THROUGHPUT OPTIMAL SCHEDULING
IN DEGRADED GAUSSIAN BROADCAST CHANNELS

BY

ATİLLA ERYILMAZ

Boğaziçi University, İstanbul, 1999

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2001

Urbana, Illinois

ABSTRACT

We consider a degraded broadcast channel in which the transmitter maintains a separate queue for each receiver. Without knowing the arrival rates to the queues, a power allocation strategy is designed such that it stabilizes the system whenever it is stabilizable.

The stability of the queues is demonstrated through a stochastic Lyapunov function argument. A fair average queueing delay property of the policy is illustrated through computer simulations.

To my family.

ACKNOWLEDGMENTS

I am grateful to my advisor, Prof. R. Srikant, for his guidance and support during the preparation of this thesis. It has been a privilege to work with him throughout my graduate studies.

I also would like to thank all the members of the ECE Department Faculty for their excellence in education and research. Their efforts build a comfortable and productive environment for study.

Finally, I am thankful to my office mates and everybody else, near or far, that has supported me in my studies.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction

It is frequently required to *broadcast* independent information to different receivers through a single transmitter over a wireless channel. The flow destined to each of the receivers is collected in a separate queue at the transmitter, where a power allocation decision is made according to a certain rule subject to average power constraints.

In this thesis, we assume that the channel is a *degraded Gaussian broadcast channel*, the capacity region of which is well defined [1], and can be achieved through *superposition coding* and *interference cancellation.* The total average power spent for the transmission is not allowed to be above some value, say P. More detailed characteristics of the degraded broadcast channel will be provided in Section 1.3.

Such a model can arise in a context where a single base station is sending separate data to each of its fixed wireless receivers. The noise power associated with each receiver would be different depending on its location. Assuming that the base station has the knowledge of the noise powers at the receivers and consequently the capacity region, we aim to come up with a power allocation strategy that *stabilizes* the queues without the knowledge of the arrival rates to the individual queues, whenever the system is stabilizable. The notion of *stability of the queues* will be clarified in the later chapters.

The thesis is organized as follows. Chapter 1 covers some related work and background information about the system model. In Chapter 2, a related deterministic fluid model is analyzed to provide insight into the evolution of the queue lengths. Chapter 3 presents the main results of the thesis for the stochastic system, where the stability of the system is proved. Simulation results and some discussions are provided in Chapter 4. Concluding remarks and possible future work is resented in Chapter 5.

1.2 Related Work

Tassiulas and Ephremides[2] considered the problem of dynamic server allocation to parallel queues competing for the attention of a single server in a purely queueing theory context. In the case of infinite buffers, they obtained necessary and sufficient conditions for the stabilizability of the system. They showed that the allocation policy that serves the longest queue at a given time stabilizes the system when stability conditions hold. We have analyzed a similar system including information theoretic considerations for broadcasting the data destined to separate users. The service rates offered in our scenario lies on the boundary of the convex Gaussian broadcast capacity region. For such a model we came up with a different power allocation strategy that is throughput optimal.

Some work has been conducted by Telatar and Gallager[3], among others, to combine queueing theory with information theory for multiaccess. In a simplified model, they have investigated the trade-offs, such as delay versus probability of error. In our work, we mainly concentrate on the stability of the queues while keeping the error probability below some threshold and propose a policy to achieve this.

It should be noted that our work assumes non-fading channel characteristics. The broadcast capacity region in the existence of fading is examined by Li and Goldsmith [4, 5]. The same idea we use in our policy can also be used in a fading channel to achieve our objective.

Moreover, Tse addressed the problem of optimal power allocation over parallel Gaussian broadcast channels in [6]. In his work, Tse presents greedy power allocation strategies to achieve all of the rates in the broadcast capacity region for a given total average power constraint.The optimal power allocation is always allocating all power, if any, to the user with the best reception in each of the component channels. An interesting feature of the optimal power allocation is that the information for no more than one user is broadcasted in each of the parallel channels, and the power allocation accross the channels is a water-filling solution. Moreover, an iterative algorithm for resource allocation is provided, which converges to the minimum total power required to support a given target rate.

Tse's policy, although optimal in the sense of maximizing the total throughput, can be unfair among users because onle the user with the best reception is allowed to transmit at any time. Such an unfairness can cause the queues of the users with bad channels to build up. This unfairness is suggected to be compensated for through the use of a different rate reward for each user. Furthermore, an iterative algorithm is offered which finds the appropriate rate rewards for minimizing the average total power consumption subject to *given* minimum rate requirements.

In our model, a *single* degraded Gaussian broadcast channel is considered, again with a total average power constraint. Contrary to Tse's approach, however, we also incorporate the queue status associated with each of the flows as well as the information theoretic analysis. Our objective is to attain a power allocation strategy among the users, which stabilizes the system *without* the knowledge of the required rates ahead of time whenever the system is stabilizable. We call such a policy as *throughput optimal.* A policy is said to be *throughput optimal*[7] if it can stabilize any system that is stabilizable under some scheduling policy. Note that Tse assumes that the rate requirements for the flows is known beforehand.

Finally, Mikhailov's work[8] is worth mentioning. He proved the stability of Markov chains in $\Re_+^N$ via a geometrical analysis under some constraints, which also hold in our case. Although his analysis is considerably simple in the 2-Dimensional case, it is difficult to extend to N-Dimensional Markov Chains. Our work takes a completely different approach and is easily extendable to N-Dimensions.

1.3 System Model

Consider a link that is sending data to N users over a degraded Gaussian broadcast channel with an upper bound of P on the average power available for transmission. Each of the N users have a buffer as shown in Fig. 1.1. The buffer length at time k for the i^{th} user is denoted by

$x_i[k]$, and the number of arrivals from the i^{th} flow at time k is denoted by $a_i[k]$. Also, let us denote the noise variance over the channel between the link and the i^{th} user by σ_i^2. Furthermore, we assume $\sigma_1^2 < \sigma_2^2 < \ldots < \sigma_N^2$. The link uses a scheduling policy and allocates a certain service rate to each user at each time instant. Let us denote the service rate offered to the i^{th} flow at time k by $\mu_i[k]$. Then the evolution of the i^{th} buffer length is given by

$$x_i[k+1] = (x_i[k] + a_i[k] - \mu_i[k])^+,$$

where $(y)^+ = y$ if $y > 0$ and zero otherwise.

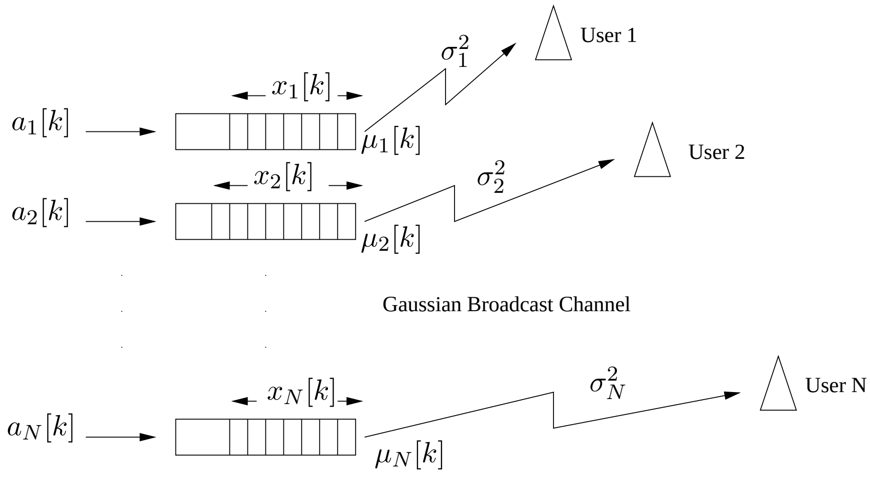

Figure 1.1 The System Model

For the degraded Gaussian broadcast channel, it is known that[1] the capacity region is defined by the following set of inequalities:

$$\mu_i[k] \leq \frac{1}{2}\log\left(1 + \frac{\alpha_i[k]P}{\sum_{j=1}^{i-1}\alpha_j[k]P + \sigma_i^2}\right), \qquad i = 1, 2, \ldots, N, \tag{1.1}$$

where $\sum_{i=1}^{N} \alpha_i[k] = 1$ for all k. Note that, the channel gain associated with user i is constant. Hence we assume that the effect of the channel gain can be absorbed in σ_i^2 and therefore is not shown explicitly in the previous equation.

The arrival process, $a_i[k]$, to queue i is assumed to be independent accross both time slots and the receivers. However, $\lambda_i := E[a_i[k]]$ could be different for each i, but the mean arrival rates, $\{\lambda_i\}$, lie within the broadcast capacity region. Further, the second moments satisfy $E[a_i^2] < A_i < \infty$ for $i = 1, 2, \ldots, N$. For such a system, our aim is to find a simple stabilizing policy that is *throughput optimal*, the definition of which is provided in Section 1.2.

As mentioned eralier, Gallager [9] derived the capacity of a broadcast channel. In addition, he also computed the error exponent as a function of the transmission rates to the receivers and the block lengths of the codes used. Thus, strictly speaking if we define a time slot as a certain number of channel uses, say K, then for a given probabilty of packet decoding error, this defines the possible rates that can be assigned to each receiver. In our model, we simply use the set of maximum achievable mutual information as the set of achievable rates. A more precise model, using the error exponents calculated by Gallager [9] can be obtained in a manner similar to Telatar and Gallager [3] for multiple access channels.

CHAPTER 2

FLUID MODEL

2.1 A Simple Fluid Model

Before we analyze the system described in Section 1.3, we will study a related fluid model, which is easy to analyze and which gives an intuitive understanding of the behavior of the system operating under this policy.

Fluid models of queueing networks are among the simplest models to analyze, owing to the fact that deterministic calculus can be applied. Dai [10] and Dai and Meyn [11] showed that stability of a fluid model implies the stability (in an appropriate sense) of related multi-class networks with random service and interarrival process of renewal type. Hajek [12] and Gamarnik [13] have also shown that the stability of networks with fluid traffic implies stability of networks with deterministically constrained traffic. However, we will not directly consider such connection between fluid models of stochastic stability, but instead we attempt to obtain some intuition the behavior of the stochastic system.

In the fluid model, the arrival processes are deterministic with a constant arrival rate of λ_i for the i^{th} flow. We recall that, in the original problem with stochastic arrival processes,

λ_i was the mean arrival rate to queue i. In the fluid model, time is no longer discrete, but is continuous. The service rate offered to the i^{th} flow at time t is denoted by $\mu_i(t)$. Thus, our continuous-time deterministic fluid model is given by

$$\begin{aligned} \dot{x}_i(t) &= \lambda_i - \mu_i(t), \qquad x_i(t) > 0 \\ &= (\lambda_i - \mu_i(t))^+, \quad x_i(t) = 0, \qquad i = 1, ..., N. \end{aligned}$$

The *stability condition* for our model requires that the mean arrival rates lie inside the broadcast capacity region. More precisely: $\exists \beta_i$, $i = 1, 2, \ldots, N$, satisfying $\sum_{i=1}^{N} \beta_i = 1$ and $0 \leq \beta_i \leq 1$, such that

$$\lambda_i < \frac{1}{2} \log \left(1 + \frac{\beta_i P}{\sum_{j=1}^{i-1} \beta_j P + \sigma_i^2} \right). \tag{2.1}$$

In Section 2.2, we present a simple scheduling policy, that has no knowledge of the arrival rate, which empties all of the buffers of the fluid model in some *finite* time, provided that the *stability condition* is satisfied. It may be possible to use this result to show that the original system is stable (i.e., it is positive recurrent) using the techniques in the papers cited earlier[10, 11], although we do not pursue this here.

Before we present our throughput-optimal policies, we note that, due to the nature of the broadcast channel, traditional *work-conserving policies* will not be throughput optimal. For example, a priority policy which works only on Queue i if it is nonempty and all the other queues with index $j < i$ are empty will not be throughput optimal. The reason is that, such policies eventually end up allocating a certain fraction of time to each user which essentially amounts to TDMA. It is well known that all points in the capacity region cannot be achieved using TDMA and that simultaneous transmission to multiple users using superposition coding is required to achieve capacity. Thus, any policy that serves only one queue at any time cannot be throughput-optimal.

2.2 A Simple Throughput Optimal Scheduling Policy

In this section, we present a simple scheduling policy, which makes the fluid model described in Section.2.1 evolve in such a way that all of the queue sizes will be zero in a *finite amount* of time provided that the system satisfies the *stability condition* specified by (2.1).

After describing the policy, a geometrical interpretation is given first to provide an intuitive understanding for the two-user case. Then we prove the stability of the system with N users using a Lyapunov analysis. And finally, we extend the results to more general scheduling policies.

Policy 1. At each time instant, t, allocate the service rates, $(\mu_1(t), \mu_2(t), ..., \mu_N(t))$, such that

$$\frac{\mu_1(t)}{x_1(t)} = \frac{\mu_2(t)}{x_2(t)} = ... = \frac{\mu_N(t)}{x_N(t)}$$

and $(\mu_1(t), \mu_2(t), ..., \mu_N(t))$ lies on the boundary of the broadcast capacity region.

Note that, this policy requires only the current queue length of each of the queues along with the capacity region.

2.2.1 Geometrical Interpretation for the 2-User Case

A two-user analysis provides insight into the evolution of the queue lengths and enables us to pictorially illustrate this evolution. For this purpose, consider the following two-queue fluid system:

$$\begin{aligned}\dot{x}_1(t) &= \lambda_1 - \mu_1(t) \\ \dot{x}_2(t) &= \lambda_2 - \mu_2(t), \end{aligned} \tag{2.2}$$

where $x_1(t)$ and $x_2(t)$ correspond to the queue lengths of Users 1 and 2, respectively. The arrival rates λ_1 and λ_2 are assumed to be constants and lie within the capacity region of a broadcast channel. The service rates $\mu_1(t)$ and $\mu_2(t)$ satisfy the following equalities:

$$\mu_1(t) = \frac{1}{2}\log\left(1 + \frac{\alpha(t)P}{\sigma_1^2}\right)$$

and

$$\mu_2(t) = \frac{1}{2}\log\left(1 + \frac{(1-\alpha(t))P}{\alpha(t)P + \sigma_2^2}\right),$$

where $\sigma_1^2 < \sigma_2^2$. That is, $\mu_1(t)$ and $\mu_2(t)$ is a point on the boundary of the capacity region. The capacity region is convex and an example is shown in Figure 2.1.

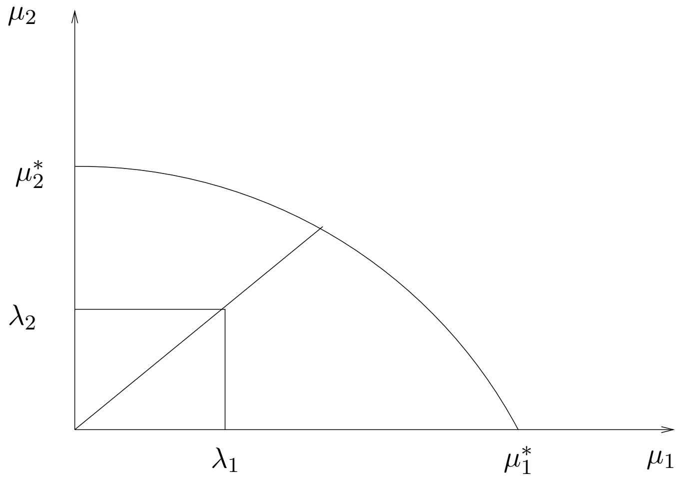

Figure 2.1 The capacity region

Along with the capacity region, the line passing through (λ_1, λ_2) is also plotted in Figure 2.1. Now, the claim is that, given any initial conditions for the queue lengths, $x_1(0) := x_{10}$ and $x_2(0) := x_{20}$, both of the queues will be emptied in a *finite* amount of time, where this time may depend on x_{10} and x_{20}. To show this, several facts will be illustrated pictorially. These are listed next, without including their proofs since the same result will be obtained using a Lyapunov argument in Subsection 2.2.2.

Fact 1. *If we assume that initially* $\frac{x_{10}}{\lambda_1} = \frac{x_{20}}{\lambda_2}$ *holds, then the scheduling policy ensures that* $\frac{x_1(t)}{\lambda_1} = \frac{x_2(t)}{\lambda_2}$ *for all* $t > 0$*. Further, both of the queue lengths* $x_1(t)$ *and* $x_2(t)$ *will decrease to hit the origin in some* finite *time.*

Proof. First of all, to get an insight, it is helpful to interpret the situation pictorially as in Figure 2.2. In the figure, the capacity region is plotted along with the rates allocated to each receiver when $\frac{x_1}{x_2} = \frac{\lambda_1}{\lambda_2}$.

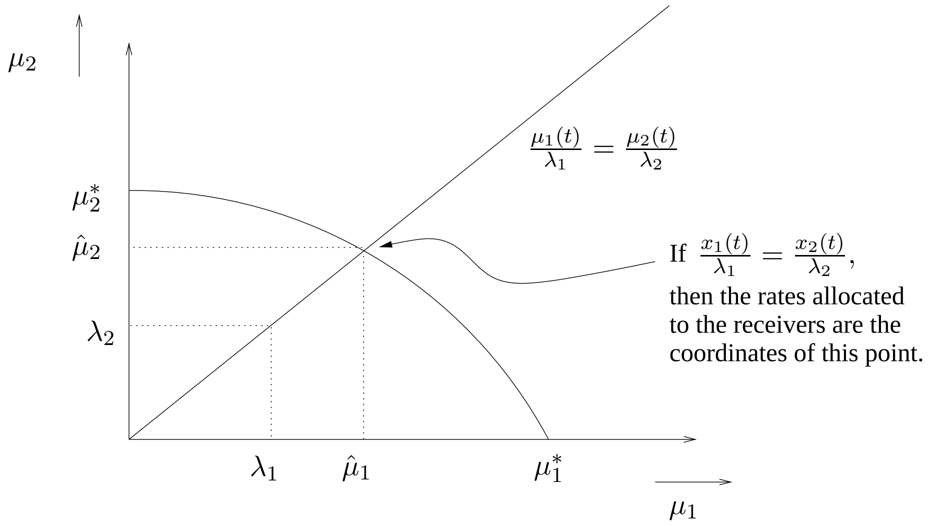

Figure 2.2 Pictorial interpretation to Fact 1

To prove the first part of the fact, let's consider the initial service rate allocation provided to user 1 and 2 respectively, μ_{10} and μ_{20}. We have

$$\frac{\mu_{10}}{\mu_{20}} = \frac{x_{10}}{x_{20}} = \frac{\lambda_1}{\lambda_2} := C$$

Hence, we see that the service rates and the arrival rates are proportional to each other when the system starts with the given condition. Now making use of the state equations given in Equation 2.2 yields:

$$\frac{x_1(t)}{x_2(t)} = \frac{x_{10} + \int_0^t (\lambda_1 - \mu_1(s))ds}{x_{20} + \int_0^t (\lambda_2 - \mu_2(s))ds} = \frac{C \cdot x_{20} + \int_0^t C \cdot (\lambda_2 - \mu_2(s))ds}{x_{20} + \int_0^t (\lambda_2 - \mu_2(s))ds} = C$$

which shows that the ratio of the two queue lengths will always be the same constant value C. Consequently, the rates allocated to the two users are always constant and are denoted by $\hat{\mu}_1$ and $\hat{\mu}_2$ in Figure 2.2. This shows that once the queue lengths hit the line given in Figure 2.2, then they must stay on that line. Now, it must be shown that the queue lengths will have a negative drift and will hit the origin in some finite time. This is obvious once we note that (λ_1, λ_2) pair lies within the capacity region. Then, $\hat{\mu}_1 > \lambda_1$ and $\hat{\mu}_2 > \lambda_2$ is obtained, which in turn implies that:

$$\dot{x}_i(t) = \lambda_i - \hat{\mu}_i < \delta \qquad\qquad i = 1, 2 \qquad \delta > 0$$

That is to say, both of the queues will have a negative drift of at least δ and hence there exists a finite time such that both of the queues will be completely emptied. $\square$

Next, we need to consider the case when $\frac{x_{10}}{\lambda_1} \neq \frac{x_{20}}{\lambda_2}$. This case is covered in Facts 2, 3 and 4.

Fact 2. *Without loss of generality, assume* $\frac{x_{10}}{\lambda_1} > \frac{x_{20}}{\lambda_2}$ *and also let* $\frac{x_{10}}{x_{20}} > \frac{\tilde{\mu}_1}{\lambda_2}$, *where* $\tilde{\mu}_1$ *is shown in Figure 2.3. Then there exists a finite* T *such that*

$$\frac{x_1(T)}{x_2(T)} = \frac{\tilde{\mu}_1}{\lambda_2}$$

.

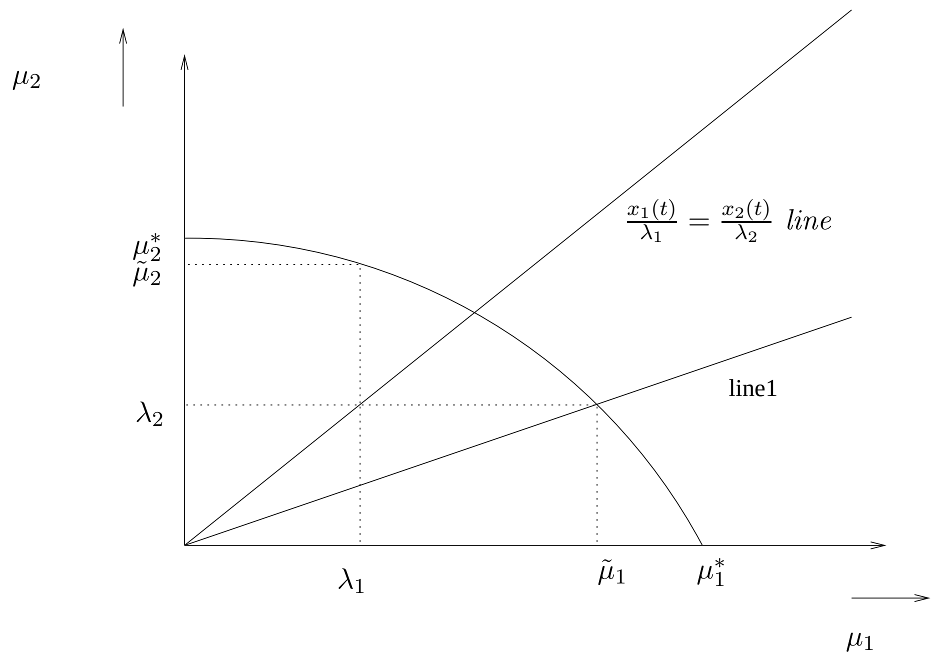

Figure 2.3 *Pictorial interpretation to Fact 2*

Proof. For the proof, let us consider the worst case scenario and upper and lower bound the queue lengths $x_1(t)$ and $x_2(t)$ as follows:

$$x_1(t) = x_{10} + \lambda_1 \cdot t - \int_0^t \mu_1(s) \cdot ds \leq x_{10} + \lambda_1 \cdot t - \tilde{\mu}_1 \cdot t$$

since $\tilde{\mu}_1$ is the minimum service that can be attained according to our assumptions. Also;

$$x_2(t) = x_{20} + \lambda_2 \cdot t - \int_0^t \mu_2(s) \cdot ds \geq x_{20} + \lambda_2 \cdot t - \lambda_2 \cdot t = x_{20}$$

since $\mu_2(s) \leq \lambda_2$ for all $s \leq T$. Using these bounds, we can see that $T < \tau$, where τ is given by:

$$\frac{x_{10} + \tilde{\mu}_1 \cdot \tau - \tilde{\mu}_1 \cdot \tau}{\tilde{\mu}_1} = \frac{x_{20}}{\lambda_2}$$

$$\tau = \tilde{\mu}_1 \cdot \frac{\left(\frac{x_{10}}{\tilde{\mu}_1} - \frac{x_{20}}{\lambda_2}\right)}{\tilde{\mu}_1 - \lambda_1}$$

which is a positive finite value since $\frac{x_{10}}{\tilde{\mu}_1} > \frac{x_{10}}{\lambda_1}$. Therefore, we have shown that $\exists T < \infty$ such that $\frac{x_1(T)}{x_2(T)} = \frac{\tilde{\mu}_1}{\lambda_2}$. ☐

The next fact covers the case when $\frac{x_{10}}{x_{20}} = \frac{\tilde{\mu}_1}{\lambda_2}$.

Fact 3. *Again without loss of generality, let us consider the case when $\frac{x_{10}}{x_{20}} = \frac{\tilde{\mu}_1}{\lambda_2}$. Then it possible to find $\bar{\mu}_1$ and $\bar{\mu}_2$ such that, there exists a finite T_1 satisfying*

$$\frac{x_1(T_1)}{\bar{\mu}_1} = \frac{x_2(T_1)}{\bar{\mu}_2}$$

where $\bar{\mu}_1 > \lambda_1 + \delta$ and $\bar{\mu}_2 > \lambda_2 + \delta$ with some finite $\delta > 0$. (Note: As can be seen from Figure 2.4, we are actually trying to show that the queue lengths will evolve from a point on line 1 to a point on line 2 in some finite time.)

Proof. The proof of this fact follows the same type of reasoning with Fact 2. Here, again a worst case scenario is being considered. Therefore, we upper bound $x_1(t)$ and lowerbound $x_2(t)$ to obtain the largest amount of time to hit line 2. During the analysis, we will find the conditions on $\bar{\mu}_1$ and $\bar{\mu}_2$ for the claim to be true.

According to our assumption, the initial conditions are such that, $\frac{x_{10}}{x_{20}} = \frac{\tilde{\mu}_1}{\lambda_2}$. First, observe that the ratio of the initial queue lengths, $\frac{x_{10}}{x_{20}}$ need always be greater $\frac{\bar{\mu}_1}{\bar{\mu}_2}$ because (x_{10}, x_{20}) lies

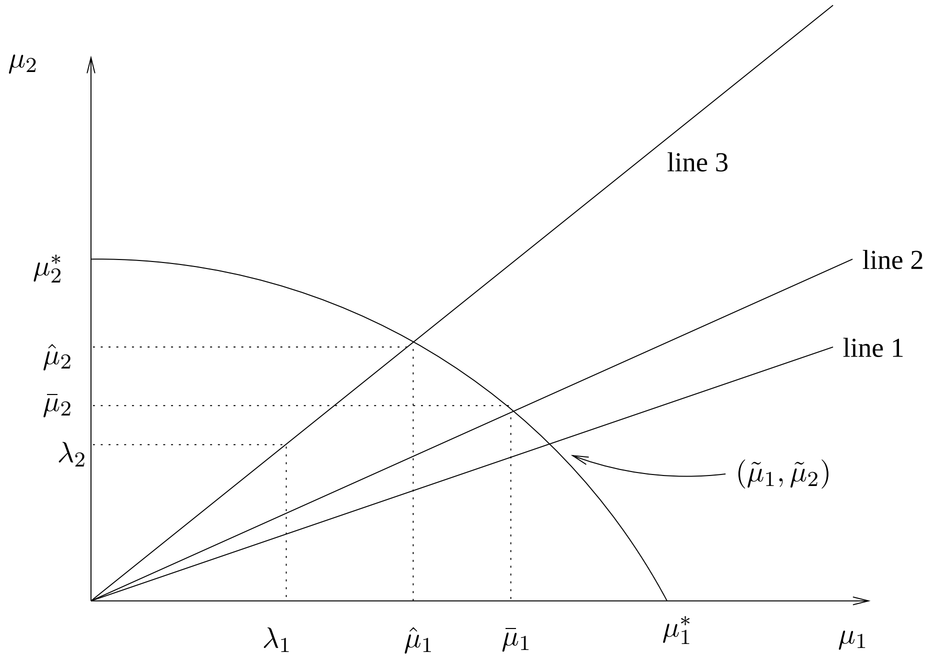

Figure 2.4 *Pictorial interpretation to Fact 3*

on line 1 which is below line 2 shown in Figure 2.4. Second, let us specify upper and lower bounds on queue lengths $x_1(t)$ and $x_2(t)$, respectively.

$$x_1(t) = x_{10} + \lambda_1 \cdot t - \int_0^t \mu_1(s) \cdot ds \leq x_{10} + \lambda_1 \cdot t - \hat{\mu}_1 \cdot t$$

since $\hat{\mu}_1$ is a lower bound on the service rate allocated to user 1. Note that $\bar{\mu}_1 > \hat{\mu}_1$ is also a lower bound, but we will take the looser bound for the analysis. Next,

$$x_2(t) = x_{20} + \lambda_2 \cdot t - \int_0^t \mu_2(s) \cdot ds \geq x_{20} + \lambda_2 \cdot t - \bar{\mu}_2 \cdot t$$

since $\bar{\mu}_2$ is the maximum service provided to user 2 in the region of concentration. Now, define the time instant τ to be the epoch when the queue lengths hit line 2. Then τ needs to satisfy:

$$\frac{x_{10} + \lambda_1 \cdot \tau - \hat{\mu}_1 \cdot \tau}{\bar{\mu}_1} = \frac{x_{20} + \lambda_2 \cdot \tau - \bar{\mu}_2 \cdot \tau}{\bar{\mu}_2}$$

$$\tau = \frac{\frac{x_{10}}{\bar{\mu}_1} - \frac{x_{20}}{\bar{\mu}_2}}{\frac{\lambda_2}{\bar{\mu}_2} - 1 + \frac{\hat{\mu}_1 - \lambda_1}{\bar{\mu}_1}}$$

We know that the numerator on the right-hand side is always positive. For τ to attain a positive value, $\bar{\mu}_1$ and $\bar{\mu}_2$ must be chosen such that the denominator is also positive. Also, note that the term $\frac{\hat{\mu}_1 - \lambda_1}{\bar{\mu}_1}$ in the denominator is always positive. Let us make $\bar{\mu}_2 \downarrow \lambda_2$ and $\bar{\mu}_1 \uparrow \tilde{\mu}_1$. Then

$$\frac{\lambda_2}{\bar{\mu}_2} - 1 \to 0 \qquad and \qquad \frac{\hat{\mu}_1 - \lambda_1}{\bar{\mu}_1} > 0.$$

Therefore, we can choose $\bar{\mu}_1$ and $\bar{\mu}_2$ such that

$$\frac{\lambda_2}{\bar{\mu}_2} - 1 + \frac{\hat{\mu}_1 - \lambda_1}{\bar{\mu}_1} > 0$$

Thus, line 2 in Figure 2.4 can be reached in finite time,$T_1 < \tau$, if $\bar{\mu}_1$ and $\bar{\mu}_2$ are chosen properly. □

Fact 4. *Once line 2 is reached, the queue lengths will never go below line 2. More precisely, given* $\frac{x_{10}}{x_{20}} = \frac{\bar{\mu}_1}{\bar{\mu}_2}$*, we have*

$$\frac{x_1(t)}{x_2(t)} \leq \frac{\bar{\mu}_1}{\bar{\mu}_2} \qquad\qquad forall \qquad t > 0$$

Proof. For the proof we will concentrate on an infinitesimal time interval, Δt, over which the service rates are constant given by $\bar{\mu}_1$ and $\bar{\mu}_2$. Then we will show that by the end of the duration Δt, the queue lengths will evolve to a point that is above line 2. Then the key point is that, the exactly same analysis is applicable to the system starting at time Δt. Thus, we can conclude that the queue lengths will never go below line 2 once it crosses line 2.

The analysis goes as follows: Let us first note that $\frac{\lambda_1}{\lambda_2} < \frac{\bar{\mu}_1}{\bar{\mu}_2}$ since line 2 lies below line 3 as shown in Figure 2.4. Now, let us claim that $\frac{x_1(\Delta t)}{x_2(\Delta t)} < \frac{\bar{\mu}_1}{\bar{\mu}_2}$, the justification of this claim is:

$$\frac{x_1(\Delta t)}{x_2(\Delta t)} = \frac{x_{10} + \lambda_1 \cdot \Delta t - \bar{\mu}_1 \cdot \Delta t}{x_{20} + \lambda_2 \cdot \Delta t - \bar{\mu}_2 \cdot \Delta t} < \frac{\bar{\mu}_1}{\bar{\mu}_2}$$

then

$$x_{10} \cdot \bar{\mu}_2 + \lambda_1 \cdot \bar{\mu}_2 \cdot \Delta t - \bar{\mu}_1 \cdot \bar{\mu}_2 \cdot \Delta t < x_{20} \cdot \bar{\mu}_1 + \lambda_2 \cdot \bar{\mu}_1 \cdot \Delta t - \bar{\mu}_1 \cdot \bar{\mu}_2 \cdot \Delta t$$

note that $x_{10} \cdot \bar{\mu}_2$ cancels $x_{20} \cdot \bar{\mu}_1$ which yields

$$\lambda_1 \cdot \bar{\mu}_2 < \lambda_2 \cdot \bar{\mu}_1$$

which is always true for any line in between line 1 and line 3. Hence, our claim is correct. This concludes the proof of Fact 4. □

Facts 3 and 4 together allow us to make the following statement: if the system ever hits line 1 in Figure 2.4, it will move to line 2, which is specified previously. And once line 2 is crossed, the system never evolves below that line. Hence, the service rates, $\mu_1(t)$ and $\mu_2(t)$, given to the buffers are greater than the corresponding arrival rates, λ_1 and λ_2 by at least a constant, positive δ. That is, $\exists \delta > 0$ such that

$$\mu_1(t) > \lambda_1 + \delta \qquad and \qquad \mu_2(t) > \lambda_2 + \delta$$

So, both of the queue lengths will decrease by a positive amount, lower bounded byδ, after the system hits line 2. As a consequence, the buffer length will be zero in a *finite* amount of time.

To sum up, Fact 1 showed that if the queue lengths ever hit the line characterized by $\frac{x_{10}}{\lambda_1} = \frac{x_{20}}{\lambda_2}$, then both of the buffer lengths will eventually be zero in a *finite* amount of time. Facts 2, 3 and 4 altogether showed that even when $\frac{x_{10}}{\lambda_1} \neq \frac{x_{20}}{\lambda_2}$, the queue lengths will evolve in such a manner that it will take a *finite* amount of time for both of the service rates, $\mu_1(t)$ and $\mu_2(t)$, to be strictly greater than the corresponding arrival rates, λ_1 and λ_2. From that point on, it will again take a *finite* time for both of the buffers to empty. Therefore, all possible initial conditions have been covered in the four facts and we were able to prove our claim of a *finite* queue draining time from any initial queue lengths.

This result can easily be demonstrated by a computer simulation. One such example is illustrated in Fig. 2.5 for the case where $\lambda_2 = 2\lambda_1$. It is observed from the figure that the queue lengths would evolve towards the straight line in the figure and would move towards the origin.

2.2.2 N-User Case Analysis

For the $N > 2$ user case, we will make use of the intuition from the two-user scenario described in the previous subsection. We will define a Lyapunov function and show that its derivative is negative. For this purpose, we make the following observation from the 2-user case: independent of the initial queue lengths, (x_{10}, x_{20}), the difference $\frac{x_1(t)}{x_2(t)} - \frac{\lambda_1}{\lambda_2} \to 0$.

Assuming that the system is stabilizable, i.e. $(\lambda_1, \lambda_2, ..., \lambda_N)$ satisfies the stability condition given by (2.1), we define the Lyapunov function $V(x_i(t), x_j(t))$ as follows:

$$V(x_i(t), x_j(t)) = \frac{1}{2}\left(\frac{x_i(t)}{x_j(t)} - \frac{\lambda_i}{\lambda_j}\right)^2 \qquad i, j \in \{1, 2, ..., N\} \qquad i \neq j. \tag{2.3}$$

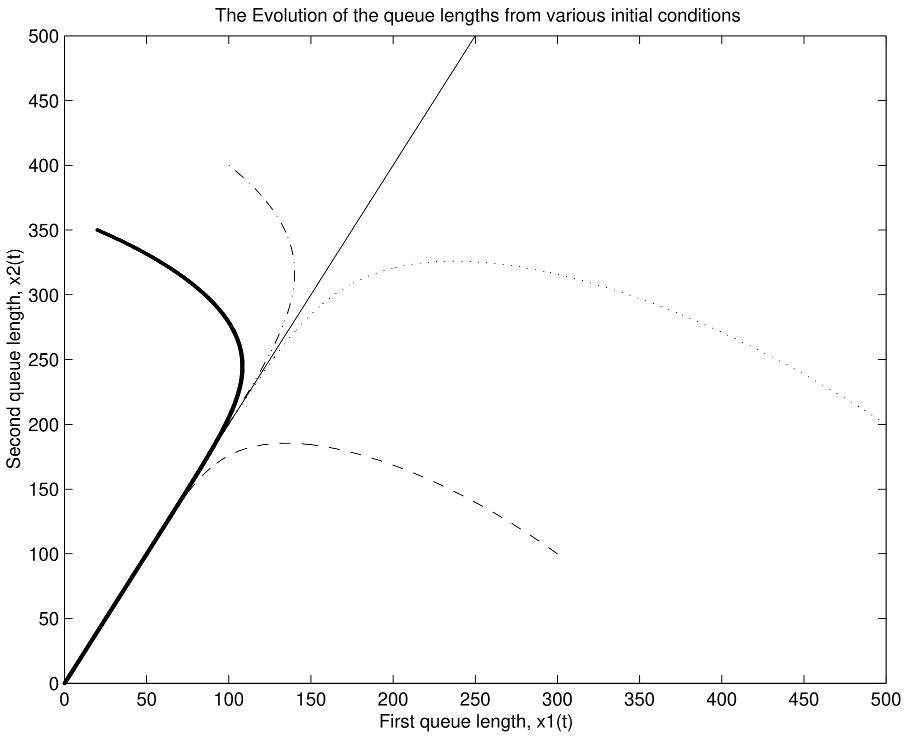

Figure 2.5 Evolution of the queue lengths from different initial starting points

Taking the derivative of (2.3) with respect to t:

$$\dot{V}(x_i, x_j) = \left(\frac{x_i}{x_j} - \frac{\lambda_i}{\lambda_j}\right)\left(\frac{\dot{x}_i x_j - \dot{x}_j x_i}{x_j^2}\right)$$

Noting that $\dot{x}_i = \lambda_i - \mu_i$,

$$\dot{V}(x_i, x_j) = \frac{1}{x_j^2}\left(\frac{x_i\lambda_j - x_j\lambda_i}{x_j\lambda_j}\right)(x_j(\lambda_i - \mu_i) - x_i(\lambda_j - \mu_j))$$

From the definition of Policy 1 in Section 2.2, it is easy to note that each flow satisfies $\mu_i = K(x_1, x_2, ..., x_N)x_i$ for some $K(x_1, x_2, ..., x_N)$. Substituting this into the previous equation yields:

$$\dot{V}(x_i, x_j) = \frac{1}{x_j^3\lambda_j}(x_i\lambda_j - x_j\lambda_i)(x_j(\lambda_i - K(x_1, x_2, ..., x_N)x_i) - x_i(\lambda_j - K(x_1, x_2, ..., x_N)x_j))$$

From this, we obtain

$$\dot{V}(x_i, x_j) = \frac{-1}{x_j^3\lambda_j}(x_i\lambda_j - x_j\lambda_i)^2 \leq 0 \tag{2.4}$$

with equality only when $\frac{x_i}{\lambda_i} = \frac{x_j}{\lambda_j}$. Also as a technicality, we note that it is assumed that none of the buffers is empty initially. This assumption is easily justifiable by noting that if any of the buffers is empty, no service will be offered to it, and it will become non-empty immediately.

Expression (2.4) implies that, for each pair of buffers, their queue length ratio has to approach their arrival rate ratio. Further, if $\frac{x_1}{\lambda_1} = \frac{x_2}{\lambda_2} = ... = \frac{x_N}{\lambda_N}$, then the system stays on that line.

To see that the queues drain in a *finite* time from *any* initial condition, we can write (2.4) as follows

$$\frac{dV(x_i(t), x_j(t))}{dt} = -\frac{\lambda_j}{x_j(t)}V(x_i(t), x_j(t))$$

$$
\begin{aligned}
\frac{dV(x_i(t), x_j(t))}{V(x_i(t), x_j(t))} &= -\frac{\lambda_j}{x_j(t)} dt \\
&\leq -\frac{\lambda_j}{x_{j0} + \lambda_j t} dt,
\end{aligned}
$$

where the last inequality is true since $x_j(t) \leq x_{j0} + \lambda_j t \quad \forall t \geq 0$. Now, taking the integral of both sides from time zero up to some $t > 0$ yields

$$
\begin{aligned}
\int_0^t \frac{dV(x_i(t), x_j(t))}{V(x_i(t), x_j(t))} &\leq -\int_0^t \frac{\lambda_j}{x_{j0} + \lambda_j t} dt \\
\ln\left(\frac{V(x_i(t), x_j(t))}{V(x_i(0), x_j(0))}\right) &\leq \ln\left(\frac{1}{x_{j0} + \lambda_j t}\right) \\
V(x_i(t), x_j(t)) &\leq \frac{V(x_i(0), x_j(0))}{x_{j0} + \lambda_j t}
\end{aligned}
$$

If we define $f(t) = \max_{j=1,\ldots,N} (x_{j0} + \lambda_j t) \quad \forall t$ we can write the last inequality as follows.

$$
V(x_i(t), x_j(t)) \leq \frac{V(x_i(0), x_j(0))}{f(t)} \qquad \forall (i, j) \in \{1, 2, ..., N\} \qquad i \neq j.
$$

Since $f(t)$ is an increasing, positive function of t, for any $\epsilon > 0$, $\exists T_\epsilon < \infty$ such that for all $t > T_\epsilon$ we have

$$
V(x_i(t), x_j(t)) < \epsilon.
$$

Hence, if we take ϵ small enough, then it is possible to get the rates as close as we like to $(\hat{\mu}_1, \hat{\mu}_2, ..., \hat{\mu}_N)$ in a finite amount of time. Also from our assumption that the arrival rates lie within the capacity region, there exists a constant $\delta > 0$ satisfying $\hat{\mu}_i > \lambda_i + \delta$. Combining the two results enable us to conclude that, by making $\epsilon > 0$ small enough it is possible to obtain service rates $\mu_i(t) > \lambda_i + \Delta$ for all $t > T_\epsilon$ and $i = 1, 2, ..., N$, where $\delta > \Delta > 0$. Hence, all of the buffers will drain at least at a positive, constant rate of Δ, which proves that they will empty in a *finite* amount of time. This proves the *throughput optimality* of Policy 1 for the general N-user case.

2.3 Weighted Throughput Optimal Scheduling Policy

In this section, we aim to improve the flexibility of Policy 1, by weighting the queue length of queue i by a constant $K_i > 0$. The definition of the new policy is provided next.

Policy 2. At each time instant, t, allocate the service rates, $(\mu_1(t), \mu_2(t), ..., \mu_N(t))$, such that

$$\frac{\mu_1(t)}{K_1 x_1(t)} = \frac{\mu_2(t)}{K_2 x_2(t)} = ... = \frac{\mu_N(t)}{K_N x_N(t)}$$

and $(\mu_1(t), \mu_2(t), ..., \mu_N(t))$ lies on the boundary of the broadcast capacity region.

Note that the simple scheduling policy analyzed in Section 2.2 is just a special case of this policy with $K_i = const$ for all $i = 1, ..., N$. For the purpose of analysis, let us take $N = 2$. Then the policy proposed gives the following relationship :

$$\frac{\mu_1(t)}{\mu_2(t)} = K \frac{x_1(t)}{x_2(t)} \qquad where \qquad K = \frac{K_1}{K_2}$$

Keeping in mind that $x_1(t)$ and $x_2(t)$ must still satisfy the state equations given in Equation 2.2, we want to find a line characterized by $\frac{x_1(t)}{x_2(t)} = c$, such that if the queue lengths ever hit (or start on) this line, they will continue to stay on it.

For the ratio $\frac{x_1(t)}{x_2(t)}$ to be constant, the ratio of the change in the queue lengths should also be equal to the same constant, i.e. $\frac{\dot{x}_1(t)}{\dot{x}_2(t)} = c$. This is due to the fact that:

$$\frac{x_1(t)}{x_2(t)} = \frac{x_{10} + \int_0^t \dot{x}_1(s)ds}{x_{20} + \int_0^t \dot{x}_2(s)ds} = \frac{cx_{20} + \int_0^t \dot{x}_1(s)ds}{x_{20} + \int_0^t \dot{x}_2(s)ds} = c$$

only when $\frac{\dot{x}_1(t)}{\dot{x}_2(t)} = c$. Then the service rates offered to the users will also be constants and are denoted by $\hat{\mu}_1$ and $\hat{\mu}_2$. Thus, the constant c must satisfy

$$c = \frac{\dot{x}_1(t)}{\dot{x}_2(t)} = \frac{\hat{\mu}_1 - \lambda_1}{\hat{\mu}_2 - \lambda_2} \qquad \textit{and} \qquad \frac{\hat{\mu}_1}{\hat{\mu}_2} = Kc. \tag{2.5}$$

Figure 2.6 illustrates the relationship between K and c. Given K, the parameter c is chosen such that the line denoted as *line*(K) is *parallel* to the line that is specified by $\frac{x_1(t)}{x_2(t)} = c$.

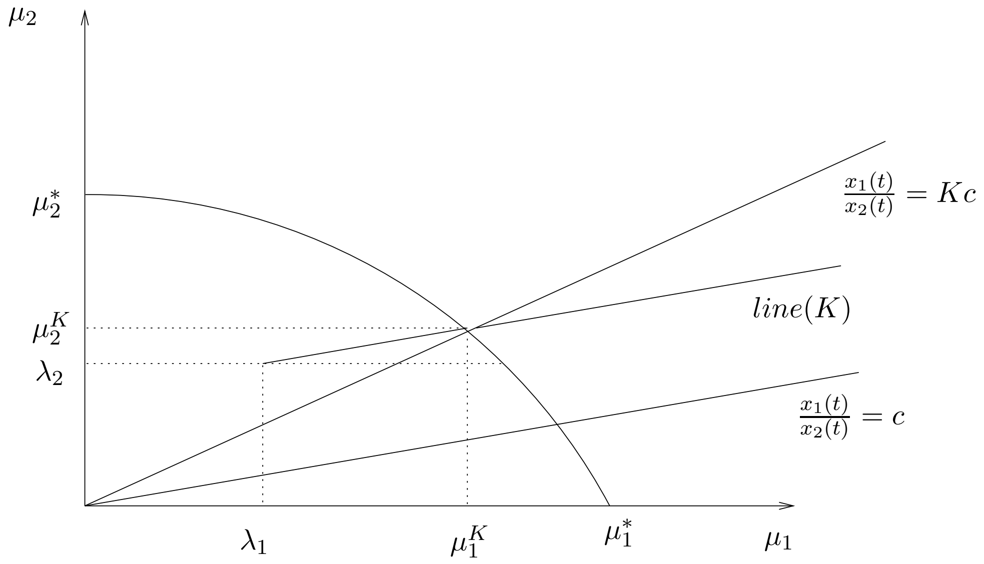

Figure 2.6 Pictorial illustration of the weighted throughput optimal policy

To show that Policy 2 is a throughput optimal policy, we will assume that (λ_1, λ_2) lies within the capacity region, and we will consider the following Lyapunov function:

$$V(x_1(t), x_2(t)) = \frac{1}{2}\left(\frac{x_1(t)}{x_2(t)} - c\right)^2. \tag{2.6}$$

Then,

$$\dot{V}(x_1, x_2) = \frac{1}{x_2^3}(x_1 - cx_2)((\lambda_1 - \mu_1)x_2 - (\lambda_2 - \mu_2)x_1). \tag{2.7}$$

At this point, we need to show that $\dot{V}(x_1, x_2) < 0$ when $\frac{x_1}{x_2} \neq c$ and $\dot{V}(x_1, x_2) = 0$ when $\frac{x_1}{x_2} = c$. For this purpose, we will consider several cases.

Case 1. If the queue lengths satisfy $\frac{x_1}{x_2} = c$, it is clear from (2.7) that $\dot{V}(x_1, x_2) = 0$. That is, the ratio $\frac{x_1}{x_2} = c$ for all $t > 0$.

Furthermore, the service rates provided to the queues will be constant, denoted by μ_1^K and μ_2^K in Figure 2.6. Using the property given by (2.5), we can easily get the following equality:

$$\frac{\mu_1^K}{\mu_2^K} = K \frac{\mu_1^K - \lambda_1}{\mu_2^K - \lambda_2}.$$

The left-hand side of this equation is always positive, hence the right-hand side needs to be positive as well. Therefore, we either have $\mu_1^K > \lambda_1$ and $\mu_2^K > \lambda_2$, or $\mu_1^K < \lambda_1$ and $\mu_2^K < \lambda_2$. But the latter case is impossible since (λ_1, λ_2) lies *within* the capacity region, whereas (μ_1^K, μ_2^K) lies *on* the boundary of the capacity region. So μ_1^K and μ_2^K are strictly greater than λ_1 and λ_2 respectively, by a constant, positive value. This observation, leads to the fact that, all of the buffer levels will be zero in a *finite* amount of time.

Case 2. The queue lengths are such that, $\frac{x_1}{x_2} > c$ and $\mu_2 < \lambda_2$, where (μ_1, μ_2) is the service rate pair lying of the boundary of the capacity region satisfying $\frac{\mu_1}{\mu_2} = K\frac{x_1}{x_2}$. Note that, since $\frac{x_1}{x_2} > c$, we have $\mu_1 > \mu_1^K > \lambda_1$ because we know that $\frac{x_1}{x_2}$ can at most attain a value of c, and if it is actually equal to c, then we proved in Case 1 that it will always stay as c.

Then under these conditions, considering the right-hand side of (2.7), we have $(x_1 - cx_2) > 0$ and $((\lambda_1 - \mu_1)x_2 - (\lambda_2 - \mu_2)x_1) < 0$ since $(\lambda_1 - \mu_1) < 0$ and $(\lambda_2 - \mu_2) > 0$. Therefore, we have showed that $\dot{V}(x_1, x_2) < 0$ as long as $\frac{x_1}{x_2} > c$ and the corresponding $\mu_2 < \lambda_2$.

Case 3. If the queue lengths are such that, $\frac{x_1}{x_2} > c$ and $\mu_2 = \lambda_2$, then clearly we would have $((\lambda_1 - \mu_1)x_2 - (\lambda_2 - \mu_2)x_1) < 0$ in (2.7), since $(\lambda_1 - \mu_1) < 0$ and $(\lambda_2 - \mu_2) = 0$. Consequently, we would have $\dot{V}(x_1, x_2) < 0$.

Case 4. If the queue lengths are such that, $\frac{x_1}{x_2} > c$ and $\mu_2 > \lambda_2$, then in order to show that $\dot{V}(x_1, x_2) < 0$ we use the following argument.

Note that, $\exists m_0 > 0$ such that, when $\frac{x_2}{x_1} = m_0 > 0$, we have $\frac{\mu_2 - \lambda_2}{\mu_1 - \lambda_1} = 0$. In this statement, $m_0 > 0$ since when $\mu_2 = \lambda_2 > 0$ we can never have $x_2 = 0$, which is the only case for which $\frac{x_2}{x_1} = 0$. And when $\frac{x_2}{x_1} = \frac{1}{c} := m_1$, we have $\frac{\mu_2^K - \lambda_2}{\mu_1^K - \lambda_1} = m_1$. We observe that the slope $\frac{\mu_2 - \lambda_2}{\mu_1 - \lambda_1}$ as a function of $\frac{x_2}{x_1}$ is continuous. And since $m_0 > 0$, a possible plot of $\frac{\mu_2 - \lambda_2}{\mu_1 - \lambda_1}$ versus $\frac{x_2}{x_1}$ is shown in Figure 2.7.

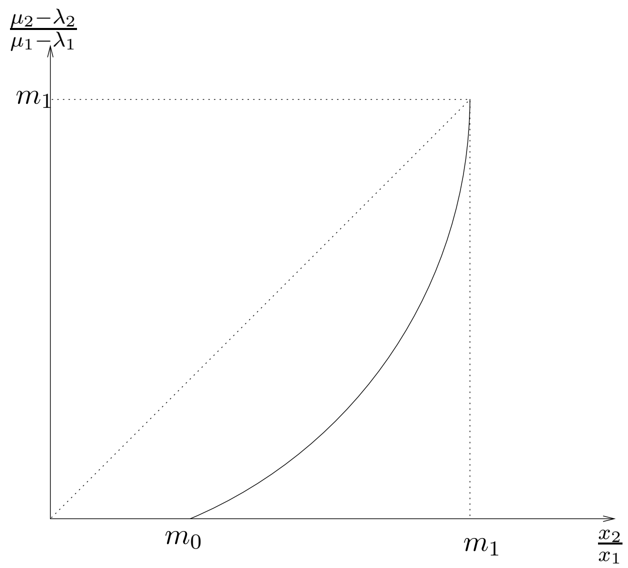

Figure 2.7 Relationship between the slopes

Due to continuity of $\frac{\mu_2-\lambda_2}{\mu_1-\lambda_1}$ as a function of $\frac{x_2}{x_1}$ and $m_0 > 0$, we obtain the following inequality: $\frac{\mu_2-\lambda_2}{\mu_1-\lambda_1} < \frac{x_2}{x_1}$ when $\frac{x_1}{x_2} > c$ and $\mu_2 > \lambda_2$, which in turn implies that $((\lambda_1-\mu_1)x_2-(\lambda_2-\mu_2)x_1) < 0$ in (2.7). Consequently, we have $\dot{V}(x_1, x_2) < 0$ for this case also.

To summarize, we have shown that the derivative of the Lyapunov function is always negative for any condition satisfying $\frac{x_1}{x_2} > c$. The case when $\frac{x_1}{x_2} < c$ follows from a similar reasoning, and gives $\dot{V}(x_1, x_2) < 0$.

In conclusion, we are able to prove that $\dot{V}(x_1, x_2) < 0$ when $\frac{x_1}{x_2} \neq c$ and $\dot{V}(x_1, x_2) = 0$ when $\frac{x_1}{x_2} = c$. Further, in Case 1, we showed that the queue lengths are emptied in a finite amount of time. Therefore, just like we have argued for the case when $K = 1$, we can come as close as we like to the line characterized with $\frac{x_1}{x_2} = c$ in some finite duration, and from that point on it will take another finite amount of time for both of the buffers to be emptied. Hence, the weighted scheduling policy is shown to be throughput optimal for the 2-user case.

The same idea can be used for the N-user case just like we have done for the simple scheduling policy in Subsection 2.2.2.

CHAPTER 3

ORIGINAL SYSTEM MODEL ANALYSIS

Motivated by the results obtained in Chapter 2, in this chapter we aim to prove that the suggested scheduling policy stabilizes the original system model described in Section 1.3. Two main theorems are given for this purpose. Before going any further, it is useful to define the Gaussian broadcast channel capacity region for N users with a total average power constraint, P, by

$$\mathcal{C}^N = \{\vec{\mu}[k] : \mu_i[k] \leq \frac{1}{2}\log\left(1 + \frac{\alpha_i[k]P}{\sum_{j=1}^{i-1}\alpha_j[k]P + \sigma_i^2}\right)\}$$

for $i = 1, \ldots, N$ and where $\sum_{i=1}^{N}\alpha_i[k] = 1 \quad \forall k$.

Theorem 1. *Considering the system described in Section 1.3, if the second moment of all the arrivals are bounded, i.e.* $E[a_i^2] < A_i < \infty$, *and* $(\lambda_1, \lambda_2, \ldots, \lambda_N)$ *lies within* $\mathcal{C}^N$ *then with the scheduling policy, which allocates rates at the beginning of each time slot in such a way that*

$$\frac{\mu_1[k]}{x_1[k]} = \frac{\mu_2[k]}{x_2[k]} = \ldots = \frac{\mu_N[k]}{x_N[k]}, \tag{3.1}$$

where $(\mu_1[k], \mu_2[k], \ldots, \mu_N[k])$ *lies on the boundary of* $\mathcal{C}^N$, *we have*

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{n=0}^{T-1} E\left[\sum_{i=1}^{N} x_i[n]\right] \leq L < \infty \tag{3.2}$$

Proof. In order to prove Theorem 1 we will make use of two different Lyapunov functions. First, we consider the Lyapunov function

$$V(\vec{x}) = \frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{j-1}(x_i\lambda_j - x_j\lambda_i)^2 \qquad where \qquad \vec{x} = (x_1, \ldots, x_N)^T. \tag{3.3}$$

Now, we will compute the conditional mean

$$J_k(\vec{x}) := E\left[V(\vec{x}[k+1]) - V(\vec{x}[k]) \quad | \quad \vec{x}[k] = \vec{x}\right] \tag{3.4}$$

Defining $\Delta x_i[k] := (x_i[k+1] - x_i[k])$ for $i = 1, \ldots, N$ and consequently substituting $x_i[k+1] = x_i[k] + \Delta x_i[k]$ in the previous equation yields

$$J_k(\vec{x}) = E\left[\frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{j-1}((x_i[k] + \Delta x_i[k])\lambda_j - (x_j[k] + \Delta x_j[k])\lambda_i)^2 - (x_i[k]\lambda_j - x_j[k]\lambda_i)^2 \quad |\vec{x}[k] = \vec{x}\right].$$

Now, since all the time parameters are k in the preceding equation, we simply drop them for easy tracking and continue with the calculation as follows

$$\begin{aligned}
&= \frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{j-1}E\left[((x_i\lambda_j - x_j\lambda_i) + (\Delta x_i\lambda_j - \Delta x_j\lambda_i))^2 - (x_i\lambda_j - x_j\lambda_i)^2 \quad | \quad \vec{x}[k] = \vec{x}\right] \\
&= \frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{j-1}E[(x_i\lambda_j - x_j\lambda_i)^2 + 2(x_i\lambda_j - x_j\lambda_i)(\Delta x_i\lambda_j - \Delta x_j\lambda_i) \\
&\quad +(\Delta x_i\lambda_j - \Delta x_j\lambda_i)^2 - (x_i\lambda_j - x_j\lambda_i)^2 \quad | \quad \vec{x}[k] = \vec{x}] \\
&= \sum_{j=1}^{N}\sum_{i=1}^{j-1}E\left[(x_i\lambda_j - x_j\lambda_i)(\Delta x_i\lambda_j - \Delta x_j\lambda_i) \quad | \quad \vec{x}[k] = \vec{x}\right] && (3.5) \\
&\quad +\sum_{j=1}^{N}\sum_{i=1}^{j-1}\frac{1}{2}E\left[(\Delta x_i\lambda_j - \Delta x_j\lambda_i)^2 \quad | \quad \vec{x}[k] = \vec{x}\right] && (3.6)
\end{aligned}$$

First, we concentrate on (3.5). Recall that the evolution of the i^{th} buffer length is given by

$$x_i[k+1] = x_i[k] + a_i[k] - \mu_i[k] + u_i[k],$$

where a_i denotes the arrival process, μ_i denotes the service rate provided to the i^{th} user according to our scheduling policy and finally u_i denotes the excess service provided in the case when i^{th} buffer is emptied. Therefore we can substitute $\Delta x_i = a_i - \mu_i + u_i$ and obtain

$$\sum_{j=1}^{N}\sum_{i=1}^{j-1} E\left[(x_i\lambda_j - x_j\lambda_i)(\Delta x_i\lambda_j - \Delta x_j\lambda_i) \quad | \quad \vec{x}[k] = \vec{x}\right]$$

$$\begin{aligned}
&= \sum_{j=1}^{N}\sum_{i=1}^{j-1}(x_i\lambda_j - x_j\lambda_i)E\left[(\Delta x_i\lambda_j - \Delta x_j\lambda_i) \quad | \quad \vec{x}[k] = \vec{x}\right] \\
&= \sum_{j=1}^{N}\sum_{i=1}^{j-1}(x_i\lambda_j - x_j\lambda_i)\left(\lambda_j\lambda_i - \lambda_j\mu_i + \lambda_j E[u_i|\vec{x}[k] = \vec{x}] - \lambda_i\lambda_j + \lambda_i\mu_j - \lambda_i E[u_j|\vec{x}[k] = \vec{x}]\right) \\
&= \sum_{j=1}^{N}\sum_{i=1}^{j-1}\left[(x_i\lambda_j - x_j\lambda_i)\left(-K(\vec{x})(x_i\lambda_j - x_j\lambda_i) + \lambda_j E[u_i|\vec{x}[k] = \vec{x}] - \lambda_i E[u_j|\vec{x}[k] = \vec{x}]\right)\right],
\end{aligned}$$

where in the last step we made use of the fact that with our policy there exists a positive value, $K(\vec{x})$, such that $\mu_i[k] = K(\vec{x})x_i[k]$ for $i = 1, \ldots, N$. So $J_k(\vec{x})$ is given by

$$\begin{aligned}
J_k(\vec{x}) &= \sum_{j=1}^{N}\sum_{i=1}^{j-1}\left[-K(\vec{x})(x_i\lambda_j - x_j\lambda_i)^2\right] \\
&\quad + \sum_{j=1}^{N}\sum_{i=1}^{j-1}\left[(x_i\lambda_j - x_j\lambda_i)(\lambda_j E[u_i|\vec{x}[k] = \vec{x}] - \lambda_i E[u_j|\vec{x}[k] = \vec{x}])\right] \\
&\quad + \sum_{j=1}^{N}\sum_{i=1}^{j-1} E\left[\frac{1}{2}(\Delta x_i\lambda_j - \Delta x_j\lambda_i)^2 \quad | \quad \vec{x}[k] = \vec{x}\right]. \qquad (3.7)
\end{aligned}$$

Next, we obtain a lower bound on $K(\vec{x})$. For this purpose, note that $K(\vec{x})$ is given by

$$K(\vec{x}) = \frac{\mu_1}{x_1} = \ldots = \frac{\mu_N}{x_N} = \frac{\sqrt{\mu_1^2 + \ldots + \mu_N^2}}{\sqrt{x_1^2 + \ldots + x_N^2}},$$

where $(\mu_1, \ldots, \mu_N)$ is a point on the *boundary* of the capacity region, $\mathcal{C}^N$, and consequently $\sqrt{\sum_{i=1}^{N}\mu_i^2}$ must be greater than a positive value, say M, for all queue lengths. With this observation, it is straight forward to lower bound $K(\vec{x})$ as follows:

$$K(\vec{x}) \geq \frac{M}{\sqrt{\sum_{i=1}^{N} x_i^2}} = \frac{M}{r} \qquad where \qquad M = \min_{\vec{\mu} \in Boundary(\mathcal{C}^N)} \sqrt{\sum_{i=1}^{N} \mu_i^2}$$

At this point it is convenient to convert the system into polar coordinates, where at each time instant, k: $r[k] = \sqrt{\sum_{i=1}^{N} x_i^2}$ gives the distance of the queue lengths from the origin, and $\theta_i[k]$ is the angle between $\vec{x}$ and $\hat{x}_i$, where $\hat{x}_i$ is the unit vector in the direction of the i^{th} axis. Hence we substitute $x_i = r\cos\theta_i, \quad i = 1, \ldots, N$ dropping the time variable, k, for easy tracking. Then (3.7) becomes

$$\begin{aligned}
J_k(\vec{x}) \quad &= \quad -Mr\sum_{j=1}^{N}\sum_{i=1}^{j-1}(\lambda_j\cos\theta_i - \lambda_i\cos\theta_j)^2 && (3.8)\\
&+ \sum_{j=1}^{N}\sum_{i=1}^{j-1}\left[r(\lambda_j\cos\theta_i - \lambda_i\cos\theta_j)(\lambda_j E[u_i|\vec{x}[k]=\vec{x}] - \lambda_i E[u_j|\vec{x}[k]=\vec{x}])\right] && (3.9)\\
&+ \frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{j-1} E\left[(\Delta x_i\lambda_j - \Delta x_j\lambda_i)^2 \quad | \quad \vec{x}[k]=\vec{x}\right]. && (3.10)
\end{aligned}$$

(3.10) is upper bounded by a finite value, say B_1, since the second moment of each the arrivals are assumed to be upper bounded in the theorem. Also for large enough r, (3.9) can never be positive. To see this let us analyze each of the terms in the summation. Taking any (i, j) pair, if the angles are such that $(\lambda_j\cos\theta_i - \lambda_i\cos\theta_j) < 0$, then $\frac{\cos\theta_i}{\cos\theta_j} = \frac{x_i}{x_j} < \frac{\lambda_i}{\lambda_j}$ which implies that even if θ_i is small enough to cause a nonzero unused service for the i^{th} queue, we can find a large enough R value such that, $\forall r > R$, both the previous inequality holds and x_j is made large enough so that $E[u_j|\vec{x}] = 0$. Thus giving

$$(\lambda_j\cos\theta_i - \lambda_i\cos\theta_j)(\lambda_j E[u_i|\vec{x}[k]=\vec{x}] - \lambda_i E[u_j|\vec{x}[k]=\vec{x}]) \leq 0.$$

A similar argument holds for the case $(\lambda_j \cos\theta_i - \lambda_i \cos\theta_j) > 0$. Hence, (3.9) can never be positive. Putting all these results together enables us to upper bound $J_k(\vec{x})$ as follows:

$$J_k(\vec{x}) \leq -Mr \sum_{j=1}^{N} \sum_{i=1}^{j-1} (\lambda_j \cos\theta_i - \lambda_i \cos\theta_j)^2 + B_1 \tag{3.11}$$

Now, we can segment the N-dimensional vector space into three regions, S_1, S_2, S_3 as follows: S_1 denotes the region for which $r \leq R < \infty$ and $0 \leq \theta_i \leq \pi/2$ for $i = 1, \ldots, N$. S_2 is the region for which $r > R > 0$ and θ_i for $i = 1, \ldots, N$ satisfies $\sum_{j=1}^{N} \sum_{i=1}^{j-1} (\lambda_j \cos(\theta_i) - \lambda_i \cos(\theta_j))^2 \leq l$, where $l > 0$ is chosen such that rate allocation, $\vec{\mu}$ for points inside S_2 yields rates that are all strictly larger than corresponding mean arrival rates, i.e. $\mu_i > \lambda_i$ for $i = 1, \ldots, N$. Such an $l > 0$ and corresponding θ_i values exist due to the convexity and continuity of the capacity region, $\mathcal{C}^N$. And finally S_3 denotes the rest of the vector space not covered by S_1 and S_2. At this point, note that $S_1 \cup S_2 \cup S_3 = \Re^N$ and they are non-intersecting. To get an insight into the segmentation, the two user case is plotted in Figure 3.1.

Arranging the terms in (3.11) by recalling the definition of $J_k(\vec{x})$ in (3.4)

$$\begin{aligned}
B_1 - E[V(\vec{x}[k+1])|\vec{x}] + V(\vec{x}[k]) &\geq Mr[k] \sum_{j=1}^{N} \sum_{i=1}^{j-1} (\lambda_j \cos\theta_i[k] - \lambda_i \cos\theta_j[k])^2 \\
&\geq Mr[k] \sum_{j=1}^{N} \sum_{i=1}^{j-1} \left[(\lambda_j \cos\theta_i[k] - \lambda_i \cos\theta_j[k])^2 I_{\vec{x}[k] \in S_3} \right] \\
&\geq Mr[k] l I_{\vec{x}[k] \in S_3},
\end{aligned}$$

where the second inequality follows from the fact that each of $(\lambda_j \cos\theta_i[k] - \lambda_i \cos\theta_j[k])^2$ is nonnegative and $I_{\vec{x}[k] \in S_1} + I_{\vec{x}[k] \in S_2} + I_{\vec{x}[k] \in S_3} = 1 \quad \forall \vec{x}[k]$. Here, I_A denotes the indicator function getting a value of 1 if the event A is true, and is zero otherwise. The third inequality is true simply because $\forall \vec{x}[k] \in S_3$ we have $(\lambda_j \cos\theta_i[k] - \lambda_i \cos\theta_j[k])^2 > l > 0$ by definition. First

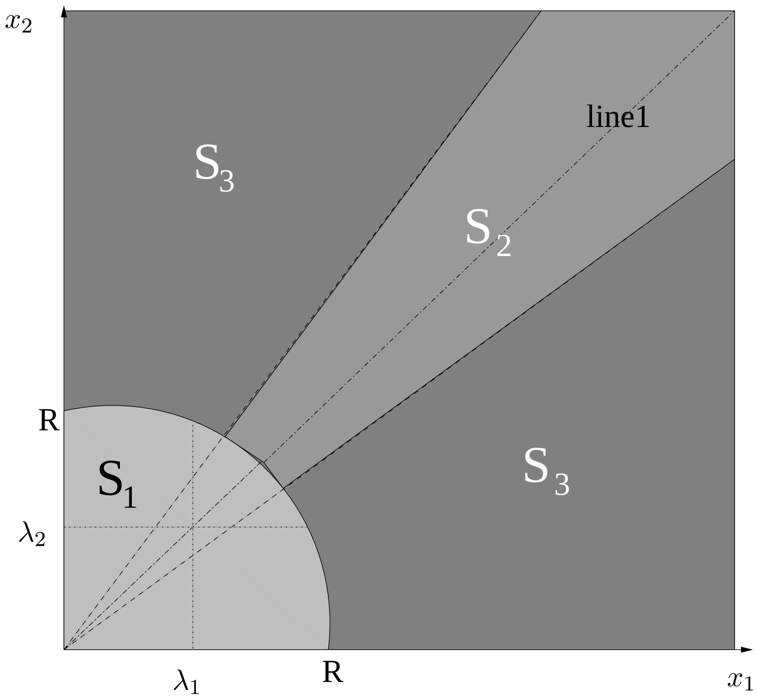

Figure 3.1 Segmentation of the $x_1 - x_2$ plane

taking expectation over $\vec{x}$ and then summing over the time instants from 0 up to $T-1$ and finally dividing by T both of the sides of the last inequality yields:

$$\begin{aligned} Ml\frac{1}{T}\sum_{n=0}^{T-1} E[r[n]I_{\vec{x}[n]\in S_3}] &\leq B_1 - \frac{1}{T}\sum_{n=0}^{T-1}[E[V(\vec{x}[n+1]) - V(\vec{x}[n])]] \\ &= B_1 - \frac{E[V(\vec{x}[T])]}{T} + \frac{E[V(\vec{x}[0])]}{T} \\ &\leq B_1 + \frac{E[V(\vec{x}[0])]}{T}. \end{aligned}$$

So we have

$$\frac{1}{T}\sum_{n=0}^{T-1} E[r[n]I_{\vec{x}[n]\in S_3}] \leq \frac{B_1}{M\cdot l} + \frac{E[V(\vec{x}[0])]}{T\cdot M\cdot l}. \tag{3.12}$$

Also by the definition of S_1 we can say that

$$\frac{1}{T}\sum_{n=0}^{T-1} E[r[n]I_{\vec{x}[n]\in S_1}] \leq R \tag{3.13}$$

Combining (3.12) and (3.13), we get

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{n=0}^{T-1} E[r[n] I_{\vec{x}[n]\in(S_1\cup S_3)}] \quad \leq \quad R + \frac{B_1}{M\cdot l} \tag{3.14}$$

This result implies that if the system is observed for a long time, the mean distance of the queue lengths from the origin is upper bounded as long as $\vec{x}[k] \in (S_1 \cup S_3)$. We will return to this result in the following pages.

Now, for the second part of the proof let us concentrate on another Lyapunov function, namely

$$W(\vec{x}) = \sum_{i=1}^{N} x_i^2 \tag{3.15}$$

Similar to the derivation in the first part of the proof we will compute the conditional mean defined by

$$U_k(\vec{x}) := E[W(\vec{x}[k+1]) - W(\vec{x}[k]) \quad | \quad \vec{x}[k] = \vec{x}]. \tag{3.16}$$

Substituting (3.15) into (3.16) yields

$$\begin{aligned} U_k(\vec{x}) \quad &= \quad E\left[\sum_{i=1}^{N}\left(x_i^2[k+1] - x_i^2[k]\right) \quad | \quad \vec{x}[k] = \vec{x}\right] \\ &= \quad \sum_{i=1}^{N} E\left[\left(x_i^2[k+1] - x_i^2[k]\right) \quad | \quad \vec{x}[k] = \vec{x}\right] \end{aligned}$$

Now, since $\Delta x_i[k] = x_i[k+1] - x_i[k]$ we get

$$\begin{aligned} U_k(\vec{x}) \quad &= \quad \sum_{i=1}^{N} E\left[(2x_i + \Delta x_i)\Delta x_i \quad | \quad \vec{x}[k] = \vec{x}\right] \\ &= \quad \sum_{i=1}^{N} 2x_i E[\Delta x_i \quad | \quad \vec{x}[k] = \vec{x}] \qquad (3.17) \\ &\quad\quad + \sum_{i=1}^{N} E\left[(\Delta x_i)^2 \quad | \quad \vec{x}[k] = \vec{x}\right] \qquad (3.18) \end{aligned}$$

If we study the two terms in the last equality separately, we can see that (3.18) can be upper bounded by a finite, positive value, say B_2, because the second moments of all the arrival processes are assumed to be upper bounded in the theorem statement. As for (3.17), we can upper bound $E[\Delta x_i \quad | \quad \vec{x}[k] = \vec{x}]$ as follows.

$$E[\Delta x_i[k] \quad | \quad \vec{x}[k] = \vec{x}] \quad \leq \quad -\delta_1 I_{\vec{x}[k] \in S_2} + \delta_2 I_{\vec{x}[k] \in (S_1 \cup S_3)}.$$

To prove this claim, note that $\Delta x_i[k] = a_i[k] - \mu_i[k] + u_i[k]$. If $\vec{x}[k] \in S_2$, then all of the queue lengths are large enough such that there can be no unused service for any user, i.e. $u_i[k] = 0 \quad \forall i$. Moreover, due to the definition of S_2, for all $\vec{x}[k] \in S_2$, $\mu_i[k]$ is larger than λ_i for all i by at least a positive, constant value, δ_1 in this case. And for all points outside of S_2, Δx_i is clearly upper bounded by a constant, positive amount, taken to be δ_2 here. Inserting this result into (3.17) gives

$$U_k(\vec{x}) = E[W(\vec{x}[k+1]) - W(\vec{x}[k]) \quad | \quad \vec{x}[k] = \vec{x}] \quad \leq \quad \sum_{i=1}^{N} \Big(-2x_i\delta_1 I_{\vec{x}[k] \in S_2} + 2x_i\delta_2 I_{\vec{x}[k] \in (S_1 \cup S_3)} \Big) + B_2$$

Next, we take the expectation of both sides over $\vec{x}$ and obtain

$$\begin{aligned} E[W(\vec{x}[k+1])] - E[W(\vec{x}[k])] \quad &\leq \quad -2\delta_1 E\left[\sum_{i=1}^{N} x_i[k] I_{\vec{x}[k] \in S_2}\right] + 2\delta_2 E\left[\sum_{i=1}^{N} x_i[k] I_{\vec{x}[k] \in (S_1 \cup S_3)}\right] + B_2 \\ &= \quad -2\delta_1 E\left[\sum_{i=1}^{N} x_i[k]\right] + 2\delta_3 E\left[\sum_{i=1}^{N} x_i[k] I_{\vec{x}[k] \in (S_1 \cup S_3)}\right] + B_2, \end{aligned}$$

where $\delta_3 = \delta_1 + \delta_2$. If we re-arrange the terms, we get

$$2\delta_1 E\left[\sum_{i=1}^{N} x_i[k]\right] \quad \leq \quad 2\delta_3 E\left[\sum_{i=1}^{N} x_i[k] I_{\vec{x}[k] \in (S_1 \cup S_3)}\right] - E[W(\vec{x}[k+1])] + E[W(\vec{x}[k])] + B_2.$$

Summing over the time instants from 0 up to $T-1$ and then dividing by T both of the sides of the last inequality yields

$$\begin{aligned}
2\delta_1 \frac{1}{T}\sum_{n=0}^{T-1} E\left[\sum_{i=1}^{N} x_i[n]\right] &\leq 2\delta_3 \frac{1}{T}\sum_{n=0}^{T-1} E\left[\sum_{i=1}^{N} x_i[n] I_{\vec{x}[n]\in(S_1\cup S_3)}\right] + \frac{E[W(\vec{x}[0])]}{T} - \frac{E[W(\vec{x}[T])]}{T} + B_2 \\
&\leq 2\delta_3 \frac{1}{T}\sum_{n=0}^{T-1} E\left[\sum_{i=1}^{N} x_i[n] I_{\vec{x}[n]\in(S_1\cup S_3)}\right] + \frac{E[W(\vec{x}[0])]}{T} + B_2 \\
\frac{1}{T}\sum_{n=0}^{T-1} E\left[\sum_{i=1}^{N} x_i[n]\right] &\leq \frac{\delta_3}{\delta_1}\frac{1}{T}\sum_{n=0}^{T-1} E\left[\sum_{i=1}^{N} x_i[n] I_{\vec{x}[n]\in(S_1\cup S_3)}\right] + \frac{E[W(\vec{x}[0])]}{2\delta_1 T} + \frac{B_2}{2\delta_1}.
\end{aligned}$$

Letting T go to ∞ yields

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{n=0}^{T-1} E\left[\sum_{i=1}^{N} x_i[n]\right] \leq \frac{\delta_3}{\delta_1}\limsup_{T\to\infty} \frac{1}{T}\sum_{n=0}^{T-1} E\left[\sum_{i=1}^{N} x_i[n] I_{\vec{x}[n]\in(S_1\cup S_3)}\right] + \frac{B_2}{2\delta_1}. \quad (3.19)$$

In order to conclude the proof, the right-hand side of (3.19) must be upper bounded. For this purpose, the result obtained in (3.14) can be used. Note that we can write each queue length as $x_i[n] = r[n]\cos(\theta_i[n])$. Therefore,

$$\begin{aligned}
\limsup_{T\to\infty} \frac{1}{T}\sum_{n=0}^{T-1} E\left[\sum_{i=1}^{N} x_i[n] I_{\vec{x}[n]\in(S_1\cup S_3)}\right] &= \limsup_{T\to\infty} \frac{1}{T}\sum_{n=0}^{T-1} E\left[r[n]\sum_{i=1}^{N} \cos(\theta_i[n]) I_{\vec{x}[n]\in(S_1\cup S_3)}\right] \\
&\leq \limsup_{T\to\infty} \frac{1}{T}\sum_{n=0}^{T-1} E\left[r[n] N I_{\vec{x}[n]\in(S_1\cup S_3)}\right] \\
&\leq RN + \frac{B_1 N}{Ml}
\end{aligned}$$

Substituting this result into (3.19) gives us the required upper bound.

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{n=0}^{T-1} E\left[\sum_{i=1}^{N} x_i[n]\right] \leq \frac{\delta_3}{\delta_1}\left(RN + \frac{B_1 N}{Ml}\right) + B_2 := L < \infty. \quad (3.20)$$

This concludes the proof of Theorem 1.

□

Up to this point, we assumed the state space is continuous, i.e. the queue lengths as well as services applied can take on any non-negative real value. Although such an assumption is reasonable for analysis purposes, in a real life scenario the queue lengths can only be non-negative integers (counted as packets, bits etc.). For large enough packet sizes, the conversion from reals to integers would cause some overhead factors in the previous derivations but the results would still hold. In the next theorem, a discrete-state system is assumed and the positive recurrence of the Markov chain is established to be correct under some further conditions.

Theorem 2. *If the queue lengths can only take discrete values, then* $(x_1, \ldots, x_N)$ *forms a discrete-state Markov Chain. For such a system, if, for each user* i*,*

$$\mathcal{P}(a_i[k] = j) > 0 \qquad \forall j \in \mathcal{Z} \backslash \mathcal{Z}^- \tag{3.21}$$

then the Markov Chain is

(i) aperiodic*,*

(ii) has a single communicating class*,*

(iii) is positive recurrent*.*

Proof. (i) Aperiodic:

As long as there exists a self loop at any state of the Markov Chain, the period of that chain can never be greater than 1. In this model $\mathcal{P}(a_i[k] = 0) > 0 \quad \forall i$, and hence there exists a self loop in all the states. So the Markov Chain is aperiodic.

(ii) Single communicating class:

From any initial state, the chain can hit the origin with some non-zero probability due to (3.21). And similarly, from zero any state can be reached with some non-zero probability because of the same reason. Hence, any state can be reached from any other state with a non-zero probability, which proves a single communicating class claim.

(iii) Positive recurrence:

For a given vector $(m_1, \ldots, m_N)$,

$$\begin{aligned}
\limsup_{T\to\infty} \frac{1}{T}\sum_{n=0}^{T-1} \mathcal{P}(x_1[n] = m_1, \ldots, x_N[n] = m_N) &\leq \limsup_{T\to\infty} \frac{1}{T}\sum_{n=0}^{T-1} \mathcal{P}\left(\sum_{i=1}^{N} x_i[n] \geq \sum_{i=1}^{N} m_i\right) \\
&\leq \limsup_{T\to\infty} \frac{1}{T}\sum_{n=0}^{T-1} \frac{E\left[\sum_{i=1}^{N} x_i[n]\right]}{\sum_{i=1}^{N} m_i} \\
&\leq \frac{L}{\sum_{i=1}^{N} m_i},
\end{aligned}$$

where the second inequality comes from Markov's Inequality, and the third inequality is obtained from Theorem 1. If we choose the set $(m_1, \ldots, m_N)$ such that $\sum_{i=1}^{N} m_i > L$, then

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{n=0}^{T-1} \mathcal{P}\left(\sum_{i=1}^{N} x_i[n] \geq \sum_{i=1}^{N} m_i\right) < 1,$$

which implies

$$\liminf_{T\to\infty} \frac{1}{T}\sum_{n=0}^{T-1} \mathcal{P}\left(\sum_{i=1}^{N} x_i[n] \leq \sum_{i=1}^{N} m_i - 1\right) > 0. \tag{3.22}$$

Since there exist a finite number of states that satisfy $\sum_{i=1}^{N} x_i[n] \leq \sum_{i=1}^{N} m_i - 1$, we can claim that $\exists (b_1, \ldots, b_N) : \sum_{i=1}^{N} b_i \leq \sum_{i=1}^{N} m_i - 1$ such that

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{n=0}^{T-1} \mathcal{P}(x_1[n] = b_1, \ldots, x_N[n] = b_N) > 0. \tag{3.23}$$

By contradiction, if such a state does not exist, then (3.22) cannot hold because there are only a finite number of states satisfying $\sum_{i=1}^{N} x_i[n] \leq \sum_{i=1}^{N} m_i - 1$. Equation (3.23) implies that $(b_1, \ldots, b_N)$ is a positive recurrent state. It also a well known fact that [14], for an aperiodic Markov Chain with a single communicating class, if any of the states is positive recurrent, then the whole Markov Chain is positive recurrent.

□

Positive recurrence property of the Markov Chain proved in Theorem 2 implies a finite mean recurrence time for all of the states. Furthermore, the chain has a stationary distribution. We say that such a Markov chain is *stable*.

CHAPTER 4

SIMULATIONS

At the beginning of each time slot, the proposed policy divides the total available power among the users in such a manner that the service provided to each receiver is proportional to the corresponding queue lengths at that instant. Such an allocation is reasonable since the queue lengths indirectly provides information on the arrival rates.

For the purpose of demonstration a two user system is simulated, where the arrival processes are taken to be independent with their means lying inside the capacity region of the broadcast channel. Through simulations it will be argued that, not only is the policy stabilizing, but also, it is fair in the sense that the mean waiting time of the packets is the same in all the queues.

As a first example let us assume that the arrival to the first queue is Poisson with mean, $\lambda_1 = 5$, whereas arrival process to the second queue is uniform with mean, $\lambda_2 = 10$. By choosing different arrival processes for separate queues, we want to show that the policy is independent of the characteristics of the arrivals, but only depend on their means. Note that, the policy need not know the arrival processes in order to operate. The mean arrival rates, (5,10), lie inside and very close to the boundary of the capacity region. The initial queue lengths are

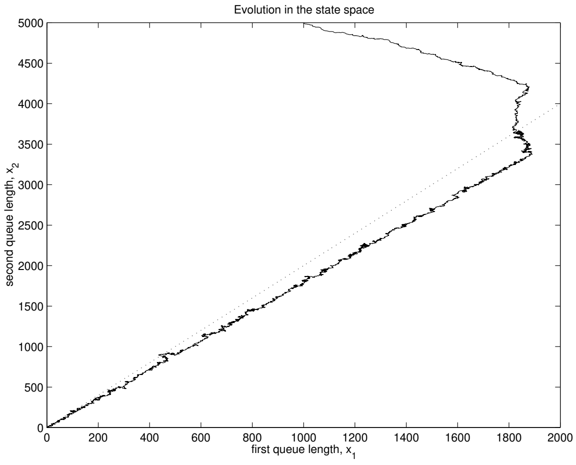

Figure 4.1 Evolution in the state space for the first system

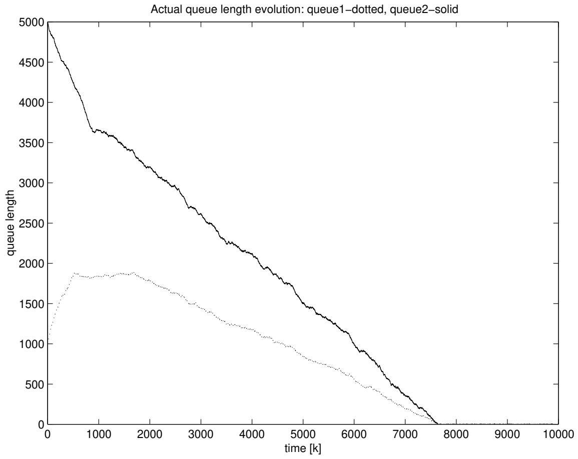

Figure 4.2 Evolution of the actual queue lengths for the first system

taken to be, $x_{10} = 1000$ and $x_{20} = 5000$. The system is observed over 10000 time slots. Then the resulting state space and separate queue length evolutions are depicted in Figures 4.1 and 4.2, respectively.

Obviously, the policy pulls both of the queue lengths toward zero and makes them stay inside a closed region around the origin. Hence, the system is stable. Moreover, from Figure 4.1, it can be seen that the queue lengths evolve toward the dotted line characterized by $\frac{x_1}{\lambda_1} = \frac{x_2}{\lambda_2}$ and evolves toward the origin afterwards. From Little's law, equalizing the fraction $\frac{x_i}{\lambda_i} \quad \forall i$ is equivalent to forcing the waiting times to be the same in each of the queues. Therefore, the policy is also useful in minimizing the maximum average delay over all queues. Graphically, the average delay experienced by the packets that has left up to some time k is plotted in Figure 4.3. It can be seen that, the average delays become quite close in a small amount of time. Also, the squared difference between the average delays are provided in Figure 4.4. It can be seen that the system is controlled such that the difference between the average delays is kept small throughout the operation.

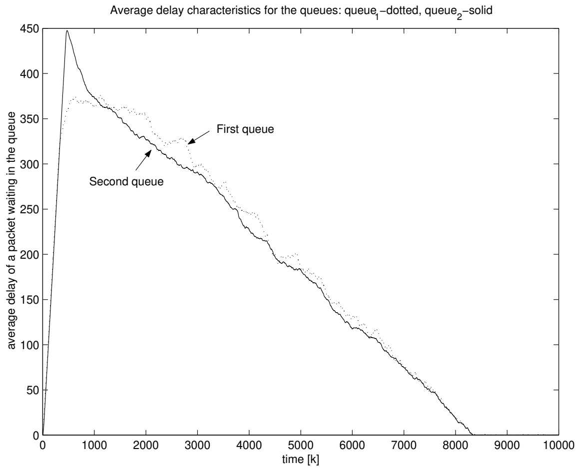

Figure 4.3 Average delay experienced by the packets for the first system

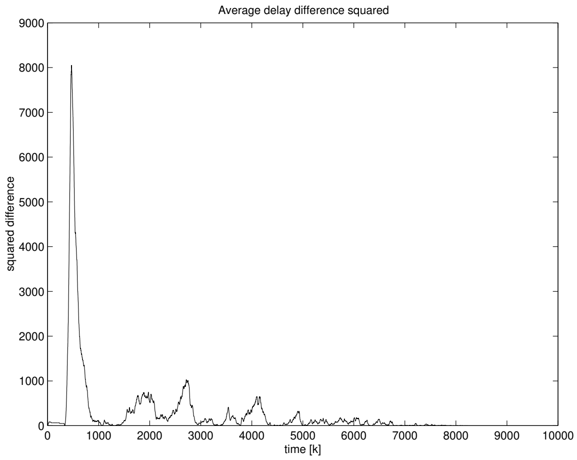

Figure 4.4 Squared difference between the average delays for the first system

Now that the fair average delay property of the policy is established via simulations, the next step is to study the effect of arrival rate locations inside the capacity region. Intuitively, we hope to observe that the queues drain faster if the rates are farther away from the capacity region boundary. For this purpose, the arrival rates in the previous simulation is reduced to $\lambda_1 = 4$ and $\lambda_2 = 7$, other parameters kept the same. Then the resulting state space and actual queue length evolution as well as average delay and squared difference between the delays are depicted in Figures 4.5, 4.6, 4.7 and 4.8, respectively.

Comparing the two simulations, one can observe that the farther away the mean arrival rates are from the capacity boundary, the faster the convergence to the vicinity of zero, as expected. In the simulations, none of the flows is favored against the others, whereas through

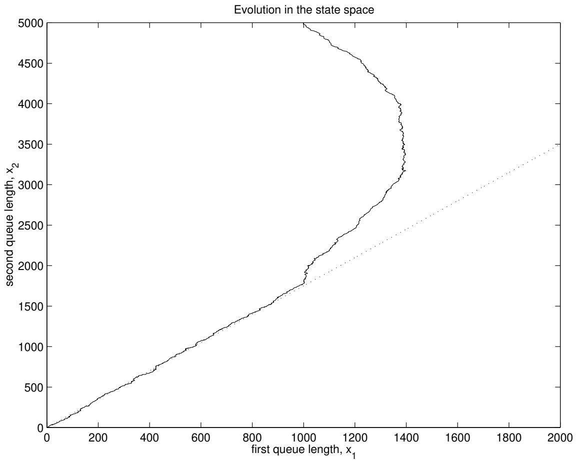

Figure 4.5 Evolution in the state space for the second system

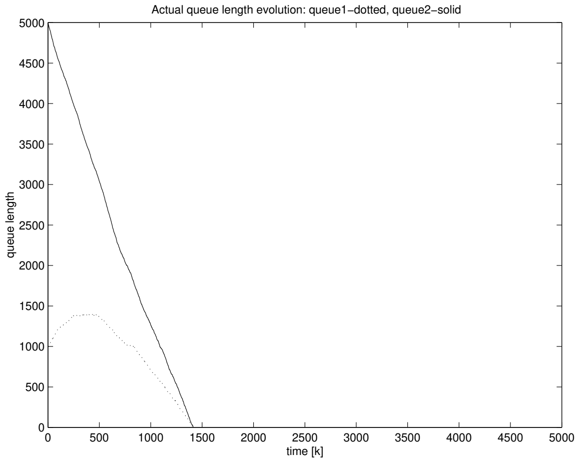

Figure 4.6 Evolution of the actual queue lengths for the second system

the use of weighting described in Section 2.3, priority may be given to any user in terms of delay experienced, while the system is still kept stable.

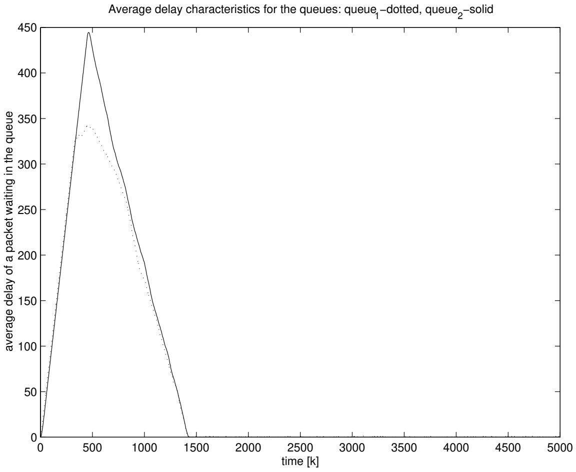

Figure 4.7 Average delay experienced by the packets for the second system

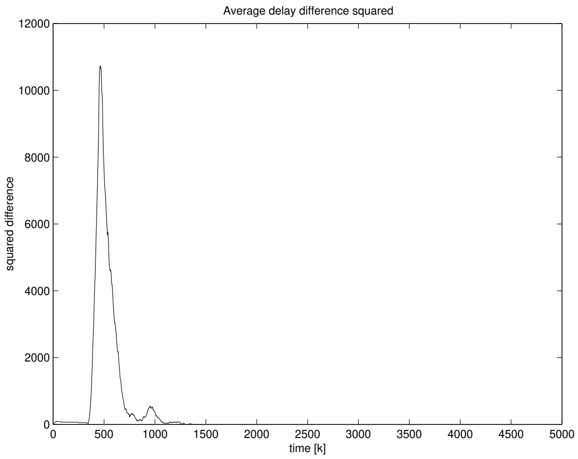

Figure 4.8 Squared difference between the average delays for the second system

CHAPTER 5

CONCLUSIONS

We have presented throughput optimal scheduling policies for a degraded Gaussian broadcast channel, which allocate power to the various receivers based on the current queue length information. We have assumed that the available transmission rates are given by the information-theoretic capacity region. To make this applicable to practical coding and modulation schemes, one has to use an error exponent argument as in Telatar and Gallager[3] or Berry[15]. Further, since power is allocated based on queue lengths, we are implicitly assuming that the codebook is changed during every time slot. Thus, the transmitter has to communicate the queue lengths to the receivers at the beginning of each time slot. We do not believe that this is a significant burden since we simply have to communicate the changes in the queue lengths at the beginning of each time slot. However, this requires further study. Finally, through computer simulations we were able to show the fairness property of these throughput optimal policies in terms of delays experienced by the packets waiting in the queues. Also by assigning appropriate weights to the queues, priorities may be given to the desired users. Author line of future work would be to extend the results to fading memoryless environments with mobile users.

REFERENCES

[1] T. Cover and J. A. Thomas, *Elements of Information Theory.* New York: John Wiley & Sons, 1991.

[2] L. Tassiulas and A. Ephremides, "Dynamic server allocation to parallel queues with randomly varying connectivity," *IEEE Transactions on Information Theory*, vol. 39, pp. 466–478, March 1993.

[3] I. E. Telatar and R. G. Gallager, "Combining queueing theory with information theory for multiaccess," *IEEE Journal on selected areas of Communications*, vol. 13, pp. 963–969, August 1995.

[4] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels i. ergodic capacity," *IEEE Transactions on Information Theory*, vol. 47, pp. 1089–1102, March 2001.

[5] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels ii. outage capacity," *IEEE Transactions on Information Theory*, vol. 47, pp. 1103–1127, March 2001.

[6] D. N. Tse, "Optimal power allocation over parallel gaussian broadcast channels," *Proceedings of Information Theory*, p. 27, 1997.

[7] S. Shakkottai and A. Stolyar, "Scheduling for multiple flows sharing a time-varying channel: The exponential rule," presented in the INFORMS Applied Probability Conference, July 2001.

[8] V. A. Mikhailov, "Geometrical analysis of the stability of markov chains in r_+^n and its application to throughput evaluation of the adaptive random multiple access algorithm," *Problemy Peredachi Informatsii*, vol. 24, pp. 61–73, January-March 1988. in Russian.

[9] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredachi Informatsii*, vol. 10, pp. 3–14, July-September 1974. in Russian.

[10] J. G. Dai, "On the positive harris recurrence for multiclass queueing networks: A unified approach via fluid limit models," *Annals of Applied Probability*, vol. 5, pp. 49–77, 1995.

[11] J. G. Dai and P. Meyn, "Convergence and moments of networks and their fluid models," *IEEE Transactions on Automatic Control*, vol. 40, pp. 1889–1904, 1995.

[12] B. Hajek, "Large bursts do not cause instability," *IEEE Transactions on Automatic Control*, vol. 5, pp. 116–118, January 2000.

[13] D. Gamarnik, "Using fluid models to prove stability of adversarial queueing networks," *IEEE Transactions on Automatic Control*, vol. 45, pp. 741–746, April 2000.

[14] R. G. Gallager, *Discrete Stochastic Processes.* Kluwer Academic Publishers, 1996.

[15] R. Berry, "Power-delay trade-offs in fading channels," Ph.D. dissertation, M.I.T., May 2000.