Delay Optimal Transmission Policy in a Wireless Multi-access Channel

Navid Ehsan, *Member, IEEE,* Tara Javidi, *Member, IEEE*

Abstract

In this note we consider the problem of delay optimal rate allocation in a (potentially asymmetric) multi-access channel. The rate feasibility region of such a network is well-studied and is shown to be of a polymatroid structure. We consider this problem with unsaturated sources, i.e. jobs arrive at sources at random times and the source has the possibility of being empty. In such a setting, all stable rate allocation policies result in a throughput matched with the average arrival rate. Hence, we are interested in rate allocation policies that minimize expected delay in the system. In this paper, we show that a policy of threshold type is optimal in minimizing the average queueing delay. We study the average delay criterion as the limit of an infinite horizon discounted cost function when the discount factor approaches 1.

Index Terms

Delay optimal policy, rate allocation, multi-access channel.

I. Introduction

In this note, we consider optimal rate assignment for a general class of multi user systems where the achievable rate region (denoted by $\mathcal{R}$) is time invariant and is given by a set of inequalities corresponding to a polymatroid structure. In other words, for a set of users denoted by $\mathcal{N}$ sharing a common channel, where r_i is the service rate for queue i, we can operate at any point in the set $\mathcal{R}$:

$$\mathcal{R} = \{\mathbf{r} : \sum_{i \in \mathcal{S}} r_i \leq k_{\mathcal{S}}, \forall \mathcal{S} \subset \mathcal{N}\}, \tag{1}$$

where $k_{\mathcal{S}}$ is a constant that depends on the subset $\mathcal{S}$.

The polymatroid structure of (1) forms a general model for the capacity and/or feasible rate region in many multi-user settings, examples of which are illustrated below .

Example 1: Additive Gaussian Multi-Access Channel with Slow Fading- The information theoretic capacity of this channel can be derived as follows [1].

$$\sum_{i \in \mathcal{S}} r_i \leq W \log_2(1 + \frac{\sum_{i \in \mathcal{S}} P_i}{N}) \ , \ \forall \ \mathcal{S} \subset \mathcal{N} \ ,$$

where W is the channel bandwidth, P_i is the received power from user i, and N is the Gaussian noise power.

Example 2: MIMO Multi-access channel with No CSI at Transmitter- Consider a Multiple Input Multiple Output (MIMO) channel where the input-output relationship is given by

$$\mathbf{y}[m] = \sum_{k=1}^{2} H_k[m]\mathbf{x}_k[m] + \mathbf{w}[m] \ ,$$

where vectors $\mathbf{y}[m]$ and $\mathbf{x}_k[m]$ represent the signal received at the MIMO receiver and the signal transmitted by user k, respectively. Matrix H_k represents the channel matrix for user k, and $\mathbf{w}[m]$ represents the Gaussian noise term. The information

Navid Ehsan is with Nextwave Broadband Inc, San Diego, CA. E-mail address: nehsan@nextwave.com

Tara Javidi is with University of California, San Diego, CA. E-mail address: tara@ece.ucsd.edu

This work was supported in part by the Center for Wireless Communications (CWC), CWC industrial sponsors LGE and VIASAT, ARO-MURI Grant No. W911NF-04-1-0224, NSF CAREER Award No. CNS-0533035, and AFOSR Grant No. FA9550-05-01-0430.

theoretic capacity of this channel for a two user scenario $\mathcal{N} = \{1, 2\}$, assuming no channel state information (CSI) at the transmitter and perfect CSI at the receiver and constant transmission power, can be derived as follows [2].

$$\begin{aligned} r_i &\leq E\left[W \log_2(1 + ||\mathbf{H}_i P_i||^2)\right] \\ r_1 + r_2 &\leq E\left[W \log_2 \det(\mathbf{I} + \frac{\mathbf{H}\mathbf{K}_\mathbf{x}\mathbf{H}^*}{N})\right], \end{aligned}$$

where W is the channel bandwidth, P_i is the vector of transmit power from user i, $\mathbf{K}_\mathbf{x}$ is the covariance matrix, and N is the Gaussian noise power.

In this paper, we are not concerned with the computation and derivation of capacity/rate region as they have been widely studied, e.g. see [1], [3], [4]. Instead we are interested in the delay optimality of various rate allocation policies when bits arrive stochastically and whose average rate of arrivals lie inside the capacity region.

When all the users are assumed to be saturated (i.e. always have packets to transmit and the queues are never empty), the class of throughput optimal policies is known. This class consists of policies operating on the dominant face of the capacity region (assuming that the rate region is time invariant). However, when stochastic arrivals are taken into account, consequently provisioning for the possibility of having empty queues, the optimal policy depends not only on the immediate aggregate throughput, but also on the queue occupancies of each user and the stochastic nature of the arrival and departure processes. The queueing stability of a multiple access channel has been studied in [5], [6]. When the arrival rates lie in the capacity region, these papers introduce policies that guarantee stability of the queues. However, these policies, in general, may not minimize the expected delay.

In this note, we are interested in those policies that minimize the expected average packet delay, when the arrival rates lie in the capacity region. We study the qualitative structure of the delay optimal policy- a function of queue lengths- which minimizes the average packet delay in a system with random arrivals. In particular, we study the optimal rate assignment in a two user system with Poisson traffic and a polymatroid capacity/rate region, i.e. when the feasible/reliable rate of serving information bits is defined by the set of inequalities in (1) for $|\mathcal{N}| = 2$.

Yeh, et. al, in [7], considered a time-varying but symmetric version of this problem and studied the optimal policy for minimizing the average *bit delay*. It was shown that the policy that serves the longest queue with the highest rate (in case of our Example 1, this translates to choosing the order of cancellation) minimizes the average bit delay. In [8] and [9] delay optimal policies have been studied with the assumption of Poisson arrival process with equal arrival rates, and exponentially distributed packets with the same parameter, and exchangeable fading process. These works are closest to our work in that their special case of time-invariant channel scenario coincides with our special case when the arrival processes as well as capacity region is symmetric. Specifically, the asymmetric arrival process, processing rate, and packet length distribution is the major difference between our work and those studied in [7], [8], [9]. In this paper, we assume a (possibly asymmetric) time-invariant rate region (this can be the model of a slow fading channel, ergodic capacity region, or the case of no channel state information at the transmitter) and consider the problem of minimizing the average job delay in presence of asymmetric arrival processes and channel conditions. We show that the optimal policy operates on the extreme points of the polymatroid, and is of a threshold nature. That is, there exists a switching curve whereby the optimal operating point is at one extreme point on the dominant face of the capacity region if the queue state vector lies on one side of the curve, and the operating point is at the other extreme point of the dominant face if the queue state vector lies on the other side of the curve. To obtain this result, we map this problem to a multi-server queueing problem and use techniques in Markov decision theory. We postpone the

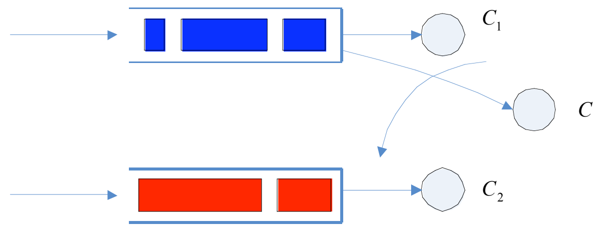

Fig. 1. The model with two dedicated servers and one shared server

discussion and citation of the related queueing and stochastic control theoretic work to Section II-A, after problem definition.

The remainder of this paper is organized as follows. In the next section, we introduce an equivalent multi-server queueing model and formulate the problem of delay optimal scheduling as a Markov decision problem with an average cost criterion. In Section III, we introduce a finite horizon discounted variant of the problem, whose solution is then used to construct a solution for the infinite horizon version of the problem (both for a discounted cost and the average cost) in Section IV. Finally, in Section V, we summarize and conclude the paper.

We close this section with a word on our notation. We will use bold face letters to represent vectors and normal-size letters to represent scalars or random variables. For example $\mathbf{x}(t) = (x_1, x_2)$ is the buffer occupancy of user 1 and 2 at time t. We use $\mathbf{e}_i$ to denote the i-th unit vector.

II. Problem Formulation and Assumptions

Consider two queues with Poisson arrivals with rates λ_1 and λ_2. Packets have a length exponentially distributed with mean L. Each queue has a dedicated server which serves the packets with constant rate C_i. In addition to the private servers, there is a server with rate C that can be allocated to either queue at any time (see Fig. 1).

Let $\mu_i = \frac{C_i}{L}$ and $\mu = \frac{C}{L}$. This model corresponds to the capacity region described in (1) for two users by putting[1],

$$\begin{aligned} \mu_1 &= k_3 - k_2 \ , \\ \mu_2 &= k_3 - k_1 \ , \\ \mu &= k_1 + k_2 - k_3 \ . \end{aligned} \tag{2}$$

By randomizing the allocation of the shared server, all the points on the boundary of the capacity region described by (1) can be achieved. In other words, due to the polymatroid structure of the capacity region, the problem of optimal rate assignment is equivalent to the queueing model in Fig. 1. Given this equivalence, the objective is to identify a server scheduling policy which at a given time determines a (potentially randomized) allocation of the shared server in Fig. 1 as to minimize the expected average delay. we formulate the following generalization of our problem:

Problem (P)

Consider the multi-server queueing problem illustrated in Fig. 1. Arrival process to queue $Q_i, i = 1, 2$ follows a Poission process with rate λ_i. Packet size is stochastic and follows an exponential distribution with mean L. Servers are deterministic with service rates C_1, C_2 and C. Consider the average cost

$$J^{\pi}_{av} = \lim_{T\to\infty} \frac{1}{T} \int_{t=0}^{\infty} c(\mathbf{x}^{\pi}(t))dt \ ,$$

[1]Note that in order to avoid triviality we have assumed $k_3 \geq k_1, k_2$, and $k_1 + k_2 \geq k_3$.

where $\mathbf{x}^{\pi}(t)$ is the state of the queues at time t under policy π and $c(\mathbf{x})$ is the instantaneous cost associated with having $\mathbf{x} = (x_1, x_2)$ packets in Q_1, and Q_2.

Find server allocation policy π^* as to minimize J^{π}_{av} in expected sense.

To keep the problem meaningful, we restrict our attention to the set of admissible arrival processes. This is the set of arrival rates for which there exists a policy which ensures positive recurrence, hence a finite delay.

Assumption 1: We assume that $\sum_i \lambda_i < \mu + \sum_i \mu_i$.

As stated earlier, it is our goal to allocate the common server in a manner as to minimize the average delay in the system. By Little's theorem we have $D = (E(x_1) + E(x_2))/(\lambda_1 + \lambda_2)$ where D is the average packet delay, x_i is the number of packets in queue i and $E(.)$ is the expectation. Therefore in order to minimize the delay it suffices to minimize the average number of packets in the system. Motivated by this observation, we restrict our choice of instantaneous cost $c(\mathbf{x})$ to a simple generalization of $x_1 + x_2$. In other words, we are only interested in cost functions $c(\mathbf{x})$ that satisfy certain conditions.

In other words, for the rest of the paper:

Assumption 2: We assume that $c(\mathbf{x}) \in \mathcal{F}$.

Definition 1: A function $f : \mathbb{Z}_+^2 \to \mathbb{R}$ belongs to the set $\mathcal{F}$ if $f(\mathbf{x})$ satisfies the following conditions:

C.1 (monotonicity or non-decreasing condition)

$f(\mathbf{x}) \leq f(\mathbf{x} + \mathbf{e}_i), \quad i \in \{1, 2\};$

C.2 (supermodularity condition)

$f(\mathbf{x} + \mathbf{e}_1) + f(\mathbf{x} + \mathbf{e}_2) \leq f(\mathbf{x}) + f(\mathbf{x} + \mathbf{e}_1 + \mathbf{e}_2);$

C.3.a (superconvexity condition)

$f(\mathbf{x} + \mathbf{e}_1) + f(\mathbf{x} + \mathbf{e}_1 + \mathbf{e}_2) \leq f(\mathbf{x} + \mathbf{e}_2) + f(\mathbf{x} + 2\mathbf{e}_1);$

C.3.b (superconvexity condition)

$f(\mathbf{x} + \mathbf{e}_2) + f(\mathbf{x} + \mathbf{e}_1 + \mathbf{e}_2) \leq f(\mathbf{x} + \mathbf{e}_1) + f(\mathbf{x} + 2\mathbf{e}_2).$

Here the terminologies follow that used in [10]. As mentioned earlier, when $c(\mathbf{x}) = x_1 + x_2$, Problem (P) is equivalent to minimizing the average number of packets, hence to minimizing the average packet delay. Note that Conditions **C.1-C.3.b** are rather benign and set $\mathcal{F}$ includes a large number of functions of interest, e.g. all functions of the form $c_1 x_1^{n_1} + c_2 x_2^{n_2}$ (for any $c_1, c_2, n_1, n_2 > 0$) satisfy these conditions. Condition **C.1** results in monotonic increasing cost functions and ensures that stability is necessary for optimality (see [11]). Conditions **C.2** and **C.3.a** result in the convexity of f in x_1; while **C.2** and **C.3.b** imply the convexity of f in x_2. An example of a function that does not satisfy the above conditions is $f(x_1, x_2) = x_1 x_2$ which fails conditions **C.3.a** and **C.3.b**.

Note that for Poisson arrivals and exponentially distributed packet length, the vector of the number of packets constitutes a state. Thus, assuming that the actual number of bits in each arriving packet is not known, in search for the optimal policy, one can restrict attention to control policies which depend on $\mathbf{x}(t) = (x_1, x_2)$. An event (potential departure or arrival) occurs in the system with rate $\Lambda = \lambda_1 + \lambda_2 + \mu_1 + \mu_2 + \mu$.

Note that in this note, we use common definitions of Markov and stationary (non-randomized) policies (e.g. see [11]):

Definition 2: A Markov policy is defined as a function from the state space to the set of all probability distributions over action space. Mathematically, a Markov policy $\pi : \mathbb{Z}_+^2 \to \mathcal{P}$, where $\mathcal{P}$ is the set of all probability distributions on the space of actions: $\{0, 1, 2\}$, corresponding to idling, allocation to Q_1, and allocation to Q_2, respectively.

In simpler and informal language, a Markov policy is a sequence (over time) of randomized functions of the current state onto the action space, that does not depend on the previous actions or the state of the system in the past. For the two user scenario, a non-idling Markov policy can be defined by a probability p, such that the policy allocates the shared server to queue 1 with probability p and allocate it to queue 2 with probability $1-p$. Therefore a Markov policy $\pi : \mathbb{Z}_+^2 \to [0,1]$ can be defined to allocate a probability p to allocating the server to queue 1, when the state is (x_1, x_2).

Definition 3: A stationary (non-randomized and non-idling) Markov policy is defined to be a function which allocates an action to each state in a stationary and deterministic manner. In other words, stationary policy $\pi : \mathbb{Z}_+^2 \to \{1,2\}$ is a mapping from state space onto the action space. Action i corresponds to allocating the shared server to queue i.

The Dynamic Programming theorem in [11] provides sufficient conditions for the optimality of a stationary (non-randomized) Markov policy for an infinite horizon MDP with expected average cost criterion and unbounded cost function over countable state space (e.g. see Theorem 4.1 in [11]). These sufficiency conditions require that (i) at least one stable policy exists, and (ii) the cost function grows large outside any compact set (hence preventing unstable policies to be optimal). It is straightforward to show that these conditions are satisfied in our problem, when restricting attention to stabilizable arrival process, i.e. $\sum_i \lambda_i \leq \mu + \sum_i \mu_i$, and class-$\mathcal{F}$ cost functions. Furthermore, due to the increasing property of function $c(.)$ restriction to non-idling policies incurs no loss of optimality. In summary, in the setting of our interest, an optimal policy can be found by restricting the search to stationary (non-randomized and non-idling) Markov policies, i.e. the optimal policy is a mapping from state space onto the action space and only changes the server allocation when the state changes, i.e. at instance of new arrivals or departures. The important consequence of this result is that restricting attention to policies which do not allow for rate splitting (i.e. simultaneous/randomized use of the shared server) do not incur any loss of optimality.

Now we are ready to study Problem (P). We sudy this problem as the limit of an infinite horizon discounted cost dynamic programming problem as the discount factor goes to one [12]. The infinite horizon discounted cost problem can be realized as the limit of the finite horizon discounted cost problem as will be discussed in Section IV. In Section III, we study the finite horizon optimization problem.

Before we proceed with the solution to Problem (P), we discuss the related work in queueing literature briefly.

A. Related Work

The problem studied in this note can be cast as a special case of the well-known *restless bandit* problem [13], [14], [15] posed by Whittle in 1988. A restless bandit is a decision problem where parallel bandits represent controlled Markov chains (queues in our problem) whose control and state result in a cost/reward. The classical multi-armed bandit is the question of how the chains, one at a time, are operated in order to arrive at the optimal additive cost/reward over the horizon of interest. While in classical multi-armed bandit the arms that are not played remain frozen in time, in the restless generalization all chains undergo state transitions (due to arrivals and dedicated servers in our setting) even when they are not played or selected (allocating the shared server in our problem). A general optimal solution is not known for this class of problems. In fact, [16] proved that this problem is complete for polynomial space. As mentioned in [16], "PSPACE-completeness does not immediately imply intractability, but it strongly suggests it. It is considered a much more convincing evidence of intractability than NP-hardness." This signifies the importance of finding structural properties (see [15], [17], [18]) of the optimal policy when dealing with cases of restless bandit problem. For instance, the threshold nature identified in this note for the optimal policy reduces the complexity of finding the optimal solution to finding the optimal threshold. Even though finding the optimal threshold is itself a difficult problem, limiting the set of policies to a threshold policy significantly reduces the complexity.

We also note that [19], [20] study the performance of specific queueing systems under threshold policies and provide methods for determining the optimal threshold in those scenarios. Unfortunately, these results only apply to the specific policies considered in those papers and cannot be generalized to other policies or scenarios. In particular, [20] studies a two-queue model with a shared server. A threshold policy is defined such that when the server is allocated to queue one, it serves this queue until it becomes empty. Then it starts serving the second queue until either it becomes empty or the first queue's occupancy exceeds a threshold T. Although our queueing model has some similarities with this work, there are some major differences that make the results in [20] inapplicable to the scenario considered here. The first difference is the nature of our results: while work in [20] involves a performance analysis of a particular family of policies, our work focuses on the structural property of a delay optimal policy. The more important difference between our work and [20] is the structure of servers. In our problem, in addition to the shared server, each queue also has a "dedicated" server. When a queue becomes empty, its dedicated server is wasted which makes finding the optimal policy for allocating the shared server considerably more complicated than that studied in [20]. Also in this paper, the server can be allocated to any queue at any given time, i.e. no restriction that the queue should become empty (or the other queue exceeds a threshold) before reallocating.

III. Finite horizon discounted cost problem

In this section we define a finite horizon version of the problem (P), where the objective is to minimize the discounted cost over a finite horizon. We define an "event" to be either arrival of a new packet or completion of service for an already existing packet.

Consider the following cost function:

$$J^{\pi}_{discounted} = \int_{t=0}^{T_K} e^{-\alpha t} c^{\pi}(\mathbf{x}(t))dt, \tag{3}$$

where T_K is the time when K^{th} event occurs. Let $\mathbf{x}_k$ be the queue length right after the k^{th} event. Using standard uniformization techniques [21], this problem can be reformulated as a discrete time optimization problem with the following cost:

$$J^{\pi}_{discrete} = \sum_{k=0}^{K-1} \beta^k c(\mathbf{x}_k) \tag{4}$$

where $\beta = \frac{\alpha}{\alpha+\Lambda}$.

Define the k step value function as follows:

$$V_k(\mathbf{x}) = \min_{\pi} E[J^{\pi}_{discrete}] = \min_{\pi} E[\sum_{k'=0}^{k-1} \beta^{k'} c(\mathbf{x}_k)|\mathbf{x}_0 = \mathbf{x}] \ , \tag{5}$$

where π belongs to the set of stationary (non-randomized and non-idling) Markov policies. This value function can be found recursively as follows [12]:

$$\begin{aligned} V_0(\mathbf{x}) &= 0 \\ V_k(\mathbf{x}) &= c(\mathbf{x}) + \beta\Lambda^{-1}\{\sum_{i=1}^{2}(\lambda_i V_{k-1}(\mathbf{x}+\mathbf{e}_i) + \mu_i V_{k-1}([\mathbf{x}-\mathbf{e}_i]^+)) + \mu \min_i \{V_{k-1}([\mathbf{x}-\mathbf{e}_i]^+)\}\} \ , \end{aligned} \tag{6}$$

where the first term corresponds to the arrival rate, second term corresponds to the service rate of the dedicated servers. The last term corresponds to the service rate of the shared server. Note that if the shared server is dedicated to user i, then the service rate of user i is equal to $\mu_i + \mu$. Therefore, the shared server should be allocated in order to minimize the last term in Equation (6).

In subsequent discussion we set $V_k(x_1, x_2) = V_k(\max(0, x_1), \max(0, x_2))$ to simplify the notation (i.e. we let $V_k(.)$ to be defined on negative values as well). Note that in order to solve (6) we only need to consider $\mathbf{x} \geq 0$.

Theorem 1: The optimal value function $V_k(.)$, belongs to $\mathcal{F}$ for all $k \geq 0$. Furthermore the optimal policy at each step is of threshold type.

Proof - Since $V_0(\mathbf{x}) = 0$, we have that $V_0(\mathbf{x}) \in \mathcal{F}$. On the other hand we have $c(.) \in \mathcal{F}$. Therefore using induction (similar to the method used in [22], [23]) it can be easily shown that $V_k(\mathbf{x}) \in \mathcal{F}$ for all $k \geq 0$.

The optimality of a threshold policy is a direct result of superconvexity property. By the first part of this theorem, $V_{k-1} \in \mathcal{F}$ for all $0 < k \leq K$. Thus by property **C.3.a** we have $V_{k-1}(\mathbf{x}+\mathbf{e}_1) + V_{k-1}(\mathbf{x}+\mathbf{e}_1+\mathbf{e}_2) \leq V_{k-1}(\mathbf{x}+2\mathbf{e}_1) + V_{k-1}(\mathbf{x}+\mathbf{e}_2)$.

By replacing $\mathbf{x}$ with $\mathbf{x}-\mathbf{e}_1-\mathbf{e}_2$ we have $V_{k-1}(\mathbf{x}-\mathbf{e}_2) + V_{k-1}(\mathbf{x}) \leq V_{k-1}(\mathbf{x}+\mathbf{e}_1-\mathbf{e}_2) + V_{k-1}(\mathbf{x}-\mathbf{e}_1)$. Rearranging, we get

$$V_{k-1}(\mathbf{x}-\mathbf{e}_2) - V_{k-1}(\mathbf{x}-\mathbf{e}_1) \leq V_{k-1}(\mathbf{x}+\mathbf{e}_1-\mathbf{e}_2) - V_{k-1}(\mathbf{x}) \ .$$

The last inequality suggests that if the left hand side is non-negative, then the right hand side is also non-negative. Therefore if the optimal decision is to allocate the shared server to the first queue when the state is $\mathbf{x}$ for some $\mathbf{x}$, then it is optimal to allocate the server to the first queue when the state is $\mathbf{x}+\mathbf{e}_1$. Similarly using $\mathbf{C.3.b}$ we can show that if the optimal decision is to allocate the shared server to the second queue when the state is $\mathbf{x}$, then it is optimal to allocate the server to the second queue when the state is $\mathbf{x}+\mathbf{e}_2$. We can then define a threshold as follows.

$$g_k(x_1) = \min\{x_2 | V_{k-1}(\mathbf{x}-\mathbf{e}_2) \leq V_{k-1}(\mathbf{x}-\mathbf{e}_1)\}, \tag{7}$$

and $g_k(x_1) = \infty$ when the above set is empty. If for time horizon k we have $x_2 \geq g_k(x_1)$ then the optimal policy is to assign the shared server at time t to queue 2, otherwise to queue 1 (if the set is empty then the threshold is infinity), proving the optimality of a threshold policy. ∎

While Theorem 1 shows that the optimal scheduler is of the threshold type, it is worth pointing out that it is in general difficult to obtain the quantitative value of the threshold. The threshold is given by Equation (7), where the current cost-to-go function needs to be calculated. This can be computationally expensive.

IV. Infinite horizon

In this section we let K (and as a result T_K) to go to infinity. For the infinite horizon case, we can define two different cost functions, the discounted cost function and the average cost function. We will study these two extensions separately.

A. Discounted cost

We define the infinite horizon discrete time discounted cost function under policy π as follows.

$$J^{\infty,\pi}_{discrete} = \lim_{K\to\infty} \sum_{k=0}^{K-1} \beta^k c(\mathbf{x}_k) \tag{8}$$

The value function can be defined as:

$$V_\infty(\mathbf{x}) = \min_\pi E[J^{\infty,\pi}_{discrete}] = \min_\pi E[\lim_{k\to\infty} \sum_{k'=0}^{k-1} \beta^{k'} c(\mathbf{x}_{k'}) | \mathbf{x}_0 = \mathbf{x}] \tag{9}$$

In this case under the additional constraint that $c(\mathbf{x}) \geq 0$ for all $\mathbf{x} \geq 0$, we have the following result.

Lemma 1: If $c(\mathbf{x}) \geq 0$ for all $\mathbf{x} \geq 0$ then we have $V_\infty(\mathbf{x}) = \lim_{k\to\infty} V_k(\mathbf{x})$.

A proof of this lemma is provided in [24]Chapter 5.4. Note that the convergence and existence of $J^{\infty,\pi}_{discrete}$ is due to the bounded nature of discounted cost.

To allow for the use of Lemma 1 and for the rest of the paper, we will make the following assumption.

Assumption 3: We will assume that $c(\mathbf{x}) \geq 0$.

Theorem 2: We have $V_\infty(.) \in \mathcal{F}$. Furthermore, the optimal policy for the infinite horizon optimization is of threshold type.

Proof - Note that the set $\mathcal{F}$ is closed under limit operation. Since $V_k \in \mathcal{F}$ for all k, therefore by Lemma 1, we have $V_\infty \in \mathcal{F}$. Therefore by superconvexity of $V_\infty(.)$ the optimality of a threshold policy is concluded. ∎

Note that in the infinite horizon scenario the policy is stationary whereas in the finite horizon case the threshold can change as the time index changes.

B. Average cost

Define the average cost as follows.

$$J_{av} = \lim_{K\to\infty} \frac{1}{K} \sum_{k=0}^{K-1} c(\mathbf{x}_k) \tag{10}$$

We consider this cost as the limit of the discounted cost problem as $\beta \to 1^-$. We define the value function the same as in the previous section, however instead of the subscript ∞ we use the subscript β in order to emphasize the dependency on β:

$$V_\beta(\mathbf{x}) = \min_\pi E[\lim_{k\to\infty} \sum_{k'=0}^{k-1} \beta^{k'} c(\mathbf{x}_{k'})|\mathbf{x}_0 = \mathbf{x}] \tag{11}$$

By Lemma 1 the function $V_\beta(\mathbf{x})$ satisfies the following recursion:

$$V_\beta(\mathbf{x}) = c(\mathbf{x}) + \beta\Lambda^{-1}\{\sum_{i=1}^{2}(\lambda_i V_\beta(\mathbf{x}+\mathbf{e}_i) + \mu_i V_\beta([\mathbf{x}-\mathbf{e}_i]^+)) + \mu \min_i\{V_\beta([\mathbf{x}-\mathbf{e}_i]^+)\}\} \tag{12}$$

Define the operator T as follows:

$$Tf(\mathbf{x}) = \sum_{i=1}^{2}(\lambda_i f(\mathbf{x}+\mathbf{e}_i) + \mu_i f([\mathbf{x}-\mathbf{e}_i]^+)) + \mu \min_i\{f([\mathbf{x}-\mathbf{e}_i]^+)\} \tag{13}$$

Then Equation (12) can be written as:

$$V_\beta(\mathbf{x}) = c_(\mathbf{x}) + \beta\Lambda^{-1}TV_\beta(\mathbf{x}) \tag{14}$$

Note that Assumption 1 guarantees the existence of a Markov stationary (non-randomized and non-idling) policy under which both queues are stable.

Theorem 3: Suppose Assumptions 1, 2, and 3 hold. Then,

(a) There exists a finite constant J^* that satisfies the following inequality:

$$J^* + h(\mathbf{x}) \geq \bar{c}(\mathbf{x}) + Th(\mathbf{x}) \ . \tag{15}$$

(b) Let π^* be a policy that minimizes the right hand side of (15). Then π^* is the optimal average cost policy.

(c) J^* is the optimum average cost.

In order to prove theorem 3 we need the following lemmas.

Lemma 2: Starting from any state $\mathbf{x}$, there exists a policy $\pi_\mathbf{x}$ so that under this policy the system will go to zero state with finite cost. We denote this finite cost by $U(\mathbf{x})$.

This is a direct result of Assumption 1 and the fact that under this assumption there exists a policy under which the system is stable.

Lemma 3: $V_\beta(\mathbf{x})$ is non-decreasing in $\mathbf{x}$. Moreover, under Assumption 1 we have

$$V_\beta(\mathbf{x}) - V_\beta(\mathbf{0}) \leq U(\mathbf{x}) \ . \tag{16}$$

Proof - In Theorem 2 we have shown that $V_\beta(\mathbf{x})$ is non-decreasing. To show that (16) holds, consider the policy π^* that follows policy $\pi_{\mathbf{x}}$ until the first time state $\mathbf{0}$ is reached and then follows the optimal policy. Therefore we have

$$V_\beta(\mathbf{x}) \leq V_\beta^{\pi^*}(\mathbf{x}) = U(\mathbf{x}) + V_\beta(\mathbf{0}),$$

thus proving the lemma. ∎

Lemma 4: Suppose $c(\mathbf{x}) \geq 0$ for all $\mathbf{x} \geq \mathbf{0}$. Then under Assumption 1, the quantity $(1-\beta)V_\beta(\mathbf{0})$ is bounded for $\beta \in (0,1)$.

Proof - Note that when $c(\mathbf{x}) \geq 0$, using Assumption 1 and Lemma 2 we have, $E^{\pi_\mathbf{0}}[c(\mathbf{x}_k)|\mathbf{x}_0 = \mathbf{0}] \leq U(\mathbf{0})$. This can be argued as follows. Under policy $\pi_\mathbf{0}$, state $\mathbf{0}$ is a recurrent state and thus any state at time t lies in between two consecutive occurrences of state $\mathbf{0}$. Since the expected sum of all costs in between those two occurrences is less than or equal to $U(\mathbf{0})$ and all costs are non-negative, the cost at each time step has to be less than or equal to $U(\mathbf{0})$. Thus we have

$$\begin{aligned} & (1-\beta)V_\beta(\mathbf{0}) \leq (1-\beta)V_\beta^{\pi_\mathbf{0}}(\mathbf{0}) = (1-\beta)E^{\pi_\mathbf{0}}[\lim_{k\to\infty}\sum_{k'=0}^{k-1}\beta^{k'}c(\mathbf{x}_u)|\mathbf{x}_0 = \mathbf{0}] \\ = \quad & (1-\beta)\lim_{k\to\infty}\sum_{k'=0}^{k-1}\beta^{k'}E^{\pi_\mathbf{0}}[c(\mathbf{x}_{k'})|\mathbf{x}_0 = \mathbf{0}] \\ \leq \quad & (1-\beta)\lim_{k\to\infty}\sum_{k'=0}^{k-1}\beta^{k'}\cdot U(\mathbf{0}) = U(\mathbf{0})\ , \end{aligned}$$

where the first inequality is due to the fact that $\pi_\mathbf{0}$ is not necessarily the optimal policy. The exchange of the limit and expectation is a result of the assumption that $c(\mathbf{x}) \geq 0$ (and consequently the fact that the sum inside the expectation is non-decreasing) and the last inequality holds by Assumption 1. ∎

Lemma 5: Let β_n be a sequence of real numbers such that $\beta_n \to 1^-$ as $n \to \infty$. Under Assumption 1, then there exists a subsequence α_n such that

$$\lim_{n\to\infty}(V_{\alpha_n}(\mathbf{x}) - V_{\alpha_n}(0)) = h(\mathbf{x})\ ,$$

where $0 \leq h(\mathbf{x}) \leq U(\mathbf{x})$ for all $\mathbf{x} > \mathbf{0}$.

Proof - Note that $h_{\beta_n} = V_{\beta_n}(\mathbf{x}) - V_{\beta_n}(0) \leq U(\mathbf{x})$ by Lemma 3. On the other hand due to monotonicity of $V_{\beta_n}(\mathbf{x})$ (same lemma) we have $h_{\beta_n} = V_{\beta_n}(\mathbf{x}) - V_{\beta_n}(0) \geq 0$. Therefore $h_{\beta_n} \in [0, U(\mathbf{x})]$. Since the space $[0, U(\mathbf{x})]$ is compact, there exists a subsequence α_n for which $h_{\alpha_n}(\mathbf{x})$ converges. We denote this value by $h(\mathbf{x})$. This holds for all $\mathbf{x} \in Z_+^2$.

Since $0 \leq h_{\alpha_n}(\mathbf{x}) \leq U(\mathbf{x})$ for all n we have $0 \leq h(\mathbf{x}) \leq U(\mathbf{x})$. ∎

Now define $h_\beta(\mathbf{x})$ as follows:

$$h_\beta(\mathbf{x}) = V_\beta(\mathbf{x}) - V_\beta(\underline{0}) \tag{17}$$

If $\beta_n \to 1^-$, then it is shown in Lemma 5 that under Assumption 1 one can find a subsequence α_n such that $\lim_{n\to\infty} h_{\alpha_n}(\mathbf{x})$ exists. We call this limit function $h(\mathbf{x})$. We then have all components needed for proof of Theorem 3.

Proof of Theorem 3 - Let β_n be a sequence of real numbers such that $\beta_n \to 1^-$ as $n \to \infty$. If we add $\beta_n V_{\beta_n}(\mathbf{0})$ to both sides of (12) we get:

$$(1-\beta_n)V_{\beta_n}(\mathbf{0}) + (V_{\beta_n}(\mathbf{x}) - V_{\beta_n}(\mathbf{0})) = c_(\mathbf{x}) + \beta_n\Lambda^{-1}T(V_{\beta_n}(\mathbf{x}) - V_{\beta_n}(\mathbf{0}))$$

By Lemma 5, there exists a sequence $\alpha_n \to 1^-$ such that $\lim_{n\to\infty}(V_{\alpha_n}(\mathbf{x}) - V_{\alpha_n}(0)) = h(\mathbf{x})$. By Lemma 4 the value $(1-\alpha_n)V_{\alpha_n}(\mathbf{0})$ is bounded. Therefore there exists a sequence $\gamma_n \to 1^-$ for which this value converges to a real number which we call J. Therefore by replacing γ_n in place of β_n in the above equation, taking the limit and using Fatou's lemma we have

$$J^* + h(\mathbf{x}) \geq \bar{c}(\mathbf{x}) + Th(\mathbf{x}) \ .$$

Now assume that policy π^* minimizes the right hand side of (15). First we show that $\bar{J}^{\pi^*} \leq J^*$. Let $\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_k$ be the (random) states that are visited during times $1, 2, \cdots, k+1$, then using (15) we have,

$$\begin{aligned} J^* + h(\mathbf{x}_0) &\geq \bar{c}(\mathbf{x}_0) + E[h(\mathbf{x}_1)|\mathbf{x}_0], \\ J^* + h(\mathbf{x}_1) &\geq \bar{c}(\mathbf{x}_1) + E[h(\mathbf{x}_2)|\mathbf{x}_1], \\ &\cdots \\ J^* + h(\mathbf{x}_{k-1}) &\geq \bar{c}(\mathbf{x}_{k-1}) + E[h(\mathbf{x}_k)|\mathbf{x}_{k-1}] \end{aligned}$$

Taking the expected value on both sides and adding the equations we get,

$$\frac{1}{k}\sum_{u=0}^{k-1} E[c(\mathbf{x}_u)] \leq J^* + \frac{(h(\mathbf{x}_1) - h(\mathbf{x}_k))}{k} \leq J^* + \frac{h(\mathbf{x}_0)}{k}, \tag{18}$$

where the second inequality is due to the fact that $h(\mathbf{x}_k) \geq 0$. Taking the limit from both sides of (18) as $t \to \infty$ and using the fact that $h(\mathbf{x}) \leq U(\mathbf{x})$ we have $\bar{J}^{\pi^*} \leq J^*$.

Now consider any other policy π'.

$$J^{\pi^*} \leq J^* \leq \limsup_{\beta \to 1^-} (1-\beta) V_\beta(\mathbf{x}) \leq \limsup_{\beta \to 1^-} (1-\beta) V_\beta^{\pi'}(\mathbf{x}) \leq J^{\pi'} \tag{19}$$

Therefore π is the optimal average cost policy. On the other hand if we let $\pi' = \pi^*$, then we scan see that J^* is the optimal average cost. ∎

From this theorem we also have the following corollary.

Corollary 1: We have $h(\mathbf{x}) \in \mathcal{F}$. Hence, the optimal policy is of threshold type.

Proof - Let β_n be a sequence such that $\beta_n \to 1^-$. Then by Lemma 5, there exists a subsequence α_n such that $\lim_{n\to\infty} h_{\alpha_n}(\mathbf{x})$ exists. On the other hand we have for any α_n,

$$h_{\alpha_n}(\mathbf{x}) = V_{\alpha_n}(\mathbf{x}) - V(\alpha_n)(\underline{0}) \ .$$

Since we have $V_{\alpha_n} \in \mathcal{F}$ for all α_n, taking the limit of both sides as $\alpha_n \to 1^-$ and noting that $\mathcal{F}$ is closed under the limit operation, we conclude that $h(\mathbf{x}) \in \mathcal{F}$.

Note that we have,

$$Th(\mathbf{x}) = \sum_{i=1}^{2} (\lambda_i h(\mathbf{x} + \mathbf{e}_i) + \mu_i h([\mathbf{x} - \mathbf{e}_i]^+)) + \mu \min_i \{h([\mathbf{x} - \mathbf{e}_i]^+)\} \ ,$$

Therefore the optimal policy allocates the server to the queue i that minimizes $h([\mathbf{x} - \mathbf{e}_i]^+)$. Using the fact that $h(.) \in \mathcal{F}$ and using similar arguments made in the proof of Theorem 1 we can show that the optimal policy is of threshold type. The threshold is given by the following equation,

$$g(x_1) = \min\{x_2 | h(\mathbf{x} - \mathbf{e}_2) \leq h(\mathbf{x} - \mathbf{e}_1)\}.$$

Note that for the average cost criterion, the threshold is stationary and does not depend on time. ∎

V. Conclusion

In this paper we considered a general potentially asymmetric multi-user system which can be used to model many communication systems with shared resources. In such a problem, we studied the problem of optimally allocating the transmission rates from the capacity region in order to minimize an average or discounted cost function. The cost at each time instant, is a function of the number of packets in each queue and for the special case of symmetric linear function results in a delay optimal strategy. In this paper, we showed that the optimal policy in such a setting is of threshold type. Note that the key issue in finding the optimal policy is the following fact. When one queue becomes empty while the other is full, the private server allocated to the empty queue is wasted. In other words, to minimize delay, it becomes important to balance the queues, in the anticipation of such an event in future. This balancing act depends on the expected arrivals in the future, holding cost, and the asymmetry in the capacity region. This intuitive observation is exactly what our threshold structure provides.

Again we emphasize that although in this paper we proved the optimality of a threshold policy, finding these thresholds can be computationally expensive. Also extending these results to more than two users still remains an open problem. The major difficulty in extending this work to more than two users, lies in the properties of discrete convexity for arbitrary dimensions. The set of properties C1-C3 define a convex curve over a three dimensional grid. Additional conditions are required to extend the definition of convex functions to spaces of higher dimension. Unfortunately, other known convexity properties do not necessarily propagate over time as required in an inductive proof such as ours. This limitation was also observed in other problems with different settings such as the problems studied in [17], [23].

Acknowledgment

The authors would like to thank the reviewers for their valuable feedback. The recommendations from the reviewers have significantly helped us to improve the quality of the paper.

References

[1] D. Tse, "Optimal power allocation over parallel gaussian broadcast channels," *Proceedings of International Symposium on Information*, June 1997.

[2] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.

[3] D. Tse and S. Hanly, "Multi-access fading channels: Part i: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2796–2815, November 1998.

[4] L. Li and A. J. Goldsmith, "Outage capacities and optimal power allocation for fading multiple-access channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1326–1347, April 2005.

[5] M. J. Neely, E. Modiano, and C. E. Rohrs, "Power allocation and routing in multibeam satellites with time-varying channels," *IEEE/ACM Tansactions on Networking*, vol. 11, no. 1, pp. 138–152, 2003.

[6] M. J. Neely, E. Modiano, and C. E. Rohrs, "Dynamic power allocation and routing for time-varying wireless networks," *IEEE Journal on Selected Areas in Communications, Special Issue on Wireless Ad-Hoc Networks*, vol. 23, no. 1, pp. 89–103, 2005.

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

[8] E. M. Yeh and A. S. Cohen, "Throughput optimal power and rate control for multiaccess and broadcast communications," *Proceedings of the 2004 International Symposium on Information Theory, Chicago, IL*, p. 112, June 2004.

[9] E. M. Yeh and A. S. Cohen, "Information theory, queueing, and resource allocation in multi-user fading communications," *Proceedings of the 2004 Conference on Information Sciences and Systems, Princeton, NJ*, pp. 1396–1401, March 2004.

[10] G. M. Koole, "Structural results for the control of queueing systems using event-based dynamic programming," *Queueing Systems*, vol. 30, pp. 323–339, 1998.

[11] V. S. Borkar, "Control of markov chains with long-run average cost criterion: the dynamic programming equations," *Siam Journal of Control and Optimization*, vol. 27, no. 3, pp. 642–657, 1989.

[12] P. R. Kumar and P. Varaiya, *Stochastic Systems, Estimation, Identification and Adaptive Control*, Prentice Hall, 1986.

[13] P. Whittle, "Restless bandits: Activity allocation in a changing world," *A Celebration of Applied Probability, ed. J. Gani, Journal of applied probability*, vol. 25A, pp. 287–298, 1988.

[14] R. Weber and G. Weiss, "On an index policy for restless bandits," *Journal of Applied Probability*, vol. 27, pp. 637–648, 1990.

[15] J. Nino-Mora, "Restless bandits, patial conservation laws, and indexability," *Advances in Applied Probability, Vol. 33, no. 1*, pp. 76–98, 2001.

[16] C. H. Papadimitriou and J. N. Tsitsiklis, "The complexity of optimal queueing network control," *Centrum voor Wiskunde en Informatica, BS-R9530*, 1995.

[17] F. J. Beutler and D. Teneketzis, "Routing in queueing networks under imperfect information: Stochastic domain and thresholds," *Stochastics and Stochastic Reports, Vol. 26*, pp. 81–100, 1989.

[18] M. Raissi-Dehkordi and J. S. Baras, "Broadcast scheduling in information delivery systems," *Proc. IEEE GLOBECOM*, 2002.

[19] T. Osogami, M. Harchol-Balter, A. Scheller-Wolf, and Li Zhang, "Exploring threshold-based policies for load sharing," *Fourty second Annual Allerton Conference on Communicatio, Control and Computing*, pp. 1012–1021, October 2004.

[20] O. J. Boxma and D. G. Down, "Dynamic server assignment in a two-queue model," *Mathematics of Operations Research, Vol. 24, No. 2*, pp. 293–305, May 1999.

[21] C. G. Cassandras and S. Lafortune, *Introduction To Discrete Event Systems*, Kluwer Academic Publishers, 1999.

[22] N. Ehsan and M. Liu, "Optimal server allocation in batches," *IEEE Journal on Selected Areas in Communications (JSAC), special issue on Nonlinear Optimization of Communication Systems*, vol. 24, no. 8, pp. 1614–1626, August 2006.

[23] B. Hajek, "Optimal control of two interacting service stations," *IEEE Trans. Auto. Control. AC-29*, pp. 491–499, 1984.

[24] D. P. Bertsekas, *Dynamic Programming, Deterministic And Stochastic Models*, Prentice Hall, 1987.

Navid Ehsan received his B.Sc. degree in 1998 from Sharif University of Technology, Tehran, Iran, and M.Sc. and PhD degrees in 2002 and 2005, respectively, from University of Michigan, Ann Arbor. He is currently with Nextwave Wireless Inc., San Diego. His research interests are in dynamic bandwidth allocation, optimal power allocation and admission control, medium access protocols for wireless systems, security, and stochastic optimization.

Tara Javidi Tara Javidi studied electrical engineering at the Sharif University of Technology from 1992 to 1996. She received her MS degrees in Electrical Engineering: Systems, and Applied Mathematics: Stochastics, from the University of Michigan, Ann Arbor, MI. She received her PhD in EECS from University of Michigan, Ann Arbor in 2002.

From 2002 to 2004, Tara was an assistant professor at the electrical engineering department, University of Washington, Seattle. She joined University of California, San Diego, in 2005, where she is currently an assistant professor of electrical and computer engineering. She was a Barbour Scholar during 1999-2000 academic year and received an NSF CAREER Award in 2004. Her research interests are in communication networks, stochastic resource allocation, and wireless communications.