Throughput-optimal scheduling for broadcast channels

Atilla Eryilmaz [a], R. Srikant [b] and James R. Perkins [c]

[a]Coordinated Science Laboratory and Department of Electrical and Computer Engineering, University of Illinois

[b]Coordinated Science Laboratory and Department of General Engineering University of Illinois

[c]Department of Manufacturing Engineering Boston University

ABSTRACT

In this paper, we consider a degraded Gaussian broadcast channel, where the transmitter maintains separate queues for each receiver. We present *throughput optimal* policies that stabilize the queues without knowing the statistics of the arrival processes to these queues.

Keywords: Throughput optimal scheduling, Gaussian broadcast channel, Fluid models, Stability

1. MODEL & INTRODUCTION

Consider a link that is sending data to N users over a 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. 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 $\sigma_1^2 < \sigma_2^2 < ...\sigma_N^2$. The link uses a scheduling policy and allocates 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.

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

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

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.

In this paper, the only assumption regarding the arrival processes is that they are ergodic Markov chains with countable state spaces, whose mean arrival rates lie within the broadcast capacity region. In Section 2, a scheduling policy that is *throughput optimal* will be given. A policy is *throughput optimal*[2] if it can stabilize any system that is stabilizable under some scheduling policy. The stability condition in this context will be provided in Section 1.1.

To analyze the system described in this section, we will make use of a fluid model described next.

Further author information:
A.E.: E-mail: eryilmaz@uiuc.edu
R.S.: E-mail: rsrikant@uiuc.edu, Ph: 217-333-2457, fax:217-244-1642
J.P.: E-mail: perkins@engc.bu.edu

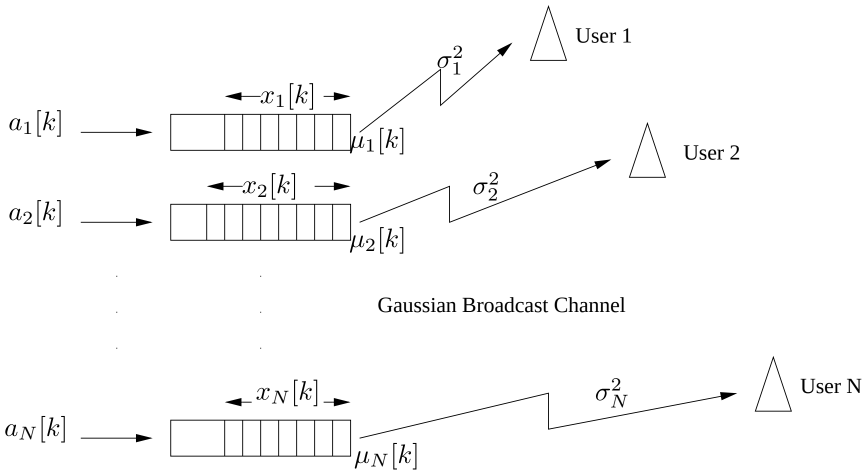

Figure 1. The Original System Model

1.1. FLUID MODEL

Fluid models of queueing networks are among the simplest models to analyze, owing to the fact that deterministic calculus can be applied. Dai[3] and Dai and Meyn[4] showed that stability of a fluid model implies the stability (in an appropriate sense) of related multiclass networks with random service and interarrival process of renewal type. Hajek[5] and Gamarnik[6] have also shown that the stability of networks with fluid traffic implies stability of networks with deterministically constrained traffic.

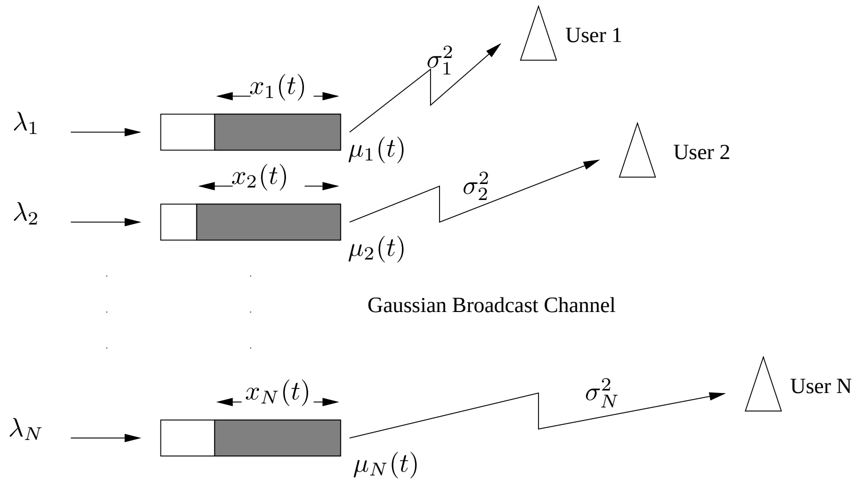

Figure 2. The Fluid Model

In our context, the fluid model for N users is given in Fig. 2. In this model, the arrival process is taken to be deterministic with a constant arrival rate of λ_i for the i^{th} flow. In the original problem with stochastic arrival processes, λ_i would be the mean arrival rate to queue i. The 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 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

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

In Section 2, we present a simple scheduling policy, that has no knowledge of the arrival statistics, which empties all of the buffers of the fluid model, in some *finite* time, provided that the *stability condition* is satisfied. This result can, in turn, be used to show that the original system is stable (i.e., it has a stationary distribution) using the techniques in the papers cited earlier,[3,4] 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. A SIMPLE THROUGHPUT OPTIMAL SCHEDULING POLICY

In this section, we present a simple scheduling policy, which makes the fluid model 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 (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 length of each of the buffers along with the capacity region.

2.1. GEOMETRICAL INTERPRETATION FOR THE TWO-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:

$$\begin{aligned} \dot{x}_1(t) &= \lambda_1 - \mu_1(t) \\ \dot{x}_2(t) &= \lambda_2 - \mu_2(t), \end{aligned} \tag{2}$$

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 Fig. 3.

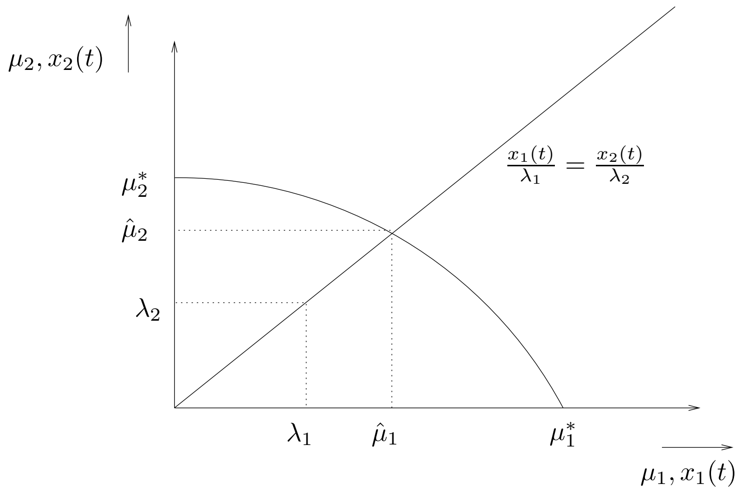

Figure 3. The capacity region

Along with the capacity region, the line passing through (λ_1, λ_2) is also plotted in Fig. 3. 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 Section 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.*

This simply says that, if the queue lengths ever hit the line given by $\frac{x_1(t)}{\lambda_1} = \frac{x_2(t)}{\lambda_2}$, which is shown in Fig. 3, then both of the buffer lengths will eventually be zero in a *finite* amount of time. From the figure, this is obvious due to the fact that $\hat{\mu}_1 > \lambda_1$ and $\hat{\mu}_2 > \lambda_2$.

FACT 2. *Without loss of generality, assume* $\frac{x_{10}}{\lambda_1} > \frac{x_{20}}{\lambda_2}$. *Then* $\frac{x_1(t)}{x_2(t)}$ *will always be decreasing and furthermore,* $\mu_1(t) > \lambda_1 + \delta$ *and* $\mu_2(t) > \lambda_2 + \delta$ *for some* $\delta > 0$ *after a finite duration.*

From Fig. 4, for any initial state that is below line 1, the policy will move it above line 1 in some finite time and from that point on the service rates, $\mu_1(t)$ and $\mu_2(t)$, will always be greater than the arrival rates, λ_1 and λ_2 by at least a constant, positive amount δ. Consequently, the queues will both be emptied in a *finite* amount of time.

This result can easily be demonstrated by a computer simulation. One such example is illustrated in Fig. 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. N-USER CASE

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 it is decreasing. 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 (1), we define the Lyapunov function $V(x_i(t), x_j(t))$ as follows:

$$V(x_i(t), x_j(t)) = \frac{1}{2}\left(\frac{x_i(t)}{x_j(t)} - \frac{\lambda_i}{\lambda_j}\right)^2 \qquad i, j \in \{1, 2, ..., N\} \qquad i \neq j. \tag{3}$$

Taking the derivative of (3) with respect to t:

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

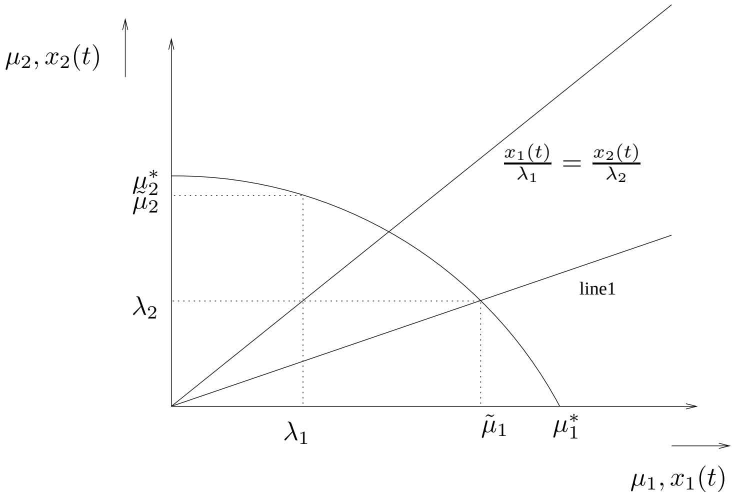

Figure 4. Pictorial interpretation to Fact 2

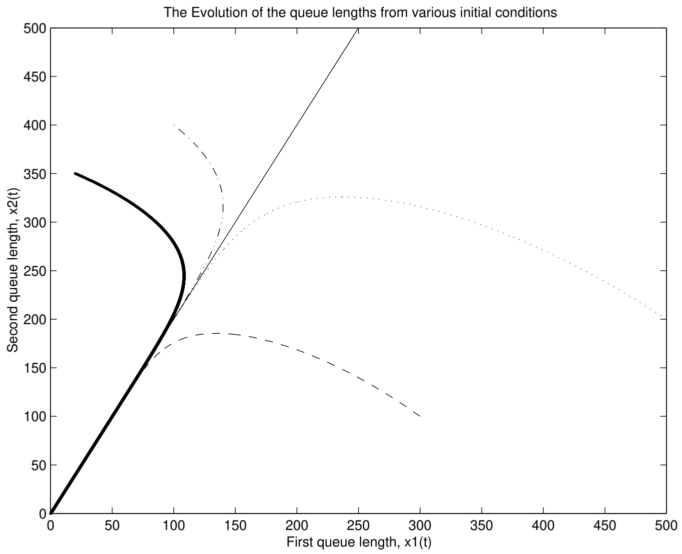

Figure 5. Evolution of the queue lengths from different initial starting points

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, it is easy to note that for each of the flows, $\mu_i = K(x_1, x_2, ..., x_N)x_i$ 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

$$\dot{V}(x_i, x_j) = \frac{-1}{x_j^3\lambda_j}(x_i\lambda_j - x_j\lambda_i)^2 \leq 0 \tag{4}$$

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 (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, the Lyapunov argument shows that,

$$\lim_{t\to\infty}\left(\frac{x_i(t)}{x_j(t)} - \frac{\lambda_i}{\lambda_j}\right)^2 = 0$$

Therefore, for $\epsilon > 0$, $\exists T_{(i,j)}(\epsilon) < \infty$ such that $\forall t > T_{(i,j)}(\epsilon)$, we have

$$-\epsilon < \left(\frac{x_i(t)}{x_j(t)} - \frac{\lambda_i}{\lambda_j}\right) < \epsilon.$$

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, and defining T as

$$T := \max_{i,j} T_{(i,j)}(\epsilon) < \infty \qquad for \qquad i, j \in \{1, 2, ..., N\} \qquad i \neq j,$$

it is possible to obtain service rates $\mu_i(t) > \lambda_i + \Delta$ for all $t > T$ 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.

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

$$\frac{\mu_1(t)}{K_1x_1(t)} = \frac{\mu_2(t)}{K_2x_2(t)} = ... = \frac{\mu_N(t)}{K_Nx_N(t)}$$

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 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, 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

$$c = \frac{\dot{x}_1(t)}{\dot{x}_2(t)} = \frac{\hat{\mu}_1 - \lambda_1}{\hat{\mu}_2 - \lambda_2} \qquad \textit{and} \qquad \frac{\hat{\mu}_1}{\hat{\mu}_2} = Kc. \tag{5}$$

Fig. 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 $\frac{x_1(t)}{x_2(t)} = c$.

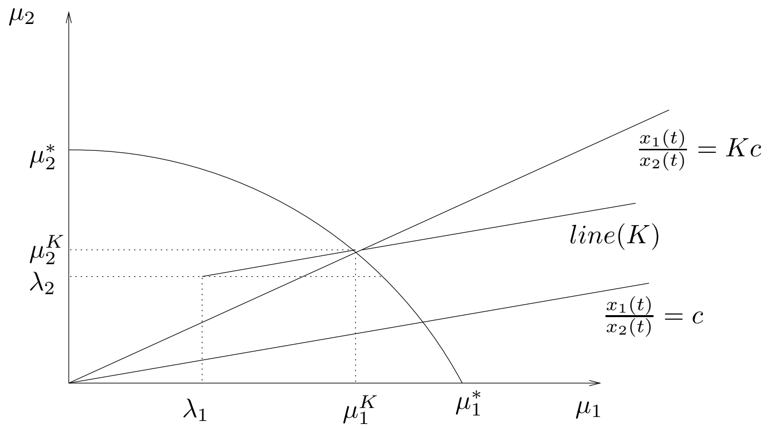

Figure 6. Pictorial illustration of the weighted throughput optimal policy

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:

$$V(x_1(t), x_2(t)) = \frac{1}{2}\left(\frac{x_1(t)}{x_2(t)} - c\right)^2. \tag{6}$$

Then,

$$\dot{V}(x_1, x_2) = \frac{1}{x_2^3}(x_1 - cx_2)((\lambda_1 - \mu_1)x_2 - (\lambda_2 - \mu_2)x_1). \tag{7}$$

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 (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 Fig. 6. Using the property given by (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 (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 (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 Fig. 7.

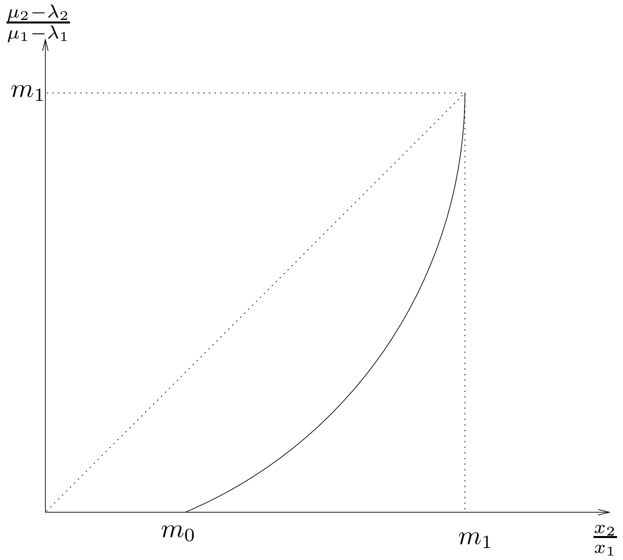

Figure 7. Relationship between the slopes

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 (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 Section 2.2.

4. CONCLUSIONS

We have presented throughput optimal scheduling policies for a 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, we have to use an error exponent analysis as in Telatar and Gallager[7] or Berry.[8] 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.

ACKNOWLEDGMENTS

This work was supported by NSF Grant ANI-9714685.

REFERENCES

1. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, 1991.
2. S. Shakkottai and A. Stolyar, "Scheduling for multiple flows sharing a time-varying channel: The exponential rule," *to appear in the INFORMS Applied Probability Conference, New York* , July 2001.
3. J. G. Dai, "On the positive harris recurrence for multiclass queueing networks: A unified approach via fluid limit models," *Annals of Applied Probability* **5**, pp. 49–77, 1995.
4. J. G. Dai and P. Meyn, "Convergence and moments of networks and their fluid models," *IEEE Transactions on Automatic Control* **40**, pp. 1889–1904, 1995.
5. B. Hajek, "Large bursts do not cause instability," *IEEE Transactions on Automatic Control* **5**, pp. 116–118, Jan. 2000.
6. D. Gamernik, "Using fluid models to prove stability of adversarial queueing networks," *IEEE Transactions on Automatic Control* **45**, pp. 741–746, April 2000.
7. I. Telatar and R. Gallager, "Combining queueing theory with information theory for multiaccess," *IEEE Journal on Selected Areas in Communications* **13**, pp. 963–969, Aug. 1995.
8. R. Berry, "Power-delay trade-offs in fading channels," *Ph.D. Thesis, M.I.T.* , June 2000.