A General Scheduling Strategy based on an Estimate of the Rate Region

A THESIS

submitted by

NAVEEN S

for the award of the degree

of

MASTER OF SCIENCE

(by Research)

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS.

JULY 2012

THESIS CERTIFICATE

This is to certify that the thesis titled **A General Scheduling Strategy based on an Estimate of the Rate Region**, submitted by **Naveen S** to the **Indian Institute of Technology Madras**, for the award of the degree of **M.S. (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,
Electrical Engineering,
IIT Madras.

Chennai 600 036.
Date:

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to my research advisor, Dr. Venkatesh Ramaiyan, for his constant support, guidance and invaluable insights into the research problem, over the entire duration of my postgraduate study. I am deeply indebted to him for the training he provided.

I thank Prof. Giridhar, Prof. Devendra Jalihal, Dr. Gaurav Raina and Dr. Krishna Sivalingam for their valuable feedback

I wish to thank Prof. Bhaskar Ramamurthi, Prof. Aravind, R., Prof. Ravinder David Koilpillai and Dr. Andrew Thangaraj for the enjoyable coursework.

I am thankful to my seniors, Vidyadar, Pavan, Jothi, Hariharan and Bama for their guidance and friendship.

I am grateful to all my friends at the institute, including Sreenath, Amarnath, Karthik, Amardeep, Sashidharan, Murali, Karthik Upadhyay and Mahesh, for their warm friendship.

Special thanks to my special friends Hariharan, Sriram, Lavanya, Deepak, Sathya, Lalitha and Varun for their enduring friendship over the years.

Finally I express my love and thanks to amma, Ruban and thatha for being the constants of my life.

ABSTRACT

In this thesis, we study a general scheduling strategy for the downlink wireless channel of a single cell of a cellular data network. A fixed number of users share the slotted wireless channel using time division multiplexing. The user queues at the base station are assumed to be saturated, i.e., the users always have data to send. The slotted wireless channel is modelled as a random and time varying process (possibly due to multipath fading). We assume that the channel state information (CSI) of every user is available at the base station at the beginning of every slot, and, the base station uses the CSI to determine a schedule in the slot. In this set up, the objective of the centralized scheduler at the base station is to implement a network quality of service (QoS) or maximize a network utility, both defined on the long time average user throughputs.

Popular scheduling strategies, like the gradient algorithm, use instantaneous channel rates and estimates of few other channel parameters, to identify a schedule in a slot. We propose a rate region based scheduler RRS that uses the entire available channel history to decide upon a schedule. In every slot, the RRS uses the available channel history to estimate the wireless channel statistics. The estimate is used to identify a rate region and an optimal rate vector in the rate region (as a function of the network QoS or the network utility). The scheduler, then, implements a schedule corresponding to the optimal rate vector in the slot. We prove that RRS is asymptotically optimal for all continuous network utilities, for some ergodic channels with discrete channel states. We note that the channel history based scheduler requires a consistency in implementation and also has poor convergence behaviour.

Then, we propose a simple *channel allocation* based scheduler, RRS-CA, which uses both schedule history and channel history, and consequently, has better convergence behaviour than RRS. We prove the asymptotic optimality of RRS-CA scheduler as well. We also propose a practical variant of the rate region based schedulers, called RRS-RA (RRS using allocated rates), which uses schedule history in the form allocated

rates. The RRS-RA is a gradient scheduler that uses a dynamic, auxiliary network utility to drive the average allocated rate (throughput) vector towards the estimated optimal rate vector. RRS-RA has several practical advantages over RRS as it makes consistent schedule decisions and has better convergence behaviour than RRS.

Using simulations, we show that the rate region based scheduling strategy provides us a common framework to implement arbitrary notions of QoS and fairness defined on the long time average user throughputs. We show that the rate region based schedulers can implement QoS defined using arbitrary continuous and some non-continuous functions, provide a parameter-less implementation and has better convergence behaviour than network utility based gradient schedulers.

The rate region based schedulers are computationally expensive to implement; the scheduler needs to compute the rate region, an optimal rate vector and a corresponding schedule in every slot. In the last chapter of the thesis, we discuss simple implementations of rate region based schedulers for max-min fairness, for the two user case and for N users. The schedulers make limited use of channel history and schedule history information, and approximates optimal rate vector and the corresponding schedule to minimize computation. Simulation results show that the performance of the proposed schedulers (simple and approximate versions) are comparable to RRS-RA schedulers and other known schedulers that implement max-min fairness. We also propose a simple technique that uses channel history information to increase the convergence rate of some schedulers in literature.

TABLE OF CONTENTS

CHAPTER 1

Introduction

We consider the downlink wireless channel of a single cell of a cellular data network. A fixed number of users share the slotted wireless channel using time division multiplexing. We assume that the user queues at the base station are saturated, i.e., the users always have data to send. The slotted wireless channel is modeled as a random and time varying process. The randomness in the channel is attributed to multipath fading and other channel effects. We assume that the channel state information (CSI) of every user is available at the base station at the beginning of every slot. Also, we assume that the base station uses the CSI to identify a user to schedule, i.e., the base station is opportunistic. In this set up, our objective is to implement a network quality of service (QoS) or maximize a network utility, both defined on the long time average user throughputs.

Data communication in a cellular setup is getting increasingly popular, especially with the advent of 3G/4G wireless standards (see e.g., IEEE 802.16m (2011), 3GPP (2011)). Key reasons to the success of such cellular data protocols are high data rates, large coverage area and support for user mobility. A big challenge in optimizing network performance in this setup involves handling wireless channel fading and inter-cell interference through appropriate feedback and scheduling mechanisms. In this work, we consider a single cell of a cellular data network. Hence, we do not study interference and are concerned only with channel fading.

A popular scheduling strategy for a fading wireless channel in this setup is opportunistic scheduling. Opportunistic schedulers make use of multiuser diversity and improves the overall network performance by scheduling users with favorable channel conditions (see Liu *et al.* (2003*a*)). However, such schedulers are known to be unfair to users with poorer channel distributions (see, e.g., Holtzman (2001)). Recently, a number of papers have proposed to implement different notions of fairness and QoS objectives among users, for example, proportional fairness (see Jalali *et al.* (2000)), minimum and maximum rate guarantees (see Liu *et al.* (2003*a*), Andrews *et al.* (2005)) and temporal fairness (see Liu *et al.* (2003*a*), Liu *et al.* (2003*b*)), while opportunistically enhancing the network performance.

Most implementations of fair schedulers on user throughputs attempt to maximize a concave and continuously differentiable utility function (that defines fairness) on the user throughputs. The scheduler in such cases is usually a gradient algorithm or is of the stochastic approximation type, and the average allocated rate (throughput) vector converges to a local optimum (which is in fact, the global optimum for strictly concave network utilities). There do not exist implementations of fairness based on non-concave or non-continuous utility functions, even though there is no reason to assume that user satisfaction metrics are concave functions of their average throughputs. Also, there has been little focus on the optimal convergence rate of fair schedulers. In this work, we study a general scheduling strategy that can implement fairness based on arbitrary continuous network utilities, including non-concave and non-differentiable functions. Using simulations, we show that the proposed scheduler has better convergence behaviour than network utility based gradient schedulers.

Popular scheduling strategies, like the gradient algorithm, use instantaneous channel rates and estimates of only a few channel parameters, to identify a schedule in a slot. In this work, we propose a rate region based scheduler, RRS, that uses the entire available channel history to determine a schedule. In every slot, RRS uses the available channel history to estimate the wireless channel statistics. The estimate of the wireless channel is used to identify a rate region and an optimal rate vector in the rate region (as a function of the QoS or the network utility). The scheduler, then, implements a schedule corresponding to the optimal rate vector in that slot. We show that RRS is asymptotically optimal for all continuous network utilities, for some ergodic channels with discrete channel states. We note that the channel history based scheduler requires a consistency in implementation and also has poor convergence behaviour.

Then, we propose a simple *channel allocation* based scheduler, RRS-CA, which uses both schedule history and channel history, and consequently, has better convergence behaviour than RRS. We prove the asymptotic optimality of RRS-CA scheduler as well. We also propose a practical variant of RRS called RRS-RA (RRS using allocated rate), which uses schedule history in the form of allocated rates. The RRS-RA is a gradient scheduler that uses a dynamic, auxiliary network utility to drive the average allocated rate (throughput) vector towards the estimated, global optimal rate vector in every slot (unlike network utility based gradient algorithms that seek local optima). RRS-RA has several practical advantages over RRS as it makes consistent schedule decisions and has better convergence behaviour than RRS.

Using simulations, we show that the rate region based scheduling strategy provides us a common framework to implement arbitrary notions of QoS and fairness defined on the long time average user throughputs. We observe that the rate region based schedulers can implement QoS defined using arbitrary continuous and some non-continuous functions, provide a parameter-less implementation and has better convergence behaviour than network utility based gradient schedulers.

The rate region based schedulers are computationally expensive to implement; the scheduler needs to compute the rate region, an optimal rate vector and a corresponding schedule in every slot. In the last chapter of this thesis, we discuss simple and computationally inexpensive implementations of rate region based schedulers for max-min fairness. Simulation results show that the performance of the simple and approximate versions are comparable to RRS-RA schedulers and other known schedulers that implement max-min fairness. Finally, we propose a simple technique that uses channel history information to improve the convergence rate of some popular schedulers for max-min fairness.

1.1 Motivation

In this section, we provide a motivation (using an example) for the rate region based scheduling strategy proposed in this work. Consider a wireless system with 4 users and a strictly concave and continuously differentiable network utility $U = \sum_{i=1}^{4} U_i$, where U_1, U_2, U_3 and U_4 are the user utility functions represented in Figure 1.1 (U_2, U_3, U_4 are logarithm functions). U_1 is a strictly concave function and has a continuous but sharply varying gradient. The $\log$ utility of users 2,3 and 4 is a strictly concave and continuously differentiable function in $\Re^4 \backslash 0^4$. The gradient scheduler (Stolyar (2005*b*)) is an asymptotically optimal scheduler for the given network utility U. However, since the scheduler uses gradient of U in its scheduling decision, its convergence behaviour is affected by the *shape* of the utility function. This effect is most pronounced when the gradient of U has sharp variations near the optimal rate vector.

In Figure 1.2 we have plotted the performance of the gradient scheduler for the utility U for the 4 users. We consider a user correlated, time independent channel model for the users, described in detail in Channel model A in Appendix E. (Please refer Section 2.4 for more information on the simulation setup). The optimal rates for the network utility U and the channel model A are marked by an o (for user 1) and by a $*$ (for users $2, 3$ and 4) in Figure 1.1. In Figure 1.2, we have plotted the time average rate vector of the gradient scheduler for a realization of the wireless channel; the gradient scheduler studied in Stolyar (2005*b*) is known to be asymptotically optimal for almost all realizations. We see from the figure that gradient scheduler exhibits slow convergence for the wireless system. The reason for this behaviour is that the gradient scheduler implements very different schedules in the neighborhood of the optimal rate vector (due to the sharp variation in the utility of user 1). This results in sub-optimal scheduling until the average rate vector $\bar{R}(t)$ settles sufficiently. Such sub-optimal performance of gradient scheduler caused by the shape of the utility function was reported in Andrews *et al.* (2005) in the context of implementing rate constraints using a modified network utility. However, the concave network utility itself could possibly have such gradient variations. Thus, gradient algorithms may not always be suitable even when the network utility is strictly concave and continuously differentiable.

A scheduling algorithm which does not rely entirely on local gradient information (and hence on the shape of the network utility) would be a better alternative in such a scenario. Particularly, a scheduler with sufficient knowledge of the entire rate region need not rely on local gradient information and can thus have a *utility-independent convergence* behaviour. In Chapter 5, Section 5.6, we report that the rate region based schedulers, in fact, can seek the global optimum without being affected by gradient variations in U. The rate region based schedulers proposed in this thesis use an estimate of the rate region of the wireless channel and seek the globally optimal rate vector directly, unlike the network utility based gradient schedulers that depend on gradient of U to implement the optimal rate vector. The above example, clearly illustrates the importance of channel history information and the necessity to seek global optimal rate vector in every slot.

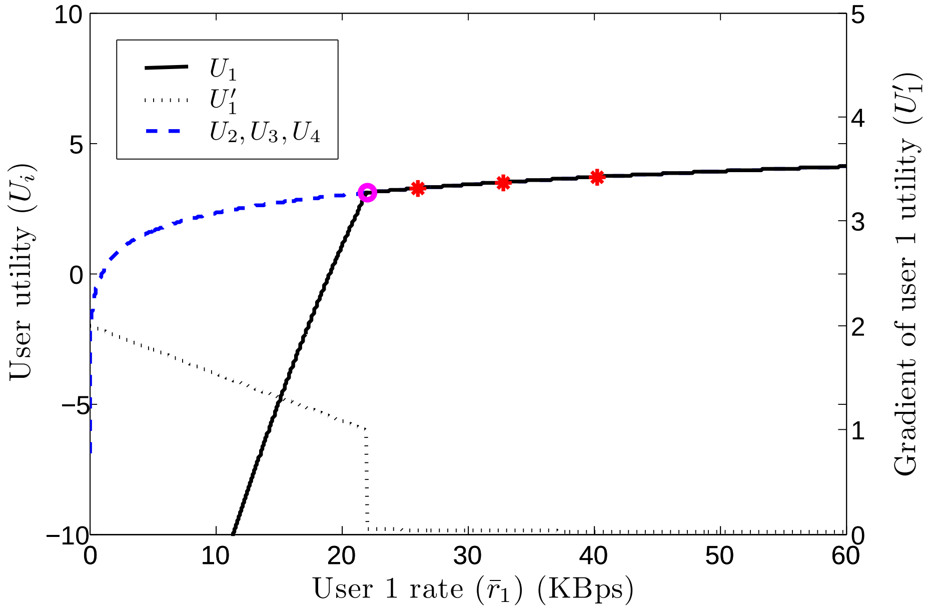

Figure 1.1: Plots of strictly concave and continuously differentiable user utility functions U_1 through U_4 with respect to the average user throughput. Also plotted in the figure is the gradient of utility function of user 1, U_1'. The optimal rates for the network utility $\sum_{i=1}^{4} U_i$ and for the channel model A (described in Appendix E), are marked by an o (user 1) and * s (user 2,3 and 4) in the figure.

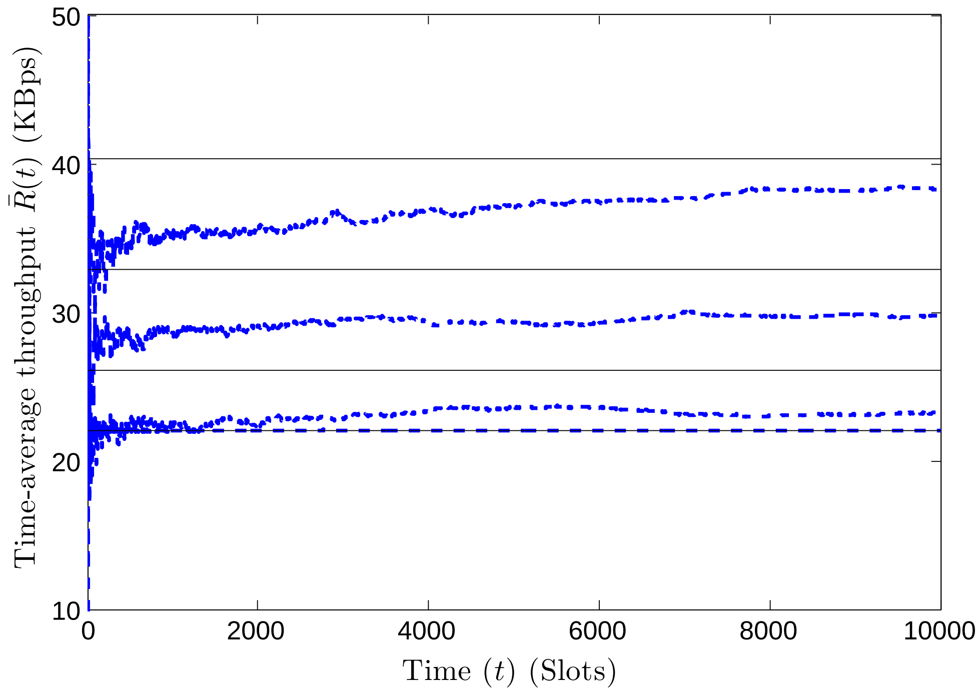

Figure 1.2: Plots of time average user throughputs of the gradient scheduler (Stolyar (2005*b*)) for the network utility $U = \sum_{i=1}^{4} U_i$ (described in Figure 1.1) and the wireless channel model A described in Appendix E. The optimal rate vector for the wireless system is marked using thin straight lines in the figure.

1.2 Related Literature

The asymptotic properties of opportunistic, fair and optimal resource sharing algorithms, for cellular data networks, were first studied in Kushner and Whiting (2002) and in Kushner and Whiting (2004) using a stochastic approximation framework. The limiting behaviour of the user throughputs of algorithms like the proportional fair schedule (see Jalali *et al.* (2000)), were shown to optimize a strictly concave and continuously differentiable network utility over the rate region of the wireless channel. In Stolyar (2005*b*), the transient dynamics of user throughputs for gradient (on a network utility) schedulers was studied using fluid sample paths and their asymptotic optimality was shown for general network scenarios. The schedulers discussed in Kushner and Whiting (2004) and Stolyar (2005*b*), use local gradient information and their applicability is restricted to concave and continuously differentiable utilities. We propose a general rate region based scheduler that can optimize any continuous network utility over the rate region. Also, our proof of the asymptotic optimality of schedulers uses techniques different from those used in Kushner and Whiting (2004) and Stolyar (2005*b*).

The gradient scheduling algorithm was generalized in Andrews *et al.* (2005) and in Stolyar (2005*a*) to optimize concave and continuously differentiable utility functions subject to throughput constraints. A general and unified framework for opportunistic scheduling that implements temporal fairness, utilitarian fairness and minimum performance guarantees was proposed in Liu *et al.* (2003*a*). The stochastic approximation based scheduler in Liu *et al.* (2003*a*) is restrictive in its implementation of utilitarian fairness and minimum performance guarantees as it implements fixed ratio of user throughputs and provides hard minimum rate guarantees only. The RRS schedulers provides a general framework that can implement arbitrary QoS including utility maximization with multiple rate constraints (see Chapter 5 for interesting QoS examples). Gains of multiuser diversity as a function of the quality of channel feedback is studied in Agrawal and Subramanian (2002). We assume that the base station has perfect CSI, but consider arbitrary definitions of QoS.

There is considerable interest and literature concerning stability of queueing networks and wireless systems. In Tassiulas and Ephremides (1992), the authors propose a maximum throughput policy for multihop radio networks, called the max-weight scheduling policy, that stabilizes any random arrival process within the rate region. A throughput optimal dynamic power allocation and routing strategy was proposed in Neely *et al.* (2005*b*) to stabilize wireless systems with time varying wireless channels and random arrivals. In Neely *et al.* (2005*a*), the authors propose a dynamic strategy to support all traffic whenever possible, and to make optimally fair decisions about the data to serve when inputs exceed network capacity. For arrival rates outside the rate region, fair rate vector is defined using a sum of concave utilities on user throughput. In our work, we assume that the users are saturated and our objective is to implement the prefered (arbitrary) QoS to the saturated users.

For systems with random arrivals, packet delay and buffer occupancy are other metrics of interest along with long time average user throughputs. In Stolyar and Ramanan (2001), the authors propose largest weighted delay first schedule that maximizes the asymptotic rate of decay of the tails of the stationary maximal weighted delay. In Andrews *et al.* (2004), a throughput optimal scheduling policy, modified-largest weighted delay first, is studied that provides minimum rate guarantees and probabilistic delay bounds on arrivals. In Shakkotai and Stolyar (2000), an exponential scheduling rule in conjunction with token based rate control for supporting a mixture of real-time and non-real time data applications is proposed. The exponential scheduling rule can implement minimum rate guarantees and provide probabilistic delay guarantees through appropriate choice of parameters. In Wu and Negi (2005) and in Shakkottai (2008), effective capacity of a wireless channel with a guaranteed QoS is studied, where QoS is based on the buffer overflow probabilities and on queue delay violation probabilities. In our work, the performance metric of our interest is the long time average user throughput and we study fairness notions based only on the average throughput vector.

Early works on fair schedulers for wireless networks such as Ng *et al.* (1998), Lu *et al.* (1999) and Nandagopal *et al.* (1999), were based on generalized processor sharing (GPS) and packetized weighted fair queueing (WFQ) principles. In Ng *et al.* (1998), the authors identify key properties for a packet fair queueing scheduler in a wireless environment and present an algorithm based on the properties. A wireless fair scheduler based on fluid fair queueing is proposed in Lu *et al.* (1999) to handle location-dependent error burst. The above schedulers based on GPS report bounded delay or fairness guarantees and their performances are essentially suboptimal. In Liu *et al.* (2003*b*), a wireless credit based fair queueing scheduler is proposed that achieves a tradeoff between fairness of temporal access independent of channel conditions and pure opportunistic scheduling. We note that the set of throughput objectives achievable with the approach is limited.

A variety of mathematical tools are available for the development and analysis of resource allocation algorithms for wireless networks. The stochastic approximation method (see Kushner and Yin (1997)), is a popular class of iterative algorithms for finding the zero of a random, unknown function. Stochastic approximation techniques are widely used in scheduling algorithms for wireless networks (see e.g., Kushner and Whiting (2002), Liu *et al.* (2003*a*)). The ODE method reported in Borkar and Meyn (2000) is a popular technique used to study the asymptotic behaviour and stability of stochastic approximation algorithms. The fluid-limit method proposed in Rybko and Stolyar (1992), helps determine limiting behaviours of a random process by studying the limiting behaviour of appropriate scaled versions of the process that are deterministic. In Stolyar (2005*b*), fluid sample paths were used to study the transient dynamics of user throughputs for gradient schedulers. Lyapunov stability theory is used in many works, including Tassiulas and Ephremides (1992), Tassiulas and Ephremides (1993) and Neely *et al.* (2005*b*), to develop stable scheduling algorithms for wireless networks with unsaturated traffic.

Sample average approximation (see Kleywegt *et al.* (2002)) is a useful technique for solving discrete stochastic optimization problems of the type $\min_{x \in \{x_1, x_2, \ldots\}} E(f(x, Y))$, where $f(\cdot, Y)$ is a random, unknown function. Sample path optimization techniques (see Robinson (1996)) can optimize the steady-state or long-time averages of dynamic systems by solving related optimization problems on finite length sample-averages. The basic principle behind our work is in spirit with sample path optimization, namely, that the optimizer(s) x of the almost sure limit, f_∞, of certain computable sequence of random functions $\{f_n\}$ can be estimated through deterministic optimization of f_n for the observed sample path. In our work, we identify the asymptotic optimal rate vector for the wireless system through the deterministic optimization of the network utility over the observed channel distribution sequence. Further, we develop control strategies that implement the optimizing rate vector in the long time average sense.

1.3 Outline

In Chapter 2, we describe the system model and assumptions. We present the rate region based scheduler RRS in Chapter 3. In Chapter 4, we propose schedule history and channel history based schedulers, RRS-CA and RRS-RA, and discuss their advantages. We report simulation results and evaluate the performance of rate region based schedulers, RRS and RRS-RA, in Chapter 5. In Chapter 6, we discuss simple implementations of rate region based max-min fair schedulers for two users and for N users. We conclude the thesis and discuss future work in Chapter 7. In Appendices A and B, we discuss the proof of the asymptotic optimality of RRS and RRS-CA schedulers. In Appendix C we discuss the rate region of ergodic channels with stationary distribution. In Appendix D, we present a simple relation between the geometry of the rate region and an important parameter of an RRS-RA scheduler. Finally, in Appendix E, we list the channel models used in our simulations.

CHAPTER 2

System Model and Assumptions

We consider the downlink wireless channel of a single cell of a cellular data network. A fixed number (N) of users, share the slotted wireless channel using time division multiplexing; a single user is scheduled in every slot and users time share the channel over slots. We assume that the user queues at the base station are saturated. In this set up, a centralized scheduler at the base station attempts to implement a network QoS or a notion of fairness.

2.1 Channel Model

We consider a random and time varying wireless channel, $\{R(t)\}_{t=1}^{\infty}$, where $R(t) = (r_1(t), \cdots, r_N(t))$ for all t. The vector $(r_1(t), \cdots, r_N(t))$ corresponds to the set of achievable rates in the slot t for any user, i.e., if user i is scheduled in the slot t, rate $r_i(t)$ is achieved for the user i in the slot without any data error or data loss. The channel state $R(t)$ is assumed to be fixed in the slot t and may vary over slots; the channel process $\{R(t)\}_{t=1}^{\infty}$ may possibly be correlated over time and across users. We assume that the channel process $\{R(t)\}_{t=1}^{\infty}$ is an ergodic random process with a discrete stationary distribution (e.g., an irreducible discrete time Markov chain). Let $R(t) \in \{\mathbf{R}_1, \mathbf{R}_2, \cdots\}$ for all t, where $\mathbf{R}_j = (r_{j,1}, \cdots, r_{j,N})$, and, let $\{\pi_j\}$ be the stationary distribution of $\{R(t)\}_{t=1}^{\infty}$ for the sample space $\{\mathbf{R}_j\}$. We will assume that $r_{j,i} < B < \infty$ for all j and all i. In this work, we assume that the channel state information $R(t)$ is available at the base station at the beginning of every slot t, and, the base station uses $R(t)$ to decide upon a schedule in the slot.

2.2 Performance Metric

Our performance metric for the saturated network is the long time average user throughput. Let $\{\mu(t)\}$ be any schedule, where $\mu(t) \in \{1, 2, \cdots, N\}$ for all t, and $\mu(t') = i'$ indicates that time-slot t' is allocated to user i' under schedule $\{\mu(t)\}$. Then, we define the long time average throughput $\bar{R}$ for the schedule $\{\mu(t)\}$ as $\bar{R}(\mu) = (\bar{r}_1(\mu), \cdots, \bar{r}_N(\mu))$, where,

$$\bar{r}_i(\mu) := \lim_{t\to\infty}\inf \bar{r}_i(\mu, t) = \lim_{t\to\infty}\inf \frac{1}{t}\sum_{s=1}^{t} r_i(s) I_{\{\mu(s)=i\}}$$

$\bar{r}_i(\mu, t)$ is the time average throughput of user i up to time t for the schedule $\{\mu(t)\}$. We define the rate region $\mathcal{C}$ of the wireless channel as the set of all long time average throughput vectors feasible with probability 1. For an ergodic channel process, the stationary distribution of the process characterizes $\mathcal{C}$ completely (see Liu *et al.* (2003*a*), Kumar *et al.* (2008)). The rate region for an ergodic wireless channel with sample space $\{\mathbf{R}_j\}$ and stationary probability distribution $\{\pi_j\}$ is given by (see Appendix C)

$$\mathcal{C} = \left\{ (\bar{r}_1, \cdots, \bar{r}_N) : \bar{r}_i = \sum_j \pi_j \, a_{j,i} \, r_{j,i}, \;\; a_{j,i} \geq 0, \sum_i a_{j,i} \leq 1, i = 1, \cdots, N \right\}$$

$\{a_{j,i}\}$s represent a stationary schedule or an average channel allocation. For example, the long time average throughput of $\sum_j \pi_j a_{j,i} r_{j,i}$ is achievable for user i with a stationary random schedule $\{a_{j,i}\}$, i.e., when we allocate channel $\mathbf{R}_j$ randomly and independently to users according to the probability distribution $(a_{j,1}, \cdots, a_{j,N})$, whenever channel $\mathbf{R}_j$ occurs. The stationary random schedule $\{a_{j,i}\}$ achieves an average channel allocation of $\{a_{j,i}\}$ in the long time average sense, i.e., a fraction $a_{j,i}$ of slots that are $\mathbf{R}_j$, is allocated to user i as time tends to infinity. The average channel allocation $a_{j,i}$ ensures the average rate $\sum_j \pi_j a_{j,i} r_{j,i}$ for an ergodic channel with probability distribution $\{\pi_j\}$. We note that every rate vector $\bar{R} = (\bar{r}_1, \cdots, \bar{r}_N)$ in the rate region $\mathcal{C}$ is achievable using a stationary random schedule. Also, we note that the rate region $\mathcal{C}$ is a convex set (see Liu *et al.* (2003*a*)).

2.3 Network Objective

In this setup, our network objective is to implement a QoS or a notion of fairness or maximize a network utility, defined on the rate region $\mathcal{C}$. For example, let $U : \mathcal{R}^N \to \mathcal{R}$ be a continuous network utility defined on the rate region $\mathcal{C}$. Then, the base station would seek to achieve (almost surely) a long time average throughput vector $\bar{R}^*$ such that

$$\bar{R}^* \in \arg\max_{\bar{R} \in \mathcal{C}} U(\bar{R})$$

$\mathcal{C}$ is a compact set (closed and bounded since $r_{j,i} < B < \infty$ for all j and i). The maximum of a continuous function is achievable in a compact set (see Rudin (1976)).

Remarks 2.3.1.

1. A number of fairness definitions can be implemented using network utilities (e.g., sum of logarithms on the user throughput achieves proportional fairness, see Kushner and Whiting (2002)). However, some network QoS such as minimum rate constraints, notions of fairness such as max-min fairness are best implemented without defining a network utility. In such cases, the scheduler would seek a long time average rate vector, $\bar{R}^*$, which is a preferred operating point in the rate region $\mathcal{C}$. Please refer to Chapter 5 for interesting definitions of QoS and their implementations.
2. For strictly concave network utilities, the objective is to maximize the concave function over a convex set. A unique maximum exists in such cases and gradient algorithms can be used to achieve the optimal throughput vector (see Kushner and Whiting (2004) and Stolyar (2005*b*)).

2.4 Simulation Setup

In Appendix E, we have listed the channel models used in our simulations. We have considered two basic types of ergodic channels in our simulations, viz., (i) user-correlated and time independent channel and (ii) time-correlated and user independent channel. We have considered four channel models, A,B,C and D in this thesis. Channel models A and C (user correlated and time independent) describe the sample space of $\{R(t)\}$ and the corresponding distribution directly; the sample space and the distribution were chosen arbitrarily. Channel models B and D (time correlated and user independent) model $\{R(t)\}$ through an underlying fast-fading wireless channel generated using the Jake's model for Rayleigh fading. (All the relevant parameters are reported in the Appendix E). We do not model shadowing and related slow-fading effects. The parameters of the Jake's model, like mean SNR, Doppler shift, etc., have been chosen arbitrarily. All the channel models have a finite state space and a well defined stationary distribution.

All the simulations reported in this thesis correspond to the performance of the schedulers for a realization of the wireless channel. We expect that all the rate region based schedulers are asymptotically optimal for almost all channel realizations (in Appendix A and in Appendix B, we have proved the almost sure optimality of RRS and RRS-CA scheduler). Hence, we use the simulation plots only to illustrate the performance of the schedulers for different network utilities and QoS. Plots that compare the performance of different schedulers use the same channel realization for all the different schedulers.

All the simulations were performed on MATLAB. Channel models B and D were implemented using the Rayleigh-fading tool of MATLAB.

2.5 Notations

N	Number of users
$R(t)$	Vector channel rates (CSI) at time t
$r_i(t)$	channel rate of user i (CSI) at time t
$\{R(t)\}_{t=1}^{\infty}$	Channel/Rate process
$r_{j,i}$	channel rate of user i, at channel state j
$\{\mathbf{R}_1, \mathbf{R}_2, \ldots\}$	Sample space of the wireless channel
$\{\pi_1, \pi_2, \ldots\}$	Stationary distribution of the wireless channel
$\mathcal{C}$	Rate region of the wireless channel
U	Network utility function
U_i	User utility function
$\pi(t)$	Estimate of the vector channel distribution at time t
$\{\pi_j(t)\}$	Estimate of the channel vector distribution at time t
$\mathcal{C}(t)$	Estimated rate region at time t
$\{\mu(t)\}$	Arbitrary schedule sequence
$\{a_{j,i}\}$	Stationary schedule
$\{\hat{a}_{j,i}(t)\}$	Channel allocation up to time t
$\bar{R}(t)$	Time average throughput (allocated rate) vector at time t
$\bar{R}(\mu, t)$	Time average throughput (allocated rate) vector for schedule μ at time t
$\bar{R}$	Long time average throughput vector

$\bar{R}(\mu)$	Long time average throughput vector for schedule μ
$\bar{r}_i(t)$	Time average throughput (allocated rate) of user i at time t
$\bar{r}_i(\mu, t)$	Time average throughput (allocated rate) of user i for schedule μ at time t
$\bar{r}_i$	Long time average throughput of user i
$\bar{r}_i(\mu)$	Long time average throughput of user i for schedule μ
$\{a^*_{j,i}\}$	Optimal stationary schedule
$\bar{R}^*(t)$	Optimal time average rate vector at time t in $\mathcal{C}(t)$
$\bar{R}^*$	Optimal long time average rate vector in $\mathcal{C}$

CHAPTER 3

A Rate Region Based Scheduler

The rate region $\mathcal{C}$ of an ergodic channel $\{R(t)\}_{t=1}^{\infty}$ with sample space $\{\mathbf{R}_j\}$ and stationary distribution $\{\pi_j\}$ is given by (see Chapter 2)

$$\mathcal{C} = \left\{ (\bar{r}_1, \cdots, \bar{r}_N) : \bar{r}_i = \sum_j \pi_j \, a_{j,i} \, r_{j,i}, \;\; a_{j,i} \geq 0, \sum_i a_{j,i} \leq 1, i = 1, \cdots, N \right\}$$

$\{a_{j,i}\}$s represent a stationary schedule or an average channel allocation. We note that every $\bar{R} \in \mathcal{C}$ is achievable using a stationary random schedule. Also, $\mathcal{C}$ is a convex and compact subset of $\mathcal{R}^N$.

Let $U : \mathcal{R}^N \rightarrow \mathcal{R}^1$ be a continuous network utility defined on the rate region $\mathcal{C}$. Our network objective is to seek a long time average throughput vector $\bar{R}^*$ such that

$$\bar{R}^* \in \arg\max_{\bar{R} \in \mathcal{C}} U(\bar{R})$$

$\mathcal{C}$ is a compact set and the maximum of a continuous function is achievable in a compact set (see Rudin (1976)). Hence, there exists an $\bar{R}^*$ in $\mathcal{C}$. In this work, we will assume that there exists a unique $\bar{R}^*$ that maximizes the network utility U over the rate region $\mathcal{C}$. (For a strictly concave network utility U, there exists a unique $\bar{R}^*$ that maximizes U in $\mathcal{C}$). $\bar{R}^*$ is the preferred operating point for the given network utility U and the channel process with stationary distribution $\{\pi_j\}$.

Suppose that the stationary distribution of the wireless channel $\{\pi_j\}$ is known apriori. Then, we can identify the rate region $\mathcal{C}$ and the optimal rate vector $\bar{R}^* \in \mathcal{C}$ that maximizes the network utility U. Let $\{a^*_{j,i}\}$ be a stationary schedule corresponding to the optimal rate vector $\bar{R}^*$, i.e.,

$$\bar{r}^*_i = \sum_j \pi_j \, a^*_{j,i} \, r_{j,i}$$

for all i. The time average throughput vector $\bar{R}^*$ is now achievable (almost surely) using the stationary random schedule $\{a^*_{j,i}\}$; we allocate channel $\mathbf{R}_j$ randomly and independently to users according to the probability distribution $(a^*_{j,1}, \cdots, a^*_{j,N})$ whenever channel $\mathbf{R}_j$ occurs (see Stolyar (2005*b*) for a discussion on a similar schedule called the static service split (SSS) schedule).

In general, the stationary distribution $\{\pi_j\}$ is unknown and hence, we would not know the optimal rate vector $\bar{R}^*$ or the optimal stationary schedule $\{a^*_{j,i}\}$.

3.1 Finite-Time Rate Region

Suppose that the stationary distribution $\{\pi_j\}$ of the wireless channel $\{R(t)\}_{t=1}^{\infty}$ is not known. We will now estimate the vector channel distribution using the available channel history $\{R(u) : 1 \leq u \leq t\}$ as follows. At time t, define $\{\pi_j(t)\}$ as

$$\pi_j(t) := \frac{1}{t}\sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}} \tag{3.1}$$

for all j. The random variable $\pi_j(t)$ is the observed fraction of channel slots that are $\mathbf{R}_j$; $\pi_j(t) \geq 0$ for all j and $\sum_j \pi_j(t) = 1$ for all t. Note that, $\{\pi_j(t)\}$ is a valid probability distribution on the channel space $\{\mathbf{R}_j\}$. For an ergodic channel process $\{R(t)\}_{t=1}^{\infty}$ with stationary distribution $\{\pi_j\}$,

$$\lim_{t\to\infty} \pi_j(t) = \pi_j \text{ a.s.}$$

for all j, i.e., the observed channel distribution converges to the stationary distribution of the wireless channel almost surely (see Wolff (1989)).

Now for every time t, we define a *finite-time* estimated rate region $\mathcal{C}(t)$, evaluated using the observed channel distribution $\{\pi_j(t)\}$ as,

$$\mathcal{C}(t) := \left\{(\bar{r}_1, \cdots, \bar{r}_N) : \bar{r}_j = \sum_j \pi_j(t)\, a_{j,i}\, r_{j,i},\ \ a_{j,i} \geq 0, \sum_i a_{j,i} \leq 1, i = 1, \cdots, N\right\} \tag{3.2}$$

$\mathcal{C}(t)$ corresponds to the set of all time average rates, $\bar{R}$, achievable over $[1, t]$, for the realization of the wireless channel ($\{R(u) : 1 \leq u \leq t\}$), possibly using non-causal schedules (a schedule based on the complete knowledge of future channel states) and fractional channel allocation (more than one user is scheduled in a slot and users time share the slot). Observe that $\mathcal{C}(t)$ is defined using $\{\pi_j(t)\}$ exactly as $\mathcal{C}$ is defined using

π; hence, $\mathcal{C}(t)$ is also the rate region for a wireless channel with stationary distribution $\{\pi_j(t)\}$. Unlike $\mathcal{C}$, this finite-time rate-region $\mathcal{C}(t)$ is a random set and depends on the sample-path (of the wireless channel) that is observed. When $\lim_{t\to\infty} \pi(t) = \pi$ a.s., we have,

$$\lim_{t\to\infty} \mathcal{C}(t) = \mathcal{C} \text{ a.s.} \tag{3.3}$$

i.e., the finite time estimated rate region of the wireless channel converges to the rate region of the wireless channel almost surely[1]. Hence, we will use $\{\pi_j(t)\}$ as our estimate of the channel distribution at time t and $\mathcal{C}(t)$ as our estimate of the rate region of the wireless channel. We will now proceed with identifying an optimal rate vector and a schedule for the network utility U in the estimated rate region $\mathcal{C}(t)$.

3.2 An Optimal Non-Causal Scheduler

We will now define an optimal (possibly) non-causal scheduler for the network utility U and the wireless channel $\{R(t)\}_{t=1}^{\infty}$, using the finite-time estimated rate-region $\mathcal{C}(t)$ as follows. Let $\bar{R}^*(t)$ be an optimal rate vector for the network utility U in the finite time rate region $\mathcal{C}(t)$, i.e.,

$$\bar{R}^*(t) \in \arg \max_{\bar{R}\in\mathcal{C}(t)} U(\bar{R}) \tag{3.4}$$

$\mathcal{C}(t)$ is a compact set in $\mathcal{R}^N$ and the maximum is achievable for a continuous function U in the compact set $\mathcal{C}(t)$. Hence, there exists a $\bar{R}^*(t) \in \mathcal{C}(t)$[2]. Let $\{a^*_{j,i}(t)\}$ be a channel allocation (or a stationary schedule) that achieves $\bar{R}^*(t)$ over $[1, t]$, i.e.,

$$\bar{r}^*_i(t) = \sum_j \pi_j(t)\, a^*_{j,i}(t)\, r_{j,i} = \sum_j \frac{1}{t} \sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}} a^*_{j,i}(t)\, r_{j,i} \tag{3.5}$$

for all i. A time average throughput of $\bar{R}^*(t)$ is achievable over $[1, t]$, if the actual channel allocation up to time t matches $\{a^*_{j,i}(t)\}$, i.e., the fraction of channel state $\mathbf{R}_j$ that was scheduled to user i, up to time t, must equal $a^*_{j,i}(t)$. In general, $\bar{R}^*(t)$ is not a feasible rate vector with causal schedulers (the sequence of schedules $\{\{a^*_{j,i}(s)\}$:

[1] Define $\mathcal{C}_{inf} := \liminf_{t\to\infty} C(t) := \{\bar{R} \in R^N : \limsup_{t\to\infty} \inf_{\bar{S}\in C(t)}\{\| \bar{R} - \bar{S} \|\} = 0\}$ and $\mathcal{C}_{sup} = \limsup_{t\to\infty} C(t) := \{\bar{R} \in R^N : \liminf_{t\to\infty} \inf_{\bar{S}\in C(t)}\{\| \bar{R} - \bar{S} \|\} = 0\}$. If $\pi(t) \to \pi$ a.s., then $\mathcal{C}_{inf} = \mathcal{C} = \mathcal{C}_{sup}$ a.s., i.e., the rate region $\mathcal{C}(t)$ converges to $\mathcal{C}$ in the Kuratowski sense, a.s. (Kuratowski (1966)).

[2] We have assumed that there exists a unique $\bar{R}^* \in \mathcal{C}$, but that does not restrict the possibility of multiple optimal rate vectors in $\mathcal{C}(t)$.

$1 \leq s \leq t\}$ is random, and the past allocations cannot be undone) and can be achieved only using a non-causal scheduler and possibly with fractional allocation of channels. Trivially, and by definition, the utility achieved by $\bar{R}^*(t)$ serves as an upper-bound for any schedule, for all time t and for every realization of the wireless channel. We formalize this in the following lemma.

Lemma 3.2.1. *For every realization of the wireless channel, the utility $U(\bar{R}^*(t))$ achieved through a (possibly) non-causal and fractional channel allocation, satisfying equations (3.4) and (3.5), is the maximum achievable utility for the wireless channel, at the time t, for the realization.* □

Remarks 3.2.1.

1. As mentioned earlier, $\bar{R}^*(t)$ may not be a feasible rate vector over $[1, t]$ using causal schedulers. However, $\bar{R}^*(t)$ is an upper-bound on the performance of all schedulers (causal and non-causal) and the utility at $\bar{R}^*(t)$ provides us a benchmark to compare the performance of other schedulers.
2. The non-causal scheduler is optimal for all t and therefore it is asymptotically optimal as well.
3. Any scheduler which can achieve $\bar{R}^*(t)$ asymptotically is an asymptotically optimal scheduler.

3.3 An Asymptotically Optimal Causal Scheduler : RRS

We will now propose the rate region based scheduler, RRS, using the finite time rate region $\mathcal{C}(t)$ and the optimal non-causal channel allocation $\{a^*_{j,i}(t)\}$ (that achieves $\bar{R}^*(t)$ over $[1, t]$) .

1. In every slot t, estimate the channel distribution $\{\pi_j(t)\}$ as

$$\pi_j(t) = \frac{1}{t} \sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}}$$

2. Compute the finite time rate region $\mathcal{C}(t)$ using $\{\pi_j(t)\}$ as

$$\mathcal{C}(t) := \left\{ (\bar{r}_1, \cdots, \bar{r}_N) : \bar{r}_i = \sum_j \pi_j(t)\, a_{j,i}\, r_{j,i},\ \ a_{j,i} \geq 0, \sum_i a_{j,i} \leq 1, i = 1, \cdots, N \right\}$$

3. Identify an optimal rate vector, $\bar{R}^*(t)$, in the estimated rate region $\mathcal{C}(t)$, where

$$\bar{R}^*(t) \in \arg \max_{\bar{R} \in \mathcal{C}(t)} U(\bar{R})$$

4. Identify an optimal channel allocation/stationary schedule $\{a^*_{j,i}(t)\}$ for $\bar{R}^*(t)$, such that

$$\bar{r}^*_i(t) = \sum_j \pi_j(t)\, a^*_{j,i}(t)\, r_{j,i}$$

5. In the slot t, implement the stationary random schedule $\{a^*_{j,i}(t)\}$, i.e., schedule channel $R(t)$ randomly and independently to a user according to the probability distribution $(a_{R(t),1}, \cdots, a_{R(t),N})$, independent of the past schedules.

In every slot t, the RRS scheduler implements the schedule $\{a^*_{j,i}(t)\}$ (the allocation that achieves $\bar{R}^*(t)$ over $[1, t]$, described in equations (3.4) and (3.5)) only for the current slot t and for the channel state $R(t)$. The following theorem proves the asymptotic optimality of RRS.

Theorem 3.3.1. *Let $U : \mathcal{R}^N \to \mathcal{R}^1$ be a continuous network utility. Let $\{R(t)\}_{t=1}^{\infty}$ be a bounded, discrete-time, irreducible, finite state, space Markov chain with stationary distribution $\{\pi_j\}$ and rate region $\mathcal{C}$. If the optimal rate vector $\bar{R}^* = \arg\max_{\bar{R}\in\mathcal{C}} U(\bar{R})$ and the optimal stationary schedule $\{a^*_{j,i}\}$ are unique, then the time average throughput achieved using RRS converges to the optimal rate vector $\bar{R}^*$ almost surely.*

$$\bar{R}(t) \to \bar{R}^* \text{ a.s.}$$

i.e., RRS is asymptotically optimal.

Proof : See Appendix A. □

Remarks 3.3.1.

1. To prove the asymptotic optimality of RRS, we require that the schedule process $\{\{a^*_{j,i}(t)\}\}_{t=1}^{\infty}$ converge to an optimal schedule $\{a^*_{j,i}\}$. We show in Appendix A that the convergence occurs whenever the optimal stationary schedule $\{a^*_{j,i}\}$ is unique. However, in general, the optimal schedule need not be unique, even when the optimal rate vector $\bar{R}^*$ is unique. We illustrate this using the following example. Consider a wireless system with 2 users. Let $\{R(t)\}_{t=1}^{\infty}$ be an i.i.d. channel process with state space $\{(100, 200), (200, 400)\}$ KBps and stationary distribution $\{0.5, 0.5\}$. Let $U = \sqrt{\bar{r}_1} + \sqrt{\bar{r}_2}$ be the continuous network utility that we seek to maximize. The unique optimal rate vector for the wireless channel is $(50, 200)$ KBps. The rate vector is achievable using the stationary schedules

$$\{a^*_{j,i}\} = \{(1, 0), (0, 1)\}$$

as well as

$$\{b^*_{j,i}\} = \{(0,1), (0.5, 0.5)\}$$

In fact, any $\lambda_1\{a^*_{j,i}\} + \lambda_2\{b^*_{j,i}\}$, where $\lambda_1, \lambda_2 \in [0,1]$ and $\lambda_1 + \lambda_2 = 1$ is an optimal stationary schedule for the system.

2. RRS requires consistency in the schedules implemented across time-slots: *Implementing any $\{a^*_{j,i}(t)\}$ that satisfies Equation 3.5 may not result in asymptotically optimal performance*. For instance, consider the previous example with multiple optimal schedules. Let schedule $\{a^*_{j,i}\}$ be implemented whenever channel state $(100, 200)$ KBps occurs and let $\{b^*_{j,i}\}$ be implemented whenever channel state $(200, 400)$ KBps occurs. This will result in a long-time average throughput of $(150, 100)$ KBps which is not utility optimal.

3. While we have assumed an unique optimal schedule, $\{a^*_{j,i}\}$, it can be seen from the proof of Theorem 3.3.1, that the convergence of the schedule sequence used under RRS, $\{\{a^*_{j,i}(t)\}\}$, to *an* optimal schedule is sufficient to ensure asymptotic optimality for the network utility. In the next Chapter 4, we propose a practical variant of RRS called RRS-RA that achieves the optimal rate vector $\bar{R}^*$ consistently, without any restrictions such as the uniqueness of the optimal schedule $\{a^*_{j,i}\}$ (used in the proof of Theorem 3.3.1).

4. The restriction on the uniqueness of the optimal schedule can easily be relaxed and a network utility of $U(\bar{R}^*) - \epsilon$ can be achieved with RRS, for any $\epsilon > 0$. The following scheduler is ϵ-optimal. Let $\{\tilde{a}_{j,i}(t)\}$ be the stationary schedule implemented in slot t and let $R(\{\tilde{a}_{j,i}(t)\}, \pi(t+1))$ be the ergodic rate for schedule $\{\tilde{a}_{j,i}(t)\}$ and distribution $\pi(t+1)$. Let $\{a^*_{j,i}(t+1)\}$ be any RRS schedule for slot $t+1$. In slot $t+1$, implement $\{a^*_{j,i}(t+1)\}$ if $\mid U(\bar{R}^*(t+1)) - U(R(\{\tilde{a}_{j,i}(t)\}, \pi(t+1))) \mid > \epsilon$. Else, continue to implement $\{\tilde{a}_{j,i}(t)\}$. Since $\bar{R}^*(t) \to \bar{R}^*$ for any sequence of RRS schedules, and consequently $U(\bar{R}^*(t)) \to U(\bar{R}^*)$, the described scheduler will have a convergent schedule sequence and will be $\epsilon-$optimal, for all continuous U.

3.4 Simulations

In Theorem 3.3.1, we proved that the rate region based scheduler RRS is asymptotically optimal for all continuous network utilities, for some ergodic channels with discrete distribution. In this section, we will present the performance of RRS scheduler for two popular notions of fairness, proportional fairness and max-min fairness. In Chapter 4 and in Chapter 5, we will present additional results and report comparison plots with RRS-RA and other schedulers from the literature.

Proportional fair rates are achieved using the sum of logarithms on the time average user throughput as the network utility; we note that the network utility is not continuous at the origin. In Figure 3.1, we report the performance of the implementation of proportional fair scheduler using RRS for a system with 4 users and for the wireless channel model A (described in Appendix E). The thin straight lines in Figure 3.1 mark the optimal rate vector $\bar{R}^*$ that achieves proportional fairness. From the figure, we observe that $\bar{R}(t)$ converges to $\bar{R}^*$ as t goes to infinity. The sum of logarithmic network utility is continuous in the positive quadrant ($\mathcal{R}^{+N} \backslash 0^N$) which ensures that RRS is asymptotically optimal.

In Figure 3.2, we plot the performance of the implementation of max-min fair scheduler using RRS for 3 users and for the channel model B (described in Appendix E). Max-min fairness is implemented by seeking a Pareto optimal $\bar{R}^*(t) \in \mathcal{C}(t)$ such that $\bar{r}_1^*(t) = \cdots = \bar{r}_N^*(t)$, without using a network utility. The thin straight line in the Figure 3.2 marks the optimal rate for the channel distribution. Here again, we observe from the plot that RRS achieves the optimal performance for max-min fairness asymptotically. For max-min fairness, there exists a unique $\bar{R}^*(t)$ in $\mathcal{C}(t)$, which converges to $\bar{R}^*$ (in $\mathcal{C}$) almost surely (for the ergodic wireless channel). Hence, RRS achieves the optimal rate vector for max-min fairness as well. Note that in literature, different utilities/QoS require different forms of schedulers, whereas RRS provides us a common framework to implement arbitrary QoS and network utilities.

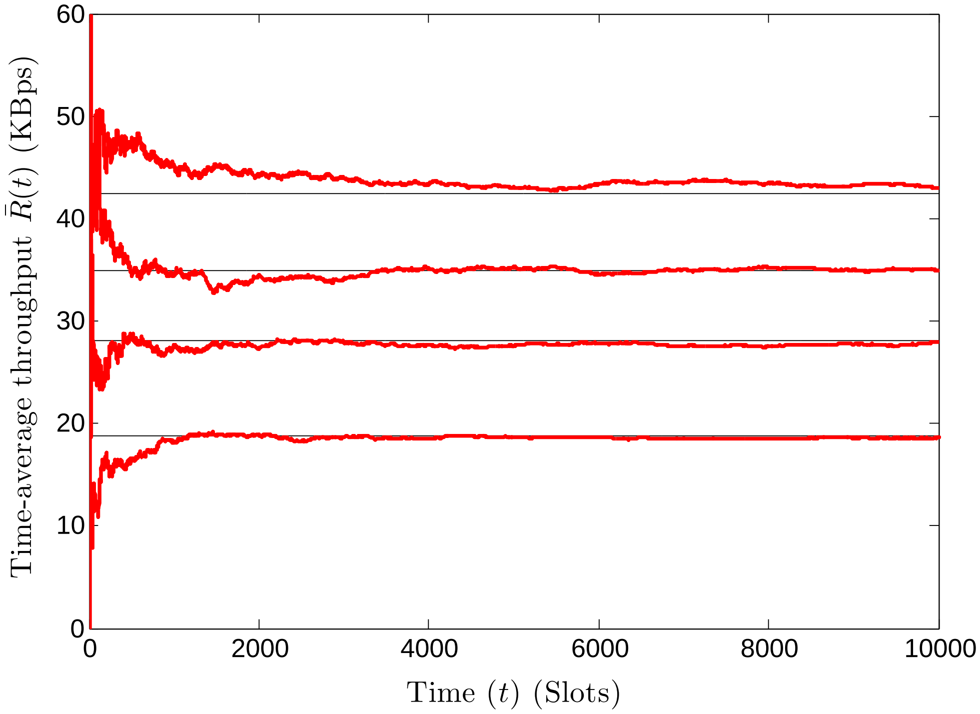

Figure 3.1: Plots of the time average user throughputs of RRS scheduler implementing proportional fairness. We consider a wireless system with $N = 4$ users and the channel corresponds to the channel model A described in Appendix E. The thin straight lines correspond to the optimal rate vector $\bar{R}^*$ for the wireless channel.

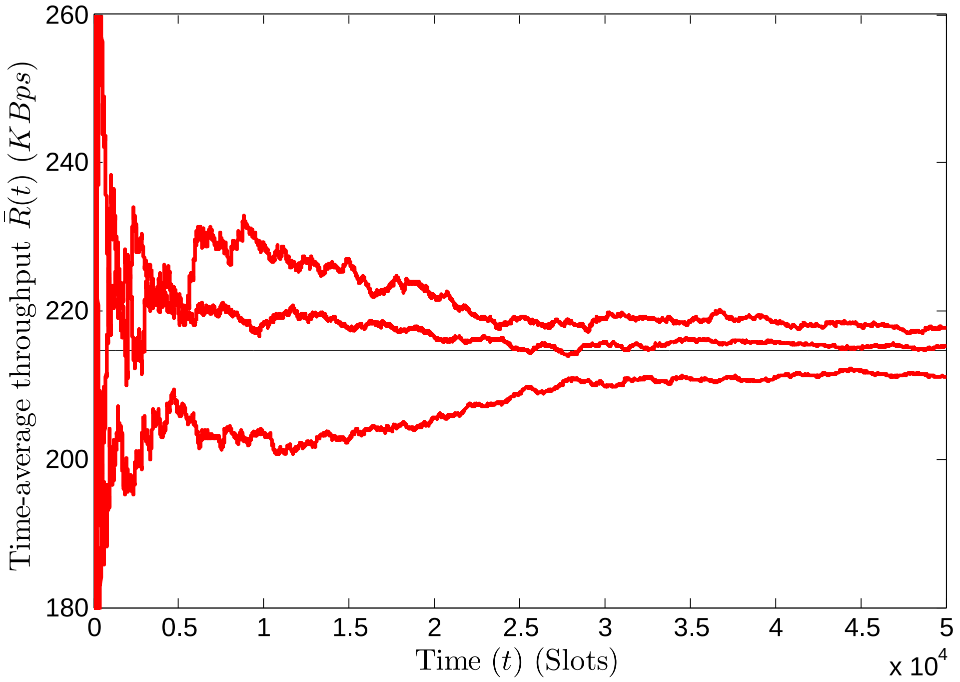

Figure 3.2: Plots of the time average user throughputs of RRS scheduler implementing max-min fairness. We consider a wireless system with $N = 3$ users and the channel corresponds to the channel model B described in Appendix E. The thin straight line corresponds to the optimal rate $\bar{R}^*$ for the wireless channel.

CHAPTER 4

Channel History and Allocation History based Scheduling

The rate region based scheduler, RRS, presented in Chapter 3, uses the available channel history to identify a schedule, but does not make use of any available information regarding schedule history and scheduled (allocated) rates. The RRS schedule at time instant t, $\{a^*_{R(t),i}(t)\}$, is a function only of the observed channel distribution $\{\pi_j(t)\}$ and the network utility U; the schedule $\{a^*_{R(t),i}(t)\}$ is chosen independent of the previous schedules $\{\{a^*_{R(s),i}(s)\} : 1 \leq s \leq t-1\}$.

The RRS scheduler seeks to achieve the time average throughput $\bar{R}^*(t)$ (and eventually $\bar{R}^*$) by using the schedule $\{a^*_{R(t),i}(t)\}$ in slot t. We note that the time average throughput $\bar{R}^*(t)$ can be achieved over $[1,t]$, only if the channel allocation up to time t is $\{a^*_{j,i}(t)\}$ for all j and i. As the schedule $\{a^*_{R(t),i}(t)\}$ is chosen independent of the previous schedules $\{\{a^*_{R(s),i}(s)\} : 1 \leq s \leq t-1\}$, the actual channel allocation up to time t with the sequence of schedules $\{\{a^*_{R(s),i}(s)\} : 1 \leq s \leq t\}$ may be very different from $\{a^*_{j,i}(t)\}$. Hence, to ensure asymptotic optimality, RRS requires that the schedules $\{a^*_{R(t),i}(t)\}$ be such that they form a consistent (convergent) sequence. In the proof of Theorem 3.3.1, we showed that any sequence of optimal schedules $\{a^*_{j,i}(t)\}$ would converge if there exists a unique optimal stationary schedule $\{a^*_{j,i}\}$. The convergent sequence of stationary schedules would then ensure that the allocated rate vector $\bar{R}(t)$ converges to $\bar{R}^*$.

In this chapter, we propose scheduling strategies that seek to achieve a time average throughput of $\bar{R}^*(t)$ over $[1,t]$ by using schedule history and scheduled rates as well. As we focus on a causal implementation at every time t (also, we cannot alter the allocated rates for time instants $1 \leq s \leq t-1$), we might expect to improve upon the performance of RRS by opting for a schedule that best implements the optimal rate vector $\bar{R}^*(t)$ over $[1,t]$ than $\{a^*_{R(t),i}(t)\}$. Of course, we also require that such a schedule (with a myopic objective of implementing $\bar{R}^*(t)$ better in every slot t) be asymptotically optimal as well. Using simulations, we will show that RRS can have poor convergence behaviour, and we can improve the performance of RRS by using schedule history and allocated rate information.

In Section 4.1, we propose a simple channel history and schedule history based strategy that seeks to achieve a time average throughput of $\bar{R}^*(t)$ over $[1,t]$, by minimizing a distance between the actual channel allocation and the optimal schedule $\{a^*_{j,i}(t)\}$. In Section 4.2, we propose a gradient algorithm based scheduling strategy that seeks to maximize a dynamic, auxiliary network utility in every slot t. The dynamic, auxiliary network utility is chosen such that the rate vector $\bar{R}^*(t)$ is the unique maximizer for the auxiliary network utility in the rate region $\mathcal{C}(t)$. The gradient algorithm would then drive the allocated rate vector $\bar{R}(t)$ towards the optimal rate vector $\bar{R}^*(t)$, and asymptotically to $\bar{R}^*$, under appropriate conditions.

4.1 A Channel Allocation based Scheduler: RRS-CA

Let $\{\mu(t)\}$ be any schedule and let $\{\hat{a}_{j,i}(\mu,t)\}$ be the actual channel allocation, where $\hat{a}_{j,i}(\mu,t)$ is the fraction of channel state $\mathbf{R}_j$ allocated to user i up to time t for the schedule $\{\mu(t)\}$, i.e.,

$$\hat{a}_{j,i}(\mu,t) = \frac{\sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}} I_{\{\mu(s)=i\}}}{\sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}}}$$

We note that $\hat{a}_{j,i}(\mu,t) \geq 0$ and $\sum_i \hat{a}_{j,i}(\mu,t) \leq 1$ for all j, i, μ and t. Hence, the actual channel allocation $\{\hat{a}_{j,i}(\mu,t)\}$ is a valid probability distribution on the schedule space and there exists a stationary schedule $\{\hat{a}_{j,i}(\mu,t)\}$ with identical probabilities. Now, the allocated rate vector, $\bar{R}(\mu,t)$, for the schedule $\{\mu(t)\}$ is given by

$$\begin{aligned}
\bar{r}_i(\mu,t) &= \frac{1}{t}\sum_{s=1}^{t}\sum_{j} I_{\{R(s)=\mathbf{R}_j\}} I_{\{\mu(s)=i\}} r_{j,i} \\
&= \sum_{j}\frac{1}{t}\sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}} I_{\{\mu(s)=i\}} r_{j,i} \\
&= \sum_{j}\frac{1}{t}\sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}} I_{\{\mu(s)=i\}} \frac{\sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}}}{\sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}}} r_{j,i}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_j \frac{1}{t}\sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}} \frac{\sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}} I_{\{\mu(s)=i\}}}{\sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}}} r_{j,i} \\
&= \sum_j \pi_j(t)\hat{a}_{j,i}(\mu,t) r_{j,i}
\end{aligned}
$$

From the above expression, we observe that the allocated rate vector over $[1,t]$, for the schedule $\{\mu(t)\}$ and channel probabilities $\{\pi_j(t)\}$ is equal to the long time average throughput achievable with a stationary random schedule $\{\hat{a}_{j,i}(\mu,t)\}$ for an ergodic channel with distribution $\{\pi_j(t)\}$. Hence, a naive strategy to drive the allocated rate vector $\bar{R}(\mu,t)$ towards the optimal rate vector $\bar{R}^*(t)$ would be to drive the actual channel allocation $\{\hat{a}_{j,i}(\mu,t)\}$ towards the optimal schedule $\{a^*_{j,i}(t)\}$. We will now define the simple channel history and schedule history based RRS-CA scheduler (RRS using channel allocation) that minimizes the Euclidean distance between the actual channel allocation $\{\hat{a}_{j,i}(t)\}$ and the optimal schedule $\{a^*_{j,i}(t)\}$ in every slot t.

1. In every slot t, estimate the channel distribution $\{\pi_j(t)\}$ as

$$
\pi_j(t) = \frac{1}{t}\sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}}
$$

2. Compute the finite time rate region $\mathcal{C}(t)$ using $\{\pi_j(t)\}$ as

$$
\mathcal{C}(t) := \left\{ (\bar{r}_1,\cdots,\bar{r}_N) : \bar{r}_i = \sum_j \pi_j(t)\, a_{j,i}\, r_{j,i}, a_{j,i} \geq 0, \sum_i a_{j,i} \leq 1, i = 1,\cdots,N \right\}
$$

3. Identify an optimal rate vector, $\bar{R}^*(t)$, in $\mathcal{C}(t)$, where

$$
\bar{R}^*(t) \in \arg\max_{\bar{R}\in\mathcal{C}(t)} U(\bar{R})
$$

4. Compute an optimal schedule $\{a^*_{j,i}(t)\}$ such that

$$
\bar{r}^*_i(t) = \sum_j \pi_j(t) a^*_{j,i}(t) r_{j,i} \ \text{ for all } i
$$

5. Define the actual channel allocation up to time $t-1$, $\{\hat{a}_{j,i}(\mu,t-1)\}$, for all j and i as

$$
\hat{a}_{j,i}(\mu,t-1) = \frac{\sum_{s=1}^{t-1} I_{\{R(s)=\mathbf{R}_j\}} I_{\{\mu(s)=i\}}}{\sum_{s=1}^{t-1} I_{\{R(s)=\mathbf{R}_j\}}}
$$

6. In every slot t, identify and implement the schedule $\mu(t) \in \{1,\cdots,N\}$ that seeks

to minimize the Euclidean distance between $\{a^*_{j,i}(t)\}$ and $\{\hat{a}_{j,i}(\mu, t)\}$:

$$\mu(t) = \arg\max_i \{a^*_{R(t),i}(t) - \hat{a}_{R(t),i}(t-1)\}$$

Remarks 4.1.1.

1. If $\{a^*_{j,i}(t)\}$ converges to $\{a^*_{j,i}\}$ as $t \to \infty$, then, we can show that $\{\hat{a}_{j,i}(t)\} \to \{a^*_{j,i}\}$ as well. Then, the allocated rate vector $\bar{R}(t)$ would also converge to the optimal rate vector $\bar{R}^*$, i.e., the RRS-CA scheduler would be asymptotically optimal.
2. For the wireless system described in Theorem 3.3.1, when $\{a^*_{j,i}\}$ is unique, we have proved that the channel allocation based scheduler RRS-CA is asymptotically optimal as well (see Appendix B).

In Figure 4.1, we plot the time average user throughput of the channel allocation based scheduler RRS-CA implementing proportional fairness. We consider a wireless system with 4 users and the channel model A described in Appendix E. Also plotted in the figure are throughput curves of the proportional fair implementation using RRS scheduler. From the figure, we observe that the RRS-CA is asymptotically optimal and it performs similar to the RRS scheduler for proportional fairness, for the channel considered.

In Figure 4.2, we plot the time average user throughput of channel allocation based scheduler RRS-CA implementing max-min fairness. Also plotted in the figure are throughput curves of a max-min fair implementation using RRS scheduler. We consider a wireless system with 3 users and the wireless channel model B described in Appendix E. From the figure, we observe that the proposed channel allocation based scheduler is asymptotically optimal and it has better convergence behaviour including reduced oscillations as compared to RRS. The rate region based scheduler RRS implements a schedule independent of the schedule history and allocated rates. Hence, the randomness introduced by the channel allocation (random schedules) and the channel evolution will settle only after large t. The channel allocation based scheduler RRS-CA, however, attempts to minimize the drift between the actual allocation and the optimal allocation in every slot t, and tends to stabilize faster. To illustrate this, we consider a *lag* metric similar to the one we use in Appendix B to prove the asymptotic optimality of RRS-CA. In each slot we define *Lag(t)*, as

$$Lag(t) = \sum_j \sum_i \left(a^*_{j,i}(t) - \hat{a}_{j,i}(t)\right)^+$$

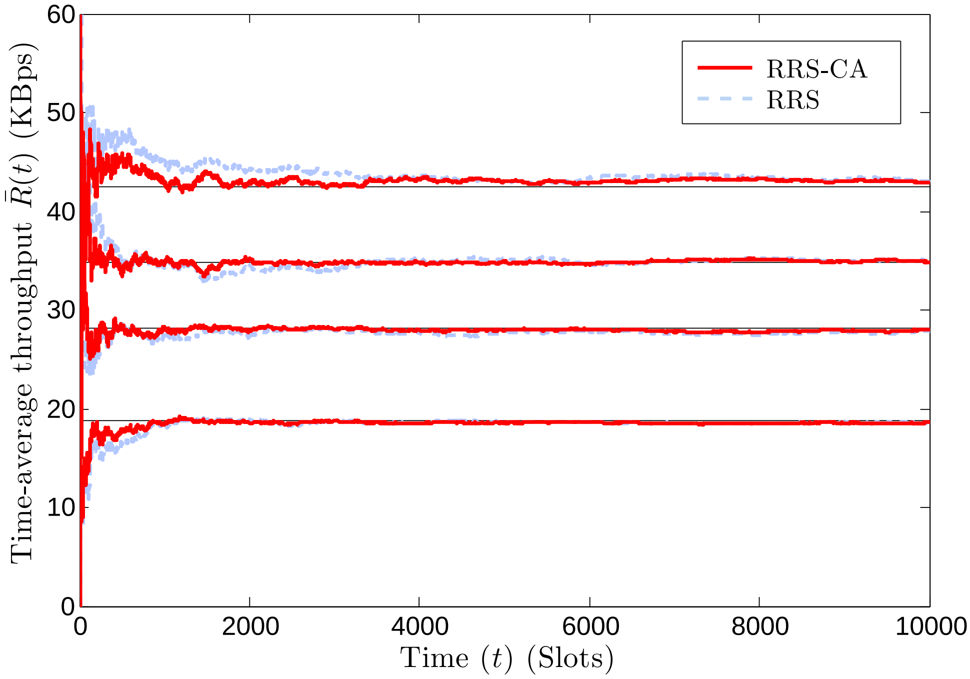

Figure 4.1: Plots of the time average user throughputs of RRS-CA and RRS, implementing proportional fairness. We consider a wireless system with 4 users and a wireless channel model A described in Appendix E. The thin straight lines correspond to the optimal rate vector $\bar{R}^*$ for the wireless channel.

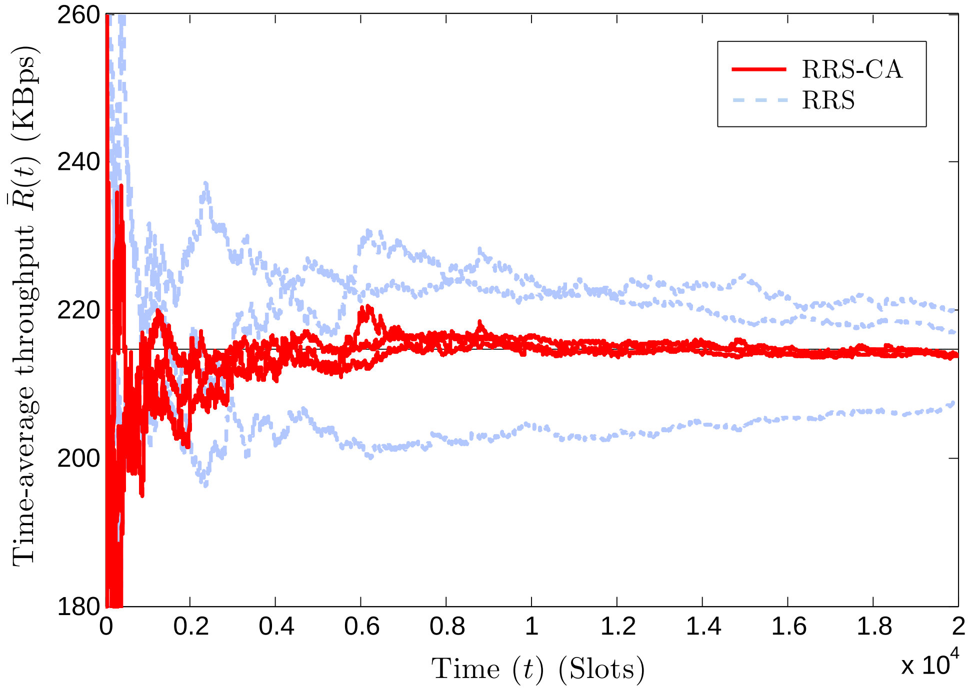

Figure 4.2: Plots of the time average user throughputs of RRS-CA and RRS, implementing max-min fairness. We consider a wireless system with 3 users and a wireless channel model B described in Appendix E. The thin straight lines correspond to the optimal rate vector $\bar{R}^*$ for the wireless channel.

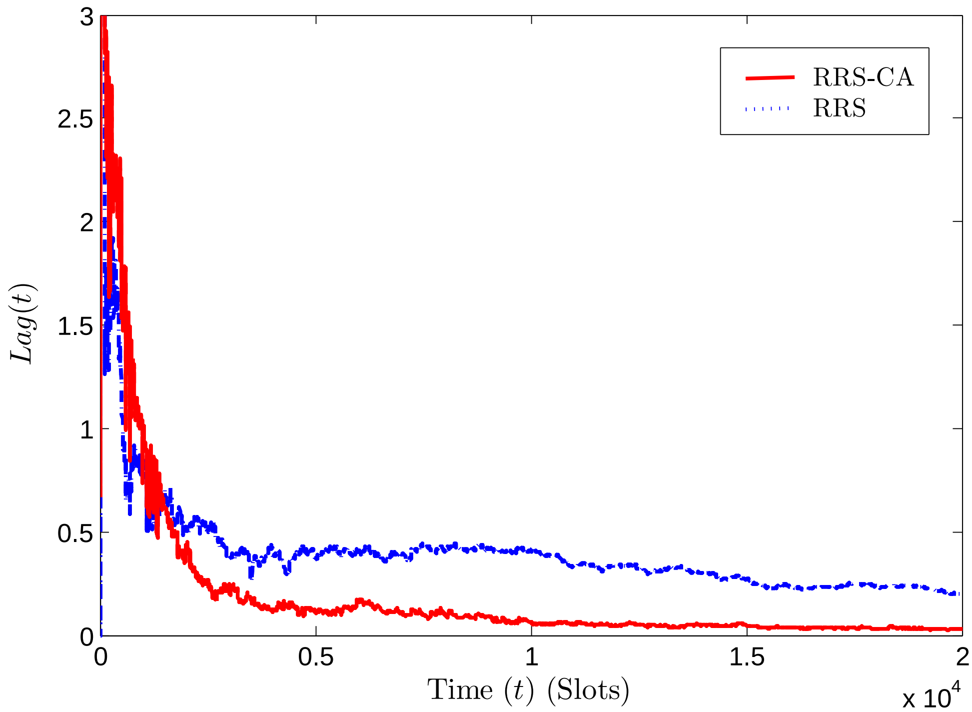

Figure 4.3: Plots of $Lag(t)$ for RRS-CA and RRS for a max-min fair implementation. We consider a wireless system with 3 users and a wireless channel model B described in Appendix E.

where $(x)^+ = \max(x, 0)$. $Lag(t)$ is a measure of the deviation of the actual channel allocation at t, $\hat{a}_{j,i}(t)$, from the optimal allocation at t, $a^*_{j,i}(t)$. Note that, $Lag(t) = 0$ implies $\bar{R}(t) = \bar{R}^*(t)$. Figure 4.3 plots $Lag(t)$ for RRS and RRS-CA for a max-min fairness implementation. We find that $Lag(t)$ for RRS-CA approaches 0 faster than in the the case of RRS. This behaviour follows directly from the form of RRS-CA and is the reason for the better convergence of RRS-CA as compared to RRS.

4.2 RRS-RA schedulers

In Section 4.1, we presented a simple channel history and schedule history based RRS-CA scheduler that would seek the time average rate vector $\bar{R}^*(t)$ by attempting to minimize the Euclidean distance between the actual channel allocation $\{\hat{a}_{j,i}(t)\}$ and the optimal schedule $\{a^*_{j,i}(t)\}$. In this section, we will seek the optimal time average rate vector $\bar{R}^*(t)$ by minimizing a distance between the optimal rate vector $\bar{R}^*(t)$ and the allocated rate vector $\bar{R}(t)$. We present a gradient scheduler that would seek the optimal rate vector $\bar{R}^*(t)$ by seeking to maximize a dynamic, auxiliary network utility in every slot t.

We consider a dynamic, auxiliary network utility function $V(\bar{R}, t) : \mathcal{C}(t) \to \mathcal{R}^1$, where $V(\bar{R}, t)$ is a strictly concave and continuously differentiable function on the user throughputs $\bar{R}$, for all t. The dynamic utility function $V(\bar{R}, t)$ is defined such that

$$\bar{R}^*(t) = \arg \max_{\bar{R} \in \mathcal{C}(t)} V(\bar{R}, t)$$

in every slot t. A gradient schedule to maximize the auxiliary network utility function is used in every slot t, i.e., the schedule in every slot is $\arg\max_i \big\{ r_i(t) \nabla_i V(\bar{R}(t-1), t) \big\}$. If $V(\bar{R}, t) \to V(\bar{R})$ (where, V is a strictly concave and continuously differentiable function such that $\bar{R}^* = \arg\max_{\bar{R} \in \mathcal{C}} V(\bar{R})$) as $t \to \infty$ (when $\pi(t) \to \pi$), then, we would expect that the gradient scheduler would be asymptotically optimal as well, i.e., the allocated rate vector $\bar{R}(t)$ would converge to the optimal rate vector $\bar{R}^*$.

4.2.1 Gradient on the network utility

Suppose that the network utility U is a strictly concave and continuously differentiable function. Then $V(\bar{R}, t) = U(\bar{R})$ is a valid auxiliary network utility function, and the RRS-RA scheduler would reduce to the popular gradient scheduling algorithm. Obviously, the auxiliary function U is independent of the observed channel distribution $\{\pi_j(t)\}$ and does not change with t. In Chapter 5, Section 5.6, we report a scenario where a dynamic, auxiliary network utility provides a significant improvement in performance as compared to a network utility based gradient scheduler.

4.2.2 Gradient on a Euclidean distance: RRS-RA-Euclid

We consider an auxiliary network utility $V(\bar{R}, t) = V\left(\bar{R}, \pi(t)\right) = \sum_{i=1}^{N} V_i\left(\bar{R}, \pi(t)\right)$ where

$$V_i\left(\bar{R}, \pi(t)\right) = -(\bar{r}_i - \bar{r}_i^*(t))^2$$

The auxiliary network utility is the negative of the square of the Euclidean distance between the rate vector $\bar{R}$ and the optimal rate vector $\bar{R}^*(t)$. RRS-RA-Euclid implements a gradient schedule in every slot for the modified network utility function $V\left(\bar{R}(t-1), \pi(t)\right)$, i.e., the schedule in every slot is $\arg\max_i \{-2(\bar{r}_i(t-1) - \bar{r}_i^*(t)) r_i(t)\}$.

We will now summarize the RRS-RA-Euclid scheduler below.

1. In every slot t, estimate the channel distribution $\{\pi_j(t)\}$ as

$$\pi_j(t) = \frac{1}{t}\sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}}$$

2. Compute the finite time rate region $\mathcal{C}(t)$ using $\{\pi_j(t)\}$ as

$$\mathcal{C}(t) := \left\{(\bar{r}_1, \cdots, \bar{r}_N) : \bar{r}_i = \sum_j \pi_j(t)\, a_{j,i}\, r_{j,i}, a_{j,i} \geq 0, \sum_i a_{j,i} \leq 1, i = 1, \cdots, N\right\}$$

3. Identify an optimal rate vector, $\bar{R}^*(t)$, in $\mathcal{C}(t)$, where

$$\bar{R}^*(t) \in \arg\max_{\bar{R}\in\mathcal{C}(t)} U(\bar{R})$$

4. In every slot, implement the schedule $\mu(t)$, where,

$$\mu(t) = \arg\max_i \{-2(\bar{r}_i(t-1) - \bar{r}_i^*(t))r_i(t)\}$$

In Figure 4.6, we plot the time average user throughput of RRS-RA-Euclid scheduler implementing proportional fairness. The figure also plots the performance of RRS and RRS-CA in comparison with the RRS-RA-Euclid scheduler. We consider a wireless system with 4 users and the channel model A described in Appendix E. In Figure 4.7, we plot the time average user throughput of RRS-RA-Euclid implementing max-min fairness. We consider a wireless system with 3 users and the channel model B described in Appendix E. The figure also plots the performance of RRS and the channel allocation based RRS-CA scheduler implementing max-min fairness. From the simulations, we observe that the RRS-RA-Euclid scheduler is asymptotically optimal and the performance shows reduced oscillations in comparison with RRS.

Remarks 4.2.1.

1. The RRS-RA-Euclid does not compute the optimal stationary schedule $\{a^*_{j,i}(t)\}$ and instead seeks the optimal rate vector $\bar{R}^*(t)$ directly. Thus, the RRS-RA-Euclid scheduler only requires that the optimal rate vector $\bar{R}^*$ be unique and does not impose restrictions on the uniqueness of the optimal stationary schedule $\{a^*_{j,i}\}$ or on the consistency in the schedule sequence.

2. The convergence behaviour of RRS-RA-Euclid near $\bar{R}^*$ is seen to become slow; in Figure 4.4 and in Figure 4.5, we note the additional delay in convergence near $\bar{R}^*$. This behaviour is a well-known feature of gradient algorithms near the optimum where the actual gradient becomes zero.

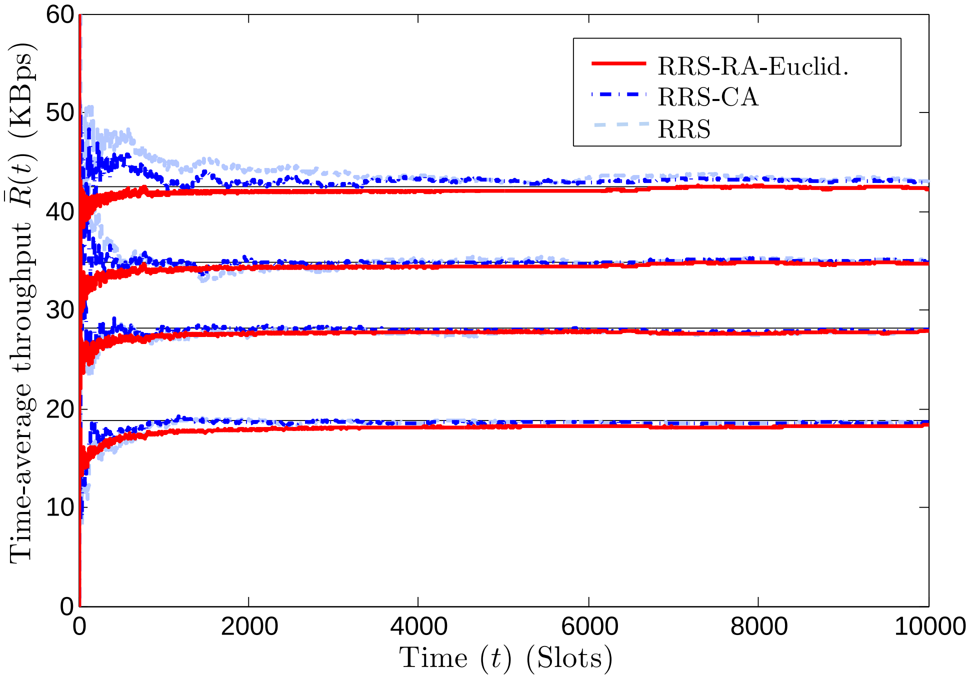

Figure 4.4: Plots of the time average user throughputs of RRS-RA-Euclid, RRS and RRS-CA implementing proportional fairness. We consider a wireless system with 4 users and a wireless channel model A described in Appendix E. The thin straight lines correspond to the optimal rate vector $\bar{R}^*$ for the wireless channel.

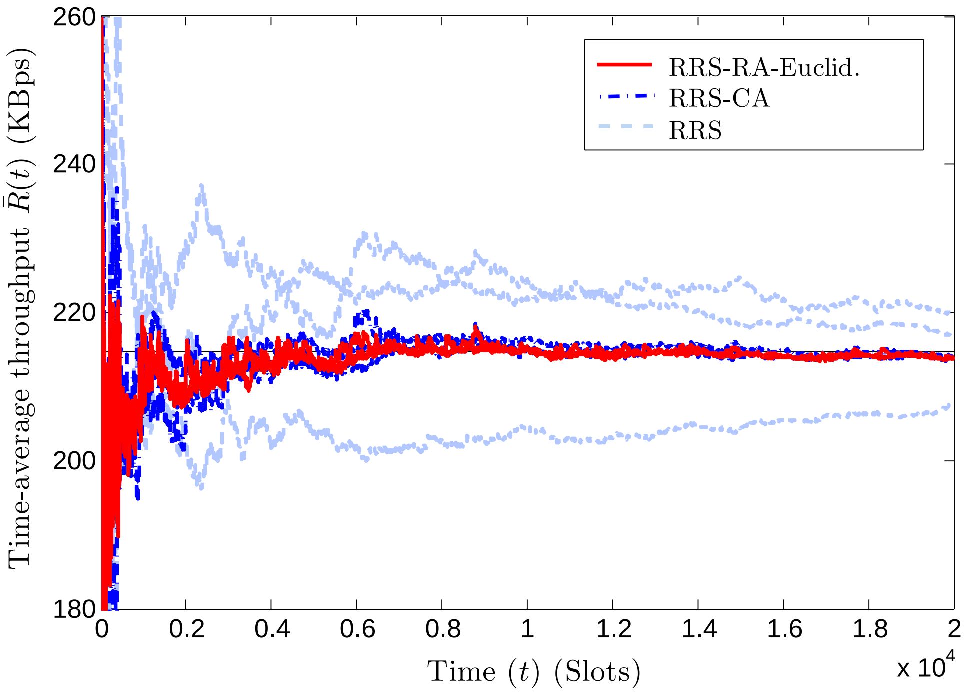

Figure 4.5: Plots of the time average user throughputs of RRS-RA-Euclid, RRS and RRS-CA implementing max-min fairness. We consider a wireless system with 3 users and a wireless channel model B described in Appendix E. The thin straight lines correspond to the optimal rate vector $\bar{R}^*$ for the wireless channel.

4.2.3 Gradient on a Sum of Weighted Logarithms: RRS-RA-WLog

In this section, we will propose weighted logarithmic utility as the dynamic, auxiliary network utility, where the weights are adjusted such that $\bar{R}^*(t)$ is the maximizer for the auxiliary utility function in the estimated rate region $\mathcal{C}(t)$.

We consider an auxiliary network utility $V\left(\bar{R},t\right) = V\left(\bar{R},\pi(t)\right) = \sum_{i=1}^{N} V_i\left(\bar{R},\pi(t)\right)$ where

$$V_i\left(\bar{R},\pi(t)\right) = w_i^*\left(\pi(t)\right)\log\left(\bar{r}_i\right)$$

RRS-RA-Wlog implements a gradient schedule in every slot t for the modified network utility function $V\left(\bar{R}(t-1),\pi(t)\right)$, i.e., the schedule in every slot is $\arg\max_i \left\{ w_i^*\left(\pi(t)\right)\frac{r_i(t)}{\bar{r}_i(t-1)}\right\}$. The weights $W^*\left(\pi(t)\right) = \left(w_1^*\left(\pi(t)\right),\cdots,w_N^*\left(\pi(t)\right)\right)$ are chosen such that

$$\bar{R}^*(t) = \arg\max_{\bar{R}\in\mathcal{C}(t)} \sum_{i=1}^{N} w_i^*\left(\pi(t)\right)\log\left(\bar{r}_i\right)$$

i.e., the optimal rate vector for a network utility $\sum_i w_i^*\left(\pi(t)\right)\log(\bar{r}_i)$ and an ergodic channel with distribution $\pi(t)$ and rate region $\mathcal{C}(t)$ is $\bar{R}^*(t)$. The weights are not unique, however, any optimal weight $W^*\left(\pi(t)\right)$ could be used to identify a schedule. When $\pi(t)\to\pi$, we expect that the weights would converge appropriately, and the gradient scheduler would then seek the optimal rate vector $\bar{R}^*$ asymptotically.

We will now summarize the RRS-RA-Wlog scheduler in detail.

1. In every slot t, estimate the channel distribution $\{\pi_j(t)\}$ as

$$\pi_j(t) = \frac{1}{t}\sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}}$$

2. Compute the finite time rate region $\mathcal{C}(t)$ using $\{\pi_j(t)\}$ as

$$\mathcal{C}(t) := \left\{(\bar{r}_1,\cdots,\bar{r}_N) : \bar{r}_i = \sum_j \pi_j(t)\, a_{j,i}\, r_{j,i}, a_{j,i}\geq 0, \sum_i a_{j,i}\leq 1, i=1,\cdots,N\right\}$$

3. Identify an optimal rate vector, $\bar{R}^*(t)$, in the rate region $\mathcal{C}(t)$, where

$$\bar{R}^*(t)\in\arg\max_{\bar{R}\in\mathcal{C}(t)} U(\bar{R})$$

4. Compute weights $W^*(t)$ such that

$$\bar{R}^*(t) = \arg\max_{\bar{R}\in\mathcal{C}(t)} \sum_i w_i^*(t)\log(\bar{r}_i)$$

5. In every slot, implement the schedule $\mu(t)$, where,

$$\mu(t) = \arg\max_i \left\{ w_i^*(t)\frac{r_i(t)}{\bar{r}_i(t-1)} \right\}$$

In Figure 4.6, we plot the time average user throughput of RRS-RA-Wlog implementing proportional fairness. We consider a wireless system with 4 users and a wireless channel model A described in Appendix E. The figure also plots the performance of RRS and RRS-RA-Euclid scheduler in comparison with the RRS-RA-Wlog scheduler. In Figure 4.7, we plot the time average user throughput of RRS-RA-Wlog implementing max-min fairness. We consider a wireless system with 3 users and the wireless channel model B described in Appendix E. The figure also plots the performance of RRS and the RRS-RA-Euclid scheduler for max-min fairness. From the simulations, we note that the RRS-RA-Wlog scheduler is asymptotically optimal and the performance of schedule history based schedulers, RRS-RA-Euclid and RRS-RA-Wlog (and RRS-CA), show reduced oscillations in comparison with the rate region based scheduler RRS.

Unlike RRS-RA-Euclid scheduler, for the RRS-RA-Wlog scheduler, the gradient at the optimal rate vector does not become zero and hence, it reports better convergence near the optimal rate vector. Of course, this limits the applicability of RRS-RA-Wlog scheduler to Pareto optimal rate vectors in the ergodic rate region. In this thesis (especially in Chapter 5), we restrict to RRS-RA-Wlog scheduler and evaluate its performance for a variety of network scenarios.

4.2.4 Comments on RRS-RA

The channel history and schedule history based scheduler RRS-RA is a practical and useful scheduler for a variety of network QoS. From simulations (reported here in Chapter 4 and in Chapter 5), we observe that the RRS-RA schedulers have better convergence behaviour than channel history based RRS scheduler. Further, RRS-RA does not compute an optimal schedule $\{a_{j,i}^*\}$ in every slot and hence, does not require that the optimal schedule be unique; In contrast, RRS and the channel allocation based RRS-CA require

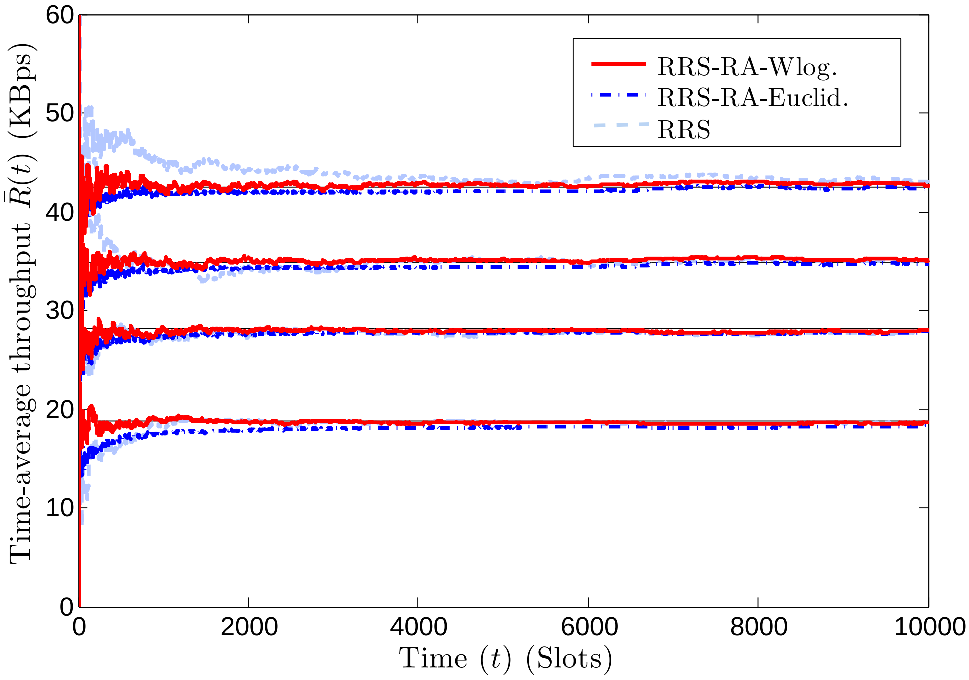

Figure 4.6: Plots of the time average user throughputs of RRS-RA-Wlog, RRS and RRS-RA-Euclid scheduler implementing proportional fairness. We consider a wireless system with 4 users and the wireless channel model A described in Appendix E. The thin straight lines correspond to the optimal rate vector $\bar{R}^*$ for the wireless channel.

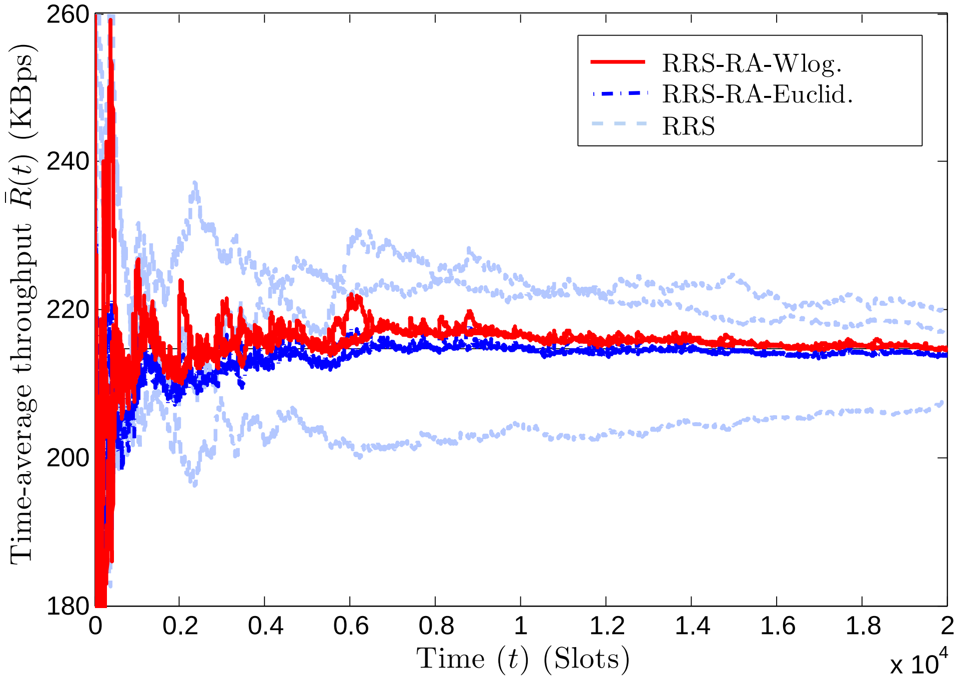

Figure 4.7: Plots of the time average user throughputs of RRS-RA-Wlog, RRS and RRS-RA-Euclid scheduler implementing max-min fairness. We consider a wireless system with 3 users and the wireless channel model B described in Appendix E. The thin straight line correspond to the optimal rate vector $\bar{R}^*$ for the wireless channel.

that the optimal schedule $\{a^*_{j,i}\}$ be unique (or, at least, we choose a consistent sequence of convergent schedules $\{a^*_{j,i}(t)\}$).

We use a strictly concave and continuously differentiable, dynamic, auxiliary utility V for the gradient algorithm. A gradient on the network utility U itself would only seek local optima (and hence is restricted to concave network utilities). Also, as reported in Chapter 5, Section 5.6, it may be useful to use an auxiliary utility V even for concave network utilities. The Euclidean and the logarithmic utility considered in RRS-RA-Euclid and RRS-RA-Wlog is appropriate but arbitrary, and we need to study optimal implementations of RRS-RA in the future.

CHAPTER 5

Performance Evaluation of Rate Region based Schedulers

The rate region based scheduling strategy provides us with a common framework to implement arbitrary definitions of network QoS and notions of fairness on the long time average user throughputs. In this chapter, we report the performance of RRS and RRS-RA, for a number of popular and interesting QoS scenarios, like, proportional fairness with minimum and maximum rate constraints, max-min fairness, maximizing a composite network utility and a non-continuous network utility. We show that the rate region based schedulers have several advantages, e.g., it provides a parameter-less implementation of the network QoS, has a convergence behaviour independent of the network utility and can adapt well with dynamic channel conditions. We have also discussed and compared the performance of the rate region based schedulers with implementations from the literature (whenever they are available).

5.1 Proportional Fairness with Rate Constraints

In Figure 5.1, we plot the time average user throughputs for proportional fair implementations with minimum and maximum rate constraints. Proportional fair rates are achieved by maximizing the sum of the logarithm of the average user throughputs subject to minimum (20 KBps) and maximum (40 KBps) rate constraints. We consider a wireless system with $N = 4$ users and the wireless channel model A described in Appendix E. In Figure 5.1, we report the performance of RRS and RRS-RA-Wlog schedulers, and compare it with an implementation from Andrews *et al.* (2005) for the network QoS. From the figure, we observe that the rate region based schedulers are asymptotically optimal and the performance (especially, RRS-RA-Wlog) is comparable with that of the GMR algorithm from Andrews *et al.* (2005).

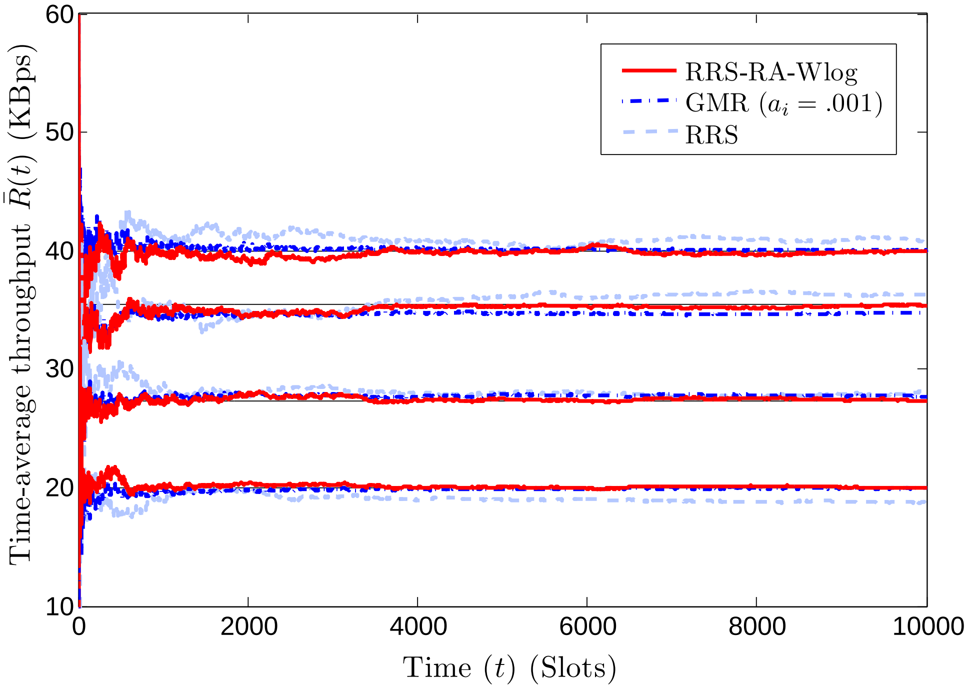

Figure 5.1: Plots of the time average user throughputs of RRS, RRS-RA-Wlog and GMR scheduler (from Andrews *et al.* (2005)) implementing proportional fairness subject to minimum (20 KBps) and maximum (40 KBps) rate constraints. We consider a wireless system with 4 users and the wireless channel model A described in Appendix E. The thin straight lines correspond to the optimal rate vector $\bar{R}^*$ for the wireless system.

5.2 Max-min Fairness

In Figure 5.2, we plot the time average user throughputs for max-min fair implementations, for a wireless system with $N = 3$ users and the wireless channel model B described in Appendix E. Max-min fair rates are achieved by seeking a Pareto optimal rate vector $\bar{R}^*(t) \in \mathcal{C}(t)$ such that $\bar{r}_i^*(t) = \bar{r}_j^*(t)$ for all $i, j = 1, 2, 3$. In Figure 5.2, we report the performance of RRS and RRS-RA-Wlog scheduler implementing max-min fairness and compare it with an implementation from Liu *et al.* (2003a) for the same network QoS. The QoS parameter δ, used in Liu *et al.* (2003a), is set to 0.00008 to optimize the performance of the max-min implementation proposed in Liu *et al.* (2003a) for the wireless system. The plots clearly show that the rate region based schedulers achieve optimal performance and the performances (especially, RRS-RA-Wlog) are comparable with the implementation reported in Liu *et al.* (2003a). In Section 5.5, we will insist that the rate region based schedulers are parameterless and are extremely useful under dynamic network and channel conditions.

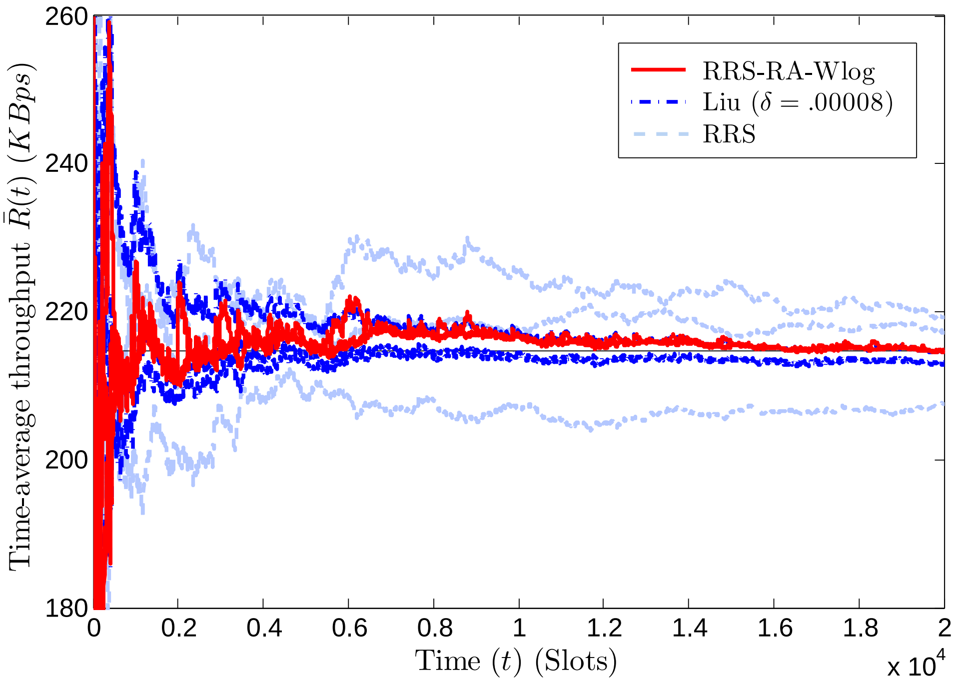

Figure 5.2: Plots of the time average user throughputs of RRS-RA-Wlog, RRS and a scheduler from Liu *et al.* (2003*a*) implementing max-min fairness. We consider a wireless system with 3 users and the wireless channel model B described in Appendix E. The thin straight line correspond to the optimal rate vector $\bar{R}^*$ for the wireless channel.

5.3 Composite Network Utility

In this section, we report the performance of the RRS-RA-Wlog scheduler implementing a composite network utility represented in Figure 5.3. We consider a wireless system with $N = 4$ users and the wireless channel model C described in Appendix E. The scheduler seeks to implement proportional fair rates if the achievable rate for every user is above the threshold (here, 16 KBps), or else, it seeks a max-min fair rate vector. The network objective is to achieve minimum performance for all users before opportunistically sharing the resources with a proportional fair utility. In Figure 5.4, we plot exponentially smoothed time average of the user throughputs for RRS-RA-Wlog scheduler, for a dynamic wireless channel. We assume that the channel distribution changes abruptly at time $t = 2000$ slots, as described in the channel model C in Appendix E. We track the dynamic channel distribution using an exponentially smoothed average as

$$\pi_j^\beta(t) = (1-\beta)\pi_j^\beta(t) + \beta I_{\{R(t)=\mathbf{R}_j\}}$$

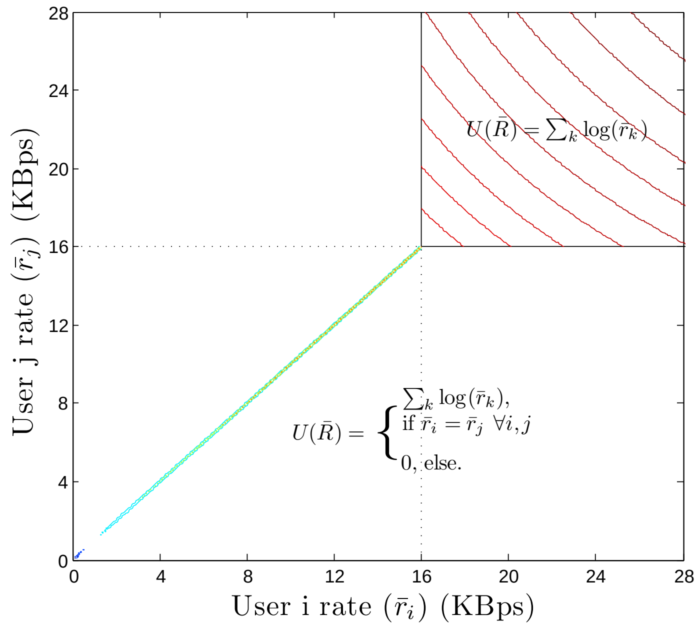

Figure 5.3: Plot of a representation of a composite network utility. The scheduler seeks proportional fair rates if the achievable rate for every user is above the threshold (16 KBps) or else it seeks a max-min fair rate vector.

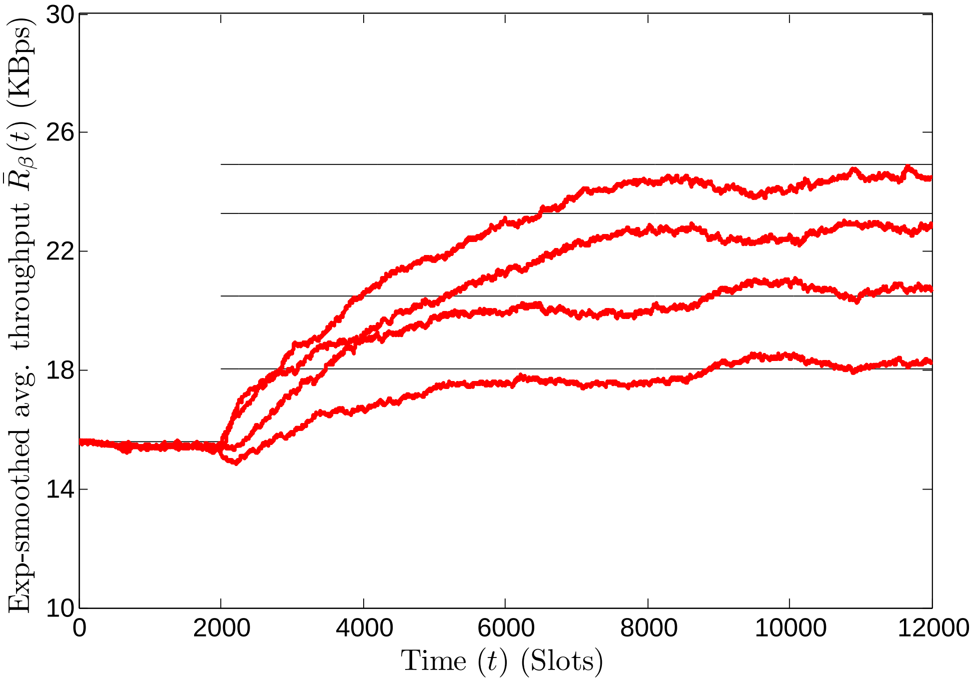

Figure 5.4: Plots of the exponentially smoothed average of user throughputs for RRS-RA-Wlog implementing the composite network utility shown in Figure 5.3. We consider a wireless system with $N = 4$ users and the wireless channel model C described in Appendix E. The channel distribution changes abruptly at $t = 2000$ slots (as described in Appendix E). The ergodic optimal rate vectors for the two channel distributions are marked by thin straight lines along the time axis. We have initialized the time average rate vector to the optimal rate vector of the wireless channel (of channel distribution before $t < 2000$ slots) at $t = 0$.

where $\{\pi_j^\beta(t)\}$ is the channel estimate at time t, and β is the averaging parameter, $0 < \beta \ll 1$. We use an exponentially smoothed average for the allocated rate vector as well. The thin straight lines in the Figure 5.4 correspond to the optimal rate vectors for the composite network utility for the two channel distributions. From the figure, we observe that the RRS-RA-Wlog scheduler achieves optimal performance for the composite network utility under dynamic channel conditions as well.

Remarks 5.3.1. *The gradient algorithm is a popular class of scheduler for concave network utilities. For general network utilities and QoS, we note that there only exists different forms (implementations) of schedulers, that estimate and seek channel and system parameters differently. In this thesis, we propose a general framework that implements arbitrary QoS including composite network utilities.*

5.4 Non-continuous Network Utility

In Figure 5.6, we report the implementation of RRS-RA-Wlog scheduler maximizing a non-continuous network utility shown in Figure 5.5. The network utility reported in Figure 5.5 is the sum of the logarithm of the user throughputs defined only in certain regions of $\mathcal{R}^2$. We consider a wireless network with $N = 2$ users and the wireless channel model C described in Appendix E. In Figure 5.6, we plot the exponentially smoothed average of the throughputs of the two users with RRS-RA-Wlog, for a dynamic channel scenario. The channel distribution changes abruptly at time $t = 2000$ slots as described in the channel model C in Appendix E, and, we track the dynamic channel distribution using an exponentially smoothed average ($\pi_j^\beta(t) = (1-\beta)\pi_j^\beta(t) + \beta I_{\{R(t)=\mathbf{R}_j\}}$). We use an exponentially smoothed average for tracking the allocated rate vector as well. The optimal rate vector for the two distributions (before and after $t = 2000$ units) are marked by a $*$ and o in Figure 5.5 and by thin straight lines in Figure 5.6. From Figure 5.6, we clearly see that the RRS-RA-Wlog can achieve optimal performance even for a non-concave and non-continuous utility function and under dynamic channel conditions.

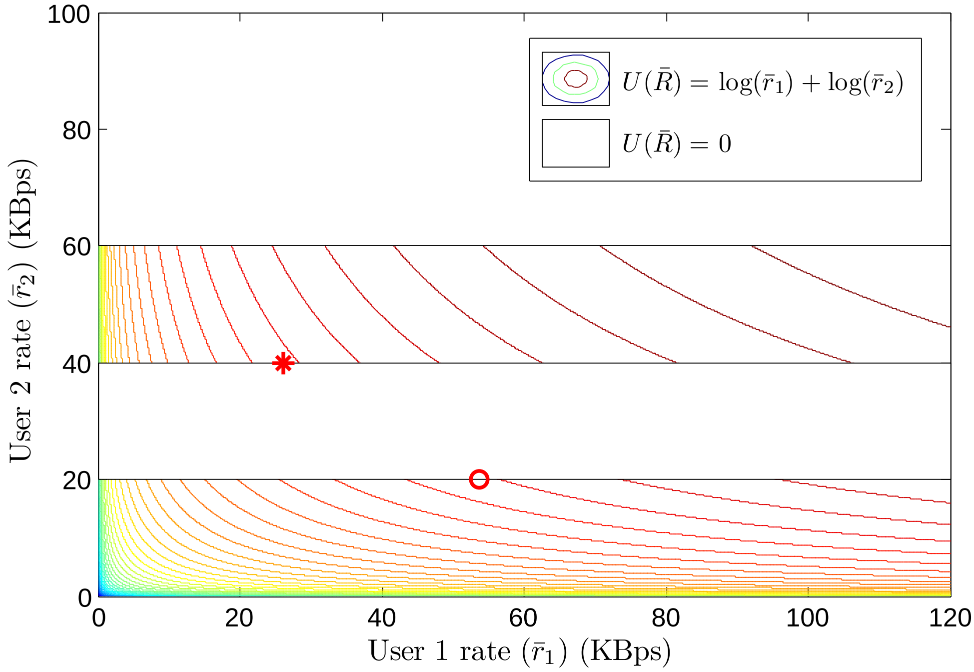

Figure 5.5: Plot of a non-concave and non-continuous network utility with respect to the average rates of the two users. The network utility is $\sum_i \log(\bar{r}_i)$ restricted to certain regions of $\Re^2$. The optimal rate vectors for the two channel distributions are marked on the network utility by a $*$ (for $t < 2000$ slots) and by an o (for $t \geq 2000$ slots).

Remarks 5.4.1. *The gradient schedulers seek local optima and hence they are restricted to concave network utilities. The rate region based schedulers, however, use the entire available channel history to estimate and seek the global optimal rate vector. Hence, they can implement arbitrary network QoS and support dynamic channel conditions as well. Most internet service providers (ISP) support discrete rates only for the DSL subscribers. Further, the user satisfaction metrics need not be concave functions of the average rates. Hence, there is hardly any reason to restrict to concave network utilities in such cases. We believe that the rate region based schedulers provides us with a powerful tool to implement a variety of such useful QoS.*

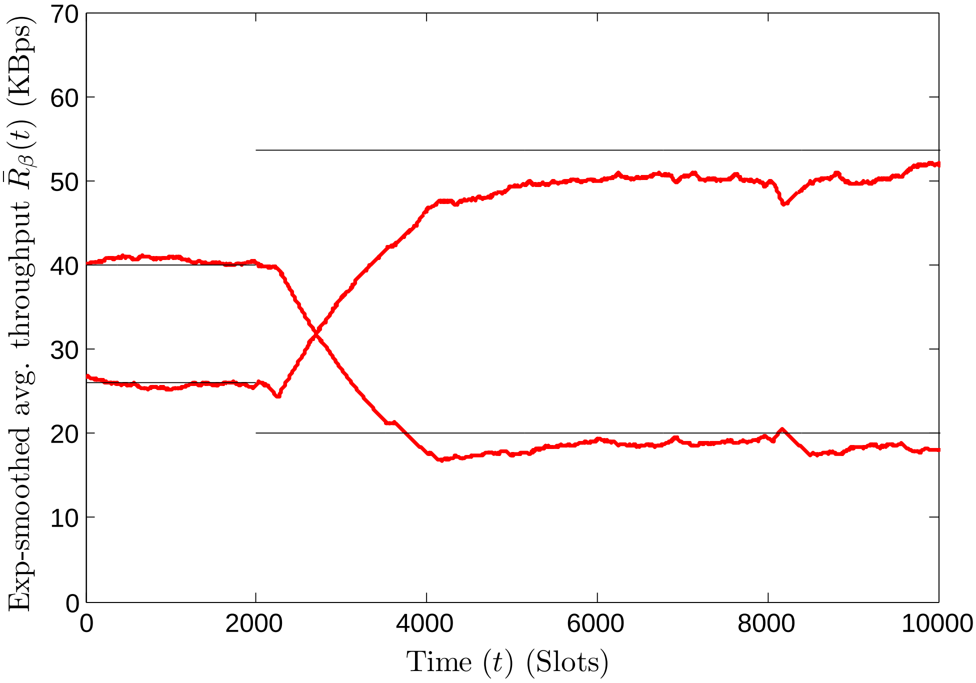

Figure 5.6: Plots of the exponentially smoothed time average user throughput of the two users for RRS-RA-Wlog scheduler implementing the network utility reported in Figure 5.5. We consider a wireless system with $N = 2$ users and the wireless channel model C described in Appendix E. The channel distribution (see Appendix E) changes abruptly at $t = 2000$ slots. The optimal rate vectors for the two channel distributions are marked by thin straight lines along the time axis. We have initialized the time average rate vector to the optimal rate vector of the wireless channel (of the channel distribution before $t < 2000$ slots) at $t = 0$.

5.5 A Parameter-less Implementation

In Figure 5.7, we plot the time average user throughputs of max-min fair implementation of RRS-RA-Wlog scheduler and compare it with an implementation from Liu *et al.* (2003*a*). We consider a wireless system with $N = 4$ users and the wireless channel model B described in Appendix E. We consider three different values for the step-size parameter δ ($\delta = 0.01, 0.001$ and 0.0001), for the implementation reported in Liu *et al.* (2003*a*). The parameter δ provides the tradeoff between convergence accuracy and convergence time. From Figure 5.7, we see that for $\delta = 0.01$, the convergence is fast but is suboptimal and for $\delta = 0.0001$, the convergence tends to be accurate but is slow. The value $\delta = 0.001$ achieves a good tradeoff between the convergence time and accuracy. The plots clearly illustrate that a wrong choice of parameters in algorithms like those in Liu *et al.* (2003*a*) could lead to poor system performance. The RRS-RA-Wlog implementation in Figure 5.7, however, achieves a performance similar to that of $\delta = 0.001$ without any input parameter; the performance of the RRS-RA-Wlog scheduler is tuned to the channel distribution in all scenarios. Another scheduler from literature with such a step-size parameter is the Gradient algorithm with Minimum/Maximum Rate constraints (GMR) proposed in Andrews *et al.* (2005). In Figure 5.8 we plot the performance of GMR for different values of the parameter $a_i(.0001, .001, .01)$. The objective is to maximize the utility $\sum_i \log(\bar{r}_i)$ subject to $\bar{r}_i > 25$ KBps for all i. Also plotted in the same figure is the performance of RRS-RA-Wlog for the same objective. We find that RRS-RA has good convergence rate to the optimal rate vector, without the need of any parameter, whereas a poor choice of parameter in GMR could lead to either slow convergence or sub-optimal performance. Thus, we see that RRS-RA-Wlog scheduler can provide a **parameter-less implementation** most appropriate for dynamic network scenarios.

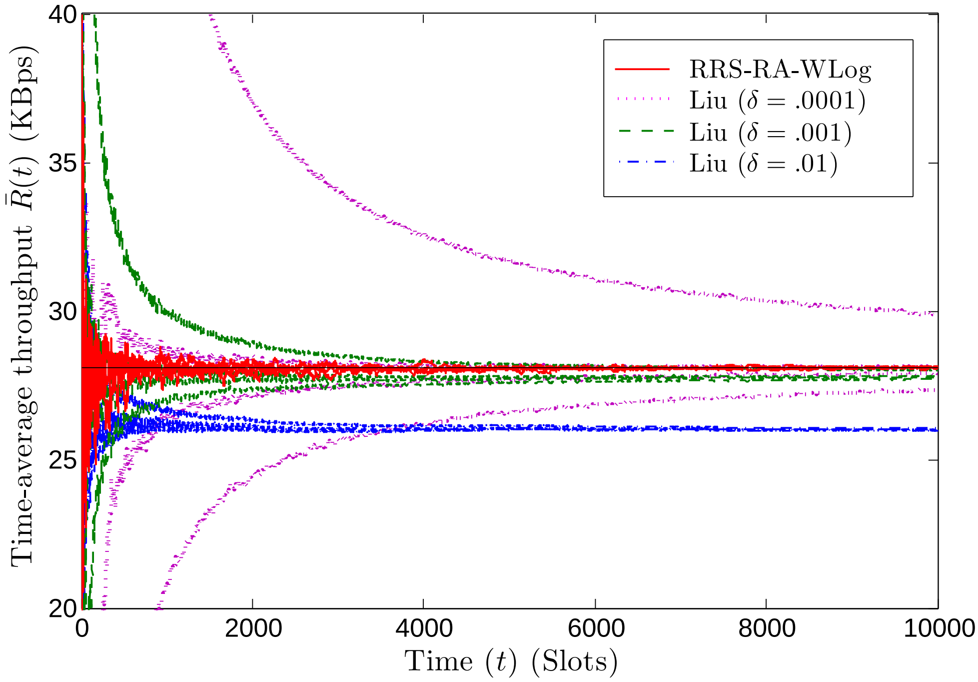

Figure 5.7: Plots of the time average user throughputs of RRS-RA-Wlog scheduler and an implementation from Liu *et al.* (2003*a*) for max-min fairness for a wireless system with $N = 4$ users and the wireless channel model A described in Appendix E. We report the performance of Liu *et al.* (2003*a*) for three different values of the step-size $\delta(\delta = .01, .001, .0001)$. The thin straight line corresponds to the optimal rates for the wireless channel.

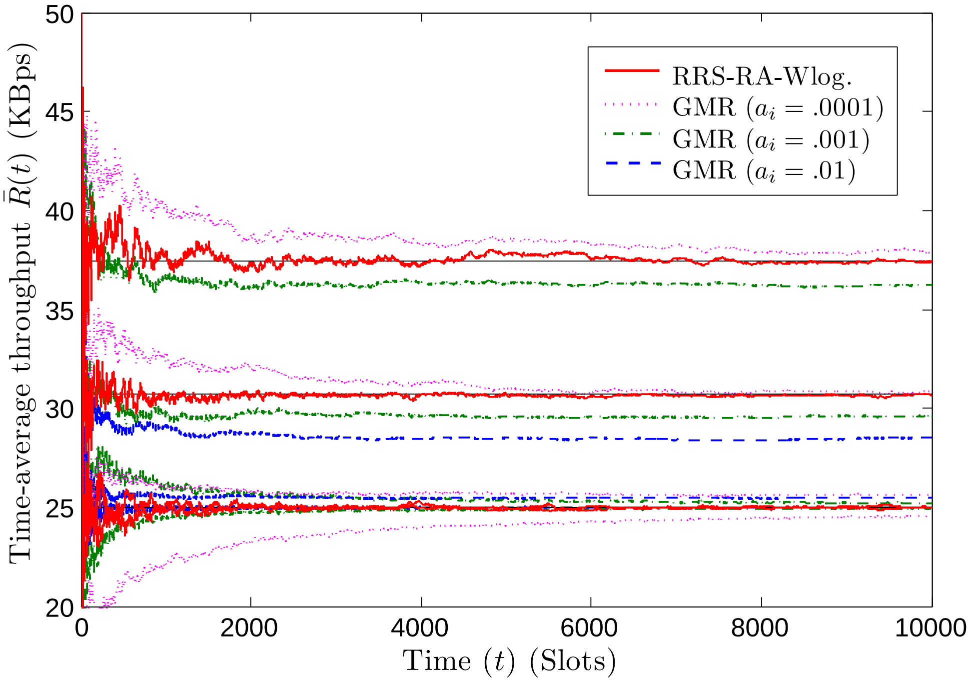

Figure 5.8: Plots of the time average user throughputs of RRS-RA-Wlog scheduler and GMR from Andrews *et al.* (2005) for maximizing $\sum_i \log(\bar{r}_i)$ subject to $\bar{r}_i > 25$ KBps $\forall i$. The wireless system has $N = 4$ users and the wireless channel model A described in Appendix E. We report the performance of Andrews *et al.* (2005) for three different values of the step-size $a_i(a_i = .01, .001, .0001)$. The thin straight lines correspond to the optimal rates for the wireless channel.

5.6 Utility Independent Convergence

In this section, we will report the applicability of RRS-RA-Wlog scheduler for a strictly concave and continuously differentiable network utility in comparison with the gradient scheduler on the network utility.

Consider a wireless system with 4 users and a strictly concave and continuously differentiable network utility $U = \sum_{i=1}^{4} U_i$, where U_1, U_2, U_3 and U_4 are the user utility functions represented in Figure 5.9 (U_2, U_3, U_4 are logarithm functions). U_1 is a strictly concave function and has a continuous but sharply varying gradient. The $\log$ utility of users 2,3 and 4 is a strictly concave and continuously differentiable function in $\Re^{+4} \setminus 0^4$. The gradient scheduler (Stolyar (2005*b*)) is an asymptotically optimal scheduler for the given network utility U. However, since the scheduler uses gradient of U in its scheduling decision, its convergence behaviour is affected by the *shape* of the utility function. This effect is most pronounced when the gradient of U has sharp variations near the optimal rate vector. In such cases, the gradient algorithm may implement very different schedules in the neighborhood of the optimal rate vector. Such behaviour was reported in Andrews *et al.* (2005) in the context of enforcing rate constraints by modifying the network utility. However, the utility function given to the scheduler itself could possibly have such sharp gradient variations (as in this case).

Since RRS-RA uses U only to calculate $\bar{R}^*(t)$ and does not use U directly in the scheduling decision, we expect it to have a *utility-independent* convergence behaviour. By this, we mean that the performance of the scheduler will depend on the network utility only to the extent that U determines $\bar{R}^*(t)$; given $\bar{R}^*(t)$, the specific shape and gradient variations of U will not affect the performance of RRS-RA. In Figure 5.10 we have plotted the performance of RRS-RA-Wlog scheduler for the network utility U. Also plotted in the same figure is the performance of the gradient scheduler for the same utility. We consider a wireless system with $N = 4$ users and the channel model A (described in Appendix E). From the figure, we find that gradient algorithm exhibits poor convergence rate as compared to RRS-RA-Wlog. This example supports the claim that rate-region based schedulers exhibit utility-independent convergence unlike gradient schedulers. The example further shows that, even in cases of strictly concave and continuously differentiable network utility, utilizing the channel history information in scheduling, may significantly improve the performance of the users.

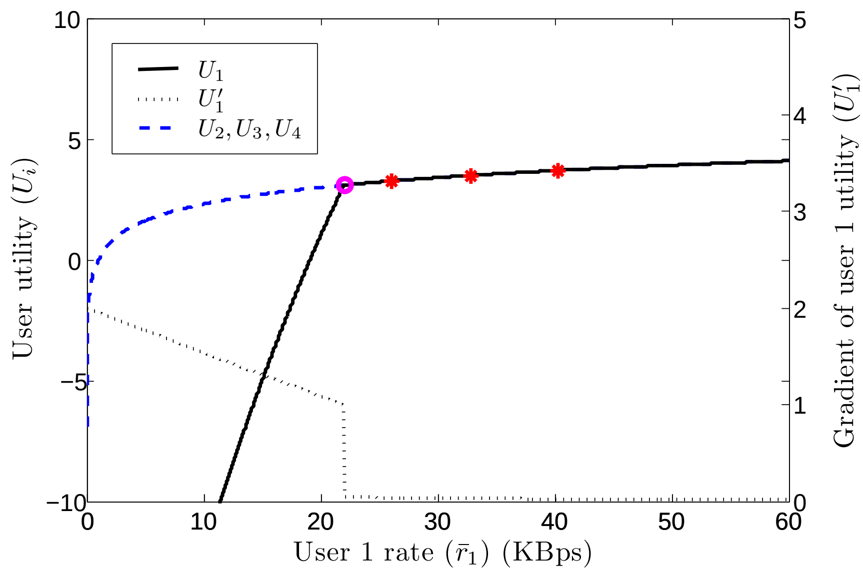

Figure 5.9: Plots of strictly concave and continuously differentiable user utility functions U_1 through U_4 with respect to the average user throughput. Also plotted in the figure is the gradient of utility function of user 1, U_1'. The optimal rates for the network utility $\sum_{i=1}^{4} U_i$ and for the channel model A (described in Appendix E), are marked by an o (user 1) and * s (user 2,3 and 4) in the figure.

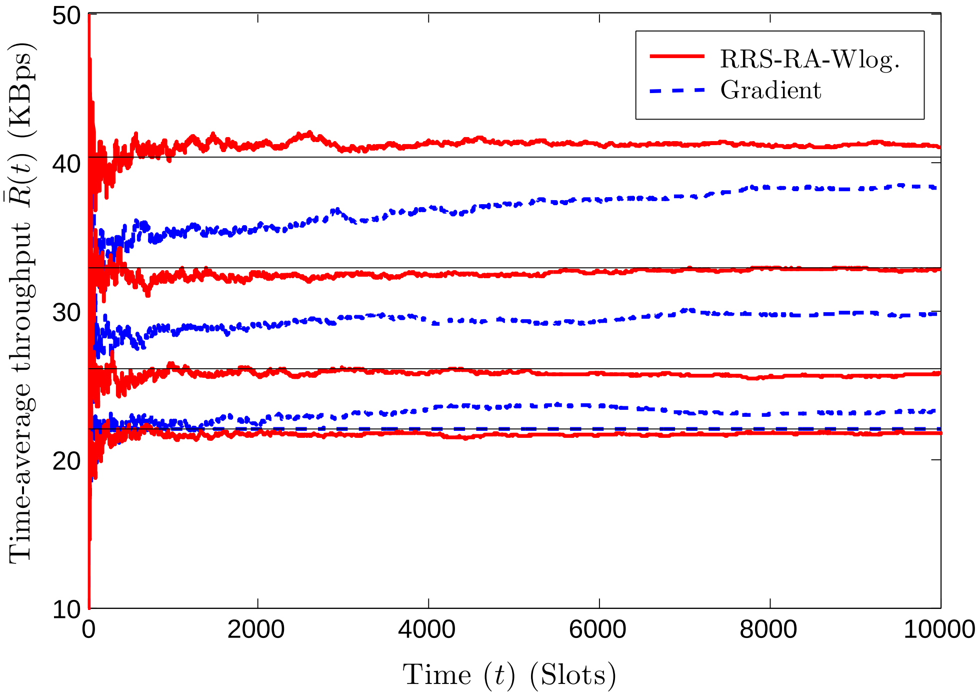

Figure 5.10: Plots of time average user throughputs for the gradient scheduler Stolyar (2005*b*) and the RRS-RA-Wlog scheduler for network utility $U = \sum_{i=1}^{4} U_i$ (reported in Figure 5.9) and the channel model A (described in Appendix E). The optimal rate vector for the system is marked by thin black lines.

5.7 A Non-Stationary Wireless Channel

In Theorem 3.3.1 and in all other simulations, we have considered an ergodic channel process with a discrete stationary distribution. We believe that the ergodicity of the channel process is a sufficient condition for the asymptotic optimality of RRS schedulers and may not be necessary. We conjecture that the convergence of the rate region $\mathcal{C}(t)$ (to some $\mathcal{C}$) is sufficient to design an asymptotically optimal scheduler.

Consider a simple wireless system with 2 users and a channel process $\{R(t)\}_{t=1}^{\infty}$ given by $R(t) = (1 - \frac{1}{t+1}, 1)$ MBps. Note that the channel process is not stationary and $\{R(t)\}_{t=1}^{\infty}$ does not have a stationary distribution. However, as shown in Figure 5.11, the estimated rate region of the wireless channel converges to the rate region of a constant wireless channel process with channel state $(1, 1)$ MBps.

In Figure 5.12, we compare the user performances of max-min fair implementations of RRS and RRS-RA-Wlog for the wireless system. We note that RRS-RA-Wlog is asymptotically optimal while RRS fails to achieve the desired time average rate vector. RRS fails to seek the optimal rate vector as the sequence of schedules $\{\{a^*_{R(t),i}(t)\}\}_{t=1}^{\infty}$

fails to implement the optimal schedule $\{a^*_{j,i}(t)\}$. The RRS-RA implementation seeks to drive the allocated rate vector towards the optimal rate vector $\bar{R}^*(t)$ in every slot t, and hence, adapts well with the channel evolution than RRS.

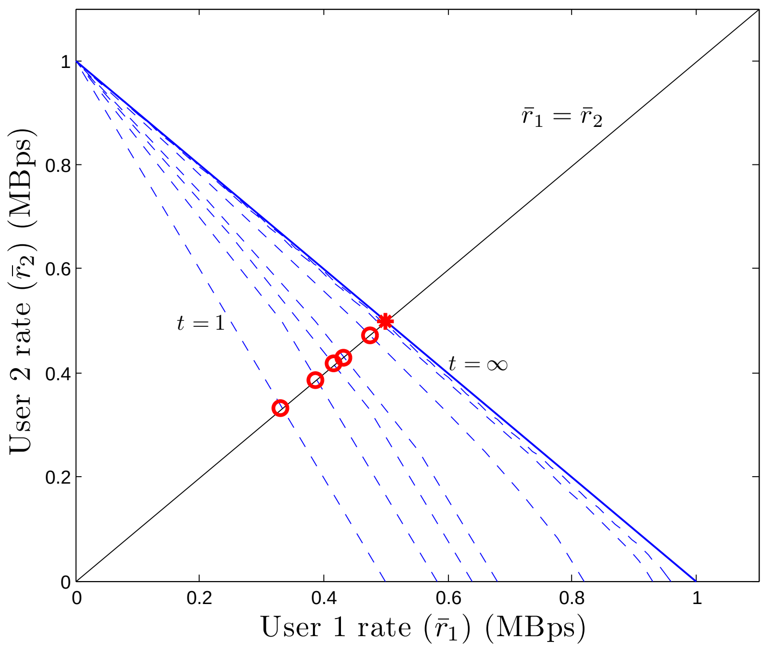

Figure 5.11: Plot showing the convergence of estimated rate region $\mathcal{C}(t)$ for the channel process given by $R(t) = \left(1 - \frac{1}{t+1}, 1\right)$. The asymptotically optimal rate vector, $\bar{R}^*$ is marked by an * and optimal rate vectors for the finite-time estimates of rate region are marked by o s.

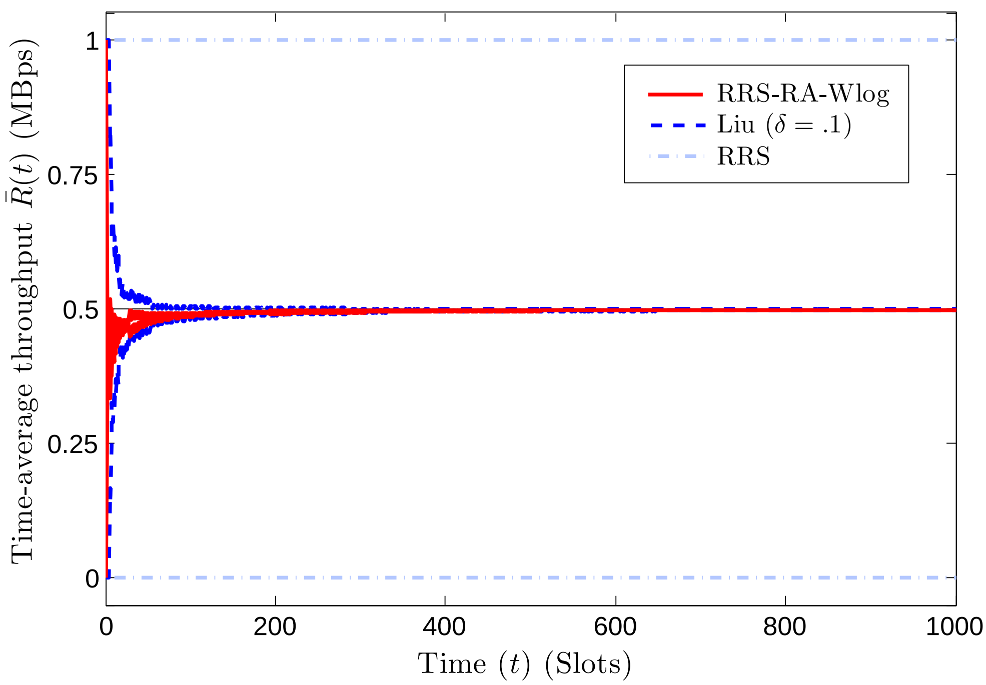

Figure 5.12: Plots of time average throughputs of 2 users for RRS-RA-Wlog, a scheduler from Liu *et al.* (2003*a*) and RRS, implementing max-min fairness. The channel process is given by $R(t) = \left(1 - \frac{1}{t+1}, 1\right)$.

CHAPTER 6

Simple Implementations of Max-Min Fairness

The rate region based schedulers, RRS, RRS-CA and RRS-RA, are computationally expensive to implement online. The scheduler needs to compute the rate region $\mathcal{C}(t)$, identify an optimal rate vector $\bar{R}^*(t)$ and a corresponding schedule in every slot t. The complexity of the optimization problem $\arg\max_{\bar{R}\in\mathcal{C}(t)} U(\bar{R})$ greatly depends on the nature of the network utility function U. If the network utility U is not concave, or, if the utility function supports multiple maxima[1], then, there may not exist simple and fast algorithms to identify the optimal rate vector $\bar{R}^*(t)$. Further, schedulers like RRS and RRS-CA requires to identify a consistent convergent sequence of schedules for asymptotic optimality. And, the RRS-RA schedulers need to solve another non-linear optimization problem to identify a schedule.

The state space of the problem and the memory requirement increases exponentially with the number of users. If we suppose that there are M channel states per user, then, for independent channels of users, we have a wireless system with M^N channel states. In Chapters 3, 4 and 5, we have restricted ourselves to simple examples with fewer users and channel states for the same reason. Our objective in this thesis is to illustrate the importance and advantages of using channel history and allocation history in scheduling and we did not concern ourselves with the complexity of the implementation.

In future, we expect to develop simple, fast and asymptotically optimal versions of the rate region based schedulers for online implementations. We list below some of the techniques useful in this regard.

1. Approximate the rate region $\mathcal{C}(t)$ of the wireless channel (e.g., discretization, channel grouping).
2. Approximate the optimal rate vector $\bar{R}^*(t)$ or the schedule (discrete search on the rate vector and schedule space)
3. Compute the network parameters at a slower rate (e.g., solve the optimization problem over multiple slots).

[1]In Theorem 3.3.1, and in this thesis, we have assumed that there exists an unique optimal rate vector $\bar{R}^*$ in the rate region of the wireless channel $\mathcal{C}$; this, however, does not imply or guarantee that there will exist a unique $\bar{R}^*(t)$ in $\mathcal{C}(t)$ for all t.

4. Independent channels across the users may help reduce computation and memory requirements.

In this chapter, we discuss two different techniques to implement max-min fairness using the estimated rate region and the allocated rates. We believe that it is more appropriate to develop simple and computationally inexpensive algorithms specific to a given network QoS than for a general scenario. In Section 6.1, we propose an asymptotically optimal scheduler for max-min fairness using a step-by-step search on the boundary of the rate region. In Section 6.2, we propose a suboptimal scheduler that approximates the optimal rate vector $\bar{R}^*(t)$ and the optimal weights $W^*(t)$ to minimize computation. In both the cases, we use RRS-RA-Wlog as our scheduler framework. Finally, we propose a simple technique that uses channel history information to increase the convergence rate of some existing schedulers for max-min fairness.

6.1 A Max-min Fair Scheduler for Two Users

The RRS-RA-Wlog scheduler schedules $\arg\max_i \left\{ w_i^*(t) \frac{r_i(t)}{\bar{r}_i(t-1)} \right\}$ in every slot t, where $W^*(t)$ is chosen such that $\bar{R}^*(t) = \arg\max_{\bar{R} \in \mathcal{C}(t)} \sum_i w_i^*(t) \log(\bar{r}_i)$.

Let $\sum_{i=1}^N m_i^*(t)\bar{r}_i = d$, be a hyperplane in $\Re^N$, tangential to $\mathcal{C}(t)$ at $\bar{R}^*(t)$. Then, $\max_{\bar{R} \in \mathcal{C}(t)} \sum m_i^*(t)\bar{r}_i^*(t) \log \bar{r}_i = \bar{R}^*(t)$. That is, $[m_1^*(t)\bar{r}_1^*(t), m_2^*(t)\bar{r}_2^*(t), \ldots m_N^*(t)\bar{r}_N^*(t)]$ is an optimal weight that can be used in RRS-RA-Wlog (see Appendix D). In the case of max-min fairness, since $\bar{r}_1^*(t) = \bar{r}_2^*(t) = \ldots = \bar{r}_N^*(t),\ \forall t$, $[m_1^*(t), m_2^*(t), \ldots, m_N^*(t)]$ is an optimal weight.

Thus, to identify optimal weights $W^*(t)$, it is sufficient to identify the slope $M^*(t)$ of a face of the finite time rate-region $\mathcal{C}(t)$ that contains $\bar{R}^*(t)$. The RRS-RA-Wlog scheduler computes the optimal weights $W^*(t)$ (equivalently, the slopes $M^*(t)$) in every slot t before identifying an schedule. In Figure 6.1, we illustrate evolution of the finite time rate region $\mathcal{C}(t)$ and identify the face of the rate region $\mathcal{C}(t)$ that contains the max-min fair rate vector.

We will now propose a simple scheduler based on RRS-RA-Wlog that uses an approximate weight $W(t)$ to identify a schedule, and achieves an asymptotically optimal performance as $W(t)$ converges to the set optimal weights at t.

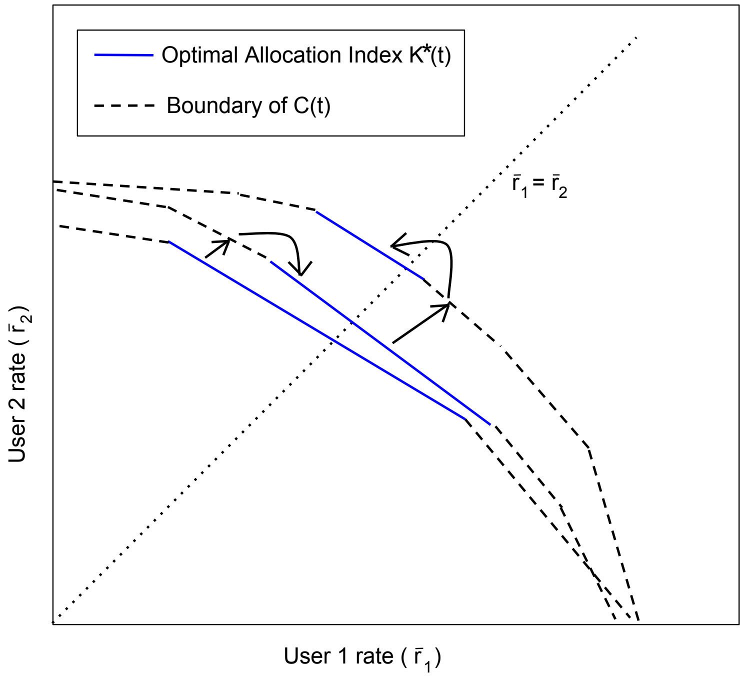

Figure 6.1: Illustration of the evolution of the finite time rate region $\mathcal{C}(t)$ and the face (slope) that contains the max-min fair rate vector.

Approximate RRS-RA-WLog algorithm for 2 users

Without loss of generality, we will assume that the channel states $\{\mathbf{R}_1, \mathbf{R}_2, \ldots \mathbf{R}_m\}$ are ordered in the decreasing order of $\frac{r_{j,1}}{r_{j,2}}$. Let K be an allocation index such that $K \in \{1, \cdots, M\}$, where channel states $(\mathbf{R}_1, \cdots, \mathbf{R}_K)$ are allocated to user 1 and channel states $(\mathbf{R}_{K+1}, \cdots, \mathbf{R}_M)$ are allocated to user 2. The ergodic rate vector achievable with the allocation K and a channel distribution $\pi(t)$ is now given by $\bar{R}_K(\pi(t)) := (\sum_{j=1}^{K} \pi_j(t)\mathbf{r}_{j,1}, \sum_{j=K+1}^{M} \pi_j(t)\mathbf{r}_{j,2})$. The RRS-RA-Wlog scheduler seeks an optimal allocation index $K^*(t)$ in every slot t such that $\sum_{j=1}^{K^*(t)-1} \pi_j(t)\mathbf{r}_{j,1} < \sum_{j=K^*(t)}^{M} \pi_j(t)\mathbf{r}_{j,2}$ and $\sum_{j=1}^{K^*(t)} \pi_j(t)\mathbf{r}_{j,1} >= \sum_{j=K^*(t)+1}^{M} \pi_j(t)\mathbf{r}_{j,2}$. The optimal weights $W^*(t)$ corresponding to the optimal allocation index $K^*(t)$ would be $W^*(t) = (\frac{1}{\mathbf{r}_{K^*(t),1}}, \frac{1}{\mathbf{r}_{K^*(t),2}})$. We propose a simple, asymptotically optimal max-min fair scheduler for the 2 users, which adapts $K(t)$ in steps to seek $K^*(t)$.

Algorithm 1 RRS-RA-WLog Scheduler for $N = 2$ users

$K(0) \in \{1, 2 \dots m\}$.

$\bar{R}_{K(0)}(\pi(t)) := (\sum_{j=1}^{K(0)} \pi_j(t)\mathbf{r}_{j,1}, \sum_{j=K(0)+1}^{M} \pi_j(t)\mathbf{r}_{j,2})$

for $t = 1$ **to** ∞ **do**

  $K(t) \leftarrow K(t-1)$

  while $\sum_{j=1}^{K(t)-1} \pi_j(t)\mathbf{r}_{j,1} > \sum_{j=K(t)}^{M} \pi_j(t)\mathbf{r}_{j,2}$ **do**

    $K(t) \leftarrow K(t) - 1$

  end while

  while $\sum_{j=1}^{K(t)} \pi_j(t)\mathbf{r}_{j,1} < \sum_{j=K(t)+1}^{M} \pi_j(t)\mathbf{r}_{j,2}$ **do**

    $K(t) \leftarrow K(t) + 1$

  end while

  $\mu(t) \leftarrow \arg\max_i \frac{r_i(t)}{r_{K(t),i}\bar{r}_i(t-1)}$

end for

The algorithm uses the weights $W(t) = (\frac{1}{\mathbf{r}_{K(t),1}}, \frac{1}{\mathbf{r}_{K(t),2}})$ in every slot t. As the rate region converges, the optimal allocation index $K(t) \to K^*(t)$ and the weights $W(t)$ converges to the set of optimal weights at t. If we choose to identify the optimal $K^*(t)$ in every slot t instead of $+/-1$ step size, we would implement RRS-RA-Wlog scheduler for the wireless system.

In Figure 6.2, we plot the slope corresponding to the channel allocation $K(t)$ $\left(\frac{\mathbf{r}_{K(t),2}}{\mathbf{r}_{K(t),1}}\right)$ for the approximate RRS-RA-Wlog scheduler. We consider a wireless system with $N = 2$ users and a wireless channel model D described in Appendix E. From the figure, we see that the slope (equivalently, the weights $W(t)$ or the allocation index $K(t)$) converges to the optimal slope (equivalently, the weights $W^*(t)$ or the allocation index $K^*(t)$) as t tends to infinity. In Figure 6.3, we report the time average user throughput of the two users for the system. The thin straight line in the figure corresponds to the optimal rate vector for the wireless system. In Figure 6.3, we also compare the performance of the approximate RRS-RA-Wlog scheduler with a max-min fair scheduler reported in Borst and Whiting (2001). The step-size parameter for the algorithm in Borst and Whiting (2001) is tuned for the wireless system. From the figure, we see that the approximate RRS-RA-Wlog scheduler reports a performance comparable to that of Borst and Whiting (2001). The ergodic rate vector for the allocation index $K(t)$ and the allocated rate vectors $\bar{R}(t)$ permit an iterative calculation and can be computed in few steps. Thus, the complexity of the approximate RRS-RA-Wlog scheduler is similar to that of the implementation from Borst and Whiting (2001).

The above implementation clearly illustrates that we can develop simple implementations of schedulers based on the estimated rate region. We believe that the search

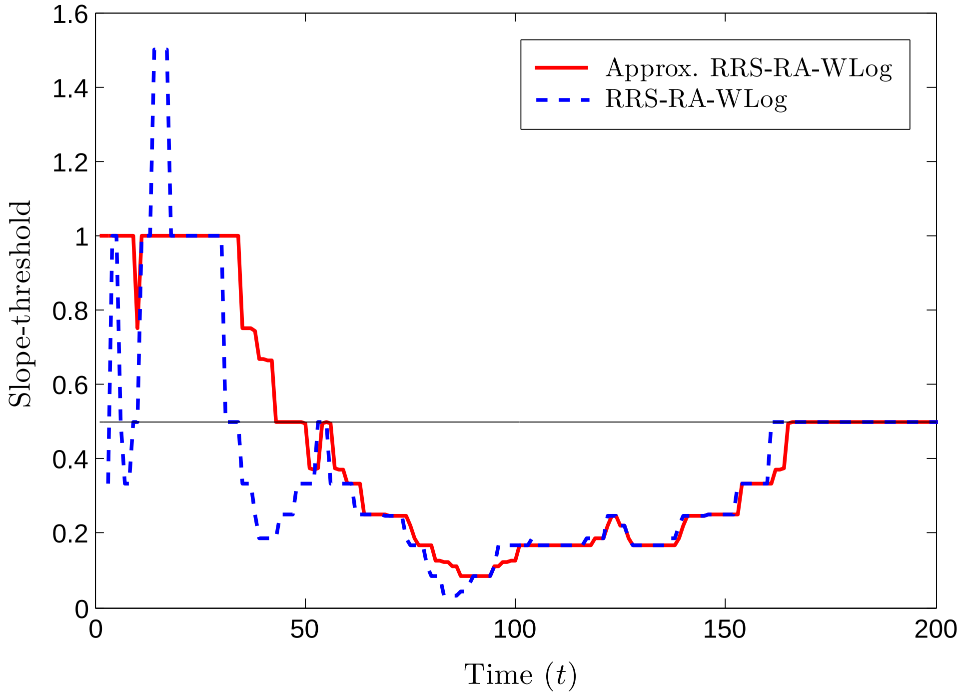

Figure 6.2: Figure plots the slope corresponding to the channel allocation $K(t)$ for the approximate RRS-RA-Wlog scheduler implementing max-min fairness. We consider a wireless system with $N = 2$ users and channel model D described in Appendix E. The thin straight line corresponds to the optimal slope for the wireless system.

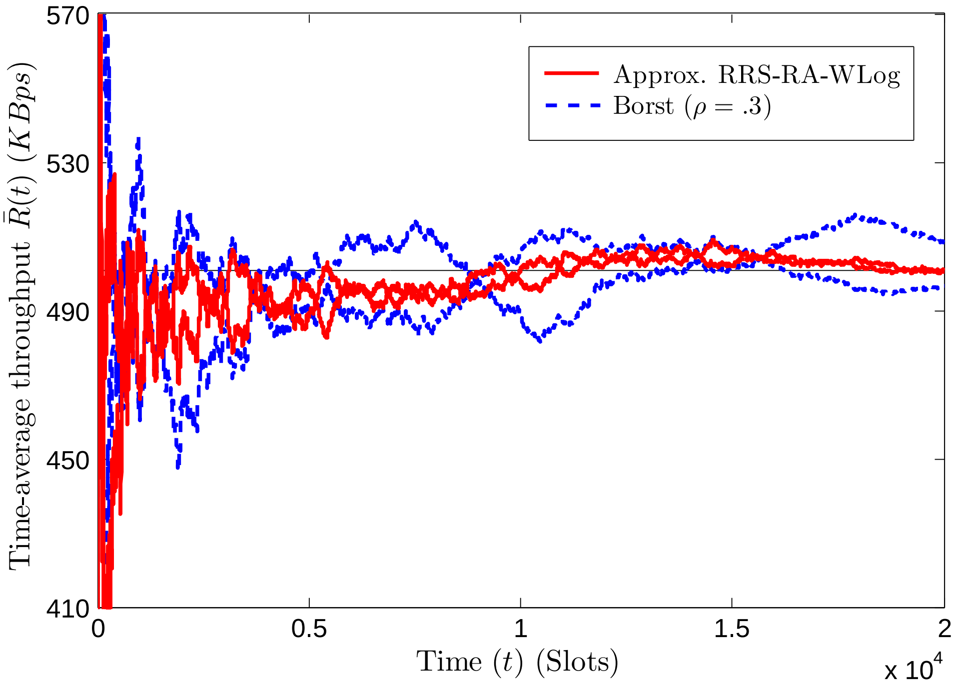

Figure 6.3: Plots of time average user throughput of the approximate RRS-RA-Wlog scheduler and a max-min fair implementation from Borst and Whiting (2001). We consider a wireless system with $N = 2$ users and the channel model D described in Appendix E. The thin straight line corresponds to the optimal rate vector for the wireless channel.

algorithm on the face of the rate region can be extended to $N > 2$ users, if a suitable ordering of channel-states (based on their *slopes*) can be found.

6.2 A Max-min Fair Scheduler for N Users

In this section, we propose a sub-optimal and an approximate implementation for max-min fairness based on the channel history and the allocated rates. We use RRS-RA-Wlog as our framework to develop the suboptimal scheduler. The proposed scheduler approximates the weights $W(t)$ in every slot t and adapts the weights in an ad hoc manner to achieve the network QoS. In every slot t,

1. The scheduler computes an estimate of the ergodic rate, $R_e(W(t-1), \pi(t))$, for the weights $W(t-1)$ and the channel distribution $\pi(t)$, where,

$$R_e(W(t-1), \pi(t)) \approx \arg \max_{\bar{R} \in \mathcal{C}(t)} \sum_i w_i(t-1) \log(\bar{r}_i)$$

2. The weights $W(\cdot)$ are then adapted based on the estimate $R_e(W(t-1), \pi(t))$ and the fairness requirement (the time average rates need to be equal for all users).
3. Schedule $\arg\max_i \frac{w_i(t) r_i(t)}{\bar{r}_i(t-1)}$ in every slot t.

We will now elaborate the above steps in detail and present the suboptimal and approximate max-min fair scheduler for N users.

Estimate of the ergodic rate vector, $R_e(W(t-1), \pi(t))$

The ergodic rate vector for the channel $\pi(t)$ and the weights $W(t-1)$ can be computed by solving the optimization problem $\arg\max_{\bar{R} \in \mathcal{C}(t)} \sum_i w_i(t-1) \log(\bar{r}_i)$. We propose a simple and a useful strategy to compute the optimizer for the strictly concave network utility using sample path optimization technique. Consider a (deterministic or a random) channel sequence with a time average channel distribution of $\pi(t)$. Then, a gradient algorithm on the channel sequence would converge to the optimizer $\arg\max_{\bar{R} \in \mathcal{C}(t)} \sum_i w_i(t-1) \log(\bar{r}_i)$. Of course, we require infinite steps to compute the exact optimizer, however, we can approximate the optimizer and obtain a $R_e(W(t-1), \pi(t))$ in fewer steps. Also, a good initial point can ensure faster convergence to the optimal rate vector.

Weight adaptation $W(t)$

We use an ad-hoc weight adaptation method, similar to the ones used in the stochastic approximation algorithms of Liu *et al.* (2003*a*),

$$w_k(t+1) = w_k(t) + \delta_k(t)\left(\frac{1}{N}\sum_i r_{e,i}(W(t),\pi(t)) - r_{e,k}(W(t),\pi(t))\right) \tag{6.1}$$

where $\delta_k(t)$ are small positive parameters. The weight adaptation is such that any difference in the ergodic rates would drive the weights of the different users appropriately to seek the max-min fair rate vector.

Even though the weight adaptation technique is similar to the one used in Liu *et al.* (2003*a*), we use $R_e(W(t),\pi(t))$ (an estimate based on the observed channel distribution) to adapt instead of the allocated rate vector $\bar{R}(t)$ used in Liu *et al.* (2003*a*).

Simulation Results

We will now present simulation results based on the suboptimal and approximate max-min fair scheduler and compare it with an implementation from Liu *et al.* (2003*a*). The estimate $R_e(W(t-1),\pi(t))$ is computed using the sample path optimization technique with 10 and 20 sample sizes in every slot t. In Figure 6.4, we plot the time average user throughputs of the approximate max-min fair scheduler and compare it with an implementation from Liu *et al.* (2003*a*). We consider a wireless system with $N = 4$ users and the channel model D described in Appendix E. In Figure 6.5, we plot the time average user throughputs of the approximate max-min fair scheduler and Liu *et al.* (2003*a*), for wireless system with $N = 10$ users and the channel model D described in Appendix E. We observe from the simulations that our approximate RRS-RA-Wlog algorithm performs comparably with the scheduler in Liu *et al.* (2003*a*).

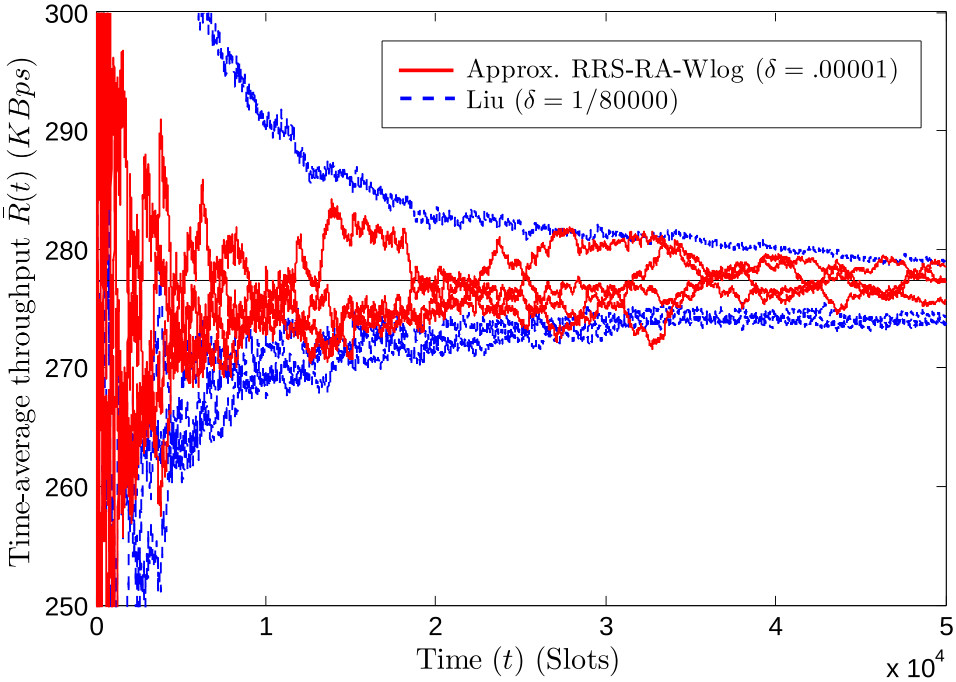

Figure 6.4: Plots of time average user throughputs of a suboptimal and approximate RRS-RA-Wlog scheduler implementing max-min fairness, and an implementation from Liu *et al.* (2003a). We consider a wireless system with $N = 4$ users and channel model D described in Appendix E. The thin straight line corresponds to the optimal rate vector for the wireless system.

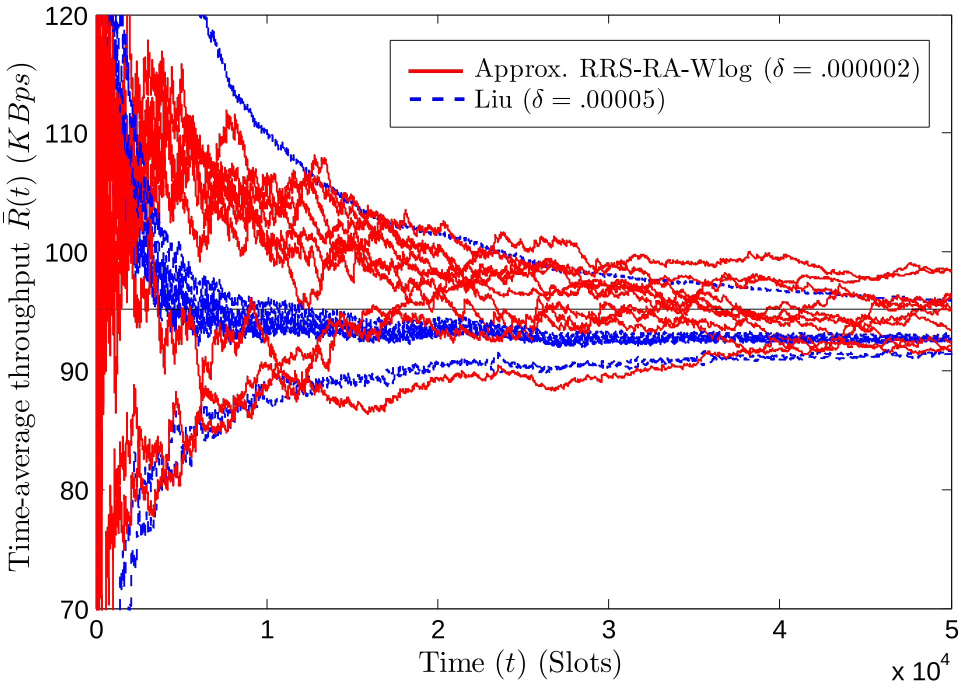

Figure 6.5: Plots of time average user throughputs of a suboptimal and approximate RRS-RA-Wlog scheduler implementing max-min fairness, and an implementation from Liu *et al.* (2003a). We consider a wireless system with $N = 4$ users and channel model D described in Appendix E. The thin straight line corresponds to the optimal rate vector for the wireless system.Equal-rates using approx.

6.3 Increasing the Convergence Rate of Schedulers using Channel history

In Section 5.5 we described the parameter-less feature of RRS-RA. A Max-min fair scheduler implementation using the utilitarian fair scheduler described in Liu *et al.* (2003*a*) is given by,

$$\mu(t) = \arg\max_i (1 + v_i^* - \frac{1}{N}\sum_j v_j^*) r_i(t) \tag{6.2}$$

where v_i^* are channel attributes estimated through the following stochastic approximation.

$$v_i(t+1) = v_i(t) + \delta(\frac{1}{N} r_{\mu(t)}(t) - I_{\{\mu(t)=i\}} r_i(t)) \tag{6.3}$$

The scheduler is asymptotically optimal for small values of the step-size parameter δ. However, as shown in figures 5.7 and 5.8, small values of δ could result in slow convergence of the scheduler. Another popular scheduler which has a similar parameter which influences convergence rate in a similar manner, is the GMR scheduler described in Andrews *et al.* (2005). The step-size parameter is a common feature of schedulers which estimate channel quantities using stochastic approximation.

Increasing the step-size, increases the rate at which the schedulers converge to their final throughputs. However, this may result in unacceptable suboptimal performance, which is again shown in Figures 5.7 and 5.8 in Section 5.5. We discussed the parameter-less RRS-RA scheduler as a better alternative in such scenarios. In this section, we present a simple method to increase the convergence rate of parameter-based schedulers, without sacrificing the optimality of the long-time average throughputs achieved. The method consists of using channel history to estimate the channel attributes used in these schedulers.

The algorithm is as follows.

1. In each slot t, simulate the channel multiple times(slots) independently and identically, with probability $\pi(t)$.
2. Calculate estimates of the channel attributes used in the parameter-based scheduler using the simulated channel instances, using sample path optimization. The parameter-based scheduler itself can be used for this estimation by fixing the step-size parameter at a small value. For instance, for the max-min fair scheduler of

Liu *et al.* (2003a), this step would calculate estimates $v_{e,i}(t)$ of v_i^* using equations 6.2 and 6.3, with a small δ.

3. Schedule slot t using the parameter-based scheduler, using the estimates obtained in the previous step.

The above algorithm uses channel history to estimate the requisite channel quantities using sample path optimization and uses it in the parameter-based scheduler. Such estimates can be calculated by using the parameter-based scheduler itself as suggested in the algorithm. The estimates can be improved by simulating the channel infinite number of times in each slot and by using a very small step-size parameter value. We observed that, by initializing the various channel attribute estimates used in the parameter-based scheduler, to the values of those estimates obtained in the previous slot, we can decrease the sample size that has to be simulated in each slot.

Simulation results

We now present simulation results on our strategy to *speed-up* the convergence rate of parameter-based schedulers using channel history. The estimates of the channel quantities used in the schedulers are computed using the sample path optimization technique with sample size 25 in every slot t. In Figure 6.6, we plot the time average user throughputs of the max-min fair scheduler from Liu *et al.* (2003a) and compare it with a version of the same scheduler *sped-up* using channel history. We consider a wireless system with $N = 4$ users and the channel model A described in Appendix E.

In Figure 6.7, we plot the time average user throughputs of the GMR scheduler from Andrews *et al.* (2005) and compare it with a version of the same scheduler *sped-up* using channel history. The wireless system has $N = 4$ users and the channel model A described in Appendix E. We observe from the simulations that the proposed channel history-based strategy increases the convergence rates of the schedulers without sacrificing the optimality of the long-time average throughputs.

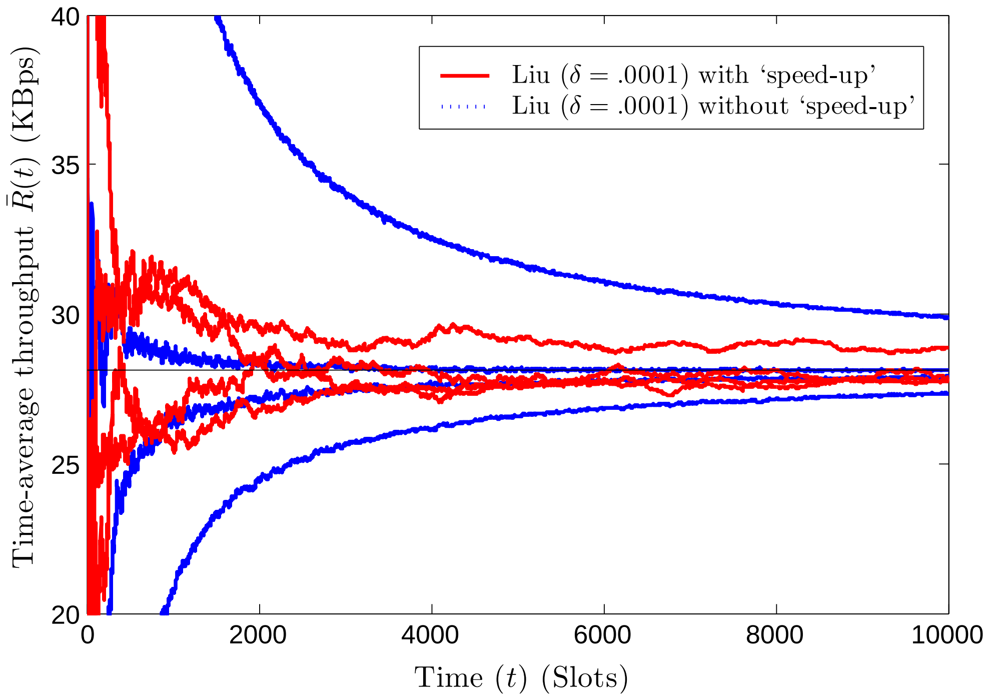

Figure 6.6: Plots of time average user throughputs of a Max-min fair scheduler from Liu *et al.* (2003*a*) and a *sped-up* version of the same scheduler. We consider a wireless system with $N = 4$ users and channel model A described in Appendix E. The thin straight line corresponds to the optimal rate vector for the wireless system.

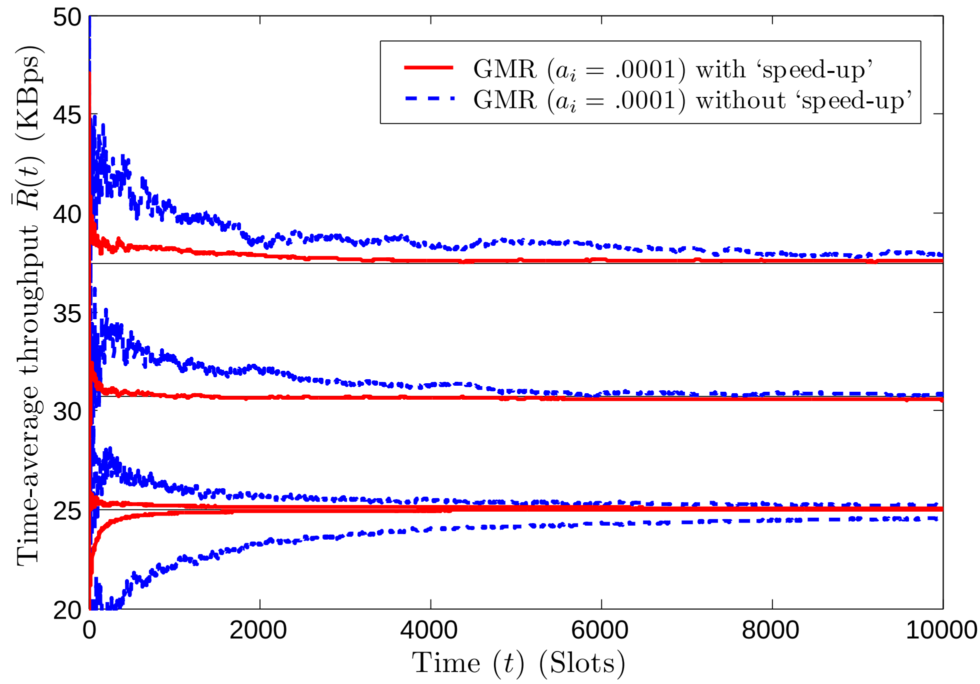

Figure 6.7: Plots of time average user throughputs of the GMR scheduler from Andrews *et al.* (2005) and a *sped-up* version of the same scheduler. We consider a wireless system with $N = 4$ users and channel model A described in Appendix E. The thin straight lines correspond to the optimal rate vector for the wireless system.

CHAPTER 7

Conclusion

In this thesis, we have studied three general scheduling strategies for cellular downlink wireless channel, RRS, RRS-CA and RRS-RA, based on an estimate of the rate region of the wireless channel. The schedulers use channel history to estimate the wireless channel statistics and the rate region of the wireless channel. The schedulers then attempt to implement a network quality of service or maximize a network utility on the estimated rate region. The schedulers seek and attempt to implement the global optimal rate vector in every slot (instead of the seeking the local optima, like in network utility based gradient schedulers), and hence, can implement non-concave network utilities and arbitrary network QoS as well.

There was a lack of general viewpoint on the design of wireless schedulers, requiring different frameworks for different network objectives. Stochastic gradient type algorithms were popular for continuous and concave network utilities. Schedulers of other simple QoS used ad hoc, adaptive learning strategies. We have studied a general scheduling methodology using an estimate of the rate region that can implement arbitrary notions of QoS and fairness defined on the long time average user throughputs.

The rate region based scheduler, RRS, uses the entire available channel history to estimate the wireless channel statistics. The estimate of the wireless channel is then used to identify a rate region and an optimal rate vector in the rate region. The optimal rate vector is obtained by maximizing the network utility in the estimated rate region. The utility at the optimal rate vector thus identified usually serves an upper bound on the network performance for the sample path and is achievable asymptotically for ergodic channels. The RRS scheduler implements the schedule corresponding to the identified optimal rate vector in every slot. We have proved that the rate region based scheduler RRS is asymptotically optimal for all continuous network utilities, for some ergodic channels with discrete channel states. We note that the channel history based scheduler requires a consistency in implementation and also has poor convergence behaviour.

Then, we proposed a simple *channel-allocation* based scheduler, RRS-CA, that uses schedule history in addition to channel history. We have proved the asymptotic optimality of RRS-CA as well, for all continuous network utilities, for some ergodic channels with discrete channel states. While RRS-CA has better convergence behaviour than RRS, we note that both RRS and RRS-CA requires consistency in the schedules implemented across time slots.

We then proposed a practical variant of RRS, called RRS-RA, that uses channel history and schedule history in the form of allocated rates, to identify a user to schedule. RRS-RA is a gradient based scheduler that uses a dynamic, auxiliary network utility to drive the allocated rate vector towards the optimal rate vector. RRS-RA has several practical advantages as it makes consistent schedule decisions and handles sample path variations better than RRS. We have discussed two implementations of RRS-RA, using a weighted logarithmic utility and a Euclidean utility, suited for non-Pareto optimal rate vectors as well.

We have studied the performance of the rate region based schedulers, RRS and RRS-RA, using simulations, for a variety of network scenarios. We observe that the rate region based strategy provides us with a general framework to implement arbitrary continuous and some non-continuous utilities, provides us a parameter-less implementation of the network QoS and has better convergence behaviour than network utility based gradient schedulers. The RRS schedulers permit implementation using an exponentially smoothed time averages and are extremely useful for dynamic channel and network scenarios. Also, we observed that the schedule history and channel history based schedulers, RRS-CA and RRS-RA, reported better convergence than channel history based scheduler RRS.

Then, we discussed simple implementations of rate region based max-min fair scheduler, for two user case and for N users. The RRS based schedulers are computationally expensive to implement, as we need to estimate the rate region, identify the optimal rate vector and the corresponding schedule in every slot. Hence, we discuss two techniques to minimize the computation needed to implement rate region based schedulers. For the two user case, we propose to search on the boundary of the rate region in steps defined by the slopes of the boundary. For the N user case, we approximate the ergodic rate vector and the ergodic schedule to minimize computation. In either case, the global optimal rate vector is estimated using the channel statistics (obtained from the channel history) and the scheduler seeks to drive the average allocated rate vector towards the global optimal rate vector. Finally, we presented a simple method, that uses channel history to increase the convergence rate of some popular schedulers in literature.

7.1 Future work

In this thesis, we have developed a powerful class of schedulers that use channel history and schedule history to implement general notions of QoS and fairness on long time average user throughputs. We have proposed the rate region based schedulers for a fixed number of saturated users and for ideal channel conditions (perfect CSI, no data error, etc). We believe that the rate region based scheduling strategy is appropriate for a number of other useful and interesting network scenarios. We list a few problems of interest here.

1) Develop simple and computationally inexpensive implementations of rate region based based schedulers for important and useful network utilities and QoS.

2) Develop channel history and schedule history based schedulers for unsaturated traffic models and for other performance metrics.

3) Use of channel history and schedule history to implement user admission control and to implement non-Pareto optimal rate vectors efficiently.

4) Develop advanced rate region based schedulers specific to the channel models and network conditions (e.g., for Markovian channels and for structured traffic)

5) RRS-RA schedulers propose one strategy to use the channel history and the schedule history information. An interesting problem would be to identify optimal use of channel history and schedule history.

6) There is very little interest or literature on scheduling strategies for non-ergodic channels. We believe that these class of problems are useful for mobile data users and for web browsing, and rate region based schedulers can be a useful tool in this scenario.

APPENDIX A

Proof of Asymptotic Optimality of RRS

Let the channel process $\{R(t)\}_{t=1}^{\infty}$ be an irreducible, discrete time Markov chain with a finite state space $\{\mathbf{R}_1, \mathbf{R}_2, \ldots \mathbf{R}_m\}$. Let $\pi = \{\pi_1, \pi_2, \cdots, \pi_m\}$ be the stationary distribution on the state space and let $\mathcal{C}$ be the rate region of the wireless channel. We assume that the network utility $U : \mathcal{R}^N \to \mathcal{R}^1$, defined on the rate region $\mathcal{C}$, is a continuous function and that there exists a unique optimal stationary schedule $\{a^*_{j,i}\}$ such that,

$$\bar{r}^*_i = \sum_{j=1}^{m} \pi_j \; a^*_{j,i} \; r_{j,i}$$

where $\bar{R}^* = \arg\max_{\bar{R} \in \mathcal{C}} U(\bar{R})$.

Let $\pi(t) = \{\pi_j(t)\}$ be the observed channel distribution at time t and let $\mathcal{C}(t)$ be the finite time rate region of the wireless channel. An irreducible, discrete time Markov chain with a finite state space is positive (see Wolff (1989), Chapter 3). Hence, $\pi(t) \to \pi$ a.s. and $\mathcal{C}(t) \to \mathcal{C}$ a.s. as well (see Chapter 3). Let $\bar{R}^*(t)$ be an optimal rate vector in the finite time rate region $\mathcal{C}(t)$ for the network utility $U(\cdot)$, and let $\{a^*_{j,i}(t)\}$ be an optimal channel allocation (or stationary schedule) that achieves $\bar{R}^*(t)$. RRS implements the stationary schedule $\{a^*_{j,i}(t)\}$ in slot t, i.e., allocates channel $R(t)$ to a user based on the probabilities $\{a^*_{R(t),1}, a^*_{R(t),2}, \cdots, a^*_{R(t),N}\}$, independent of the previous schedules.

For convenience, we will denote the rate corresponding to a stationary distribution π and a stationary schedule $\{a_{j,i}\}$ by $\bar{R}(\pi, \{a_{j,i}\})$. For example, $\bar{R}(\pi, \{a^*_{j,i}\}) = \bar{R}^*$ and $\bar{R}(\pi(t), \{a^*_{j,i}(t)\}) = \bar{R}^*(t)$. Then, $U(\bar{R}(\pi, \{a_{j,i}\}))$ will denote the utility achieved with the rate vector $\bar{R}(\pi, \{a_{j,i}\})$. The maximum utility achievable for the wireless channel is $U(\bar{R}(\pi, \{a^*_{j,i}\}))$.

Lemma A.0.1. *Let $\{R(t)\}_{t=1}^{\infty}$ be an irreducible, discrete time Markov chain with a finite state space. Let $U : \mathcal{R}^N \to \mathcal{R}^1$ be a continuous network utility, and let the optimal stationary schedule $\{a^*_{j,i}\}$ for the channel and the network utility be unique. Then, the sequence of schedules $\{\{a^*_{j,i}(t)\}\}_{t=1}^{\infty}$ of RRS converges to the optimal stationary* *schedule* $\{a^*_{j,i}\}$ *almost surely, i.e.,*

$$\{a^*_{j,i}(t)\} \rightarrow \{a^*_{j,i}\} \text{ a.s.}$$

Proof. An irreducible, discrete time Markov chain with a finite state space is positive (see Wolff (1989), Chapter 3). Hence, $\pi(t) \rightarrow \pi$ (i.e., $\pi_j(t) \rightarrow \pi_j$ for all $j = 1, 2, \cdots, m$) almost surely. Consider a realization ω (sample path) with channel sequence $\{R(t, \omega)\}$ such that $\pi(t, \omega) \rightarrow \pi$ (i.e., $\pi_j(t, \omega) \rightarrow \pi_j$ for all $j = 1, 2, \cdots, m$).

Consider a sequence of channel schedules $\{\{a^*_{j,i}(t, \omega)\}\}_{t=1}^{\infty}$, for RRS, for the sample path. For any ω and for all t, $0 \leq a^*_{j,i}(t, \omega) \leq 1$, for all $j = 1, 2, \cdots, m$ and $i = 1, 2, \cdots, N$. Hence, $\{a^*_{j,i}(t, \omega)\} \in [0, 1]^{m \times N}$ for all t, i.e., the sequence of stationary schedules $\{\{a^*_{j,i}(t, \omega)\}\}_{t=1}^{\infty}$ lie in $[0, 1]^{m \times N}$. Hence, there exists limit points of $\{\{a^*_{j,i}(t, \omega)\}\}_{t=1}^{\infty}$, all in $[0, 1]^{m \times N}$ (see Rudin (1976)). Let $\{a'_{j,i}(\omega)\}$ be a limit point of the channel schedule sequence $\{\{a^*_{j,i}(t, \omega)\}\}_{t=1}^{\infty}$. We will now prove that $\{a'_{j,i}(\omega)\}$ is in fact $\{a^*_{j,i}\}$.

Suppose that $\{a'_{j,i}(\omega)\} \neq \{a^*_{j,i}\}$. Consider a subsequence of channel schedules $\{\{a^*_{j,i}(t_k, \omega)\}\}_{t_k}$, $\{t_k\} \subset \{1, 2, \cdots\}$ such that $\{a^*_{j,i}(t_k, \omega)\} \rightarrow \{a'_{j,i}(\omega)\}$ as $t_k \rightarrow \infty$. Clearly, $\pi(t_k, \omega) \rightarrow \pi$ as $t_k \rightarrow \infty$. We now have,

$$U(\bar{R}(\pi(t_k, \omega), a^*_{j,i}(t_k, \omega))) \geq U(\bar{R}(\pi(t_k, \omega), a^*_{j,i})) \text{ (by the definition of RRS)}$$

Taking limit $t_k \rightarrow \infty$, we have,

$$\begin{aligned} \lim_{t_k \rightarrow \infty} U(\bar{R}(\pi(t_k, \omega), a^*_{j,i}(t_k, \omega))) &\geq \lim_{t_k \rightarrow \infty} U(\bar{R}(\pi(t_k, \omega), a^*_{j,i})) \\ \Rightarrow U(\bar{R}(\pi, a'_{j,i}(\omega))) &\geq U(\bar{R}(\pi, a^*_{j,i})) \text{ (since } U, \bar{R} \text{ are continuous)} \end{aligned}$$

This is a contradiction, since $\{a^*_{j,i}\}$ is the unique stationary schedule that maximizes the network utility U over the rate region $\mathcal{C}$ (corresponding to the distribution π). Hence, $\{a'_{j,i}(\omega)\} = \{a^*_{j,i}\}$, or, the sequence of channel schedules of RRS converges to the optimal schedule $\{a^*_{j,i}\}$ (whenever $\pi(t, \omega) \rightarrow \pi$). Since $\Pr(\{\omega : \pi(t, \omega) \rightarrow \pi\}) = 1$, it follows that $\{\{a^*_{j,i}(t)\}\}_{t=1}^{\infty} \rightarrow \{a^*_{j,i}\}$ almost surely. □

Now, we state and prove the main theorem concerning the asymptotic optimality of RRS.

Theorem A.0.1. *Let $\{R(t)\}_{t=1}^{\infty}$ be an irreducible, finite state space, discrete time Markov chain with a stationary distribution $\{\pi_j\}$. Let $U : \mathcal{R}^N \rightarrow \mathcal{R}^1$ be a continuous network utility. Let $\mathcal{C}$ be the rate region of the wireless channel and let $\bar{R}^*$ be the optimal rate vector such that $\bar{R}^* = \arg\max_{\bar{R}\in\mathcal{C}} U(\bar{R})$. Let the optimal stationary schedule $\{a^*_{j,i}\}$ that achieves the optimal rate vector $\bar{R}^*$ be unique. Then, the time average throughput achieved using RRS converges to the optimal rate vector $\bar{R}^*$ almost surely,*

$$\bar{R}(t) \rightarrow \bar{R}^* \text{ a.s.}$$

i.e., RRS is asymptotically optimal.

Proof. We will prove the optimality of RRS using an implementation of the algorithm.

RRS implements the stationary schedule $\{a^*_{j,i}(t)\}$ in slot t by allocating channel $R(t)$ to a user based on the probabilities $\{a^*_{R(t),1}(t), a^*_{R(t),2}(t), \cdots, a^*_{R(t),N}(t)\}$, independent of the previous schedules. For example, let $I(t)$ be a Uniform random variable in $[0,1)$ chosen independently, then, RRS will schedule user i if,
$I(t) \in \Big[\sum_{k=1}^{i-1} a^*_{R(t),k}(t), \sum_{k=1}^{i} a^*_{R(t),k}(t)\Big)$.

Let $\{I_j(t)\}_{t=1}^{\infty}$ be a sequence of i.i.d. random variables with a Uniform distribution in $[0,1)$, for all $j = 1, 2, \cdots, m$. Let $M_j(t) := \sum_{s=1}^{t} 1_{\{R(s)=\mathbf{R}_j\}}$ be the count of the number of occurrences of channel $\mathbf{R}_j$ up to time t. We will assume that RRS schedules user i at time t (i.e., allocates channel $R(t)$ at time t to user i) if $\sum_{k=1}^{i-1} a^*_{R(t),k}(t) \leq I_{R(t)}(M_{R(t)}(t)) < \sum_{k=1}^{i} a^*_{R(t),k}(t)$ (i.e., we assume that the sequence of i.i.d. Uniform random variables $\{I_j(t)\}_{t=1}^{\infty}$ is used to identify an user to schedule whenever channel $\mathbf{R}_j$ occurs).

Consider $\{I_j(t)\}_{t=1}^{\infty}$. For any rational $q, q \in \mathcal{Q} \cap [0,1]$, from the strong law of large numbers, we have,

$$\mathsf{Pr}\left(\left\{\omega : \frac{1}{t}\sum_{s=1}^{t} 1_{\{I_j(s,\omega)<q\}} \rightarrow q\right\}\right) = 1$$

Taking a countable intersection of probability one sets, for every $q \in \mathcal{Q} \cap [0,1]$, we have,

$$\mathsf{Pr}\left(\left\{\omega : \frac{1}{t}\sum_{s=1}^{t} 1_{\{I_j(s,\omega)<q\}} \to q, q \in \mathcal{Q} \cap [0,1]\right\}\right) = 1$$

Using a continuity argument, we get,

$$\mathsf{Pr}\left(\left\{\omega : \frac{1}{t}\sum_{s=1}^{t} 1_{\{I_j(s,\omega)<q\}} \to q, q \in [0,1]\right\}\right) = 1$$

Taking a finite intersection of the above (probability one) sets, for all channel states j, we have,

$$\mathsf{Pr}\left(\left\{\omega : \frac{1}{t}\sum_{s=1}^{t} 1_{\{I_j(s,\omega)<q\}} \to q, q \in [0,1], j = 1, 2, \cdots, m\right\}\right) = 1 \tag{A.1}$$

From LemmaA.0.1, we have

$$\mathsf{Pr}\left(\{\omega : \pi(t,\omega) \to \pi, \{a^*_{j,i}(t,\omega)\} \to \{a^*_{j,i}\}\}\right) = 1 \tag{A.2}$$

Combining the results from equations (A.1) and (A.2), we have,

$$\mathsf{Pr}\left(\left\{\omega : \pi(t,\omega) \to \pi, \{a^*_{j,i}(t,\omega)\} \to \{a^*_{j,i}\}, \frac{1}{t}\sum_{s=1}^{t} 1_{\{I_j(s,\omega)<q\}} \to q, q \in [0,1], \forall j\right\}\right) = 1 \tag{A.3}$$

Now, consider a realization ω (sample path) from the above set, such that, $\pi(t,\omega) \to \pi$, $\{a^*_{j,i}(t,\omega)\} \to \{a^*_{j,i}\}$ and $\frac{1}{t}\sum_{s=1}^{t} 1_{\{I_j(s,\omega)<q\}} \to q$ for all $q \in [0,1]$ and for all $j = 1, 2, \cdots, m$. For such an ω, we will now show that the time average rate $\bar{R}(t,\omega) \to \bar{R}^*$.

Without loss of generality, consider user 1. For a channel sequence $\{R(t)\}_{t=1}^{\infty}$, user 1 is scheduled in slot t if $0 \le I_{R(t)}(M_{R(t)}(t)) < a^*_{R(t),1}(t)$, where $M_{R(t)} := \sum_{s=1}^{t} I_{\{R(s)=R(t)\}}$ is the number of occurrences of the channel state "$R(t)$" up to time t. We note again that the sequence of random variables used to schedule user 1, in any channel state j, is an i.i.d. sequence. The time average rate of user 1 for the realization ω is now given by,

$$\bar{r}_1(t,\omega) \quad = \quad \frac{1}{t}\sum_{s=1}^{t}\sum_{j=1}^{m} 1_{\{R(s,\omega)=\mathbf{R}_j\}} 1_{\{\mu(s,\omega)=1\}} \mathbf{r}_{j,1}$$

Rearranging terms, we get,

$$\bar{r}_1(t,\omega) \;=\; \sum_{j=1}^{m}\left(\frac{1}{t}\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}} 1_{\{\mu(s,\omega)=1\}}\right)\mathbf{r}_{j,1}$$

Rewriting the schedule $\mu(s,\omega)$ in terms of the uniform random variables, we get,

$$\bar{r}_1(t,\omega) \;=\; \sum_{j=1}^{m}\left(\frac{1}{t}\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}} 1_{\{I_j(M_j(s,\omega),\omega)<a^*_{j,1}(s,\omega)\}}\right)\mathbf{r}_{j,1}$$

Multiplying and dividing each term (inside the parenthesis) by $\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}$, we get,

$$\bar{r}_1(t,\omega) \;=\; \sum_{j=1}^{m}\left(\frac{1}{t}\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}} 1_{\{I_j(M_j(s,\omega),\omega)<a^*_{j,1}(s,\omega)\}}\frac{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}}{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}}\right)\mathbf{r}_{j,1}$$

Rearranging terms within the parenthesis, we have,

$$\begin{aligned}
\bar{r}_1(t,\omega) \;&=\; \sum_{j=1}^{m}\left(\frac{1}{t}\frac{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}} 1_{\{I_j(M_j(s,\omega),\omega)<a^*_{j,1}(s,\omega)\}}}{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}}\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}\right)\mathbf{r}_{j,1} \\
&=\; \sum_{j=1}^{m}\left(\frac{1}{t}\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}\frac{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}} 1_{\{I_j(M_j(s,\omega),\omega)<a^*_{j,1}(s,\omega)\}}}{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}}\right)\mathbf{r}_{j,1}
\end{aligned}$$

Taking limit $t \to \infty$, we have,

$$\begin{aligned}
\lim_{t\to\infty}\bar{r}_1(t,\omega) \;&=\; \lim_{t\to\infty}\sum_{j=1}^{m}\left(\frac{1}{t}\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}\frac{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}} 1_{\{I_j(M_j(s,\omega),\omega)<a^*_{j,1}(s,\omega)\}}}{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}}\right)\mathbf{r}_{j,1} \\
&=\; \sum_{j=1}^{m}\lim_{t\to\infty}\left(\frac{1}{t}\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}\frac{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}} 1_{\{I_j(M_j(s,\omega),\omega)<a^*_{j,1}(s,\omega)\}}}{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}}\right)\mathbf{r}_{j,1}
\end{aligned} \tag{A.4}$$

We will now evaluate the limit of the expression inside the paranthesis in equation (A.4). Since $\pi(t,\omega) \to \pi$ (by hypothesis), we have, for all j,

$$\lim_{t\to\infty}\frac{1}{t}\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}} = \pi_j \tag{A.5}$$

Now, we will compute the limit of the other term in the above expression,

$$\lim_{t\to\infty} \frac{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}} 1_{\{I_j(M_j(s,\omega),\omega)<a_{j,1}^*(s,\omega)\}}}{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}}$$

Fix an $\epsilon > 0$. By hypothesis, $\{a_{j,i}^*(t,\omega)\} \to \{a_{j,i}^*\}$, hence, there exists a $T_\epsilon(\omega) \in \{1, 2, \cdots\}, T_\epsilon(\omega) < \infty$ such that $|a_{j,i}^*(t,\omega) - a_{j,i}^*| < \epsilon$ for all $t > T_\epsilon(\omega)$. We will now rewrite the above limit in two terms (for large t) as,

$$\lim_{t\to\infty} \Big(\frac{\sum_{s=1}^{T_\epsilon(\omega)} 1_{\{R(s,\omega)=\mathbf{R}_j\}} 1_{\{I_j(M_j(s,\omega),\omega)<a_{j,1}^*(s,\omega)\}}}{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}} + \frac{\sum_{s=T_\epsilon(\omega)+1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}} 1_{\{I_j(M_j(s,\omega),\omega)<a_{j,1}^*(s,\omega)\}}}{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}} \Big) \tag{A.6}$$

The denominator term in the above expression, $\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}$ is $M_j(t,\omega)$, the count of the number of occurrences of channel state $\mathbf{R}_j$ up to time t in the sample path. By hypothesis, $\lim_{t\to\infty} \frac{1}{t} \sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}} = \pi_j$ and for a positive DTMC, $\pi_j > 0$ for all j (see Wolff (1989), Chapter 3). Then, $\lim_{t\to\infty} \frac{M_j(t,\omega)}{t} = \pi_j > 0$ implies that $M_j(t,\omega) \to \infty$ as $t \to \infty$. $M_j(t,\omega) \to \infty$ means that the channel state $\mathbf{R}_j$ occurs infinitely often.

The limit of the first term in the expression (A.6) evaluates as

$$\lim_{t\to\infty} \frac{\sum_{s=1}^{T_\epsilon(\omega)} 1_{\{R(s,\omega)=\mathbf{R}_j\}} 1_{\{I_j(M_j(s,\omega),\omega)<a_{j,1}^*(s,\omega)\}}}{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}} = 0$$

since $\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}} \to \infty$ and the numerator is bounded above by $T_\epsilon(\omega)$. Now, consider the second term in the expression (A.6),

$$\frac{\sum_{s=T_\epsilon(\omega)+1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}} 1_{\{I_j(M_j(s,\omega),\omega)<a_{j,1}^*(s,\omega)\}}}{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}}$$

For $s > T_\epsilon(\omega)$, $a_{j,1}^*(s,\omega) \geq a_{j,1}^* - \epsilon$ (by assumption). Hence, the above term is lower bounded by,

$$\frac{\sum_{s=T_\epsilon(\omega)+1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}} 1_{\{I_j(M_j(s,\omega),\omega)<a_{j,1}^*-\epsilon\}}}{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}}$$

Rewriting the above expression in terms of $M_j(t,\omega)$, and simplifying it, we have,

$$\frac{\sum_{s=M_j(T_\epsilon(\omega)+1,\omega)}^{M_j(t,\omega)} 1_{\{I_j(s,\omega)<a^*_{j,1}-\epsilon\}}}{M_j(t,\omega)}$$

Taking limit $t \to \infty$, hence, $M_j(t,\omega) \to \infty$ and by the hypothesis (in equations (A.1) and (A.3)), we have,

$$\lim_{t\to\infty} \frac{\sum_{s=M_j(T_\epsilon(\omega)+1,\omega)}^{M_j(t,\omega)} 1_{\{I_j(s,\omega)<a^*_{j,1}-\epsilon\}}}{M_j(t,\omega)} = a^*_{j,1} - \epsilon$$

Hence,

$$\lim_{t\to\infty} \frac{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}} 1_{\{I_j(M_j(s,\omega),\omega)<a^*_{j,1}(s,\omega)\}}}{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}} \geq a^*_{j,1} - \epsilon \tag{A.7}$$

Substituting (A.5) and (A.7) in (A.4), we get,

$$\lim_{t\to\infty} \bar{r}_1(t,\omega) \;\; \geq \;\; \sum_j \pi_j \; (a^*_{j,1} - \epsilon) \; \mathbf{r}_{j,1}$$

Similarly, we can show that

$$\lim_{t\to\infty} \bar{r}_1(t,\omega) \;\; \leq \;\; \sum_j \pi_j \; (a^*_{j,1} + \epsilon) \; \mathbf{r}_{j,1}$$

Since, ϵ is arbitrary and $\mathbf{r}_{j,1}$ is bounded (for all j), we have,

$$\lim_{t\to\infty} \bar{r}_1(t,\omega) = \sum_j \pi_j \; a^*_{j,1} \; \mathbf{r}_{j,1} = \bar{r}^*_1$$

Thus, for every ω measured in equation (A.3), we have $\bar{r}_1(t,\omega) \to \bar{r}^*_1$, i.e., $\bar{r}_1(t) \to \bar{r}^*_1$ almost surely. User 1 was an arbitrary choice, and hence, we have $\bar{R}(t) \to \bar{R}^*$ almost surely, or, RRS is asymptotically optimal. $\square$

APPENDIX B

Proof of Asymptotic Optimality of RRS-CA

Let the channel process $\{R(t)\}_{t=1}^{\infty}$ be an irreducible, discrete time Markov chain with a finite state space $\{\mathbf{R}_1, \mathbf{R}_2, \dots \mathbf{R}_m\}$. Let $\pi = \{\pi_1, \pi_2, \cdots, \pi_m\}$ be the stationary distribution on the state space and let $\mathcal{C}$ be the rate region of the wireless channel. We assume that the network utility $U : \mathcal{R}^N \to \mathcal{R}^1$, defined on the rate region $\mathcal{C}$, is a continuous function and that there exists a unique optimal stationary schedule $\{a^*_{j,i}\}$ such that,

$$\bar{r}^*_i = \sum_{j=1}^{m} \pi_j \; a^*_{j,i} \; r_{j,i}$$

where $\bar{R}^* = \arg\max_{\bar{R}\in\mathcal{C}} U(\bar{R})$.

Let $\pi(t) = \{\pi_j(t)\}$ be the observed channel distribution at time t and let $\mathcal{C}(t)$ be the finite time rate region of the wireless channel. An irreducible, discrete time Markov chain with a finite state space is positive (see Wolff (1989), Chapter 3). Hence, $\pi(t) \to \pi$ a.s. and $\mathcal{C}(t) \to \mathcal{C}$ a.s. as well (see Chapter 3). Let $\bar{R}^*(t)$ be an optimal rate vector in the finite time rate region $\mathcal{C}(t)$ for the network utility $U(\cdot)$, and let $\{a^*_{j,i}(t)\}$ be an optimal channel allocation (or stationary schedule) that achieves $\bar{R}^*(t)$. Let $\{\hat{a}_{j,i}(t)\}$ be the actual channel allocation up to time t. Then, the scheduling strategy of the channel allocation based RRS-CA scheduler is given by,

$$\mu(t) = \arg\max_{i} \left\{ a^*_{R(t),i}(t) - \hat{a}_{R(t),i}(t-1) \right\} \qquad \text{(B.1)}$$

The following theorem proves the asymptotic optimality of the channel allocation based RRS-CA scheduler.

Theorem B.0.2. *Let $\{R(t)\}_{t=1}^{\infty}$ be an irreducible, finite state space, discrete time Markov chain with a stationary distribution $\{\pi_j\}$. Let $U : \mathcal{R}^N \to \mathcal{R}^1$ be a continuous network utility. Let $\mathcal{C}$ be the rate region of the wireless channel and let $\bar{R}^*$ be the optimal rate vector such that $\bar{R}^* = \arg\max_{\bar{R}\in\mathcal{C}} U(\bar{R})$. Let the optimal stationary schedule $\{a^*_{j,i}\}$ that achieves the optimal rate vector $\bar{R}^*$ be unique. Then, the time average throughput* *achieved using RRS-CA converges to the optimal rate vector $\bar{R}^*$ almost surely,*

$$\bar{R}(t) \to \bar{R}^* \textit{ a.s.}$$

i.e., RRS-CA is asymptotically optimal.

Proof. Consider a realization ω (sample path) such that $\pi(t,\omega) \to \pi$ and $\{a^*_{j,i}(t,\omega)\} \to \{a^*_{j,i}\}$. From Appendix A, Lemma A.0.1, we know that

$$\mathsf{Pr}\left(\left\{\omega : \pi(t,\omega) \to \pi, \{a^*_{j,i}(t,\omega)\} \to \{a^*_{j,i}\}\right\}\right) = 1 \tag{B.2}$$

Define $l_j(t,\omega)$ as,

$$l_j(t,\omega) := \sum_{i=1}^{N} \left(a^*_{j,i} - \hat{a}_{j,i}(t,\omega)\right)^+$$

where $\hat{a}_{j,i}(t,\omega) = \frac{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}} 1_{\{\mu(s,\omega)=i\}}}{\sum_{s=1}^{t} 1_{\{R(s,\omega)=\mathbf{R}_j\}}}$ is the fraction of channel $\mathbf{R}_j$ allocated to user i up to time t in the realization ω and $(x)^+ = \max(x,0)$. Define $L_j(t,\omega) := \{i : a^*_{j,i} - \hat{a}_{j,i}(t,\omega) \geq 0\}$. Then, $l_j(t,\omega) = \sum_{i \in L_j(t,\omega)} (a^*_{j,i} - \hat{a}_{j,i}(t,\omega))$. $l_j(t,\omega)$ measures a certain "lag" in the service provided to the users for the channel state j, at time t. $l_j(t,\omega) \geq 0$ for all t and for all ω. Further, $l_j(t,\omega) = 0$ only if the actual channel allocation $\{\hat{a}_{j,i}(t,\omega)\}$ matches the optimal channel allocation $\{a^*_{j,i}(t,\omega)\}$ for channel j (at time t and realization ω).

The optimal rate vector $\bar{R}^*$ and the actual allocated time average rate vector $\bar{R}(t)$ are defined as $\bar{r}^*_i = \sum_j \pi_j \; a^*_{j,i} \; r_{j,i}$ and $\bar{r}_i(t) = \sum_j \pi_j(t) \; \hat{a}_{j,i}(t) \; r_{j,i}$ (see Chapter 4, Section 4.1), hence, to prove the asymptotic optimality of RRS-CA, it is sufficient to show that $\{\hat{a}_{j,i}(t)\} \to \{a^*_{j,i}\}$ almost surely, i.e., $\{\hat{a}_{j,i}(t,\omega)\} \to \{a^*_{j,i}\}$ for all ω measured in (B.2). We will now show that $\{\hat{a}_{j,i}(t,\omega)\} \to \{a^*_{j,i}\}$ by showing that $\sum_j l_j(t,\omega) \to 0$.

A finite state space, irreducible, discrete time Markov chain is positive (see Wolff (1989), Chapter 3). Hence, $\pi_j > 0$ for all j. Hence, for all ω measured in equation (B.2), every channel j occurs infinitely often (since $\pi_j(t,\omega) \to \pi_j > 0$). Without loss of generality, consider the channel state $\mathbf{R}_1$. Let $\{t_1, t_2, \cdots\} \subset \{1, 2, \cdots\}$ correspond to the subsequence of the natural numbers such that $R(t_k,\omega) = \mathbf{R}_1$; t_k marks the k^{th} occurrence of channel state $\mathbf{R}_1$ in the sequence $\{R(t,\omega)\}_{t=1}^{\infty}$. Note that $l_1(t,\omega)$ changes only at channel instants such that $R(t,\omega) = \mathbf{R}_1$ and remains unchanged otherwise (i.e.,

$l_1(t,\omega) = l_1(t_k,\omega)$ for all $t_k < t < t_{k+1}$).

Fix an $\epsilon > 0$. If $\{a^*_{j,i}(t,\omega)\} \to \{a^*_{j,i}\}$, then there exists a $T_\epsilon(\omega) < \infty$ such that $|a^*_{j,i}(t,\omega) - a^*_{j,i}| < \epsilon$ for all $t > T_\epsilon(\omega)$, for all j and i. Hence, $|a^*_{1,i}(t,\omega) - a^*_{1,i}| < \epsilon$ for all $t > T_\epsilon(\omega)$.

Consider an integer K such that $\frac{1}{K} < \epsilon$, $t_K > T_\epsilon(\omega)$ and $l_1(t_K,\omega) > 3N\epsilon$, where N is the number of users in the system. t_K corresponds to the time of the K^{th} occurrence of the channel state $\mathbf{R}_1$ in ω. $l_1(t_K,\omega) > 3N\epsilon$ implies that the lag in the service for channel state $\mathbf{R}_1$ at time t_K (at the K^{th} occurrence of the channel $\mathbf{R}_1$) is greater than $3N\epsilon$. Trivially, if there does not exist such a K, then $l_1(t,\omega) \leq 3N\epsilon$ for all large t, and since ϵ is arbitrary, this implies that $l_1(t,\omega) \to 0$ (i.e., $\{a^*_{1,i}(t,\omega)\} \to \{a^*_{1,i}\}$ or RRS-CA is asymptotically optimal).

Let $i' = \arg\max_i \left\{a^*_{1,i} - \hat{a}_{1,i}(t_K,\omega)\right\}$. Since $l_1(t_K,\omega) > 3N\epsilon$, we have,

$$a^*_{1,i'} - \hat{a}_{1,i'}(t_K,\omega) > 3\epsilon \tag{B.3}$$

Since $\mid a^*_{1,i} - a^*_{1,i}(t,\omega) \mid < \epsilon$ for all $t \geq t_K > T_\epsilon(\omega)$ and for all i, we have, $\mid a^*_{1,i'} - a^*_{1,i'}(t_{K+1},\omega) \mid < \epsilon$. Using this in equation (B.3), we get,

$$a^*_{1,i'}(t_{K+1},\omega) - \hat{a}_{1,i'}(t_K,\omega) \quad > \quad 2\epsilon$$

Hence,

$$\max_i \left\{a^*_{1,i}(t_{K+1},\omega) - \hat{a}_{1,i}(t_K,\omega)\right\} \quad > \quad 2\epsilon$$

Let $\mu(t_{K+1},\omega)$ be the user scheduled by RRS-CA (see equation (B.1)), then,

$$\mu(t_{K+1},\omega) = \arg\max_i \left\{a^*_{1,i}(t_{K+1},\omega) - \hat{a}_{1,i}(t_K,\omega)\right\}$$

Since, $\mid a^*_{1,\mu(t_{K+1},\omega)} - a^*_{1,\mu(t_{K+1},\omega)}(t_{K+1},\omega) \mid < \epsilon$, we have,

$$a^*_{1,\mu(t_{K+1},\omega)} - \hat{a}_{1,\mu(t_{K+1},\omega)}(t_K,\omega) > \epsilon$$

Hence, the "lag", at time t_K, of the user scheduled at time t_{K+1} is at least ϵ. We will now compute the lag of the same user at time t_{K+1} (after channel allocation).

$$a^*_{1,\mu(t_{K+1},\omega)} - \hat{a}_{1,\mu(t_{K+1},\omega)}(t_{K+1},\omega) = a^*_{1,\mu(t_{K+1})} - \left(\hat{a}_{1,\mu(t_{K+1})}(t_K,\omega)\frac{K}{K+1} + \frac{1}{K+1}\right)$$

Since $a^*_{1,\mu(t_{K+1},\omega)} - \hat{a}_{1,\mu(t_{K+1},\omega)}(t_K,\omega) > \epsilon$ and $\frac{1}{K+1} < \epsilon$ (by assumption), we have,

$$a^*_{1,\mu(t_{K+1},\omega)} - \hat{a}_{1,\mu(t_{K+1},\omega)}(t_{K+1},\omega) > 0$$

i.e., the lag in service, at time t_{K+1}, for the user scheduled at time t_{K+1} remains positive. Hence,

$$\mu(t_{K+1},\omega) \in L_1(t_{K+1},\omega) \tag{B.4}$$

Now we will expand $l_1(t_{K+1},\omega)$ in terms of $l_1(t_K,\omega)$. Expanding $\hat{a}_{1,i}(t_{K+1},\omega)$ in terms of $\hat{a}_{1,i}(t_K,\omega)$ and simplifying it, we get,

$$\begin{aligned}
l_1(t_{K+1},\omega) &= \sum_{i\in L_1(t_{K+1},\omega)} \left(a^*_{1,i} - \hat{a}_{1,i}(t_{K+1},\omega)\right) \\
&= \sum_{i\in L_1(t_{K+1},\omega)} \left(a^*_{1,i} - \left(\hat{a}_{1,i}(t_K,\omega)\frac{K}{K+1} + \frac{1}{K+1}1_{\{\mu(t_{K+1},\omega)=i\}}\right)\right) \\
&= \sum_{i\in L_1(t_{K+1},\omega)} \left(a^*_{1,i} - \hat{a}_{1,i}(t_K,\omega) + \frac{\hat{a}_{1,i}(t_K,\omega) - 1_{\{\mu(t_{K+1},\omega)=i\}}}{K+1}\right) \\
&= \sum_{i\in L_1(t_{K+1},\omega)} \left(a^*_{1,i} - \hat{a}_{1,i}(t_K,\omega)\right) + \sum_{i\in L_1(t_{K+1},\omega)} \left(\frac{\hat{a}_{1,i}(t_K,\omega) - 1_{\{\mu(t_{K+1},\omega)=i\}}}{K+1}\right)
\end{aligned}$$

Since $\mu(t_{K+1},\omega) \in L_1(t_{K+1},\omega)$ (from equation (B.4)), we have, $\sum_{i\in L_1(t_{K+1},\omega)} 1_{\{\mu(t_{K+1},\omega)=i\}} = 1$. Substituting in the above equation, we get,

$$l_1(t_{K+1},\omega) = \sum_{i\in L_1(t_{K+1},\omega)} \left(a^*_{1,i} - \hat{a}_{1,i}(t_K,\omega)\right) + \left(\frac{\sum_{i\in L_1(t_{K+1},\omega)} \hat{a}_{1,i}(t_K,\omega) - 1}{K+1}\right) \tag{B.5}$$

By definition, we have, $L_1(t_K,\omega) = \{i : a^*_{1,i} - \hat{a}_{1,i}(t_K,\omega) \geq 0\}$ and $a^*_{1,i} - \hat{a}_{1,i}(t_K,\omega) < 0$ for all $i \notin L_1(t_K,\omega)$. Therefore,

$$\begin{aligned}
\sum_{i\in L_1(t_{K+1},\omega)} \left(a^*_{1,i} - \hat{a}_{1,i}(t_K,\omega)\right) &\leq \sum_{i\in L_1(t_K,\omega)} \left(a^*_{1,i} - \hat{a}_{1,i}(t_K,\omega)\right) \\
&= l_1(t_K,\omega)
\end{aligned} \tag{B.6}$$

. Substituting (B.6) in (B.5), we get,

$$\begin{aligned} l_1(t_{K+1},\omega) &\leq l_1(t_K,\omega) + \left(\frac{\sum_{i\in L_1(t_{K+1},\omega)} \hat{a}_{1,i}(t_K,\omega) - 1}{K+1} \right) \\ \Rightarrow l_1(t_K,\omega) - l_1(t_{K+1},\omega) &\geq \frac{1 - \sum_{i\in L_1(t_{K+1},\omega)} \hat{a}_{1,i}(t_K,\omega)}{K+1} \end{aligned} \quad \text{(B.7)}$$

Consider the user i' such that $i' = \arg\min_i \left\{a^*_{1,i} - \hat{a}_{1,i}(t_K,\omega)\right\}$. Since $l_1(t_K,\omega) > 3N\epsilon$, we require $\hat{a}_{1,i'}(t_K,\omega) - a^*_{1,i'} > 3\epsilon$ and $\hat{a}_{1,i'}(t_K,\omega) - a^*_{1,i'}(t_{K+1},\omega) > 2\epsilon$. User i' is not scheduled at t_{K+1}, hence, $\hat{a}_{1,i'}(t_{K+1},\omega) = \hat{a}_{1,i'}(t_K,\omega)\frac{K}{K+1}$. Since $\frac{1}{K} < \epsilon$, we have, $\hat{a}_{1,i'}(t_{K+1},\omega) - a^*_{1,i'} > 2\epsilon$. Therefore $i' \notin L_1(t_{K+1},\omega)$ and also $\hat{a}_{1,i'}(t_K,\omega) > 3\epsilon$, which implies that,

$$1 - \sum_{i\in L_1(t_{K+1},\omega)} \hat{a}_{1,i}(t_K,\omega) = \sum_{i\notin L_1(t_{T+1},\omega)} \hat{a}_{1,i}(t_K,\omega) > 3\epsilon \quad \text{(B.8)}$$

Using B.8 in B.7, we get,

$$l_1(t_K,\omega) - l_1(t_{K+1},\omega) > \frac{3\epsilon}{K+1} \quad \text{(B.9)}$$

i.e., the lag in service, measured for channel state $\mathbf{R}_1$, decreases by, at least, $\frac{3\epsilon}{K+1}$ if the lag, to begin with, is greater than $3N\epsilon$. Since $\sum_{t=K}^{\infty} \frac{1}{t} = \infty$ for all K, and since channel state $\mathbf{R}_1$ occurs infinitely often (in all ω measured in equation (B.2)), it follows from equation (B.9), that $l_1(t,\omega)$ cannot remain greater than $3N\epsilon$ for all $t > t_K$; there exists a $K' > K$ (with $t_{K'} > t_K$) such that $l_1(t_{K'},\omega) \leq 3N\epsilon$. Also, since $|l_1(t_s,\omega) - l_1(t_{s+1},\omega)| \leq \frac{1}{s+1}$, we have, $|l_1(t,\omega) - l_1(t+1,\omega)| < \epsilon$ for all $t > t_K$. Combining the above arguments, we have, $l_1(t,\omega) \leq 3N\epsilon + \epsilon$ for all $t > t_{K'}$. Since $\epsilon(> 0)$ can be arbitrarily small, we have $\lim_{t\to\infty} l_1(t,\omega) = 0$, i.e., $\{\hat{a}_{1,i}(t,\omega)\} \to \{a^*_{1,i}\}$. Generalizing the above argument to arbitrary channel states, we have, $\{\hat{a}_{j,i}(t,\omega)\} \to \{a^*_{j,i}\}$, and hence, $\bar{R}(t,\omega) \to \bar{R}^*$ (for all ω measured in equation (B.2)). Hence, $\bar{R}(t) \to \bar{R}^*$ almost surely, i.e., RRS-CA is asymptotically optimal. □

APPENDIX C

Rate Region for Ergodic channels with Stationary distribution

We define the rate region $\mathcal{C}$ of the wireless channel as the set of all long time average throughput vectors feasible with probability 1. For an ergodic channel process, the stationary distribution of the process characterizes $\mathcal{C}$ completely (see Liu *et al.* (2003*a*), Kumar *et al.* (2008)). We will show now that the rate region for an ergodic wireless channel with sample space $\{\mathbf{R}_j\}$ and stationary probability distribution $\{\pi_j\}$ is given by

$$\mathcal{C} = \left\{ (\bar{r}_1, \cdots, \bar{r}_N) : \bar{r}_i = \sum_j \pi_j \, a_{j,i} \, r_{j,i}, \;\; a_{j,i} \geq 0, \sum_i a_{j,i} \leq 1, i = 1, \cdots, N \right\} \tag{C.1}$$

$\{a_{j,i}\}$s represent a stationary schedule or an average channel allocation. We will show that for an ergodic channel process, for any $\mu, \bar{R}(\mu)$ is dominated by some $\bar{R}'$, where $\bar{r}_i' = \sum_j \pi_j \, a_{j,i} \, r_{j,i}, \;\; a_{j,i} \geq 0, \sum_i a_{j,i} \leq 1, i = 1, \cdots, N$.

Consider a sample path ω such that $\pi(t, \omega) \to \pi$.

We have each $r_{j,i} \leq B \;\; \forall i, j$.

$\therefore \bar{R}(\mu, t, \omega)$ has limit points.

Let $\bar{R}'$ be a limit point of $\bar{R}(\mu, t, \omega)$.

Let $\bar{R}(\mu, k_1, \omega) \to \bar{R}'$ along $k_1 \in \mathbb{K}_1$, where $\mathbb{K}_1$ is a sub-sequence of $\mathbb{N}$.

Further, let $\{\hat{a}_{j,i}(\mu, k_2, \omega)\} \to \{\hat{a}_{i,j}'\}$ along $k_2 \in \mathbb{K}_2$, where $\mathbb{K}_2$ is a sub-sequence of $\mathbb{K}_1$. We have,

$$\begin{aligned} \lim_{k_2 \to \infty, k_2 \in \mathbb{K}_2} \sum_{j=1}^{M} \hat{a}_{j,i}(\mu, k_2, \omega) \pi_i(k_2, \omega) r_{j,i} &= \sum \hat{a}_{j,i}' \pi_j r_{j,i} \\ &= \bar{r}_i' \end{aligned}$$

We have $\hat{a}_{j,i}' \geq 0 \; \forall i, j, \; \sum_j \hat{a}_{j,i}' = 1$. Thus the allocation $\{\hat{a}_{j,i}'\}$ is a valid stationary schedule as well. Consider the long time average throughput vector corresponding to the stationary schedule $\{\hat{a}'_{j,i}\}$.

$$
\begin{aligned}
\bar{r}_i(\{\hat{a}'_{j,i}\}) &= \lim_{t\to\infty} \frac{1}{t} \sum_{s=1}^{t} \sum_{j} I_{\{R(s)=\mathbf{R}_j\}} I_{\{\mu(s)=i\}} r_{j,i} \\
&= \sum_{j} \lim_{t\to\infty} \frac{1}{t} \sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}} I_{\{\mu(s)=i\}} r_{j,i} \\
&= \sum_{j} \lim_{t\to\infty} \frac{1}{t} \sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}} I_{\{\mu(s)=i\}} \frac{\sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}}}{\sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}}} r_{j,i} \\
&= \sum_{j} \left(\lim_{t\to\infty} \frac{1}{t} \sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}} \right) \left(\lim_{t\to\infty} \frac{\sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}} I_{\{\mu(s)=i\}}}{\sum_{s=1}^{t} I_{\{R(s)=\mathbf{R}_j\}}} \right) r_{j,i} \\
&= \sum_{j} \pi_j \hat{a}'_{j,i} r_{j,i} \quad \text{a.s (by strong law of large numbers Wolff (1989))} \\
&= \bar{r}'_i
\end{aligned}
$$

That is, the rate vector $\bar{R}'$ is feasible almost surely, with a stationary schedule $\{\hat{a}'_{j,i}\}$. Thus for any limit point $\bar{R}'$ of $\bar{R}(\mu, t, \omega)$, there exists atleast one corresponding stationary schedule $\{\hat{a}'_{j,i}\}$. Now, for any limit point $\bar{R}'$ of $\bar{R}(\mu, t, \omega)$, $\liminf \bar{r}_i(\mu, t, \omega) \leq \bar{r}'_i$. Thus for any μ, there exists atleast one stationary schedule, $\{\hat{a}'_{j,i}\}$, such that $\bar{r}_i(\mu, t, \omega) \leq \bar{r}_i(\{\hat{a}'_{j,i}\}),\ \forall i$, where $\bar{R}(\{\hat{a}'_{j,i}\})$ can be achieved almost surely with the schedule. Thus we have proved equation C.1

APPENDIX D

Determination of Optimal Weight $W^*(t)$ for RRS-RA-Wlog

Let surface H, given by $\sum_{i=1}^{N} m_i^* \bar{r}_i = d$, be a hyperplane in $\Re^N$, tangential to $\mathcal{C}$ at $\bar{R}^*$. We will show that $\max_{\bar{R}\in\mathcal{C}} \sum m_i^* \bar{r}_i^* \log \bar{r}_i = \bar{R}^*$. That is, $[m_1^*\bar{r}_1^*, m_2^*\bar{r}_2^*, \ldots m_N^*\bar{r}_N^*]$ is an optimal weight that can be used in RRS-RA-Wlog.

By property of the proportional fair point, we have,

$$\sum_{i=1}^{N} \frac{w_i^*(\bar{r}_i - \bar{r}_i^*)}{\bar{r}_i^*} \leq 0 \tag{D.1}$$

for any $\bar{R} \in \mathcal{C}$. This property is necessary and sufficient for a point to be the unique maximum of $\sum w_i^* \log \bar{r}_i$ in $\mathcal{C}$ (see Kelly (1997)).

Let $\bar{R}$ be any point in $\mathcal{C}$. Since $\mathcal{C}$ is convex, $\bar{R}$ lies in $\sum_i m_i^* r_i <= d$. Also, since $\bar{R}^*$ lies on H, we have $\sum_i m_i^* \bar{r}_i^* = d$. Therefore,

$$\sum_i m_i^*(\bar{r}_i - \bar{r}_i^*) \leq 0 \tag{D.2}$$

$$\Rightarrow \sum_i \frac{m_i^* \bar{r}_i^*(\bar{r}_i - \bar{r}_i^*)}{\bar{r}_i^*} \leq 0 \tag{D.3}$$

Comparing with condition D.1, we get the required result. In general, there can be more than one hyperplane tangential to $\mathcal{C}$ at $\bar{R}^*$. Therefore the weight W^* need not unique upto a scaling factor.

APPENDIX E

Channel Models used in Simulations

E.1 Channel Model A: Correlated Users, Time Independent

We consider a 10 state user correlated sample space for $R(t)$ with a uniform distribution on the sample space. The sample space of the process $R(t)$, for all t is (in KBps), $\Omega_{R(t)} = \{(108, 60, 180, 84)$, $(96, 68, 171, 77)$, $(66, 68, 135, 133)$, $(114, 52, 108, 77)$, $(66, 44, 162, 133)$, $(108, 48, 153, 140)$, $(66, 80, 99, 105)$, $(78, 80, 171, 112)$, $(96, 64, 144, 98)$, $(114, 48, 153, 140)\}$
The vector channel $\{R(t)\}$ is assumed to be i.i.d. over the slots. The sample space of the channel was generated arbitrarily.

E.2 Channel Model B: Independent Users, Time Correlated

We consider independent Rayleigh-fading channels for 3 users, with the following parameters: Mean gain (in dB) $= (5, 3, 2)$, Doppler shift (in Hz) $= (50, 90, 40)$ and Slot duration (in msec) $= 1.5$. The mapping from minimum SNR to maximum supportable transmission rate is described in Table E.1.

The SNR to transmission rate mapping is based on a sub-set of the one used in CDMA/HDR Bender *et al.* (2000).

SNR range (dB)	Transmission Rate (KBps)
< -8.5	0
-8.5 to -1	102
-1 to 9.5	614
≥ 9.5	2457

Table E.1: Channel model B: Minimum SNR to maximum supportable transmission rate

E.3 Channel Model C: Correlated Users, Time Independent

We consider a 10 state user correlated sample space for $R(t)$ with a uniform distribution on the sample space. The sample space of the process $R(t)$ (in KBps) is $\Omega_{R(t)} = \{(95, 17, 44, 23), (16, 6, 70, 81), (40, 88, 26, 36), (100, 33, 8, 88), (19, 66, 29, 19), (43, 56, 96, 59), (27, 92, 86, 7), (33, 42, 3, 69), (11, 44, 45, 78), (9, 76, 13, 100)\}$.

The vector channel $\{R(t)\}$ is assumed to be i.i.d. over slots. The sample space of the channel was generated arbitrarily.

In the simulation discussed in Section 5.3, the sample space of $R(t)$ for $t < 2000$ slots is $\Omega_{R(t)} = \{(95, 17, 44, 23), (16, 6, 70, 81), (40, 88, 26, 36), (100, 33, 8, 88), (19, 66, 29, 19), (43, 56, 96, 59), (27, 92, 86, 7), (33, 42, 3, 69), (11, 44, 45, 78), (9, 76, 13, 100)\}$ and the sample space of $R(t)$ for $t >= 2000$ slots is $\Omega_{R(t)} = \{(95, 17, 44, 23), (16, 6, 70, 81), (40, 88, 26, 36), (100, 33, 8, 88), (19, 66, 29, 19), (43, 56, 96, 59), (27, 92, 86, 7)\}$.
We assume a uniform distribution on the sample space for all t. The vector channel $\{R(t)\}$ is assumed to be independent over slots. The sample space of the channel was generated arbitrarily.

In the simulation discussed in Section 5.4, only the first two users (and their corresponding channel states) are considered. The sample space of $R(t)$ for $t < 2000$ slots is $\Omega_{R(t)} = \{(95, 17), (16, 6), (40, 88), (100, 33), (19, 66), (43, 56), (27, 92), (33, 42), (11, 44), (9, 76)\}$ and the sample space of $R(t)$ for $t >= 2000$ slots is $\Omega_{R(t)} = \{(95, 17), (16, 6), (40, 88), (100, 33), (19, 66)\}$. We assume a uniform distribution on the sample space for all t. The vector channel $\{R(t)\}$ is assumed to be independent over slots. The sample space of the channel was generated arbitrarily.

E.4 Channel Model D: Independent Users, Time Correlated

We consider independent Rayleigh-fading channels for 10 users with the following parameters: Mean gain (in dB) $= (2, 5, 1, 1, 5, 5, 10, -5, -5, 5)$, Doppler shift (in Hz) $= (50, 90, 50, 90, 60, 100, 50, 70, 50, 90)$ and Slot duration (in msec) $= 1.5$. The mapping from minimum SNR to maximum supportable transmission rate is described in Table E.2. The SNR to transmission rate mapping is based on the one used in CDMA/HDR Bender *et al.* (2000).

In the simulation discussed in Section 6.1, only the first two users are considered.

In the simulation discussed in Section 6.2, in Figure 6.4, only the first 4 users are considered.

SNR range (dB)	Transmission Rate (KBps)
< -12.5	0
-12.5 to -9.5	38
-9.5 to -8.5	76
-8.5 to -6.5	102
-6.5 to -5.7	153
-5.7 to -4	204
-4 to -1	307
-1 to 1.3	614
1.3 to 3.0	921
3.0 to 7.2	1228
7.2 to 9.5	1843
≥ 9.5	2457

Table E.2: Channel model D: Minimum SNR to maximum supportable transmission rate

REFERENCES

1. **3GPP** (2011). 3GPP LTE-Advanced, http://www.3gpp.org/lte-advanced.

2. **Agrawal, R.** and **V. Subramanian**, Optimality of certain channel aware scheduling policies. *In Proceedings of the 40th Allerton Conference on Communication, Control, and Computing*. 2002.

3. **Andrews, M.**, **K. Kumaran**, **K. Ramanan**, **A. Stolyar**, **R. Vijayakumar**, and **P. Whiting** (2004). Scheduling in a queueing system with asynchronously varying service rates. *Probability in the Engineering and Informational Sciences*, **18**, 191–217.

4. **Andrews, M.**, **L. Qian**, and **A. Stolyar**, Optimal utility based multi-user throughput allocation subject to throughput constraints. *In Proceedings of the 24th Annual Joint Conference of the IEEE Conference on Computer Communications Societies, Infocom*, volume 4. 2005.

5. **Bender, P.**, **P. Black**, **M. Grob**, **R. Padovani**, **N. Sindhushayana**, and **A. Viterbi** (2000). CDMA/HDR a bandwidth efficient high speed wireless data service for nomadic users. *IEEE Communications magazine*, **38**, 70–77.

6. **Borkar, V.** and **S. Meyn** (2000). The O.D.E. method for convergence of stochastic approximation and reinforcement learning. *SIAM Journal on Control and Optimization*, **38**, 447–469.

7. **Borst, S.** and **P. Whiting**, Dynamic rate control algorithms for HDR throughput optimization. *In Proceedings of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, volume 2. 2001.

8. **Holtzman, J.**, Asymptotic analysis of proportional fair algorithm. *In Proceedings of the IEEE PIMRC*, volume 2. 2001.

9. **Jalali, A.**, **R. Padovani**, and **R. Pankaj**, Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system. *In Proceedings of the 51st IEEE Vehicular Technology Conference*, volume 3. 2000.

10. **Kelly, F.** (1997). Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, **8**, 33–37.

11. **Kleywegt, A.**, **A. Shapiro**, and **T. H. de Mello** (2002). The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, **12**, 479–502.

12. **Kumar, A.**, **M. D**, and **J. Kuri** (2008). Wireless networking.

13. **Kuratowski, K.**, *Topology, Volumes I and II*. New York: Academic Press, 1966.

14. **Kushner, H.** and **P. Whiting**, Asymptotic properties of proportional-fair sharing algorithms. *In Proceedings of the 40th Allerton Conference on Communication, Control, and Computing*. 2002.

15. **Kushner, H.** and **P. Whiting** (2004). Convergence of proportional-fair sharing algorithms under general conditions. *IEEE Transactions on Wireless Communications*, **3**, 1250–1259.

16. **Kushner, H.** and **G. Yin**, *Stochastic Approximation Algorithms and Applications*. Springer-Verlag New York, 1997.

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

18. **Liu, Y.**, **S. Gruhl**, and **E. Knightly** (2003*b*). Wcfq: An opportunistic wireless scheduler with statistical fairness bounds. *IEEE Transactions on Wireless Communications*, **2**, 1017–1028.

19. **Lu, S.**, **V. Bharghavan**, and **R. Srikant** (1999). Fair scheduling in wireless packet networks. *IEEE/ACM Transactions on Networking*, **7**, 473–489.

20. **IEEE 802.16m** (2011). IEEE 802.16m-2011 - IEEE standard for local and metropolitan area networks - part 16: Air interface for broadband wireless access systems. amendment 3: Advanced air interface.

21. **Nandagopal, T.**, **S. Lu**, and **V. Bharghavan** (1999). A unified architecture for the design and evaluation of wireless fair queueing algorithms. *Wireless Networks*, **8**, 231–247.

22. **Neely, M.**, **E. Modiano**, and **C.-P. Li**, Fairness and optimal stochastic control for heterogeneous networks. *In Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies(INFOCOM)*, volume 3. 2005*a*.

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

24. **Ng, E.**, **I. Stoica**, and **H. Zhang**, Packet fair queueing algorithms for wireless networks with location-dependent errors. *In Proceedings of the 17th Annual Joint Conference of the IEEE Computer and Communications Societies(INFOCOM)*, volume 3. 1998.

25. **Robinson, S.** (1996). Analysis of sample-path optimization. *Mathematics of Operations Research*, **21**, 513–528.

26. **Rudin, W.** (1976). Principles of mathematical analysis, third edition.

27. **Rybko, A.** and **A. Stolyar** (1992). Ergodicity of stochastic processes describing the operation of open queueing networks. *Problems of Information Transmission*, **28**, 199–220.

28. **Shakkotai, S.** and **A. Stolyar**, Scheduling algorithms for a mixture of real-time and nonreal-time data in HDR. *In Proceedings of the International Teletraffic Congress, ITC*. 2000.

29. **Shakkottai, S.** (2008). Effective capacity and QoS for wireless scheduling. *IEEE Transactions on Automatic Control*, **53**, 749–761.

30. **Stolyar, A.** (2005*a*). Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm. *Queueing Systems*, **50**, 401–457.

31. **Stolyar, A.** (2005*b*). On the asymptotic optimality of the gradient scheduling algorithm for multiuser throughput allocation. *Operations Research*, **53**, 12–25.

32. **Stolyar, A.** and **K. Ramanan** (2001). Largest weighted delay first scheduling: Large deviations and optimality. *The Annals of Applied Probability*, **11**, 1–48.

33. **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**, 1936–1948.

34. **Tassiulas, L.** and **A. Ephremides** (1993). Dynamic server allocation to parallel queues with randomly varying connectivity. *IEEE Transactions on Information Theory*, **39**, 466–478.

35. **Wolff, R. W.** (1989). Stochastic modeling and the theory of queues.

36. **Wu, D.** and **R. Negi** (2005). Utilizing multiuser diversity for efficient support of quality of service over a fading channel. *IEEE Transactions on Vehicular Technology*, **54**, 1198–1206.