Delay Minimization in Multiple Access Channels

Jing Yang Sennur Ulukus
Department of Electrical and Computer Engineering
University of Maryland, College Park, MD 20742
yangjing@umd.edu *ulukus@umd.edu*

***Abstract*—We investigate a delay minimization problem in a multiple access wireless communication system. We consider a discrete-time non-fading additive white Gaussian noise (AWGN) multiple access channel. In each slot, bits arrive at the transmitters randomly according to some distribution, which is i.i.d. from user to user and from slot to slot. Each transmitter has an average power constraint of P. Our goal is to allocate rates to users, from the multiple access capacity region, based on their current queue lengths, in order to minimize the average delay of the system. We formulate the problem as a Markov decision problem (MDP) with an average cost criterion. We first show that the value function is increasing, symmetric and convex in the queue length vector. Taking advantage of these properties, we show that the optimal rate allocation policy is one which tries to equalize the queue lengths as much as possible in each slot, while working on the dominant face of the capacity region.**

I. Introduction

Traditional information theory investigates transmission problems from a physical layer perspective. In the simplified source-channel-destination model, information-theoretic approaches assume the availability of an infinite number of bits at the transmitter before the transmission starts. The burstiness of the arrivals and the associated issue of delay are mostly ignored. In contrast, network theory gives sophisticated analysis of network layer issues, such as random arrivals and network delay. However, in network-theoretic approaches, the underlying physical layer model is usually very simplified, e.g., in most approaches simultaneous transmissions are not allowed, and even when they are allowed, a collision channel model is used, which is too simplistic to capture what can be achieved in the physical layer from an information-theoretic perspective.

In recent years, many authors have taken efforts to bridge the gap between information theory and network theory [1]. Reference [2] addresses the delay issue for an additive Gaussian noise multiple access channel. Packets with random sizes arrive according to a Poisson process, and are transmitted out immediately with a fixed power. At the physical layer, the receiver decodes a packet while treating other transmissions as noise. Consequently, the service rate becomes a function of the number of active users in the system. Reference [2] derives the relationship between the average delay and a fixed probability of error requirement. References [3], [4] and [5] consider a discrete-time model for a power-constrained single-user communication channel. Random arrivals queue at the transmitter to wait to be transmitted. In each slot, the transmitter adapts its service rate, i.e., transmission rate, according to the queue length and the channel state, as well as the average power constraint, to minimize the average delay. Reference [3] formulates the problem as a dynamic programming problem and develops a delay-power tradeoff curve. References [4] and [5] determine some structural properties of the optimal power/rate allocation policy.

Reference [6] uses a continuous-time queueing model to model the network layer behavior of a multiple access system. The packets arrive at the transmitters according to independent Poisson processes, and the packet lengths are exponentially distributed. The physical layer is modelled as an additive Gaussian noise channel, whose capacity region is a pentagon for the two-user case. The goal of [6] is to select an operating rate point inside the multiple access capacity region, as a function of the current queue lengths, in order to minimize the average packet delay. The transmission rates selected from the capacity region serve as the current service rates of the queues. Reference [6] develops the longer-queue-higher-rate (LQHR) allocation strategy, which selects an extreme point in the capacity region of the multiple access channel (i.e., a corner point of the pentagon). Reference [6] shows that LQHR minimizes the average delay of a symmetric system. Reference [7] extends [6] to a potentially asymmetric setting, and proves that the delay-optimal policy has a threshold (switch) structure. Reference [8] develops a policy named "modified LQHR" which works at a corner point of the pentagon when the queue lengths are different, and switches to the mid-point of the dominant face of the pentagon when the queue lengths become equal. The "modified LQHR" algorithm is shown to minimize the average bit delay in the system. The third chapter of [9] extends "modified LQHR" to an M-user scenario.

In this paper, we consider a similar delay minimization problem. In order to track the relationship between the average delay and the transmission rates more accurately and also to consider more general arrivals, we adopt a discrete-time queueing model and consider the problem from a bit perspective rather than a packet perspective. We partition the time into small slots. In each slot, bits arrive at the transmitters randomly according to some general distribution. At the beginning of each slot, we allocate transmission rates from within the multiple access capacity region to the users, based on their current queue lengths, to minimize the average delay. In our model, the number of bits transmitted in each slot is equal to the product of the transmission rate and

This work was supported by NSF Grants CCF 04-47613, CCF 05-14846, CNS 07-16311 and CCF 07-29127.

the number of channel uses in each slot. We formulate the problem as an average cost Markov decision problem (MDP). We first analyze the corresponding discounted cost MDP, and obtain some properties of the value function. Based on these properties, we prove that the delay optimal rate allocation policy for this discounted MDP is to equalize the queue lengths in each slot as much as possible. We then prove that this "queue balancing policy" is optimal for the average cost MDP as well.

Essentially, both the "modified LQHR" and our policy aim to balance the queue lengths as well as to maximize the throughput at any time. However, the continuous model in [8], [9] allows the rates to be changed at any time, while our model allows us to make decisions only at the beginning of each slot. Consequently, the resulting optimal policies are different: The operating point of the "modified LQHR" algorithm is either one of the corner points or the mid-point of the dominant face of the pentagon, while the "queue balancing policy" here may operate at any point on the dominant face of the pentagon.

II. System Model and Problem Formulation

A. Physical Layer Model

We consider a two-user AWGN multiple access system

$$Y = X_1 + X_2 + Z \tag{1}$$

where X_i is the signal of user i, and Z is a Gaussian noise with zero-mean and variance σ^2. In this multiple access system, the capacity region is given by [10]

$$R_1 \leq \frac{1}{2}\log\left(1+\frac{P_1}{\sigma^2}\right) \triangleq C_1 \tag{2}$$

$$R_2 \leq \frac{1}{2}\log\left(1+\frac{P_2}{\sigma^2}\right) \triangleq C_2 \tag{3}$$

$$R_1 + R_2 \leq \frac{1}{2}\log\left(1+\frac{P_1+P_2}{\sigma^2}\right) \triangleq C_s \tag{4}$$

The capacity region is a pentagon, as shown in Fig. 1. In this paper we consider a symmetric two-user system, where $P_1 = P_2 = P$. Our results can be generalized to the symmetric K-user case.

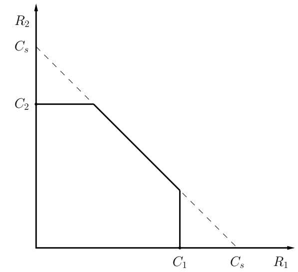

Fig. 1. The capacity region for a two-user multiple access system.

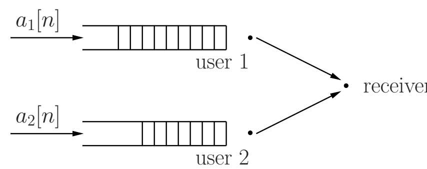

Fig. 2. System model.

B. Medium Access Control (MAC) Layer Model

In the MAC layer, we assume that the bits arrive at the transmitters in random numbers in each slot, see Fig. 2. Let $a_1[n]$ and $a_2[n]$ denote the number of bits arriving at the first and the second transmitter, respectively, during time slot n. Here, $a_1[n]$ and $a_2[n]$ are two independent random variables with a common distribution F_a. We assume that the arrivals are i.i.d. in n.

There is an infinite capacity buffer at each transmitter to store the bits. Let $q_1[n]$ and $q_2[n]$ denote the number of bits in the first and the second buffer, respectively, at the beginning of the nth slot. At the beginning of each slot, the transmitters decide on how many bits to transmit in this slot based on the current lengths of the two queues. Let $d_1[n]$ and $d_2[n]$ denote the number of bits to be transmitted from the first and the second queue, respectively, in the nth time slot. Let us define $\mathbf{q}[n] \triangleq (q_1[n], q_2[n])$, $\mathbf{d}[n] \triangleq (d_1[n], d_2[n])$, and $\mathbf{a}[n] \triangleq (a_1[n], a_2[n])$. Then, the queue lengths evolve according to

$$\mathbf{q}[n+1] = (\mathbf{q}[n] - \mathbf{d}[n])^+ + \mathbf{a}[n] \tag{5}$$

where $(x)^+$ denotes $\max(0, x)$.

If the number of channel uses in a slot is τ, the transmission rate of user i becomes $R_i[n] = d_i[n]/\tau$. Consequently, the actual rates of the users that need to be selected from the capacity region described by (2)-(4), are proportional to $d_1[n]$ and $d_2[n]$, and therefore, $(d_1[n], d_2[n])$ can be viewed as (scaled) rates. In order to simplify the notation, we will call $d_i[n] = R_i[n]\tau$ as the *rate* of user i for slot n. The corresponding scaled capacity region that (d_1, d_2) should reside in is described by (2)-(4) by multiplying right hand sides by τ.

C. Formulation as an MDP

According to Little's law [11], minimizing the average delay in the system is equivalent to minimizing the average number of bits in the system, which is the average sum of queue lengths. If the system starts from state $\mathbf{q}[1]$, the delay minimization problem is to obtain optimal policy $\mathbf{d}[n]$, $n = 1, 2, \ldots$ to minimize

$$\limsup_{N\to\infty} \frac{1}{N} E\left[\sum_{n=1}^{N}(q_1[n] + q_2[n])\right] \tag{6}$$

Therefore, this problem can be formulated as a standard average cost MDP. The state space consists of all possible queue length vectors, while the policy space is the set of operating points within the multiple access capacity region. In principle, the values of $q_i[n]$, $d_i[n]$ can only be integers, however, for practical applications, one bit is a fine enough precision that we can use a fluid model to reasonably approximate the original discrete-state system.

III. The Discounted Cost Problem

Instead of considering the minimization problem with the average cost criterion in (6) directly, we first consider the following minimization problem with a total discounted cost criterion

$$E\left[\sum_{n=1}^{\infty}\beta^n(q_1[n]+q_2[n])\right] \tag{7}$$

where $0<\beta<1$ is the discount factor. We will return to the average cost criterion in (6) by letting β go to 1.

Let us define $V^\beta(\mathbf{q})$ to be the total discounted cost starting from an initial state $\mathbf{q}$. Then, for the optimization problem with criterion (7), $V^\beta(\mathbf{q})$ must satisfy the following optimality condition [12]

$$V^\beta(\mathbf{q})=\min_{\mathbf{d}\in C}\left\{q_1+q_2+\beta E\left[V^\beta((\mathbf{q}-\mathbf{d})^++\mathbf{a})\right]\right\} \tag{8}$$

We will first start with a discounted cost problem over finite horizon N. For this problem with an initial state $\mathbf{q}$, the dynamic programming formulation is

$$V_N^\beta(\mathbf{q})=\min_{\mathbf{d}\in C}\left\{q_1+q_2+\beta E\left[V_{N-1}^\beta((\mathbf{q}-\mathbf{d})^++\mathbf{a})\right]\right\} \tag{9}$$

with $V_0^\beta(\cdot)=0$. Since the instantaneous cost $q_1[n]+q_2[n]$ is positive, and the policy space is finite [12]

$$V_N^\beta(\mathbf{q})\to V^\beta(\mathbf{q}) \quad \text{as} \quad N\to\infty \tag{10}$$

where $V^\beta(\cdot)$ is the unique bounded solution of (8).

In the following, we will analyze the discounted cost problem and obtain structural properties of the value function $V^\beta(\mathbf{q})$. We will find these structural properties of $V^\beta(\mathbf{q})$ by examining the structural properties of the finite-horizon discounted cost problem $V_N^\beta(\mathbf{q})$.

Lemma 1 *$V^\beta(\mathbf{q})$ is increasing in q_1 and q_2.*

Proof: From (10), we know that proving $V^\beta(\mathbf{q})$ is increasing in q_1 and q_2 is equivalent to proving $V_N^\beta(\mathbf{q})$ is increasing in q_1 and q_2 for every N. We prove this through induction. First, when $N=0,1$, this is trivially true. Next, we assume that it is true for $N-1$. We will prove that $V_N^\beta(q_1+1,q_2)>V_N^\beta(q_1,q_2)$ for any positive (q_1,q_2).

$$\begin{aligned}
&V_N^\beta(q_1+1,q_2)\\
&\quad=q_1+q_2+1+\beta E\Big[V_{N-1}^\beta((q_1+1-d_1^*)^++a_1,\\
&\qquad\qquad(q_2-d_2^*)^++a_2)\Big] &(11)\\
&\quad\geq q_1+q_2+1+\beta E\Big[V_{N-1}^\beta((q_1-d_1^*)^++a_1,\\
&\qquad\qquad(q_2-d_2^*)^++a_2)\Big] &(12)\\
&\quad>\min_{\mathbf{d}\in C}\Big\{(q_1+q_2)+\beta E\Big[V_{N-1}^\beta\big((q_1-d_1)^++a_1,\\
&\qquad\qquad(q_2-d_2)^++a_2\big)\Big]\Big\} &(13)\\
&\quad=V_N^\beta(q_1,q_2) &(14)
\end{aligned}$$

where (d_1^*,d_2^*) in (11) is the point within the capacity region that minimizes $V_N^\beta(q_1+1,q_2)$, and (12) follows from the assumption that $V_{N-1}^\beta(q_1,q_2)$ is increasing for every q_1. Therefore, $V_N^\beta(\mathbf{q})$ is increasing in q_1 for every N. Using (10), this implies that $V^\beta(\mathbf{q})$ is increasing in q_1. Now, following the same procedure for q_2, we can prove that $V^\beta(\mathbf{q})$ is increasing in q_2 as well. ■

Lemma 2 *In (8), the optimal operating point $\mathbf{d}$ must be on the boundary of the capacity region C.*

Proof: For an initial state $\mathbf{q}$, if the optimal operating point $\mathbf{d}=(d_1,d_2)$ is not on the boundary of the capacity region but on the interior of the capacity region, then, we can always find points $\bar{\mathbf{d}}=(d_1',d_2)$, $\tilde{\mathbf{d}}=(d_1,d_2')$ that are on the boundary of the capacity region with $d_1'>d_1$, $d_2'>d_2$. Note that $\bar{\mathbf{d}}\geq\mathbf{d}$ and $\tilde{\mathbf{d}}\geq\mathbf{d}$. Then, by Lemma 1, we have

$$E\left[V^\beta((\mathbf{q}-\bar{\mathbf{d}})^++\mathbf{a})\right]\leq E\left[V^\beta((\mathbf{q}-\mathbf{d})^++\mathbf{a})\right] \tag{15}$$

and

$$E\left[V^\beta((\mathbf{q}-\tilde{\mathbf{d}})^++\mathbf{a})\right]\leq E\left[V^\beta((\mathbf{q}-\mathbf{d})^++\mathbf{a})\right] \tag{16}$$

This contradicts the optimality of $\mathbf{d}$. Thus, $\mathbf{d}$ must be on the boundary of the capacity region. ■

Lemma 3 *$V^\beta(\mathbf{q})$ is symmetric and jointly convex in $\mathbf{q}$.*

Proof: The symmetry property can be proved by induction. Note that $V_N^\beta(\mathbf{q})$ is symmetric for $N=0,1$. Assuming that $V_{N-1}^\beta(\mathbf{q})$ is symmetric, it is easy to see that $V_N^\beta(\mathbf{q})$ would be symmetric. Now, taking the limit $N\to\infty$, it follows that $V^\beta(\mathbf{q})$ is symmetric.

We prove the convexity of $V^\beta(\mathbf{q})$ through induction as well. When $N=0,1$, it is trivial to see that $V_N^\beta(\mathbf{q})$ is convex in $\mathbf{q}$. Next, we assume that $V_{N-1}^\beta(\mathbf{q})$ is convex in $\mathbf{q}$. Given two different queue length vectors $\mathbf{x}\triangleq(x_1,x_2)$ and $\mathbf{y}\triangleq(y_1,y_2)$, we have

$$\begin{aligned}
&\lambda V_N^\beta(\mathbf{x})+(1-\lambda)V_N^\beta(\mathbf{y})\\
&=\lambda(x_1+x_2)+(1-\lambda)(y_1+y_2)\\
&\quad+\lambda\beta E\Big[V_{N-1}^\beta((\mathbf{x}-\mathbf{b}^*)^++\mathbf{a})\Big]\\
&\quad+(1-\lambda)\beta E\Big[V_{N-1}^\beta((\mathbf{y}-\mathbf{d}^*)^++\mathbf{a})\Big] &(17)\\
&\geq\lambda(x_1+x_2)+(1-\lambda)(y_1+y_2)+\\
&\quad\beta E\Big[V_{N-1}^\beta\big(\lambda(\mathbf{x}-\mathbf{b}^*)^++(1-\lambda)(\mathbf{y}-\mathbf{d}^*)^++\mathbf{a}\big)\Big] &(18)\\
&\geq\lambda(x_1+x_2)+(1-\lambda)(y_1+y_2)+\\
&\quad\beta E\Big[V_{N-1}^\beta\big(\big(\lambda(\mathbf{x}-\mathbf{b}^*)+(1-\lambda)(\mathbf{y}-\mathbf{d}^*)\big)^++\mathbf{a}\big)\Big] &(19)\\
&\geq\min_{\mathbf{d}\in C}\Big\{\lambda(x_1+x_2)+(1-\lambda)(y_1+y_2)+\\
&\quad\beta E\Big[V_{N-1}^\beta\big(\big(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}-\mathbf{d}\big)^++\mathbf{a}\big)\Big]\Big\} &(20)\\
&=V_N^\beta(\lambda\mathbf{x}+(1-\lambda)\mathbf{y}) &(21)
\end{aligned}$$

where $\mathbf{b}^*$ and $\mathbf{d}^*$ are the minimizers for $V_N^\beta(\mathbf{x})$ and $V_N^\beta(\mathbf{y})$, respectively. Here, (18) follows from the assumption of the convexity of $V_{N-1}^\beta(\cdot)$, (19) follows from the convexity of the function $(\cdot)^+$, and (20) is valid because $\mathbf{b}^*, \mathbf{d}^* \in C$, and C is a convex set, implying $\lambda\mathbf{b}^* + (1-\lambda)\mathbf{d}^* \in C$. ■

Before we move on to the next structural property of the function $V^\beta(\mathbf{q})$, we need to introduce the concepts of majorization and Schur-convexity.

Definition 1 ([13]) *Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we say that $\mathbf{x}$ majorizes $\mathbf{y}$, and we write $\mathbf{x} \succeq \mathbf{y}$, if*

$$\sum_{i=1}^{k} x_i \geq \sum_{i=1}^{k} y_i, \quad k \in \{1, \ldots, d-1\} \tag{22}$$

$$\sum_{i=1}^{d} x_i = \sum_{i=1}^{d} y_i \tag{23}$$

where x_i and y_i are the ith largest elements of $\mathbf{x}$ and $\mathbf{y}$, respectively.

Definition 2 ([13]) *A function $f : \mathbb{R}^d \to \mathbb{R}$ is said to be Schur-convex when $\mathbf{x} \succeq \mathbf{y}$ implies $f(\mathbf{x}) \geq f(\mathbf{y})$.*

A function is Schur-convex if it is symmetric and convex [13]. Using Lemma 3, we conclude that $V^\beta(\mathbf{q})$ is Schur-convex. However, given that $\mathbf{x} \succeq \mathbf{y}$, we cannot directly claim that $V^\beta(\mathbf{x}+\mathbf{a}) \geq V^\beta(\mathbf{y}+\mathbf{a})$ for every $\mathbf{a}$. This is because the randomness of $\mathbf{a}$ may reverse the majorization relationship between $\mathbf{x}+\mathbf{a}$ and $\mathbf{y}+\mathbf{a}$. However, provided that $V^\beta(\mathbf{q})$ is symmetric and convex, and $\mathbf{a}$ has i.i.d. components, we can prove that $E[V^\beta(\mathbf{x}+\mathbf{a})] \geq E[V^\beta(\mathbf{y}+\mathbf{a})]$ if $\mathbf{x} \succeq \mathbf{y}$.

Lemma 4 *For i.i.d. a_is $\mathbf{x} \succeq \mathbf{y}$ implies $E[V^\beta(\mathbf{x}+\mathbf{a})] \geq E[V^\beta(\mathbf{y}+\mathbf{a})]$.*

Proof: When $a_1 = a_2$, clearly, $\mathbf{x}+\mathbf{a} \succeq \mathbf{y}+\mathbf{a}$, and $V^\beta(\mathbf{x}+\mathbf{a}) \geq V^\beta(\mathbf{y}+\mathbf{a})$. When $a_1 \neq a_2$, we evaluate the functions $V^\beta(\mathbf{x}+\mathbf{a})$ and $V^\beta(\mathbf{y}+\mathbf{a})$ at two symmetric points (c_1, c_2) and (c_2, c_1). In order to simply the notation, for any vector $\mathbf{v} = (v_1, v_2)$, we define $\check{\mathbf{v}} = (v_2, v_1)$. Because a_is are i.i.d., the two points $\mathbf{c}$, $\check{\mathbf{c}}$ have the same probability mass. Without loss of generality, we assume $c_1 > c_2$, $x_1 \geq x_2$, $y_1 \geq y_2$. Since $\mathbf{x} \succeq \mathbf{y}$, we have $x_1 \geq y_1 \geq y_2 \geq x_2$.

Consider four vectors $(\mathbf{x}+\mathbf{c})$, $(\check{\mathbf{x}}+\mathbf{c})$, $(\mathbf{y}+\mathbf{c})$, $(\check{\mathbf{y}}+\mathbf{c})$. We see that they are four points on the line $q_1 + q_2 = x_1 + x_2 + c_1 + c_2$. Moreover, since $x_1 \geq y_1 \geq y_2 \geq x_2$, $(\mathbf{x}+\mathbf{c})$ and $(\check{\mathbf{x}}+\mathbf{c})$ are the two outer points, and the mid-point of these two points is the same as the mid-point of the other two points. Since $V^\beta(\mathbf{q})$ is convex, we have

$$V^\beta(\mathbf{x}+\mathbf{c}) + V^\beta(\check{\mathbf{x}}+\mathbf{c}) \geq V^\beta(\mathbf{y}+\mathbf{c}) + V^\beta(\check{\mathbf{y}}+\mathbf{c}) \tag{24}$$

We also note that because of the symmetry property of $V^\beta(\mathbf{q})$ we have $V^\beta(\check{\mathbf{x}}+\mathbf{c}) = V^\beta(\mathbf{x}+\check{\mathbf{c}})$. Similarly, we have $V^\beta(\check{\mathbf{y}}+\mathbf{c}) = V^\beta(\mathbf{y}+\check{\mathbf{c}})$. Therefore, (24) is equivalent to

$$V^\beta(\mathbf{x}+\mathbf{c}) + V^\beta(\mathbf{x}+\check{\mathbf{c}}) \geq V^\beta(\mathbf{y}+\mathbf{c}) + V^\beta(\mathbf{y}+\check{\mathbf{c}}) \tag{25}$$

Integrating over a_1, a_2, we get

$$\begin{aligned}
&E[V^\beta(\mathbf{x}+\mathbf{a})] \\
&= \int_{a_1>a_2} V^\beta(\mathbf{x}+\mathbf{a}) + \int_{a_1<a_2} V^\beta(\mathbf{x}+\mathbf{a}) + \int_{a_1=a_2} V^\beta(\mathbf{x}+\mathbf{a}) \\
&= \int_{a_1<a_2} (V^\beta(\mathbf{x}+\mathbf{a}) + V^\beta(\mathbf{x}+\check{\mathbf{a}})) + \int_{a_1=a_2} V^\beta(\mathbf{x}+\mathbf{a}) \\
&\geq \int_{a_1<a_2} (V^\beta(\mathbf{y}+\mathbf{a}) + V^\beta(\mathbf{y}+\check{\mathbf{a}})) + \int_{a_1=a_2} V^\beta(\mathbf{y}+\mathbf{a}) \\
&= E\left[V^\beta(\mathbf{y}+\mathbf{a})\right]
\end{aligned}$$

where the inequality follows from (25). ■

We now combine Lemmas 1 through 4 to obtain the main result of this paper which is given in Theorem 1.

Theorem 1 *To minimize the average delay, in each slot, the transmitters should choose an operating point on the dominant face of the capacity region that equalizes the queue lengths. If no such operating point exists, the transmitters should operate at a corner point which minimizes the queue length difference.*

Proof: We know from Lemma 2 that, in each slot, the transmitters must operate on the dominant face (sum-rate constrained face) of the multiple access capacity region.

First, we prove that if there exists a point on the dominant face that equalizes the queue lengths, then this point must be the optimal operating point. Given queue lengths $\mathbf{q} = (q_1, q_2)$, let $\mathbf{d} = (d_1, d_2)$ be such a point, i.e., $(q_1-d_1)^+ = (q_2-d_2)^+$. If $(q_1-d_1)^+ = (q_2-d_2)^+ = 0$, then, clearly, $\mathbf{d}$ is the optimal operating point. We consider the case when $q_1-d_1 = q_2-d_2 > 0$. To prove the claim by contradiction, let us assume that $\mathbf{d}$ is not optimal, but $\mathbf{b} = (b_1, b_2)$ is the optimal point on the dominant face. Since both $\mathbf{d}$ and $\mathbf{b}$ are on the dominant face of the capacity region: $d_1+d_2 = b_1+b_2$. Since with a fixed sum, the vector with identical components is majorized by any other vector [13], we have $(q_1-b_1, q_2-b_2) \succeq (q_1-d_1, q_2-d_2)$. Without loss of generality, we assume $q_1-b_1 > q_2-b_2$, i.e., $q_1-b_1 > q_1-d_1 = q_2-d_2 > q_2-b_2$. If $q_2-b_2 \geq 0$, we have $((q_1-b_1)^+, (q_2-b_2)^+) \succeq ((q_1-d_1)^+, (q_2-d_2)^+)$, and using Lemma 4, this implies

$$E[V^\beta((\mathbf{q}-\mathbf{b})^+ + \mathbf{a})] \geq E[V^\beta((\mathbf{q}-\mathbf{d})^+ + \mathbf{a})] \tag{26}$$

On the other hand, if $q_2-b_2 < 0$, we have

$$\begin{aligned}
&E[V^\beta((\mathbf{q}-\mathbf{b})^+ + \mathbf{a})] \\
&\quad = E[V^\beta((q_1-b_1)+a_1, a_2)] && (27)\\
&\quad \geq E[V^\beta(q_1-d_1+a_1, d_1-b_1+a_2)] && (28)\\
&\quad = E[V^\beta(q_1-d_1+a_1, b_2-d_2+a_2)] && (29)\\
&\quad > E[V^\beta(q_1-d_1+a_1, q_2-d_2+a_2)] && (30)\\
&\quad = E[V^\beta((\mathbf{q}-\mathbf{d})^+ + \mathbf{a})] && (31)
\end{aligned}$$

where (28) follows from $(q_1-b_1, 0) \succeq (q_1-d_1, d_1-b_1)$ and Lemma 4, (29) follows from the fact that $d_1+d_2 = b_1+b_2$, and (30) is valid because we assumed that $q_2-b_2 < 0$, thus $q_2-d_2 > b_2-d_2$, and we apply Lemma 1. The results in

(26) and (31) contradict the optimality of $\mathbf{b}$, and therefore, $\mathbf{d}$ must be the optimal operating point.

Next, we prove that if there does not exist a point on the dominant face of the capacity region which equalizes the queue lengths, then the optimal operating point must be one of the corner points. Let us assume that the optimal operating point $\mathbf{d} = (d_1, d_2)$ is not a corner point, and without loss of generality, let us assume that $(q_1 - d_1)^+ > (q_2 - d_2)^+$. If $q_1 - d_1 > q_2 - d_2 \geq 0$, we can always find a small enough $\delta > 0$, such that the operating point $(d_1 + \delta, d_2 - \delta)$ is also on the dominant face, and $q_1 - (d_1 + \delta) > q_2 - (d_2 - \delta) > 0$. Since $(q_1 - d_1, q_2 - d_2) \succeq (q_1 - (d_1 + \delta), q_2 - (d_2 - \delta))$, based on Lemma 4, we have $E[V^\beta((\mathbf{q} - \mathbf{d})^+ + \mathbf{a})] \geq E[V^\beta(q_1 - (d_1 + \delta) + a_1, q_2 - (d_2 - \delta) + a_2)]$, and this contradicts the optimality of $\mathbf{d}$. On the other hand, if $q_1 - d_1 > 0 > q_2 - d_2$, we can also find a small enough $\delta > 0$, such that $q_1 - (d_1 + \delta) > 0 \geq q_2 - (d_2 - \delta)$, and $(d_1 + \delta, d_2 - \delta)$ is on the dominant face as well. Therefore, we have $0 < q_1 - (d_1 + \delta) < q_1 - d_1$, and $(q_2 - d_2)^+ = (q_2 - d_2 + \delta)^+ = 0$. According to Lemma 1, we have $V^\beta(q_1 - d_1 + a_1, a_2) > V^\beta(q_1 - (d_1 + \delta) + a_1, a_2)$ for any value of a_1 and a_2. Therefore, $E[V^\beta(q_1 - d_1 + a_1, a_2)] > E[V^\beta(q_1 - (d_1 + \delta) + a_1, a_2)]$, and this contradicts the optimality of $\mathbf{d}$. Hence, the optimal operating point, in this case, must be one of the corner points. ■

Using Theorem 1, we express the optimal operating point $\mathbf{d}^* = (d_1^*, d_2^*)$ as a function of the queue lengths $\mathbf{q} = (q_1, q_2)$

$$\mathbf{d}^* = \begin{cases} \left(\frac{q_1 - q_2 + C_s}{2}, \frac{q_2 - q_1 + C_s}{2}\right), & |q_1 - q_2| < 2C_1 - C_s \\ (C_1, C_s - C_1), & q_1 - q_2 > 2C_1 - C_s \\ (C_s - C_2, C_2), & q_1 - q_2 < C_s - 2C_1 \end{cases}$$

This optimal rate allocation scheme works on the dominant face of the capacity region and therefore maximizes the number of bits transmitted in each slot; and, at the same time, it tries to balance the queue lengths as much as possible, which, in turn, minimizes the probability that any one of the queues becomes empty in the upcoming slots. When a queue becomes empty, the system resources cannot be utilized most efficiently, as even though the user with an empty queue has power to transmit, it does not have any bits to transmit.

Finally, while we developed Theorem 1 for the discounted cost criterion, we can find a sub-sequence of discount factors β_n such that $\beta_n \to 1$ as $n \to \infty$. Therefore, the policy we developed is optimal for the average cost problem as well.

IV. Simulation Results

We consider a two-user AWGN multiple access channel, with $C_1 = C_2 = 20$ bits/slot and $C_s = 30$ bits/slot. The number of bits arriving at the transmitters in each slot follows a Poisson distribution with parameter λ. We compare two policies: the optimal policy developed in this paper which tries to balance the queue lengths in each slot and the LQHR algorithm developed in [6] which chooses a corner point of the capacity region and allocates the larger rate to the longer queue. We plot the average delay versus λ in Fig. 3.

We observe that when λ is small, both the LQHR policy and the queue balancing policy yield delay close to one slot,

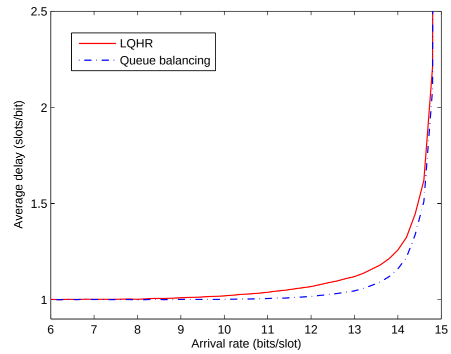

Fig. 3. Average delay versus arrival arrival rate.

and the difference between these two policies is insignificant. This is because, the system has a light traffic, and both policies empty both queues in almost all slots. When λ becomes very close to the boundary of the capacity region, the average delay grows rapidly under both policies, and again the difference between the two policies becomes insignificant. This is because, the system has a heavy traffic, and the probability that the queues become empty is very small under both policies, and the actual number of departures in each slot is almost the same for both policies. When λ is neither very small nor very large, the queue balancing policy outperforms the LQHR policy significantly. This is because, equalizing the queue lengths minimizes the probability that one queue is large while the other queue is empty or close to empty, and consequently utilizes the system resources more efficiently.

References

[1] A. Ephremides and B. Hajek, "Information theory and communication networks: an unconsummated union," *IEEE Transactions on Information Theory*, vol. 44, pp. 2416–2434, Oct 1998.

[2] I. E. Telatar and R. G. Gallager, "Combining queueing theory with information theory for multiaccess," *IEEE Journal on Selected Areas in Communications*, vol. 13, pp. 963–969, Aug 1995.

[3] R. A. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," *IEEE Transactions on Information Theory*, vol. 48, pp. 1135–1149, May 2002.

[4] I. Bettesh and S. Shamai, "Optimal power and rate control for minimal average delay: The single-user case," *IEEE Transactions on Information Theory*, vol. 52, pp. 4115–4141, Sep 2006.

[5] M. Goyal, A. Kumar, and V. Sharma, "Power constrained and delay optimal policies for scheduling transmission over a fading channel," *IEEE Infocom*, vol. 1, pp. 311– 320, March/Apr 2003.

[6] E. Yeh, "Delay-optimal rate allocation in multiaccess communications: A cross-layer view," *Multimedia Signal Processing, 2002 IEEE Workshop on*, pp. 404– 407, Dec 2002.

[7] N. Ehsan and T. Javidi, "Delay optimal transmission policy in a wireless multi-access channel," *IEEE Transactions on Information Theory*, no. 8, pp. 3745–3751, Aug 2008.

[8] E. Yeh, "Minimum delay multiaccess communication for general packet length distributions," *Proceedings of the Allerton Conference on Communication, Control, and Computing*, pp. 1536–1545, Sep/Oct 2004.

[9] S. Musy, "Delay and coding in multiple-user communications," Ph.D. dissertation, EPFL, 2007.

[10] T. M. Cover and J. A. Thomas, *Element of Information Theory*. New York: John Wiley and Sons, Inc, 1991.

[11] V. G. Kulkarni, *Modeling and analysis of stochastic systems*. Chapman and Hall/CRC, 1995.

[12] J. Walrand, *An introduction to queueing networks*. Prentice Hall, 1988.

[13] A. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its applications*. New York: Academic Press, 1979.