find the boundary points defined by the above parameters. First, we find the boundary points for the case when user-2 helps user-1 (but user-1 does not help user-2). The expected rates achievable by user 1 and user 2 for a fixed $(K, \alpha_1, \alpha_2, \beta_1, \beta_2)$, under the assumption that user-2 helps user-1, the users are equipped with infinite buffers and the inter-user channel power gain h_3 being a constant, can be obtained from (3.19),(3.20) and is given by (see also (3.11)):

$$\begin{aligned} \mathbb{E}[R_1] &= \mathbb{E}\left[\log_2\left(1 + h_1 P\right) \mathcal{I}_{\left(h_2 < \frac{(1+h_1 P)^K - 1}{(P/\beta_2)}\right)} \right] + \\ &\quad \min\left\{ \mathbb{E}\left[(1-\alpha_2)\beta_2 \log_2\left(1 + h_2 \frac{P}{\beta_2}\right) \mathcal{I}_{\left(h_2 > \frac{(1+h_1 P)^K - 1}{(P/\beta_2)}\right)} \right], \right. \\ &\quad \left. (1-\beta_2)\log_2\left(1 + h_3 \frac{P_2}{1-\beta_2}\right) \right\} \qquad (3.21) \\ \mathbb{E}[R_2] &= \alpha_2 \beta_2\, \mathbb{E}\left[\log_2\left(1 + h_2 \frac{P}{\beta_2}\right) \mathcal{I}_{\left(h_2 > \frac{(1+h_1 P)^K - 1}{(P/\beta_2)}\right)} \right] \qquad (3.22) \end{aligned}$$

As in [8], [45], we assume that the rate that is transmitted by the BS to user 2 for carrying user 1's data, is no more than (or in best case equals to) what user 2 can relay to user 1, so that the resources are not wasted. It follows from this assumption, that α_2, β_2 are to be chosen such that[4]

$$\mathbb{E}\Big[(1-\alpha_2)\beta_2 \log_2\Big(1+h_2\frac{P}{\beta_2}\Big)\mathcal{I}_{\left(h_2 > \frac{(1+h_1P)^K-1}{(P/\beta_2)}\right)}\Big] = (1-\beta_2)\log_2\Big(1+h_3\frac{P_2}{1-\beta_2}\Big) \quad (3.23)$$

For $0 \leq \alpha_2 \leq 1$, the value of β_2 which satisfies (3.23) can be found, and it can be shown easily that a solution always exists, which satisfies $0 \leq \beta_2 \leq 1$. Define the following:

$$\begin{aligned} \beta_2^*(\alpha_2) \quad &: \quad \text{the value of } \beta_2 \text{which satisfies (3.23), for a fixed } \alpha_2 \\ R_{21}\big(\beta_2^*(\alpha_2)\big) \quad &= \quad \big(1-\beta_2^*(\alpha_2)\big)\log_2\Big(1+h_3\frac{P_2}{1-\beta_2^*(\alpha_2)}\Big) \end{aligned}$$

The boundary points of the achievable rate region for a particular value of K, for the class of optimal cooperative scheduling scheme is:

$$\begin{aligned} \mathbb{E}[R_1] \quad &= \quad \mathbb{E}\Big[\log_2\Big(1+h_1P\Big)\mathcal{I}_{\left(h_2 < \frac{(1+h_1P)^K-1}{(P/\beta_2)}\right)}\Big] + R_{21}\big(\beta_2^*(\alpha_2)\big) \\ \mathbb{E}[R_2] \quad &= \quad \alpha_2\beta_2^*(\alpha_2)\,\mathbb{E}\Big[\log_2\Big(1+h_2\frac{P}{\beta_2^*(\alpha_2)}\Big)\mathcal{I}_{\left(h_2 > \frac{(1+h_1P)^K-1}{(P/\beta_2)}\right)}\Big] \end{aligned}$$

Notice, that when $\alpha_2 = 0$ (i.e, when user 2 is fully helping user 1), then from (3.21),(3.22)) and (3.23), it follows that the *maximum additional rate* (due to user 2), user 1 can get is $R_{21}\big(\beta_2^*(0)\big)$, with $\mathbb{E}[\mathrm{R}_2] = 0$. The boundary points for all values of K ($0 \leq K < \infty$) can be found and the union of all such points is the achievable rate region for the case when user 2 helps user 1. Similar analysis can be carried out for the case when user 1 helps user 2. The achievable rate region of the cooperation

[4] For the case when h_3 is random, (3.23) is replaced by

$$\mathbb{E}\Big[(1-\alpha_2)\beta_2 \log_2\Big(1+h_2\frac{P}{\beta_2}\Big)\mathcal{I}_{\left(h_2 > \frac{(1+h_1P)^K-1}{(P/\beta_2)}\right)}\Big] = \mathbb{E}\big[(1-\beta_2)\log_2\Big(1+h_3\frac{P_2}{1-\beta_2}\Big)\big].$$

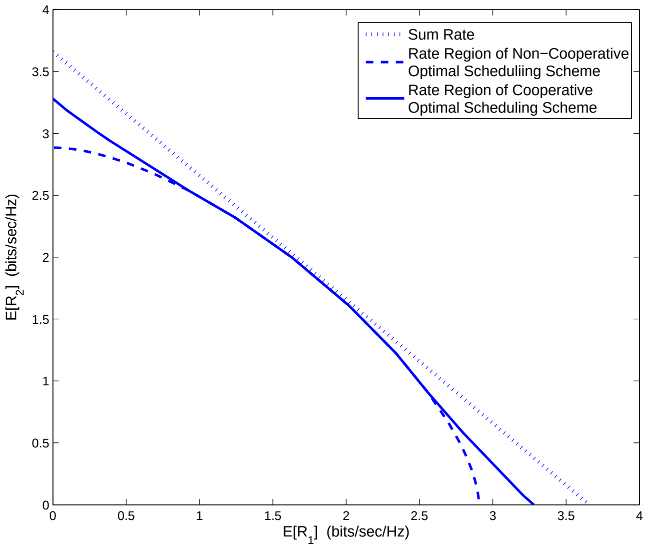


Fig. 10. Rate region of cooperation scheme with the following parameters. $P = 10, P_1 = P_2 = 0.4, \mathbb{E}[h_1] = 1, \mathbb{E}[h_2] = 1, h_3 = 10.$

scheme is given in Fig. 10, Fig. 11, and Fig. 12.

We conclude this section by giving some examples of resource allocation parameters. Suppose for example that the BS services user-2 at a particular time slot. Then, the amount of user-1's data that the BS needs to send to user-2 (so that user-2 could relay user-1's data) is a function of how much bandwidth is allocated for the inter-user channel. If no bandwidth is allocated for the inter-user channel, then it is easy to see that the BS should not send any of user-1's data so that the resource allocated for carrying user-1's data could as well be given to user-2 thereby increasing user-2's throughput. Also it is to be noted that once we allocate some bandwidth for the inter-user channel, then the amount of bandwidth that can be allocated for the link between the BS and user-2 will decrease (notice that both the user's data need to be carried in the link between the BS and user-2), which in turn would mean lesser throughput for user-2. So, we see that there is a tradeoff between the parameters α_2

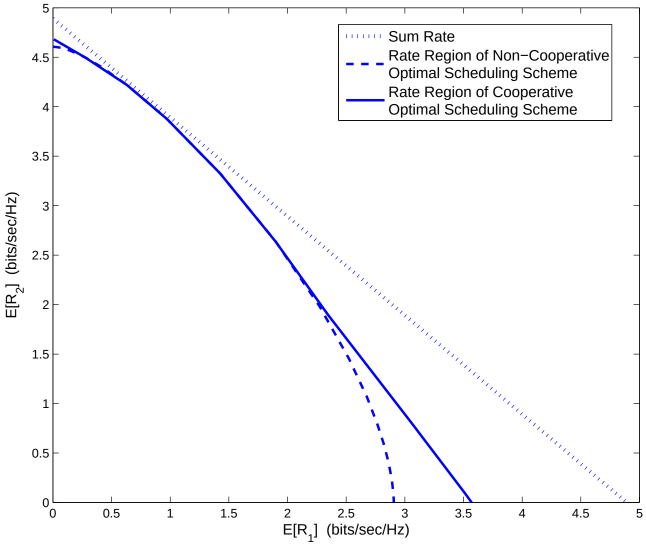


Fig. 11. Rate region of cooperation scheme with the following parameters. $P = 10, P_1 = P_2 = 0.4, \mathbb{E}[h_1] = 1, \mathbb{E}[h_2] = 4, h_3 = 10.$

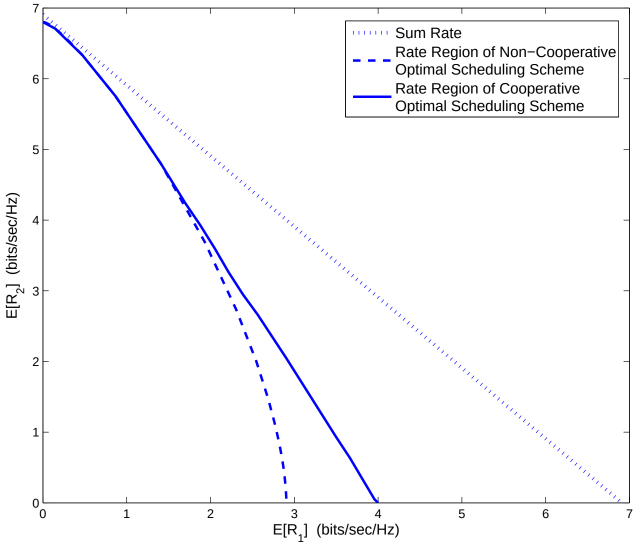


Fig. 12. Rate region of cooperation scheme with the following parameters. $P = 10, P_1 = P_2 = 0.4, \mathbb{E}[h_1] = 1, \mathbb{E}[h_2] = 20, h_3 = 20.$

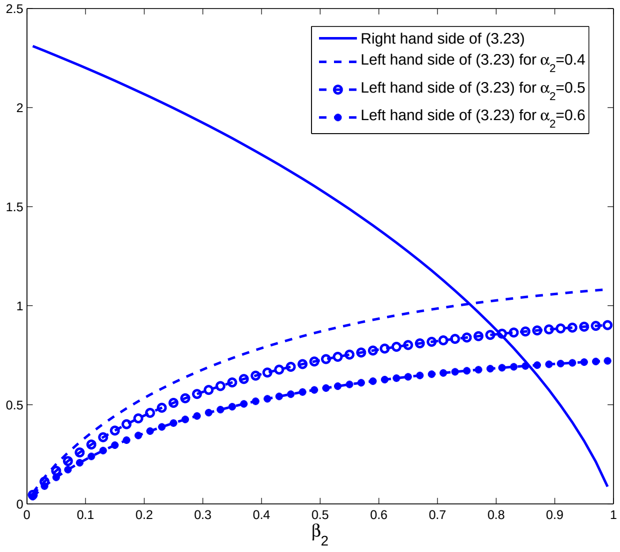


Fig. 13. The LHS of (3.23) for $\alpha_2 = 0.4, 0.5, 0.6$ and the RHS of (3.23) for $K = 1$. From the figure one can obtain the corresponding value of β_2 for $\alpha_2 = 0.4, 0.5, 0.6$ to be $0.754, 0.806, 0.856$. Here $P = 10, P_1 = P_2 = 0.4$ and $\gamma_1 = \gamma_2 = 1$, $h_3 = 10$.

and β_2. Though the trade-off is easy to see for boundary points (i.e., when $\alpha_2 = 1$ or $\beta_2 = 1$), for general values of $\alpha_2 \in (0, 1)$, the corresponding value of β_2 can be found from (3.23). Fig. 13 shows the left hand side (LHS) and right hand side (RHS) of (3.23) for $\alpha_2 = 0.4, 0.5, 0.6$, and $K = 1$. We see that there is a unique solution β_2 for every value of α_2, as was claimed earlier.

E. Discussions

One aspect of the derived cooperative scheduling scheme is that, one does not have to service a user by taking into account the instantaneous value of inter-user channel power gain (h_3), but instead, to consider the class of non-cooperative scheduling scheme and then allocate resources over a long time scale, based on (3.23). It is important to see that the cooperative scheduling scheme does depend on h_3 through the resource allocation parameters as in (3.23). One may note that in order to find the boundary of the rate region (for example, the case when user-2 helps user-1, but not vice versa, which is outlined in Section 3), a constrained optimization over the parameters (α_2, β_2, K) where $0 \leq \alpha_2, \beta_2 \leq 1$ and $0 \leq K < \infty$ needs to be performed. In what follows, a simple iterative algorithm is presented for finding the resource allocation parameters and the scheduling scheme.

As an example, consider determining the resource allocation parameters and the scheduling scheme for the following cooperative scenario

$$\begin{aligned} &\max && \mathbb{E}[R_1], \\ &\text{subject to} && \mathbb{E}[R_2] = C. \end{aligned}$$

where C is a constant. From (3.21), (3.22) and (3.23), the above problem reduces to

$$\max \quad \mathbb{E}\left[\log_2\left(1 + h_1 P\right)\mathcal{I}_{\left(h_2 < \frac{(1+h_1P)^K - 1}{(P/\beta_2)}\right)}\right] + (1-\beta_2)\log_2\left(1 + h_3\frac{P_2}{1-\beta_2}\right), \quad (3.24)$$

$$\text{subject to} \quad \beta_2\,\mathbb{E}\left[\log_2\left(1 + h_2\frac{P}{\beta_2}\right)\mathcal{I}_{\left(h_2 > \frac{(1+h_1P)^K - 1}{(P/\beta_2)}\right)}\right] - (1-\beta_2)\log_2\left(1 + h_3\frac{P_2}{1-\beta_2}\right) = C. \quad (3.25)$$

Recall that β_2 is the bandwidth allocated for carrying user-1's as well as user-2's data from the BS when user-2 is serviced and K defines the scheduling scheme according to (3.2). The higher the value of K, the lower are the chances of user-2 accessing the channel. The feasible values of β_2 and K which satisfy (3.25) with $C = 0.5$ are depicted in Fig. 14. The reason why K is a monotonic increasing function of β_2 can be explained as follows. As β_2 increases, the 'bit pipe' for carrying the data from the BS to user-2 becomes larger, and hence user-2 needs to access the channel a comparatively fewer number of times for the constraint in (3.25) to be met (i.e., which implies K is higher). The constraint between K and β_2 in (3.25) being nonconvex (see Fig. 14) and the non-existence of a closed form expression between K and β_2 makes the optimization problem in (3.24) difficult. One can always perform an exhaustive search over all feasible values of K and β_2, however in what follows a simple iterative algorithm is presented to obtain the maximizing value of β_2 and K in (3.24) from (3.25), by linearizing the constraint between K and β_2 of the form $\tilde{K}(\beta_2) = a_2\,\beta_2 + a_1$.

Step 1) Initialize the maximizing value of β_2 in (3.24), $\beta_2^{max} = 0.5$.

Step 2) Form $\tilde{K}(\beta_2) = a_2\beta_2 + a_1$, with a_1, a_2 chosen to match $\frac{dK}{d\beta_2}\Big|_{\beta_2^{max}}$ and $K(\beta_2^{max})$ which can be obtained from the constraint between K and β_2 in (3.25).

Step 3) Update β_2^{max} by finding the new value of β_2 that maximizes (3.24) using the locally linearized constraint $\tilde{K}(\beta_2)$ obtained in step (2), in place of (3.25).

Step 4) Repeat steps (2) and (3) until convergence occurs.

It was observed that the above algorithm typically converged in about eight to ten iterations and that the numerical computations needed for this algorithm were much less than performing an exhaustive search. An example is now given that provides the resource allocation parameters and the scheduling scheme for both the cooperative and the non-cooperative case. Consider determining the resource allocation parameters and the scheduling scheme when $\mathbb{E}[R_2] = 0.5$ bits/sec/Hz for the system parameters $P = 10$, $P_1 = P_2 = 0.4$, $\mathbb{E}[h_1] = 1$, $\mathbb{E}[h_2] = 4$ and $h_3 = 10$. For the non-cooperative case, the optimal scheduling scheme can be obtained by letting $\alpha_2 = \beta_2 = 1$ in (3.22), which yields $K = 4.5$ with the rate attained by user 1, $\mathbb{E}[R_1] = 2.8$ bits/sec/Hz. For the cooperative scenario the resource allocation parameters and the scheduling scheme that are evaluated by the algorithm given above and

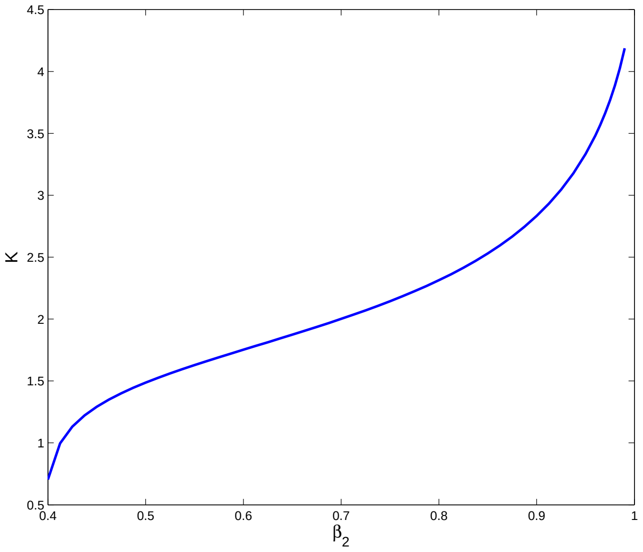


Fig. 14. The plot of feasible values of β_2 and K that satisfy (3.25) for $C = 0.5$, $P = 10$, $P_2 = 0.4$, $\mathbb{E}[h_1] = 1$, $\mathbb{E}[h_2] = 4$, $h_3 = 10$.

(3.23) turn out to be $\beta_2 = 0.74$, $\alpha_2 = 0.32$, $K = 2.1$ with the rate attained by user 1, $\mathbb{E}[R_1] = 3.2$ bits/sec/Hz. It is important to note that the optimal cooperative and non-cooperative scheduling scheme (parameterized by K) are not the same. In this example, the capability of the two users to cooperate has improved user-1's throughput by 14 %. When $\mathbb{E}[R_2] = 0$, the resource allocation parameters for cooperative case is $\beta_2 = 0.67$, $\alpha_2 = 0$, $K = 2.47$, with $\mathbb{E}[R_1] = 3.6$ bits/sec/Hz, whereas for the non-cooperative case it is $\mathbb{E}[R_1] = 2.9$ bits/sec/Hz (with K arbitrarily large). For this case the user-1's throughput has increased by 24 % through user cooperation (the maximum that user-1 can achieve, since $\mathbb{E}[R_2] = 0$). When $\mathbb{E}[R_2] = 1.3$ bits/sec/Hz, the resource allocation parameters for both the cooperative and non-cooperative case turn out to be the same (i.e., $\alpha_2 = \beta_2 = 1, K = 0.35$), with $\mathbb{E}[R_1] = 1.2$ bits/sec/Hz in which case there are no gains that can be obtained by the cooperation scheme.

F. Conclusion

We have derived a class of optimal scheduling schemes that achieve the boundary points of the rate region for the non-cooperative case. Hence we are able to determine the necessary scheduling scheme to achieve any rate pair on the boundary. Furthermore, it was shown that linear scheduling can be considered to be nearly optimal for all practical purposes. It was established that optimal cooperative scheduling schemes belong to the same class as that of optimal non-cooperative scheduling schemes. Based on the class of optimal cooperative scheduling schemes, the achievable rate region was found, and it is observed that a cooperative scheduling system enlarges the achievable rate region. Finally, a simple iterative algorithm was presented to find the resource allocation parameters and the scheduling scheme for a cooperative scenario. We conclude this chapter by providing a intuitive explanation of the proposed cooperative scheduling scheme.

Note that the opportunistic scheme (i.e., $K = 1$ in (3.2)) maximizes the sum-rate of the system (i.e., $R_1 + R_2$). If the goal is to maximize R_2, subject to $R_1 = C$, and, suppose that the opportunistic scheme produces a rate for user-1 such that $R_1 < C$. Then the non-cooperative system must adjust its scheduling scheme by lowering the parameter K to some level $K = K_{nc}$ such that the user-1's constraint is satisfied. However user-2's throughput will decrease at the same time. In the sense of the sum-rate, this scheduling scheme is inefficient, but in the non-cooperative environment, this is the only mechanism that is available to meet the constraint.

Next, consider a two-user system where user-2 helps user-1. The additional rate that user-2 provides user-1 through the inter-user channel, bends the scheduling scheme back towards the opportunistic scheme. This will decrease the data rate that is directly communicated to user-1, but will increase the data rate available to user-2 (by more than R_1 is decreased). Hence, user-2 can then use some of the extra rate to allow user-1 to make up for its lost rate and still have some left over. But the catch here is that, this comes at an additional cost because there are now two transmitters (that is, the BS and user-2) and the bandwidth must be shared. As the rate of the data that must be forwarded from user-2 to user-1 increases, the bandwidth needed for the inter-user link increases (i.e., β_2 decreases) and hence the bandwidth available for the link from BS to user-2 decreases. So, not only does user-2 have to give up some of its bit pipe from the BS to carry data for user-1, but the size of the bit pipe gets squeezed in order to make room for the inter-user channel. At some point, the cost associated with cooperation becomes greater than the benefit of using a more efficient non-cooperative scheduling scheme at which point cooperative scheduling scheme does not yield any gains beyond the gains provided by the non-cooperative scheduling scheme.

CHAPTER IV

THE EFFECTIVE CAPACITY OF A TIME DIVISION DOWNLINK SCHEDULING SYSTEM

We derived the optimal scheduling scheme for the downlink system in Chapter III without taking into account any constraints on the quality of service (QoS) requirements. In this chapter, we extend the optimal scheduling scheme presented in Chapter III, to the case when QoS constraints are imposed on each user. The QoS requirement is specified in terms of the asymptotic decay rate of the buffer occupancy. Furthermore, the effective capacity of this system is characterized.

A. Introduction

In Chapter III we studied the throughput that the users can achieve under the assumption that users can wait for an arbitrary amount of time (i.e., when users can tolerate an arbitrary amount of delay). But in practical systems, for example, in voice applications, usually the users need to get their data (or packets) within a certain time range (for example, 50ms to 100ms). We formulate a problem that takes into account these ‘practical’ constraints and then seek what throughput the users achieve.

Future wireless communication networks are expected to support high traffic and at the same time provide reliable service to delay sensitive applications. Hence, there has been an increasing interest in the recent past for understanding and evaluating the performance of wireless communication systems not only based on Shannon capacity, but also on QoS in terms of delay sensitivity. Such an analysis becomes essential if one considers real-time services like multimedia video conference or VoIP because the key QoS metric is to ensure delay bounds rather than achieving spectral efficiency. There has been extensive study devoted to understanding how to design systems to enhance spectral efficiency, for example [50]. Information theory has been the framework in evaluating these techniques. However, information theory does not take into account QoS service requirements, and hence it becomes necessary to seek tools that can help understand the interplay between throughput and QoS. There have been some recent contributions in trying to address these issues in different ways. For example, [41] addresses the power-delay tradeoff. In [12], [51] and [52], effective capacity has been used as a tool in studying the performance of resource allocation of a point to point system and MIMO systems with delay constraints. In this chapter, effective capacity is also used to study the problem of finding the capacity region and the optimal scheduling scheme for a two user downlink system with QoS constraints. The insights learned from the two user problem are then used to develop a sub-optimal scheduling scheme for the multiuser scenario.

This chapter is organized as follows. The system model is detailed in Section B. The optimal scheduling scheme with QoS constraints and the capacity region is derived in Section C for the two-user case. Section D contains numerical results. A simple multiuser scheduling method based on the optimal two-user scheduling scheme is proposed in Section E. In Section F, numerical study is performed on mapping the QoS requirement, θ, and the delay profile (in ms), where it is shown that there does not exist a unique value of θ which yields a desired delay profile and that the delay profile depends on the system parameters. It is also shown in this section that for a typical delay profile desired in voice-type applications, a frequency division multiaccess (FDMA) scheme yields better throughput than the TDMA based scheme. Finally Section G concludes the chapter.

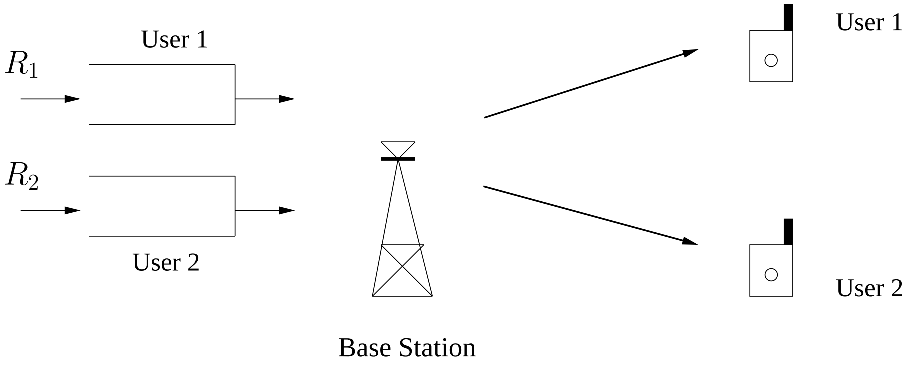


Fig. 15. Downlink system model with QoS constraints.

B. Statement of the Problem and System Model

Consider the system shown in Fig. 15 where there is a BS sending independent information to two wireless users. There is a buffer associated with each user at the BS into which messages intended for that user arrive at a particular rate. Furthermore, there is a QoS constraint, θ, for each buffer which indicates the exponential decay rate of the buffer overflow probability. The wireless channels between the BS and the users are subject to i.i.d Rayleigh block fading and there is an average power constraint of P at the BS during each block. Let a_i be the channel gain of the i^{th} user which is complex Gaussian with zero mean and unit variance. Accordingly, the magnitude, $|a_i|$, follows a Rayleigh distribution. $|a_i|$ changes independently from block to block and is perfectly known at the BS. Let T_i be the block length over which the i^{th} user's fading process remains constant and can be interpreted as the coherence time of the i^{th} user's channel. The system is operated in a TD mode where in each time slot only one of the wireless users gets service. Therefore, if the i^{th} user's information is scheduled to be transmitted during a block, then the received signal of the corresponding user can be written as:

$$y_i = a_i\sqrt{P}S_i + n_i.$$

where S_i is the transmitted signal of unit power, and n_i is the additive white gaussian noise (AWGN) with zero mean and variance N_oW (N_o is the power spectral density and W is the bandwidth). Accordingly, the service rate provided to user i at time m, is $R_i[m] = W\log(1+|a_i|^2\frac{P}{N_oW})$ and an amount equal to this gets depleted from the i^{th} user's buffer. Let $\gamma_i = \mathbb{E}[|a_i|^2]$ $(i = 1, 2)$ be the expected channel power gains of the links from the BS to the wireless users. In this chapter, the term 'scheduling scheme' refers to the manner in which the BS services the users. For simplicity, henceforth let $|a_i|^2 = h_i$ $(i = 1, 2)$.

Given the fact that the BS accurately knows the channel power gains of each user, the coherence times and the QoS requirement, θ, it can devise strategies as to how the users are to be serviced. One scheme would to be service a user solely based on the channel power gain (e.g., opportunistic scheme or the more general class of scheduling schemes presented in Chapter III). This chapter develops optimal schemes that makes use of all the available information such as the coherence times and the QoS constraint as well as the channel power gain in servicing a user and it is shown that these schemes naturally outperform the ones which does not take into account all this information.

C. Optimal Scheduling Scheme with QoS constraints and the Effective Capacity

As in [25], the problem of finding an optimal scheduling scheme reduces to finding a function $h_2 = g(h_1)$ in the $h_1 - h_2$ plane such that user-1 is serviced if $h_2 < g(h_1)$ and user-2 is serviced if $h_2 > g(h_1)$. Here h_1, h_2 are the channel power gains of user-1 and user-2, respectively. More specifically, scheduling schemes are desired which achieve a point on the boundary of the capacity region such that the effective capacity of user-2 is maximized, given that the effective capacity of user-1 is fixed (see also [25]).

These points on the boundary of the capacity region can be obtained from the well known optimization problem:

$$\max \qquad \beta \mathrm{E}_{C_1}(\theta) + (1-\beta)\mathrm{E}_{C_2}(\theta), \tag{4.1}$$

where,

$$\mathrm{E}_{C_1}(\theta) = -\frac{1}{\theta T_1} \log \mathbb{E}\left[e^{-\theta T_1 R_1[m]}\right], \qquad \mathrm{E}_{C_2}(\theta) = -\frac{1}{\theta T_2} \log \mathbb{E}\left[e^{-\theta T_2 R_2[m]}\right],$$

are the effective capacities of user-1 and user-2 respectively, $R_i[m] = W \log_2 \left(1 + h_i \frac{P}{N_o W}\right)$ is the service that user i gets at time m, and $\beta \in [0,1]$. Since only one user is serviced at time m, whenever, $R_1[m] > 0$, it follows that $R_2[m] = 0$ and vice-versa. Hence, the problem in (4.1), can be written as

$$\begin{aligned} \max \qquad & -\tfrac{\beta}{\theta T_1} \log \mathbb{E}\left[e^{-\theta T_1 \left(R_1[m]=0\right)} \mathcal{I}_{U_2} + e^{-\theta T_1 \left(R_1[m]>0\right)} \mathcal{I}_{U_1}\right] - \\ & \tfrac{(1-\beta)}{\theta T_2} \log \mathbb{E}\left[e^{-\theta T_2 \left(R_2[m]=0\right)} \mathcal{I}_{U_1} + e^{-\theta T_2 \left(R_2[m]>0\right)} \mathcal{I}_{U_2}\right], \end{aligned} \tag{4.2}$$

where, U_i is the region in the $h_1 - h_2$ plane over which user i is serviced and $\mathcal{I}_{U_i} (i = 1, 2)$ represents the indicator function, defined by

$$\mathcal{I}_{U_i} = \begin{cases} 1 & \text{if} \quad (h_1, h_2) \in U_i, \\ 0 & \text{if} \quad (h_1, h_2) \notin U_i \end{cases}$$

Equation (4.2), can be further simplified to

$$\begin{aligned} &\min \\ &\frac{\beta}{\theta T_1} \log \left(\int\limits_{h_1=0}^{\infty} \int\limits_{h_2=g(h_1)}^{\infty} p_1(h_1) p_2(h_2)\, dh_1\, dh_2 + \int\limits_{h_1=0}^{\infty} \int\limits_{h_2=0}^{g(h_1)} e^{-\theta T_1 R_1[m]} p_1(h_1) p_2(h_2)\, dh_1\, dh_2 \right) + \\ &\frac{(1-\beta)}{\theta T_2} \log \left(\int\limits_{h_1=0}^{\infty} \int\limits_{h_2=0}^{g(h_1)} p_1(h_1) p_2(h_2)\, dh_1\, dh_2 + \int\limits_{h_1=0}^{\infty} \int\limits_{h_2=g(h_1)}^{\infty} e^{-\theta T_2 R_2[m]} p_1(h_1) p_2(h_2)\, dh_1\, dh_2 \right) \end{aligned}$$

where, $p_i(h_i) = \frac{1}{\gamma_i}\exp\big(-\frac{h_i}{\gamma_i}\big)$ is the pdf of an exponentially distributed random variable with mean γ_i. After some algebra, the above reduces to:

$$\min \quad \frac{\beta}{\theta T_1}\log\Big(\int\limits_{h_1=0}^{\infty} f_1\big(h_1, g(h_1)\big)\,dh_1\Big) + \frac{(1-\beta)}{\theta T_2}\log\Big(\int\limits_{h_1=0}^{\infty} f_2\big(h_1, g(h_1)\big)\,dh_1\Big), \tag{4.3}$$

where,

$$f_1\big(h_1, g(h_1)\big) = \frac{1}{\gamma_1}\exp\big(-\frac{h_1}{\gamma_1}\big)\Bigg[\Big(1+h_1\frac{P}{N_oW}\Big)^{-k_1}\Big(1-\exp\big(-\frac{g(h_1)}{\gamma_2}\big)\Big)+\exp\big(-\frac{g(h_1)}{\gamma_2}\big)\Bigg],$$

$$f_2\big(h_1, g(h_1)\big) = \frac{1}{\gamma_1}\exp\big(-\frac{h_1}{\gamma_1}\big)\Bigg[1-\exp\big(-\frac{g(h_1)}{\gamma_2}\big)+\frac{1}{\gamma_2}F\big(k_2, g(h_1)\big)\Bigg], \tag{4.4}$$

and

$$F(k_2, g(h_1)) = \int\limits_{h_2=g(h_1)}^{\infty}\Big(1+h_2\frac{P}{N_oW}\Big)^{-k_2}\exp\big(-\frac{h_2}{\gamma_2}\big)\,dh_2; \quad k_i = \theta T_i W M \ \ (i=1,2),$$

and $M = \log_2 e$. Note that the integrals defined in (4.3) and (4.4) are bounded for $k_i \in [0,\infty)$ $(i = 1, 2)$. The optimization problem in (4.3) is a variational problem which can be solved by using the technique of calculus of variations [49] as follows. Let

$$J\big(g_1(h_1)\big) = \frac{\beta}{\theta T_1}\log\Big(\int\limits_{h_1=0}^{\infty} f_1\big(h_1, g_1(h_1)\big)\,dh_1\Big) + \frac{(1-\beta)}{\theta T_2}\log\Big(\int\limits_{h_1=0}^{\infty} f_2\big(h_1, g_1(h_1)\big)\,dh_1\Big). \tag{4.5}$$

As in Arfken ([49],Chap. 17), let, $g_1(h_1) = g(h_1)+s\eta(h_1)$, where, $g(h_1)$ is the function which minimizes (4.5), s is any constant, and $\eta(h_1)$ represents arbitrary variation. As the objective function in (4.5) is not dependent on the derivatives of the unknown function $g_1(h_1)$, Euler's method of finding the unknown function reduces to [49]

$$\frac{d}{ds}\Big[J\big(g_1(h_1)\big)\Big]_{s=0} = 0, \tag{4.6}$$

$$\frac{d}{ds}\Big[J\big(g_1(h_1)\big)\Big] =$$
$$\frac{\beta}{\theta T_1}\frac{\int\limits_{h_1=0}^{\infty}\frac{d}{dg_1}\big(f_1\big(h_1,g_1(h_1)\big)\big)\big(\frac{dg_1(h_1)}{ds}\big)dh_1}{\int\limits_{h_1=0}^{\infty}f_1\big(h_1,g_1(h_1)\big)dh_1} + \frac{(1-\beta)}{\theta T_2}\frac{\int\limits_{h_1=0}^{\infty}\frac{d}{dg_1}\big(f_2\big(h_1,g_1(h_1)\big)\big)\big(\frac{dg_1(h_1)}{ds}\big)dh_1}{\int\limits_{h_1=0}^{\infty}f_2\big(h_1,g_1(h_1)\big)dh_1}. \tag{4.7}$$

Define the following (for $i = 1, 2$):

$$\frac{d}{dg}\big(f_i(h_1,g(h_1))\big) = f'_{ig}\big(h_1,g(h_1)\big); \quad C_i = \int\limits_{h_1=0}^{\infty} f_i\big(h_1,g(h_1)\big)dh_1; \quad C = \frac{T_1C_1}{T_2C_2}. \tag{4.8}$$

By noting that $\frac{dg_1(h_1)}{ds} = \eta(h_1)$, evaluating (4.7) at $s = 0$ and using (4.6) and (4.8) yields

$$\int\limits_{h_1=0}^{\infty}\left(\left(\frac{\beta}{\theta T_1C_1}\right)f'_{1g}\big(h_1,g(h_1)\big) + \left(\frac{(1-\beta)}{\theta T_2C_2}\right)f'_{2g}\big(h_1,g(h_1)\big)\right)\eta(h_1)\,dh_1 = 0. \tag{4.9}$$

Noting that (4.9) needs to be satisfied for any $\eta(h_1)$, it follows that

$$\left(\tfrac{\beta}{\theta T_1C_1}\right)f'_{1g}\big(h_1,g(h_1)\big) + \left(\tfrac{(1-\beta)}{\theta T_2C_2}\right)f'_{2g}\big(h_1,g(h_1)\big) = 0, \tag{4.10}$$
$$\Rightarrow \frac{f'_{1g}}{f'_{2g}} = -\tfrac{1-\beta}{\beta}C.$$

Recall that the functions $f_i(i = 1, 2)$ are defined in (4.4). Defining K as below and after some algebra, (4.10), works out to be

$$\frac{1-\left(1+h_1\frac{P}{N_oW}\right)^{-\theta T_1MW}}{1-\left(1+g(h_1)\frac{P}{N_oW}\right)^{-\theta T_2MW}} = \frac{1-\beta}{\beta}C = K, \tag{4.11}$$

$$g(h_1) = \frac{\left(1+\frac{1}{K}\left[\left(1+\frac{P}{N_oW}h_1\right)^{-\theta T_1MW}-1\right]\right)^{\frac{-1}{\theta T_2MW}}-1}{P/N_oW}. \tag{4.12}$$

As β varies from 0 to 1, K varies from 0 to ∞ (see (4.11)). One can obtain a class of functions by changing K in (4.12) from zero to infinity, which in turn can be used in (4.3) to obtain the effective capacity. (Notice that $C > 0$ can be easily shown). For a given β, there exists an optimal value of K that depends on C (see (4.11)) and it is to be noted that C in turn depends on the function $g(h_1)$ in (4.12) (see also (4.8) and (4.4)). For the trivial case when $\beta = 0$ (i.e, when only the effective capacity of user-2 is maximized), then the optimal solution is to always service user-2 which implies that the function $g(h_1) = 0$ can be obtained from (4.12). Similarly, when $\beta = 1$, $g(h_1) = \infty$ can be obtained by solving $f'_{1g}\big(h1, g(h_1)\big) = 0$ in (4.10). For the general case when $\beta \in (0, 1)$, the following is a simple algorithm to evaluate the value of C and hence the optimal scheduling scheme (i.e., K in (4.11)).

Step 1) Initialize K to some positive value (e.g., $K = 1$).

Step 2) Form $g(h_1)$ from (4.12).

Step 3) Find C_1, C_2, and hence C from (4.8) by using $g(h_1)$ obtained in step (2).

Step 4) For the assumed value of β, update K according to $K = \frac{1-\beta}{\beta}C$ using the value of C obtained in step (3).

Step 5) Repeat steps (2) to (4) until convergence occurs.

It was observed that the above algorithm typically converged in about six to ten iterations. Taking a closer look at (4.11), in the limit as θ goes to zero,

$$\begin{aligned} \lim_{\theta \to 0} \left[\frac{\left(1 + h_1 \frac{P}{N_o W}\right)^{-\theta T_1 M W} - 1}{\left(1 + g(h_1) \frac{P}{N_o W}\right)^{-\theta T_2 M W} - 1} \right] &= \lim_{\theta \to 0} \frac{1-\beta}{\beta} C \\ \frac{\log\left(1 + h_1 \frac{P}{N_o W}\right)}{\log\left(1 + g(h_1) \frac{P}{N_o W}\right)} &= \frac{1-\beta}{\beta} \end{aligned} \tag{4.13}$$

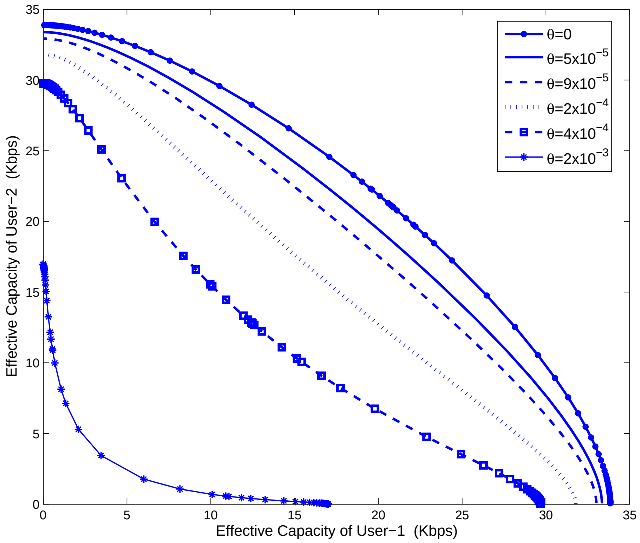


Fig. 16. Effective capacity for the cases when $\theta = 0, 5\times10^{-5}, 9\times10^{-5}, 2\times10^{-4}, 4\times10^{-4}$ and 2×10^{-3} for the system parameters $T_1 = T_2 = 100$ ms, $\gamma_1 = \gamma_2 = 1$ and $\frac{P}{N_oW} = 15$.

$$g(h_1) \quad = \quad \frac{\left(1 + h_1 \frac{P}{N_oW}\right)^{\frac{\beta}{1-\beta}} - 1}{P/N_oW} \tag{4.14}$$

This is precisely the optimal scheduling scheme derived in Chapter III. Notice that the LHS of (4.13) follows from L'Hospital rule and the constant C is easily shown to be $\frac{T_1}{T_2}$, when $\theta \longrightarrow 0$ from (4.8) and (4.4). Letting $\beta = 0.5$ in (4.14) gives the well known opportunistic scheduling scheme $g(h_1) = h_1$.

D. Numerical Results

The effective capacities are computed using the optimal scheduling scheme for the case when $\theta = 0$ (Shannon ergodic capacity) and $\theta = 5 \times 10^{-5}, 9 \times 10^{-5}, 2 \times 10^{-4}, 4 \times 10^{-4}$ and 2×10^{-3} for the system parameters $\frac{P}{N_oW} = 15$, $T_1 = T_2 = 100$ ms, $W = 10$ KHz, and $\gamma_1 = \gamma_2 = 1$ in Fig. 16. We can clearly see the loss in achievable throughputs

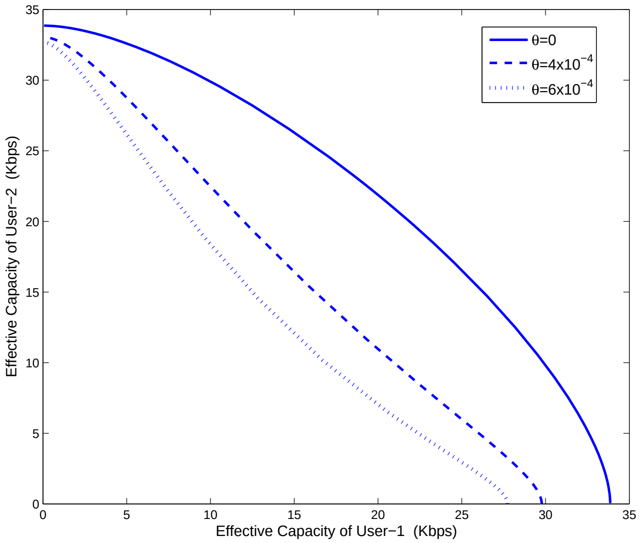


Fig. 17. Effective capacity for the case when $\theta = 0, 4 \times 10^{-4}$ and 6×10^{-4} for the system parameters T_1=100 ms, T_2=20 ms, $\gamma_1 = \gamma_2$=1 and $\frac{P}{N_o W} = 15$.

of both users as θ increases. The reason for this behaviour may be explained due to the fact that as θ increases, the input arrival rate must be controlled, which in turn limits the buffer length and hence the delay. The case when $\theta = 0$ eliminates the constraint on the buffer delay and depends only on the service process, which produces the Shannon ergodic capacity. Further, we see that as θ becomes large (e.g., $\theta = 2 \times 10^{-3}$), the ability of the system to provide equitable rates to users keeps decreasing. Hence for tight delay constraints, sharing resources between the users becomes unproductive. Fig. 17 shows the effective capacity when user-1 and user-2 have different coherence times, $T_1 = 100$ms and $T_2 = 20$ms. Although the average channel power gains of both users are the same (i.e., $\gamma_1 = \gamma_2 = 1$), the throughputs of the users differ. This can be attributed to the fact that user-1's channel changes slowly w.r.t user-2's channel and hence the throughput offered to user-1 is less than that of user-2. If we consider this phenomenon from the buffer perspective at the

BS, the buffer corresponding to the user with the larger coherence time can support lower arrival rates than the user with the channel having a smaller coherence time. This phenomenon is nicely captured in this framework, which the Shannon ergodic capacity cannot, as it does not account for delay constraints.

E. Extensions to a Multiuser Scenario

The problem of finding the optimal scheduling scheme for the multiuser case can be written as

$$\max \qquad \sum_{i=1}^{N} \beta_i E_{C_i}(\theta) \qquad \text{such that} \quad \sum_{i=1}^{N} \beta_i = 1. \tag{4.15}$$

where N represents the total number of users in the system, $E_{C_i}(\theta) = -\frac{1}{\theta T_i} \log \left[e^{-\theta T_i R_i[m]} \right]$ represents the effective capacity of user i, T_i represents the coherence time of the i^{th} user's channel, and $R_i[m]$ is the service rate of user i at the m^{th} time instant. Let h_i be the channel power gain of user i. Similar to the idea presented in Section C on the operational interpretation of (4.1) for the two user case, the problem in (4.15) can be thought of splitting the $\big(h_1, h_2 \ldots h_N\big)$ plane into N regions $\big(U_1, U_2 \ldots U_N\big)$, which requires finding a distinct set of $(N-1)$ surfaces. This problem appears to be mathematically intractable even for the case when $N = 3$. In what follows, a simple suboptimal scheme is presented for multiuser scheduling based on the optimal two user scheme derived in (4.11) and (4.12).

For $j = 1, 2 \ldots N$, let

$$\tilde{g}\big(h_j, T_j\big) = \frac{\beta_j}{T_j} \Bigg(1 - \Big(1 + h_j P \Big)^{-\theta T_j M W} \Bigg) \tag{4.16}$$

Recall that the optimal scheduling scheme for the two user case in (4.11), is to service user i for which

$$i = \arg\max_{j} \big(\ell_j \ \tilde{g}\big(h_j, T_j\big) \big) \tag{4.17}$$

where $\beta_1 = \beta$, $\beta_2 = 1 - \beta$, $\ell_1 = 1$ and $\ell_2 = \frac{C_1}{C_2}$ was obtained from C (see (4.8)) by the algorithm outlined in Section C.

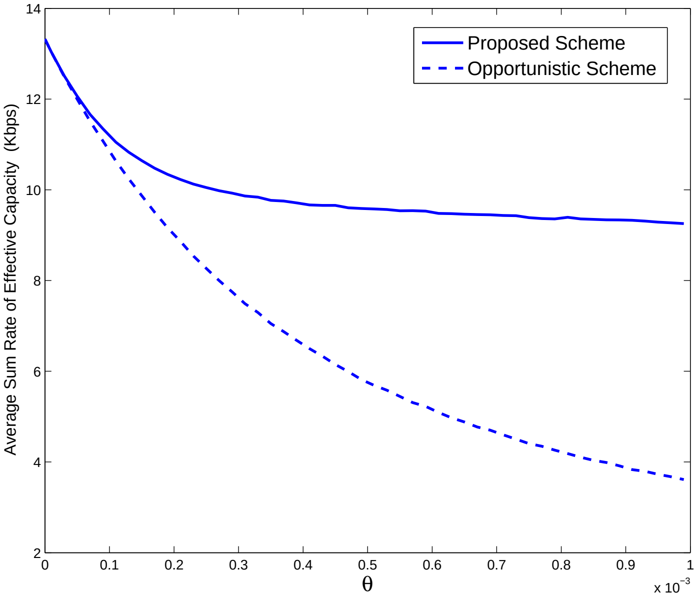


Fig. 18. Average sum rate of the opportunistic and the proposed scheme for the case when $N = 5$ with $T_1 = 20$ ms, T_2=70 ms, T_3=80 ms, T_4=100 ms, T_5=120 ms and $\frac{P}{N_o W} = 50$.

For the N user problem, consider maximizing the sum rate of the effective capacities (i.e., $\beta_j = 1/N \;\; \forall\, j = 1, 2, \ldots N$). For a given $(\ell_1, \ell_2 \ldots \ell_N)$, $(T_1, T_2 \ldots T_N)$ and $(h_1, h_2 \ldots h_N)$, let $\mathcal{R}^{sum}\big(\ell_1, \ell_2 \ldots \ell_N\big)$ denote the achievable sum rate by using the scheme in (4.17). Then the possible set of sum rates for the scheme in (4.17) can be written as:

$$\mathfrak{R}^{sum} = \bigcup_{\substack{\left(\ell_1, \ell_2 \ldots \ell_N\right) \\ \ell_j \in [0,\infty] \;\; \forall\, j = 1,2\ldots N}} \mathcal{R}^{sum}\big(\ell_1, \ell_2 \ldots \ell_N\big) \tag{4.18}$$

Since maximizing the sum rate of effective capacities implies maximizing the

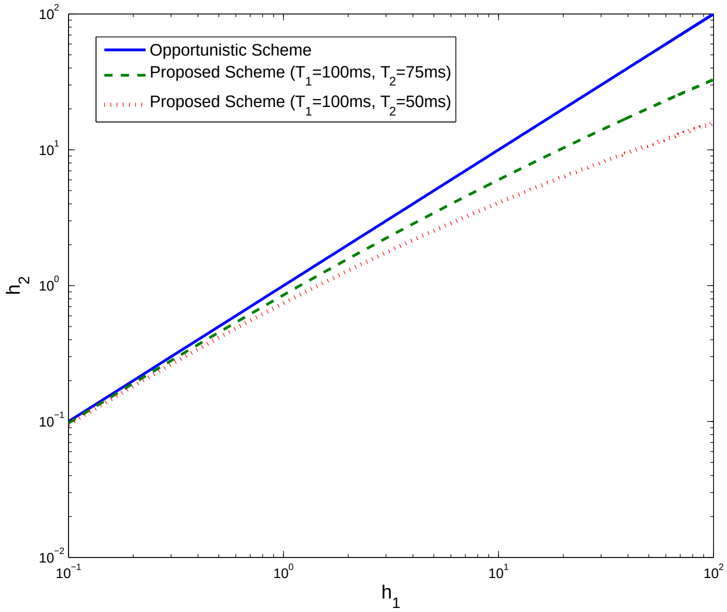


Fig. 19. Decision region for maximizing the sum rate of the opportunistic and the proposed scheme for the two user problem when $\theta = 10^{-4}$.

sum input arrival rates, a natural scheme to accomplish this is to maximize the sum output service rates (see also Fig. 15), for which the opportunistic scheduling scheme is well known to be optimal. The average sum rate of effective capacities achieved by the proposed scheduling scheme in (4.17) for $\ell_j = 1$ $(j = 1, 2 \ldots N)$, and that of the opportunistic scheme are compared in Fig. 18 for the case when $N = 5$ with $T_1 = 20$ ms, $T_2 = 70$ ms, $T_3 = 80$ ms, $T_4 = 100$ ms, $T_5 = 120$ ms. The reason that the proposed scheme yields a higher sum rate throughput can be attributed to the fact that it takes into account the channel power gains and delay constraints as well as the channel coherence times in servicing a user, unlike the opportunistic scheme which services the user solely based on the channel power gains. This can also be seen from the optimal decision region for maximizing the sum rate of effective capacities of the opportunistic and the proposed scheme depicted in Fig. 19 for the two user case. Further, when $\ell_j = 1, T_j = T$ $(j = 1, 2, \ldots N)$, the proposed scheme in (4.17) reduces to the opportunistic scheme (which implies that the opportunistic scheme is a subset of the proposed scheme).

F. Discussions

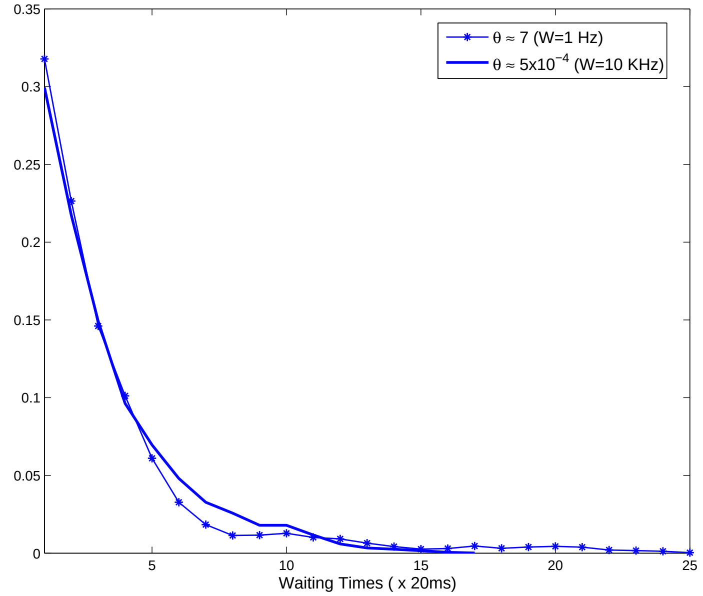


Fig. 20. Waiting time distribution of bits for different values of θ and the bandwidth W.

From a practical standpoint, one is interested in knowing what values of θ yields a desired delay profile of bits. Intuitively speaking this depends on the system parameters such as the bandwidth, SNR etc. The waiting time distribution of bits is depicted in Fig. 20 when $\theta = 7$ for $W = 1$ Hz and when $\theta = 5 \times 10^{-4}$ for $W = 10$ KHz with the other system parameters the same as was described in Section D. We infer from this plot there is no one universal value of θ that would yield a desired delay profile and that it depends on the parameters of the system.

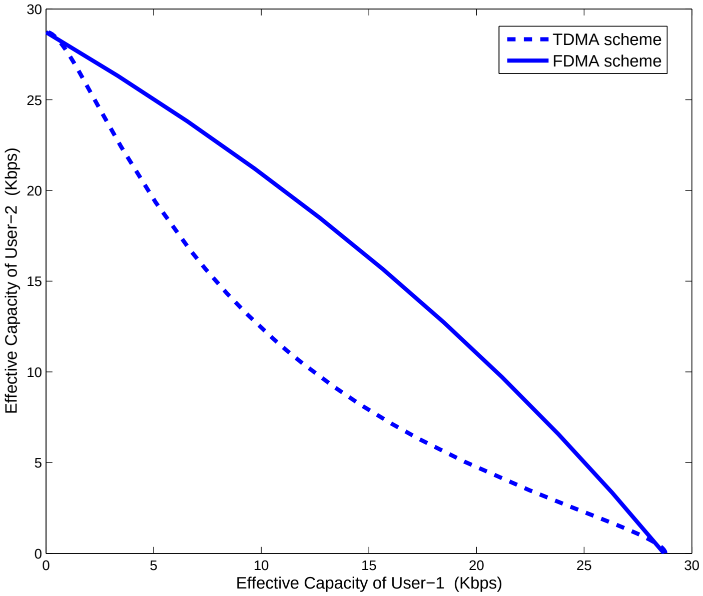


Fig. 21. Rate region for a desired delay profile (shown in Fig. 20) achieved by a TDMA and FDMA scheme.

Suppose, if the users are serviced by a FDMA scheme by splitting the bandwidth into two, would it be possible to get better rates than the TDMA scheme for a desired delay profile? For example, if one is interested in achieving the delay profile shown in Fig. 20 (which typically might correspond to a voice system), what are the admissible rates that a TDMA and a FDMA scheme can support? The rate region of a TDMA scheme when $W = 10$ KHz (with the other system parameters being the same as was described in Section D) is compared to a FDMA scheme with the users endowed with ηW KHz and $(1-\eta)W$ KHz respectively (where $0 \leq \eta \leq 1$) and with the same SNR as the time division scheme in Fig. 21. This study suggests that the FDMA schemes achieves better throughput for users than the TDMA scheme.

G. Conclusion

This chapter characterizes the effective capacity of a downlink scheduling system with QoS constraints. The optimal scheduling scheme without QoS constraints and the opportunistic scheduling schemes fall out as special cases of a more generalized class of scheduling schemes derived in this chapter. Thus, this chapter has unified the class of capacity achieving scheduling schemes for a given QoS constraint as well as demonstrates how the effect of various channel parameters and the QoS constraint influence the effective capacity. Based on the optimal two user scheduling scheme, a sub-optimal, yet effective scheme for multiuser scheduling is proposed and shown to dramatically increase the achievable sum rates as compared to the commonly used opportunistic scheduling. Finally, we conclude that for delay constrained applications, FDMA schemes are better than TDMA schemes in terms of getting better throughput to users.

CHAPTER V

THE ADMISSIBLE REGION OF A COGNITIVE RADIO SYSTEM

In previous chapters, we have looked at systems wherein users are allocated a licensed channel (i.e., specific time slots or separate frequency bands) over which communication takes place. Though the users are allocated a licensed channel, it has been found that in practical systems, users always do not have data to send; so the licensed channel is unused most of the times. This has motivated the FCC to propose a system wherein the unused channel could be utilized by users who really need them, thereby making efficient reuse of the licensed spectrum. This has lead to the evolution of cognitive radios.

In this chapter we study such a communication scenario in which one primary (licensed) and one secondary (unlicensed) user wish to communicate to different receivers under a quality of service (QoS) constraint as was done in Chapter IV. The QoS requirement is specified in terms of asymptotic decay rate of buffer occupancy. Two schemes are proposed by which such a system could operate from a link layer perspective and the admissible rate tuple under QoS constraint is characterized. In the first scheme, the cognitive radio is modeled as a strict priority queuing system, wherein the secondary user is serviced only when the primary user has no data to transmit. In this scheme the service provided to the secondary user depends on the input arrival rate of the primary user aside from the channel service rate. In the second scheme, the cognitive radio is modeled as a non-strict priority queuing system, wherein the primary and the secondary users are simultaneously provided service (i.e., there is no dependence on when service is provided to the secondary user on the primary user, however the amount of service provided to the primary and the secondary user can vary).

A. Introduction

As the licensed spectrum is under-utilized most of the times, the FCC recently recommended that spectral efficiency could be enhanced by deploying additional wireless devices (i.e., secondary users) that can coexist with the primary users. A natural way by which such a system could function is that a secondary user could transmit data whenever the primary user does not transmit. This further implies that the secondary user must be able to sense to a reasonable level of accuracy, the presence/absence of the primary user. Misdetection by the secondary user potentially interferes with the primary user which is undesirable. False alarms by the secondary user could lead to decreased spectral efficiency. The research in [53], [54] addresses these issues. For the purpose of this chapter, we assume that a perfect sensing algorithm is in place.

Prior research on these systems were more from a physical layer perspective. For example, achievable rates of cognitive radio channels are discussed in [55] and [56] addresses maximum rates that primary and secondary users can get under the stipulation that they coexist.

In this chapter, we take a totally different point of view on the cognitive radio system by looking at it from a link layer perspective. We assume that there is one primary and secondary user and that each user is endowed with an infinite length buffer. Furthermore, there is a QoS constraint (as discussed in Chapter IV) denoted by (θ_1, θ_2), where θ_i $(i = 1, 2)$ is the asymptotic decay rate of buffer occupancy of the primary and the secondary user respectively. The input arrival rates into each buffer is assumed to be constant. The amount of packets taken out from each of the user's buffer is a function of whether the user is serviced and the state of the channel. The primary question addressed in this chapter is the following: Given the QoS constraint and the service process of the channel what is the maximum admissible rate when:

- Secondary user is serviced only when the primary user does not have any data to transmit.
- Both primary and secondary users are serviced simultaneously with varying service rates.

B. Statement of the Problem

The systems of interest are shown in Fig. 22 and Fig. 23. In system-I (shown in Fig. 22), the cognitive radio is modeled as a strict priority system. In this model, the primary and secondary users are equipped with separate infinite buffers into which data arrives at rates r_1 and r_2 respectively. The primary and secondary users are serviced by a wireless channel which is modeled as an 'ON-OFF' process (i.e., Gilbert-Elliott model). Accordingly, the service rate of the channel is c when the channel is in the 'ON' state, and zero when the channel is in the 'OFF' state. The transition rate from 'ON' to 'OFF' is α, while the transition rate from 'OFF' to 'ON' is denoted by β. As long as the there is data to be transmitted by the primary user, the primary user is always given priority over the secondary (cognitive) user. The secondary user is serviced only when the primary user does not have anything to send (i.e., when the primary user's buffer is zero). Let $X(t)$ and $Y(t)$ denote the buffer contents of the primary and secondary user respectively at time t. The queue length process $X(t)$ and $Y(t)$ can be shown to converge in distribution to random variables $X(\infty)$ and $Y(\infty)$ such that the following holds:

$$-\lim_{x\to\infty}\frac{\log(\Pr\{X(\infty)>x\})}{x}=\theta_1$$

and

$$-\lim_{y\to\infty}\frac{\log(\Pr\{Y(\infty)>y\})}{y}=\theta_2$$

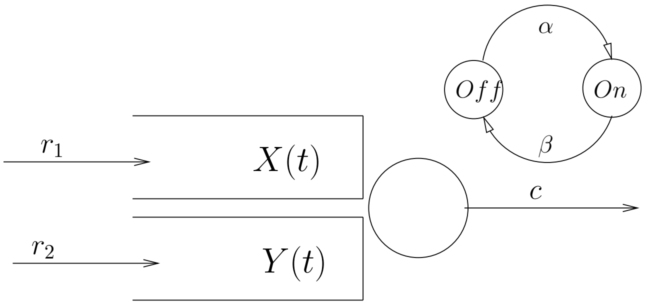


Fig. 22. Priority based queuing model of a cognitive radio system (system-I).

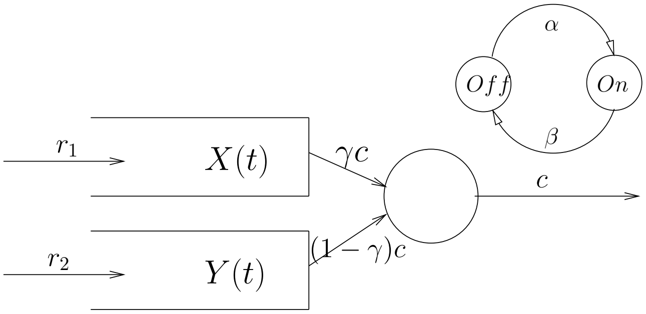


Fig. 23. Non-priority based queuing model of a cognitive radio system (system-II).

θ_1, θ_2 are sometimes called the QOS exponents [52] of the system, which depends on the arrival rates r_1 and r_2 . The higher the value of $\theta_i, (i = 1, 2)$, the more stringent is the QoS requirement. Given (θ_1, θ_2), we seek the maximum admissible rate tuple (r_1, r_2) so as to fulfill the QoS exponent criterion. In system-II (shown in Fig. 23), we again have separate infinite buffers for the primary and secondary user into which data arrives at rates r_1 and r_2 respectively. Both the primary and secondary user are serviced by a wireless channel which is modeled similar to the one detailed in the system-I above. Unlike in system-I, where the secondary user is served only when there is no data to be transmitted by the primary user, in this case, both primary and the secondary users are served simultaneously by varying the level of service rate $\gamma \in (0, 1)$ such that primary and secondary users are serviced by rates γc and $(1-\gamma)c$

respectively when the channel is in the 'ON' state as shown in Fig. 23. Here again we seek the rate tuple (r_1, r_2) such that the QoS exponent criterion is satisfied. In what follows, we derive bounds on the arrival rates for system-I and system-II.

C. Arrival Bounds on System-I

We first consider the model shown in Fig. 22. Stability of the system demands that $r_1 + r_2 < c\frac{\alpha}{\alpha+\beta}$.

Lemma V.1. *The systems shown in Fig. 22 and Fig. 24 are equivalent.*

Proof. First, consider the system shown in Fig. 22. Let $S_1(t)$ denote the state of the channel (i.e., $S_1(t) = 0$ denotes the channel is 'OFF', while $S_1(t) = 1$ denotes the channel is 'ON'). The evolution of the buffer contents $X(t), Y(t)$ at time t can then be written as:

$$\begin{gathered} \frac{dX(t)}{dt} = \begin{cases} r_1 & S_1(t) = 0 \\ r_1 - c & S_1(t) = 1, X(t) > 0 \\ 0 & S_1(t) = 1, X(t) = 0 \end{cases} \\ \frac{dY(t)}{dt} = \begin{cases} r_2 & S_1(t) = 0 \\ r_2 - (c - r_1) & S_1(t) = 1, X(t) = 0, Y(t) > 0 \\ 0 & S_1(t) = 1, X(t) = 0, Y(t) = 0 \end{cases} \end{gathered} \tag{5.1}$$

Now, consider the system shown in Fig. 24. Let $S_2(t)$ denote the state of the equiv-

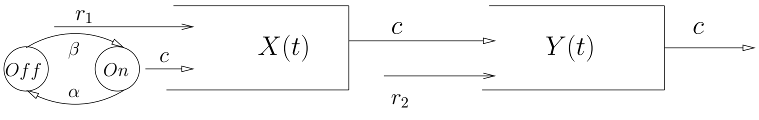


Fig. 24. Equivalent of system model-I of a cognitive radio.

alent source at the input of the high priority buffer. The equivalent source produces data at rate c when $S_2(t) = 1$ and zero when $S_2(t) = 0$. The evolution of the buffer contents $X(t), Y(t)$ for this case can be written as:

$$\frac{dX(t)}{dt} = \begin{cases} r_1 - c & S_2(t) = 0, X(t) > 0 \\ 0 & S_2(t) = 0, X(t) = 0 \\ r_1 & S_2(t) = 1 \end{cases}$$

$$\frac{dY(t)}{dt} = \begin{cases} r_2 - (c - r_1) & S_2(t) = 0, X(t) = 0, Y(t) > 0 \\ 0 & S_2(t) = 0, X(t) = 0, Y(t) = 0 \\ r_2 & S_2(t) = 1 \end{cases} \tag{5.2}$$

From (5.1) and (5.2), it is clear that the evolution of the buffers are the same when $S_2(t)$ in (5.2) is 'ON', while $S_1(t)$ in (5.1) is 'OFF' and vice-versa. Hence, the system in Fig. 22 with a markov modulated service wireless channel with transition rate parameters (α, β), is equivalent to a system with a *compensating source* as shown in Fig. 24, having parameters (β, α). □

This equivalence makes it easier to use the well developed theory of effective bandwidth [29], [57], to get bounds on the arrival rates for the problem posed in Fig. 22.

1. Bounds on Arrival Rates

Having established the equivalence, we henceforth focus on the system depicted in Fig. 24. The admission criterion using the theory of effective bandwidth demands that the following be satisfied [57]. For the primary user:

$$r_1 + eb_1(\theta_1) < c \tag{5.3}$$

where $eb_1(\cdot)$ is the effective bandwidth of the equivalent source which was compensated for the wireless channel given by

$$eb_1(\theta_1) = \frac{1}{2\theta_1}\big(c\theta_1 - \alpha - \beta + \sqrt{(c\theta_1 - \alpha - \beta)^2 + 4\beta c\theta_1}\big) \tag{5.4}$$

[57], [58], and θ_1 is the QoS exponent requirement of the primary user's buffer. As mentioned previously, the secondary user's QoS also depends on the QoS requirement of the primary user. Accordingly, it can be shown that the following condition needs to be satisfied for the secondary user's buffer [57]:

$$r_2 + eb_s(\theta_2) < c \tag{5.5}$$

where $eb_s(\cdot)$ denotes the effective bandwidth of the departure process of the primary user's buffer (which is also equal to the effective bandwidth of the arrival process into the secondary user's buffer), which is given by [58]:

$$eb_s(\theta_2) = \begin{cases} r_1 + eb_1(\theta_2) & 0 \le \theta_2 \le \theta^* \\ c - \frac{\theta^*}{\theta_2}\big(c - (r_1 + eb_1(\theta^*))\big) & \theta_2 > \theta^* \end{cases} \tag{5.6}$$

where θ^* is the solution to,

$$\frac{d}{d\theta}\big(r_1\,\theta + \theta\, eb_1(\theta)\big) = c \tag{5.7}$$

[58], and, $eb_1(\theta)$ is given by (5.4). From (5.7) one obtains $\theta^* = \frac{\alpha-\beta}{c} + \frac{c-2r_1}{c}\sqrt{\frac{\alpha\beta}{r_1(c-r_1)}}$. Hence the admissible rate tuple (r_1, r_2) can be obtained from (5.3) and (5.5). It is important to note that (5.5) depends on r_1 as well. The admissible rate tuple (r_1, r_2) is shown in Fig. 25 for QoS exponent $(\theta_1 = 8, \theta_2 = 4)$. Without loss of generality, we take the service rate of the channel when it is 'ON' to be $c = 1$. It is not surprising to see that the channel correlation $(\alpha + \beta)$, does have an impact on the admissible rates. The waiting time distribution of bits for the primary and the secondary user

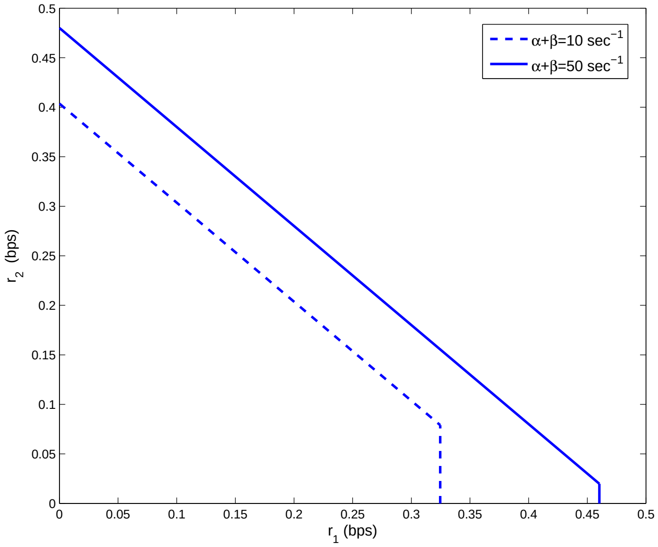


Fig. 25. Admissible rates for different channel correlation rates $(\alpha + \beta)$ with $(\theta_1 = 8, \theta_2 = 4)$ and $c = 1$.

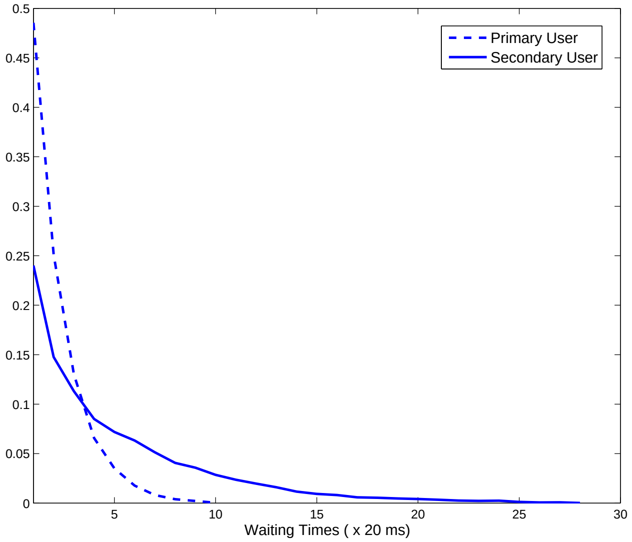


Fig. 26. Waiting time distribution of bits of the primary and secondary user for system-I when $\theta_1 = 8,\ \theta_2 = 4$.

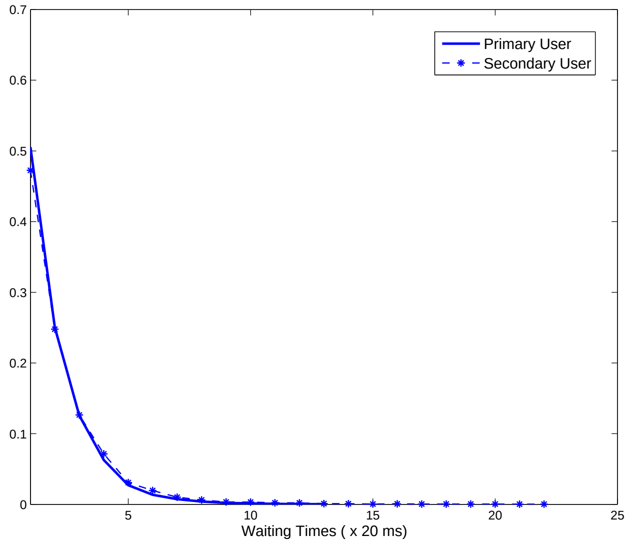


Fig. 27. Waiting time distribution of bits of the primary and secondary user for system-I when $\theta_1 = 50,\ \theta_2 = 40$.

for various values of (θ_1, θ_2) are depicted in Fig. 26 and Fig. 27 for $(\alpha + \beta) = 50$ sec^{-1}.

2. Finite Buffer Case

Another way to think of this problem formulation is to seek an admissible rate tuple (r_1, r_2) such that $\mathcal{E}X \leq C_1$ and $\mathcal{E}Y \leq C_2$, where $\mathcal{E}X, \mathcal{E}Y$ represents the steady state expected buffer contents of the high priority (primary user) and low priority buffer (secondary user's buffer) respectively and C_1, C_2 are arbitrary constants which may depend on the buffer capacity of the primary and the secondary user.

Hence the problem reduces to finding the expected value of the buffer contents at steady state. We use tools from *poisson driven stochastic differential equations* to be able to solve the problem at hand. The derivation presented here proceeds along similar lines to the ones in [30] and [59].

An ON-OFF Markov modulated process can be represented by:

$$dp(t) = (1-p)dN_1 - pdN_2$$

where $p(t) \in \{0,1\}$ and N_1, N_2 are counters of rate β and α respectively. Then, the evolution of the primary and secondary user's buffers can be written as:

$$\begin{aligned} dX &= -c\mathcal{I}_X dt + cpdt + r_1 dt \\ dY &= -c\mathcal{I}_Y dt + c\mathcal{I}_X dt + r_2 dt \end{aligned} \tag{5.8}$$

where

$$\mathcal{I}_X = \begin{cases} 1 & X > 0 \\ 0 & \text{otherwise} \end{cases}$$

By using the Ito's rule for jump process [30], we obtain the following:

$$\begin{aligned} dX^2 &= (-2cX\mathcal{I}_X + 2cpX + 2Xr_1)dt \\ dY^2 &= (-2cY\mathcal{I}_Y + 2cY\mathcal{I}_X + 2Yr_2)dt \end{aligned}$$

Using the method to calculate expectations from [30], we get:

$$\begin{aligned} \frac{d}{dt}\big(\mathcal{E}X^2\big) &= 2(r_1 - c)\mathcal{E}X + 2c\mathcal{E}(pX) \\ \frac{d}{dt}\big(\mathcal{E}Y^2\big) &= 2(r_2 - c)\mathcal{E}Y + 2c\mathcal{E}(\mathcal{I}_X Y) \end{aligned} \tag{5.9}$$

Noting that at steady state $\frac{d}{dt}\big(\mathcal{E}X^2\big) = \frac{d}{dt}\big(\mathcal{E}Y^2\big) = 0$, we obtain:

$$\mathcal{E}(Y\mathcal{I}_X) = \frac{c - r_2}{c}\mathcal{E}Y \tag{5.10}$$

$$\mathcal{E}(pX) = \frac{c - r_1}{c}\mathcal{E}X \tag{5.11}$$

Now, we have the following:

$$d(XY) = XdY + YdX$$

$$\begin{aligned}
&= (-cX\mathcal{I}_Y + cX\mathcal{I}_X + Xr_2 - cY\mathcal{I}_X + cYp + r_1Y)dt \\
d(pX) &= pdX + Xdp \\
&= (-cp\mathcal{I}_X + cp^2 + r_1p + X(1-p)dN_1 - XpdN_2)dt \qquad (5.12) \\
\frac{d}{dt}\big(\mathcal{E}XY\big) &= (c+r_2)\mathcal{E}X - c\mathcal{E}(Y\mathcal{I}_X) - c\mathcal{E}(X\mathcal{I}_Y) + c\mathcal{E}(pY) + r_1\mathcal{E}Y \\
\frac{d}{dt}\big(\mathcal{E}(pX)\big) &= -c\mathcal{E}(p\mathcal{I}_X) + (c+r_1)\frac{\beta}{\alpha+\beta} + \beta\mathcal{E}X - \mathcal{E}(pX)(\alpha+\beta) \\
&\overset{\text{(a)}}{=} r_1\frac{\beta}{\alpha+\beta} + \beta\mathcal{E}X - \mathcal{E}(pX)(\alpha+\beta)
\end{aligned}$$

where (a) follows from the fact that $\mathcal{E}(p\mathcal{I}_X) = \mathcal{E}X$. This is because, when $p(t) = 1$, the primary user's buffer content is always positive as the input rate $(c+r_1)$ is greater than the service rate c. Again by making use of the fact that $\frac{d}{dt}\big(\mathcal{E}(Xp)\big) = 0$ and from (5.10), we get the expected value for the primary user's content to be:

$$\mathcal{E}X = \frac{r_1c}{c\alpha - r_1(\alpha+\beta)}\frac{\beta}{\alpha+\beta} \qquad (5.13)$$

Along similar lines to the idea presented above we obtain the following:

$$\frac{d}{dt}\mathcal{E}(pY) = r_2\frac{\beta}{\alpha+\beta} + \beta\mathcal{E}Y - \mathcal{E}(pY)(\alpha+\beta)$$

By equating $\frac{d}{dt}\mathcal{E}(pY) = 0$, we have

$$\mathcal{E}(pY) = \frac{r_2\beta}{(\alpha+\beta)^2} - \frac{\beta}{\alpha+\beta}\mathcal{E}Y \qquad (5.14)$$

Making use of the fact that $X > 0$ implies that $Y > 0$ (this is because, when $X > 0$, the service rate of the primary user's buffer will be c, which means that the input rate to the secondary user's buffer will be at rate $r_2 + c$, which is greater than the service rate c of the secondary user's buffer) which implies $\mathcal{E}(X\mathcal{I}_Y) = \mathcal{E}X$ and from (5.10), (5.12), (5.13) and (5.14) we get the expected value of the secondary user's

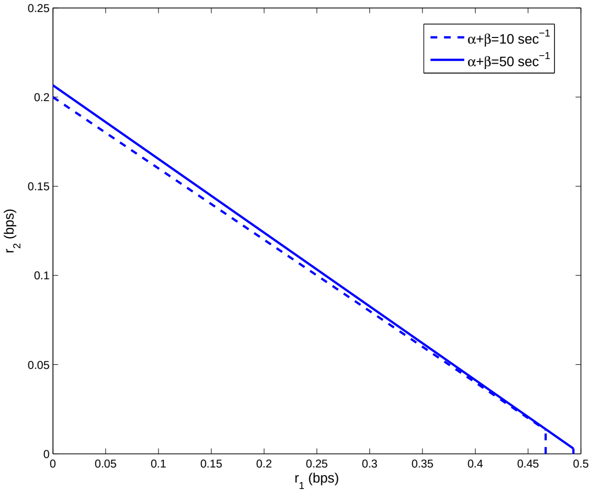


Fig. 28. Bounds on the admissible rates for system-I for the finite buffer case when $C_1 = 0.7$, $C_2 = 0.5$ and $c = 1$.

buffer content to be:

$$\mathcal{E}Y = \frac{1}{(\alpha+\beta)}\Big[\frac{c\,r_2\,\beta + (\alpha+\beta)^2\,r_2\,\mathcal{E}X}{c\,\alpha - (r_1+r_2)(\alpha+\beta)}\Big] \tag{5.15}$$

Hence from (5.13) and (5.15), the admissible rate tuple (r_1, r_2) which satisfies $\mathcal{E}X \leq C_1$ and $\mathcal{E}Y \leq C_2$ can be determined.

The admissible rate is shown in Fig. 28, when $C_1 = 0.7$, $C_2 = 0.5$ and when the service rate of the channel when 'ON' is $c = 1$.

D. Arrival Bounds on System-II

We now consider the system model depicted in Fig. 23. It can be easily noticed that once the fraction of service rate ($\gamma_1 = \gamma, \gamma_2 = 1-\gamma$) are fixed, the buffers can be analyzed independently. As done in Lemma V.1, Section C, we can establish the equivalence of systems in Fig. 23 and Fig. 29 with $\gamma_1 = \gamma$ and $\gamma_2 = 1-\gamma$. Based on

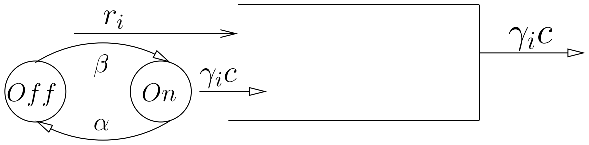


Fig. 29. Equivalent of system model-II (shown in Fig. 23) of a cognitive radio with $\gamma_1 = \gamma$ and $\gamma_2 = 1 - \gamma$.

the QoS exponents (θ_1, θ_2), the admissible rate tuple (r_1, r_2) turns out to be:

$$\begin{array}{rcl} r_1 + eb_1(\theta_1) & < & \gamma c \\ r_2 + eb_2(\theta_2) & < & (1-\gamma)c \end{array}$$

which yields:

$$\begin{array}{rcl} r_1 & < & \gamma c - eb_1(\theta_1) \\ r_2 & < & (1-\gamma)c - eb_2(\theta_2) \end{array} \tag{5.16}$$

where, in this case $eb_1(\theta_1) = \frac{1}{2\theta_1}\Big(\gamma c\theta_1 - \alpha - \beta + \sqrt{(\gamma c\theta_1 - \alpha - \beta)^2 + 4\beta\gamma c\theta_1}\Big)$ and $eb_2(\theta_2) = \frac{1}{2\theta_2}\Big((1-\gamma)c\theta_2 - \alpha - \beta + \sqrt{((1-\gamma)c\theta_2 - \alpha - \beta)^2 + 4\beta(1-\gamma)c\theta_1}\Big)$. The admissible rate tuple (r_1, r_2) is shown in the Fig. 30 and the waiting time distribution when $\gamma = 0.5$, $\theta_1 = \theta_2 = 50$ is shown in Fig. 31.

E. Discussions

The numerical results given in Fig. 25 and Fig. 30 for system-I and system-II respectively suggests that when the QoS constraints are loose (i.e., θ_1, θ_2 are low), the throughput provided by system-I is roughly the same as that of system-II for most of the rate tuples. On the other hand, when the QoS constraints become tight (i.e,. θ_1, θ_2 are high), the throughput of system-II is better than that of system-I for most of the useful operating points as shown in Fig. 32. Hence, from a practical perspective, it may be beneficial to operate a cognitive radio system by giving access to both

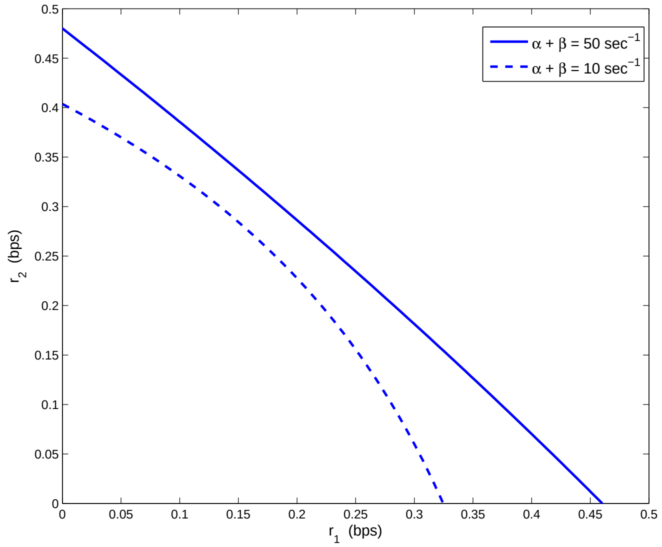


Fig. 30. Bounds on the admissible rates for system-II (shown in Fig. 23), when ($\theta_1 = 8$, $\theta_2 = 4$) and $c = 1$.

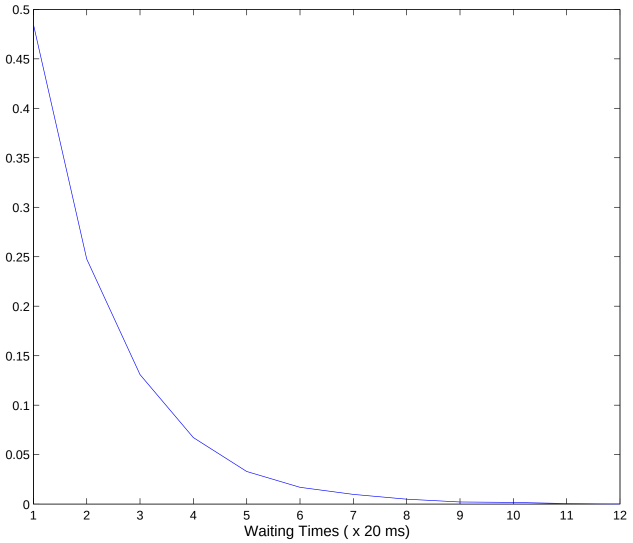


Fig. 31. Waiting time distribution of bits of the primary and secondary user for system-II when $\theta_1 = \theta_2 = 50$ and $\gamma = 0.5$.

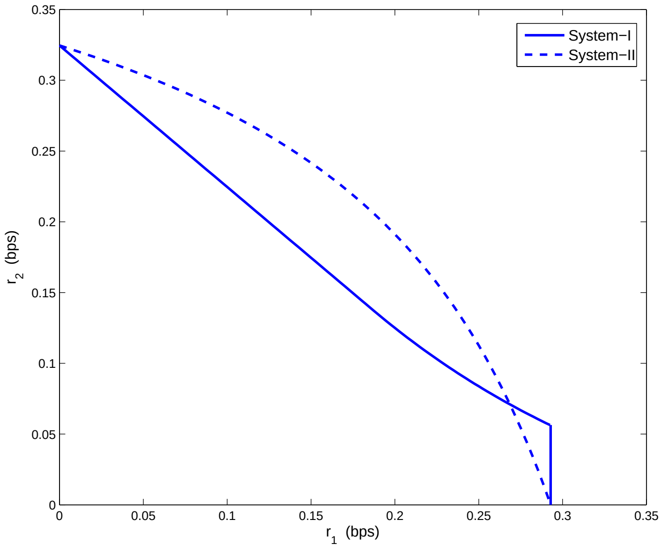


Fig. 32. Comparison of admissible rates of system-I and system-II when $\theta_1 = 50$, $\theta_2 = 40$ and $\alpha + \beta = 50\,\text{sec}^{-1}$.

primary and secondary users simultaneously by varying their service rates appropriately. Furthermore, for the finite buffer case, the numerical results in Fig. 28 suggests that the impact of channel correlation on the throughput of the users becomes less significant as compared to the case when users are endowed with infinite length buffer shown in Fig. 25 and Fig. 30.

F. Conclusion

In this chapter we have characterized the admissible region of a cognitive radio system by modeling it as a strict priority system (wherein the secondary user is serviced only when the primary user has nothing to transmit) and a non-strict priority system (wherein service is provided simultaneously to both primary and the secondary user by varying their service rates). The results of this chapter suggests that operating the cognitive radio system by providing access to both primary and secondary users by varying their service rates appropriately is beneficial for most operating points of interest.

CHAPTER VI

ON THE SECRECY CAPACITY OF A SOURCE CODING PROBLEM

In previous chapters, we have looked into problems wherein information needed to be communicated reliably. In this chapter we take a different point of view in that the information not only needs to be communicated reliably but also in a secure manner. We study the 'additional resources' that are required to satisfy the constraint of 'secrecy' in this chapter.

More specifically, we treat the problem of communicating multiple sources to $(L + 1)$ receivers amidst an eavesdropper. A direct channel is available from the source encoder to each of the L receivers, while the $(L + 1)^{th}$ receiver has access to the outputs of L channels (i.e., there is no direct channel from the source encoder to the $(L + 1)^{th}$ receiver). The eavesdropper has access to a subset of size m (where $m < L$), and the legitimate receiver should be able to reconstruct the source from any subset of size n (where $m < n \leq L$). The goal of the source encoder is to encode the sources in such a way so as to communicate the sources reliably to the $(L + 1)^{th}$ receiver, with minimum overhead. The rate region of such a system is found and achievability is established using simple linear network codes.

A. Introduction

In this chapter, we first consider the following communication problem: An information source needs to be sent from P to Q. We assume that there is no direct way of communicating from P to Q, however there are L "agents" that can help carry the information, to which both P and Q have access to. Depending on the "reliability" that P has on each of the agents, the information is securely encoded into L packets, possibly of different size, with each agent carrying a packet, with the additional stipulation that Q should understand completely, whatever P sends. The goal of P is then to encode the information source to the fidelity requirements of Q, with minimum overhead, by taking into account the following situations:

- The agents clandestinely collaborate to pool their information, in which case the agents might be able to decode partly/completely, though they may not be able to decode individually.
- Q be able to decode completely from a subset of agents whose size is larger than the size of the coalition that the agents make to clandestinely decode the data.

One can think of the situation mentioned above with P being a source encoder, Q being a legitimate receiver and the "agents" as the eavesdroppers.

Alternately, one could think of the aforementioned situation wherein there are $(L+1)$ receivers, out of which L receivers each has a direct communication channel from the source encoder, while the $(L+1)^{th}$ receiver (i.e., henceforth the legitimate receiver) has access to L channels, though there is no direct channel from it to the source encoder. The eavesdropper has access to a subset of m channels, and it is desired that the legitimate receiver be able to completely decode the information from a subset of size n of the L-direct channels (i.e., $m < n \leq L$). We will henceforth use this interpretation in this chapter.

Our main result is an exact characterization of the rate region for simple cases of the problem considered above. For more general cases, we characterize the exact sum rate and construct achievable schemes by using simple linear network codes which are shown to provide the desired level of secrecy. A central theme which emerges in showing the achievability is the following: Any "partial" secure code is constructed by simply concatenating a "fully secure" code and a "totally insecure" code.

Next, we consider the problem of communicating multiple sources to a receiver amidst an eavesdropper. There are totally $L-m$ sources to be communicated to the receiver with the stipulation that the receiver needs to be able to reconstruct sources $\{1, 2 \ldots k\}$, by accessing any subset of $(m+k)$ channels $(1 \leq k \leq L-m)$. We obtain the coding rate region for the case when $L = 3, m = 1$, and show that separate secure encoding of sources is optimal for this problem. For the case of arbitrary m, L we derive a lower bound on the sum rate and the result implies a separation scheme to achieve this lower bound.

Notable results in this topic can be found in [60], [61], [62], [63], [64] and [65].

We have organized the results in this chapter as follows. In Section B, we present the problem and the results for the single source case. These results are extended to multiple sources in Section C. Finally, we present the conclusion in Section D.

B. The Single Source Problem with Eavesdropper and Diversity Coding

The problem considered in this section is shown in Fig. 33. A source encoder is presented with a sequence of i.i.d source letters $\{X_k\}$ drawn from the alphabet $\mathfrak{X}$. For each block of N letters (N arbitrary), there are L outputs, $f_l(X^N) = W_l \in \{1, 2 \ldots 2^{NR_l}\}$ for $l = 1, 2 \ldots L$. The codeword W_l is sent through the l^{th} $(l = 1, 2 \ldots L)$ communication channel. The eavesdropper has access to m channels $(m < L)$, while the legitimate receiver (i.e., Receiver-0 in Fig. 33) has access to all the L channels. The source is to be reconstructed at the legitimate receiver with a distortion of D_0 by accessing any subset of n channels. To make this problem interesting, we need to have $m < n \leq L$. We denote $R(\mathrm{D}_0)$ to be minimum number of bits required to reconstruct the source with distortion D_0 [66]. Let $\mathfrak{B}$ denote the collection of all subsets of channels that the legitimate receiver may be able to reconstruct the source

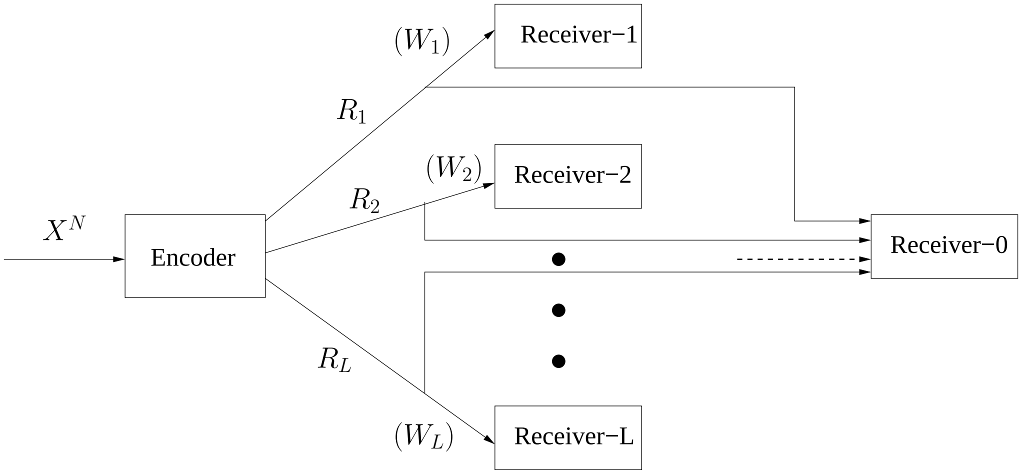


Fig. 33. The single-source problem considered in this chapter. Receiver-0 is the legitimate receiver. Codeword sent on channel l is denoted by $W_l \in \{1, 2 \ldots 2^{nR_l}\}$ $(l = 1, 2 \ldots L)$.

from. Clearly, $|\mathfrak{B}| = \binom{L}{n}$, where $|\mathfrak{B}|$ represents the cardinality of set $\mathfrak{B}$. Let $\mathfrak{A}$ denote the collection of all subsets of channels that the eavesdropper has access to.

For any $\mathcal{B} \in \mathfrak{B}$, consider all collection of subsets of size m and denote it by $\Upsilon\big(\mathcal{B}\big)$. Clearly,

$$\mathfrak{A} = \bigcup_{\forall\, \mathcal{B} \in \mathfrak{B}} \Upsilon\big(\mathcal{B}\big)$$

We assume that the eavesdropper can access any member of $\mathfrak{A}$, but no more than one member of $\mathfrak{A}$. For $A \in \mathfrak{A}$, (and hence $A \subset \mathcal{B}$ for one or more $\mathcal{B} \in \mathfrak{B}$) let Y_A be the codeword transmitted on the channel corresponding to the elements in A. It will be convenient to think of Y_A as a vector whose j^{th} element denotes the codeword $\big($i.e., W_l $(l \in \{1, 2 \ldots L\})\big)$ sent on the channel corresponding to the j^{th} element of A.

We define the equivocation rate (i.e., level of secrecy) when the eavesdropper has access to any $A \in \mathfrak{A}$ to be:

$$\Delta_A = \frac{1}{N} H(X^N | Y_A) \tag{6.1}$$

and it is desired that $\Delta_A \geq h_A$, for a given h_A.

To clarify the notations introduced, we give an example here. For the case, when $L = 4, n = 3, m = 2$, we have $\mathfrak{B} = \big\{(1,2,3),(1,2,4),(1,3,4),(2,3,4)\big\}$ and $\mathfrak{A} = \big\{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\big\}$. For $A = (1,3)$, $Y_A = \left[\begin{array}{cc} W_1 & W_3 \end{array}\right]$.

The primary problem addressed in this section is the following:

- For the case when $h_A = h$ and $\Delta_A \geq h$, $\forall A \in \mathfrak{A}$ and for any $\mathcal{B} \in \mathfrak{B}$, we characterize the sum-rate and propose simple schemes which achieve the lower bound.

The case when $n = L$ can be thought of the situation wherein the legitimate receiver needs to access all the channels to reconstruct the source, while for the case when $n < L$, we impose a diversity coding from the legitimate receiver perspective.

Before we begin addressing the aforementioned problems, we characterize the rate region and achievability for a few simple representative cases with asymmetrical secrecy requirements as well, which gives considerable insight for solving the more generalized problem mentioned above. We consider the following scenarios:

- $m = 1$ and $n = L = 2$.
- $m = 2$ and $n = L = 3$.
- $m = 1$,$n = 2$ and $L = 3$.

1. $m = 1$, $n = L = 2$.

In this case, we have, $\mathfrak{A} = \{\{1\},\{2\}\}$ and $\mathfrak{B} = \{(1,2)\}$. The desired level of secrecy required by each channel for a given h_1, h_2 is given by:

$$\frac{1}{N} H(X^N|W_i) \geq h_i \quad \text{for } i = 1, 2. \tag{6.2}$$

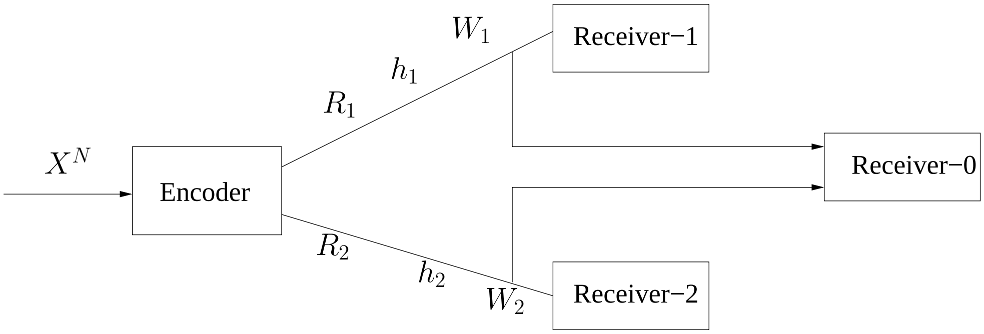


Fig. 34. The case when $m = 1$ and $n = L = 2$. The secrecy constraints on the two channels are given by h_1 and h_2.

where, W_i is the codeword sent along channel i. The goal is to determine (R_1, R_2, h_1, h_2) with a distortion requirement of D_0 desired by the legitimate receiver, such that (6.2) is satisfied. See also Fig. 34.

Theorem VI.1. *The rate region* (R_1, R_2, h_1, h_2)*, for a distortion requirement of* D_0 *desired by the legitimate receiver, is given by the set of rate pairs* (R_1, R_2) *satisfying the following:*

$$R_1 \geq \left(h_2 - [H(X) - R(D_0)]\right)^+ = Z_1 \tag{6.3}$$

$$R_2 \geq \left(h_1 - [H(X) - R(D_0)]\right)^+ = Z_2 \tag{6.4}$$

$$R_1 + R_2 \geq R(D_0) \tag{6.5}$$

where, $(x)^+ = \max(x, 0)$

Proof. Refer the generalized proof given in Theorem VI.5 for (6.3) and (6.4). (6.5) follows naturally. □

Remark 1. We can explain intuitively as to why (6.3) or (6.4) should hold. Suppose, $(H(X) - h_2)$ is the 'effective' information bits (i.e., after taking into account the secrecy level of channel 2) that is sent through channel 2 . Then, it is not difficult

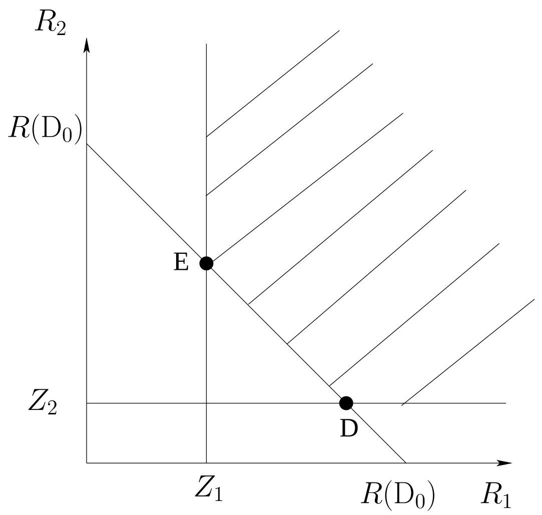


Fig. 35. The rate region for the case when $Z_1 + Z_2 \leq R(\mathrm{D}_0)$.

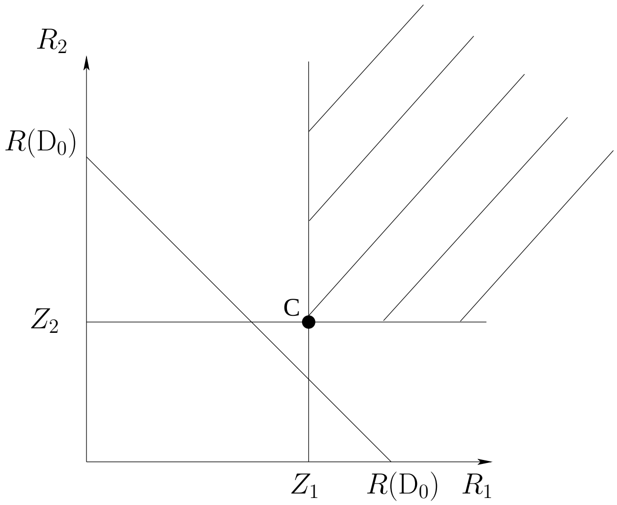


Fig. 36. The rate region for the case when $Z_1 + Z_2 > R(\mathrm{D}_0)$.

to see that the rest of the information bits $\big($i.e., $R(\mathrm{D}_0) - [H(X) - h_2]\big)$ has to be at least supplied by channel 1 for (6.5) to be satisfied, which is precisely (6.3).

a. Achievability of Theorem VI.1

Depending on the secrecy requirements of the channels, we have two cases:

<u>**Case 1:**</u> $(Z_1 + Z_2) \leq R(\mathrm{D}_0)$

This case (shown in Fig.35) arises when the security requirements (h_1, h_2) of both channels are small. It is clear from Fig.35, that we need only to show achievability of points D and E as the other points along the line DE can be achieved by time sharing between points D and E. We will show this only for point D as similar ideas can be used for showing the achievability of point E.

- Source coding (i.e., rate distortion coding) is first performed. Then, the $R(\mathrm{D}_0)$ information bits are split into two parts such that $R_1 = (H(X) - h_1)$ bits are sent through channel 1, while the rest of the information bits $R_2 = Z_2$ are sent through channel 2.

It only remains to show that point D (i.e., a split of $\big(R_1 = (H(X) - h_1),\, R_2 = Z_2\big)$ in Fig.35 satisfies the desired secrecy requirements, which is proved in the following lemma.

Lemma VI.1. *The assignment of* $\big(R_1 = (H(X) - h_1),\, R_2 = Z_2\big)$ *bits (i.e., point D in Fig.35) satisfies the secrecy requirements, in (6.2).*

Proof.

$$\begin{aligned}
\frac{1}{N}H(X^N|W_1) &= \frac{1}{N}\big[H(X^N, W_1) - H(W_1)\big] \\
&\geq \frac{1}{N}\big[H(X^N) - H(W_1)\big] \\
&\geq \big[H(X) - R_1\big] \\
&\geq h_1
\end{aligned}$$

Following similar steps as above, we get:

$$\frac{1}{N}H(X^N|W_2) \geq \big[H(X) - R_2\big] \overset{(b)}{\geq} h_2$$

where, (b) follows from the fact that $Z_1 + Z_2 \leq R(\mathrm{D}_0)$. □

<u>Case 2:</u> $(Z_1 + Z_2) > R(\mathrm{D}_0)$

This case arises (shown in Fig.36) when the secrecy requirements of the channels are high. As shown in Fig.36, we clearly need more than $R(\mathrm{D}_0)$ bits in this regime. We need to show only the achievability of point C in Fig.36 $\big($i.e., $\big(R_1 = Z_1, R_2 = Z_2\big)\big)$, which follows:

1. Generate $R_k = Z_1 + Z_2 - R(\mathrm{D}_0) = h_1 + h_2 - 2H(X) + R(\mathrm{D}_0)$ bits *independent* of the $R(\mathrm{D}_0)$ information bits. We call these *key* bits. We denote the key bits

by $\boldsymbol{\mathcal{K}}$, whose length is R_k.

2. Denote the information bits by $\{I_i\}$ for $i = 1, 2, \ldots R(\mathrm{D}_0)$, and let $\ell_1 = H(X) - h_1$ and $\ell_2 = H(X) - h_2$. Let $\boldsymbol{\mathcal{I}_s} = \{I_i\}_{i=1}^{R_k}$, $\boldsymbol{\mathcal{I}_1} = \{I_i\}_{R_k+1}^{R_k+\ell_1+1}$ and $\boldsymbol{\mathcal{I}_2} = \{I_i\}_{R_k+\ell_1+2}^{R_k+\ell_1+\ell_2+2}$.

3. Mask $\boldsymbol{\mathcal{I}_s}$ with $\boldsymbol{\mathcal{K}}$ and concatenate it with $\boldsymbol{\mathcal{I}_1}$ and place this in channel 1. Concatenate $\boldsymbol{\mathcal{K}}$ with $\boldsymbol{\mathcal{I}_2}$ and place it in channel 2. Hence, our code looks like:

$$\begin{array}{lllll} \mathbf{C}_1 : & [& \boldsymbol{\mathcal{I}_1} & \boldsymbol{\mathcal{I}_s} \oplus \boldsymbol{\mathcal{K}} &] \\ \mathbf{C}_2 : & [& \boldsymbol{\mathcal{I}_2} & \boldsymbol{\mathcal{K}} &] \\ & & \underbrace{\qquad\qquad}_{\text{Non-Secure Code}} & \underbrace{\qquad\qquad}_{\text{Perfectly Secure Code}} & \end{array} \tag{6.6}$$

It is easy to see that the code construction in (6.6) meets the secrecy requirements in (6.2) and also satisfies (6.3) and (6.4) with equality.

Remark 2. Alternately, (6.6) could have been constructed by placing $\boldsymbol{\mathcal{K}}$ in $\mathbf{C_1}$ and $\boldsymbol{\mathcal{I}_s} \oplus \boldsymbol{\mathcal{K}}$ in $\mathbf{C}_2$. For the case of perfect secrecy (i.e., $h_1 = h_2 = H(X)$), $\boldsymbol{\mathcal{K}}$ can be sent in one channel and $\boldsymbol{\mathcal{I}_s} \oplus \boldsymbol{\mathcal{K}}$ can be sent in the other channel, where $\boldsymbol{\mathcal{K}}$ and $\boldsymbol{\mathcal{I}_s}$, each are of length $R(\mathrm{D}_0)$ bits. We see that for the case of perfect secrecy, the overhead we need is at least as much as the number of information bits, which agrees with Shannon's results [14].

The above observations naturally genealize to the case when there are L receivers, with an eavesdropper having access to any one of the L channels, and the legitimate receiver be able to reconstruct the source from all the $n = L$ channels. In this case, $\mathfrak{A} = \big\{\{1\}, \{2\}, \ldots \{L\}\big\}$, $\mathfrak{B} = \big\{(1, 2 \ldots L)\big\}$.

Theorem VI.2. *The rate region $(R_1, R_2, \ldots R_L, h_1, h_2 \ldots h_L)$, for a distortion requirement of D_0 desired by the legitimate receiver, with h_i defined as in (6.2) for $i = 1, 2, \ldots L$, is given by the set of rate pairs $(R_1, R_2, \ldots R_L)$ satisfying the following:*

$$R_1 + R_2 + \ldots + R_{L-2} + R_{L-1} \geq \ \big(h_L - \big[H(X) - R(D_0)\big]\big)^+ \ = P_1$$

$$\begin{aligned} R_1 + R_2 + \ldots + R_{L-2} + R_L &\geq \big(h_{L-1} - [H(X) - R(D_0)]\big)^+ &= P_2 \\ R_1 + R_2 + \ldots + R_{L-3} + R_{L-1} + R_L &\geq \big(h_{L-2} - [H(X) - R(D_0)]\big)^+ &= P_3 \\ &\vdots & \quad (6.7) \\ R_1 + R_3 + \ldots + R_{L-1} + R_L &\geq \big(h_2 - [H(X) - R(D_0)]\big)^+ &= P_{L-1} \\ R_2 + R_3 + \ldots + R_{L-1} + R_L &\geq \big(h_1 - [H(X) - R(D_0)]\big)^+ &= P_L \\ R_1 + R_2 + \ldots + R_{L-1} + R_L &\geq R(D_0) & \quad (6.8) \end{aligned}$$

Proof. Refer the generalized proof given in Theorem VI.5 □

b. Achievability of Theorem VI.2

By adding all the L equations in (6.7) we get:

$$R_1 + R_2 + \ldots R_L \geq \frac{1}{L-1} \sum_{i=1}^{L} P_i = P \tag{6.9}$$

From (6.8) and (6.9), we see that we again have two cases to address:

<u>Case 1</u>: $P \leq R(\mathrm{D}_0)$

In this regime we have,

$$\sum_{i=1}^{L} \big(H(X) - h_i\big) \geq R(\mathrm{D}_0). \tag{6.10}$$

From a closer look of (6.10), we infer that the sum of the net 'effective' information bits (i.e., information bits after taking the secrecy constraints into account), of all the L channels is at least the required $R(\mathrm{D}_0)$ bits that the legitimate receiver needs, which can be done as was seen earlier in the two channel case (i.e., $m = 1, n = L = 2$) by splitting the bits. Clearly, there are a myriad of possibilities in this regime. A very simple scheme is as follows:

1. For any $F \subset \{1, 2 \ldots L\}$, for which $\sum_{i \in F} \big(H(X) - h_i - \epsilon_i\big) = R(\mathrm{D}_0)$, (where $\epsilon_i \geq 0 \; \forall i \in \{1, 2 \ldots L\}$), send $\big(H(X) - h_i\big)$ bits in the i^{th} channel, and none in the rest of the channels.

<u>**Case 2**</u>: $P > R(\mathrm{D}_0)$

In this case we have $\sum_{i=1}^{L} \big(H(X) - h_i\big) < R(\mathrm{D}_0)$. Furthermore, we need to have $h_i \geq \big[H(X) - R(\mathrm{D}_0)\big]$ for $i = 1, 2, \ldots L$ (see (6.7)). Suppose not, we can easily argue that there exists at least one $i \in \{1, 2, \ldots L\}$ such that $h_i \leq \big[H(X) - R(\mathrm{D}_0)\big]$, which implies that all the $R(\mathrm{D}_0)$ bits can be sent through the i^{th} channel and none through the rest of the channels, (which satisfies the secrecy constraints) and clearly makes this case trivial.

The following is an achievable scheme in the regime when $P_i > 0 \; \forall i \in \{1, 2, \ldots L\}$:

1. Generate, $R_k = P - R(\mathrm{D}_0) = \frac{1}{L-1}\big[\sum_{i=1}^{L} \big(h_i - H(X)\big) + R(\mathrm{D}_0)\big]$ key bits independent of the information bits. Denote the key bits by $\boldsymbol{\mathcal{K}}$.

2. Letting $\ell_i = H(X) - h_i$, represent L distinct set of information bits by $\boldsymbol{\mathcal{I}}_i$, each of length ℓ_i for $i = 1, 2 \ldots L$.

3. Represent another $(L - 1)$ distinct set of information bits by $\boldsymbol{\mathcal{J}}_i$ for $i = 1, 2, \ldots (L-1)$, each of length R_k, which are different from $\{\boldsymbol{\mathcal{I}}_i\}_{i=1}^{L}$. Letting $\mathbf{C}_j$ represent the code sent through channel j for $j = 1, 2, \ldots L$, we then have:

$$\mathbf{C_1} = \big[\; \boldsymbol{\mathcal{I}}_1 \;\; \boldsymbol{\mathcal{J}}_1 \oplus \boldsymbol{\mathcal{K}} \;\big], \quad \mathbf{C_2} = \big[\; \boldsymbol{\mathcal{I}}_2 \;\; \boldsymbol{\mathcal{J}}_2 \oplus \boldsymbol{\mathcal{K}} \;\big], \quad \ldots \quad \mathbf{C_L} = \big[\; \boldsymbol{\mathcal{I}}_L \;\; \boldsymbol{\mathcal{K}} \;\big].$$

It is easy to verify that (6.7) is satisfied with equality.

2. $m = 2$, $n = L = 3$

In this case, the eavesdropper has access to any two out of total of three channels and hence we have $\mathfrak{A} = \{(1,2),(1,3),(2,3)\}$, $\mathfrak{B} = \{(1,2,3)\}$. The desired level of secrecy for a given h_{12}, h_{13}, h_{23} are given by:

$$\frac{1}{N}H(X^N|W_i, W_j) \geq h_{ij} \quad 1 \leq i < j \leq 3 \tag{6.11}$$

where W_i is the codeword sent along channel i.

Theorem VI.3. *The rate region* $(R_1, R_2, R_3, h_{12}, h_{13}, h_{23})$*, for a distortion requirement of* D_0 *desired by the legitimate receiver is given by the set of rate pairs* (R_1, R_2, R_3) *satisfying the following:*

$$\begin{aligned} R_1 &\geq \Big(h_{23} - \big[H(X) - R(D_0)\big]\Big)^+ = S_1 && (6.12)\\ R_2 &\geq \Big(h_{13} - \big[H(X) - R(D_0)\big]\Big)^+ = S_2 && (6.13)\\ R_3 &\geq \Big(h_{12} - \big[H(X) - R(D_0)\big]\Big)^+ = S_3 && (6.14)\\ R_1 + R_2 + R_3 &\geq R(D_0) && (6.15) \end{aligned}$$

Proof. Refer the generalized proof given in Thorem VI.5 for (6.12), (6.13) and (6.14). (6.15) follows naturally. □

a. Achievability of Theorem VI.3

We have two cases:

<u>**Case 1:**</u> $S_1 + S_2 + S_3 \leq R(\mathrm{D}_0)$

In this regime, we have:

$$\big(H(X) - h_{12}\big) + \big(H(X) - h_{13}\big) + \big(H(X) - h_{23}\big) \geq 2R(\mathrm{D}_0) \tag{6.16}$$

It is clear from Theorem VI.3 that, all we need to show is that the points $\big(R_1 = S_1, R_2 = S_2, R_3 = R(\mathrm{D}_0) - (S_1 + S_2)\big)$, $\big(R_1 = S_1, R_2 = R(\mathrm{D}_0) - (S_1 + S_3), R_3 = S_3\big)$ and $\big(R_1 = R(\mathrm{D}_0) - (S_2 + S_3), R_2 = S_2, R_3 = S_3\big)$ achieve the desired secrecy levels. We will do it only for one point as the others points can be shown in a similar manner.

Lemma VI.2. *The assignment of* $\big(R_1 = S_1, R_2 = S_2, R_3 = R(D_0) - (S_1 + S_2)\big)$ *information bits achieves the secrecy level in (6.11).*

Proof.

$$\begin{aligned}
\frac{1}{N}H(X^N|W_1W_2) &\geq \frac{1}{N}\big[H(X^N, W_1, W_2) - H(W_1, W_2)\big] \\
&\geq \frac{1}{N}\big[H(X^N) - H(W_1, W_2)\big] \\
&\geq \frac{1}{N}\big[H(X^N) - H(W_1) - H(W_2)\big] \\
&\geq H(X) - R_1 - R_2 \\
&\overset{(a)}{\geq} h_{12}
\end{aligned}$$

By following similar steps as above

$$\begin{aligned}
\frac{1}{N}H(X^N|W_1, W_3) &\geq H(X) - R_1 - R_3 \overset{(a)}{\geq} h_{13} \\
\frac{1}{N}H(X^N|W_2, W_3) &\geq H(X) - R_2 - R_3 \overset{(a)}{\geq} h_{23}
\end{aligned}$$

where (a) follows from (6.16). □

<u>Case 2:</u> $S_1 + S_2 + S_3 > R(\mathrm{D}_0)$

In this regime, we need to show the achievability of $\big(R_1 = S_1, R_2 = S_2, R_3 = S_3\big)$ in (6.12), (6.13) and (6.14). Without loss of generality, we assume that $h_{23} \geq h_{13} \geq h_{12}$, which implies $R_1 \geq R_2 \geq R_3$.

We note that the maximum key bits (i.e., additional bits) that we are allowed to use in this regime is,

$$R_k = S_1 + S_2 + S_3 - R(\mathrm{D}_0) = \big(h_{12} + h_{13} + h_{23} - 3H(X) + 2R(\mathrm{D}_0)\big) \geq 0 \qquad (6.17)$$

which clearly depends on the desired secrecy levels (h_{12}, h_{13}, h_{23}), for a fixed $H(X)$ and $R(\mathrm{D}_0)$. Furthermore, we observe that for the case of perfect secrecy (i.e., $h_{12} = h_{13} = h_{23} = H(X)$), the total key bits we need is $2R(\mathrm{D}_0)$ (i.e., overhead of factor two). Hence, we have $0 \leq R_k \leq 2R(\mathrm{D}_0)$. Depending on R_k, we will explicitly show a code construction in two regimes, namely:

- $R(\mathrm{D}_0) \leq R_k \leq 2R(\mathrm{D}_0)$

- $0 \leq R_k < R(\mathrm{D}_0)$

We begin with the case when $R(\mathrm{D}_0) \leq R_k \leq 2R(\mathrm{D}_0)$.

1. Letting $M_k = R_k - R(\mathrm{D}_0)$, we write $(R_1 = S_1, R_2 = S_2, R_3 = S_3)$ in (6.12), (6.13) and (6.14) as:

$$\begin{aligned} R_1 &= \big(H(X) - h_{12}\big) + \big(H(X) - h_{13}\big) + M_k \\ R_2 &= \big(H(X) - h_{12}\big) + \big(H(X) - h_{23}\big) + M_k \qquad (6.18) \\ R_3 &= \big(H(X) - h_{13}\big) + \big(H(X) - h_{23}\big) + M_k \end{aligned}$$

2. Represent a distinct set of information bits of length M_k by $\boldsymbol{\mathcal{I}_s}$. The idea now is to make $\boldsymbol{\mathcal{I}_s}$ of length M_k perfectly secure, which by our previous observation requires an overhead of $2M_k$ bits.

3. Generate key bits $\boldsymbol{\mathcal{K}_s^1}$ and $\boldsymbol{\mathcal{K}_s^2}$ each of length M_k independently of each other and independent of the information bits. Place $\boldsymbol{\mathcal{I}_s}$, $\boldsymbol{\mathcal{I}_s} \oplus \boldsymbol{\mathcal{K}_s^1}$ and $\boldsymbol{\mathcal{I}_s} \oplus \boldsymbol{\mathcal{K}_s^2}$ in different channels.

4. Denote $\mathrm{M_{rem}}$ to be the remaining information bits to code at this point. We then have, $\mathrm{M_{rem}} = R(\mathrm{D}_0) - M_k = \big(H(X)-h_{12}\big)+\big(H(X)-h_{13}\big)+\big(H(X)-h_{23}\big)$. We also note that the remaining key bits we can use is $R_k - 2M_k = \mathrm{M_{rem}}$ as well. (Recall, that we had already used $2M_k$ key bits for constructing the perfect secure code).

5. Generate key bits $\boldsymbol{\mathcal{K}}^{\mathbf{12}}$, $\boldsymbol{\mathcal{K}}^{\mathbf{13}}$, $\boldsymbol{\mathcal{K}}^{\mathbf{23}}$ of length $\big(H(X)-h_{12}\big), \big(H(X)-h_{13}\big), \big(H(X)-h_{23}\big)$ respectively, independent of each other and independent of $\boldsymbol{\mathcal{K}}_{\boldsymbol{s}}^{\mathbf{1}}$, $\boldsymbol{\mathcal{K}}_{\boldsymbol{s}}^{\mathbf{2}}$ and the information bits. Represent distinct sets of information bits by $\boldsymbol{\mathcal{I}}^{\mathbf{12}}$, $\boldsymbol{\mathcal{I}}^{\mathbf{13}}$, $\boldsymbol{\mathcal{I}}^{\mathbf{23}}$ of length $\big(H(X)-h_{12}\big), \big(H(X)-h_{13}\big), \big(H(X)-h_{23}\big)$ respectively, which are different from $\boldsymbol{\mathcal{I}_s}$. Place $\boldsymbol{\mathcal{K}}^{\boldsymbol{ij}}$, $\boldsymbol{\mathcal{I}}^{\boldsymbol{ij}} \oplus \boldsymbol{\mathcal{K}}^{\boldsymbol{ij}}$ in channel i (channel j) and channel j (channel i) respectively for all $1 \le i < j \le 3$. It is instructive to see that the code construction given below follows the representation of R_1, R_2, R_3 in (6.18):

$$\begin{array}{lcccccc}
\mathbf{C}_1: & [& \boldsymbol{\mathcal{I}}^{\mathbf{12}} \oplus \boldsymbol{\mathcal{K}}^{\mathbf{12}} & \boldsymbol{\mathcal{K}}^{\mathbf{13}} & \boldsymbol{\mathcal{I}}^{\boldsymbol{s}} \oplus \boldsymbol{\mathcal{K}}_{\boldsymbol{s}}^{\mathbf{1}} \oplus \boldsymbol{\mathcal{K}}_{\boldsymbol{s}}^{\mathbf{2}} &] \\
\mathbf{C}_2: & [& \boldsymbol{\mathcal{K}}^{\mathbf{12}} & \boldsymbol{\mathcal{I}}^{\mathbf{23}} \oplus \boldsymbol{\mathcal{K}}^{\mathbf{23}} & \boldsymbol{\mathcal{K}}_{\boldsymbol{s}}^{\mathbf{1}} &] \\
\mathbf{C}_3: & [& \boldsymbol{\mathcal{I}}^{\mathbf{13}} \oplus \boldsymbol{\mathcal{K}}^{\mathbf{13}} & \boldsymbol{\mathcal{K}}^{\mathbf{23}} & \boldsymbol{\mathcal{K}}_{\boldsymbol{s}}^{\mathbf{2}} &]
\end{array} \tag{6.19}$$

It is interesting to see that the code construction in (6.19) provides *perfect security* for the case when eavesdropper has access to only one instead of two channels. This is expected, as we need $R(\mathrm{D}_0)$ key bits to provide perfect security to $R(\mathrm{D}_0)$ information bits, provided the eavesdropper has access to only one channel, from our study of the case $m = 1, n = L = 2$.

Now, we turn to case when $0 \le R_k < R(\mathrm{D}_0)$. It is easy to see that in this regime some of the bits must go unprotected even if the eavesdropper has access to one channel, much like the structure in (6.6). Listed below are the possibilities that need to be considered in constructing a code for this case:

(i) With the availability of R_k key bits, construct a perfect secure code, by providing perfect security to $\frac{R_k}{2}$ information bits, much like the third column in the structure of (6.19) and distribute the rest of the information bits in the channels depending on rate constraints. This further implies that the minimum rate constrained channel (which in our case is channel 3; recall the assumption $R_1 \geq R_2 \geq R_3$), should at least support $\frac{R_k}{2}$ bits.

(ii) If the condition (i) listed above fails, the next possibility is to use R_k key bits to provide security to R_k information bits across the other two channels (i.e., in our case channels 1 and 2), and distribute the rest of the information bits depending on rate constraints. This naturally implies that $R_2 = S_2$ must at least support a rate of R_k.

(iii) If condition (i) and (ii) listed above fails, then, it can be easily argued that the only way is to code jointly across channels 1,3 and channels 1,2.

We can succinctly represent the above possible cases by rewriting (6.12), (6.13) and (6.14) as:

$$\begin{aligned} R_1 &= \delta_1 + \eta + \gamma_1 \\ R_2 &= \delta_2 + \gamma_1 \\ R_3 &= \delta_3 + \eta + \gamma_2 \end{aligned} \tag{6.20}$$

where $\delta_1, \delta_2, \delta_3, \eta, \gamma_1$ and γ_2 for the cases (i), (ii) and (iii) discussed above are given in Table I. (Note that R_k and S_i ($i = 1, 2, 3$) are defined as in (6.17), (6.12), (6.13) and (6.14) respectively). The achievability in this regime is as follows:

1. Generate a distinct set of information bits $\{\boldsymbol{\mathcal{I}}_i\}$ ($i = 1, 2 \ldots 5$) of lengths $\delta_1, \delta_2, \delta_3, \eta$ and γ_1 respectively.

Table I. The case when $0 \leq R_k < R(\mathrm{D}_0)$ when $m = 2, n = L = 3$.

Condition	δ_1	δ_2	δ_3	η	γ_1	γ_2
(i) $S_3 \geq \frac{R_k}{2}$	$S_1 - \frac{R_k}{2}$	$S_2 - \frac{R_k}{2}$	$S_3 - \frac{R_k}{2}$	0	$\frac{R_k}{2}$	$\frac{R_k}{2}$
(ii) $S_2 < R_k$	$S_1 - R_k$	$S_2 - R_k$	S_3	0	R_k	0
(iii) $S_2 < R_k$ & $S_3 < \frac{R_k}{2}$	$S_1 - R_k$	$S_2 - R_k + S_3$	0	S_3	$R_k - S_3$	0

2. Generate distinct set of key bits $\{\boldsymbol{\mathcal{K}}_i\}$ $(i = 1, 2, 3)$ of lengths γ_1, γ_2 and η respectively, independent of each other and the information bits. Letting $\mathbf{C}_i$ be the code sent through channel i, we have:

$$\begin{array}{ll} \mathbf{C}_1: & \begin{bmatrix} \boldsymbol{\mathcal{I}_1} & \boldsymbol{\mathcal{I}_4} \oplus \boldsymbol{\mathcal{K}_3} & \boldsymbol{\mathcal{I}_5} \oplus \boldsymbol{\mathcal{K}_1} \oplus \boldsymbol{\mathcal{K}_2} \end{bmatrix} \\ \mathbf{C}_2: & \begin{bmatrix} \boldsymbol{\mathcal{I}_2} & & \boldsymbol{\mathcal{K}_1} \end{bmatrix} \\ \mathbf{C}_3: & \begin{bmatrix} \boldsymbol{\mathcal{I}_3} & \boldsymbol{\mathcal{K}_3} & \boldsymbol{\mathcal{K}_2} \end{bmatrix} \end{array} \tag{6.21}$$

It is possible to show that the secrecy requirements are met for the code in (6.21) for all the cases outlined in Table I.

3. $m = 1, n = 2, L = 3$

In this case, the eavesdropper has access to any one of three channels and the legitimate receiver needs to be able to reconstruct the source from any of the two channels. Here, we have $\mathfrak{A} = \big\{\{1\}, \{2\}, \{3\}\big\}$ and $\mathfrak{B} = \{(1, 2), (1, 3), (2, 3)\}$. The desired level of secrecy for a given h_1, h_2, h_3 are given by:

$$\frac{1}{N} H(X^N | W_i) \geq h_i \quad \text{for } i = 1, 2, 3. \tag{6.22}$$

where W_i is the codeword sent through the i^{th} channel.

Theorem VI.4. *The rate region* $(R_1, R_2, R_3, h_1, h_2, h_3)$*, for a distortion requirement of* D_0 *desired by the legitimate receiver is given by the set of rate pairs* (R_1, R_2, R_3) *satisfying the following:*

$$\begin{aligned} R_1 &\geq \max\{h_2, h_3\} - \big[H(X) - R(D_0)\big] && (6.23)\\ R_2 &\geq \max\{h_1, h_3\} - \big[H(X) - R(D_0)\big] && (6.24)\\ R_3 &\geq \max\{h_1, h_2\} - \big[H(X) - R(D_0)\big] && (6.25)\\ R_i + R_j &\geq R(D_0) \quad \textit{for } 1 \leq i < j \leq 3 && (6.26) \end{aligned}$$

Proof. We will prove only (6.23), as (6.24) and (6.25) can be proved similarly. (6.26) follows naturally.

$$\begin{aligned} NR_1 &\geq H(W_1)\\ &\overset{(a)}{\geq} H(W_1|W_2)\\ &\geq H(W_1|W_2) - H(W_1|W_2, X^N)\\ &= I(X^N; W_1|W_2)\\ &= I(X^N; W_1, W_2) - I(X^N; W_2)\\ &= I(X^N; W_1, W_2) - H(X^N) + H(X^N|W_2)\\ &\overset{(b)}{\geq} N\big[R(\mathrm{D}_0) - H(X) + h_2\big] && (6.27) \end{aligned}$$

Following similar steps as above

$$\begin{aligned} NR_1 &\geq H(W_1)\\ &\geq H(W_1|W_3) - H(W_1|W_3, X^N)\\ &\overset{(b)}{\geq} N\big[R(\mathrm{D}_0) - H(X) + h_3\big] && (6.28) \end{aligned}$$

where (a) follows from the fact that conditioning reduces entropy and (b) follows due to the fact that $R(\mathrm{D}_0)$ is the minimum number of bits required to describe a source at distortion D_0, $\{X\}$ being i.i.d and the secrecy requirements in (6.22). From (6.27) and (6.28), we obtain (6.23). □

a. Achievability of Theorem VI.4

From (6.26), we get:

$$R_1 + R_2 + R_3 \geq \frac{3R(\mathrm{D}_0)}{2} \tag{6.29}$$

It is easy to see that any two out of three equations in (6.23), (6.24), (6.25) needs to be the same. Without loss of generality, we let $h_1 > h_2 > h_3$. Then (6.23), (6.24), (6.25) reduces to:

$$\begin{aligned} R_1 \quad &\geq h_2 - \big[H(X) - R(\mathrm{D}_0)\big] \quad = T_1 && (6.30) \\ R_2 = R_3 \quad &\geq h_1 - \big[H(X) - R(\mathrm{D}_0)\big] \quad = T_2 && (6.31) \end{aligned}$$

Henceforth we will appeal to (6.29), (6.30), (6.31), (6.26) for showing the achievability. Here again, we have two cases to address depending on T_1, T_2.

<u>**Case 1:**</u> $T_1 + T_2 \leq R(D_0)$

It turns out that we need to show achievability of two points as described below. However, we will not prove that the points satisfy the secrecy constraints, as similar ideas (refer to the proof of Lemma VI.1) can be used here as well.

- We first consider the case when $T_2 < \frac{R(D_0)}{2}$. Due to the fact that $T_1 < T_2$ and that (6.26) has to be satisfied, the only point achievable in this case has to be $\big(R_1 = R_2 = R_3 = \frac{R(D_0)}{2}\big)$, which satisfies (6.29) with equality. A simple achievable scheme is the following:

  1. Represent distinct sets of information bits by $\boldsymbol{\mathcal{I}}_1$ and $\boldsymbol{\mathcal{I}}_2$ each of length $\frac{R(D_0)}{2}$.

2. Send $\boldsymbol{\mathcal{I}}_1$ trough channel 1, $\boldsymbol{\mathcal{I}}_2$ through channel 2 and $\boldsymbol{\mathcal{I}_1} \oplus \boldsymbol{\mathcal{I}}_2$ through channel 3.

- Now consider the case when $T_2 > \frac{R(D_0)}{2}$ and of course $T_1 + T_2 < R(\mathrm{D}_0)$, which implies $T_1 < \frac{R(D_0)}{2}$. Clearly, the achievable point which satisfies the secrecy constraints is $\big(R_1 = H(X) - h_1, R_2 = R_3 = T_2\big)$ which satisfies (6.29) with equality. A simple achievable scheme is the following:

1. Represent distinct set of information bits by $\boldsymbol{\mathcal{I}}_1, \boldsymbol{\mathcal{I}}_2$ each of length $\big(H(X) - h_1\big)$ bits. Represent another distinct set of information bits of length $\big(R(\mathrm{D}_0) - 2[H(X) - h_1]\big)$ by $\boldsymbol{\mathcal{I}}_3$. Letting $\mathbf{C}_i$ be the code sent through channel i, we have:

$$\begin{array}{ll} \mathbf{C}_1 : & [\quad \boldsymbol{\mathcal{I}_1} \quad] \\ \mathbf{C}_2 : & [\quad \boldsymbol{\mathcal{I}_2} \quad \boldsymbol{\mathcal{I}_3} \quad] \\ \mathbf{C}_3 : & [\quad \boldsymbol{\mathcal{I}_1} \oplus \boldsymbol{\mathcal{I}_2} \quad \boldsymbol{\mathcal{I}_3} \quad] \end{array} \tag{6.32}$$

<u>**Case 2:**</u> $T_1 + T_2 > R(D_0)$

We show a simple scheme to achieve the point $\big(R_1 = T_1, R_2 = R_3 = T_2\big)$ which follows:

1. Generate a key $\boldsymbol{\mathcal{K}}$ of length $R_k = T_1 + T_2 - R(\mathrm{D}_0) = \big(h_1 + h_2 - 2H(X) + R(\mathrm{D}_0)\big)$, independent of the information bits. We note that for perfect secrecy the amount of key bits required is $R(\mathrm{D}_0)$ (i.e., overhead of factor one). As was done before we represent (6.30), (6.31) as :

$$\begin{aligned} R_1 &= \big(H(X) - h_1\big) + R_k \\ R_2 &= \big(h_1 - h_2\big) + \big(H(X) - h_1\big) + R_k \\ R_3 &= \big(h_1 - h_2\big) + \big(H(X) - h_1\big) + R_k \end{aligned} \tag{6.33}$$

2. The idea now is to use $\boldsymbol{\mathcal{K}}$ of length R_k to provide security to a total of $2R_k$ information bits of which only R_k are distinct. (Note that we are doing a diversity coding here). Denote distinct information bits of length R_k by $\boldsymbol{\mathcal{I}_s}$. Place $\boldsymbol{\mathcal{K}}$, $\boldsymbol{\mathcal{I}_s} \oplus \boldsymbol{\mathcal{K}}$, $\boldsymbol{\mathcal{I}_s} \oplus \alpha\boldsymbol{\mathcal{K}}$ in different channels. Here, α represents an element from Galois Field.

3. At this point we still have $\big(H(X) - h_1\big) + \big(H(X) - h_2\big)$ distinct information bits that have to be coded in such a way that it can be decoded from any two of the three channels. Denote distinct sets of information bits by $\boldsymbol{\mathcal{I}_1}$, $\boldsymbol{\mathcal{I}_2}$ each of length $\big(H(X) - h_1\big)$. Let $\boldsymbol{\mathcal{I}_3}$ represent a distinct set of information bits of length $(h_1 - h_2)$. Place $\boldsymbol{\mathcal{I}_1}$ in channel 1, concatenate $\boldsymbol{\mathcal{I}_2}$ with $\boldsymbol{\mathcal{I}_3}$ and place in channel 2, and finally place $\boldsymbol{\mathcal{I}_3}$ concatenated with $\boldsymbol{\mathcal{I}_1} \oplus \boldsymbol{\mathcal{I}_2}$ in channel 3.

The code construction in (6.34) follows the representation in (6.33). Furthermore, it is important to note that for $\boldsymbol{\mathcal{I}_s}$ to be decodable by any two of the three channels, α has to be chosen such that the column corresponding to the perfect secure code in (6.34) which involves the variables $\boldsymbol{\mathcal{I}_s}$ and $\boldsymbol{\mathcal{K}}$ are to be *linearly independent*. More on choosing α is outlined in Section 4. It is also interesting to note that (6.34) is simply a concatenation of (6.32) with a perfect secure code.

$$\begin{array}{lccccr}
\mathbf{C}_1: & [& \multicolumn{2}{c}{\boldsymbol{\mathcal{I}_1}} & \boldsymbol{\mathcal{K}} &] \\
\mathbf{C}_2: & [& \boldsymbol{\mathcal{I}_3} & \boldsymbol{\mathcal{I}_2} & \boldsymbol{\mathcal{I}_s} \oplus \boldsymbol{\mathcal{K}} &] \\
\mathbf{C}_3: & [& \boldsymbol{\mathcal{I}_3} & \boldsymbol{\mathcal{I}_1} \oplus \boldsymbol{\mathcal{I}_2} & \boldsymbol{\mathcal{I}_s} \oplus \alpha\boldsymbol{\mathcal{K}} &] \\
 & & \multicolumn{2}{c}{\underbrace{\hphantom{\boldsymbol{\mathcal{I}_3} \qquad \boldsymbol{\mathcal{I}_1} \oplus \boldsymbol{\mathcal{I}_2}}}_{\text{Non-Secure Code}}} & \underbrace{\hphantom{\boldsymbol{\mathcal{I}_s} \oplus \alpha\boldsymbol{\mathcal{K}}}}_{\text{Perfect Secure Code}} &
\end{array} \tag{6.34}$$

The following gives some central themes that have emerged from the study of these simple, yet representative cases:

(a) The ‘extra’ bits beyond $R(\mathrm{D}_0)$ information bits, if required, are the ‘key’ bits that have to be generated independently of the information bits, and these needs to be conveyed to the the legitimate receiver.

(b) It is interesting to note that the overhead factor of the key bits that is required for providing perfect secrecy for the case $m = 1, n = L = 2$ was *one*; for the case $m = 2, n = L = 3$ was *two*; while for $m = 1, n = 2, L = 3$ was *one*. These observations are made more precise in Section 4.

(c) Depending on the availability of ‘key’ bits, some of the information bits are provided perfect secrecy using these key bits by constructing a simple network code and the rest of the information bits are again network coded and concatenated with the secure network code, which was observed to provide the desired level of secrecy.

4. Generalizations to Arbitrary m, n, L

Here, we state precisely some of the observations that were made the Section 1, Section 2 and Section 3. We focus on the symmetrical secrecy requirement which facilitates derivations of lower bounds on sum-rate. Finally we propose simple schemes to achieve the lower bound. We first introduce some notations pertinent to this section. Note that $\mathfrak{A}$, $\mathfrak{B}$ and $\Upsilon\big(\mathcal{B}\big)$ for any $\mathcal{B} \in \mathfrak{B}$ are all defined in Section B. Let $\mathfrak{R}_A$ denote the sum rate of channels corresponding to elements in A, and $\mathcal{H}_A$ denote the sum of the entropies of the codewords corresponding to the elements in A. For example, if $A = (1, 2, 4)$, $\mathfrak{R}_A = R_1 + R_2 + R_4$, while $\mathcal{H}_A = H(W_1) + H(W_2) + H(W_4)$, where R_1, R_2, R_4 are the rates of channels $1, 2, 4$ respectively, and W_1, W_2, W_4 are the messages sent along channels $1, 2, 4$ respectively.

Theorem VI.5. *For a fixed $\mathcal{B} \in \mathfrak{B}$ and any $A \in \Upsilon\big(\mathcal{B}\big)$, the following holds:*

$$\mathfrak{R}_{A^c} \geq h_A - \big(H(X) - R(D_0)\big) \tag{6.35}$$

where, $A^c = \mathcal{B} \setminus A$.

Proof.

$$\begin{aligned}
N\mathfrak{R}_{A^c} &\geq \mathcal{H}_{A^c} \\
&\geq H(Y_{A^c}) \\
&\overset{(a)}{\geq} H(Y_{A^c}|Y_A) \\
&\geq H(Y_{A^c}|Y_A) - H(Y_{A^c}|Y_A, X^N) \\
&= I(Y_{A^c}; X^N|Y_A) \\
&= I(Y_{\mathcal{B}}; X^N) - I(Y_A; X^N) \\
&= I(Y_{\mathcal{B}}; X^N) - H(X^N) + H(X^N|Y_A) \\
&\overset{(b)}{\geq} N\big[R(\mathrm{D}_0) - H(X) + h_A\big]
\end{aligned}$$

where (a) follows from the fact that conditioning reduces entropy, (b) follows from the fact that $R(\mathrm{D}_0)$ is the minimum number of bits required to describe a source at distortion D_0, $\{X\}$ being i.i.d and the secrecy requirement in (6.1). □

Lemma VI.3. *The sum rate of L channels without secrecy constraints (i.e., $m = 0, n \leq L$) satisfies:*

$$\sum_{i=1}^{L} R_i \geq \frac{L}{n} R(D_0) \tag{6.36}$$

Proof. For every $\mathcal{B} \in \mathfrak{B}$, the following has to be true:

$$\mathfrak{R}_{\mathcal{B}} \geq R(\mathrm{D}_0)$$

Summing over all $\mathcal{B} \in \mathfrak{B}$ we get:

$$\binom{L-1}{n-1}\sum_{i=1}^{L} R_i \geq \binom{L}{n} R(\mathrm{D}_0)$$

which yields (6.36). □

Lemma VI.4. *The sum rate of L channels with secrecy constraints (i.e., $m \neq 0$, $m < n \leq L$) , for the case when $h_A = h$, $\forall A \in \mathfrak{A}$ satisfies:*

$$\sum_{i=1}^{L} R_i \geq \frac{L}{n-m}\left[h - \big(H(X) - R(D_0)\big)\right] \tag{6.37}$$

Proof. From Theorem VI.5, we have for any $A \in \Upsilon\big(\mathcal{B}\big)$, and for a fixed $\mathcal{B} \in \mathfrak{B}$:

$$\mathfrak{R}_{A^c} \geq h_A - \big(H(X) - R(\mathrm{D}_0)\big)$$

Summing the above over all $\Upsilon\big(\mathcal{B}\big)$, for a fixed $\mathcal{B} \in \mathfrak{B}$, together with the condition $h_A = h$, $\forall A \in \mathfrak{A}$ yields:

$$\begin{aligned} \binom{n-1}{n-m-1}\mathfrak{R}_{\mathcal{B}} &\geq \binom{n}{m}\left[h - \big(H(X) - R(\mathrm{D}_0)\big)\right] \\ \Rightarrow \quad \mathfrak{R}_{\mathcal{B}} &\geq \frac{n}{n-m}\left[h - \big(H(X) - R(\mathrm{D}_0)\big)\right] \end{aligned}$$

Summing the above over all $\mathcal{B} \in \mathfrak{B}$,

$$\binom{L-1}{n-1}\sum_{i=1}^{L} R_i \geq \binom{L}{n}\frac{n}{n-m}\big(h - \left[H(X) - R(\mathrm{D}_0)\right]\big)$$

which yields (6.37). □

Let us take a close look at (6.36). If $n = L$ as in [67] (i.e., if we do not impose diversity coding), then the sum-rate in (6.36)) reduces to the total information bits of $R(\mathrm{D}_0)$. Hence, we see that the $\frac{L}{n}$ factor overhead in (6.36) is precisely due to the fact that the legitimate receiver is to be endowed with diversity coding. On the other hand, the lower bound in Lemma VI.4 (refer (6.37)) imposes secrecy constraints on the channels, as well as diversity coding on the legitimate receiver. These results gives a clear picture on the impact of secure and diversity coding on the sum-rate. Further, we assume that

$$\frac{L}{n}R(\mathrm{D}_0) \leq \frac{L}{n-m}\left[h - \big(H(X) - R(\mathrm{D}_0)\big)\right] \tag{6.38}$$

If do not impose the above assumption, then, all we need to achieve is the lower bound in (6.36) (i.e., without the secrecy constraints), which makes the problem less interesting. Now, it only remains the show the lower bound in (6.37) is achievable. Letting R_k represent the minimum additional bits, we then have the following:

Lemma VI.5. *The minimum number of additional bits (i.e., key bits) with the eavesdropper being able to access any of the $\binom{L}{m}$ channels and the legitimate receiver being able to decode from any of the $\binom{L}{n}$ channels with $m < n \leq L$ is given by*

$$R_k = \frac{L}{n(n-m)}\Big[n\big(h - H(X)\big) + mR(D_0)\Big] \tag{6.39}$$

Here again, we see that the number of key bits required is scaled by the factor $\frac{L}{n}$ (i.e., diversity gain factor). We will first show that, we can construct a perfect secure code which can provide a diversity gain as well, for the legitimate receiver. Then, we will address the construction of a partial secure code that provides diversity gain.

Lemma VI.6. *There exists a code which is perfectly secure and provides diversity gain for the legitimate receiver, which achieves the lower bound in Lemma VI.4.*

Proof. The code is constructed as follows:

1. For the case of perfect secrecy, we first note from Lemma VI.5 that we have a total of $R_k = \frac{L}{n}\frac{m}{n-m}R(\mathrm{D}_0)$ key bits. Hence, the number of distinct key bits we need is $\frac{m}{n-m}R(\mathrm{D}_0)$. Generate key bits $\boldsymbol{\mathcal{K}}_i$ for $i = 1, 2 \ldots m$, each of length $\frac{R(D_0)}{n-m}$ independently of each other and the information bits.

2. Generate distinct information bits $\boldsymbol{\mathcal{I}}_j$, each of length $\frac{R(D_0)}{n-m}$ for $j = 1, 2 \ldots n-m$.

   Note that, at this point we have represented all our distinct information bits.

Denote,

$$\mathbf{S} \;=\; \begin{bmatrix} \boldsymbol{\mathcal{K}}_1 & \boldsymbol{\mathcal{K}}_2 & \ldots & \boldsymbol{\mathcal{K}}_m & \boldsymbol{\mathcal{I}}_1 & \boldsymbol{\mathcal{I}}_2 & \ldots & \boldsymbol{\mathcal{I}}_{n-m} \end{bmatrix}^{\mathbf{T}}$$

The code can then be written as :

$$\mathbf{C}\big(R(\mathrm{D}_0)\big) = \mathbf{AS} \tag{6.40}$$

where, $\mathbf{A}$ is a $L \times n$ matrix, which takes the form:

$$\mathbf{A} \;=\; \begin{bmatrix} \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1,n-1} & \alpha_{1,n} \\ \alpha_{21} & \alpha_{22} & \ldots & \ldots & \alpha_{2,n} \\ & \vdots & & \ddots & \\ \alpha_{L,1} & \alpha_{L,2} & \ldots & \ldots & \alpha_{L,n} \end{bmatrix} \tag{6.41}$$

and $\mathbf{C}\big(R(\mathrm{D}_0)\big)$ represents a code that can provide perfect secrecy and diversity gain for $R(\mathrm{D}_0)$ information bits. In order for the legitimate receiver be able to recover $\mathbf{S}$, from any of the $\binom{L}{n}$ submatrices of $\mathbf{A}$, each of size $n \times n$, it must be true that all such submatrices must be invertible. But, this can be easily guaranteed by choosing $\alpha_{i,j}$ from a large field for $1 \le i \le L,\;\; 1 \le j \le n$ such that $\mathbf{A}$ has all its rows linearly independent. It is easy to see that for the case when no diversity coding is required (i.e., $n = L$), simply delete the last $(L-n)$ rows of matrix $\mathbf{A}$. When the eavesdropper accesses m channels at a time, he has n unknowns to determine $(m < n)$ as in [67]. □

Another important aspect in achieving the lower bound in (6.37) is that the elements $\alpha_{i,j}$ in matrix $\mathbf{A}$ should be chosen such that we do not increase the rate of the resultant code. Note that it may not always be possible to do this as it depends on m, n, L and the secrecy requirements. The best way to ensure such that we operate on the lower bound is to use the channel once per G input codewords after collecting enough information bits (where, G may be an arbitrary function of m, n, L and can be determined empirically). It is clear that this scheme will incur a delay.

Lemma VI.7. *A partially secure code can be obtained by concatenating a perfect secure network code with another network code of information bits, which at the same time provides diversity gain for the legitimate receiver and achieves the lower bound in Lemma VI.4.*

Proof. The code is constructed as follows:

1. From Lemma VI.5, and by our previous observation from Lemma VI.6 on the amount of key bits required for providing perfect secrecy, we infer that for the present case (i.e., when $h \neq H(X)$), the amount of *distinct* information bits that can be provided perfect secrecy is: $\big[R(\mathrm{D}_0) - \big(H(X) - h\big)\frac{n}{m}\big]$.
2. Hence we see that we need to code an additional $\big(H(X) - h\big)\frac{n}{m}$ information bits. Generate $\boldsymbol{\mathcal{J}}_i$ distinct information bits each of length $\frac{\left(H(X)-h\right)}{m}$ for $i = 1, 2 \ldots n$.

Denote:

$$\tilde{\mathbf{S}} = \begin{bmatrix} \boldsymbol{\mathcal{J}}_1 & \boldsymbol{\mathcal{J}}_2 & \ldots & \boldsymbol{\mathcal{J}}_n \end{bmatrix}^{\mathbf{T}}$$

The network code of information bits is given by:

$$\tilde{\mathbf{C}} = \tilde{\mathbf{A}}\tilde{\mathbf{S}}$$

where, $\tilde{\mathbf{A}}$ is a $L \times n$ matrix, which can be represented as:

$$\tilde{\mathbf{A}} = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1,n-1} & \beta_{1,n} \\ \beta_{21} & \beta_{22} & \cdots & \cdots & \beta_{2,n} \\ & \vdots & & \ddots & \\ \beta_{L,1} & \beta_{L,2} & \cdots & \cdots & \beta_{L,n} \end{bmatrix}$$

Here again, we need to make sure that all the rows of $\tilde{\mathbf{A}}$ are linearly independent for the legitimate receiver to be able to decode the rest of the $\big(H(X) - h\big)\frac{n}{m}$ information bits. Hence, the partial secure code is given by:

$$\mathbf{R} = \begin{bmatrix} \quad \tilde{\mathbf{C}} & \mathbf{C}\Big(R(\mathrm{D}_0) - (H(X) - h)\frac{n}{m}\Big) \quad \end{bmatrix} \tag{6.42}$$

where, $\mathbf{R}$ is the final partial secure code and $\mathbf{C}\big(R(\mathrm{D}_0) - (H(X) - h)\frac{n}{m}\big)$ is the perfect secure code which can provide perfect secrecy to $\big(R(\mathrm{D}_0) - (H(X) - h)\frac{n}{m}\big)$ information bits. Note that its construction was provided in the proof of Lemma VI.6. □

C. Multiple Sources with Diversity Coding and Secrecy Constraints

In Section B, we looked at the problem of communicating *one* source in presence of an eavesdropper with the legitimate receiver being able to reconstruct the source from a subset whose size is larger than the size of the subset that the eavesdropper has access to. In this section, we look at a natural extension of the problem treated in Section B: communicating *multiple* sources with an eavesdropper and the legitimate receiver being endowed with diversity coding. The motivation for studying this problem comes from the following:

Consider the scenario where the eavesdropper has access to a subset of size m out of a total of L available channels and that the legitimate receiver should be able to reconstruct the source whose quality depends on the number of the channels that it has access to. For example, if the legitimate receiver accesses any subset of size $(m + 1)$ channels, it should be able to reconstruct the source to an *acceptable level* of distortion (i.e., basic quality), with the quality of reconstruction of the source increasing with increasing k, as the legitimate receiver gets to access $(m+k)$ channels. Roughly speaking, this situation is more or less similar to the problem of successive refinement with secrecy constraints. However for this problem setup, the point of view of successive refinement is as follows: We assume that an i.i.d source S can be split into several i.i.d sources $S_1, S_2 \dots S_k$ with decreasing level of importance. For example, S_1 could represent the bits corresponding to the low frequency coefficients of the image S, while sources $S_2, S_3 \dots S_k$ could represent the bits corresponding to the $(k-1)^{th}$ *detail* coefficients, in a wavelet-type expansion. Hence, the problem then reduces to communicating *multiple* sources with an eavesdropper, with the legitimate receiver endowed with diversity coding. This scheme is sometimes called multilevel diversity coding [68].

In a multilevel diversity coding system, information sources are encoded by many encoders and there are multiple decoders with each decoder having access to a subset of encoders [68]. The goal of the multilevel diversity coding system is that, the encoding needs to be done in such a way that each decoder be able to reconstruct the source either perfectly or with some distortion. In the symmetrical multilevel diversity coding problem treated by Roche and Yeung [69], the decoders are partitioned into multiple levels and that the decoder belonging to a particular level be able to reconstruct a predetermined number of sources.

We consider a more generalized version of the problem treated in [69] in this section: There are multiple information sources that have to be communicated to a legitimate receiver through the total available L separate communication channels. The eavesdropper has access to any subset of m channels ($m < L$), while the legitimate receiver has access to all the L channels. There are a total of $L-m$ sources that has to be communicated to the legitimate receiver. The source encoder encodes the sources into L packets, possibly of different size and sends them each, through a channel. The encoding is done in such a way that, any subset of m packets does not convey any information about any of the sources, while the sources $\{1, 2, \dots k\}(1 \leq k \leq L-m)$

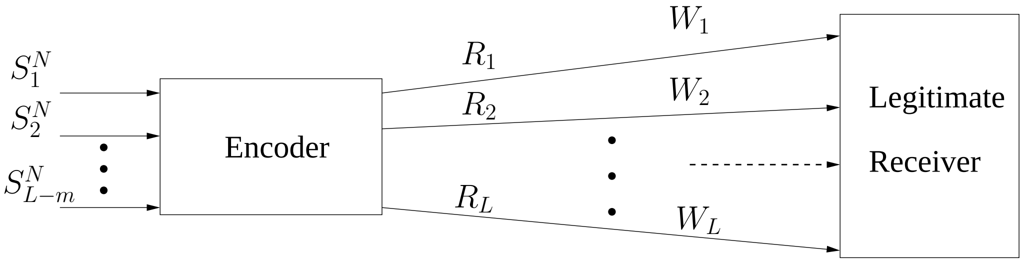


Fig. 37. The general problem considered in this section. Codeword sent on channel l is denoted by $W_l \in \{1, 2 \dots 2^{nR_l}\}$ $(l = 1, 2 \dots L)$.

can be reconstructed perfectly by accessing any subset of $(m+k)$ channels. The case of $m = 0$ is treated in [69].

Our main result is an exact characterization of the rate region for the case when $L = 3$ and $m = 1$, with two sources. Furthermore, we show that a 'separate' secure coding of information sources is optimal for this problem. For the more general case when L and m are arbitrary, we characterize a lower bound on the sum-rate and the result implies a natural separation strategy to achieve this lower bound.

1. System Model

A source encoder is presented with a sequence of source letters from i.i.d sources $\{S_1, S_2 \dots S_{L-m}\}$, drawn from the alphabet $\mathfrak{X}_1, \mathfrak{X}_2 \dots \mathfrak{X}_{L-m}$ respectively, as shown in Fig. 37. For each block of N letters from each of the sources (N arbitrary), there are L outputs, $f_l(S_1^N, S_2^N \dots S_{L-m}^N) = W_l \in \{1, 2 \dots 2^{NR_l}\}$ for $l = 1, 2 \dots L$. The codeword W_l is sent through the l^{th} $(l = 1, 2 \dots L)$ communication channel.

Let $\mathfrak{A}$ denote the collection of all subsets of channels of size m, which the eavesdropper has access to. We assume that the eavesdropper can access any member of $\mathfrak{A}$, but no more than one member of $\mathfrak{A}$. For $A \in \mathfrak{A}$, let Y_A be the codeword transmitted on the channel corresponding to the elements in A. It is desired that the eavesdropper does not get any information on any of the sources when it has access to Y_A, for all

$A \in \mathfrak{A}$. Denote the collection of all subsets of sources of size $\{1, 2 \ldots L - m\}$, by $\mathcal{S}$. For any $S \in \mathcal{S}$, the equivocation rate needs to satisfy the following:

$$H(S|Y_A) = H(S) \qquad \forall S \in \mathcal{S} \text{ and } \forall A \in \mathfrak{A} \tag{6.43}$$

Denote by $\mathcal{Z}_{m+k}$ which contains the collection of all subsets of channels of size $(m+k)$ $(1 \leq k \leq L - m)$. Since the legitimate receiver requires that sources $\{1, 2 \ldots k\}$ need to be reconstructed perfectly when it accesses any $B \in \mathcal{Z}_{m+k}$, it follows that:

$$H(S_1^N, S_2^N \ldots S_k^N | Y_B) = 0 \qquad \forall B \in \mathcal{Z}_{m+k} \quad \text{and} \quad 1 \leq k \leq L - m \tag{6.44}$$

where, Y_B is the codeword transmitted on the channels corresponding to elements in B. For simplicity, we henceforth denote S_k^N $(1 \leq k \leq L - m)$ by $\mathbf{S_k}$. We also define the following:

$$x \oplus y = \begin{cases} x + y & \text{if } x + y \leq 3 \\ x + y - 3 & \text{if } x + y > 3 \end{cases}$$

2. The Rate Region for the case when $L = 3$ and $m = 1$

In this case, we have two sources S_1, S_2 with a total of three communication channels and the eavesdropper having access to any one channel (See Fig. 37).

Theorem VI.6. *The rate region for the case with $L = 3$, $m = 1$, with two sources S_1, S_2 is given by the rate tuple (R_1, R_2, R_3) which satisfy the following:*

$$R_i \geq H(S_1) \quad \textit{for } i = 1, 2, 3. \tag{6.45}$$

$$R_i + R_j \geq 2H(S_1) + H(S_2) \quad \textit{for } 1 \leq i < j \leq 3 \tag{6.46}$$

$$R_1 + R_2 + R_3 \geq 3H(S_1) + \frac{3}{2}H(S_2) \tag{6.47}$$

We first prove some lemmas which will be useful in proving Theorem VI.6.

Lemma VI.8. *For* $1 \leq i \leq 3$,

$$H(W_i) \geq H(\mathbf{S_1}) + H(W_i|\mathbf{S_1}, W_{i\oplus 1})$$

Proof.

$$\begin{aligned} H(W_i) &\geq H(W_i|W_{i\oplus 1}) \\ &= H(W_i|W_{i\oplus 1}) + H(\mathbf{S_1}|W_i, W_{i\oplus 1}) \\ &= H(\mathbf{S_1}, W_i|W_{i\oplus 1}) \\ &= H(\mathbf{S_1}|W_{i\oplus 1}) + H(W_i|\mathbf{S_1}, W_{i\oplus 1}) \\ &= H(\mathbf{S_1}) + H(W_i|\mathbf{S_1}, W_{i\oplus 1}) \end{aligned}$$

where the first equality follows from the fact that the source $\mathbf{S_1}$ must be decodable by any of the two distinct channels and the last equality follows due to the secrecy constraints. □

Lemma VI.9. *For* $1 \leq i \leq 3$,

$$H(W_i|\mathbf{S_1}, W_{i\oplus 1}) + H(W_{i\oplus 1}|\mathbf{S_1}, W_{i\oplus 2}) \geq H(\mathbf{S_2})$$

Proof.

$$\begin{aligned} H(W_i|\mathbf{S_1}, W_{i\oplus 1}) + H(W_{i\oplus 1}|\mathbf{S_1}, W_{i\oplus 2}) &\geq H(W_i|\mathbf{S_1}, W_{i\oplus 1}, W_{i\oplus 2}) + H(W_{i\oplus 1}|\mathbf{S_1}, W_{i\oplus 2}) \\ &= H(W_i, W_{i\oplus 1}|\mathbf{S_1}, W_{i\oplus 2}) \\ &\overset{(a)}{=} \Big(H(W_i, W_{i\oplus 1}|\mathbf{S_1}, W_{i\oplus 2}) \ + \\ &\quad H(\mathbf{S_2}|W_i, W_{i\oplus 1}, W_{i\oplus 2}, \mathbf{S_1})\Big) \\ &= H(\mathbf{S_2}, W_i, W_{i\oplus 1}|\mathbf{S_1}, W_{i\oplus 2}) \\ &\geq H(\mathbf{S_2}|\mathbf{S_1}, W_{i\oplus 2}) \\ &= H(\mathbf{S_1}, \mathbf{S_2}|W_{i\oplus 2}) - H(\mathbf{S_1}|W_{i\oplus 2}) \end{aligned}$$

$$
\begin{aligned}
&\overset{(b)}{=} \quad H(\mathbf{S_1}, \mathbf{S_2}) - H(\mathbf{S_1}) \\
&\overset{(c)}{=} \quad H(\mathbf{S_2})
\end{aligned}
$$

where, (a) follows due to the fact that the source $\mathbf{S_2}$ needs to be decodable from all three channels, (b) due to secrecy constraints and (c) due to the sources being i.i.d. □

Now, we begin proving Theorem VI.6. (6.45) can be proved as follows:

$$NR_i \geq H(W_i) \geq H(\mathbf{S_1}) = NH(S_1)$$

where, the second inequality follows from Lemma VI.8. Now, we prove (6.46).

$$
\begin{aligned}
N(R_i + R_j) &\geq H(W_i) + H(W_j) \\
&\geq 2H(\mathbf{S_1}) + H(W_i|\mathbf{S_1}, W_{i\oplus 1}) + H(W_j|\mathbf{S_1}, W_{j\oplus 1}) \\
&\geq N(2H(S_1) + H(S_2))
\end{aligned}
$$

where the second inequality follows from Lemma VI.8, and the last inequality follows from Lemma VI.9. Finally, (6.47) follows from:

$$
\begin{aligned}
N(R_1 + R_2 + R_3) &\geq H(W_1) + H(W_2) + H(W_3) \\
&\geq 3H(\mathbf{S_1}) + H(W_1|\mathbf{S_1}, W_2) + H(W_2|\mathbf{S_1}, W_3) + H(W_3|\mathbf{S_1}, W_1) \\
&= 3H(\mathbf{S_1}) + \frac{1}{2}\Big[\Big(H(W_1|\mathbf{S_1}, W_2) + H(W_2|\mathbf{S_1}, W_3)\Big) + \\
&\Big(H(W_2|\mathbf{S_1}, W_3) + H(W_3|\mathbf{S_1}, W_1)\Big) + \Big(H(W_3|\mathbf{S_1}, W_1) + H(W_1|\mathbf{S_1}, W_2)\Big)\Big] \\
&\geq N(3H(S_1) + \frac{3}{2}H(S_2))
\end{aligned}
$$

where the second inequality follows from Lemma VI.8 and the last inequality follows from Lemma VI.9.

a. Separate Secure Coding of Sources is Optimal

We will show that, we can achieve the lower bounds in Theorem VI.6, by designing secure codes for sources S_1 and S_2 separately. We will use the technique in [69] to prove this claim. We write R_i $(1 \leq i \leq 3)$ as follows:

$$R_i = r_i^1 + r_i^2 \qquad 1 \leq i \leq 3 \tag{6.48}$$

where, r_i^1 and r_i^2 are the rate constraints for sources S_1 and S_2 respectively. Denote n to be number of channels from which the source S_i $(i = 1, 2)$ needs to be reconstructed. For source S_1, we then have $n = 2$, $L = 3$ and $m = 1$. From Chapter VI (Theorem VI.5 with $\mathrm{D}_0 = 0$), the rate region is given by the rate tuple (r_1^1, r_2^1, r_3^1) which satisfy the following:

$$r_i^1 \geq H(S_1) \quad for \quad 1 \leq i \leq 3 \tag{6.49}$$

For the source S_2, we have $n = L = 3$ and $m = 1$. The rate region is given by the rate tuple (r_1^2, r_2^2, r_3^2) which satisfy the following (from Chapter VI, Theorem VI.5 with $\mathrm{D}_0 = 0$):

$$\begin{aligned} r_1^2 + r_2^2 &\geq H(S_2) \\ r_2^2 + r_3^2 &\geq H(S_2) \\ r_1^2 + r_3^2 &\geq H(S_2) \end{aligned} \tag{6.50}$$

By taking into account (6.48) through (6.50) and also by the fact that:

$$r_1^2 + r_2^2 + r_3^2 \geq \frac{3}{2} H(S_2) \tag{6.51}$$

(which follows by summing all equations in (6.50)), we get the rate tuples (R_1, R_2, R_3) that satisfy Theorem VI.6.

3. Sum-Rate Lowerbound for Arbitrary L and m

In this section, we derive the sum-rate lowerbound for the case depicted in Fig. 37. The derivation closely follows the technique presented in [69], with the secrecy constraints enforced.

Theorem VI.7. *For the case of arbitrary L and m, with $L-m$ sources, the sum-rate satisfies the following:*

$$R_1 + \ldots R_L \geq L \sum_{i=m+1}^{L} \frac{H(S_{i-m})}{i-m}$$

Proof. Let T denote a subset of $\{1, 2 \ldots L\}$. We will prove that for $m < r \leq L$, the following holds:

$$\begin{aligned} \sum_{i=1}^{L} H(W_i) \quad &\geq \quad NL \sum_{i=m+1}^{r} \frac{H(S_{i-m})}{i-m} + \left\{ \frac{L}{\binom{L}{r+1-m}} \times \right. \\ &\left. \sum_{T:|T|=r+1-m} \frac{H(W_j, j \in T | \mathbf{S_1}, \mathbf{S_2} \ldots \mathbf{S_{r-m}}, Y_A, A \in \mathfrak{A}^T)}{r+1-m} \right\} \end{aligned} \tag{6.52}$$

where, $\mathfrak{A}^T$ denotes the collection of all subsets of channels that the eavesdropper has access to, which does *not* contain j for $j \in T$. Clearly, $|\mathfrak{A}^T| = \binom{L-|T|}{m}$, where $|\mathfrak{A}^T|$ represents the cardinality of the set $\mathfrak{A}^T$. Assume that (6.52) is true for $r = p$. We will now show that (6.52) holds for $r = p + 1$.

$$\begin{aligned} \sum_{i=1}^{L} H(W_i) \quad &\geq \quad NL \sum_{i=m+1}^{p} \frac{H(S_{i-m})}{i-m} + \left\{ \frac{L}{\binom{L}{p+1-m}} \times \right. \\ &\left. \sum_{T:|T|=p+1-m} \frac{H(W_j, j \in T | \mathbf{S_1}, \mathbf{S_2} \ldots \mathbf{S_{p-m}}, Y_A, A \in \mathfrak{A}^T)}{p+1-m} \right\} \\ &\overset{(a)}{=} \quad NL \sum_{i=m+1}^{p} \frac{H(S_{i-m})}{i-m} + \left\{ \frac{L}{\binom{L}{p+1-m}} \times \right. \end{aligned}$$

$$\sum_{T:|T|=p+1-m}\Big[\frac{H(W_j, j \in T|\mathbf{S_1}, \mathbf{S_2}\ldots\mathbf{S_{p-m}}, Y_A, A \in \mathfrak{A}^T)}{p+1-m} + \frac{H(\mathbf{S_{p+1-m}}|\mathbf{S_1}, \mathbf{S_2}\ldots\mathbf{S_{p-m}}, W_j, Y_A, j \in T, A \in \mathfrak{A}^T)}{p+1-m}\Big]\Big\}$$

$$= NL\sum_{i=m+1}^{p}\frac{H(S_{i-m})}{i-m} + \Big\{\frac{L}{\binom{L}{p+1-m}} \times \sum_{T:|T|=p+1-m}\frac{H(\mathbf{S_{p+1-m}}, W_j, j \in T|\mathbf{S_1}, \mathbf{S_2}\ldots\mathbf{S_{p-m}}, Y_A, A \in \mathfrak{A}^T)}{p+1-m}\Big\}$$

$$= NL\sum_{i=m+1}^{p}\frac{H(S_{i-m})}{i-m} + \Big\{\frac{L}{\binom{L}{p+1-m}} \times \sum_{T:|T|=p+1-m}\Big[\frac{H(\mathbf{S_{p+1-m}}|\mathbf{S_1}, \mathbf{S_2}\ldots\mathbf{S_{p-m}}, Y_A, A \in \mathfrak{A}^T)}{p+1-m} + \frac{H(W_j, j \in T|\mathbf{S_1}, \mathbf{S_2}\ldots\mathbf{S_{p+1-m}}, Y_A, A \in \mathfrak{A}^T)}{p+1-m}\Big]\Big\}$$

$$= NL\sum_{i=m+1}^{p}\frac{H(S_{i-m})}{i-m} + \frac{L}{\binom{L}{p+1-m}}\sum_{T:|T|=p+1-m}\Big[\frac{H(\mathbf{S_1}, \ldots\mathbf{S_{p+1-m}}|Y_A, A \in \mathfrak{A}^T)}{p+1-m} - \frac{H(\mathbf{S_1}, \ldots\mathbf{S_{p-m}}|Y_A, A \in \mathfrak{A}^T)}{p+1-m} + \frac{H(W_j, j \in T|\mathbf{S_1}, \mathbf{S_2}\ldots\mathbf{S_{p+1-m}}, Y_A, A \in \mathfrak{A}^T)}{p+1-m}\Big]$$

$$\overset{(b)}{=} NL\sum_{i=m+1}^{p}\frac{H(S_{i-m})}{i-m} + \frac{L}{\binom{L}{p+1-m}}\sum_{T:|T|=p+1-m}\Big[\frac{H(\mathbf{S_{p+1-m}})}{p+1-m} + \frac{H(W_j, j \in T|\mathbf{S_1}, \mathbf{S_2}\ldots\mathbf{S_{p+1-m}}, Y_A, A \in \mathfrak{A}^T)}{p+1-m}\Big]$$

$$= NL\sum_{i=m+1}^{p+1}\frac{H(S_{i-m})}{i-m} + \Big\{\frac{L}{\binom{L}{p+1-m}} \times \sum_{T:|T|=p+1-m}\Big[\frac{H(W_j, j \in T|\mathbf{S_1}, \mathbf{S_2}\ldots\mathbf{S_{p+1-m}}, Y_A, A \in \mathfrak{A}^T)}{p+1-m}\Big]\Big\}$$

$$\overset{(c)}{\geq} NL\sum_{i=m+1}^{p+1}\frac{H(S_{i-m})}{i-m} + \Big\{\frac{L}{\binom{L}{p+2-m}} \times \sum_{T:|T|=p+2-m}\frac{H(W_j, j \in T|\mathbf{S_1}, \mathbf{S_2}\ldots\mathbf{S_{p+1-m}}, Y_A, A \in \mathfrak{A}^T)}{p+2-m}\Big\}$$

where (a) follows from the decodability of the source $\mathbf{S_{p+1-m}}$ given a subset of message $\{W_1, \dots W_L\}$ of size $(p+1)$, (b) follows due to secrecy constraints and the fact that all sources are i.i.d and (c) from the application of Han's inequality [66]. □

D. Conclusion

We have addressed the secrecy capacity of a source coding problem wherein information needs to be transmitted from a source to a legitimate receiver amidst an eavesdropper with the legitimate receiver endowed with diversity coding. The rate region was precisely characterized for simple cases and schemes using linear network codes were proposed to achieve the lower bound. We showed that we could construct a partial secure code by simply concatenating a perfect secure network code with another network code of information bits. For the case of arbitrary m, n, L, we derived a lower bound on the sum-rate and proposed simple schemes to achieve it. Furthermore, the concept of secure-diverse coding was extended for the problem with multiple sources. Again, the rate region was derived for a simple case, while a sum-rate lower bound was provided for the general case.

CHAPTER VII

CONCLUSION

In this dissertation, we have looked at some fundamental limits of cooperative scheduling systems, cognitive radio systems and the secrecy capacity of a source coding problem.

In Chapter II, we have given a brief introduction to the general principles of cooperative and secrecy systems. Chapter III dealt with finding optimal scheduling schemes for a downlink (i.e., non-cooperative) and cooperative systems. It was interesting to note that the optimal scheduling scheme for a cooperative system belongs to the same class as that of non-cooperative systems. Furthermore, it was shown that the cooperative scheduling system enlarges the achievable rate region. The concept of finding optimal scheduling scheme was extended to the case of delay constrained systems in Chapter IV. Thus the results of Chapter III and Chapter IV unifies the capacity achieving scheduling schemes for any delay constraints.

In Chapter V, we analyzed the cognitive radio system from a link layer perspective. To this end, we modeled the cognitive radio as a priority queuing system (by treating the secondary user as a lower priority user), with constant input arrival rate and variable service rate (to model the time varying wireless channel). By using a simple dual structure of the counterpart priority queuing system (i.e., with variable input rate and constant service rate), we characterized the input rates that the primary and secondary user could achieve for the original cognitive radio system.

Chapter VI looked at secrecy capacity of a source coding problem for single and multiple sources. We started looking at the rate region of a problem wherein a source needs to be transmitted to a legitimate receiver through a total of L channels with the eavesdropper having access to any subset of m channels and the legitimate receiver being able to decode the source completely from any subset of n channels (where $m < n \leq L$). We proposed simple schemes that the encoder may use to communicate with the legitimate receiver using linear network codes. We extended this problem to the case when there are multiple sources to be communicated to the legitimate receiver. The upshot of the study for this problem is the following: Performing separate secure coding of individual sources is optimal for the case of two sources and in general for achieving the sum-rate point for arbitrary number of sources.

It may be noted in Chapter VI, that the level of secrecy for a source was fixed (i.e., the problem that we have addressed is that *each* source has to be perfectly secure to any subset of m channels irrespective of the source). However, we can view this problem in a different perspective. Suppose that the secrecy level desired for each source is different. For example, if the source S_k, for $k < L$ has to be secure against k out of a total of L channels. Then we could ask what would be the optimal scheme to communicate in this scenario? One possible solution is to construct a separate code for each of the sources as we did in Chapter VI. But another line of thought is the following: Note that the secrecy level of S_k increases with k (akin to the *degraded broadcast* framework where there is a strict ordering in the channel of users). A natural question to ask where there is an ordering in the secrecy levels of users is the following: Can we construct a *embedded secure code* wherein the weaker secure code is embedded in a stronger secure code (analogous to the code construction of degraded broadcast channels, wherein the message of the weaker user is embedded in the message of the stronger user)? We strongly conjecture the existence of embedded secure codes, however, neither we have been able to prove the converse nor come up with a scheme for this scenario. It would be interesting to answer the non-existence or existence of secure embedded codes.

REFERENCES

[1] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, pp. 623–656, Oct. 1948.

[2] L. Zheng and D. Tse, “Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels,” *IEEE Transactions on Information Theory*, vol. 49, pp. 1073–1096, May 2003.

[3] S. M. Alamouti, “A simple transmit diversity technique for wireless communications,” *IEEE Journal on Select Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[4] A. Sendonaris, E. Erkip and B. Aazhang, “User cooperation diversity–Part I: System description,” *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1938, Nov. 2003.

[5] A. Sendonaris, E. Erkip and B. Aazhang, “User cooperation diversity–Part II: Implementation aspects and performance analysis,” *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1939–1948, Nov. 2003.

[6] J. N. Laneman, D. N. C. Tse and G. W. Wornell, “Cooperative diversity in wireless networks: Efficient protocols and outage behaviour,” *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 3518–3539, Dec. 2005.

[7] J. N. Laneman and G. W. Wornell, “Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks,” *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.

[8] Lingjia Liu, J. F. Chamberland and S. L. Miller, "User cooperation in the absence of phase information at the transmitters," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1197–1206, Mar. 2008.

[9] P. Viswanath, D. N. C. Tse and R. Laroia "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277–1294, June 2002.

[10] R. Berry and R. Gallager "Communication over fading channels with delay constraints," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1135–1149, May 2002.

[11] L. Liu, P. Parag, J. Tang, W. Y. Chen and J. F. Chamberland "Resource allocation and quality of service evaluation for wireless communication systems using fluid models," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1767–1777, May 2007.

[12] L. Liu and J. F. Chamberland, "On the effective capacities of multiple-antenna Gaussian channels," *Proc. ISIT*, July 2008, pp. 2583–2587.

[13] L. Liu, P. Parag and J. F. Chamberland "Quality of service analysis for wireless user-cooperation networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3833–3842, Oct. 2007.

[14] C. E. Shannon, "Communication theory of secrecy systems," *The Bell System Technical Journal*, vol. 28, pp. 656–715, Oct. 1949.

[15] A. D. Wyner, "The wire-tap channel," *The Bell System Technical Journal*, vol. 54, pp. 1355–1387, Oct. 1975.

[16] S. K. L. Cheong and Martin E. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. Inf. Theory*, vol. 24, no. 4, pp. 451–456, July 1978.

[17] P. K. Gopala, L. Lai and Hesham El Gamal "On the secrecy capacity of fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4687–4698, Oct. 2008.

[18] Y. Liang and H. V. Poor, "Multiple access channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 54, pp. 976–1002, Mar. 2008.

[19] E. Tekin and A. Yener, "The Gaussian multiple access wire-tap channel," *IEEE Trans. Inf. Theory*, vol. 54, pp. 5747–5755, Dec. 2008.

[20] G. Bagherikaram, A. S. Motahari and A. K. Khandani "The secrecy rate region of the broadcast channel," *http://arxiv.org/abs/0806.4200*, Oct. 2009.

[21] T. Liu and S. Shamai "A note on the secrecy capacity of the multi-antenna wiretap channel," *http://arxiv.org/abs/0710.4105*, Oct. 2009.

[22] A. Khisti and G. Wornell, "Secure transmission with multiple antennas: The MISOME wiretap channel," *http://arxiv.org/abs/0708.4219*, Oct. 2009.

[23] Y. Liang, A. Somekh-Baruch, H. V. Poor, S. Shamai (Shitz) and S. Verdu , "Capacity of cognitive interference channels with and without secrecy," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 604–619, Feb. 2009.

[24] A. Balasubramanian, L. Liu, S. L. Miller and J. F. Chamberland, "The rate region of a cooperative scheduling system," *Proceedings of the Conference on Information Sciences and Systems (CISS)*, Baltimore, MD, Mar. 2007, pp. 576–580.

[25] A. Balasubramanian, L. Liu and Scott Miller, "The rate region of a cooperative scheduling system," *IEEE Trans. Wireless Commun.*, in press.

[26] A. Balasubramanian and S. L. Miller, "On optimal scheduling for time division systems with quality of service constraints," *Proceedings of the Conference on*

Information Sciences and Systems (CISS), Baltimore, MD, Mar. 2009, pp. 719–722.

[27] A. Balasubramanian and S. L. Miller, "The effective capacity of a time division downlink scheduling system," *IEEE Trans. Commun.*, in press.

[28] A. Balasubramanian, T. Liu and S. L. Miller, "Multilevel diversity coding with secrecy constraints," *Proceedings of the Information Theory and Applications (ITA) Workshop*, La Jolla, CA, Feb. 2009, pp. 216–219.

[29] Anwar I. Elwalid and D. Mitra, "Effective bandwidth of genral Markovian traffic sources and admission control of high speed networks," *IEEE Trans. Networking*, vol. 1, no. 3, pp. 329–343, June 1993.

[30] R. W. Brockett, "Lecture notes: Stochastic control," Harvard University, Cambridge, MA, Tech. Rep., 1983.

[31] R. Knopp and P. Humblet, "Information capacity and power control in single cell multiuser communications," *Proc. IEEE Int.Computer Conference (ICC '95)*, Seattle WA, June 1995, pp. 331–335.

[32] X. Liu, E. Chong and N. Shroff, "Opportunistic transmission scheduling with resource sharing constraints in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, pp. 2053–2064, Oct. 2001.

[33] M. Sharif and Babak Hassibi, "A delay analysis for opportunistic transmission in fading broadcast channels," *Proc. IEEE INFOCOM*, March 2005, vol. 4, pp. 2720–2730.

[34] M. Sharif and B. Hassibi, "Delay considerations for opportunistic scheduling in broadcast fading channels," *IEEE Trans. Wireless Communications*, vol. 6, no. 6,

pp. 3353–3363, Sep. 2007.

[35] E. Yeh and A. S. Cohen, “Throughput and delay optimal resource allocation in multiaccess fading channels,” *Proc. ISIT*, July 2004, pp. 245–245.

[36] M. Agrawal and A. Puri, “Base station scheduling of requests with fixed deadlines,” *Proc. INFOCOM*, June 2002, vol. 2, pp. 487–496.

[37] A. Eryilmaz, R. Srikant, and J. Perkins, “Stable scheduling policies for broadcast channels,” *Proc. ISIT*, July 2002, pp. 382–382.

[38] A. Ganti, E. Modiano, and J. Tsitsiklis, “Optimal transmission scheduling in symmetric communication models with intermittent connectivity,” *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 998–1008, Mar. 2007.

[39] H. Viswanathan and K. Kumaran, “Rate scheduling in multiple antenna downlink wireless systems,” *IEEE Trans. Communications*, vol. 53, no. 4, pp. 645–655, Apr. 2005.

[40] X. Qin and R. Berry, “Exploiting multiuser diversity for medium access control in wireless networks,” *Proc. INFOCOM*, Apr. 2003, vol. 2, pp. 1084–1094.

[41] P. Liu, R. Berry and M. Honig, “A fluid analysis of a utility-based wireless scheduling policy,” *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 2872–2889, July 2006.

[42] V. Tsibonis and L. Georgiadis, “Optimal downlink scheduling policies for slotted wireless time-varying channels,” *IEEE Trans. Wireless Communications*, vol. 4, no. 4, pp. 1808–1817, July 2005.

[43] D. N. C. Tse and P. Viswanath, *Fundamentals of wireless communications*. London, U.K.: Cambridge University Press, 2005.

[44] L. Li and Andrea Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels-Part I: Ergodic capacity," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1083–1102, Mar. 2001.

[45] L. Liu, J. F. Chamberland and Scott Miller, "The uplink achievable rate region of a user cooperation scheme," *Proceedings on the 9th Canadian Workshop on Information Theory*, Montreal, Quebec, June 2005, pp. 163–166.

[46] Y. Liang, and V. V. Veeravalli, "Cooperative relay broadcast channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 900–928, Mar. 2007.

[47] D. Tse, "Optimal power allocation over parallel Gaussian channels," *Proc. ISIT*, Ulm, Germany, July 2004, pp. 27–27.

[48] S. Vishwanath, N. Jindal and A. Goldsmith, "Duality, achievable rates and sum rate capacity of Gaussian MIMO broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[49] G. B. Arfken, *Mathematical methods for physicist.* Orlando: Academic Press, 1985.

[50] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[51] Jia Tang and Xi Zhang, "Quality-of-service driven power and rate adaptation for multichannel communications over wireless links," *IEEE Tran on Wireless Commun.*, vol. 6, no. 12, pp. 4349–4360, Dec. 2007.

[52] D. Wu and R. Negi, "Effective capacity: A wireless link model for support of quality of service," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 630–643, July 2003.

[53] R. Tandra and A. Sahai, “SNR walls for signal detection,” *IEEE Journal on Special Topics in Signal Processing*, vol. 2, pp. 4–17, Feb. 2008.

[54] G. Atia, A. Sahai and V. Saligrama, “Spectrum enforcement and liability assignment in cognitve radio systems,” *IEEE Symposium on Dynamic Spectrum Access Networks*, Oct. 2008, pp. 1–12.

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

[56] A. Jovicic and P. Viswanath, “Cognitive radio: An information-theoretic perspective,” *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 3945–3958, Sep. 2009.

[57] G. de Veciana, C. Courcoubetis and J. Warland, “Effective bandwidths for networks: A decomposition approach to resource management,” *INFOCOM 1994*, June 1994, vol. 2, pp. 466–473.

[58] V. G. Kulkarni and N. Gautam, “Admission control of multi-class traffic with service priorities in high-speed networks,” *Queuing Systems: Theory and Applications*, vol. 27, no. 1-2, pp. 79–97, Dec. 1997.

[59] R. W. Brockett, W. Gong and Y. Guo, “Stochastic analysis for fluid queuing systems,” *Proceedings of the* 38th *Conference on Decision & Control*, Dec. 1999, vol. 3, pp. 3077–3082.

[60] R. Ahlswede and I. Csiszár, “Common randomness in information theory and cryptography, Part I: Secret sharing,” *IEEE Trans. Inf. Theory*, vol. 39, pp. 1121–1132, July 1993.

[61] I. Csiszár and P. Narayan, “Secrecy capacities for multiple terminals,” *IEEE Trans. Inf. Theory*, vol. 50, no.12, pp. 3047–3061, Dec. 2004.

[62] U. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 734–742, May 1993.

[63] L. Ozarow, "On a source-coding problem with two channels and three receivers," *The Bell System Tech Journal*, vol. 59, no. 10, pp. 1909–1921, Dec. 1980.

[64] D. Gündüz, E. Erkip and H. Vincent Poor, "Lossless compression with security constraints," *Proc. of ISIT 2008*, June 2008, pp. 111–115.

[65] W. Luh and D. Kundur, "Distributed secret sharing for discrete memoryless networks," *IEEE Trans. Information Forensics and Security*, vol. 3, no. 3, pp. 424–430, Sep. 2008.

[66] T. M.Cover and J. A. Thomas, *Elements of information theory*. New York: Wiley, 1991

[67] N. Cai and R. W. Yeung, "Secure network coding," *http://iest2.ie.cuhk.edu.hk/ whyeung/publications*, Oct. 2009.

[68] R. W. Yeung, "Multilevel diversity coding with distortion," *IEEE Trans. Inf. Theory*, vol. 41, no. 2, pp. 412–422, May 1995.

[69] J. R. Roche, R. W. Yeung and K. P. Hau, "Symmetrical multilevel diversity coding," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 1059–1064, May 1997.

VITA

Anantharaman Balasubramanian was born in Chennai, India. He received his bachelor's degree from the College of Engineering, Chennai, and his master's degree from the Indian Institute of Science, Bangalore, all in Electrical Engineering. In December 2009, he graduated with his Ph.D. in Electrical & Computer Engineering at Texas A&M University. His research interests are in the broad areas of signal processing and communications.

His address is: 241G WERC, Department of Electrical and Computer Engineering, 3128 TAMU, College Station, TX 77843-3128. He can be contacted through email: anantharamb@gmail.com.

The typist for this dissertation was Anantharaman Balasubramanian.