Contention Resolution for Opportunistic Scheduling and General Utility Maximization

A THESIS

submitted by

VAISHAKH J

for the award of the degree

of

MASTER OF SCIENCE

(by Research)

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS.

OCTOBER 2013

THESIS CERTIFICATE

This is to certify that the thesis titled **Contention Resolution for Opportunistic Scheduling and General Utility Maximization**, submitted by **Vaishakh J** to the Indian Institute of Technology Madras for the award of the Degree of **Master of Science by Research**, is a bona fide record of the research work done by him under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University, for the award of any Degree or Diploma.

Dr. Venkatesh Ramaiyan
Research Guide
Assistant Professor
Dept. of Electrical Engineering
IIT Madras
Chennai 600 036.

Place: Chennai
Date: 14 October 2013

ACKNOWLEDGEMENTS

I would like to convey my inexpressible gratitude to my guide, Dr. Venkatesh Ramaiyan for his constant support, and invaluable advice throughout the duration of my thesis work. I am extremely fortunate to do masters under his supervision. The interest he put in my work has been a constant source of inspiration for me. I also would like to thank my committee members Dr. Srikrishna Bhashyam and Dr. Krishna Moorthy Sivalingam for the insightful discussions during the committee meeting.

I sincerely appreciate the support and encouragement from my parents. I would like to thank my sister and brother-in-law for encouraging me to pursue the career of my interest.

I also want to thank the staff of electrical engineering department for providing and maintaining the facilities in lab without fail. I also want to thank my lab members Easwar, Suman, Sundaram, Ram and Vishal for having a good time together. I would also like to thank Dr. T. M. Muruganandam and Arun Prakash for the intense Hapkido sessions. Besides, my sincerest thanks to my friends Ajesh, Anoop, Arun, Deepak, Haseen, Ishaque, Jayesh, Jithin, Kuttalu, Nikhil, Sandeep, Sreekanth, Umesh and Vineed for making my stay at IITM an enjoyable one.

ABSTRACT

We study the contention resolution problem for opportunistic scheduling in a cellular network scenario. We consider a single cell of a cellular wireless network with a fixed number of users. We assume that time is slotted and the base station schedules a single user in a slot. The wireless channel between the base station and the users is assumed to fade randomly. In this setup, the base station aims to resolve the contention and schedule the user with the favourable channel condition in every slot. In this thesis, we study splitting based contention resolution strategies. In particular, we study a generalization of the opportunistic splitting algorithm called the maximal probability allocation (MPA) scheme for contention resolution. MPA attempts to maximize the probability of success in contention resolution in every attempt. We characterize the performance of the greedy MPA strategy and comment on its optimality for a variety of network scenarios.

Opportunistic contention resolution by splitting involves identifying a channel threshold between the user with best channel and the second best channel. Hence, the problem of opportunistic splitting permits a formulation as a source coding problem for the random threshold and we can relate the average delay (of the contention resolution) with the entropy (of the threshold). In our work, we study the correlation between the average delay of a contention resolution strategy and the average entropy of the strategy as well.

In Chapter 2, we study the performance of MPA for i.i.d. channel of users. We show that MPA need not be delay optimal or entropy optimal but is a very good approximation for the i.i.d. case. Further, we observe that the entropy optimal strategy need not be delay optimal. In Chapter 3, we study the performance of MPA for non-identically distributed channels and for correlated wireless channels. Here again, we show that MPA need not be delay optimal or entropy optimal. We characterize the performance of MPA as a function of the channel characteristics. For example, we show that the average delay of contention resolution for the non-identically distributed channel is always lesser than the average delay of contention resolution for the i.i.d. case. Also, we

show that MPA can be a severely suboptimal strategy especially for correlated wireless channel distributions. In all the cases, we compare the average delay of the contention resolution with the entropy of the strategy to understand the applicability of the entropy minimization framework for the contention resolution problem. We observe that there is a good correlation between the average delay of a strategy and its entropy and we obtain a bound between them as well. We also study a formulation of the entropy minimization problem and comment on the feasibility of identifying delay minimizing solutions.

In Chapter 4, we extend the study to different network scenarios including different thresholds and multiple feedback case and study generalizations of the maximal probability allocation strategy. Using simulations, we also report the performance of the MPA strategy for a variety of network scenarios and compare it with other popular contention resolution strategies such as polling and channel gain based random access as well.

Finally, in Chapter 5, we study general utility maximization in a cellular setup using an estimate of the rate region. We generalize the applicability of a rate region based scheduler RRS and propose its use for a number of interesting network scenarios like utility maximization for extended arrival rates, for feedback controlled arrival process and for the energy minimization problem.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ACK	Acknowledgment
AMC	Adaptive modulation and coding
AP	Access point
CDF	Cumulative distribution function
CDMA	Code division multiple access
CSI	Channel state information
CSMA	Carrier sense multiple access
HDR	High data rate
i.i.d.	Independent and identically distributed
Kbps	Kilo bits per second
LTE	Long term evolution
MAC	Medium access control
Mbps	Mega bits per second
MPA	Maximal probability allocation
O-CSMA	Opportunistic CSMA
OSA	Opportunistic splitting algorithm
PDF	Probability density function
PF	Proportional fair
PHY	Physical layer
Q-CSMA	Quantile-based CSMA
QoS	Quality of service
SNR	Signal to noise ratio
RHS	Right hand side
RRS	Rate region scheduler
WLAN	Wireless local area network

CHAPTER 1

Introduction

The rise in popularity of 3G/4G cellular wireless standards such as WiMAX IEEE802.16 (2006) and LTE 3GPP (2012) and the widespread deployment of IEEE 802.11 WLANs IEEE802.11 (2012) has led to easy and convenient broadband wireless data access for the users. The increasing demand in wireless access due to the ever increasing number of wireless users and applications coupled with the limitations of the wireless channel necessitates judicious and optimal use of the available wireless resources. Unlike wireline systems, the capacity of the wireless channel cannot be increased arbitrarily. Further, the wireless channel exhibits temporal variations in the channel capacity. The fading statistics seen by users need not be identically distributed and can be time varying as well. These unique characteristics of wireless channel makes the resource allocation problem challenging.

One of the popular and an optimal scheduling strategy for data traffic in wireless networks is the opportunistic scheduling strategy. Opportunistic scheduling (see Knopp and Humblet (1995)) involves scheduling a user when the channel condition of the user is the most favorable among the other users. This scheme exploits the time varying nature of the channel, and improves system throughput using multiuser diversity. Generalizations to opportunistic scheduling attempts to provide fairness to channel access while opportunistically scheduling the users (e.g., Jalali *et al.* (2000)).

Implementation of an opportunistic scheduler involves identifying a user with the highest metric. The metric usually depends on the application, the quality of service requirement or the network utility function. In opportunistic scheduling, the metric is the instantaneous channel rate. i.e., in every slot, the user with the highest supported rate is scheduled. The time overhead in finding the user with best channel increases linearly with the number of users in the network. Feedback overhead can be reduced by polling a subset of users in the network, and allocating resources opportunistically among the users in the subset (e.g., Gopalan *et al.* (2012)). In such schemes, opportunism is exploited partially, and hence it degrades the achievable rate region. Contention based feedback mechanism is an alternate strategy to reduce feedback overhead, where users contend for a common resource opportunistically. In contention based schemes, there is a possibility of not resolving contention over the duration of a time slot (especially when channel state values are correlated). But in polling, some degree of opportunism is always exploited.

In Qin and Berry (2004), Qin and Berry propose a contention based splitting algorithm for opportunistic scheduling where users transmit its channel state information to the base station if its instantaneous rate is above a certain threshold. For i.i.d. channels and for any number of users, they showed that opportunistic contention resolution can be resolved on an average in 2.5 minislots with feedback from the base station; this is a considerable improvement in feedback overhead compared to the polling scheme. In reality, the channel states of users need not be identically distributed or independent. The distributions are usually asymmetric due to the difference in users' location, mobility pattern, shadowing etc. Users near the base station generally experience a better channel compared to the users near the cell edge. For this reason, we study the performance of splitting algorithms for generalized channel states including correlated channels. We also propose generalizations of splitting algorithms and discuss an entropy based approach to identify thresholds to minimize average delay.

Most implementations of opportunistic schedulers involves a gradient scheduler that seeks to maximize a given concave network utility in the long term average sense, e.g., Stolyar (2005). However, the gradient based schedulers fails to maximize non-differentiable or non-concave utilities and their applicability is limited. In Naveen and Ramaiyan (2013), Naveen and Ramaiyan propose a scheduling algorithm RRS, based on the estimated rate region, which can maximize general network utilities. The algorithm is proposed for saturated traffic of a fixed number of users. Using simulations, we show that RRS can be used in unsaturated network scenarios as well. In this work, we study the application of RRS for network utility maximization for extended arrival rates (outside the rate region), for feedback controlled arrival process and for the energy minimization problem.

1.1 Related Literature

A lot of research has been carried out in the area of scheduling under fading channel. The capacity of a single link of a fading wireless channel with average power constraint was studied in Goldsmith and Varaiya (1997), under different channel side information. For a multiuser network, Knopp and Humblet (1995) proposed a power control scheme to improve the system capacity. The scheduling scheme, known as max-rate, suffered from being unfair to users with poorer channel conditions. The issue of fairness was addressed in Jalali *et al.* (2000), where an implementation of proportional fairness was studied. Implementation of other fairness notions like minimum throughput guarantee, temporal fairness are discussed in Liu *et al.* (2003); Andrews *et al.* (2005), etc. In Liu *et al.* (2003), Liu et al discuss a framework for opportunistic scheduling and propose opportunistic scheduling strategies that guarantee minimum throughput and time fairness. Algorithm for optimizing concave utilities with lower and upper throughout bounds is studied in Andrews *et al.* (2005). In Stolyar (2005), Stolyar discusses the asymptotic optimality of gradient scheduling algorithms for multiuser networks using fluid sample path techniques.

The use of channel state information for stabilizing the queues in a network has also attracted considerable attention. In Tassiulas and Ephremides (1992), Tassiulas and Ephremides propose a max-weight scheduling policy that can stabilize all arrival rates within the rate region of the wireless network. In Neely *et al.* (2005), Neely et al studied a dynamic power allocation and routing strategy for multihop wireless network that generalized the network model studied in Tassiulas and Ephremides (1992). In Shakkottai and Stolyar (2002), Shakkottai and Stolyar proposed a delay optimal stabilizing strategy called the Exponential rule for a multiuser wireless network. An energy optimal policy for a time varying wireless newtork was studied in Neely (2006). In Neely *et al.* (2008), Neely proposed a method of dynamic resource allocation for all traffic whenever possible, maximizing a concave utility function on the rate region. A work on utility maximization for feedback controlled arrival process is studied in Eryilmaz and Srikant (2005). The feedback is a function of the queue length, and the scheduler allocates resources depending on the instantaneous channel rate. In Chapter 5, we discuss an implementation based on RRS (see Naveen and Ramaiyan (2013)) for maximizing general network utilities in a variety of such network scenarios.

Schedulers that use channel state information may require a complete feedback of the information from all the users to the base station at the beginning of every slot. Polling or piggy-backing is a simple technique that can permit such feedback from the users. While polling is an effective strategy to feed back CSI, the feedback load may limit the performance of the system especially for large N. In Gesbert and Alouini (2004), authors provided a theoretical analysis of feedback load and proposed a scheme to reduce the feedback overhead. In the scheme discussed in Gesbert and Alouini (2004), only the users with channel quality above a threshold are allowed to transmit. Thresholds are optimized to achieve some outage probability. If no user is above threshold, a random user is selected. In Sanayei and Nosratinia (2005), the effect of 1 bit feedback was studied. The authors showed that the same capacity growth can be achieved with 1 bit feedback. The effect of the 1 bit feedback with possibilities of error was studied in Xue and Kaiser (2007).

A random access feedback protocol is proposed in Tang and Heath (2005) for contention resolution. Users send a feedback message in a minislot with some fixed probability. The total number of feedback slot is fixed. A successful transmission happens if exactly one user transmits. Otherwise the scheduler polls a user randomly. The model was extended to include multiple access in So (2009). The model of channel state dependent random access was originally studied in Aloha networks. In Jahn and Bottcher (1993), the authors used a channel state dependent access probability for slotted ALOHA protocol. The channel access probabilities were chosen in a heuristic manner. Qin and Berry, in Qin and Berry (2001), proposed a channel state dependent slotted ALOHA protocol with transmission power constraint. A modified slotted ALOHA protocol with a CSI dependent channel access probability was studied in Adireddy and Tong (2005). Asymptotic throughput is evaluated by considering both population dependent and population independent channel access functions. In Kim *et al.* (2011), Kim et al proposed and analyzed two variants of channel aware slotted CSMA, Opportunistic CSMA(O-CSMA) and Quantile-based CSMA(Q-CSMA).

Timer based scheme is another method to reduce the feedback load for contention resolution. In timer based schemes, each node selects a timer depending on its metric and will attempt after the expiry of the timer. A single node transmitting in a slot is considered to have won the contention. In Bletsas *et al.* (2006), an inverse timer mapping is proposed for opportunistic relaying. An optimal scheme in this scenario was studied in Shah *et al.* (2010). It is shown that the optimal mapping maps the metric to discrete timer values.

Splitting is a popular contention resolution strategy originally proposed for scheduling random arrivals in a network. The idea of splitting was first proposed in the context of ALOHA networks by Gallager in Gallager (1978). In Arrow *et al.* (1981), Arrow et al, studied the problem of resolving the user with highest value in a sample using binary type questions. The nature of the optimal sequence of questions for i.i.d. sample values was characterized in Anantharam and Varaiya (1986). In Qin and Berry (2004), Qin and Berry used the idea of splitting for opportunistic contention resolution. They proposed the opportunistic splitting algorithm for i.i.d. wireless channel and characterized the average delay performance of the algorithm. The performance of splitting for a heterogeneous wireless channel with fairness constraint is studied in Qin and Berry (2006). Unlike the work in Arrow *et al.* (1981) which assumed the knowledge of the exact number of users contending at a time, Qin and Berry (2004) assumed a ternary feedback model and studied the performance of the greedy strategy.

In Ramaiyan (2013), Ramaiyan proposes a generalization of the opportunistic splitting algorithm called the maximal probability allocation (MPA) scheme. In this thesis, we study the delay performance of MPA strategy for a variety of network scenarios and comment on its optimality. The work Ramaiyan (2013) also discusses a source coding framework relevant to the opportunistic contention resolution problem. The average delay to resolve contention is related to the entropy of the contention resolution strategy. In this thesis, we characterize the correlation between the average delay and the entropy. We also formulate a entropy minimization framework and characterize MPA as a local minima of the optimization problem.

In Gopalan *et al.* (2012), Aditya et al studied the effect of partial feedback scheme and proposed an algorithm which can stabilize the arrivals in the achievable rate region. The degradation in the capacity region due to lack of full feedback was also characterised. Another work on joint scheduling and channel probing algorithm is reported in Karaca *et al.* (2012). In this thesis, we assume that the contention resolution is perfect. In a chapter on performance evaluation, using simulations, we report the impact of partial resolution on the throughput performance.

1.2 Outline of the Thesis

In Chapter 2, we describe the maximal probability allocation strategy and the source coding framework for the contention resolution problem (from Ramaiyan (2013)). We study the optimality of MPA for i.i.d. channel scenario. We characterize the contention resolution strategy based on the average delay performance as well as the entropy of the strategy.

In Chapter 3, we study the performance of MPA for non i.i.d. channel scenarios. We consider independent and non-identically distributed channel as well as correlated wireless channel. We characterize the performance of the MPA as a function of the channel distributions. We study the correlation between the average delay of a strategy and its entropy and we obtain a bound between them as well. We also propose a formulation of the entropy minimization problem and comment on the feasibility of identifying delay minimizing solutions.

In Chapter 4, we extend the study to different network scenarios including multiple thresholds and multiple feedback cases and propose generalizations of the maximal probability allocation strategy. Using simulations, we characterize the performance of MPA for a variety of network scenarios and compare the performance with other contention resolution strategies such as polling and channel value based random access.

Finally, in Chapter 5, we study general utility maximization problem in a cellular setup using a rate region based scheduler called RRS. We generalize the applicability of RRS and propose its use for a number of interesting network scenarios like utility maximization for extended arrival rates, for feedback controlled arrival process and for the energy minimization problem.

In Chapter 6, we conclude the thesis.

CHAPTER 2

Maximal Probability Allocation for i.i.d. Channel

In this chapter, we will study the performance of the maximal probability allocation scheme, proposed in Ramaiyan (2013), for the i.i.d. channel. In this thesis, we will restrict to splitting based algorithms for opportunistic contention resolution. In Section 2.2, we will first discuss a popular splitting based contention resolution algorithm called the opportunistic splitting algorithm (OSA) studied in Qin and Berry (2004). Then, in Section 2.3, we will consider a generalization of OSA called the maximal probability allocation (MPA) strategy, and comment on the source coding framework for opportunistic contention resolution (all from Ramaiyan (2013)). In Section 2.4, we will study the optimality of the MPA strategy in terms of the average delay to resolve contention and the entropy of the threshold.

2.1 Network Model

We consider a single cell of a cellular wireless network with a base station or an access point and a fixed number of users, N, in an infrastructure setup of traffic. We assume that the wireless channel is slotted and the nodes are synchronized to the slots. The users time share the slotted wireless channel and the base station seeks to schedule a user with the favorable channel condition in any slot. We assume that the slots are further divided into mini-slots and the users contend in such mini-slots for channel access.

Channel Model

We consider a fading wireless channel between the base station and the wireless users. Let $(H_1(t), \cdots, H_N(t))$ be the channel vector at time t, where $H_i(t)$ is the channel gain between the base station and user i at time t. In this chapter, we assume that the channel gains are i.i.d. over users and time slots with a common continuous distribution $F(\cdot)$. In Chapters 3 and 4, we will extend our study to correlated and non identical distributions for the users. We assume that the channel state remains constant in a slot and varies i.i.d. over time slots for all the users.

We assume that the users have complete information about the channel at the beginning of every slot. For example, the base station can broadcast a pilot signal at the beginning of every slot. Users estimate their channel gain H_i in that slot using the pilot signal. The base station seeks to schedule the user with highest channel gain, i.e.,

$$\arg\max\{H_1, H_2, \ldots, H_N\} \tag{2.1}$$

where the channel information is available only with the users. In this setup, we consider a splitting based contention resolution algorithm during the minislots to identify the user with the best channel. In this thesis, we assume that the base station and the users knows the common channel distribution as well as the number of contending users (N) in the network.

Define $X_i := F(H_i)$, the cumulative distribution value of i^{th} user in a slot. Then, the vector $(X_1, X_2, \ldots, X_N)$ is i.i.d. Uniform in $[0,\ 1]$ for any channel (continuous) distribution $F(\cdot)$. Therefore, finding the user with highest channel rate is equivalent to finding,

$$\arg\max\{X_1, X_2, \ldots, X_N\} \tag{2.2}$$

Hence, without loss of generality, we will assume that the channel gain is distributed as i.i.d. Uniform in $[0,\ 1]$, and we will consider $(X_1, X_2, \ldots, X_N)$ as the channel vector in a slot.

Contention Model

Each time slot is divided into smaller time units called minislots (e.g., there can be K minislots in a slot). Users contend for the channel in the minislots aided by feedback from the base station. We assume that the duration of a minislot is of the order of the round trip time. Users can transmit MAC packets to the base station and receive feedback from the base station within the minislot. We further assume that the feedback is received by all nodes at the end of a minislot without any error. The contention process can take a random number of minislots (k) to find the user with the best channel. The selected user is allowed to transmit data to the base station in the remaining time

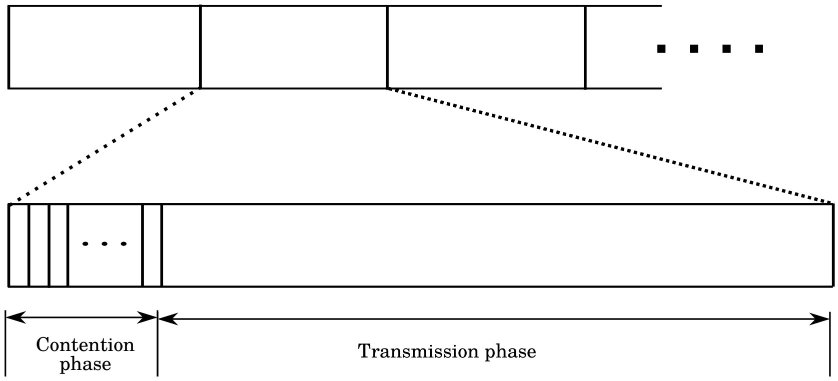

Figure 2.1: Structure of a slot and the minislots. We assume that the users contend in the minislots and the successful user is allocated the remainder of the slot for data transmission.

of the slot. In Figure 2.1, we have illustrated the structure of the slot and the minislots.

2.2 Opportunistic Splitting

We will now discuss the opportunistic splitting algorithm (OSA), proposed in Qin and Berry (2004), to resolve contention. We note again that the splitting algorithm for i.i.d. block fading channel assumes that every user has perfect knowledge of the total number of users and its channel state at the beginning of every slot.

In every minislot, OSA identifies a continuous range, $(y_{min}, y_{max}] \subset [0, 1]$ to aid in contention. User(s) with channel gain values in the range $(y_{min}, y_{max}]$ transmit MAC packets to the base station in that minislot. After receiving any MAC packets from the user(s), the base station responds with a feedback on the contention status in the minislot to all the users. The feedback will be either 0, 1 or e, indicating whether the minislot was idle, success or collision respectively. A feedback of 1 means that a single MAC packet was received. The contention is considered resolved in this case and the lone user is allocated the remaining time in the slot to transmit data. If the feedback is either 0 or e, the continuous range is suitably adjusted depending on the information gained from the feedback in the previous slots and the contention process continues.

Algorithm 2.1 describes the pseudo-code of OSA for i.i.d. channel statistics with fixed and known number of users N. In the pseudo-code, f denotes the feedback and k denotes the number of minislots used for contention resolution.

Algorithm 2.1 OSA

Initialize: $y_{low} = 0, y_{min} = 1 - \frac{1}{N}, y_{max} = 1$
Initialize: $f = 0, k = 1$
  while $f \neq 1$ and $k \leq K$ **do**
    $f = (0, 1, e)$ feedback from $(y_{min},\ y_{max}]$
    if $f = e$ **then**
      $y_{low} = y_{min}$
      $y_{min} = (\frac{y_{min}+y_{max}}{2})$
    end if
    if $f = 0$ **then**
      $y_{max} = y_{min}$
      if $y_{low} \neq 0$ **then**
        $y_{min} = \frac{(y_{low}+y_{max})}{2}$
      else
        $y_{min} = y_{max}\left(1 - \frac{1}{N}\right)$
      end if
    end if
    $k = k + 1$
  end while

Remark 2.1.

1. We assume that the base station can identify a single successful transmission, an idle minislot or a collision and feedback the information to the users. This is the ternary feedback model studied in detail in this thesis. In Section 2.4.1, we also discuss a complete feedback model where the base station can identify the exact number of contending users in a minislot even for a collision channel.

2. OSA aims to chose a continuous range such that the probability of only one user transmitting a MAC packet in a minislot is maximized. In the event of a collision in the previous minislot, the algorithm acknowledges the fact that the chance of two users involved in a collision is high and suggests the optimal strategy for the two users.

3. The average delay of the OSA was characterized in Qin and Berry (2004) and it was shown that the average delay is less than 2.5017 mini slots, independent of the number of users and the channel distribution.

2.2.1 Maximal Probability Allocation

OSA approximated the choice of the contention range in the event of a collision. In Ramaiyan (2013), the authors chose to identify the optimal range for contention that maximizes the probability of success in the event of collision as well as an idle minislot. The following theorem from Ramaiyan (2013), identifies the optimal range for the following minislot.

Theorem 2.1. *Given N users and thresholds $(y_{min}, y_{max}]$, the contention range that maximizes the probability of success is $(y, y_{max}]$, where y is the unique stationary point of $(y_{max} - y)(y^{N-1} - y_{min}^{N-1})$.*

Using numerical results, it was shown in Ramaiyan (2013), that MPA improves the performance with the optimal choice of the contention range and that OSA is nearly optimal in its choice of contention range for the i.i.d. channel.

2.3 A Source Coding Framework

We will now discuss the source coding framework for the opportunistic contention resolution problem, originally proposed in Ramaiyan (2013). Let $(X_1, X_2, \dots, X_N)$ correspond to the vector of i.i.d. channel gain values of users in a slot, and let $(Y_1, Y_2, \dots, Y_N)$ be the ordered N tuple such that $Y_1 \leq \cdots \leq Y_{N-1} \leq Y_N$. The contention resolution algorithm seeks to identify the user with the highest channel gain by identifying a continuous range $(y, y_{max}]$ in every minislot such that a single user is in the continuous range. We note here that y_{max} is essentially 1 for opportunistic scheduling and hence, the algorithm seeks to identify a threshold $y \in [0, 1]$ such that a single user has a channel gain greater than y. In other words, the scheduler resolves the contention by identifying a threshold y between Y_{N-1} and Y_N, i.e., $Y_{N-1} < y \leq Y_N$. As Y_{N-1} and Y_N are random variables, we note that the threshold Y (such that $Y_{N-1} \leq Y \leq Y_N$) that resolves contention is random as well.

Define $\Omega_Y = \{y_1, y_2, \dots\}$ as the set of thresholds that resolve contention (i.e., sample space of Y) and let $\mathbf{P}_Y = \{p_1, p_2, \dots\}$ be a probability assignment, where $p_i := Pr(Y = y_i)$ is the probability that the threshold y_i resolves contention in a slot. We expect that the probabilities sum up to one, i.e., $\sum_i p_i = 1$, or else the average delay to resolve contention will be infinity. We note here that it is sufficient to describe a probability mass function for Y as the set of thresholds possible in this network model is discrete.

The thresholds are fed back by the base station using ternary alphabet of $(0, 1, e)$. Every threshold is uniquely identified by a finite sequence of $(0, 1, e)$. The random sequence fed back to the users corresponds to the random threshold Y that resolves the

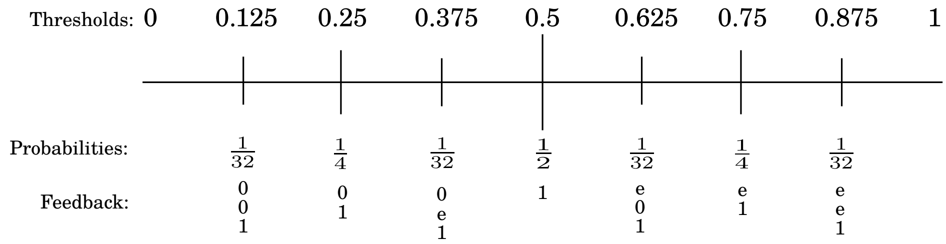

Figure 2.2: The set of MPA thresholds $\{y_i\}$, their probabilities $\{p_{y_i}\}$ and the corresponding feedback from the base station for a wireless network with $N = 2$ users and i.i.d. channel.

contention in the slot. In Figure 2.2, we illustrate the source coding framework for the contention resolution problem for $N = 2$ users and for i.i.d. channel. For $N = 2$, the MPA thresholds are $\Omega_Y = \{\frac{1}{2}, \frac{3}{4}, \frac{1}{4}, \frac{7}{8}, \frac{5}{8}, \frac{3}{8}, \frac{1}{8}, \ldots\}$ and the corresponding probabilities are $\mathbf{P}_Y = \{\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{32}, \frac{1}{32}, \frac{1}{32}, \frac{1}{32}, \ldots\}$ respectively. Note that the probabilities sum up to one. In the Figure 2.2, we have also shown the feedback corresponding to the thresholds as $\{1, e1, 01, ee1, e01, 0e1, 001\}$. Notice that the feedback from the base station can be interpreted as the binary representation of the threshold (interpreting $e = 1$) separating the highest channel gain from the second highest value. For $N > 2$, the feedback from the base station will be a weighted binary (actually ternary) sequence uniquely identifying the random threshold.

Clearly, the uncertainty in Y should be a measure of the description length of the random variable Y and hence, the average number of minislots required to resolve contention (the number of symbols required to describe a threshold Y is the same as the number of minislots required to resolve contention with the threshold Y). Hence, the entropy of the random variable Y (in appropriate alphabets) should approximate the average number of minislots required to resolve contention. The entropy of the random variable Y is defined as

$$H(\{p_i\}) = -\sum_i p_i \log_2(p_i) \tag{2.3}$$

We would expect that minimizing the entropy of a strategy should seek to minimize the average delay to resolve contention as well. In this thesis, we will bound the difference between the expected delay to resolve contention for a strategy and its entropy. In Chapter 3, we will formulate the entropy minimization problem and characterize MPA strategy as a solution to the optimization problem.

2.4 Optimality of MPA

We will now study the optimality of the greedy contention resolution strategy MPA. We will study both the delay optimality as well as the entropy optimality of the MPA strategy for the i.i.d. channel.

Lemma 2.1. *Consider $N = 2$. MPA is average delay optimal and entropy optimal for the ternary collision feedback model.*

Proof. Let D^* be the optimal average delay for the contention resolution problem. Let $(y^*, 1]$ be the contention range in the first minislot for the delay minimizing strategy. Conditioned on the first minislot, the optimal average delay for the two user network, D^*, can be written as

$$D^* = 2y^*(1 - y^*) + (1 + D^*)((y^*)^2 + (1 - y^*)^2) \tag{2.4}$$

Rewriting the above equation, we have,

$$D^* = \frac{1}{2y^*(1 - y^*)} \tag{2.5}$$

The above expression is minimized at $y^* = \frac{1}{2}$ (see Figure 2.3) and the optimal value D^* is 2 minislots. MPA seeks the threshold $y \in [0, 1]$ maximizing the probability of success $2y(1 - y)$ in the first minislot. Hence, the threshold for MPA is also $\frac{1}{2}$ and we know that the average delay for MPA scheme is also 2 minislots. Therefore, MPA is an average delay optimal strategy for $N = 2$ users and for i.i.d. channel.

We will now prove the entropy optimality of the MPA strategy for $N = 2$ users and for i.i.d. channel. Let E^* be the minimum entropy with a contention resolution strategy and let $\{y_i^*\}$ be the thresholds with the corresponding probabilities $\{p_i^*\}$ (where we assume that $\sum_i p_i^* = 1$). Without loss of generality, let y_1^* be the first threshold and $0 < p_1^* < 1$ be the corresponding probability, where $p_1^* = 2y_1^*(1 - y_1^*)$. We can write E^* as

$$E^* = H(\{p_i^*\}) = -p_1^* \log_2(p_1^*) - \sum_{\{i:y_i^* < y_1^*\}} p_i^* \log_2(p_i^*) - \sum_{\{i:y_i^* > y_1^*\}} p_i^* \log_2(p_i^*) \tag{2.6}$$

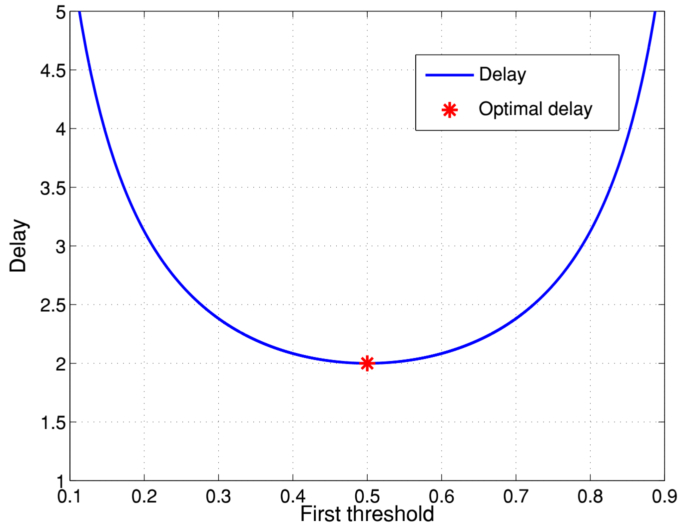

Figure 2.3: Average delay D (see equation 2.5) as a function of the first threshold y for $N = 2$ users and i.i.d channel.

Consider the set of thresholds on the left of y_1^*, $\{y_i^* : y_i^* < y_1^*\}$; we will re-label them as $\{l_i^*\}$ with the corresponding probabilities $\{q_i^*\}$. We know that the set of thresholds $\{l_i^*\}$ is countable. Further, we know that $\sum_i q_i^* = (y_1^*)^2$ and their contribution to the entropy E^* is $-\sum_i q_i^* \log_2(q_i^*)$.

Multiplying $\frac{1}{y_1^*}$ to the thresholds $\{l_i^*\}$ would give us a sequence of thresholds $\{\frac{l_i^*}{y_1^*}\}$ corresponding to a contention resolution strategy for the interval $[0, 1]$. This is possible due to the fact that both idle and collision involves exactly 2 users. The probability associated with the new set of thresholds $\{\frac{l_i^*}{y_1^*}\}$ would now be $\{\frac{q_i^*}{(y_1^*)^2}\}$ (which sum up to one). We can now compute the entropy of the strategy as

$$\begin{aligned} H(\{\frac{q_i^*}{(y_1^*)^2}\}) &= -\sum_i \frac{q_i^*}{(y_1^*)^2} \log_2\left(\frac{q_i^*}{(y_1^*)^2}\right) \\ &= -\frac{1}{(y_1^*)^2}(\sum_i q_i^* \log_2(q_i^*) + \sum_i q_i^* \log_2\left(\frac{1}{(y_1^*)^2}\right)) \\ &= -\frac{1}{(y_1^*)^2}(\sum_i q_i^* \log_2(q_i^*) + \log_2\left(\frac{1}{(y_1^*)^2}\right)(y_1^*)^2) \end{aligned}$$

Suppose that the new set of thresholds are not the same as the original set of thresholds $\{y_i^*\}$. Then, from the optimality of the original thresholds $\{y_i^*\}$ for the contention resolution problem, we have,

$$-\sum_i p_i^* \log_2(p_i^*) \leq -\frac{1}{(y_1^*)^2}(\sum_i q_i^* \log_2(q_i^*) + \log_2\left(\frac{1}{(y_1^*)^2}\right)(y_1^*)^2)$$

Rewriting the equations and with a little rearrangement, we get,

$$\begin{aligned}
-\sum_i (y_1^*)^2 p_i^* \log_2(p_i^*) + \log_2\left(\frac{1}{(y_1^*)^2}\right)(y_1^*)^2 &\leq -\sum_i q_i^* \log_2(q_i^*) \\
-\sum_i (y_1^*)^2 p_i^* \log_2(p_i^*) + (y_1^*)^2 \log_2\left(\frac{1}{(y_1^*)^2}\right)\sum_i p_i^* &\leq -\sum_i q_i^* \log_2(q_i^*) \\
-\sum_i (y_1^*)^2 p_i^* \log_2(p_i^*(y_1^*)^2) &\leq -\sum_i q_i^* \log_2(q_i^*)
\end{aligned}$$

Note that the left hand side of the above expression can be viewed as a valid strategy for the left side of the threshold y_1^* different from the threshold $\{l_i^*\}$ that contributes to the optimal entropy E^*. The threshold points would be $\{y_i^* y_1^*\}$ and their corresponding probabilities would be $\{p_i^*(y_1^*)^2\}$ (which sums up to $(y_1^*)^2$). The last expression implies that there exists an optimal strategy for the left hand side (of y_1^*) that can minimize the entropy further than E^* which is a contradiction. Hence, we need that the strategy on the left side of the threshold y_1^* is symmetrical to the strategy considered for the original problem. Hence, $\{l_i^*\}$ is the same as $\{y_i^* y_1^*\}$. A similar argument holds for the right hand side as well.

Now, the expression for the optimal E^* can be written as follows.

$$\begin{aligned}
E^* &= -p_1^* \log_2(p_1^*) - \sum_{\{i:y_i^*<y_1^*\}} p_i^* \log_2(p_i^*) - \sum_{\{i:y_i^*>y_1^*\}} p_i^* \log_2(p_i^*) \\
&= -p_1^* \log_2(p_1^*) - \sum_i p_i^*(y_1^*)^2 \log_2(p_i^*(y_1^*)^2) - \sum_i p_i^*(1-y_1^*)^2 \log_2(p_i^*(1-y_1^*)^2)
\end{aligned}$$

Using the definition $H(\{p_i\}) := -\sum_i p_i \log_2(p_i)$ for any probability distribution $\{p_i\}$, the optimal expression for the entropy E^* can be written as

$$E^* = H(\{p_i^*\}) = -p_1^* \log_2(p_1^*) + H(\{p_i^*(y_1^*)^2\}) + H(\{p_i^*(1-y_1^*)^2\})$$

Expanding the above expression in terms of the threshold y_1^*, we have,

$$H(\{p_i^*\}) = -2y_1^*(1-y_1^*)\log_2(2y_1^*(1-y_1^*)) + H(\{p_i^*(y_1^*)^2\}) + H(\{p_i^*(1-y_1^*)^2\}) \quad (2.7)$$

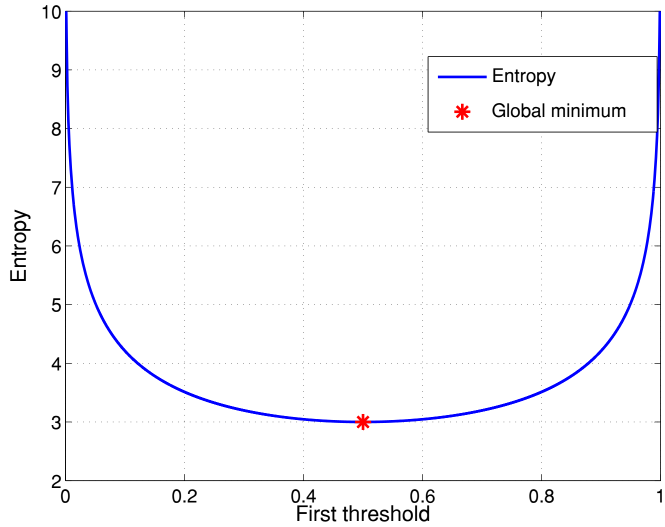

Figure 2.4: Entropy E (see equation 2.8) as a function of the first threshold y_1 for $N = 2$ users and i.i.d channel.

From the definition of $H(\cdot)$, we have,

$$\begin{aligned} H(\{ap_i^*\}) &= -\sum_i ap_i^* \log_2(ap_i^*) \\ &= -a(\sum_i p_i^* \log_2(p_i^*) + \log_2(a)\sum_i p_i^*) \\ &= aH(\{p_i^*\}) - a\log_2(a) \end{aligned}$$

Substituting in the above equation (2.7) and with a little rearrangement, we have,

$$E^* = H(\{p_i^*\}) = \frac{-2y_1^*(1-y_1^*)\log_2(2y_1^*(1-y_1^*)) - (y_1^*)^2\log_2((y_1^*)^2) - (1-y_1^*)^2\log_2(1-y_1^*)^2}{2y_1^*(1-y_1^*)} \tag{2.8}$$

Figure 2.4 shows the variation of the optimal entropy with first threshold y_i^*. Equation (2.8) is minimized at $y_1^* = \frac{1}{2}$, and value of E^* is 3. MPA also partitions the continuous range at $\frac{1}{2}$ and has an entropy of 3, i.e., MPA is an entropy optimal strategy as well. $\square$

For general N, we conjecture using simulations that MPA may not be an optimal strategy.

Proposition 2.1. *Suppose $N > 2$. MPA is neither delay optimal nor entropy optimal for the ternary feedback collision model.*

Remark 2.2.

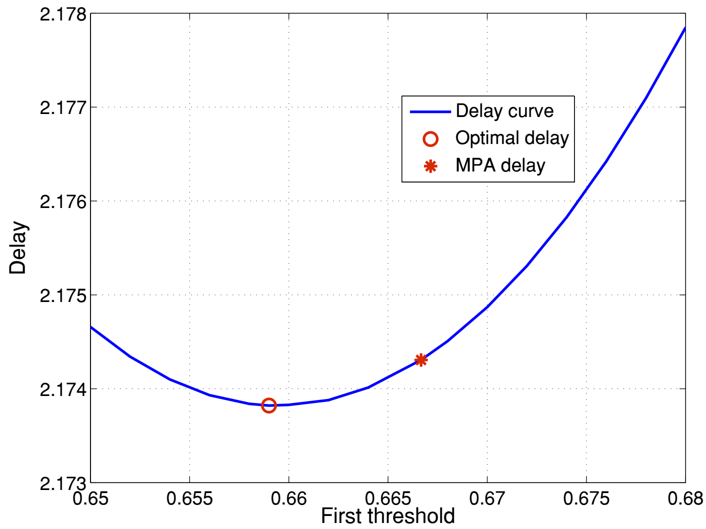

Figure 2.5: Plot of the average delay for a contention resolution strategy as a function of the first threshold. The contention resolution strategy uses MPA for the remaining minislots. We assume that $N = 3$ users and the channel is i.i.d.

1. Through numerical evaluation reported in Figure 2.5 and 2.6 for $N = 3$ users and i.i.d. channel, we tend to believe that the above proposition is true. In the figures, we plot the performance of a strategy with a choice for the first threshold and MPA for the remainder of the minislots. We note that MPA is neither delay optimal nor entropy optimal. Further, we also observe that the entropy optimal strategy is not delay optimal as well.

2. We have observed from a number of simulations that the performance of MPA is approximately optimal for the i.i.d. channel scenario with ternary feedback. In Qin and Berry (2004), the authors show that the performance of OSA/MPA is nearly optimal for i.i.d. channel by obtaining tight lower and upper bounds for the contention resolution problem.

3. In Chapter 3, we show that the performance of MPA for the non-i.i.d. channel case with ternary feedback can be arbitrarily bad with respect to an optimal algorithm.

2.4.1 Complete Feedback Case

We will now study the optimality of MPA for the complete feedback model. Suppose that the base station can feedback the exact number of users k involved in a collision. Here, $k = 0$ would correspond to an idle minislot and $k = 1$ would correspond to a success and $k > 1$ (where $k > 1$) would mean a collision. When there is a collision involving k users, the threshold would be chosen by the MPA with the knowledge of

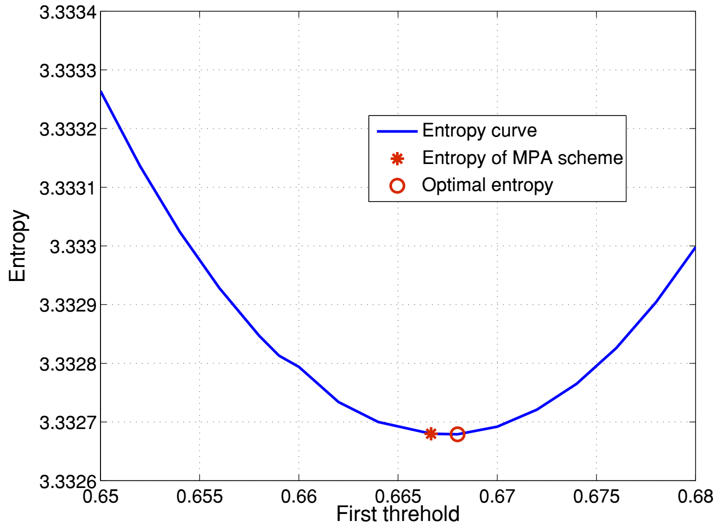

Figure 2.6: Plot of the entropy for a contention resolution strategy as a function of the first threshold. The contention resolution strategy uses MPA for the remaining slots. We assume that $N = 3$ users and the user channel is i.i.d.

k. The MPA contention range for an interval $(y_{min}, y_{max}]$ would be $(y_{min} + (y_{max} - y_{min})(1 - \frac{1}{k}), y_{max}]$. For $N = 2$ users, the ternary feedback collision model is the same as the complete feedback collision model as every collision involves exactly two users. The following lemma characterizes the performance of MPA for the complete feedback case.

Lemma 2.2. *Suppose $N > 2$. MPA is neither delay optimal nor entropy optimal for the complete feedback collision model.*

Proof. Let N be the number of users in the network. The average delay to resolve contention, D_N^*, can be expressed in terms of the first threshold $y, 0 \leq y \leq 1$, as follows.

$$\begin{aligned} D_N^* &= \min_{y \in [0,1]} \left\{ p_I(y)(1 + D_N^*) + p_S(y) + p_C(y)(1 + \sum_{k=2}^{N} p_c^k(y) D_k^*) \right\} \\ &= \min_{y \in [0,1]} \{ y^N(1 + D_N^*) + Ny(1-y)^{N-1} + (1 - y^N - Ny(1-y)^{N-1}) \times \\ &\qquad (1 + \sum_{k=2}^{N} \frac{NCk(1-y)^k y^{N-k}}{1 - y^N - Ny(1-y)^{N-1}} D_k^*) \right\} \end{aligned} \tag{2.9}$$

where $p_I(y) = y^N$ is the probability that the channel was idle with the threshold $y \in$

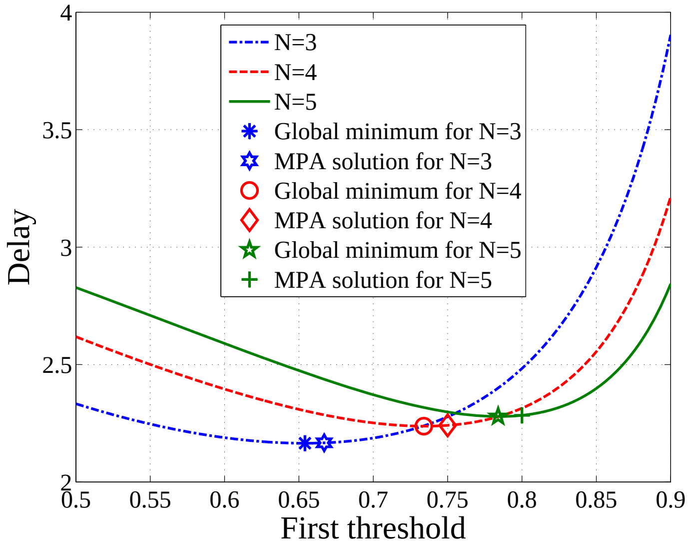

Figure 2.7: Plot of D_N as a function of the first threshold y for three different set of users $N = 3$, $N = 4$, $N = 5$ and for i.i.d. channel.

$[0, 1]$, $p_S(y) = Ny(1 - y)^{(N-1)}$ is the probability that the minislot had a success and $p_C(y)$ is the probability that the minislot involved a collision. Further, $p_c^k(y)$ is the probability that the minislot involved a collision with k users given that the minislot was a collision and D_k^* is the optimal average delay to resolve collision with k users.

In the Figure 2.7, we plot D_N as a function of the first threshold for three different values of N. We note that the strategy maximizing the probability of success in the first minislot (marked as MPA solution in the figure) does not minimize the overall average delay of contention resolution. We made a similar observation for the entropy of the strategies as well. □

2.5 Conclusion

In this chapter, we have studied the performance of the greedy MPA strategy and discussed its optimality for the i.i.d. channel case. We note that MPA is entropy optimal as well as delay optimal for $N = 2$ users. However, the strategy is neither delay or entropy optimal for general N especially for the complete feedback case. Further, we note that the entropy optimal strategy need not be the same as the delay optimal strategy.

CHAPTER 3

Maximal Probability Allocation for General Channel

In Chapter 2, we studied the performance of the maximal probability allocation strategy for i.i.d. channel. In practice, the users channel state distributions can be non-identically distributed and even correlated. The asymmetry in channel distribution can arise due to the difference in users location, shadowing, mobility pattern etc, and the presence of strong inter-cell interference may induce correlation among the users. Further, the utility of a user may be different from other users. This will also create asymmetry in the metric maximized by the base station in a slot. In this chapter, we will study the performance of MPA for non-i.i.d. channel and comment on its optimality. In Section 3.1, we will study the performance of MPA and other contention resolution strategies for independent and non-identically distributed channel and for correlated channel. In Section 3.2, we will discuss a entropy minimization problem for the source coding framework.

3.1 Contention Resolution for Non-i.i.d. Channel

We will first study the performance of MPA for a two user network. We consider two users with independent channel statistics communicating with the base station. We will assume arbitrary continuous channel distributions for the two users. We will also assume that the base station and the users have perfect knowledge of the distribution of all the users. In this setup, we can use MPA to find the user with highest channel rate. Let (X_1, X_2) be the vector of channel gain of the two users. MPA attempts to identify a threshold Y such that $\min(X_1, X_2) < Y \leq \max(X_1.X_2)$. The following lemma characterizes the delay of MPA for the network scenario.

Lemma 3.1. *Let $N = 2$. The expected delay to resolve contention using MPA for independent and non-identically distributed channel is upper bounded by 2 minislots. Also, the minimum entropy of a strategy is upper bounded by 3.*

Proof. Let $F_1(\cdot)$ and $F_2(\cdot)$ be the CDF of the wireless channel of the two users. Let y be a threshold in $(0, 1]$. The probability of success in a minislot for a given threshold y is then given by

$$p_s(y) = F_1(y)(1 - F_2(y)) + (1 - F_1(y))F_2(y)$$

due to independence of the channel of the two users. Consider a $y \in (0, 1]$ such that $F_1(y) = \frac{1}{2}$. Then, the probability of success is given by

$$p_s(y) = \frac{1}{2}(1 - F_2(y)) + \frac{1}{2}F_2(y) = \frac{1}{2}$$

Hence, $\max_{\{y \in (0,1]\}} p_s(y) \geq \frac{1}{2}$. The argument can be extended to arbitrary intervals $(y_{min}, y_{max}]$ with the corresponding conditional CDFs. Thus, the conditional probability of success in any minislot is at least $\frac{1}{2}$. Hence, the average delay to resolve contention for the independent and non-identically distributed channel is at most 2 minislots.

We will now show that the entropy of the above strategy is upper bounded by 3. Let $\{y_{(d,f)}\}$ be the set of thresholds for the contention resolution strategy and let $\{p_{(d,f)}\}$ be the corresponding probabilities. Threshold $y_{(d,f)}$ corresponds to a threshold used in minislot d with the past feedback of f. Hence, if the threshold $y_{(d,f)}$ resolves the contention, then, the random delay for the contention is d minislots. The entropy of the contention resolution strategy is a function of the probability distribution $\{p_{(d,f)}\}$ and is given as

$$H(\{p_{(d,f)}\}) = p_{(1,-)} \log_2\left(\frac{1}{p_{(1,-)}}\right) + p_{(2,0)} \log_2\left(\frac{1}{p_{(2,0)}}\right) + p_{(2,e)} \log_2\left(\frac{1}{p_{(2,e)}}\right) + \\ p_{(3,00)} \log_2\left(\frac{1}{p_{(3,00)}}\right) + \cdots p_{(3,0e)} \log_2\left(\frac{1}{p_{(3,0e)}}\right) + \tag{3.1}$$

Define

$$\begin{aligned} P_1 &= p_{(1,-)} \\ P_2 &= p_{(2,0)} + p_{(2,e)} \\ P_3 &= p_{(3,00)} + p_{(3,0e)} + p_{(3,e0)} + p_{(3,ee)} \\ &\cdots \end{aligned}$$

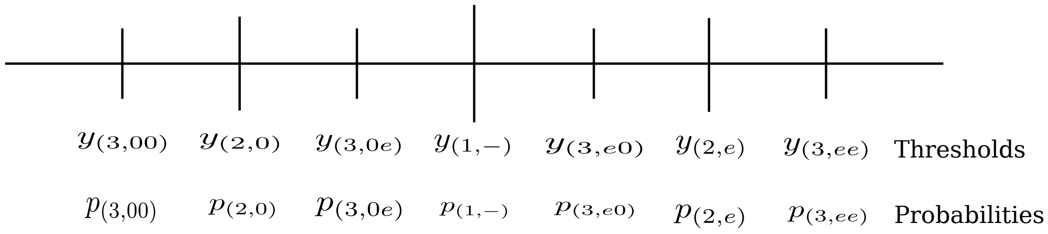

Figure 3.1: Thresholds $\{y_{(d,f)}\}$ of a contention resolution strategy where threshold $y_{(d,f)}$ is used in minislot d for a past feedback of f.

i.e. P_i is the probability that contention is resolved in the i^{th} minislot with the strategy. The following inequality holds because $-P\log_2(P)$ is a concave function in P.

$$H(\{p_{(d,f)}\}) \leq \; P_1 \log_2\left(\frac{1}{P_1}\right) + 2\left(\frac{P_2}{2}\right)\log_2\left(\frac{2}{P_2}\right) + 4\left(\frac{P_3}{4}\right)\log_2\left(\frac{4}{P_3}\right) + \cdots$$

Further, for the above strategy, we have, $P_1 \geq \frac{1}{2}, P_2 \geq (1-P_1)\frac{1}{2}, P_3 \geq (1-P_1)(1-P_2)\frac{1}{2}, \cdots$ (since the conditional probability of success is at least $\frac{1}{2}$ for any feedback in a minislot). Hence, the entropy of the strategy is upper bounded by

$$H(\{p_{(d,f)}\}) \leq \frac{1}{2}\log_2\left(\frac{1}{\frac{1}{2}}\right) + 2\left(\frac{\frac{1}{4}}{2}\right)\log_2\left(\frac{2}{\frac{1}{4}}\right) + 4\left(\frac{\frac{1}{8}}{4}\right)\log_2\left(\frac{4}{\frac{1}{8}}\right) + \cdots = 3$$

□

In Table 3.1, we show the performance of MPA for few cases of independent channel distributions. We note that the average delay is less than 2 for all non-identically distributed channel scenarios.

User1 distribution	User2 distribution	Delay	Entropy
Rayleigh with $\sigma = 1$	Rayleigh with $\sigma = 2$	1.745	2.442
Rayleigh with $\sigma = 1$	Rayleigh with $\sigma = 4$	1.364	1.417
Rayleigh with $\sigma = 1$	Uniform in [0, 1]	1.629	2.099
Uniform in [0, 1]	Uniform in [0, 0.5]	1.780	2.448

Table 3.1: Average delay and entropy of MPA for $N = 2$ users with independent and non-identically distributed channel.

3.1.1 Additive and Multiplicative Scaling

As a special case, for $N = 2$, we numerically evaluate the average delay and entropy of MPA when users take the same distribution with different scaling - multiplicative scaling and additive scaling. The use of multiplicative scaling is motivated by the implementation of fairness metrics such as proportional fairness Jalali *et al.* (2000), and additive scaling factor is motivated from implementations like in Liu *et al.* (2003).

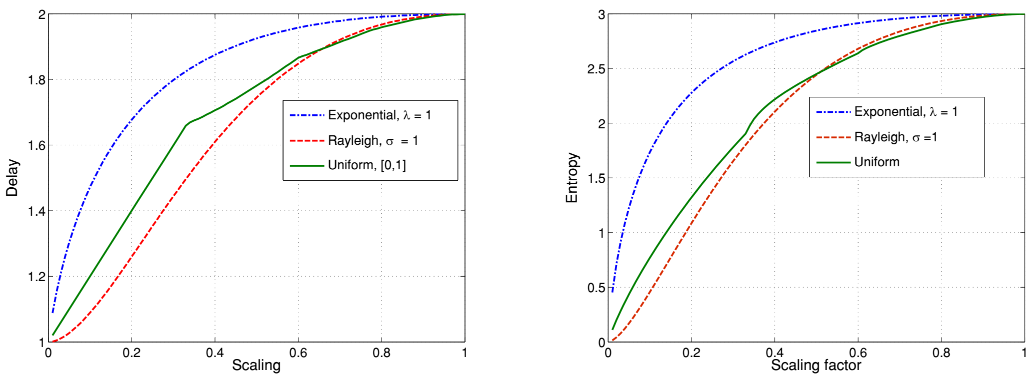

Figure 3.2: Variation of average delay and entropy for MPA with multiplicative scaling for three different distributions and $N = 2$ users.

In Figure 3.2, we plot the average delay and entropy respectively for three different distributions (exponential, Rayleigh and Uniform) as a function of the multiplicative scaling parameter. For example, if the channel of users 1 is Exponential with mean λ, a scaling of 0.5 implies that the channel of user 2 is Exponential with mean 0.5λ independent of the user 1. Note that the average delay (and entropy) is a monotone function of the scaling parameter with the average delay being the highest for identically distributed channel. Also, notice that the entropy and delay of the MPA strategy has similar behaviour as a function of the scaling parameter.

In Figure 3.3, we plot the average delay and entropy respectively for three different distributions (exponential, Rayleigh and Uniform) as a function of the additive scaling parameter. For example, if the channel of users 1 is Exponential with mean λ, an additive scaling of 0.5 implies that the channel of user 2 is Exponential with mean $0.5 + \lambda$ independent of the user 1. We note that the performance with additive scaling is also monotone (the average delay is maximum for identically distributed channel). Further, the average delay and entropy of the MPA strategy has similar behaviour with the scaling parameter.

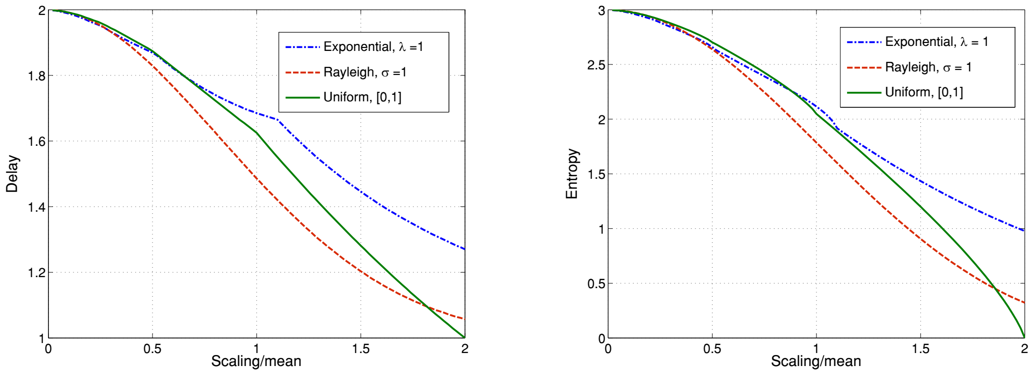

Figure 3.3: Variation of average delay and entropy with additive scaling for three different distributions and $N = 2$ users.

The average delay and entropy are not necessarily monotone functions of scaling parameters in general. The following example illustrates the case. Let the probability density function of user 1 be as described below.

$$f_{X_1}(x) = \begin{cases} \frac{1}{4} & \text{if } x \in [2,4) \cup [8,10), \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

Let the distribution of user 2 be identical to that of aX_1 (or $a + X_1$) but independent of the actual distribution of X_1. Consider a multiplicative scaling of $a = 0.5$. Contention can be resolved in one minislot by setting the threshold 4. In this case the user with the highest channel can be determined even if the minislot is a collision or success. If the first minislot is a collision, user 1 has the best value. If the first minislot is idle, then again, user 1 has the best channel (in $[2, 4)$ where as the rate of user 2 is in $[1, 2)$). In the case of a single user transmission, the lone user is declared the winner of the contention. However, for a multiplicative scaling of $a = \frac{1}{4}$ or $a = 1$, contention resolution requires more than a single minislot as the support of the two user distributions has overlap. A similar observation can be made with additive scaling of 0, 2 and 6 for the user distributions. The delay and entropy in the case of additive scaling is illustrated in figure 3.4.

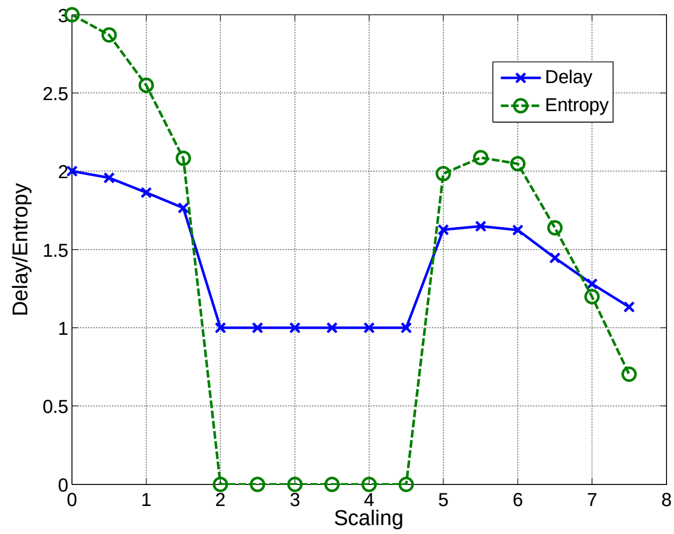

Figure 3.4: Variation of average delay and entropy with additive scaling for the probability density function described in equation 3.2 for $N = 2$ users.

3.1.2 Performance of MPA for $N > 2$

In Lemma 3.1, we showed that, for $N = 2$ users, the average delay/entropy of MPA to resolve contention for independent and non-identical channel distribution is less than the average delay/entropy for i.i.d. channel. For $N > 2$, we conjecture that the delay/entropy of MPA is maximum when the users channel states are distributed independently with identical distribution. Intuitively, we would expect that, for independent distributions, uncertainty (entropy) should increase with symmetry in the distribution. We make the observation through numerical verification for a variety of independent and non-identically distributed channel distributions for different values of N. In Tables 3.2 and 3.3, we have given the average delay of MPA for $N = 3$ users and $N = 4$ users for few cases. Notice that the average delay is maximum when the channel distributions are identically distributed.

User1	User2	User3	Delay	Entropy
Rayleigh, $\sigma = 1$	Rayleigh, $\sigma = 1$	Rayleigh, $\sigma = 1$	2.174	3.332
Rayleigh, $\sigma = 1$	Rayleigh, $\sigma = 1$	Exponential, $\lambda = 1$	2.144	3.277
Rayleigh, $\sigma = 1$	Exponential, $\lambda = 1$	Exponential, $\lambda = 1$	2.119	3.230
Rayleigh, $\sigma = 1$	Exponential, $\lambda = 1$	Uniform in $[2, 4]$	1.440	1.568

Table 3.2: Average delay and entropy of MPA for independent and non-identically distributed channel and $N = 3$ users.

User 1	User 2	User 3	User4	Delay	Entropy
Rayleigh, $\sigma = 1$	Rayleigh, $\sigma = 1$	Rayleigh, $\sigma = 1$	Rayleigh, $\sigma = 1$	2.250	3.469
Rayleigh, $\sigma = 1$	Rayleigh, $\sigma = 1$	Rayleigh, $\sigma = 1$	Exponential, λ =1	2.235	3.442
Rayleigh, $\sigma = 1$	Rayleigh, $\sigma = 1$	Uniform in $[2, 4]$	Exponential, λ =1	1.528	1.867
Rayleigh, $\sigma = 1$	Uniform in $[2, 4]$	Uniform in $[2, 4]$	Exponential, λ =1	2.042	3.083

Table 3.3: Average delay and entropy of MPA for independent and non-identically distributed channel and $N = 4$ users.

3.1.3 Optimality of MPA

In the earlier sections, we characterized the performance of MPA in comparison with the i.i.d. channel. In this section, we will study the optimality of MPA using an example. We will consider a two user wireless network with Uniform distribution and with a multiplicative scaling. Let (X_1, X_2) be the user channel gains where the distribution of X_1 is $U(0, 1]$ and the distribution of X_2 is $U(0, a]$ for $0 \leq a \leq 1$.

Suppose $0 \leq a \leq \frac{1}{3}$. The first MPA threshold will then be a and the probability of success in the first minislot will be $(1 - a)$; the probability of the minislot being idle is a and there cannot be a collision in the first minislot with threshold a. If the first minislot is idle, then, the system is i.i.d. after the idle slot, and it will take 2 minislots on an average to resolve contention. Therefore, the expected delay for MPA is equal to $1 \times (1 - a) + (1 + 2) \times a = 1 + 2a$ for $a \leq \frac{1}{3}$.

Consider a scaling of a such that $\frac{1}{3} \leq a \leq 1$. Let y_1 be the threshold that maximizes the probability of success in the first minislot. Then,

$$y_1 \quad = \quad \arg\max_{y \in (0,a]} \left(\frac{y}{a}\right)(1 - y) + y\left(1 - \frac{y}{a}\right)$$

Differentiating the right hand side expression, we get,

$$y_1 = \frac{1 + a}{4}$$

Now, we will find a range of scaling for a, $\frac{1}{3} \leq a \leq 1$, such that conditioned on the event of collision in the first minislot with threshold y_1, the contention resolution problem has the form with distributions (X, aX) where $0 \leq a \leq \frac{1}{3}$ in the future minislots. To find the maximum value of a for such a case, we require,

$$\begin{aligned} \frac{a - y_1}{1 - y_1} &= \frac{1}{3} \\ \text{Therefore, } a &= \frac{1 + 2y_1}{3} \end{aligned} \tag{3.3}$$

Substituting $y_1 = \frac{1+a}{4}$, in equation (3.3), we get

$$a = 0.6 \tag{3.4}$$

Therefore, for any $a \in (\frac{1}{3}, 0.6]$, the expression for delay of MPA can be written as,

$$D = y_1\left(1 - \frac{y_1}{a}\right) + (1 - y_1)\frac{y_1}{a} + (1+2)\frac{y_1^2}{a} + \left[1 + 1 + 2\left(\frac{a - y_1}{1 - y_1}\right)\right]\left(1 - \frac{y_1}{a}\right)(1 - y_1) \tag{3.5}$$

Substituting $y_1 = (1 + a)/4$ in equation (3.5), we get the delay associated with MPA scheme.

We will now provide a counter example to show the sub-optimality of MPA strategy. Consider the two user example with multiplicative scaling of $a = 0.5$ with the Uniform distribution (i.e., $X_1 \approx U(0, 1]$ and $X_2 \approx U(0, 0.5]$ and X_1 is independent of X_2). In the first mini slot, we use threshold y_1, and if the first minislot is idle or collision, we follow MPA splitting strategy in the future minislots. The delay in finding the user with highest metric is now a function of the first threshold, y_1 , and we can minimize the delay by choosing an optimal y_1. Note that MPA is a strategy at $y_1 = \frac{1+0.5}{4}$ (for $a = 0.5$). The average delay to resolve contention can now be written as

$$\begin{aligned} D(y_1) &= y_1(1 - 2y_1) + (1 - y_1)(2y_1) + 6y_1^2 \\ &\quad + \left(1 + 1 + 2\left(\frac{\frac{1}{2} - y_1}{1 - y_1}\right)\right)(1 - 2y_1)(1 - y_1) \\ &= 3y_1 + 2y_1^2 + (3 - 4y_1)(1 - 2y_1) \\ &= 3 - 7y_1 + 10y_1^2 \end{aligned}$$

The delay expression thus simplifies to $10y_1^2 - 7y_1 + 3$ and the optimal $y_1 = 0.35$ and the average delay of the strategy with the modified MPA strategy is 1.775, whereas the delay corresponding to MPA scheme is 1.781 with the MPA threshold 0.375.

The entropy expression for this example conditioned on the first threshold, y_1, can be simplified to

$$\begin{aligned} H(y_1) = &- \left(\frac{y_1(0.5-y_1)}{0.5} + \frac{y_1(1-y_1)}{0.5}\right) \log_2\left(\frac{y_1(0.5-y_1)}{0.5} + \frac{y_1(1-y_1)}{0.5}\right) \\ &- \left(\frac{y_1^2}{0.5}\right)\log_2\left(\frac{y_1^2}{0.5}\right) + 3\left(\frac{y_1^2}{0.5}\right) \\ &- \left(\frac{0.5(0.5-y_1)(1-y_1)}{0.5(1-y_1)}\right)\log_2\left(\frac{0.5(0.5-y_1)(1-y_1)}{0.5(1-y_1)}\right) \\ &- \left(\frac{(0.5-y_1)^2}{0.5}\right)\log_2\left(\frac{(0.5-y_1)^2}{0.5}\right) + 3\left(\frac{(0.5-y_1)^2}{0.5}\right) \end{aligned} \tag{3.6}$$

The expression in 3.6 is minimized at $y_1 = 0.352$. This shows that MPA is neither delay optimal nor entropy optimal in this case. Also note that entropy optimal solution is different from delay optimal solution.

3.1.4 Correlated Wireless Channel

In Lemma 3.1, we proved that the delay of MPA for independent channel distributions is upper bounded by the average delay performance of i.i.d. channel. In Ramaiyan (2013), an example was presented to show the poor performance of MPA when the channel distributions are correlated. The example considers 2 users having a discrete distribution with correlation among them. The distribution has a sample space $\Omega_H = \{(4,2),(4,6),(8,6),(8,10),(12,10),(12,14),(16,14)\}$ and joint probabilities $p_H = \{\frac{1}{7}-6\epsilon, \frac{1}{7}-5\epsilon, \frac{1}{7}-4\epsilon, \frac{1}{7}-3\epsilon, \frac{1}{7}-2\epsilon, \frac{1}{7}-\epsilon, \frac{1}{7}+21\epsilon\},\ 0 < \epsilon << 1$. If we restrict to integer thresholds, MPA thresholds form the sequence $15, 13, 11, 9, 7, 5$ and 3, with an average delay of $\frac{27}{4} \approx 4$. If there are k channel states, the average delay to resolve contention will be $\frac{k}{2}$ approximately. The average delay associated with this scheme scales linearly with the number of channel states in the channel distribution. However, in this scenario, there exist a strategy which can resolve contention of the order of time which is logarithmic in number of intervals in the channel distribution. In this example, consider the threshold value 9 in the first minislot. If it is a collision, use 13 as the next threshold, and use 7 if the first minislot is idle. If the first two minislot is a collision, use 15 as the threshold, and continue this strategy until a single user transmits. The associated delay of this scheme is approximately 3. In general, the delay of this scheme is of the order of $\log_2(k)$. Thus, MPA can be arbitrarily bad when the user channel distribution is correlated. In the next section, we discuss an entropy minimization formulation which gives alternate solutions other than MPA. We formulate the entropy minimization problem as a concave minimization problem. The MPA solution is identified as a local minimizer to the optimization problem.

3.2 Average Delay and Entropy

We will now relate the average delay of a strategy and its entropy. Consider a contention resolution strategy such that the set of threshold are $\{y_{(d,f)}\}$ and the corresponding probabilities are $\{p_{(d,f)}\}$. Threshold $y_{(d,f)}$ corresponds to a threshold used in minislot d with the past feedback of f. Hence, if the threshold $y_{(d,f)}$ resolves the contention, then, the random delay for the contention is d minislots. In Figure 3.1, we have illustrated a set of thresholds $\{y_{(d,f)}\}$ for a contention resolution strategy. The average delay of the contention resolution strategy, D, is then given by,

$$D = p_{(1,-)} + 2p_{(2,e)} + 2p_{(2,0)} + 3p_{(3,ee)} + 3p_{(3,e0)} + 3p_{(3,0e)} + 3p_{(3,00)} + \cdots$$

The entropy of the contention resolution strategy is a function of the probability distribution $\{p_{(d,f)}\}$ and is given as

$$\begin{aligned} H(\{p_{(d,f)}\}) = p_{(1,-)} \log_2 \left(\frac{1}{p_{(1,-)}}\right) + p_{(2,0)} \log_2 \left(\frac{1}{p_{(2,0)}}\right) + p_{(2,e)} \log_2 \left(\frac{1}{p_{(2,e)}}\right) + \\ p_{(3,00)} \log_2 \left(\frac{1}{p_{(3,00)}}\right) + \cdots p_{(3,0e)} \log_2 \left(\frac{1}{p_{(3,0e)}}\right) + \end{aligned} \tag{3.7}$$

Define

$$\begin{aligned} P_1 =& p_{(1,-)} \\ P_2 =& p_{(2,0)} + p_{(2,e)} \\ P_3 =& p_{(3,00)} + p_{(3,0e)} + p_{(3,e0)} + p_{(3,ee)} \\ & \ldots \end{aligned}$$

i.e. P_i is the probability that contention is resolved in the i^{th} minislot with the strategy. The following inequality holds because $-P\log_2(P)$ is a concave function in P.

$$H(\{p_{(d,f)}\}) \leq \; P_1 \log_2\left(\frac{1}{P_1}\right) + 2\left(\frac{P_2}{2}\right)\log_2\left(\frac{2}{P_2}\right) + 4\left(\frac{P_3}{4}\right)\log_2\left(\frac{4}{P_3}\right) + \cdots$$

Simplifying the right hand side expression, we get,

$$\begin{aligned} H(\{p_{(d,f)}\}) &\leq P_1 \log_2\left(\frac{1}{P_1}\right) + P_2 \log_2\left(\frac{2}{P_2}\right) + P_3 \log_2\left(\frac{4}{P_3}\right) + \cdots \\ &= P_1 \log_2\left(\frac{1}{P_1}\right) + P_2 \log_2\left(\frac{1}{P_2}\right) + P_3 \log_2\left(\frac{1}{P_3}\right) + \cdots \\ &\quad + P_2 \log_2(2) + P_3 \log_2(4) + \cdots \end{aligned} \tag{3.8}$$

By the definition of entropy, $\sum_i P_i \log_2(1/P_i)$ is the entropy of the probability distribution $\{P_i\}$. Then,

$$H(\{p_{(d,f)}\}) \leq \; H(\{P_i\}) + P_2 + 2P_3 + 3P_4 + \cdots \tag{3.9}$$

We know that the average code word length of a prefix code is lower bounded by its entropy (see Cover and Thomas (2012)). We show that $H(\{P_i\})$ is a lower bound on delay of the scheme using this result. By mapping success to 1, both collision and idle to 0, we can form a prefix code with probability distribution $\{P_i\}$. The possible code words are $\{1, 01, 001, 0001, \ldots\}$ with probability $\{P_1, P_2, P_3, \ldots\}$ Average code word length of this code is same as the average delay of MPA scheme. So,

$$H(\{P_i\}) \leq D \tag{3.10}$$

Using 3.9 and 3.10, we get

$$\begin{aligned} H(\{p_{(d,f)}\}) &\leq D + P_2 + 2P_3 + 3P_4 + \cdots \\ &= D + P_1 + 2P_2 + 3P_3 + \cdots - (P_1 + P_2 + P_3 + \cdots) \\ &= D + D - 1 \\ &= 2D - 1 \end{aligned} \tag{3.11}$$

where, D is the average number of minislots required to resolve contention. Thus, the entropy of the threshold random variable and the average delay to resolve contention are bounded and hence, they showed similar characteristics in a variety of network scenarios. Also, for $N = 2$ users with i.i.d. channel, we note that the inequality for the MPA strategy is exact with the average delay being 2 and the entropy of the strategy being 3.

3.2.1 Entropy Minimization

We will now discuss a entropy minimization framework for the wireless network. We will restrict the formulation to a discrete finite state channel (so that the set of thresholds are also finite). Suppose that there are N users and M discrete channel rates, $\Omega_R = \{r_1, r_2, \ldots, r_M\}$, i.e., the channel of every user is restricted to one of the M states. Without loss of generality, let us assume that $r_1 < r_2 < \cdots < r_M$. For an arbitrary distribution of the vector channel, let $(X_1, \cdots, X_{N-1}, X_N)$ be the channel state vector in a slot and let $(Y_1, \cdots, Y_{N-1}, Y_N)$ be the ordered channel state vector in the slot. The contention resolution algorithm seeks to find a threshold Y such that $Y_{N-1} < Y \leq Y_N$. Consider the set of all possible two-tuples for (Y_{N-1}, Y_N)

$$\Omega = \{(r_1, r_1), (r_1, r_2), (r_1, r_3), \cdots, (r_1, r_M), (r_2, r_2)(r_2, r_3), \cdots, (r_{M-1}, r_M), (r_M, r_M)\}$$

Let the probability associated with a two-tuple be $\mathsf{P}(r_i, r_j)$ such that $\sum_{i,j:j\geq i} \mathsf{P}(r_i, r_j) = 1$.

Without loss of generality, we will assume that $\mathsf{P}(Y_N = Y_{N-1}) = 0$ in the above framework and we will restrict to two-tuples such that $Y_{N-1} < Y_N$. Then, a two-tuple (r_i, r_j) for (Y_{N-1}, Y_N) can be resolved by any threshold y such that $r_i < y \leq r_j$. Hence, the set of thresholds that can resolve the two-tuple (r_i, r_j) is given by $\{r_{i+1}, r_{i+2}, \cdots, r_j\}$. Let $I_y(r_i, r_j)$ be the event that the two-tuple (r_i, r_j) is resolved by the threshold y. Then, we require that

$$\sum_{y=r_{i+1}}^{r_j} I_y(r_i, r_j) = 1$$

and $I_y(r_i, r_j) \in \{0, 1\}$. Define P_y as the probability that the threshold $y \in \Omega_R$ resolves contention.

$$P_y = \sum_{(r_i, r_j)\in\Omega} \mathsf{P}(r_i, r_j) I_y(r_i, r_j)$$

Then, the entropy of the assignment is defined as $\sum_{y\in\Omega_R} P_y \log_2\left(\frac{1}{P_y}\right)$ Our objective is to find an integer threshold arrangement for the two-tuple such that the entropy of the random threshold is minimized. The entropy minimization problem can be formulated as follows.

$$
\begin{aligned}
\min \ & -\sum_{y \in \Omega_R} P_y \log_2(P_y) \\
\text{s.t.} \ & P_y = \sum_{(r_i,r_j)\in\Omega} \mathsf{P}(r_i, r_j) I_y(r_i, r_j) \\
& I_y(r_i, r_j) \geq 0 \ \forall \ (r_i, r_j) \in \Omega, \ y \in (r_i, r_j] \\
& \sum_{y=r_{i+1}}^{r_j} I_y(r_i, r_j) = 1
\end{aligned}
\tag{3.12}
$$

We can relax the integer assumption on $I_y(\cdot,\cdot)$ and let random assignment of the two-tuple to a threshold. $-P_y \log_2(P_y)$ is a concave function of P_y and hence, the objective $-\sum_y P_y \log_2(P_y)$ is a concave function of the set of P_y. By definition, P_y is a linear combination of $I_y(r_i, r_j)$, and therefore the objective function is concave in $I_y(r_i, r_j)$ (see Luenberger and Ye (2008)). Thus, we have a concave minimization problem for the entropy minimization formulation. Thus, starting at some random initial point can give a solution which is a local minima for the formulation (3.12). Note that the integer relaxation is not an issue as there exists integer solutions for the problem, i.e., $I_y(\cdot,\cdot) \in \{0, 1\}$ because (3.12) is a concave minimization over a convex set, which gives an extreme point as solution (Luenberger and Ye (2008)).

Example 3.1. We will now discuss the formulation using an example. Consider a three user wireless network $N = 3$ with four channel states $M = 4$. Without loss of generality, we will assume that $\Omega_R = \{0, 1, 2, 3\}$ and the sample space of the two-tuple as $\Omega = \{(0, 1), (0, 2), (0, 3), (1, 2), (1, 3), (2, 3)\}$ with the corresponding probabilities $\{0.0714, 0.0714, 0.0714, 0.2143, 0.2143, 0.3571\}$. We solve the optimization problem (3.14) to get the thresholds with

$$
\begin{aligned}
P_1 &= P(0,1)I_1(0,1) + P(0,2)I_1(0,2) + P(0,3)I_1(0,3) \\
P_2 &= P(0,2)I_2(0,2) + P(0,3)I_2(0,3) + P(1,2)I_2(1,2) + P(1,3)I_2(1,3) \\
P_3 &= P(0,3)I_3(0,3) + P(1,3)I_3(1,3) + P(2,3)I_3(2,3)
\end{aligned}
\tag{3.13}
$$

The entropy minimization problem is then given as

$$
\begin{aligned}
\min \; & -P_1 \log_2(P_1) - P_2 \log_2(P_2) - P_3 \log_2(P_3) \\
\text{s.t. } & I_y(\cdot,\cdot) \geq 0 \\
& I_1(0,1) = 1 \\
& I_1(0,2) + I_2(0,2) = 1 \\
& I_1(0,3) + I_2(0,3) + I_3(0,3) = 1 \\
& I_2(1,2) = 1 \\
& I_2(1,3) + I_3(1,3) = 1 \\
& I_3(2,3) = 1
\end{aligned}
\tag{3.14}
$$

A local minimum solution (the MPA solution) to the problem is given by the strategy $I_1(0,1) = I_2(0,2) = I_3(0,3) = I_2(1,2) = I_3(1,3) = I_3(2,3) = 1$ and the other variables are zero. The entropy of the strategy is 1.1981. The average delay associated with this (MPA) strategy is $\frac{20}{14}$ minislots. The formulation permits other solutions as well. For example, another strategy that is a local minima of the optimization problem is $I_1(0,1) = I_2(0,2) = I_2(0,3) = I_2(1,2) = I_2(1,3) = I_3(2,3) = 1$ and the other variables are zero. The entropy of the strategy is 1.2638 and the average delay of the strategy is $\frac{20}{14}$ as well. Thus, we note that the entropy minimization problem helps us identify many solutions for the contention resolution problem. Further, we note that the MPA is a local minima of the concave minimization problem and hence, we may only expect to develop bounded-error or suboptimal strategies for the contention resolution problem.

3.3 Conclusion

We have studied the delay and entropy performance of MPA for general channel scenario. We showed that the delay and entropy of MPA is maximum when channel states are independent and identically distributed. For correlated wireless channel, we observe that MPA can perform arbitrarily bad. We present a relation between the average delay of a strategy and its entropy. We have also proposed an entropy minimization framework for discrete channel distributions, which can be used to obtain solutions other than

MPA for the contention resolution problem.

CHAPTER 4

Performance Evaluation of MPA

In this chapter, we will discuss generalizations of the maximal probability allocation scheme and also compare the performance of MPA with other contention resolution strategies. We first discuss generalizations of MPA such as those that permits different thresholds for different users and allows aggregating contention feedback. In the later part of this chapter, we report simulation results evaluating the performance of MPA and comparing them with other popular contention resolution strategies such as polling and timer-based contention resolution schemes. We also study the performance of MPA when there is an error in estimating the number of users, N, and distribution of users' channel SNR.

4.1 Generalizations of MPA

We will now discuss generalizations of MPA strategy that permit different thresholds for users and aggregates user feedback.

4.1.1 Unequal thresholds

In a minislot, MPA uses the same threshold for all users. We will now discuss the need to permit different thresholds for users depending on their channel distribution. With the help of an example, we will show the improvement achieved in the average delay by using different threshold for users depending on their distribution.

Consider a two user wireless network with user 1 and user 2 distributed Uniformly in $[0, 1]$ and $[0.5, 1.5]$ respectively as shown in Figure 4.1. Suppose that in the first minislot, user 1 uses 0.5 as the threshold and user 2 uses 1 as the threshold (i.e., the users transmit a contention packet if their channel realization is higher than their individual threshold). Notice that, in this scenario, the user with highest channel rate can be determined if the feedback in the first minislot is a collision (user 2 wins the contention), idle (user 2 wins

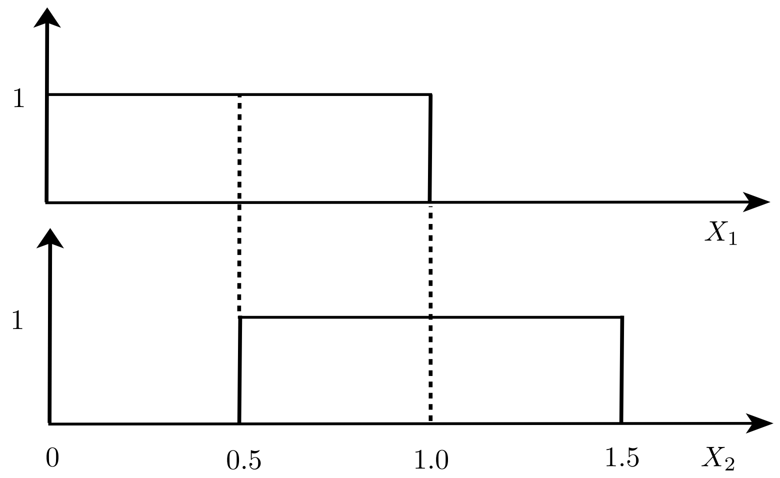

Figure 4.1: Plot of the probability density function of channel state of two users. We assume that the channel realizations of the two users are independent of each other. The dotted lines mark the first thresholds for the two users that minimizes the average delay to resolve contention.

the contention) or a successful transmission by user 2 (user 2 wins the contention). In the case when user 1 transmits a contention packet and user 2 stays idle in the first minislot, the contention can further be resolved using MPA in the interval $[0.5, 1)$, and it takes additional two minislots on an average to resolve the contention. The proposed algorithm, thus, has an average delay of $\frac{3}{4} \times 1 + \frac{1}{4} \times (1 + 2) = 1.5$ minislots. Here, we note that the average delay of MPA (from the first minislot) with identical thresholds has a larger average delay of 1.625 minislots. Even in terms of entropy, the algorithm with unequal thresholds has a significant improvement with an entropy of 1.3113 in comparison to 2.048 in the case of MPA.

4.1.2 Multiple feedback

In this section, we will show for a correlated channel that the feedback from different thresholds can be aggregated to improve the delay performance of MPA. We provide an example to illustrate this, where we consider identical distribution with correlation among users (see Figure 4.2). We assume that $N = 2$ and the channel values of the two user's occur together in the intervals $\{[0, 1), [1, 2), [2, 3), [3, 4)\}$ with probability $\{(0.25 - 2\epsilon), (0.25 - \epsilon), (0.25 + \epsilon), (0.25 + 2\epsilon)\}$ respectively, for $0 < \epsilon << 1$. Users are correlated in such a way that both users are present only in one interval. In an interval, the channel values are assumed to be independent of each other.

The first MPA threshold for this example will be at channel value 3.5. If the minislot

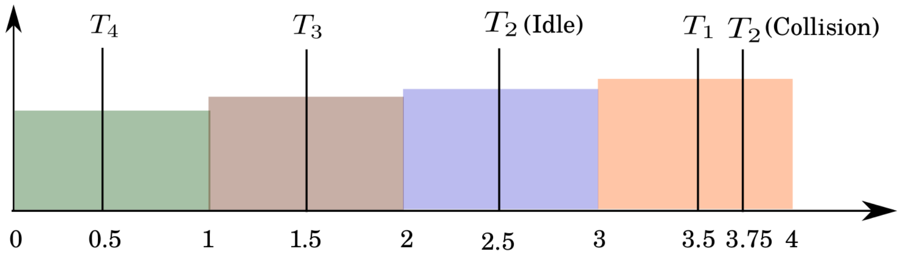

Figure 4.2: Correlated wireless channel for two users. The channel values of the two users occur together in $[0, 1)$ or $[1, 2)$ or $[2, 3)$ or $[3, 4)$ with probability $0.25 - 2\epsilon$, $0.25 - \epsilon$, $0.25 + \epsilon$ and $0.25 + 2\epsilon$ respectively. In an interval, the channel values are assumed to be independent of each other.

is a collision, it would take an average of 2 minislots to resolve contention further. If the first minislot is idle, MPA sets 2.5 as next threshold and so on until the contention is resolved. The average delay associated with MPA can be shown to scale linearly with the number of intervals in the channel distribution. However, in this scenario, there exist strategies which can resolve contention with a fixed duration (on an average). We note that the users channel are correlated to intervals and are indepedent and identically distributed in an interval. So, we propose to consider multiple contention ranges simultaneously in the first minislot, $[0.5, 1)$, $[1.5, 2)$, $[2.5, 3)$ and $[3.5, 4)$. Note that the users are in only one of the intervals and in any interval, the above thresholds correspond to the optimal threshold for MPA with two users. The proposed strategy initiates a MPA in every threshold. Since the users are correlated to an interval, only one MPA will be effective and hence, the average delay to resolve contention is 2 minislots independent of the number of intervals.

4.1.3 No feedback to indicate collision

The duration of a minislot for the contention resolution scheme with triple feedback can be computed as, $t_{\text{pkt}} + t_{\text{ack}} + 2t_{\text{pd}} + t_{\text{aRxTxTurnaroundTime}}$, where t_{pkt} is the MAC contention packet transmission duration, t_{ack} is the ACK transmission delay and t_{pd} is the maximum propagation delay in the network. In this example, we use the term ACK for the packet transmitted by the base station containing the triple feedback information. Also, we assume that ACK is transmitted immediately after receiving MAC packets from users and no additional duration is assumed. Assuming 20 byte length for the MAC and the ACK packet, for a 1 Mbps link with users located 1 Km away from the AP, we have, $t_{\text{pkt}} = t_{\text{ack}} = 160\mu s$, and $t_{\text{pd}} = 3.3\mu s$. $t_{\text{aRxTxturnaroundTime}}$ is typically $5\mu s$.

Taking this into account, we propose a strategy for contention resolution, which uses no ACK to indicate collision, and thereby saving the time required to resolve contention. The algorithm works as follows. If the transmission is idle or success AP broadcasts an ACK packet indicating whether the minislot was idle or collision. After receiving the feedback from AP, users adjust their new threshold and transmit in the next slot. For the event of collision, no feedback is given by AP. Mobile hosts wait for a $t_{\text{PhyRxIndication}}$ for $t_{\text{pd}} + t_{\text{aRxTxTurnaroundTime}} + \delta$ after $t_{\text{pkt}} + t_{\text{p}}$ seconds (where $\delta < t_{\text{ACK}}$. We assume $\delta = 20\mu s$). If PHY did not indicate any reception during that time, users assume that the current minislot is a collision and adjust the new threshold accordingly. Thus, the duration of the minislot in which collision happens reduces to $t_{\text{pkt}} + 2t_{\text{pd}} + t_{\text{aRxTXturnaroundtime}} + \delta(160\mu s + 6.6\mu s + 5\mu s + 20\mu s = 191.6\mu s)$. The duration of regular minislot is $331.6\mu s$. Thus, for the modified scheme an i.i.d. 2 user system takes $593.2\mu s(2 \times (\frac{3}{4} \times 331.6\mu s + \frac{1}{4} \times 191.6\mu s))$ on average to select the best user. Without the modified algorithm, the average time required for contention resolution is $663.2\mu s(2 \times 331.6\mu s)$.

4.2 Simulation Results

In this section, we will now report the performance of MPA using simulations for a variety of network scenarios. We will also compare the performance of MPA with other contention resolution strategies such as polling and channel-gain based timer schemes. In all our simulations in this section, we will consider N users in the network. For splitting, we assume that a slot is divided into K minislots and the first set of k minislots are used for splitting. If contention is not resolved in the first k minislots, a user is scheduled randomly for the remainder of the slots. We have considered the CDMA/HDR channel model reported in Bender *et al.* (2000) for all our simulations. We use three different probability assignment ($\{p_1\}$, $\{p_2\}$ and $\{p_3\}$) for the AMC levels as shown in table 4.1. For example, if a user takes channel distribution $\{p_1\}$, its supported data rate (in Kbps) in a slot is 1228.8 with probability $\frac{3}{11}$, 1843.2 with probability $\frac{3}{11}$, and 2457.7 with probability $\frac{5}{11}$.

Data Rate	$\{p_1\}$	$\{p_2\}$	$\{p_3\}$
38.4	0	1/11	5/11
76.8	0	1/11	3/11
102.6	0	1/11	3/11
153.6	0	1/11	0
204.8	0	1/11	0
307.2	0	1/11	0
614.4	0	1/11	0
921.6	0	1/11	0
1228.8	3/11	1/11	0
1843.2	3/11	1/11	0
2457.7	5/11	1/11	0

Table 4.1: Supported data rates (in Kbps) and probability assignment for the simulations reported in this section (from Bender *et al.* (2000)).

4.2.1 Average Delay

We will now report the delay performance of MPA for identically distributed and non-identically distributed user channels. In Figure 4.3, we plot the CDF of the number of the minislots required to resolve contention for the i.i.d. and non-i.i.d. case. We consider a wireless network with $N = 20$ users, with a slot length of $K = 50$ minislots where the first set of $k = 20$ minislots are used for contention. We assume that if the contention is not resolved within the k minislots, a random user is scheduled for the reminder of the slots. For the i.i.d. case, we assume that the users' channel is distributed as $\{p_2\}$ (see Table 4.1). For non identically distributed channel (channel distributions are still independent across users and time) case, $N_1 = 1$ channel is distributed as $\{p_1\}$, $N_2 = 9$ channels are distributed as $\{p_2\}$ and $N_3 = 10$ channels are distributed as $\{p_3\}$. From the Figure 4.3, we observe that the i.i.d. case takes longer time to resolve contention. Also note that contention is resolved in few minislots with high probability in either case.

4.2.2 Average Throughput

We will now study the average throughput performance of MPA and compare it with other contention resolution strategies. In Figure 4.4, we plot the long run average throughput of the splitting strategy as a function of k, for different values of N. Also, in the plot, we have shown the optimal performance of the polling strategy for the number

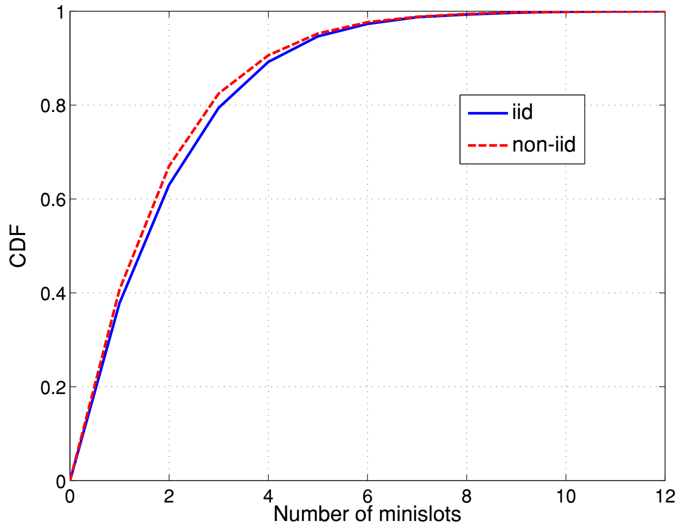

Figure 4.3: CDF of number of minislots required for resolving contention with i.i.d. and non identical channel state distribution.

of users N. From the figure, we observe that the average throughput increases with increase with k and N. Note that the improvement in throughput decreases with increasing value of k implying that the splitting algorithm resolves contention well within few minislots with high probability. The increase in throughput with N is attributed to the multiuser diversity gain. In Figure 4.4, we have compared the average throughput achieved by the splitting strategy with the optimal polling scheme; the number of users to be polled in a slot is adjusted to maximize the system throughput. For example, for $N = 20$ users, with distribution $\{p_2\}$, polling 8 users achieves maximum throughput (the optimal value is computed numerically). From the figure, we note that splitting can significantly improve the network performance as compared to the best polling strategy.

4.2.3 Imperfect Information

In Figure 4.5, we compare the average throughput performance of the different contention resolution strategies under imperfect channel knowledge scenarios. We assume that the users have perfect knowledge of the number of users N but have imperfect information about the channel distribution of the users. In this scenario, we study the performance of the splitting strategy with triple feedback and compare it with polling and a timer-based random access scheme (see Shah *et al.* (2010)). There are $N = 20$

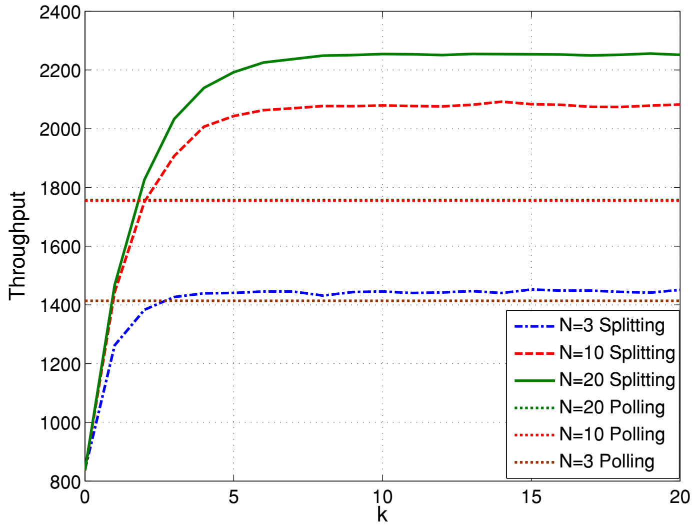

Figure 4.4: Plots of the average throughput of the users with a splitting strategy as a function of k (number of minislots used for contention resolution) for different values of N. The straight lines in the plot correspond to the average throughput performance of the optimal polling strategy for different values of N.

users connected to the access point and the users has the perfect knowledge of N. We assume that scheduler assumes independent channel realizations from the distribution $\{p_2\}$ for all the users while the actual channel is independent realizations from the distribution $\{p_3\}$. In this case, splitting takes a longer time to resolve contention, and therefore the throughput achieved is less in comparison to the perfect channel case. In Figure 4.5 we also plot the average throughput achieved by splitting and other selection schemes for the network scenario. The random selection strategy reported in the figure randomly selects one of the 20 users and schedules in the slot. We note that the simple random selection strategy performs better than the other strategies as it has less overhead. We also compare with the performance of the timer based selection scheme studied in Shah *et al.* (2010). The average throughput obtained using the timer-based scheme is similar to splitting for the network scenario. This is due to fact that both splitting and timer based scheme fails to resolve the best user in the allotted time interval ($k = 20$ for splitting, $\frac{T_{max}}{\Delta} = 40$ for timer based selection scheme. See Shah *et al.* (2010) for details).

In Figure 4.6, we report the average delay performance of MPA for i.i.d. case with imperfect knowledge of the number of users (N) in the network. Suppose that the number of users in the network is random and Uniformly distributed in $\{1, 2, \ldots, 40\}$.

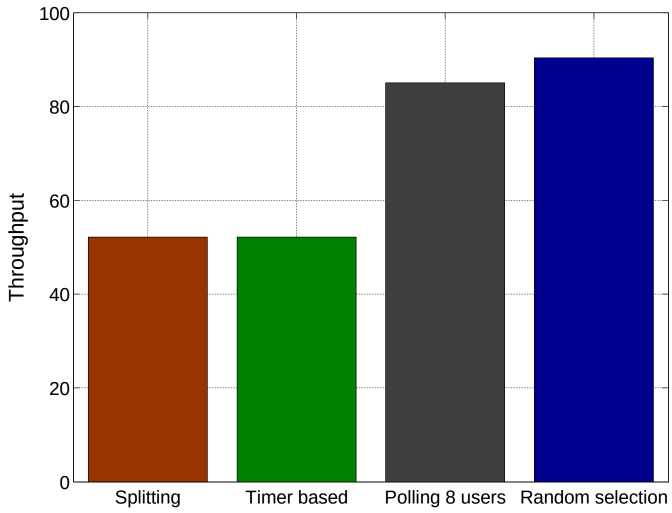

Figure 4.5: Average throughput of splitting, polling, timer-based scheme and random selection for a network with 20 users and with imperfect channel information.

We assume that the user channel distributions are i.i.d. with the distribution $\{p_2\}$ and that the user channel distributions are known at the schedulers. In Figure 4.6, we plot the CDF of the number of minislots required to resolve contention with splitting algorithm with an estimate of $N = 20$ (floor of the average value of N). We observe from the plot that the average delay of splitting algorithm increases significantly with imperfect knowledge of N.

4.3 Conclusion

We have discussed generalizations of the greedy MPA strategy by allowing different thresholds and by aggregating feedback of the users. We note that the optimal greedy strategy for contention resolution depends not only on the channel distribution but also on the network parameters and the metric of interest. Using simulations, we also evaluated the performance of MPA for a variety of network scenarios including networks with imperfect estimates of N and channel distribution. We observed that the MPA is effective even with few minislots dedicated for contention resolution. We also compared the performance of MPA with other contention resolution strategies and identified network scenarios when MPA was optimal.

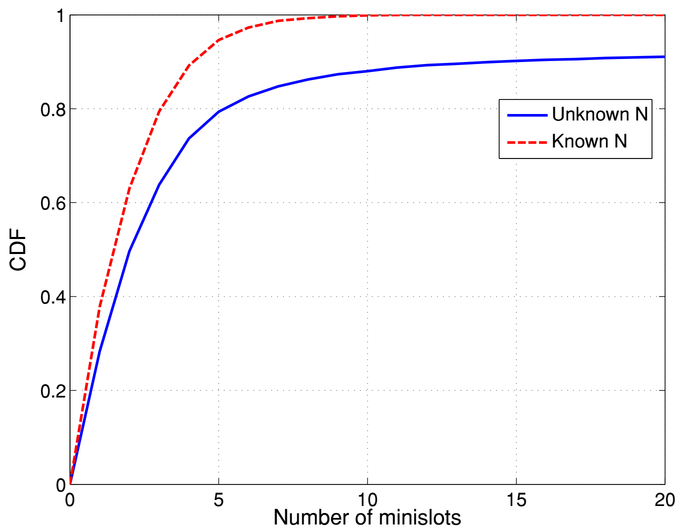

Figure 4.6: CDF of number of minislots required for resolving contention with unknown N.

CHAPTER 5

A Rate Region based Scheduler for Unsaturated Traffic

In Chapters 2, 3 and 4, we studied a problem of resolving the user with the favorable channel condition. In Chapter 2, we studied the problem of opportunistic contention resolution for i.i.d. channel. In Chapter 3, we studied the problem of opportunistic contention resolution with non-identically distributed channel gains and studied the effect of additive and multiplicative scaling on the performance of MPA. Such models are relevant when different users have different network utility functions or QoS objectives. The focus of the earlier chapters was to identify the user with best metric in every slot. In this chapter, we study the problem of identifying the appropriate metric in every slot to support a given quality of service.

We consider a single cell of a cellular data network with a base station and N wireless users. We assume that the channel is slotted and the channel fades randomly over the slots. Let $\bar{R}(t) = (R_1(t), \cdots, R_N(t))$ be the channel state vector in slot t, where $R_i(t)$ is the rate that can be supported for user i in slot if user i is scheduled in a slot. We assume that a single user can be scheduled in the slot and the base station seeks to schedule a user such that the network objective is maximized or the QoS is delivered. In this setup, we assume that perfect CSI is available at the users and the base station and we seek to identify a strategy over the slots such that the desired QoS can be implemented.

We assume that the channel process $\bar{R}(t) = (R_1(t), \cdots, R_N(t))$ is a discrete time Markov chain with a discrete state space $\Omega_R = \{\bar{r}_1, \bar{r}_2, \cdots, \bar{r}_M\}$ where $\bar{r}_j = (r_{j,1}, r_{j,2}, \ldots, r_{j,N})$ is a vector channel state for the network and $r_{j,i}$ is the rate that user i can achieve (if scheduled) when the channel state is $\bar{r}_j$. Let the corresponding stationary probability distribution for the wireless channel be $\mathsf{P}_R = \{\pi_1, \pi_2, \cdots, \pi_M\}$. The set of long term average rate vectors feasible in such a network is defined as the rate region of the wireless network and is described completely using the channel distribution (see Naveen and Ramaiyan (2013)). Let $\mathcal{C}$ be the rate region of the network. Then,

$$\mathcal{C} = \left\{ (r_1, \cdots, r_N) : r_i = \sum_j \pi_j \times a_{ji} \times r_{ji};\ a_{ji} \geq 0;\ \sum_i a_{ji} = 1 \right\} \tag{5.1}$$

where, π_i is the stationary distribution of the wireless channel, a_{ji} is the fraction of time channel j (i.e., $\bar{r}_j$) is allocated to user i and $r_{j,i}$ is the rate achieved for user i when channel j is allocated to user i. If the channel statistics $\{\pi_i\}$ and the instantaneous channel rates $\bar{R}(t)$ are known, then, there exist a stationary schedule that can achieve every rate in the rate region $\mathcal{C}$.

The rate region $\mathcal{C}$ is a bounded convex set, and for strictly concave utility functions, the utility maximizing point on the rate region is unique. Let $\bar{r}^*$ be a point on the rate region at which some concave network utility is maximized and let $\{a^*_{j,i}\}$ be the stationary schedule that achieves $\bar{r}^*$. If the steady state distribution of the channel is known a priori, schedulers can maximize general network utilities on the rate region. In cellular data networks, base station does not have the knowledge of the channel distribution. Base station has the channel state information in every slot only. For concave utilities we can achieve $\bar{r}^*$ in the long term using gradient scheduling algorithms, e.g., Stolyar (2005). But gradient based scheduling algorithms cannot maximize non-concave or non-differentiable network utilities.

In Naveen and Ramaiyan (2013), Naveen and Venkatesh proposed a rate region based scheduler that uses the history of the instantaneous channel state information to make scheduling decision in every slot. In Naveen and Ramaiyan (2013), the authors studied a rate region based scheduler called RRS for a discrete time Markov wireless channel. The rate region based scheduler would estimate the distribution of the wireless channel from the history of the wireless channel as

$$\pi_j(t) = \frac{1}{t} \sum_{k=1}^{t} I_{\{\bar{R}(t) = \bar{r}_j\}}$$

Then, RRS would estimate the rate region of the wireless channel as a function of the estimated channel distribution $\{\pi(t)\}$ as given below.

$$\mathcal{C}(t) = \left\{ (r_1, \cdots, r_N) : r_i = \sum_j \pi_j(t) \times a_{ji} \times r_{ji};\ a_{ji} \geq 0;\ \sum_i a_{ji} = 1 \right\}$$

For the estimated rate region, the RRS scheduler would now identify the operating rate vector $R^*(t)$ corresponding to the desired QoS or the network utility. Then, in every slot t, RRS would implement the stationary schedule $\{a^*_{j,i}(t)\}$ corresponding to the desired operating point $\bar{R}^*(t)$. In Naveen and Ramaiyan (2013), the authors showed that the RRS scheduler is asymptotically optimal for general network utility and QoS.

In Naveen and Ramaiyan (2013), the authors focussed on the implementation of RRS for saturated traffic. They showed that RRS could maximize general network utility functions including non-concave utility functions, provides a parameter-less implementation and has better rate of convergence among other useful properties. In this thesis, we extend the applicability of the scheduler for unsaturated traffic conditions. We show that the RRS scheduler can be used to maximize network utility even when the arrival rate is outside the rate region, or for arrival process controlled using a feedback and also for energy optimization problems for unsaturated traffic. Our exercise demonstrates the general applicability and usefulness of RRS for a variety of network and channel conditions.

The chapter is organized as follows. In section 5.1, we consider a traffic model with average arrival rate outside the rate region and we study a general network utility maximization problem using RRS. In section 5.2, we describe a scheduling strategy based on RRS for utility maximization in systems with feedback based arrival process. In section 5.3, we extend RRS for the energy minimization problem.

5.1 Arrival Rate outside the Rate Region

Opportunistic scheduling algorithms in the context of wireless networks perform two important roles - stabilize an arrival process supported within the rate region and provide fair throughput to the users. Throughput optimal schedulers attempt to serve every arrival rate vector inside the capacity region while optimizing the average queue in the buffer. Fair schedulers seek to achieve utility optimal average throughput vectors within the rate region and are usually not delay concerned. In this section, we seek to maximize a general network utility for an arrival process that is outside the capacity region.

In Neely *et al.* (2008), Neely et al showed that neither throughput optimal schedulers nor fair schedulers can be used for all possible arrival rates. Many schedulers which use queue length information can serve any arrival rate inside the rate region. But when the arrival rate goes outside the rate region, the operating point shifts to some unfair point on the capacity region. Also, fair schedulers like proportional fair (PF) scheduler cannot stabilize all arrival rates inside the capacity region. They proposed a scheduling strategy that maximizes a concave network utility for an extended arrival process. In this section, we propose the application of RRS for maximizing a general network utility for an arrival process outside the rate region. By simulation, we show the usefulness of RRS in obtaining a fair operating point for the network. Unlike the scheduler described in Neely *et al.* (2008), we use a parameter less implementation and can maximize non-concave utility functions on the rate region.

Consider an average arrival rate vector $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)$ strictly outside the rate region and our objective is to support a fraction of the arrival rate maximizing some network utility. Let $U()$ be the utility function on the rate region. The network optimization problem can be formulated as shown below.

$$\begin{aligned} &\text{Maximize: } U(r_1, \cdots, r_N) \\ &\text{Subject to: } (r_1, \cdots, r_N) \in \mathcal{C} \\ &\qquad 0 \leq (r_1, \cdots, r_N) \leq (\lambda_1, \cdots, \lambda_N) \end{aligned} \tag{5.2}$$

In every slot, RRS estimates the channel distribution, $\pi(t)$, and computes the estimated rate region $\mathcal{C}(t)$. Further, we will estimate the time average arrival rate till time t, $\lambda(t)$, from the arrival history. We will now solve the following optimization problem in every slot to identify the optimal rate vector (and the stationary schedule) that maximizes the network utility subject to constraint imposed by the arrival process.

$$\begin{aligned} &\text{Maximize: } U(r_1, \cdots, r_N) \\ &\text{Subject to: } (r_1, r_2, \ldots, r_N) \in \mathcal{C}(t) \\ &\qquad 0 \leq (r_1, \ldots, r_N) \leq (\lambda_1(t), \cdots, \lambda_N(t)) \end{aligned} \tag{5.3}$$

RRS computes the stationary schedule that solves the above optimization problem and uses the schedule in the slot for the channel. By simulation, we show that as $t \to \infty$, the average allocated rate converges to the optimal allocation vector, $\bar{r}^*$, which is the solution to the optimization problem 5.2. We illustrate our implementation with the help of the following example.

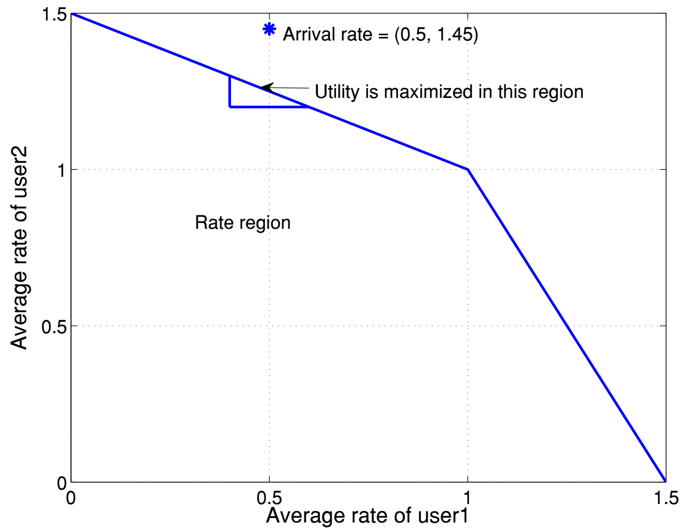

Figure 5.1: Plot of the rate region for the wireless network with $R(t) \in \{(1,2),(2,1)\}$ Mbps with probability $\{\frac{1}{2}, \frac{1}{2}\}$. The average arrival rate of the Poisson process to the two queues and the utility optimizing rate vector is also indicated in the figure.

Example 5.1. Consider a base station communicating with two mobile hosts through a shared wireless channel with time varying capacity . Assume that the channel state in each slot is i.i.d. across time and correlated across users. In any time slot, the channel state vector is in $\{(1,2),(2,1)\}$ Mbps with probability $\{\frac{1}{2}, \frac{1}{2}\}$. We consider two Poisson arrivals with mean arrival rate 0.5 Mbps and 1.45 Mbps to queues corresponding to user 1 and user 2 respectively. The throughput utility function of user 1 and user 2 is defined as $U(r_1, r_2) = U_1(r_1) + U_2(r_2)$, where,

$$U_1(r_1) = \begin{cases} 1 & \text{if } r_1 \geq 0.4, \\ 0 & \text{else} \end{cases} \tag{5.4}$$

$$U_2(r_2) = \begin{cases} 1 & \text{if } r_2 \geq 1.2, \\ 0 & \text{else} \end{cases} \tag{5.5}$$

Where, r_1 and r_2 are the allocated rates to user 1 and user 2.

In Figure 5.1), we show the rate region for the wireless network, the average arrival rate vector and the optimal operating point $\bar{r}^*$ that maximizes the network utility. Note that the arrival process is strictly outside the rate region. The network utility is maximized in the region $\{r_1 \geq 0.4, r_2 \geq 1.2\} \cap \mathcal{C}$, where $\mathcal{C}$ is the rate region for the given

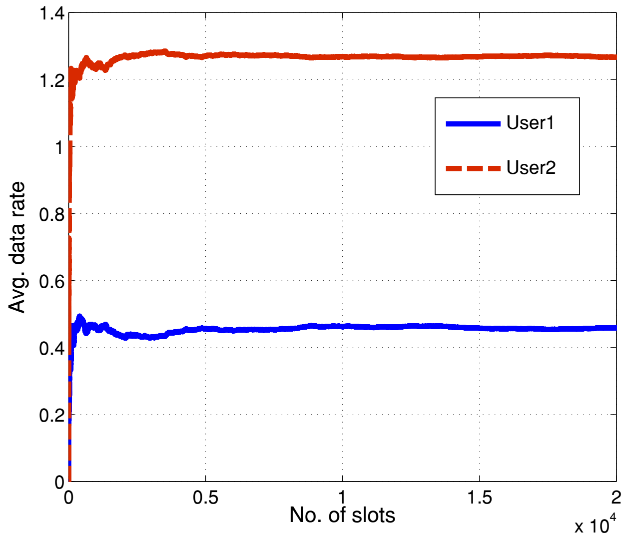

Figure 5.2: Long term average throughput achieved by RRS for the network model shown in Figure 5.1.

channel statistics. In Figure 5.2, we report the performance of the RRS for the extended arrival process. RRS seeks an operating point such that the network utility is maximized constrained to the arrival process. The choice of discontinuous utility function was made to show that RRS imposes no restriction on the nature of the utility functions.

5.2 Feedback based arrival process

In this section, we will study the use of RRS for an arrival process controlled by a feedback. In Eryilmaz and Srikant (2005), Eryilmaz and Srikant studied a traffic model in which data arrival rate is controlled by a feedback from the base station. The feedback is a function of the MAC layer queue length. More the queue length, lesser will be the arrival rate. In such systems, the authors showed that a local queue length based scheduler can achieve fair resource allocation. By controlling parameters in the feedback, they showed that closeness to the optimal allocation can be achieved with trade-off in MAC layer queueing delay. In this context, we propose a scheduling algorithm based on RRS to maximize general utility functions on the rate region. Similar to the section on the extended arrival rates, in every slot, RRS estimates the rate region of the wireless channel and seeks a stationary schedule that maximizes the network utility on the estimated rate region. Using simulations, we show that RRS is asymptotically optimal

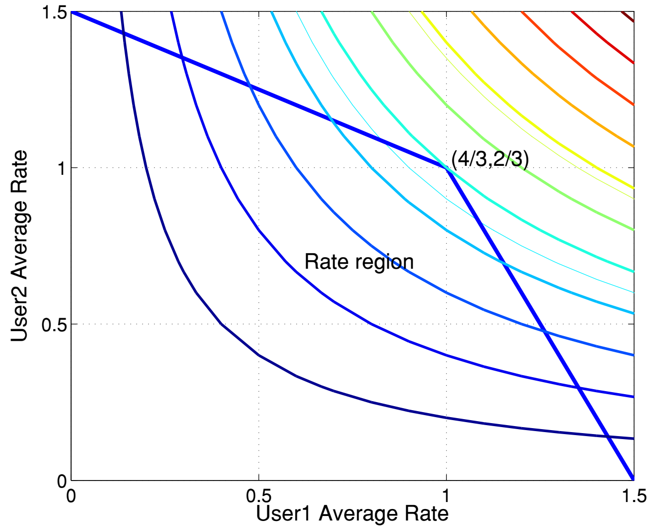

Figure 5.3: Plot of the rate region (Area under the thick blue line is the rate region) for a wireless channel with states $\{(1,2),(2,1)\}$ Mbps and probabilities $(\frac{1}{3},\frac{2}{3})$. The optimal operating point for the proportional fair utility is $(\frac{4}{3},\frac{2}{3})$ Mbps.

and the allocated rate converges to the optimal operating point.

Example 5.2. Consider a two user downlink with channel rate vector $\{(1,2),(2,1)\}$ Mbps, which occur with probability $(\frac{1}{3},\frac{2}{3})$ in each slot. The base station maintains separate queues for the two users. Arriving packets are stored in the queue until it is served. Arrival occurs at the beginning of each slot, which is controlled by the feedback given in the previous slot. We model arrivals as a random process with mean equal to the feedback value. In simulations, Poisson arrivals with mean arrival rate $a_i(t)$ are used,

$$a_i(t) = \min\left(\frac{k}{x_i(t)}, M\right), \text{for } i \in \{1,2\} \tag{5.6}$$

where, $x_i(t)$ is the queue length of user i at time t, and k, M are arbitrarily chosen constants. In Figure 5.3, we plot the rate region for the wireless network and the network utility function $U = \log(r_1) + \log(r_2)$. The network utility is maximized at $(\frac{4}{3},\frac{2}{3})$ in the rate region of the wireless channel. The long term average rate achieved using RRS is reported in Figure 5.4, which shows that the allocated rates using RRS converges to the optimal operating point for the network utility function in the wireless network.

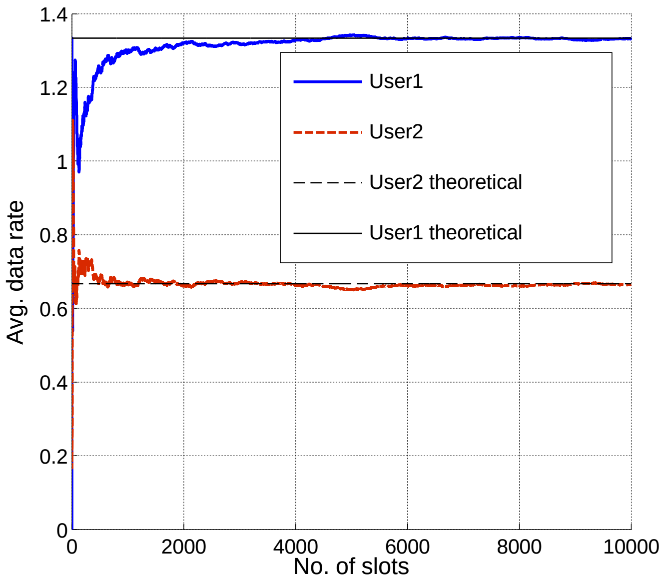

Figure 5.4: Plot of the time average data rate of users with RRS for the feedback controlled arrival process network illustrated in Figure 5.3.

5.3 Energy optimal schedule

In this section, we develop a throughput optimal scheduling policy based on RRS, which minimizes the average power expenditure. Consider an average arrival rate within the capacity region of the wireless channel. From the description of the channel distribution and power allocation policy, we can compute the minimum power required to support the given arrival process. In our work, we use RRS to estimate the capacity region based on the channel history. We will then estimate the average arrival rate and compute the optimal power allocation policy needed to support the estimated arrival process while minimizing the power expenditure.

We assume an Ergodic arrival process with mean arrival rate, λ, inside the rate region. The channel process is assumed to be Ergodic with finite state space. In any slot, the maximum data rate the channel can support is a function of the current SNR (channel state), and the transmission power. We assume that the base station can transmit data in finite number of power levels, which give rise to a discontinuous rate-power curve. A similar model is considered in Neely (2006), with an objective to minimize the average power expenditure. We propose a generic solution based on RRS and demonstrate its applicability using an example.

Define $\alpha_{j,i,k}$ as the fraction of time channel state $\bar{r}_j$ is allocated to user i with a rate $r_{j,i,k}$ using a power P_k. To minimize the average power consumed, we need to find the optimal $\{\alpha_{j,i,k}\}$ (schedules), subject to the stability constraint. This problem can be formulated as follows.

$$\begin{aligned} \min \; & \sum_j \pi_j \sum_{i,k} \alpha_{j,i,k} P_k \\ \text{s.t.} \; & \alpha_{j,i,k} \geq 0 \,\forall\, i, j, k, \\ & \sum_{i,k} \alpha_{j,i,k} = 1 \,\forall\, j, \\ & \lambda_k(t) \leq \sum_{j,i} \alpha_{j,i,k} r_{j,i,k} \end{aligned} \tag{5.7}$$

The last inequality guarantees the stability for the arrival process. Here, we assume that the transmission rate depends only on the transmission power and channel state of the user and the allocated power $r_{j,i,k}$. In our proposed extension to RRS, we solve the optimization problem in every slot using the estimated channel distribution and the average arrival rate and implement the optimal strategy in the slot.

Example 5.3. Consider a two user example in cellular downlink. Channel condition for user 1 and 2 can be in GOOD or BAD state. We assume that the channel condition is correlated across users with state vector $R(t) \in \{(\text{BAD, GOOD}), (\text{GOOD, BAD})\}$ with probability distribution $\{0.5, 0.5\}$. Arrival occur at the beginning of each slot with mean arrival rate vector $\lambda = (0.75, 0.75)$. In simulations, we use Poisson random variable to model the arrival process. The base station uses a binary power control scheme. i.e. we use an ON/OFF model. If the base station is ON, it transmits at a power of 1 Watt, and this corresponds to a peak rate of 2 Mbps if the channel state is GOOD, and 1 Mbps if the channel state is BAD. We do not spend any power when the power control is OFF, i.e., $P_1 = 1$ and $P_2 = 0$. Note that, for the given channel process and the power control policy, the average arrival rate vector is inside the rate region. The average power minimization problem for this particular example can be formulated as follows. Trivially, $\alpha_{1,1,2} = \alpha_{1,2,2} = \alpha_{2,1,2} = \alpha_{2,2,2} = 0$ because $P_2 = 0$. Substituting

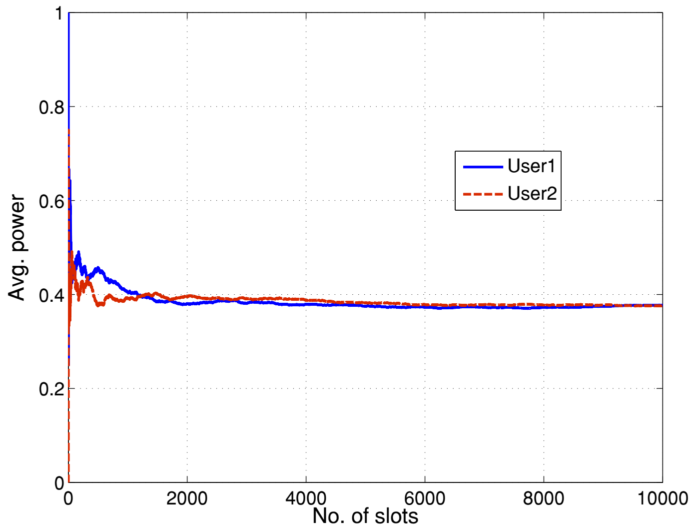

Figure 5.5: Plot of the average power spent by RRS to support an arrival rate. The thin straight line indicates the minimum power required to support the arrival process.

for P_1, P_2 and $\{r_{j,i,k}\}$, the optimization problem simplifies as shown below.

$$\begin{aligned} \min\ & \alpha_{1,1,1} + \alpha_{1,2,1} + \alpha_{2,1,1} + \alpha_{2,2,1} \\ \text{s.t. } & \alpha_{1,1,1} + \alpha_{1,2,1} \leq 1, \\ & \alpha_{2,1,1} + \alpha_{2,2,1} \leq 1, \\ & 0.75 \leq \alpha_{1,1,1} + 2\alpha_{2,1,1}, \\ & 0.75 \leq 2\alpha_{1,2,1} + \alpha_{2,2,1} \end{aligned} \tag{5.8}$$

Solution to the optimization problem 5.8 yields a minimum average power vector equal to $P_{min} = (0.375W, 0.375W)$.

The plot of the average power spent as a function of time, for RRS, is reported in Figure 5.5. The figure shows that RRS stabilizes and can achieve energy optimal schedule for an arrival rate within the capacity region. By estimating the rate region and the average arrival rate in each slot, the RRS algorithm finds the energy optimal schedule in every slot. Due to randomness in arrival and channel process, in some initial slots, the estimated average arrival rate need not be inside the estimated rate region. In this case, we schedule the non-empty queue which maximize the services rate in the slot.

5.4 Conclusion

In this chapter, we have discussed the use of RRS scheduler to provide QoS in wireless networks under unsaturated traffic conditions. We have demonstrated the applicability of RRS for extended arrival rates, feedback controlled arrival process and for energy optimal strategies. Our use of RRS exploits the arrival and channel history in making scheduling decision. The scheme is useful in the sense that it imposes no restrictions on the nature of QoS. This translates to larger freedom to the network operator for implementing a variety of services available in modern wireless systems.

CHAPTER 6

Conclusion

In this thesis, we have studied the performance of a greedy opportunistic contention resolution strategy, MPA, for a variety of network scenarios. For i.i.d. user channel, we observed that the performance of MPA is approximately optimal. In particular, for the $N = 2$ user case, MPA is both average delay optimal and entropy optimal. For $N > 2$ and for the complete feedback model, we showed that MPA was slightly suboptimal with respect to delay. We also observed that the entropy optimal strategy and the delay optimal strategy were not the same.

The performance of MPA was also studied for non-identically distributed channel and correlated wireless channel. Using analysis and simulation, we showed that the average delay to contention resolution decreases with asymmetry of the channel distribution. For $N = 2$ users and for independent and non-identically distributed channel, we proved that the average delay is upper bounded by 2 minislots (the average delay of the optimal scheme for the i.i.d. case). Using numerical work, we noted that the observation holds for large N as well. For the special case of multiplicative and additive scaling of distributions, we characterized the average delay of MPA as a function of the scaling. For a correlated wireless channel, we observed that MPA can be strictly suboptimal and discussed the need for caution.

The opportunistic splitting strategy attempts to identify a random threshold between the user with the best channel and the user with the second best channel. This permits a source coding problem on the random threshold with the entropy of the threshold related to the average delay to resolve contention. Throughout the thesis, we studied the entropy of the contention resolution strategy along with the average delay of the strategy. We noted that there is a strong correlation between the entropy and the average delay. We also obtained a bound between the entropy and the average delay. This motivated us to study a simple entropy minimization formulation to identify average delay optimal strategies. We observed that the entropy minimization formulation is a concave minimization problem. Using an example, we noted that MPA is a local minima and there can be many solutions to the problem. Thus, the average delay minimization is a difficult problem to study and we may only aim to obtain bounded-optimal strategies for a network scenario.

In chapter 4, we studied generalizations of the MPA strategy using examples. We studied network scenarios that permit different thresholds for users and contention aggregation. The examples illustrated the necessity to model the contention resolution problem based on the correct network objective. Using simulations, we evaluated the performance of MPA for a variety of network scenarios and compared it with other contention resolution strategies such as polling and channel gain based random access. We identified scenarios where splitting based algorithms were a good choice in comparison with the other strategies.

In Chapter 5, we discussed the applicability of a known opportunistic scheduling strategy called RRS for the unsaturated traffic model. We studied RRS for a number of interesting network scenarios such as utility maximization for extended arrival rates, for feedback controlled arrival process and for the energy minimization problem. Using simulations, we showed that RRS can be a good choice of a scheduler in a variety of network and channel scenarios.

REFERENCES

1. **3GPP** (2012). Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification. *3GPP, Tech. Rep. TS 36.321*.

2. **Adireddy, S.** and **L. Tong** (2005). Exploiting decentralized channel state information for random access. *IEEE Transactions on Information Theory*, **51**(2), 537–561.

3. **Anantharam, V.** and **P. Varaiya** (1986). An optimal strategy for a conflict resolution problem. *Systems & control letters*, **7**(4), 329–332.

4. **Andrews, M.**, **L. Qian**, and **A. Stolyar** (2005). Optimal utility based multi-user throughput allocation subject to throughput constraints. *24th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM*, **4**, 2415–2424.

5. **Arrow, K. J.**, **L. Pesotchinsky**, and **M. Sobel** (1981). On partitioning a Sample with Binary-type Questions in lieu of Collecting Observations. *Journal of the American Statistical Association*, **76**(374), 402–409.

6. **Bender, P.**, **P. Black**, **M. Grob**, **R. Padovani**, **N. Sindhushyana**, and **S. Viterbi** (2000). CDMA/HDR: a Bandwidth Efficient high speed Wireless Data Service for Nomadic users. *IEEE Communications Magazine*, **38**(7), 70–77.

7. **Bletsas, A.**, **A. Khisti**, **D. P. Reed**, and **A. Lippman** (2006). A Simple Cooperative Diversity Method Based on Network Path Selection. *Selected Areas in Communications, IEEE Journal on*, **24**(3), 659–672.

8. **Cover, T. M.** and **J. A. Thomas**, *Elements of information theory*. John Wiley & Sons, 2012.

9. **Eryilmaz, A.** and **R. Srikant** (2005). Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control. *24th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM*, **3**, 1794–1803.

10. **Gallager, R. G.** (1978). Conflict resolution in random access broadcast networks. *Proc. of the AFOSR Workshop in Communication Theory and Applications*, 74–76.

11. **Gesbert, D.** and **M.-S. Alouini** (2004). How much Feedback is Multi-user Diversity really worth? *IEEE International Conference on Communications*.

12. **Goldsmith, A. J.** and **P. P. Varaiya** (1997). Capacity of fading channels with channel side information. *IEEE Transactions on Information Theory*, **43**, 1986–1992.

13. **Gopalan, A.**, **C. Caramanis**, and **S. Shakkottai** (2012). On Wireless Scheduling With Partial Channel-State Information. *IEEE Transactions on Information Theory*, **58**, 403–420.

14. **IEEE802.11** (2012). Ieee standard for information technology–telecommunications and information exchange between systems local and metropolitan area networks–specific requirements part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications. *(Revision of IEEE Std 802.11-2007)*.

15. **IEEE802.16** (2006). IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2. *IEEE Std 802.16e-2005 and IEEE Std 802.16-2004/Cor 1-2005 (Amendment and Corrigendum to IEEE Std 802.16-2004)*, 1–822.

16. **Jahn, A.** and **A. Bottcher** (1993). Improvement of slotted aloha for land-mobile satellite communications, using channel state information. *43rd IEEE Conference on Vehicular Technology*, 339–342.

17. **Jalali, A.**, **R. Padovani**, and **R. Pankaj** (2000). Data throughput of cdma-hdr a high efficiency high data rate personal communication wireless system. *IEEE Vehicular Technology Conference Proceedings*, **3**, 1854–1858.

18. **Karaca, M.**, **Y. Sarikaya**, **O. Ercetin**, **T. Alpcan**, and **H. Boche** (2012). Throughput optimal scheduling with dynamic channel feedback. *arXiv preprint arXiv:1203.5362*.

19. **Kim, Y.**, **F. Baccelli**, and **G. De Veciana** (2011). Spatial reuse and fairness of mobile ad-hoc networks with channel-aware csma protocols. *International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, 360–365.

20. **Knopp, R.** and **P. A. Humblet** (1995). Information capacity and power control in single-cell multiuser communications. *IEEE International Conference on Communications*, **1**, 331–335.

21. **Liu, X.**, **E. K. P. Chong**, and **N. B. Shroff** (2003). A framework for opportunistic scheduling in wireless networks. *Computer Networks*, **41**, 451–474.

22. **Luenberger, D. G.** and **Y. Ye**, *Linear and Nonlinear Programming*, volume 116. Springer, 2008.

23. **Naveen, S.** and **V. Ramaiyan** (2013). A General Scheduling Strategy based on an Estimate of the Rate Region. *Fifth International Conference on Communication Systems and Networks (COMSNETS)*, 1–8.

24. **Neely, M. J.** (2006). Energy Optimal Control for Time-Varying Wireless Networks. *IEEE Transactions on Information Theory*, **52**, 2915–2934.

25. **Neely, M. J.**, **E. Modiano**, and **C.-P. Li** (2008). Fairness and optimal stochastic control for heterogeneous networks. *IEEE/ACM Transactions on Networking*, **16**(2), 396–409.

26. **Neely, M. J.**, **E. Modiano**, and **C. E. Rohrs** (2005). Dynamic power allocation and routing for time-varying wireless networks. *IEEE Journal on Selected Areas in Communications*, **23**(1), 89–103.

27. **Qin, X.** and **R. Berry** (2001). Exploiting multiuser diversity in wireless aloha networks. *Proceedings of the annual Allerton conference on communication control and computing*, **39**(2), 793–794.

28. **Qin, X.** and **R. Berry** (2004). Opportunistic splitting algorithms for wireless networks. *Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies, INFOCOM*, **3**, 1662–1672.

29. **Qin, X.** and **R. Berry** (2006). Opportunistic splitting algorithms for wireless networks with fairness constraints. *4th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt)*, 1–8.

30. **Ramaiyan, V.** (2013). An Information Theoretic Point of View to Contention Resolution. *arXiv preprint arXiv:1307.8230*.

31. **Sanayei, S.** and **A. Nosratinia** (2005). Exploiting multiuser diversity with only 1-bit feedback. *IEEE Wireless Communications and Networking Conference*, **2**, 978–983.

32. **Shah, V.**, **N. B. Mehta**, and **R. Yim** (2010). A complete characterization of an optimal timer based selection scheme. *IEEE International Conference on Communications (ICC)*, 1–5.

33. **Shakkottai, S.** and **A. L. Stolyar** (2002). Scheduling for multiple flows sharing a time-varying channel: The exponential rule. *Translations of the American Mathematical Society-Series 2*, **207**, 185–202.

34. **So, J.** (2009). Opportunistic feedback with multiple classes in wireless systems. *IEEE Communications Letters*, **13**(6), 384–386.

35. **Stolyar, A. L.** (2005). On the Asymptotic Optimality of the Gradient Scheduling Algorithm for Multiuser Throughput Allocation. *Operations Research*, **53**, 12–25.

36. **Tang, T.** and **R. W. Heath** (2005). Opportunistic Feedback for Downlink Multiuser Diversity. *IEEE Communications Letters*, **9**(10), 948–950.

37. **Tassiulas, L.** and **A. Ephremides** (1992). Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Transactions on Automatic Control*, **37**(12), 1936–1948.

38. **Xue, Y.** and **T. Kaiser** (2007). Exploiting Multiuser Diversity with Imperfect one-bit Channel State Feedback. *IEEE Transactions on Vehicular Technology*, **56**(1), 183–193.

LIST OF PAPERS BASED ON THESIS

1. Venkatesh R and Vaishakh J, An Information Theoretic Point of View to Contention Resolution, *submitted to COMSNETS 2014*.
2. Vaishakh J and Venkatesh R, A Rate Region Based Scheduler for Unsaturated Traffic, *submitted to NCC 2014*.