Optimal Cross Layer Scheduling of Transmissions over a Fading Multiaccess Channel

Munish Goyal, Anurag Kumar, Vinod Sharma
Department of Electrical Communication Engg
Indian Institute of Science, Bangalore, India
Email: goel.munish@gmail.com,{anurag, vinod}@ece.iisc.ernet.in

***Abstract*— We consider the problem of several users transmitting packets to a base station, and study an optimal scheduling formulation involving three communication layers, namely, the medium access control, link and physical layers. We assume Markov models for the packet arrival processes and the channel gain processes. Perfect channel state information is assumed to be available at the transmitter and the receiver. The transmissions are subject to a long-run average transmitter power constraint. The control problem is to assign power and rate dynamically as a function of the fading and the queue lengths so as to minimize a weighted sum of long run average packet transmission delays.**

First, we study the problem for a single user system and obtain structural properties of the optimal policy. We obtain numerical results for the delay-power tradeoff. Then, we consider the multiuser system and obtain a value iteration algorithm for computing the optimal policy. We identify that the problem is computationally intractable and consider an approximation for the cost to go function. The approximating function provides a tight upped bound for the optimal cost function and thus a one-step value iteration could result in a close to optimal policy. A one-step value iteration is then carried out to improve upon the policy. We obtain structural properties of the one-step iterated policy and show that the resulting policy can be obtained via solving a fixed point iteration on a family of suitably modified single user optimal control policies.

Keywords: power and rate control in wireless networks, constrained Markov decision processes

I. Introduction

The dream of "anyone, anywhere, anytime communication" can only be realized by the widespread deployment of high quality wireless access networks. The traditional approach to network architecture is based on a stack of protocol layers with well defined layer functionality and interlayer interfaces. This flexible and transparent approach is mainly responsible for the success of today's wired networks. In the context of wireless networks, however, it has been observed that truly efficient use of the wireless communication resources (spectrum and power) requires adaptability to changing channel and network characteristics in all layers. This leads to a concept called *cross-layer design*, the idea of joint optimization at two or more of the layers of communication. Such an approach can help address the unique challenges of the wireless environment such as the time-varying or fading nature of wireless channels, and the limited battery energy available in wireless handsets.

One of the main ideas of cross layer design is to permit the exchange of information across layers, something that would be considered a "layer violation" in traditional design. This additional information is used by the layers to better adapt to varying transmission conditions. In this work, we will concentrate on cross layer design problems involving three of the wireless communication system layers: the data link layer, the medium access control (MAC) layer and the physical layer. The approach combines fundamental communication limits of physical layer capacity (captured via the information theoretic channel capacity), with a higher layer quality of service measure, namely, long run average packet delay. We consider random packet arrivals and queueing at the transmitters and incorporate the effect of temporal variations in the channel. The physical layer constraint is the average transmission energy available (which relates to battery life). The control variables are the amount of data released from the link layer and the transmission power used at the physical layer. An information theory based analysis helps in obtaining the limits of what could possibly be achieved using an efficient channel coding-decoding scheme, and also provides insight into good rate and power control policies.

Background Literature: In [18] Tse and Hanly considered a resource allocation problem for a multiaccess fading channel with the objective of maximizing the throughput capacity, and in the sequel [10], they went on to discuss the capacity region when users need delay guarantees. They considered a framework in which delay guarantees are achieved if each user transmits at a fixed guaranteed rate. This is a restrictive assumption and can lead to the wastage of resources. Collins and Cruz [7] considered a rate and power scheduling problem for a point-to-point wireless system with the objective of minimizing average transmitter power subject to an average transmission delay constraint. They assumed that the received power is always constant. Our research has been in the spirit of Berry and Gallager [3] who considered a problem similar to the one considered in [7] but without the constant received power assumption. Berry and Gallager [3] obtain structural results exhibiting a tradeoff between the optimal transmitter power and the mean queueing delay. They show that the optimal power versus the optimal delay curve is convex, and as the average power available for transmission increases, the achievable mean delay decreases. They also provide some structural results for the optimal policy that achieves any point on the power-delay curve. In [2], Berry further considered the wireless multiaccess fading system with the objective of minimizing a weighted sum of average packet transmission delays and transmission power. The author obtained a class of

simple rate and power scheduling policies that were shown to be nearly optimal when the average delay constraint is large. A recent survey paper [4] discusses fundamental limits of cross-layer design algorithms for multiaccess wireless networks. The survey paper contains a good list of references on cross-layer design problems in wireless networks. The work presented in this paper is an extension of our earlier work reported in [12] on power and delay optimal transmission policies for a wireless link. In this paper we extend the work in [12] to the Markov arrival and fading settings and also to the multiuser case.

Contributions: In this paper, we provide several new results on the single user problem introduced in [3] and then we develop the multiuser problem. We cast the single user problem as a constrained Markov decision process and draw upon results from average cost Markov decision theory to establish detailed structural results for the optimal policy in the single user case. Our approach permits us to go beyond the results in [3] in the following ways.

1) We obtain a complete structural characterization of the policy in the i.i.d. arrival and fading case. Further new structural results are obtained in the Markov arrivals and fading case in Theorem 3.2.
2) In Section III-E, we utilise a relative value iteration algorithm to obtain numerical results for the average cost problem.
3) We then develop the constrained Markov decision problem approach for the multiuser case. In Section IV, we approximate the cost to go function by replacing it with an additive separable function that tightly upper bounds the original cost to go function. A one-step value iteration is then carried out to obtain an improved policy which could be close to optimal since the cost to go function tightly upper bounds the original cost to go function.
4) We obtain structural properties of the one-step iterated policy and show that the policy is obtained essentially via a fixed point iteration on a family of single user control policies.
5) In a special case of on-off control, the control policy of a tagged user, given the control decisions of all the other users, is shown to possess a simple threshold form.

Paper Outline: This paper is organized as follows. In Section II, we discuss the model of the system under consideration and formulate the controller objectives as a constrained optimization problem. In Section III, we analyse the single user system with a mean delay objective. We use a result from [15] to convert the single user problem into a family of unconstrained optimization problems. This unconstrained problem is a Markov decision problem (MDP) with the average cost criterion. We show the existence of stationary average cost optimal policies and their structural properties in Section III-B. A corresponding discounted cost problem is studied in Appendix C. In Section III-C, we obtain conditions for the existence of a Lagrange multiplier such that the optimal policy corresponding to that value for the multiplier is also optimal for the original constrained MDP. We provide numerical results for the optimal policy and obtain the power-delay tradeoff curve in Section III-E.

We analyse the M user mean delay minimization problem in Section IV. We observe that the general problem is too complicated to work with. We approximate the cost to go function with an additive separable function and carry out a one-step value iteration to derive the control policy. The proofs of theorems are given in Appendix D.

II. The System Model

We consider a discrete time model of a multiaccess fading channel with M users communicating to a receiver as shown in Figure 1. Time is divided into slots of length τ units each. Let N be the number of channel uses per slot. Packets generated at the higher layer arrive at the link layer at the end of every slot as shown in the top part of Figure 1 (the process $A[n]$). The link layer is modeled as a queue of infinite capacity where the arriving packets are kept before forwarding to the medium access control (MAC) layer. A controller co-located at the receiver decides about the number of packets to be forwarded to the MAC for transmission in any given slot. The decision is based on the number of packets buffered at each user's queue and the channel gain (attenuation) for each transmit-receive pair. The queue length information is communicated through the header of the last packet transmitted in a slot whereas the channel gain for each transmit-receive pair is measured at the receiver. The control decisions are communicated to the transmitters during a control period at the beginning of each slot (see Figure 1). In practice, WIMAX based on an orthogonal frequency division multiple access (OFDMA) scheme provides such a mechanism for exchange of information between a centralized controller and spatially distributed transmitters. The physical layer has the responsibility of transmitting bits on the multipath fading wireless channel. The transmission power is constrained by battery life.

We use bold symbols to represent vectors of length M (the number of users); i.e. $\mathbf{x}$ represents $\{x_1, x_2, \cdots, x_M\}$. Each source generates fixed size packets (each of length b bits) according to a finite state ergodic Markov chain $\mathbf{A}[n]$; let P_a be the transition probability matrix. At time $n\tau$, let $\mathbf{Q}[n]$ be the queue length and $\mathbf{R}[n]$ be the number of packets to be released in the current slot as per the control decision. The evolution equation of the buffer length process is given by

$$\mathbf{Q}[n+1] = (\mathbf{Q}[n] - \mathbf{R}[n])^+ + \mathbf{A}[n], \tag{1}$$

where $(x)^+$ is a notation for $\max\{x, 0\}$.

We assume a long run average transmitter power constraint of $\bar{P}$ at each transmitter. Let the transmitted signal from the i^{th} transmitter be $X_i[n]$. Let $H_i[n]$ be the fading process seen by the i^{th} user's transmission and $Z[n]$ be an additive white Gaussian noise (receiver noise) process with zero mean and variance σ^2. The signal received at the receiver is then

$$Y[n] = \sum_{i=1}^{M} \sqrt{H_i[n]} X_i[n] + Z[n].$$

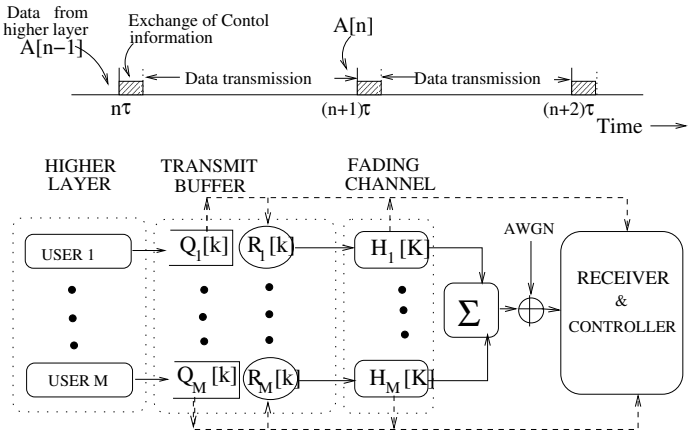

Fig. 1. System model for an M user multiaccess fading channel. The controller, based on the channel gain vector $\mathbf{h}[k]$ and the queue length vector $\mathbf{q}[k]$, decides upon $r_i[k]$, the number of packets to be transmitted during slot k and $p_i[k]$, the transmitter power during slot k from the i^{th} user.

We assume that the channel state information is available perfectly at both the transmitter as well as the receiver end. We model the fading process as block-fading where the channel gain (fade) stays constant over the duration of a slot, i.e, for τ time units. The channel gain process $H_i[n]$, embedded at the slot boundaries, is assumed to be a finite state ergodic Markov chain; let P_h be the transition probability matrix. We assume, for each n, $(A_1[n], H_1[n], A_1[n], H_1[n], \ldots, A_M[n], H_M[n])$ are mutually independent random variables. We will make an additional assumption that the fading is bounded away from zero and takes values in a finite subset of $[h_0, 1]$ where $h_0 > 0$.

We address the optimal power and rate allocation problem for the M user multiaccess system with the objective of minimizing mean packet transmission delay subject to an average power constraint. The receiver acts as a central controller, which, depending upon the transmitter buffer lengths and the channel gains of each user, allocates the packet transmission rates $\mathbf{R}[n]$ and powers $\mathbf{P}[n]$ to individual users. Based on the control decisions, the controller informs the physical layer as to what transmission rate and power to use. The physical layer then encodes the data coming from the MAC layer at that rate and transmits the encoded data over the channel at the scheduled power level.

The random processes $\mathbf{R}[k], \mathbf{Q}[k], \mathbf{P}[k]$ correspond to packet transmission rate vector, queue length vector and power allocation vector respectively. We have the following scheduling constraints. First, there is the natural constraint that $\mathbf{R}[k] \leq \mathbf{Q}[k]$, where the representation $\mathbf{R}[k] \leq \mathbf{Q}[k]$ means $R_i[k] \in \{0, 1, \cdots, Q_i[k]\}$ for $i \in \{1, 2, \cdots, M\}$. Further, we will have the multiple users capacity constraint in each slot, i.e., the allowed values of $\mathbf{R}[k]$ given a power allocation vector $\mathbf{P}[k]$. The capacity of a multiaccess system depends upon the decoding scheme employed at the receiver. Given a power vector $\mathbf{p}$ and a channel gain vector $\mathbf{h}$, we assume that the maximum number of packets $\mathbf{r}$ that can be transmitted "reliably" for a system that employs successive decoding should satisfy the capacity constraint $\mathbf{r} \in C_g(\mathbf{h}, \mathbf{p})$ where $C_g(\mathbf{h}, \mathbf{p})$ is the set of rate vectors satisfying,

$$\sum_{j \in S} r_j \leq \frac{1}{\theta} \ln \left(1 + \frac{\sum_{j \in S} h_j p_j}{\sigma^2} \right) \tag{2}$$

for every $S \subset \{1, 2, \cdots, M\}$ and $\theta := \frac{2 \ln(2) b}{N}$.

Remark 2.1: Note that, by our model, a codeword can at the most stretch to a slot to ensure decoding at the end of every slot. This assumption of finite length code would result in a strictly positive decoding error probability [8]. Since the capacity function is averaged only over Gaussian noise, a relatively short code block lengths are required to approximate the asymptotic capacity results. Under an assumption that the coding/scheduling frame time is shorter than the channel coherence time and the number of channel uses in this time are large enough for reliable communication, the capacity formula could be a reasonable approximation for the purpose of exploring this problem [3]. Further, We need to pay a marginal rate or a power penalty to achieve a target decoding error probability while using codewords of length N (channel uses per slot). The penalty factors could also be incorporated easily by using the error exponent bounds on the probability of error [8]. ■

The Multiuser Problem: At time $n\tau$, the state of the system is represented by $\mathbf{X}[n] := (\mathbf{Q}[n], \mathbf{H}[n], \mathbf{A}[n])$. Recall that the process $\mathbf{Q}[n]$ is the queue length process at time instant $n\tau$. At the n^{th} decision instant (time instant $n\tau$, $n \geq 0$), the controller decides upon the number of packets $\mathbf{R}[n]$ to be transmitted in the current slot and $\mathbf{P}[n]$, the transmitter power required for reliable transmission depending on the entire history of state evolution, i.e., $\mathbf{X}[k]$ for $k = \{0, 1, 2, \cdots, n\}$ that minimizes a weighted sum of a long run average delay subject to the average power constraint $\bar{\mathbf{P}}$. Now since delay is related to the amount of data in the buffer by Little's formula [20], the control objective is equivalent to minimizing a weighted sum of the average queue lengths. The controller's objective is to obtain an optimal sequence of pairs $(\mathbf{R}[n], \mathbf{P}[n])$ that solves the following optimization problem.

$$\min \left\{ \limsup_n \frac{1}{n} \mathbb{E} \sum_{k=0}^{n-1} \sum_{i=1}^{M} \omega_i Q_i[k] \right\}$$

subject to

$$\mathbf{R}[k] \leq \mathbf{Q}[k] \text{ and } \mathbf{R}[k] \in C_g(\mathbf{H}[k], \mathbf{P}[k]); \text{ for } k \geq 0$$

$$\limsup_n \frac{1}{n} \mathbb{E} \left[\sum_{k=0}^{n-1} \mathbf{P}[k] \right] \leq \bar{\mathbf{P}}, \tag{3}$$

where w_i are nonnegative weights associated with user i and define the relative importance of user i over other users.

The Single User Problem: We first analyse the single user system with one transmitter and one receiver (controller) and then specialize the results obtained to the multiuser scenario employing successive decoding at the receiver. The single user model is shown in Figure 2.

In the single user setting, the power, p, required for reliable transmission of r packets in a slot gets fixed as $p = \frac{\sigma^2}{h}(e^{\theta r} - 1)$. Thus the control objective is to obtain the sequence $R[n]$ as a function of $\{\mathbf{X}[0], \mathbf{X}[1], \cdots, \mathbf{X}[n]\}$, that solves the following

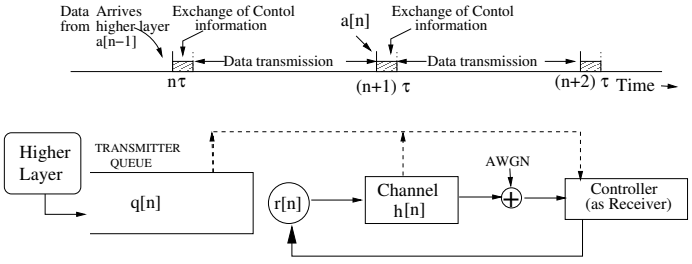

Fig. 2. Model of a single user point to point wireless link.

optimization problem.

$$\min \left\{ \limsup_n \frac{1}{n} \mathbb{E} \sum_{k=0}^{n-1} Q[k] \right\}$$
$$\text{subject to } R[k] \leq Q[k] \text{ for } k \geq 0$$
$$\text{and } \limsup_n \frac{1}{n} \mathbb{E} \left[\sum_{k=0}^{n-1} \frac{\sigma^2}{H[k]} (e^{\theta R[k]} - 1) \right] \leq \bar{P}. \quad (4)$$

III. The Single User System

The single user control problem, as stated in Equation 4, is a constrained dynamic optimization problem. We first convert it into a family of unconstrained optimization problems and analyse them. The unconstrained optimization problems belong to a category of average cost Markov decision problems (MDP). We characterize the optimal policies for these MDPs. We then show how these policies result in a solution to the original constrained problem.

A. Formulation as a Markov Decision Process (MDP)

Let $\{X[n], n \in \{0, 1, 2, \cdots\}\}$ denote a controlled Markov chain, with state space $\mathcal{X} = \mathbb{Z}^+ \times (h_0, 1] \times \mathbb{Z}^+$, and action space $\mathbb{Z}^+$, where $\mathbb{Z}^+$ denotes the set of nonnegative integers. The set of feasible actions r in state $x = (q, h, a)$ is the set of all integers belonging to $\mathcal{R}(x) = \{0, 1, \cdots, q\}$. Let $\mathcal{K}$ be the set of all feasible state-action pairs. The transition kernel on $\mathcal{X}$ given an element $(x, r) \in \mathcal{K}$ is denoted by Γ. Define the mapping $p : \mathcal{K} \to \mathbb{R}^+$ by $p(x, r) = \frac{\sigma^2}{h}(e^{\theta r} - 1)$, the power required to transmit r in a slot with $\theta = \frac{2 \ln(2) b}{N}$.

Define a policy $\pi = (\pi_0, \pi_1, \pi_2, \cdots)$ that at time instant n generates an action $r[n]$ depending upon the entire history of the process, i.e., at decision instant $n \in \{0, 1, 2, \cdots\}$, π_n is a mapping from $\mathcal{K}^n \times \mathcal{X}$ to $\mathcal{R}(X[n])$. Let Π be the space of all such policies. A stationary policy is of the form $\pi = (f, f, f, \cdots)$ where f is a measurable mapping from $\mathcal{X}$ to $\mathcal{R}(X[n])$. For a policy $\pi \in \Pi$, and initial state $x \in \mathcal{X}$, we define two cost functions B_x^π, the buffer cost, and K_x^π, the power cost by,

$$B_x^\pi = \limsup_n \frac{1}{n} \mathbb{E}_x^\pi \sum_{k=0}^{n-1} Q[k];$$

$$K_x^\pi = \limsup_n \frac{1}{n} \mathbb{E}_x^\pi \sum_{k=0}^{n-1} p(X[k], R[k]).$$

Given the power constraint $\bar{P} > 0$, denote by $\Pi_{\bar{P}}$ the set of all admissible control policies $\pi \in \Pi$ which satisfy the long run transmitter power constraint $K_x^\pi \leq \bar{P}$. Then the controller objective can be restated as a constrained optimization problem (CP) defined as,

$$(CP): \quad \text{Minimize} \quad B_x^\pi \quad \text{subject to} \quad \pi \in \Pi_{\bar{P}} \quad (5)$$

The problem (CP) can be converted into a family of unconstrained optimization problems through a Lagrangian relaxation [15]. For every $\beta > 0$, define a mapping $c_\beta : \mathcal{K} \to \mathcal{R}^+$ as $c_\beta(x, r) = q + \beta p(x, r)$. Define a corresponding functional for any policy $\pi \in \Pi$ by,

$$J_\beta^\pi(x) = \limsup_n \frac{1}{n} \mathbb{E}_x^\pi \sum_{k=0}^{n-1} c_\beta(X[k], R[k]).$$

Given $\beta > 0$, define the unconstrained problem (UP_β)

$$(UP_\beta): \quad \text{Minimize} \quad J_\beta^\pi(x) \quad \text{subject to} \quad \pi \in \Pi \quad (6)$$

The following theorem gives sufficient conditions under which an optimal policy for an unconstrained problem is also optimal for the original constrained control problem (CP).

Theorem 3.1: [15] Let, for some $\beta > 0$, $\pi^* \in \Pi$ be the policy that solves the unconstrained problem UP_β such that π^* yields the expressions B^{π^*} and K^{π^*} as limits for all $x \in \mathcal{X}$, and in addition, for all x, $K^{\pi^*} = \bar{P}$. Then the policy π^* is optimal for the constrained problem (CP).

Proof: See [15] ■

We analyse the problem (UP_β) in Section III-B. We verify that the conditions stated in the hypothesis of the Theorem 3.1 are valid and thus obtain the constrained solution in Section III-C.

B. Structure of the Optimal Policy for UP_β

The problem (UP_β) is a standard Markov decision problem with an average cost criterion. We now study the unconstrained problem and obtain the structural properties of the optimal policy. *As $\beta > 0$ is fixed for the analysis in this section*, for notational simplicity *we suppress the subscript β*. Define a discounted cost MDP with discount factor $\alpha \in (0, 1)$ corresponding to the problem (UP), for each initial state $x = (q, h, a)$, with value function,

$$V_\alpha(q, h, a) = \min_{\pi \in \Pi} \mathbb{E}_x^\pi \Big[\sum_{k=0}^{\infty} \alpha^k \left(Q[k] + \beta p(X[k], R[k]) \right) \Big]. \quad (7)$$

We call the optimal solution for the discounted problem a discount optimal policy. The following lemma states that the average cost problem can be studied as a limit of discounted cost problems as the discount factor α increases to one and also proves its existence.

Lemma 3.1: There exists a stationary deterministic policy $r(q, h, a)$ that solves the unconstrained problem UP for each $\beta > 0$. The stationary optimal policy that solves the unconstrained problem UP is limit discount optimal in the sense that the policy can be obtained as a limit of discount optimal policies as the discount factor increases to one.

Proof: (See Appendix B) ■

First, we study the discounted cost problem and obtain structural properties of the discount optimal policy (see Appendix C). In light of the result of Lemma 3.1, using the structural properties of the discount optimal policies, we obtain structural properties of $r(q,h,a)$ for each (q,h,a). For a state-action pair $(x=(q,h,a),r)$, define $u := q-r, u \in \{0,1,\cdots,q\}$, as the number of packets *not served* when the system is in state x. Thus $u(q,h,a) = q - r(q,h,a)$ also defines the optimal stationary policy for the single user problem. We note that the value function $V_\alpha(q,h,a)$ is convex in q (refer Lemma C-1 in Appendix C). Define a differential of the value function as $G_\alpha(q,h,a) = V_\alpha(q,h,a) - V_\alpha(q-1,h,a)$ and

$$Z(q,h,a) = e^{\theta q} \lim_{\alpha \to 1} \alpha \mathbb{E}_{h,a}[G_\alpha(q+A,H,A)],$$

where $\mathbb{E}_{h,a}[\cdot]$ denotes expectation with respect to the transition probability of H and A with initial state h and a respectively.

Note that $Z(q,h,a)$ is monotone increasing in q as the value function $V_\alpha(\cdot)$ is convex in q. Define a function $u^*(q,h,a)$ as the value of u that solves the following inequalities, for given (q,h,a),

$$Z(u,h,a) \le \frac{\beta\sigma^2}{h} e^{\theta q}(e^\theta - 1) \le Z(u+1,h,a). \quad (8)$$

The optimal policy $u(q,h,a)$ equals $\min\{u^*(q,h,a), q\}$. In section III-D, we state an algorithm to compute $u^*(q,h,a)$ and hence the optimal policy using the above relation.

The following theorem gives the structural properties of the optimal policy.

Theorem 3.2: Structural properties of the stationary optimal policy for (UP):

i) The optimal policy $u(q,h,a)$ is monotonic nondecreasing in q and $r(q,h,a) = q - u(q,h,a)$ is monotonic nondecreasing in q as well.

ii) The function $u^*(q,h,a)$ is bounded below by,

$$\frac{1}{\theta}\ln\left(\frac{\sqrt{1+4\frac{\beta^2\sigma^4}{h}\eta(h,a)e^{\theta q}(e^\theta-1)}-1}{2\beta\sigma^2\eta(h,a)}\right) - 1$$

and bounded above by,

$$\frac{1}{\theta}\ln\left(\frac{\beta\sigma^2}{h}e^{\theta q}(e^\theta-1)\right),$$

where $\eta(h,a) = \mathbb{E}_{h,a}\left[\frac{e^{\theta A}}{H}\right]$ and $\theta = \frac{2\ln(2)b}{N}$.

iii) The optimal number of packets transmitted $r(q,h,a)$ increases to infinity as q increases to infinity.

iv) The optimal solution $u(q,h,a)$ is monotone nondecreasing with β (the power price).

v) If the fading and the arrival processes are i.i.d:

a) The function $Z(q,h,a)$ is only a function of q and $W(y)$ defined as the value of u that solves $Z(u) \le e^{\theta y} \le Z(u+1)$ is a monotone nondecreasing in y.

b) Given any (q,h), the optimal policy $u(q,h)$, is

$$u(q,h) = \min\left\{q, W\left(q - \frac{1}{\theta}\ln\left(\frac{h/\beta\sigma^2}{(e^\theta-1)}\right)\right)\right\}.$$

c) The optimal solution is (see Figure 4):

$$u(q,h) = \begin{cases} 0 & \text{if} \quad \frac{e^{\theta q}}{h} < \frac{Z(0)}{\beta\sigma^2(e^\theta-1)} \\ q & \text{if} \quad \frac{e^{\theta q}}{h} > \frac{Z(q)}{\beta\sigma^2(e^\theta-1)} \\ u^*(q,h) & \text{o.w.} \end{cases}$$

Proof: Let $x=(q,h,a)$. Denote by $u_\alpha(x)$, the discount optimal policy. Theorem A-2 states that the average cost optimal policy $u(x)$ is the limit of discount optimal policies which might be optimal for some close neighbourhood of x rather than x itself. Lemma D-1 provides a stronger result that $u(x)$ is limit of discount optimal policies $u_\alpha(x)$ as discount α increases to one. The results of the theorem now follow from the analysis of the discounted cost problem (see Appendix C).

i) Follows from Theorem C-1 and Theorem C-2.
ii) Follows from Theorem C-3
iii) Follows from Lemma C-2
iv) Follows from Theorem C-5
v) Follows from Theorem C-6

■

Figure 3 depicts the structure of the optimal policy depicted in terms of the unsent data $u(q,h,a)$.

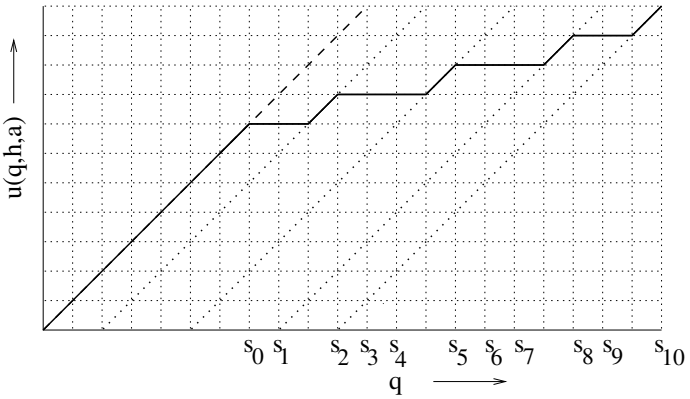

Fig. 3. The structure of the optimal amount of buffer not served $u(q,h,a)$ versus q, for a fixed h and a. The dark curve plots a typical policy. Define $s_i(h,a) := \max\{q : u^*(q,h,a) \ge q - i\}$. For $q \le s_0$, no packets are transmitted. The number of packets transmitted for $q \in (s_i+1, s_{i+1})$ is $i+1$. Thus, for $q \in (s_1+1, s_2)$, it is optimal to serve two packets.

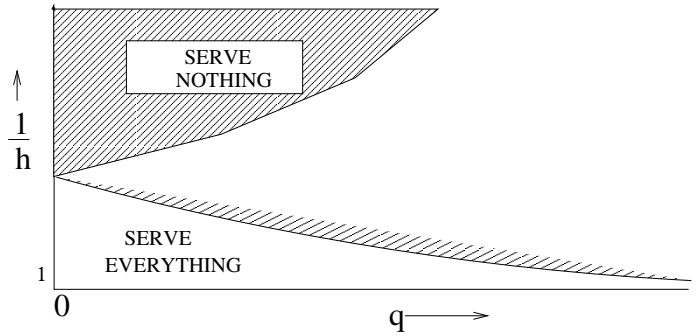

Fig. 4. Depiction of the optimal policy for the scenario when the channel gain process and the arrival processes are i.i.d.

Discussion: The structural properties of the average optimal policy stated in Theorem 3.2 goes beyond the results stated in [2] and [3].

i) Given (h, a), it follows from Theorem 3.2(i) that for any q_1 and q_2 satisfying $q_1 < q_2$, we have $q_1 - u(q_1) \leq q_2 - u(q_2)$, i.e., $u(q_2) - u(q_1) \leq q_2 - q_1$ and $u(q_1) \leq u(q_2)$. This implies that the number of packets transmitted $r(q, h, a)$ grows at a rate slower than q. These characteristics of the optimal policy are shown in Figure 3.

ii) The lower bound for the optimal policy is especially useful in the sense that it provides information about the rate of growth of transmission rate with the queue length.

iii) The transmission rate does not saturate to a level as the queue length increases to infinity. The larger the queue the larger the number of packets transmitted.

iv) Given (q, h, a), the optimal number of packets transmitted $r(q, h, a)$ is nonincreasing in β. This is natural to expect since the larger the power price β, the higher is the transmission cost.

v) The following observations can be made for the i.i.d. case,

   a) it is optimal not to serve anything when the channel is bad (i.e., $\frac{1}{h}$ is large); for each q there is a small enough h such that for channel pairs worse than this, it is optimal not to serve.
   b) when the channel is good, it is optimal to serve everything until a value of the buffer size q that increases with increasing h;
   c) even under poor channel conditions, as q increases it becomes optimal to serve data as the delay becomes costlier than power.

Note that Figure 3 would represent the policy for a small value of h, with nothing being served until $q \geq s_0$.

C. The Power Constrained Delay Optimal Policy

We have given structural results for the optimal policy for the unconstrained problem (UP_β). Now (invoking Theorem 3.1) we show that there exists a $\beta > 0$ for which the optimal policy obtained above is also optimal for the constrained problem (CP). We reintroduce the dependence on the multiplier β. Recollect that the solution to the problem (UP_β) is $r_\beta(q, h, a) = q - u_\beta(q, h, a)$. We show that the conditions under which Theorem 3.1 holds are satisfied. First, we need to show that for each $\beta > 0$, the $\limsup$ and $\liminf$ are equal. This is true if the controlled chain is ergodic as it would imply $u_\beta(q, h, a)$ yields the expressions B^{u_β} and K^{u_β} as limits. Theorem 3.2(ii) states that the optimal number of packets transmitted increases to infinity as q tends to infinity. By a standard drift argument, it is easy to show [12] that the system is ergodic for all finite arrival rates. This is natural to expect as the relaxed problem is an unconstrained system and one can carry any arrival rate by spending more and more power.

Next, we need to show the existence of a β such that the average power cost is equal to the power constraint, i.e., $K^{u_\beta} = \bar{P}$. We know that the policy $r^{st}(h) = \frac{1}{\theta}\left(\ln\left(\frac{h}{\lambda\sigma^2}\right)\right)^+$ is stabilising [9], where λ solves for the power constraint $\bar{P}$. Define

$$\mathbb{P}^{st}(R) = \min\left\{\mathbb{E}\left(\frac{\sigma^2}{H}\left(e^{\theta r^{st}(H)} - 1\right)\right) : \mathbb{E}[r^{st}(H)] \geq R\right\},$$

$$\mathbb{R}^{st}(P) = \max\left\{\mathbb{E}[r^{st}(H)] : \mathbb{E}\left(\frac{\sigma^2}{H}\left(e^{\theta r^{st}(H)} - 1\right)\right) \leq P\right\},$$

If the mean arrival rate $\mathbb{E}[A] < \mathbb{R}^{st}(\bar{P})$, the mean queue length is finite under the stabilizing policy. The mean delay increases as β increases since the power gets costlier whereas the power cost decreases to $\mathbb{P}^{st}(\mathbb{E}[A])$ as β increases to infinity. Note that $\mathbb{P}^{st}(\mathbb{E}[A])$ is the minimum power required to keep the queue stable for an arrival rate $\mathbb{E}[A]$ and we need $\mathbb{P}^{st}(\mathbb{E}[A]) < \bar{P}$. It has been shown [3] that the average power is monotone nonincreasing convex function of mean delay. Further, it is easy to see that the average transmitter power required is monotone nonincreasing in β and converges to $\mathbb{P}^{st}(\mathbb{E}[A])$ as $\beta \to \infty$. If this power cost function is continuous in β, there always exists a $\beta > 0$ such that the average power cost for the optimal policy corresponding to that β equals $\bar{P}$. But a monotone function may have jump discontinuities and thus there may not be a value of β for which the average power constraint is satisfied with equality. This is a very standard situation that even arises in knapsack packing problems.

In case there is no β for which the average power constraint is satisfied with equality, we have two possible solutions. The first one is to change the power constraint itself by choosing one of the nearest (usually a lower one) number for which there is a β satisfying the average power cost with equality and say that there is no advantage in having a constraint value larger than that number. Thus the hypothesis of Theorem 3.1 would be satisfied and we have an optimal solution for the constrained problem. The second approach is to define a randomized policy. Since there always exist a β for which the power cost $K^{u_\beta} < \bar{P}$ because otherwise it contradicts the existence of stabilizing policy. Thus define β_0 as the smallest value of β for which $K^{u_\beta} \leq \bar{P}$. If the equality holds then we are done. But due to possibility of a discontinuity at β_0, $K^{u_{\beta_0^+}} < \bar{P}$ and $K^{u_{\beta_0^-}} > \bar{P}$. Define a new randomized policy that randomizes between $u_{\beta_0^+}$ and $u_{\beta_0^-}$ and the probabilities are chosen so that the power constraint is met with equality. This randomized policy defines a constrained solution.

Remark 3.1: The monotonic nature of optimal delay and optimal power usage with respect to beta yields a simple iterative algorithm to compute an appropriate choice for beta that satisfies the average power constraint (or delay constraint). Start with an arbitrary choice of β such that $\beta > 0$ and compute the optimal policy and the long run average power required. If the average power required is more(less) than the constraint, decrease(increase) the value of beta and recompute. Repeat till we converge to a value of β where monotonicity property guarantees the convergence of this iteration. If there is a discontinuity, a randomized policy needs to be considered as discussed and explained above.

D. An Algorithm for Computing $u^*(\cdot)$

In order to compute $u^*(q, h, a)$, as per the definition in Equation 19 we need to compute $Z(q, h, a)$. To compute

$Z(q,h,a)$ we need an algorithm to compute $V_\alpha(q,h,a)$ as defined by Equation 7 for each $\alpha \in (0,1)$. Consider the following iterative algorithm to compute $V_\alpha(q,h,a)$. We suppress the subscript α. For $n \geq 0$,

$$V_n(q,h,a) = \min_{u\in\{0,1,\cdots,q\}} \left\{ q + \frac{\beta\sigma^2}{h}\left(e^{\theta(q-u)}-1\right) + \alpha\mathbb{E}_{h,a}[V_{n-1}(u+A,H,A)]\right\}, \quad (9)$$

with $V_0(q,h,a)=0$. Let $G_n(q,h,a) = V_n(q,h,a) - V_n(q-1,h,a)$ and $Z_n(q,h,a) = e^{\theta q}\mathbb{E}_{h,a}[G_n(q+A,H,A)]$. Define $u_n^*(q,h,a)$ be the value of u that solves the following inequalities,

$$\alpha Z_n(u,h,a) \leq \frac{\beta\sigma^2}{h}e^{\theta q}(e^\theta - 1) \leq \alpha Z_n(u+1,h,a).$$

Note that $Z(q,h,a) = \lim_{\alpha\to 1, n\to\infty} Z_n(q,h,a)$. Define $s_{(i,n)}(h,a) = \max\{q : u_n^*(q,h,a) \geq q-i\}$. Thus based on the constrained solution as defined earlier, given a value of (h,a), we have,

- For $q \leq s_{(0,n)}$ (no packet transmission ($r(\cdot)=0$)),
$$G_{n+1}(q,h,a) = 1 + \alpha\mathbb{E}_{h,a}[G_n(q+A,H,A)]$$
- For $q = s_{(i,n)}+1$ and $i \in \{0,1,2,\cdots\}$ implying that the number of packets transmitted at q is one larger than those transmitted for a queue length of $q-1$ (See Figure 3),
$$G_{n+1}(q,h,a) = 1 + \frac{\beta\sigma^2}{h}\left(e^{\theta(i+1)} - e^{\theta(i)}\right)$$
- For $q \in \{s_{(i,n)}+2, \cdots, s_{(i+1,n)}\}$ and $i \in \{0,1,2,\cdots\}$ implying that given i, the number of packets transmitted for this range of queue lengths is the same,
$$G_{n+1}(q,h,a) = 1 + \alpha\mathbb{E}_{h,a}[G_n(u_n^*(q,h,a)+A,H,A)]$$
- Further by definition,
$$Z_{n+1}(q,h,a) = e^{\theta q}\mathbb{E}_{h,a}[G_{n+1}(q+A,H,A)]$$

The sequence $u_n^*(q,h,a)$ converges to the optimal solution $u^*(q,h,a)$ in the limit as n tends to ∞ followed by the limit as α increases to 1.

E. Numerical Evaluation

We numerically evaluate the optimal policy and the power-delay trade-off as the multiplier β is varied. We will use an average cost value iteration algorithm for numerical computation. The value iteration algorithm is similar to the iteration in Equation 9 with α set to 1. One expects that if $V_n(x) - V_{n-1}(x)$ converges and is independent of x then the limiting policy is the average cost optimal and the limiting difference would be the average cost. The convergence of such an algorithm is known to be considerably difficult to analyse. Chen and Meyn [6] have given sufficient condition for the convergence of the value iteration algorithm for problems arising in queueing networks. The essential idea is to initialize the algorithm with a value function corresponding to a stable policy, i.e., $V_0(x)$ needs to be appropriately chosen. They give counterexamples to show that the value iteration if initialized with $V_0(x)=0$ may never converge.

Theorem 3.3: Suppose (h,a) assumes only finitely many values and there exists a pair (h_0,a_0) among possible pairs of (h,a) for which the transition probability matrix for h and a has a positive entry at the diagonal. Let there be a positive probability of arrivals a being equal to 0. Further, suppose there exists a t such that for any given $\bar{\eta} > 0$ and starting the dynamic system in state $q \in \{0,\cdots,\bar{\eta}\}$ and in any (h,a), we can reach $z=(0,h_0,a_0)$ at time t with a positive probability. Then the value iteration algorithm if initialized with $V_0(x)$ (the cost corresponding to a policy $r(q,h,a)=q$) results in convergence and the limiting policy is optimal. Further, $V_n(z) - V_{n-1}(z)$ converges to the optimal average cost.

Proof: See Appendix D ■

We consider the following numerical example. The ambient noise power $\sigma^2 = 1$. The number of channel uses per slot is $N=10$. The channel gain process is assumed to be i.i.d. and h take values in the set $\{.4,.7,1\}$ with probabilities $\{.3,.4,.3\}$ respectively. The packet arrival process is also assumed to be i.i.d. and takes values $\{0,100\}$ with probabilities $\{.5,.5\}$ respectively.

In order to verfiy the convergence of the numerical algorithm, we note that the hypothesis of Theorem 3.3 is satisfied. Since the arrival process is i.i.d., the optimal control related functions are independent of variable a. The value function $V_0(q,h)$ is $q + \frac{\beta\sigma^2}{h}(e^{\theta q}-1) + \mathbb{E}V_0[A,H]$ with $G_0(q,h) = V_0(q,h) - V_0(q-1,h) = 1 + \frac{\beta\sigma^2}{h}e^{\theta q}(1-e^{-\theta})$. Set state $z=(0,1)$ and $V_{n+1}(z)$ is computed from $G_n(q,h)$ as follows.

$$V_{n+1}(0,h) = \mathbb{E}[V_n(A,H)] = \mathbb{E}[\sum_{k=1}^{A} G_n(k,H) + V_n(0,H)].$$

The optimal average cost is thus the limit of $V_{n+1}(z) - V_n(z)$ as n tends to infinity.

The plot for optimal number of packets transmitted versus the queue length for various values of β and channel gain equal to 1 are shown in Figure 5. The optimal policy $r(q,h)$ for other values of h is $r\left(\left(q - \frac{1}{\theta}\ln\frac{1}{h}\right)^+\right)$. The power and delay versus β is shown in Figure 6 and the power-delay tradeoff curve is shown in Figure 7. Recall that the mean queue length is proportional to the mean transmission delay.

IV. The Multiaccess System

We now consider the M user delay minimization problem with an average power constraint (Equation 3). Based on the structural properties for the single user system, we wish to obtain similar structural results for the multiaccess system. We convert the problem into a family of unconstrained problems (see [15]) by associating multipliers $\lambda_i, i \in \{1,2,\cdots,M\}$ with the average power constraints. The controller objective is to allocate optimal rate vector $\mathbf{R}[n]$ and a power vector $\mathbf{P}[n]$ given the state $\mathbf{X}[n] = (\mathbf{Q}[n], \mathbf{H}[n], \mathbf{A}[n])$ while minimizing a weighted sum of the average delay and the average power.

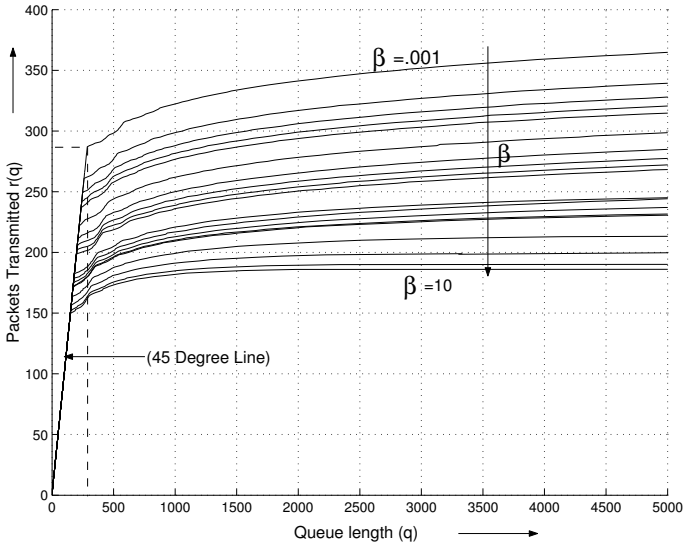

Fig. 5. The optimal number of packets transmitted $r(q)$ versus q for various values of the Lagrange multiplier β (power price) for the numerical example in Section III-E. Note that $r(q) = q$ for lower values of q and this 45 degree line is shown. The values of β chosen are $((2n \ mod \ 9) + 1) \times 10^{-3+\frac{2n}{9}}$ for $n = \{0, 1, 2, \cdots, 18\}$. As β increases, the number of packets transmitted decreases.

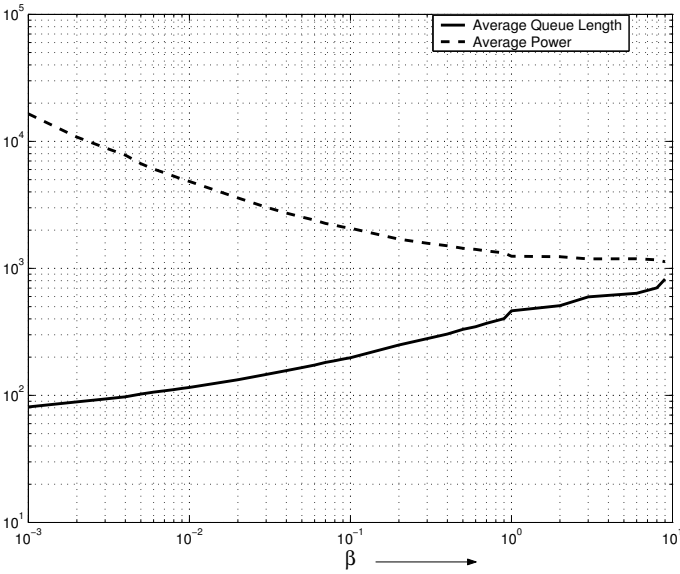

Fig. 6. The average transmitter power required and the mean queue length for various choices of β for the numerical example in Section III-E. Given a power constraint, one can find a smallest value of β for which the average power constraint is satisfied and the policy corresponding to that β would yields the minimum mean queue length as shown in the plot against that choice of β. Similarly, if the mean queue length constraint is given, one can find a largest value of β for which the queue length constraint is met and the policy corresponding to that β would yield a minimum average power required as per the curve shown in the plot against that choice of β.

We have the following formulation.

$$\min_{\mathbf{R}[k],\mathbf{P}[k]} \left\{ \limsup_n \frac{1}{n} \mathbb{E}_{\mathbf{x}} \sum_{k=0}^{n-1} \sum_{i=1}^{M} \left(\omega_i Q_i[k] + \lambda_i P_i[k] \right) \right\}, \quad (10)$$

subject to

$$\mathbf{R}[\mathbf{k}] \in C_g(\mathbf{H}[k], \mathbf{P}[k]);$$

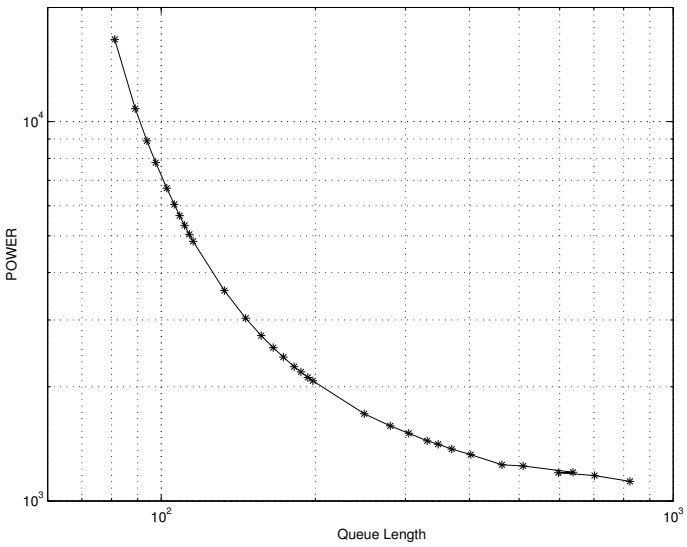

Fig. 7. Power-Delay trade-off curve for the numerical example in Section III-E

$$R_i[k] \in \{0, 1, \cdots, Q_i[k]\} \text{ for } i \in \{1, 2, \cdots, M\},$$

where subscript $\mathbf{x}$ denotes the initial state of the system.

This is a standard average cost constrained Markov decision problem. As before we consider the corresponding discounted cost problem.

$$\min_{\mathbf{R}[k],\mathbf{P}[k]} \left\{ \mathbb{E}_{\mathbf{x}} \sum_{k=0}^{\infty} \sum_{i=1}^{M} \alpha^k \left(\omega_i Q_i[k] + \lambda_i P_i[k] \right) \right\}, \quad (11)$$

subject to

$$\mathbf{R}[\mathbf{k}] \in C_g(\mathbf{h}[k], \mathbf{P}[k]);$$
$$R_i[k] \in \{0, 1, \cdots, Q_i[k]\} \text{ for } i \in \{1, 2, \cdots, M\}.$$

The corresponding discounted cost optimality equation is,

$$V_\alpha(\mathbf{q}, \mathbf{h}, \mathbf{a}) = \min_{(\mathbf{r},\mathbf{p}):\{\mathbf{r} \le \mathbf{q}; \mathbf{r} \in C_g(\mathbf{h},\mathbf{p})\}} \left\{ \sum_{i=1}^{M} (\omega_i q_i + \lambda_i p_i) + \alpha \mathbb{E}_{\mathbf{h},\mathbf{a}}[V_\alpha(\mathbf{q} - \mathbf{r} + \mathbf{A}, \mathbf{H}, \mathbf{A}))] \right\}, \quad (12)$$

where $\mathbb{E}_{\mathbf{h},\mathbf{a}}[f(\mathbf{A}, \mathbf{H})]$ denote the expectation of $f(\cdot, \cdot)$ conditioned upon $(\mathbf{h}, \mathbf{a})$, and $V_\alpha(\mathbf{x})$ is the discounted cost value function with discount factor $\alpha \in (0, 1)$ when starting in state $\mathbf{x}$.

A. Analysis of the Multiuser Problem

Observe that if we fix $\mathbf{r}$, the objective function in Equation 12 is minimized by that choice of $\mathbf{p}$ which solves

$$\min_{\mathbf{p}} \left\{ \sum_{i=1}^{M} \lambda_i p_i; \quad \text{subject to } \{\mathbf{p} : \mathbf{r} \in C_g(\mathbf{h}, \mathbf{p})\} \right\}.$$

This is true since the cost to go only depends upon the choice of $\mathbf{r}$. But we know (see [18]) that, given $\mathbf{h}$,$\mathbf{r}$, and $\boldsymbol{\lambda}$, and reindexing the users so that

$$\frac{\lambda_1}{h_1} \ge \frac{\lambda_2}{h_2} \ge \cdots \ge \frac{\lambda_M}{h_M},$$

the optimal value for the above problem is,

$$\sum_{i=1}^{M} \frac{\sigma^2 \lambda_i}{h_i} \left\{ e^{\theta(\sum_{k=1}^{i} r_k)} - e^{\theta(\sum_{k=1}^{i-1} r_k)} \right\}. \tag{13}$$

Recall that the decoding is sequential where a user with the lowest value of $\frac{\lambda}{h}$ is decoded first and the decoded signal is then subtracted from the received signal (see [18]). While decoding a signal, the interference only comes from transmissions of users having higher value of $\frac{\lambda}{h}$ than the user whose signal is being decoded. For each ordering of $\frac{\lambda_i}{h_i}$, we can similarly obtain the optimal cost. There are at most $M!$ distinct orderings possible. We enumerate the possible orderings and define $\nu(k,i)$, the index of the user with the i^{th} order in the k^{th} ordering. As an example, for $M = 2$, there are two possible orderings $(1,2)$ and $(2,1)$ indexed by $k = 1$ and $k = 2$ respectively. Then $\nu(1,1) = 1$, $\nu(1,2) = 2$, $\nu(2,1) = 2$ and $\nu(2,2) = 1$. Given $\mathbf{r}$, and $\mathbf{h}$ satisfying order k, the optimal value of $\boldsymbol{\lambda} \cdot \mathbf{p}$ for the k^{th} ordering is $\sum_{i=1}^{M} \frac{\sigma^2 \lambda_{\nu(k,i)}}{h_{\nu(k,i)}} \left\{ e^{\theta(\sum_{j=1}^{i} r_{\nu(k,j)})} - e^{\theta(\sum_{j=1}^{i-1} r_{\nu(k,j)})} \right\}$. We can rewrite the optimal value for the k^{th} ordering in the following convenient form,

$$\sum_{i=1}^{M} \sigma^2 \left(\frac{\lambda_{\nu(k,i)}}{h_{\nu(k,i)}} - \frac{\lambda_{\nu(k,i+1)}}{h_{\nu(k,i+1)}} \right) e^{\theta(\sum_{j=1}^{i} r_{\nu(k,j)})} - \frac{\sigma^2 \lambda_{\nu(k,1)}}{h_{\nu(k,1)}}, \tag{14}$$

where, by convention, $\frac{\lambda_{\nu(k,M+1)}}{h_{\nu(k,M+1)}}$ is zero.

The set of all possible channel gain vectors can be partitioned into $M!$ subsets such that each subset corresponds to one of the ordering, i.e., k^{th} ordering corresponds to k^{th} subset say $\mathcal{H}_k$ with $\bigcup_{k=1}^{M!} \mathcal{H}_k = \mathcal{H}$, the whole set. The channel gain transition probability matrix needs to be redefined. Let $P_{\mathbf{h},k}(\mathbf{H})$ define the probability that the next state of the channel gain vector is $\mathbf{H} \in \mathcal{H}_k$ given that the current channel gain vector is $\mathbf{h}$. Let $\mathbb{E}_{\mathbf{h},k}[f(H)]$ defined the conditional expectation of $f(\cdot)$ with respect to this transition probability matrix.

We define $M!$ value functions indexed by k say $V_k(\mathbf{q}, \mathbf{h}, \mathbf{a})$ where k signifies the fact that the channel gain vector $\mathbf{h} \in \mathcal{H}_k$. We now have a family of $M!$ coupled discounted cost optimality equations corresponding to the optimality equation of Equation 12. We drop the subscript α for convenience. Given $\mathbf{q}, \mathbf{a}$ and $\mathbf{h} \in \mathcal{H}_k$ for $k \in \{1, 2, \cdots, M!\}$, we have

$$V_k(\mathbf{q}, \mathbf{h}, \mathbf{a}) = \min_{\mathbf{r} \leq \mathbf{q}} \left\{ \sum_{i=1}^{M} \left(\omega_i q_i + \sigma^2 \frac{\lambda_{\nu(k,i)}}{h_{\nu(k,i)}} \left\{ e^{\theta \sum_{j=1}^{i} r_{\nu(k,j)}} - e^{\theta \sum_{j=1}^{i-1} r_{\nu(k,j)}} \right\} \right) + \alpha \sum_{l=1}^{M!} \mathbb{E}_{\mathbf{h},l,\mathbf{a}}[V_l(\mathbf{q} - \mathbf{r} + \mathbf{A}, \mathbf{H}, \mathbf{A})] \right\}, \tag{15}$$

where we note that the rate-power constraint has been eliminated. The corresponding value iteration algorithm, for $k \in \{1, 2, \cdots, M!\}$, is given by,

$$V_{k,n}(\mathbf{q}, \mathbf{h}, \mathbf{a}) = \min_{\mathbf{r} \leq \mathbf{q}} \left\{ \sum_{i=1}^{M} \left(\omega_i q_i + \sigma^2 \frac{\lambda_{\nu(k,i)}}{h_{\nu(k,i)}} \left\{ e^{\theta \sum_{j=1}^{i} r_{\nu(k,i)}} - e^{\theta \sum_{j=1}^{i-1} r_{\nu(k,j)}} \right\} \right) + \alpha \sum_{l=1}^{M!} \mathbb{E}_{\mathbf{h},l,\mathbf{a}}[V_{l,n-1}(\mathbf{q} - \mathbf{r} + \mathbf{A}, \mathbf{H}, \mathbf{A})] \right\},$$

where $V_{l,0}(\mathbf{x}) = 0$ for all $\mathbf{x}$. Recall from MDP theory that the first expression within the parentheses on the right hand side of the above value iteration is called the single stage cost while the second expression is called the cost to go. The convergence of the algorithm can be easily shown as in single user case.

The above problem appears intractable for structural analysis mainly due to a high degree of nonlinearity and coupling of the single stage cost function resulting in a highly complex cost to go expression $\sum_{l=1}^{M!} \mathbb{E}_{\mathbf{h},l,\mathbf{a}}[V_l(\mathbf{q} - \mathbf{r} + \mathbf{A}, \mathbf{H}, \mathbf{A})]$. Further, the huge state space associated with the above said control problem renders it impractical and computationally inefficient. Hence, there is a need for some near-optimal approximating scheme to it that addresses these analytical and computational difficulties.

We proceed as follows:

i) Obtain an additive separable approximation of the cost to go expression also known as the value function and study its structural properties. By an additive separable function we mean that it can be divided into additive terms with each being a function of only one user's variables. Such approximations exist in the literature [1].
ii) Carry out one-step value iteration with the approximated cost to go function and obtain the structural properties of the resulting policy. This idea of one-step of value iteration was introduced by Krishnan and Ott [13], and has been used in many papers since then. The remarkable fact about applying one-step value iteration is that the resulting policy could be very close to optimal if the cost to go function is chosen appropriately.

We assume that available rate set is bounded above. This is not an unrealistic assumption owing to the fact that only finite rate codewords are practical. Without loss of generality we assume that the bound is the same for all the users.

B. Cost to Go Approximation

Let $\hat{r}$ be the upper bound on the available rates set. Using the result obtained in Equation 13, the discounted cost problem (Equation 11) can be restated as,

$$\min_{\mathbf{R}[n]} \left\{ \mathbb{E}_{\mathbf{x}} \sum_{n=0}^{\infty} \sum_{i=1}^{M} \alpha^n \left(\omega_i Q_i[n] + \frac{\sigma^2 \lambda_{\nu(k[n],i)}}{H_{\nu(k[n],i)}} \{ e^{\theta \sum_{j=1}^{i} R_{\nu(k[n],j)}[n]} - e^{\theta \sum_{j=1}^{i-1} R_{\nu(k[n],j)}[n]} \} \right) \right\} \tag{16}$$

subject to $R_i[n] \in \min\{\hat{r}, \{0, 1, \cdots, Q_i[n]\}\}$ for $i \in \{1, 2, \cdots, M\}$, where $k[n]$ is the ordering during n^{th} slot.

In order to obtain the additive separable cost to go function, we first replace, for each $i \in \{1, 2, \cdots, M\}$, the rates $R_{\nu(k[n],j)}[n]$ with $\hat{r}$ for $j \neq i$, and obtain a reasonable upper bound for the above objective function. This replacement is equivalent to an assumption that every transmission sees maximum possible interference. For example, a user with the highest value of $\frac{\lambda}{h}$ observes no interference whereas a user with smallest value of $\frac{\lambda}{h}$ assumes that all the other users transmit at the highest possible rate $\hat{r}$. This will be the scenario when users are unaware of the queue length of other users. Next, we optimize the resulting function and obtain a tight uniform bound for the cost to go function. Define

$\mu(k,i) = \{j : \nu(k,j) = i\}$, the order of user i in the k^{th} ordering. We obtain the following optimization problem and refer to it as a cost bounding problem.

$$\min_{\mathbf{R}[n]} \left\{ \mathbb{E}_{\mathbf{x}} \sum_{n=0}^{\infty} \sum_{i=1}^{M} \alpha^n \left(\omega_i Q_i[n] + \frac{\sigma^2 \lambda_i}{H_i[n]} e^{\theta(\mu(k[n],i)-1)\hat{r}} (e^{R_i[n]} - 1) \right) \right\}.$$

Note that $e^{\theta(\mu(k[n],i)-1)\hat{r}}$ is the total power received from all the users to be decoded after decoding user i with the assumption that the users transmit at the rate $\hat{r}$, the maximum allowed transmission rate. The term $\frac{\sigma^2\lambda_i}{H_i[n]} e^{\theta(\mu(k[n],i)-1)\hat{r}}$ upper bounds the other user interference to user i since $\mu(k[n],i)-1$ is the number of users decoded after decoding the signal of user i. Observe that the objective function of the cost bounding problem can be separated user wise and yields a cost that upper bounds the minimal cost achievable by the original discounted cost problem (Equation 16).

We now study the cost bounding problem and obtain a value function which will serve as the approximate cost to go function. Since the objective function is separable, we tag user 1 and analyse the corresponding discounted cost optimality equation. The state vector includes the tagged user's queue length q_1, the arrival state a_1 and the channel gain vector $\mathbf{h}$ since the ordering k is determined by the complete channel gain vector. For notational simplicity, we represent $\mu(k,1)$ by $\mu(k)$ and set the weight of user 1, $\omega_1 = 1$. Define $\beta(\mathbf{h})$, for $\mathbf{h} \in \mathcal{H}_k$ representing k^{th} ordering, as,

$$\beta(\mathbf{h}) = \frac{\sigma^2 \lambda_1}{h_1} e^{\theta(\mu(k)-1)\hat{r}}.$$

We drop the subscript 1 for notational simplicity. The discounted cost optimality equation is given by,

$$V(q,\mathbf{h},a) = q + \min_{r \le \min\{q,\hat{r}\}} \{\beta(\mathbf{h})(e^{\theta r} - 1) + \alpha \mathbb{E}_{\mathbf{h},a}[V((q-r)^+ + A, \mathbf{H}, A)], \}, \quad (17)$$

Note that the Equation 17 has a close resemblance to the single user problem (Equation B-1 in the Appendix) where the only difference is that the rate is now constrained to $\hat{r}$ and $\beta(\mathbf{h})$ has replaced $\frac{\beta}{h}$. Along the lines of results for the single user problem, it is easy to prove the following result.

Theorem 4.1: The value function $V(q,\mathbf{h},a)$ is convex and monotone nondecreasing in q.

We now carry out one-step value iteration with $V(\cdot)$ serving as the approximating cost to go function.

C. One-Step Value Iteration

Recall the actual discounted cost problem (Equation 16) restated below for convenience. Given $\mathbf{x} = (\mathbf{q},\mathbf{h},\mathbf{a})$,

$$V(\mathbf{x}) = \min_{\mathbf{R}[n]} \left\{ \mathbb{E}_{\mathbf{x}} \sum_{n=0}^{\infty} \sum_{i=1}^{M} \alpha^n \left(\omega_i Q_i[n] + \frac{\sigma^2 \lambda_{\nu(k[n],i)}}{H_{\nu(k[n],i)}} \{ e^{\theta \sum_{j=1}^{i} R_{\nu(k[n],j)}[n]} - e^{\theta \sum_{j=1}^{i-1} R_{\nu(k[n],j)}[n]} \right) \right\}.$$

We write the above objective function as a sum of two components, namely, $n = 0$ representing the first stage cost and $n > 0$ representing the aggregate expected cost to go. By definition of the initial state of the system $\mathbf{x} = (\mathbf{q},\mathbf{h},\mathbf{a})$, we have $Q_i[0] = q_i$, $H_i[0] = h_i$ and $A_i[0] = a_i$ for $i = \{1,2,\cdots,M\}$. Let the ordering $k[0]$, corresponding to $\mathbf{h}$, be k. Thus, given a rate vector $\mathbf{r} \le \min\{\mathbf{q},\hat{\mathbf{r}}\}$, the first stage cost equals

$$\sum_{i=1}^{M} \left(\omega_i q_i + \sigma^2 \frac{\lambda_{\nu(k,i)}}{h_{\nu(k,i)}} \left\{ e^{\theta \sum_{j=1}^{i} r_{\nu(k,j)}} - e^{\theta \sum_{j=1}^{i-1} r_{\nu(k,j)}} \right\} \right).$$

Now that the vector $\mathbf{r}$ is the transmission rate vector for the first stage, the queue length vector at the end of the first stage equals $\mathbf{q} - \mathbf{r} + \mathbf{a}$. The channel gain vector and the arrival rate vector for the second stage will be random variables with transition density conditional on $\mathbf{h}$ and $\mathbf{a}$ respectively. The aggregate expected cost to go thus equals

$$\alpha \mathbb{E}_{\mathbf{h},\mathbf{a}} V(\mathbf{q} - \mathbf{r} + \mathbf{a}, \mathbf{H}, \mathbf{A}).$$

We substitute the above cost to go function with the separable approximation carried out in Section IV-B. The approximated cost to go function is,

$$\sum_{i=1}^{M} \alpha \mathbb{E}_{\mathbf{h},a_i} V_i(q_i - r_i + A_i, \mathbf{H}, A_i),$$

where $V_i(\cdot)$ is as obtained in Equation 17 and satisfies the properties stated in Theorem 4.1. The approximation as stated earlier provides a tight upper bound for the actual cost to go function and a one step optimization would further result in a close to optimal solution.

Given any $\mathbf{q},\mathbf{a}$ and $\mathbf{h} \in \mathcal{H}_k$, the one-step value iteration is to obtain the rate vector $\mathbf{r} \le \min\{\mathbf{q},\hat{\mathbf{r}}\}$ that minimizes,

$$\sum_{i=1}^{M} \left(\omega_i q_i + \sigma^2 \frac{\lambda_{\nu(k,i)}}{h_{\nu(k,i)}} \left\{ e^{\theta \sum_{j=1}^{i} r_{\nu(k,j)}} - e^{\theta \sum_{j=1}^{i-1} r_{\nu(k,j)}} \right\} + \alpha \mathbb{E}_{\mathbf{h},a_i} V_i(q_i - r_i + A_i, \mathbf{H}, A_i) \right). \quad (18)$$

This is an approximate multiuser problem. The solution that we obtain by solving this problem will provide a close upper bound to the optimal system performance. We refer to the solution to this problem as an one-step iterated policy.

Theorem 4.1 implies that the objective function (Equation 18) is strictly convex in $\mathbf{r}$ which, along with the fact the decision space in compact, implies the existence of a unique minimizer $\mathbf{r}^*$. We now derive structural properties of the one-step iterated policy. Since $\mathbf{h}$ is given and fixed, without loss of generality we assume that $\mathbf{h} \in \mathcal{H}_k$ and the ordering k is such that $\nu(k,i) = i$. Define $r_{-i} := \{r_j, j \ne i\}$. Given r_{-i}, we study the structural properties of r_i and show that the optimal one-step iterated policy is the solution to a fixed point equation.

We factor out terms involving r_i (the decision variable) in the objective function.

$$\min_{r_i \le \{q_i,\hat{r}\}} \left\{ e^{\theta r_i} \left[e^{\theta \sum_{l=1}^{i-1} r_l} \left(\frac{\sigma^2 \lambda_i}{h_i} + \sum_{j=i+1}^{M} \frac{\sigma^2 \lambda_j}{h_j} (e^{\theta \sum_{l=i+1}^{j} r_l} - e^{\theta \sum_{l=i+1}^{j-1} r_l}) \right) \right] + \alpha \mathbb{E}_{\mathbf{h},a_i} V_i(q_i - r_i + A_i, \mathbf{H}, A_i) \right\},$$

where V_i is the solution of Equation 17. If we denote the expression within the square brackets by $g(\mathbf{r}_{-i})$, the optimization problem becomes,

$$\min_{r_i \le \{q_i, \hat{r}\}} \left\{ e^{\theta r_i} g(\mathbf{r}_{-i}) + \alpha \mathbb{E}_{\mathbf{h},a_i} V_i(q_i - r_i + A_i, \mathbf{H}, A_i) \right\}.$$

The analysis of the problem is similar to the analysis of the single user problem. As in Section III-B, we define for $i = \{1, 2, \cdots, M\}$ a differential of the value function as $G_i(q_i, \mathbf{h}, a_i) = V_i(q_i, \mathbf{h}, a_i) - V_i(q_i - 1, \mathbf{h}, a_i)$ and

$$Z_i(q_i, \mathbf{h}, a_i) = e^{\theta q_i} \alpha \mathbb{E}_{\mathbf{h},a_i}[G(q_i + A_i, \mathbf{H}, A_i)],$$

where $\mathbb{E}_{\mathbf{h},a_i}[\cdot]$ denotes expectation with respect to the transition probability of H and A_i with initial state $\mathbf{h}$ and a_i respectively.

Note that $Z(q_i, \mathbf{h}, a_i)$ is monotone increasing in q_i as the value function $V(q_i, \mathbf{h}, a_i)$ is convex in q_i. Define a function $u_i^*(q_i, h_i, a_i)$ as the value of u that solves the following inequalities, for given $(q_i, \mathbf{h}, a_i)$,

$$Z_i(u, \mathbf{h}, a_i) \le e^{\theta q_i}(e^{\theta} - 1) \le Z_i(u + 1, \mathbf{h}, a_i). \tag{19}$$

The minimizing policy $r_i(q_i, \mathbf{h}, a_i)$ equals

$$\min \left\{ \hat{r}, \max \left\{ q_i - u_i^* \left(q_i + \frac{1}{\theta} \log(g(\mathbf{r}_{-i})), \mathbf{h}, a_i \right), 0 \right\} \right\}.$$

We state the following results.

Theorem 4.2: Given the transmission rates of all the other users $\mathbf{r}_{-i}$ and the channel gain vector, the structural properties of the optimal one-step iterated policy for user i is as follows.

i) The optimal policy $r_i(q_i, \mathbf{h}, a_i)$ is monotonic nondecreasing in q_i and $u_i(q_i, \mathbf{h}, a_i) = q_i - r_i(q_i, \mathbf{h}, a_i)$ is monotonic nondecreasing in q_i as well.
ii) The optimal number of packets transmitted $r_i(q_i, \mathbf{h}, a_i)$ increases to $\hat{r}$ as q_i increases to infinity.
iii) The optimal solution $r_i(q_i, \mathbf{h}, a_i)$ is monotone nonincreasing in $g(\mathbf{r}_{-i})$.

Proof: The proofs of i) and ii) are along the lines of single user discounted cost analysis (Refer Appendix C). Result iii) follows from the fact that as $g(\mathbf{r}_{-i})$ increases, the first stage cost increases while the cost to go remains the same. ■

Remark 4.1: According to result iii), if any of the $r_j, j \neq i$ increases, the function $g(\mathbf{r}_{-i})$ increases and hence r_i decreases.

We obtain similar results by tagging other users one-by-one. Given the system state $(\mathbf{q}, \mathbf{h}, \mathbf{a})$, the unique minimizer $\mathbf{r}^* = \{r_1, r_2, \cdots, r_M\}$ is obtained by solving the following family of nonlinear equations. For each $i = \{1, 2, \cdots, M\}$, we solve,

$$r_i = \min \left\{ \hat{r}, \max \left\{ q_i - u_i^* \left(q_i + \frac{1}{\theta} \log(g(\mathbf{r}_{-i})), \right), 0 \right\} \right\}.$$

We now turn our attention to a special case of on-off control.

D. On-Off Control

In practice, one cannot change the coding rate every slot and in most systems, only one code book is available at each of the transmitters implying only one transmission rate. In this section, we assume that only one transmission rate is possible. The transmitter is allowed to decide whether to transmit at that rate, or not to transmit at all in a given slot. We call a user as ON over a slot if it transmits at the allowed rate otherwise the user is OFF, i.e., it does not transmit anything in that slot. At the start of each slot, depending upon the state of the system, the controller make the on-off decisions.

1) Analysis of Cost Bounding Problem: We solve the cost bounding problem to get an approximate cost to go function. Tag user 1. Based on the system state $(q, \mathbf{h}, a)$, the tagged user may decide to be ON and transmit at a fixed rate $r_1 = r$ or it may just decide to be OFF meaning it does not transmit $r_1 = 0$. Given the system state, let the ordering be k and define $\mu(k) := \{i : \nu(k, i) = 1\}$, the index of the tagged user under k^{th} ordering. Recall earlier definition of $\beta(\mathbf{h})$.

- If **ON**: The queue length at the end of current slot would be $(q - r)^+ + A$ where A is a random variable with probability distribution corresponding to the a^{th} row of the arrival state transition probability matrix. The costs incurred are the holding cost q and the power price $\beta(\mathbf{h})(e^{\theta r} - 1)$.
- If **OFF**: The queue length at the end of current slot would be $q + A$ where A is a random variable with probability distribution corresponding to the a^{th} row of the arrival state transition probability matrix. The costs incurred would be the holding cost q.

The above problem can now be formulated as the following Markov decision problem. Let $V(q, \mathbf{h}, a)$ be the discounted cost value function. The discounted cost optimality equation is,

$$V(q, \mathbf{h}, a) = q + \min \{ \beta(\mathbf{h})(e^{\theta r} - 1) + \alpha \mathbb{E}_{\mathbf{h},a}[V((q - r)^+ + A, \mathbf{H}, A)], \alpha \mathbb{E}_{\mathbf{h},a}[V(q + A, \mathbf{H}, A)] \}, \tag{20}$$

where by definition the optimal action is ON if $\beta(\mathbf{h})(e^{\theta r} - 1) + \alpha \mathbb{E}_{\mathbf{h},a}[V((q - r)^+ + A, \mathbf{H}, A)]$ is less than $\alpha \mathbb{E}_{\mathbf{h},a}[V(q + A, \mathbf{H}, A)]$, and is OFF otherwise. Note that $\beta(\mathbf{h}) > 0$. Consider the corresponding discounted cost value iteration algorithm. Let $V_0(q, \mathbf{h}, a) = 0$. Then for $n \ge 1$,

$$V_{n+1}(q, \mathbf{h}, a) = q + \min \{ \beta(\mathbf{h})(e^{\theta r} - 1) + \alpha \mathbb{E}_{\mathbf{h},a}[V_n((q - r)^+ + A, \mathbf{H}, A)], \alpha \mathbb{E}_{\mathbf{h},a}[V_n(q + A, \mathbf{H}, A)] \}. \tag{21}$$

Define $\mathbf{x} = (\mathbf{h}, a)$ and $W_n(q, \mathbf{x}) = \mathbb{E}_{\mathbf{h},a}[V_n(q + A, \mathbf{H}, A)]$. Let $W(q, \mathbf{x})$ be the limiting function. Thus

$$V_{n+1}(q, \mathbf{x}) = q + \min\{\beta(\mathbf{h})(e^{\theta r} - 1) + \alpha W_n((q - r)^+, \mathbf{x}), \alpha W_n(q, \mathbf{x})\}. \tag{22}$$

We now state structural results. The proofs are given in Appendix D.

Theorem 4.3: Given $\mathbf{x}$, the function $W(q, \mathbf{x})$ is monotone increasing in q. ■

Theorem 4.4: Given $\mathbf{x}$, the difference $W(q, \mathbf{x}) - W((q - r)^+, \mathbf{x})$ is nondecreasing in q. ■

The channel gain vector $\mathbf{h}$ appearing in the state space does not contain any more information then contained in $\beta(\mathbf{h})$. Since the set $\mathcal{H}$ is finite, we enumerate it. Define $\zeta^j = \beta(\mathbf{h}_j)$. The transition probability matrix for the Markov chain $\mathbf{h}$, yields a new Markov chain $Z[n]$ with states space ζ^j for all $j \in \{1, 2, \cdots, |\mathcal{H}|\}$. Let P_ζ be the transition probability matrix. The new state space is (q, ζ, a). The optimality equation is

$$V_{n+1}(q,\zeta,a) = q + \min\{\zeta(e^{\theta r}-1) + \alpha W_n((q-r)^+,\zeta,a), \alpha W_n(q,\zeta,a)\}, \quad (23)$$

where $W_n(q,\zeta,a) = \mathbb{E}_{\zeta,a}[V_n(q+A,Z,A)]$.

Theorem 4.5: The difference $\frac{W(q+r,\zeta,a)-W(q,\zeta,a)}{\zeta}$ is nonincreasing in ζ. ■

2) One-Step Value Iteration: As discussed for the general multiuser case, the cost to go function is approximated by $W(q,\zeta,a)$. Without loss of generality, let $\mathbf{h} \in \mathcal{H}_k$ with $\nu(k,i) = i$. The one-step control problem is to obtain $r_i \in \{0, r\}$ for $i \in \{1, 2, \cdots, M\}$ that minimizes,

$$\sum_{i=1}^{M} \left(\omega_i q_i + \sigma^2 \frac{\lambda_i}{h_i} \left\{ e^{\theta \sum_{j=1}^{i} r_j} - e^{\theta \sum_{j=1}^{i-1} r_j} \right\} + \alpha \sum_{i=1}^{M} W_i(q_i - r_i + a_i, \zeta_i, a_i) \right), \quad (24)$$

where $\zeta_i = \beta_i(\mathbf{h})$. We now obtain the structural properties of one-step iterated ON/OFF policy. Define $r_{-i} = \{r_j, j \neq i\}$ and

$$g(r_{-i}) = e^{\theta \sum_{l=1}^{i-1} r_l} \left(\frac{\sigma^2 \lambda_i}{h_i} + \sum_{j=i+1}^{M} \frac{\sigma^2 \lambda_j}{h_j} \left(e^{\theta \sum_{l=i+1}^{j} r_l} - e^{\theta \sum_{l=i+1}^{j-1} r_l} \right) \right).$$

Thus given r_{-i}, the one step control problem is to obtain $r_i \in \{0, r\}$ for $i \in \{1, 2, \cdots, M\}$ that minimizes,

$$e^{\theta r_i} g(r_{-i}) + \alpha \sum_{i=1}^{M} W_i(q_i - r_i + a_i, \zeta_i, a_i)$$

Theorem 4.6: Given any i and r_{-i}, the optimal one-step iterated policy r_i is a threshold policy, i.e., for each $(\mathbf{h}, \mathbf{a})$ there exists a threshold $q_i^*(\mathbf{h}, \mathbf{a})$ such that the user is ON if and only if $q_i \geq q_i^*(\mathbf{h}, \mathbf{a})$.

Proof: Result follows directly from Theorem 4.4. ■

Theorem 4.7: Given i, the threshold q_i^* increases with $g(r_{-i})$.

Proof: As $g(r_{-i})$ increases, the first stage cost increases while the cost to go remains the same. ■

Remark 4.2: Given the channel gain vector and the arrival vector, if $r_i = r$ for all $i \neq j$, $g(r_{-j})$ is the maximum and hence the ON threshold q_j^* is the largest. Since user j transmits only if q_j is larger than q_j^*, this largest threshold suggests that user j transmits if its queue length is larger than the largest threshold irrespective of the decision of the other users.

V. Conclusion

We considered the mean packet transmission delay optimization problem for a multiaccess fading communication system with the objective of designing cross-layer algorithms by jointly optimizing two or more of the layers of the communication system. The control objective is to minimize the mean packet transmission delay subject to a long run average transmission power constraint. First, we studied the problem for a single user system by formulating it as a constrained Markov decision problem. We analysed a family of corresponding unconstrained problems and obtained structural properties of the optimal policy in Section III. Further, given (h, a), the channel state and the number of arrivals, the optimal policy $r(q,h,a)$ is monotone nondecreasing in q; $q - r(q,h,a)$ is monotone nondecreasing in q; and for any state vector (q,h,a), $r(q,h,a)$ is monotone nonincreasing in the power price (shadow) β; $r(q,h,a)$ goes to zero as the price goes to infinity. Particularly for the i.i.d. fading and arrival model, we could completely characterize the optimal policy (see Figure 4). We showed the existence of a problem among the family of unconstrained problems such that the optimal policy for that problem also solves the constrained problem. We numerically evaluated the optimal policy and obtained an optimal delay-power tradeoff curve.

Next, we studied the multiuser problem and obtained the value iteration algorithm for computing the optimal policy. Due to high complexity of the cost to go function, we approximate it with an additive separable function that upper bounds the original cost function. A one-step value iteration is then carried out to obtain a 'good' control policy. We studied the structural properties of the control policy and showed that the control policy is obtained by solving M nonlinear equations. We also studied an on-off multiuser system and showed that the control policy is of threshold form.

References

[1] Adelman Daniel and A. Mersereau,"Relaxations of weakly coupled stochastic dynamic programs," *Accepted in Operations Research*. Available for download at first author's website.

[2] Randall A. Berry, "Power and Delay Trade-offs in Fading Channels," *Ph.D. Thesis*, MIT, June 2000.

[3] Randall A. Berry, R. G. Gallager, "Communication over Fading Channels with Delay Constraints," *IEEE Transaction on Information Theory*, vol. 48, no. 5, 1135-1149, May 2002.

[4] Randall A. Berry, Edmund Yeh, "Cross-layer Wireless Resource Allocation: Fundamental Performance Limits," *IEEE Signal Processing Magazine*, 21(5):59-68, 2004.

[5] D. P. Bertsekas, "Dynamic Programming and Optimal Control (Vol I,II)," Athena Scientific, Massachusetts, 2001.

[6] Rong Rong Chen, S. Meyn, "Value Iteration and Optimization of Multiclass Queueing Networks," *Queueing System and Applications*, 32(1):65-97, 1999.

[7] B. Collins and R. Cruz,"Transmission Policies for Time Varying Channels with Average Delay Constraints," in *Proc. Allerton Conf. on Commun. Control*, 1999.

[8] R. G. Gallager, "Information Theory and Reliable Communication," New York: Wiley, 1968.

[9] A. J. Goldsmith, P. Varaiya, "Capacity of Fading Channels with Channel Side Information," *IEEE Transaction on Information Theory*, vol. 43, no. 6, 1986-1992, Nov 1997.

[10] Stephen V Hanly . and David N. C. Tse , "Multi-access fading channels: Part II: Delay limited capacities," *IEEE Trans. on Info. Theory*, 44(7):2816-2831, 1998.

[11] Munish Goyal, Anurag Kumar, Vinod Sharma, "Optimal Resource Allocation Policies for a Multiple-Access Fading Channel with a Quality of Service Constraint," IEEE International Symposium on Information Theory 2002, Lausanne, Switzerland.
[12] Munish Goyal, Anurag Kumar, Vinod Sharma, " Power Constrained and Delay Optimal Policies for Scheduling Transmissions over a Fading Channel," Summary appeared in IEEE Information Theory Workshop, 2003, Full Version appeared in IEEE Infocom 2003, San Francisco, USA.
[13] Krishnan K.R. and T.J. Ott," On state-dependent routing for telephone traffic: Theory and Results,"*IEEE CDC*, 2124-2128, 1986.
[14] O. H. Lerma, J. B. Lassere," Discrete-time Markov control processes: Basic optimality criteria," Springer Verlag, New York, 1996.
[15] D J. Ma, A.M.Makowski, A.Shwartz, "Estimation and Optimal Control for Constrained Markov Chains," *IEEE Conf. on Decision and Control*, Dec 1986.
[16] S. P. Meyn, R. L. Tweedie, "Markov Chains and Stochastic Stability," Springer-Verlag, London, 1993.
[17] Manfred Schal, "Average Optimality in Dynamic Programming with General State Space," *Mathematics of Operations Research*, vol. 18, no. 1, 163-172, Feb 1993.
[18] David N. C. Tse and Stephen V. Hanly, "Multi-access fading channels: Part I: Polymatroid structure, optimal resource allocation and throughput capacities ," *IEEE Trans. on Info. Theory*, 44(7):2796-2815, 1998.
[19] J. Walrand," An Introduction to Queueing Networks,". Prentice-Hall, New Jersey, 1988.
[20] R. W. Wolff, "Stochastic Modeling and the Theory of Queues," Prentice Hall, New Jersey, 1988.

APPENDIX A
MDP: SOME KEY RESULTS

Consider a discrete time controlled Markov process $\{X[n], A[n], n \in \{0, 1, 2, \cdots\}\}$, with state space $\mathcal{X}$, and action space $\mathcal{A}$. The set of feasible actions a in state $x \in \mathcal{X}$ is $A(x) \subset \mathcal{A}$. Let $\mathcal{K}$ be the set of all feasible state-action pairs. The transition kernel on $\mathcal{X}$ given an element $(x, a) \in \mathcal{K}$ is denoted by Γ. The immediate one step cost $c : \mathcal{K} \to \mathbb{R}^+$ is $c(x, a)$. At time n, given $x[n]$ and $a[n]$, the system moves to the next state $x[n + 1]$, a $\mathcal{X}$ valued random variable with distribution $\Gamma(\mathcal{B}|(x, a)) := Prob(x[n + 1] \in \mathcal{B}|x[n] = x, a[n] = a)$ and a cost $c(x, a)$ is incurred. A policy $\pi = \{a[n]\}$ generates at time n an action $a[n]$ depending upon the entire history of the process, i.e., at decision instant $n \in \{0, 1, 2, \cdots\}$, π_n is a mapping from $\mathcal{K}^n \times \mathcal{X}$ to $A(x)$. Let Π be the space of all such policies. A stationary policy is a measurable mapping $f : \{x \in \mathcal{X} : x \to \mathcal{A}(x)\}$, i.e., $\pi = \{f, f, \cdots\}$ is a stationary policy. In addition to the dynamic system and a set of policies, we need a performance criterion, also called objective function. The average cost Markov decision problem (MDP) is to minimize,

$$J^\pi(x) = \limsup_n \frac{1}{n} \mathbb{E}_x^\pi \sum_{k=0}^{n-1} c(X[k], A[k]),$$

over all $\pi \in \Pi$. The discounted cost problem is,

$$V_\alpha(x) = \min_{\pi \in \Pi} \mathbb{E}_x^\pi \left[\sum_{k=0}^{\infty} \alpha^k c(X[k], A[k]) \right],$$

where $\alpha \in (0, 1)$. Dynamic programming is an algorithm to compute the optimal value functions $V^*(x)$ or $J^*(x)$ and the corresponding optimal policies $\pi^* \in \Pi$.

Definition 1.1: A function $c(x, a)$ is said to be *inf-compact* on $\mathcal{K}$ if for every $x \in \mathcal{X}$ and $r \in \mathbb{R}$, the set $\{a \in A(x) : c(x, a) \leq r\}$ is compact.

First, we give sufficient conditions for the existence of a stationary discounted cost optimal policy and that it can be obtained using a value iteration algorithm.

Theorem A-1: [14] Suppose

D1. $c(x, a)$ is lower semi-continuous, nonnegative and inf-compact (see Definition 1.1) on $\mathcal{K}$
D2. Γ is strongly continuous
D3. There exists a policy π such that the discounted cost is finite for each initial state $x \in \mathcal{X}$

then the discount value function $V^*(x)$ is the minimal solution of the following optimality equation (DCOE)

$$v(x) := \min_{a \in A(x)} \left\{ c(x, a) + \alpha \int_{\mathcal{X}} v(y) Q(dy|(x, a)) \right\}.$$

Further, the minimizer to the DCOE is the stationary discount optimal policy $f^*(x)$. The following iterative algorithm called the discounted cost value iteration algorithm converges to $V^*(x)$, the discount value function and satisfies the DCOE.

$$v_n(x) := \min_{a \in A(x)} \left\{ c(x, a) + \alpha \int_{\mathcal{X}} v_{n-1}(y) Q(dy|(x, a)) \right\}, \tag{A-1}$$

where $v_0(x) := 0$.

The average cost problems are normally studied as a limit of discount cost problems as the discount factor α tends to one. We need some conditions to ensure such a convergence. We use subscript α to denote discount value function or a discount optimal policy.

Lemma A-1: [[17], Proposition 2.1] Suppose

W1. $\mathcal{X}$ is a locally compact space with a countable base.
W2. $A(x)$, the set of feasible actions in state x, is a compact subset of $\mathcal{A}$ (the action space), and the multifunction $x \to A(x)$ is upper semi-continuous.
W3. Q is continuous in a with respect to weak convergence in the space of probability measures.
W4. $c(x, a)$ is lower semi-continuous

then there exists a discounted cost stationary optimal policy f_α for each $\alpha \in (0, 1)$.

Now we state a result related to the existence of stationary average optimal policies which can be obtained as limit of discounted cost optimal policies f_α. Define

$$w_\alpha(x) = V_\alpha(x) - \inf_{x \in \mathcal{X}} V_\alpha(x).$$

Theorem A-2: [[17],Theorem 3.8] Suppose there exists a policy Ψ and an initial state $x \in \mathcal{X}$ such that the average cost $J^\Psi(x) < \infty$. Further suppose $\sup_{\alpha < 1} w_\alpha(x) < \infty$ for all $x \in \mathcal{X}$ and the hypothesis of Lemma A-1 is satisfied. Then there exists a stationary policy f_1 which is average cost optimal and the optimal cost is independent of the initial state. Also f_1 is *limit discount optimal* in the sense that, for any $x \in \mathcal{X}$ and given any sequence of discount factors converging to one, there exists a subsequence $\{\alpha_m\}$ of discount factors and a sequence $x_m \to x$ such that $f_1(x) = \lim_{m \to \infty} f_{\alpha_m}(x_m)$.

Remark A-1: In Theorem A-2, the subsequence of discount factors depends upon the choice of x. ■

The following theorem gives a useful bound on $w_\alpha(x)$.

Theorem A-3: [17] Given a constant $\eta > 0$, define a stopping time ς as,

$$\varsigma = \inf\{n \geq 0, V_\alpha(X[n]) \leq V_\alpha(x_0) + \eta\},$$

where $V_\alpha(X[n])$ is the discount value function for discount α and x_0 is a state for which $V_\alpha(x)$ is minimum. Define $w_\alpha(x) = V_\alpha(x) - V_\alpha(x_0)$. Then for x in the state space, we have

$$w_\alpha(x) \leq \eta + \inf_\pi \mathbb{E}_x^\pi \left[\sum_{n=0}^{\varsigma-1} c(X[n], A[n])\right].$$

APPENDIX B
PROOF OF LEMMA 3.1

Consider the discounted cost MDP and the corresponding value function as defined in Equation 7. First, we verify the conditions W of Lemma A-1 for the existence of stationary discount cost optimal policies. Condition $W1$ holds since the set of feasible actions in any state is finite. All functions are continuous since the underlying topology is discrete and thus implying $W2, W3, W4$. The single stage cost $c(\cdot)$ is non-negative. Next we verify applicability of Theorem A-2 (see Appendix A) to ascertain the existence of a stationary average optimal policy and that it can be obtained as a limit of discount optimal policies as the discount factor increases to one.

The first hypothesis of Theorem A-2 should hold in most practical problems because otherwise the cost is infinite for any choice of the policy, and thus any policy is optimal. To verify that $\sup_{\alpha<1} w_\alpha(x) < \infty$ for $x \in \mathcal{X}$, consider the following discounted cost optimality equation (DCOE) for $V_\alpha(x)$,

$$V_\alpha(q,h,a) = \min_{r\in\{0,1,\cdots,q\}} \Big\{q + \beta\frac{\sigma^2}{h}(e^{\theta r} - 1) + \alpha\mathbb{E}_{h,a}[V_\alpha(q - r + A, H, A))]\Big\}, \quad \text{(B-1)}$$

where $\mathbb{E}_{h,a}[f(H, A)]$ denote the expectation of $f(\cdot,\cdot)$ conditioned upon (h, a).

Given (h, a), $V_\alpha(q, h, a)$ is clearly increasing in q since the larger the initial buffer the larger will be the cost to go. We will later prove this as well. Thus $\arg\ \inf_{y\in\mathcal{X}} V_\alpha(y) = (0, h_0, a_0) := x_0$, i.e., the infimum is achieved when the system starts with an empty buffer, and for some channel state h_0 and arrival state a_0. Also when the buffer is empty, the set of feasible actions is $\{0\}$ and $c(x_0, 0) = 0$, we have, $V_\alpha(x_0) = \alpha\mathbb{E}_{h_0,a_0}[V_\alpha(A, H, A)]$. In addition, as the policy $r(q, h, a) = q$ is feasible for state (q, h, a), we have from B-1 that

$$V_\alpha(x) \leq q + \frac{\beta\sigma^2}{h}(e^{\theta q} - 1) + \alpha\mathbb{E}_{h,a}[V_\alpha(A, H, A)].$$

Let the system start in state (a, h, a) and we take an action $r[n] = a[n]$ at time $n\tau$ for all n. Thus at any time instant $n\tau$, the state would be $(a[n], h[n], a[n])$. Let $\nu(h, a)$ be the expected number of slots to hit the state (a_0, h_0, a_0) for the first time when starting in state (a, h, a). Note that $\nu(h, a)$ depends entirely on the transition probability matrices. Note that $\nu(h, a)$ is finite for all values of h and a. Define $c_{\max} = \max_{h,a}\Big\{a + \frac{\beta\sigma^2}{h}(e^{\theta a} - 1)\Big\}$. Then from Wald's lemma [20] we have

$$\begin{aligned}
&\alpha\mathbb{E}_{h,a}[V_\alpha(A, H, A)] \\
&\quad\leq\ c_{\max}\nu(h, a) + \alpha\mathbb{E}_{h_0,a_0}[V_\alpha(A, H, A)] \\
&\quad\leq\ c_{\max}\nu(h, a) + V_\alpha(x_0).
\end{aligned}$$

Thus substituting the above expression in the upper bound for $V_\alpha(x)$ we get,

$$V_\alpha(x) \leq q + \frac{\beta\sigma^2}{h}(e^{\theta q} - 1) + c_{\max}\nu(h, a) + V_\alpha(x_0).$$

Implying for all $x \in \mathcal{X}$,

$$w_\alpha(x) := V_\alpha(x) - V_\alpha(x_0) \leq q + \frac{\beta\sigma^2}{h}(e^{\theta q} - 1) + c_{\max}\nu(h, a) < \infty.$$

Thus all the conditions of Theorem A-2 are satisfied. *We have proved the existence of a stationary average optimal policy which can be obtained as a limit of discount optimal policies as described in Theorem A-2.*

APPENDIX C
ANALYSIS OF THE UNCONSTRAINED PROBLEM

We now study the discounted cost MDP and obtain the structural properties of discount optimal policies. As the discount factor α is fixed for the analysis of the remaining section, we drop the subscript α. Recall the definition of u as defined in Section III-B. The discounted cost optimality equation (DCOE) in the control variable u is,

$$V(q,h,a) = \min_{u\in\{0,1,\cdots,q\}} \Big\{q + \frac{\beta\sigma^2}{h}\left(e^{\theta(q-u)} - 1\right) + \alpha\mathbb{E}_{h,a}[V(u + A, H, A)]\Big\} \quad \text{(C-1)}$$

and the corresponding value iteration algorithm is,

$$V_n(q,h,a) = \min_{u\in\{0,1,\cdots,q\}} \Big\{q + \frac{\beta\sigma^2}{h}\left(e^{\theta(q-u)} - 1\right) + \alpha\mathbb{E}_{h,a}[V_{n-1}(u + A, H, A)]\Big\}, \quad \text{(C-2)}$$

with $V_0(q, h, a) = 0$. It can be easily seen that the control problem satisfies the hypothesis of the Theorem A-1 and thus the convergence of the value iteration. We now provide structural results.

Lemma C-1: The value function $V(q, h, a)$ is an increasing convex function of q for each (h, a). ■

Proof: It is easy to see that the value function $V(q, h, a)$ is increasing in q. Consider the value iteration algorithm (Equation C-2). For $n = 0$, $V_0(q, h, a) = 0$. This implies that $V_1(q, h, a) = q$ and thus increasing. Let $V_{n-1}(q, h, a)$ be increasing in q. Fix (h, a). The set of feasible actions in state $q+1$ is $\{0, 1, \cdots, q+1\}$ whereas for state q, the feasible action set is $\{0, 1, \cdots, q\}$. We show that for any action in state $q + 1$, the value $V_n(q, h, a)$ is less than $V_n(q + 1, h, a)$. Let the optimal action in state $q + 1$ be $u \in \{0, 1, \cdots, q\}$. Thus

$$V_n(q+1,h,a) = q + 1 + \frac{\beta\sigma^2}{h}\left(e^{\theta(q+1-u)} - 1\right) + \alpha\mathbb{E}_{h,a}[V_{n-1}(u + A, H, A)].$$

Since this u is also feasible for state q, we have,

$$\begin{aligned} & V_n(q,h,a) \\ &\leq \quad q+\frac{\beta\sigma^2}{h}\left(e^{\theta(q-u)}-1\right)+\alpha\mathbb{E}_{h,a}[V_{n-1}(u+A,H,A)] \\ &\leq \quad V_n(q+1,h,a). \end{aligned}$$

Now let the optimal action in state $q+1$ be $u=q+1$. Thus $V_n(q+1,h,a) = q+1+\alpha\mathbb{E}_{h,a}[V_{n-1}(q+1+A,H,A)]$. Also since $u=q$ is a feasible action in state q, we have $V_n(q,h,a) = q+\alpha\mathbb{E}_{h,a}[V_{n-1}(q+A,H,A)]$. Now from the increasing property of $V_{n-1}(q,h,a)$ in q, we have the result. Induction hypothesis implies that $V(q,h,a)$ is increasing in q.

We prove convexity of $V(q,h,a)$ by induction. For $n=0$, $V_0(q,h,a)=0$ and hence convex. Assume $V_{n-1}(q,h,a)$ is convex in q. Fix (q,h,a). Let u_1 and u_2 be the optimal policy for $q-1$ and $q+1$.

$$\begin{aligned} & V_n(q+1,h,a)+V_n(q-1,h,a) \\ &= \quad 2q+\frac{\beta\sigma^2}{h}\left(e^{\theta(q-1-u_1)}+e^{\theta(q+1-u_2)}-2\right) \\ &\quad +\alpha\mathbb{E}_{h,a}[V_{n-1}(u_1+A,H,A)+V_{n-1}(u_2+A,H,A)], \\ &\geq \quad 2q+\tfrac{\beta\sigma^2}{h}\left(e^{\theta(q-\lfloor\frac{u_1+u_2}{2}\rfloor)}-1\right)+\tfrac{\beta\sigma^2}{h}\left(e^{\theta(q-\lceil\frac{u_1+u_2}{2}\rceil)}-1\right) \\ &\quad +\alpha\mathbb{E}_{h,a}[V_{n-1}\left(\lfloor\tfrac{u_1+u_2}{2}\rfloor+A,H,A\right)+V_{n-1}\left(\lceil\tfrac{u_1+u_2}{2}\rceil+A,H,A\right)], \\ &\geq^{*} \quad 2V_n(q,h,a), \end{aligned}$$

where we use the fact that the function $e^{\theta(q-u)}$ is convex in u and for a convex function $f(x)$, the following is true.

$$f(x_1)+f(x_2)\geq f\left(\left\lfloor\frac{x_1+x_2}{2}\right\rfloor\right)+f\left(\left\lceil\frac{x_1+x_2}{2}\right\rceil\right).$$

Also $*$ follows since the policies $\lceil\frac{u_1+u_2}{2}\rceil$ and $\lfloor\frac{u_1+u_2}{2}\rfloor$ are feasible for q. The results follows from induction. ∎

Define $G(q,h,a) = V(q,h,a)-V(q-1,h,a)$ and $Z(q,h,a)=e^{\theta q}\mathbb{E}_{h,a}[G(q+A,H,A)]$.

Corollary C-1: $G(q,h,a)$ is nondecreasing in q. ∎

The unconstrained minimizer $u^*(q,h,a)$ (i.e., u do not satisfy $u\in\{0,1,\cdots,q\}$) of Equation (C-1) is the value of u that solves the following inequalities,

$$\alpha Z(u,h,a)\leq\frac{\beta\sigma^2}{h}e^{\theta q}(e^{\theta}-1)\leq\alpha Z(u+1,h,a). \quad \text{(C-3)}$$

It is clear that for (q,h,a) satisfying $\frac{\beta\sigma^2}{h}e^{\theta q}(e^{\theta}-1)<\alpha Z(0,h,a)$, the solution is $u^*(q,h,a)=0$ and for those (q,h,a) satisfying $\alpha Z(u+1,h,a)<\frac{\beta\sigma^2}{h}e^{\theta q}(e^{\theta}-1)$ for all u, $u^*(q,h,a)=\infty$. By convexity, the solution for the constrained problem ($u\in\{0,1,\cdots,q\}$) is, $u(q,h,a)=\min\{q,u^*(q,h,a)\}$. For a given value for the pair (h,a), we have the following monotonicity results.

Theorem C-1: The optimal policy $u(q,h,a) := q-r(q,h,a)$ is nondecreasing in q.

Proof: We need to show that $u(q,h,a)$ is nondecreasing in q. Since (h,a) are fixed, we suppress their dependence. We argue by contradiction. Let there be q_1 and q_2 such that $q_1<q_2$ but $u(q_1)>u(q_2)$. Thus a policy which uses $u(q_2)$ in state q_1 and $u(q_1)$ in state q_2 is feasible. Since $u(\cdot)$ is optimal, it follows that

$$\begin{aligned} & q_1+\frac{\beta\sigma^2}{h}\left(e^{\theta(q_1-u(q_1))}-1\right)+\alpha\mathbb{E}_{h,a}[V(u(q_1)+A,H,A)]< \\ & \quad q_1+\frac{\beta\sigma^2}{h}\left(e^{\theta(q_1-u(q_2))}-1\right)+\alpha\mathbb{E}_{h,a}[V(u(q_2)+A,H,A)] \\ & q_2+\frac{\beta\sigma^2}{h}\left(e^{\theta(q_2-u(q_2))}-1\right)+\alpha\mathbb{E}_{h,a}[V(u(q_2)+A,H,A)]\leq \\ & \quad q_2+\frac{\beta\sigma^2}{h}\left(e^{\theta(q_2-u(q_1)))}-1\right)+\alpha\mathbb{E}_{h,a}[V(u(q_1)+A,H,A)]. \end{aligned}$$

Adding the two equations we get,

$$e^{\theta q_1}(e^{-\theta u(q_1)}-e^{-\theta u(q_2)})<e^{\theta q_2}(e^{-\theta u(q_1)}-e^{-\theta u(q_2)})$$

But since $u(q_1)>u(q_2)$, it implies $q_1>q_2$, which is a contradiction. Thus $u(q_1)\leq u(q_2)$. ∎

Theorem C-2: The optimal rate allocation policy $r(q,h,a)=s-u(q,h,a)$ is nondecreasing in q.

Proof: We need to show that $r(q,h,a)$ is nondecreasing in q. We again suppress the dependence on (h,a). We prove by contradiction. Let there be q_1 and q_2 such that $q_1<q_2$ but $r(q_1)>r(q_2)$. The policy that takes an action $r(q_2)$ in state q_1 and $r(q_1)$ in state q_2 is also feasible. Since $r(\cdot)$ is optimal, it follows that

$$\begin{aligned} & q_1+\frac{\beta\sigma^2}{h}\left(e^{\theta r(q_1)}-1\right)+\alpha\mathbb{E}_{h,a}[V(q_1-r(q_1)+A,H,A)]< \\ & \quad q_1+\frac{\beta\sigma^2}{h}\left(e^{\theta r(q_2)}-1\right)+\alpha\mathbb{E}_{h,a}[V(q_1-r(q_2)+A,H,A)] \\ & q_2+\frac{\beta\sigma^2}{h}\left(e^{\theta r(q_2)}-1\right)+\alpha\mathbb{E}_{h,a}[V(q_2-r(q_2)+A,H,A)]\leq \\ & \quad q_2+\frac{\beta\sigma^2}{h}\left(e^{\theta r(q_1)}-1\right)+\alpha\mathbb{E}_{h,a}[V(q_2-r(q_1)+A,H,A)], \end{aligned}$$

Now by adding the two equations we get,

$$\begin{aligned} & \mathbb{E}_{h,a}[V(q_1-r(q_2)+A,H,A)]-\mathbb{E}_{h,a}[V(q_1-r(q_1)+A,H,A)]> \\ & \quad \mathbb{E}_{h,a}[V(q_2-r(q_2)+A,H,A)]-\mathbb{E}_{h,a}[V(q_2-r(q_1)+A,H,A)] \end{aligned}$$

Since $V(q,h,a)$ is convex in q, the function $\mathbb{E}_{h,a}[V(y+A,H,A)]$ is convex in y. But the above relation contradicts the convexity of $\mathbb{E}_{h,a}[V(y+A,H,A)]$ in y. Hence proved. ∎

Theorem C-3: The function $u^*(q,h,a)$ as defined earlier satisfies the following bounds.

$$\frac{1}{\theta}\ln(f(q,h))-1\leq u^*(q,h,a)\leq\frac{1}{\theta}\ln\left(\frac{\beta\sigma^2}{h\alpha}e^{\theta q}(e^{\theta}-1)\right),$$

where $\eta(h,a)=\mathbb{E}_{h,a}\left[\frac{e^{\theta A}}{H}\right]$ and

$$f(q,h,a)=\frac{\sqrt{1+4\frac{\beta^2\sigma^4}{\alpha h}\eta(h,a)e^{\theta q}(e^{\theta}-1)}-1}{2\beta\sigma^2\eta(h,a)}.$$

. *Proof:* We need to show the desired bounds on $r(q,h,a)$. Consider the functions $G(q,h,a)$ and $V(q,h,a)$. Consider a feasible policy that serves everything, i.e., $u(x)=0$ for all $x\in\mathcal{X}$.

$$V(q,h,a)\leq q+\frac{\beta\sigma^2}{h}(e^{\theta q}-1)+\alpha\mathbb{E}_{h,a}[V(A,H,A)];$$

Also $V(0,h,a)=\alpha\mathbb{E}_{h,a}[V(A,H,A)]$. Since $V(q,h,a)$ is increasing in q, we also have $V(q,h,a)\geq q+\alpha\mathbb{E}_{h,a}[V(A,H,A)]$ independent of the choice of the policy.

Thus $G(q,h,a) \leq 1 + \frac{\beta\sigma^2}{h}(e^{\theta q}-1)$. Convexity of $V(q,h,a)$ in q implies that $G(q,h,a)$ is monotone nondecreasing. Thus as $G(1,h,a) \geq 1$ implies $G(q,h,a) \geq 1$ for all q. Hence for $Z(u,h,a)$ as defined earlier, we have

$$e^{\theta u} \leq Z(u,h,a) \leq e^{\theta u}\left(1+\beta\sigma^2 e^{\theta u}\mathbb{E}_{h,a}\left[\frac{e^{\theta A}}{H}\right]\right).$$

Fix (h,a). Define $\eta := \mathbb{E}_{h,a}\left[\frac{e^{\theta A}}{H}\right]$. Now consider Equation C-3. The bounds for $Z(u,h,a)$ would result in bounds on $u^*(q,h,a)$. First consider the upper bound of $Z(u,h,a)$. This would yield a lower bound on $u^*(q,h,a)$.

$$\alpha e^{\theta u}(1+\beta\sigma^2 e^{\theta u}\eta) \leq \frac{\beta\sigma^2}{h}e^{\theta q}(e^{\theta}-1) \leq \alpha e^{\theta(u+1)}(1+\beta\sigma^2 e^{\theta(u+1)}\eta).$$

Solving the above equations, we get

$$u^*(q,h,a) \geq \frac{1}{\theta}\ln\left(\frac{\sqrt{1+4\frac{\beta^2\sigma^4}{\alpha h}\eta e^{\theta q}(e^{\theta}-1)}-1}{2\beta\sigma^2\eta}\right)-1.$$

Now the lower bound of $Z(u,h,a)$ would yield an upper bound on $u^*(q,h,a)$, i.e., $u \leq \frac{1}{\theta}\ln\left(\frac{\beta\sigma^2}{\alpha h}e^{\theta q}(e^{\theta}-1)\right)$. The function $f(q,h,a)$ as defined earlier is $\frac{\sqrt{1+4\frac{\beta^2\sigma^4}{\alpha h}\eta e^{\theta q}(e^{\theta}-1)}-1}{2\beta\sigma^2\eta}$. ∎

Lemma C-2: Given $\epsilon > 0$, there exists $q^* < \infty$ such that the the optimal number of packets transmitted $r(q,h,a)$ is greater than $\left(\left\lceil\frac{1}{\theta}\ln\left(\frac{\alpha h}{(1-\alpha)\beta\sigma^2(e^{\theta}-1)}\right)\right\rceil\right)-\epsilon$ for $q > q^*$.

Proof: We need to show that the existence of a q^* such that $r(q,h,a)$ is larger than a number for $q > q^*$. Consider the algorithm with $G_0(q,h,a)=0$. This implies that $Z_0(q,h,a)=0$ and hence $u_0^*(q,h,a)=\infty$ for all (q,h,a) and $s_{(i,0)}=\infty$ for $i=\{0,1,2,\cdots\}$. It follows from the algorithm that $G_1(q,h,a)=1$. Thus

$$u_1^*(q,h,a) = q - \left\lceil\frac{1}{\theta}\ln\left(\frac{\alpha h}{\beta\sigma^2(e^{\theta}-1)}\right)\right\rceil;$$

$$r_1(q,h,a) = \min\left\{\left(\left\lceil\frac{1}{\theta}\ln\left(\frac{\alpha h}{\beta\sigma^2(e^{\theta}-1)}\right)\right\rceil\right)^+, q\right\}.$$

Define $L_n(h) = \left(\left\lceil\frac{1}{\theta}\ln\left(\frac{\alpha(1-\alpha^n)h}{(1-\alpha)\beta\sigma^2(e^{\theta}-1)}\right)\right\rceil\right)^+$ and $L_n = \max L_n(h)$. Thus for $q > L_1$, $r_1(q,h,a) = L_1(h)$ for all a. Moreover, $G_2(q,h,a) = 1+\alpha$ for $q > L_1$ and for all (h,a).

Iterating one step further, we get,

$$u_2^*(q,h,a) = q - \left\lceil\frac{1}{\theta}\ln\left(\frac{\alpha(1+\alpha)h}{\beta\sigma^2(e^{\theta}-1)}\right)\right\rceil \text{ if } u_2^*(q,h,a) > L_1.$$

Thus for $q > L_2 + L_1$, $r_2(q,h,a) = L_2(h)$ for all a. Moreover, $G_3(q,h,a) = 1+\alpha+\alpha^2$ for $q > L_2+L_1$ and all (h,a). Thus we get $r_n(q,h,a) = L_n(h)$ for $q > \sum_{k=1}^{n} L_k$. As we know $r_n(q,h,a)$ converges to optimal $r(q,h,a)$ as n goes to infinity and L_N converges to L_∞, find N large enough such that $r(q,h,a) > r_N(q,h,a) - \frac{\epsilon}{2}$, and $L_N > L_\infty - \frac{\epsilon}{2}$ for $q = \sum_{k=1}^{N} L_k$. Define $q^* = \sum_{k=1}^{N} L_k$. Now the result follows form the monotone increasing property of $r(q,h,a)$ in q. ∎

Theorem C-4: As $\beta \to 0$, the solution $u^*(q,h,a) \to 0$ and $u^*(q,h,a) \geq q$ (nothing is transmitted) for all q if $\beta > \frac{e^{\theta}\alpha\sigma^2}{(e^{\theta}-1)(1-\alpha)}$.

Proof: We need to show the behaviour of the optimal policy as β decreases to zero or increases to infinity. As $\beta \to 0$, Equation B-1 implies that the cost of serving decreases to zero except that the constraint should be satisfied. Thus the solution would be to serve as much as possible, i.e., $u(q,h,a)=0$. Thus the action is to transmit all buffered packets. To show the other part, observe from Equation C-3 that it is enough to show $G(q,h,a) \leq \frac{1}{1-\alpha}$. Since $G(q,h,a) \leq \frac{1}{1-\alpha}$ would imply $Z(u,h,a) \leq \frac{e^{\theta u}}{1-\alpha}$ and hence $u^*(q,h,a) > q$. We show the above upper bound for $G(q,h,a)$ by induction. Since $G_0(q,h,a)=0$, if $\beta > \frac{e^{\theta}\alpha\sigma^2}{(e^{\theta}-1)(1-\alpha)}$, then $u_0^*(q,h,a)=\infty$ and $G_1(q,h,a)=1$. Let $G_n(q,h,a) = \frac{1-\alpha^n}{1-\alpha}$. Thus $Z_n(u,h,a) = e^{\theta u}\frac{1-\alpha^n}{1-\alpha}$. This implies that $u_n^*(q,h,a) \geq q$ for all q and hence $G_{n+1}(q,h,a) = \frac{1-\alpha^{n+1}}{1-\alpha}$. By induction hypothesis it follows that $G_n(q,h,a)$ increases to $\frac{1}{1-\alpha}$ and $u^*(q,h,a) \geq q$ for all q. Thus nothing in transmitted for any state vector (q,h,a). ∎

Theorem C-5: (Parametric Monotonicity) The unconstrained minimizer $u^*(q,h,a)$ is monotonically nondecreasing with β.

Proof: We need to prove that the optimal policy $u^*(q,h,a)$ is monotone nondecreasing in β. We introduce the parameter β as a variable in the functions defined earlier to indicate its dependence. Observe that the recursive algorithm stated for $G_n(q,h,a)$ in Section III-D is equivalent to the following recursion (obtained by dividing throughout by β as $\beta > 0$). Initialize $G_0(q,h,a,\beta)=0$. Let $u_n^*(q,h,a,\beta)$ be the value of u that solves the following inequalities,

$$\alpha Z_n(u,h,a,\beta) \leq \frac{1}{h}e^{\theta q}(e^{\theta}-1) \leq \alpha Z_n(u+1,h,a,\beta).$$

Let $s_{(i,n)} = \max\{q : u_n^*(q,h,a,\beta) \geq q-i\}$. The algorithm for computing $u^*(q,h,a)$ as stated in Section III-D can be rewritten as,

- For $q \leq s_{(0,n)}$,
$$G_{n+1}(q,h,a,\beta) = \frac{1}{\beta} + \alpha\mathbb{E}_{h,a}[G_n(q+A,H,A,\beta)]$$
- For $q = s_{(i,n)}+1$ for $i \in \{0,1,2,\cdots\}$,
$$G_{n+1}(q,h,a,\beta) = \frac{1}{\beta} + \frac{\sigma^2}{h}\left(e^{\theta(i+1)} - e^{\theta(i)}\right)$$
- For $q \in \{s_{(i,n)}+2,\cdots,s_{(i+1,n)}\}$ and $i \in \{0,1,2,\cdots\}$,
$$G_{n+1}(q,h,a,\beta) = \frac{1}{\beta} + \alpha\mathbb{E}_{h,a}[G_n(u_n^*(q,h,a)+A,H,A,\beta))]$$
- Further by definition,
$$Z_{n+1}(q,h,a,\beta) = e^{\theta q}\mathbb{E}_{h,a}[G_{n+1}(q+A,H,A,\beta)]$$

We fix (h,a). Using Corollary C-1 (similar result holds for each n as well), it follows that in order to show that $u_n^*(q,h,a,\beta)$ is monotonically nondecreasing in β, it is enough to show that the function $G_n(q,h,a,\beta)$ is nonincreasing in β for all n. We show this by induction. The function $G_0(u,h,a,\beta)=0$. Let $G_n(q,h,a,\beta)$ be nonincreasing in β.

This implies $Z_n(u,h,a,\beta)$ is nonincreasing in β and hence, $u_n^*(q,h,a,\beta)$ is monotone nondecreasing in β. Thus $s_{(i,n)}$ is nondecreasing in β. Also by monotonicity of $u^*(q,h,a)$ in q, $s_{(i+1,n)} > s_{(i,n)}$. The values of β for which $q = s_{(i,n)} + 1$ is an interval. Further, the values of β for which $q \in (s_{(i,n)}+2, \cdots, s_{(i+1,n)})$ is also an interval with the left end point corresponding to $s_{(i+1,n)}$. Now, given (q,h,a), the above recursive algorithm seen as a function of β is,

- For β satisfying $q \in (s_{(i+1,n)}, \cdots, s_{(i,n)}+2)$ and $i \in \{0,1,2,\cdots\}$,

$$G_{n+1}(q,h,a,\beta) = \frac{1}{\beta} + \alpha \mathbb{E}_{h,a}[G_n(q-(i+1)+A, H, A, \beta)],$$

since in this range of β, $u_n^*(q,h,a,\beta) = q-(i+1)$.

- For β such that $q = s_{(i,n)}+1$ and $i \in \{0,1,2,\cdots\}$,

$$G_{n+1}(q,h,a,\beta) = \frac{1}{\beta} + \frac{\sigma^2}{h}\left(e^{\theta(i+1)} - e^{\theta(i)}\right)$$

- For β satisfying $q \le s_{(0,n)}$,

$$G_{n+1}(q,h,a,\beta) = \frac{1}{\beta} + \alpha \mathbb{E}_{h,a}[G_n(q+A, H, A, \beta)]$$

If we show that the following inequalities hold,

i) For β satisfying last bullet above, $\frac{\sigma^2}{h}\left(e^\theta - 1\right) \ge \alpha \mathbb{E}_{h,a}[G_n(q+A,H,A,\beta)]$

ii) For β satisfying first bullet above, $\frac{\sigma^2}{h} e^{\theta(i+1)}\left(e^\theta - 1\right) \ge \alpha \mathbb{E}_{h,a}[G_n(s_{(i+1,n)} - (i+1) + A, H, A, \beta)]$

iii) For β satisfying second bullet above, $\alpha \mathbb{E}_{h,a}[G_n(s_{(i,n)} + 2 - (i+1) + A, H, A, \beta)] \ge \frac{\sigma^2}{h} e^{\theta(i)}\left(e^\theta - 1\right)$

then we are done. Since each of these components are decreasing in β, pasting them together would result in $G_{n+1}(q,h,a,\beta)$ nonincreasing in β and hence the desired result would follow from induction hypothesis.

Over the region where β satisfies item 1, $u_n^*(q,h,a,\beta) \ge q$, i.e.,

$$\alpha Z_n(q,h,a,\beta) = \alpha e^{\theta q} \mathbb{E}_{h,a}[G_n(q+A,H,A,\beta)] \le \frac{\sigma^2}{h} e^{\theta q}\left(e^\theta - 1\right).$$

This implies that inequality 1 is true.

Over the region where β satisfies item 2, $i+1$ is the optimal solution for $q = s_{(i+1,n)}$. Thus,

$$\alpha Z_n(s_{(i+1,n)} - (i+1), h, a, \beta) \le \frac{\sigma^2}{h} e^{\theta(s_{(i+1,n)})}\left(e^\theta - 1\right),$$

implying the inequality of item 2.

Over the region where β satisfies item 3, $i+1$ is the optimal solution for $q = s_{(i,n)} + 1$. Thus,

$$\alpha Z_n(s_{(i,n)} + 1 - (i+1) + 1, h, a, \beta) \le \frac{\sigma^2}{h} e^{\theta(s_{(i,n)}+1)}\left(e^\theta - 1\right),$$

Implying,

$$\alpha e^{\theta(s_{(i,n)}+1-(i+1)+1)} \mathbb{E}_{h,a}[G_n(s_{(i,n)} + 2 - (i+1), h, a, \beta)] \le \frac{\sigma^2}{h} e^{\theta(s_{(i,n)}+1)}\left(e^\theta - 1\right),$$

implying the inequality of item 2.

The result now follows from induction. ∎

Corollary C-2: The optimal policy $r(x)$ is monotone nonincreasing in β. ∎

Consider a special case where the arrival and the channel gain processes are independent and identically distributed (i.i.d.). Consider Equation C-3. Note that under the i.i.d. assumption, the function $Z(q,h,a)$ is independent of (h,a) say $Z(q)$ and $G(q,h,a)$ is independent of a say $G(q,h)$. The unconstrained minimizer $u^*(q,h,a)$ is a function of (q,h) only, say, $u^*(q,h)$. Define $u(y)$, for $y \ge 0$ real, a value of u that solves $Z(u) \le e^{\theta y} \le Z(u+1)$. Then

$$u^*(q,h) = u\left(q - \frac{1}{\theta}\ln\left(\frac{\alpha h}{\beta\sigma^2(e^\theta - 1)}\right)\right). \tag{C-4}$$

This solution is depicted in Figure 4. The following theorem states the optimal solution.

Theorem C-6: Let $u^*(q,h,a)$ be as defined by Equation C-3. If the arrival process $A[n]$ and the channel gain process $H[n]$ are i.i.d., the constrained optimal solution is $u(q,h) = 0$ for $\frac{1}{h} < \frac{\alpha Z(0) e^{-\theta q}}{\beta\sigma^2(e^\theta - 1)}$; $u(q,h) = q$ for $\frac{1}{h} > \frac{\alpha \mathbb{E}[G(q+A,H)]}{\beta\sigma^2(e^\theta - 1)}$ and $u(q,h) = u^*(q,h)$ otherwise.

APPENDIX D
PROOFS OF THEOREMS

Lemma D-1: Given any sequence of discount factors α converging to one, there exists a subsequence α_n converging to 1 such that the average cost optimal policy $u_1(q,h,a)$ *for any* (q,h,a) can be obtained as a pointwise limit of discount optimal policies, i.e., $u_1(q,h,a) = \lim_n u_{\alpha_n}(q,h,a)$.

Proof: Since the state space is countable (q, a are integer valued and h take values from a finite set), we can directly modify the convergence result stated in Theorem A-2 to the following result. Given x and a sequence of discount factors converging to one, there exists a subsequence α_n such that $u_{\alpha_n}(q,h,a) \to u_1(q,h,a)$ as $n \to \infty$ but the subsequence may depend upon the choice of $x = (q,h,a)$.

Enumerate the possible choices of x. Given x_1, let $\{\alpha_{1n}\}$ be a subsequence such that $u_{\alpha_{1n}}(x_1) \to u_1(x_1)$. Take x_2 and find a subsequence $\{\alpha_{2n}\} \subset \{\alpha_{1n}\}$ such that $u_{\alpha_{2n}}(x_2) \to u_1(x_2)$. Also we have $u_{\alpha_{2n}}(x_1) \to u_1(x_1)$. Keep on doing this till the state space is exhausted. By Cantor diagonalization procedure, we get a sequence $\{\alpha_n\}$ such that $u_{\alpha_n}(x) \to u_1(x)$ for all x. ∎

Theorem D-1: [6] A controlled chain $X[n]$ is $c(x,\pi)$ regular if for some state z, $\mathbb{E}_z^\pi[\sum_{n=0}^{n_z-1} c(X[n],\pi)] < \infty$, where n_z is the first return time to z when starting in z and $c(x,r)$ is the one stage cost function as defined earlier. The policy π is regular if $X[n]$ is $c(x,\pi)$ regular. A function $f(x)$ is norm-like if $\{x : f(x) < \Delta\}$ is finite for each Δ finite. Define a resolvent kernel for policy π as $M_\pi = \sum_{t=0}^\infty 2^{-(t+1)} P_\pi^t$ where P_π^t is the t step transition operator under policy π. We need the following conditions.

A1. There exists a regular policy π^{-1}. Let $V_0(x)$ be the value function corresponding to π^{-1}. Further, for the optimal policy π^*, $\lim_n \frac{1}{n} E^{\pi^*} V_0(X[n]) = 0$.

A2. The function $c(x,\cdot)$ is norm-like for each x. Further, there exist a norm-like function $\underline{c}(x)$ such that $c(x,\pi) \ge \underline{c}(x)$ for all regular policies π.

A3. For any action r, there is a positive probability of returning to state z in the next step when starting in state

z. Define $S_0 = \{x : \underline{c}(x) < \bar{\eta}$ where $\bar{\eta}$ is the average cost under policy π^{-1}. Further, we need $M_\pi(x, z) > \delta$ for some $\delta > 0$ and all Markovian policies π and $x \in S_0$. If the value iteration algorithm is initialized with $V_0(x)$ and the conditions A1-A3 holds, then $V_n(z) - V_{n-1}(z)$ converges to the optimal average cost and the limiting policy would be the average cost optimal policy.

Proof of Theorem 3.3

Proof: Define an initializing policy $r^{-1}(q, h, a) = q$. We must first verify conditions A1-A3 of Theorem D-1. Note that the policy is regular. The value function $V_0(x)$ is $q + \frac{\beta\sigma^2}{h}(e^{\theta q} - 1) + \mathbb{E}_{h,a}V[A, H, A]$, where $\mathbb{E}_{h,a}V[A, H, A]$ is a finite number. Thus the assumption A1 holds. The cost function $c(x, r) = q + \frac{\beta\sigma^2}{h}(e^{\theta r} - 1)$ is definitely norm-like in r. Define $\underline{c}(x) = q$, a norm-like function. Thus $c(x, \pi) \geq \underline{c}(x)$ for all π implying A2. Fix $z = (0, h_0, a_0)$, where (h_0, a_0) are those values of (h, a) for which the transition probability matrix for h and a have a positive entry at the diagonal. We assume there exists one such pair. We further assume that the arrivals a can be 0 with a positive probability. We do not need A3 to be satisfied for all policies but only for those policies that would arise during the iteration. Note that the condition required regarding the resolvent kernel would hold if we show the existence of a t such that starting in a state $q \in \{0, \cdots, \bar{\eta}\}$ and any (h, a) we can reach z at time t with a positive probability. Such an irreducibility condition needs to be imposed on the underlying system. Thus the condition A3 is also satisfied under this irreducibility condition. The result now follows from Theorem D-1. ■

Proof of Theorem 4.3

Proof: We need to show that the function $W_n(q, \mathbf{x})$ is monotone increasing in q for each n. The function $V_0(\cdot) = 0$. Thus $W_0(\cdot) = 0$. This implies $V_1(q, \mathbf{x}) = q$ and hence $W_1(q, \mathbf{x}) = q + \mathbb{E}_a[A]$ is increasing in q. Let $W_n(q, \mathbf{x})$ is monotone increasing in q. The two expression within the braces in Equation 22 are monotone nondecreasing and hence the minimum is also nondecreasing. It follows that the function $V_{n+1}(q, \mathbf{x})$ is increasing in q. Thus $W_{n+1}(q, \mathbf{x})$ is increasing in q as it is a convex combination of increasing functions. ■

Proof of Theorem 4.4

Proof: If $q < r$ then Theorem 4.3 implies that the result holds. Thus we look for $q \geq r$ or show that $W_n(q + r, \mathbf{x}) - W_n(q, \mathbf{x})$ is nondecreasing in q for all $q \geq 0$. For $n = 1$, we certainly have a threshold policy since $W_1(q, \mathbf{x}) = q + \mathbb{E}[A]$. By induction hypothesis, let $W_{n-1}(\cdot)$ has the desired property. Extrapolate the function $W_{n-1}(q, \mathbf{x})$ such that $W_{n-1}(q, \mathbf{x}) = W_{n-1}(0, \mathbf{x})$ for $q \in [-r, 0]$.

Fix $\mathbf{x}$ and drop it as an argument. Define $C = \beta(\mathbf{h})(e^{\theta r} - 1)$ and

$$G_n(q) = \min\left\{C + \alpha W_{n-1}((q - r)^+)\ , \alpha W_{n-1}(q)\right\}.$$

It is enough to show that $G_n(q+r) - G_n(q)$ is nondecreasing in q for all nonnegative q. Let q^* be the minimizing threshold at the n^{th} stage. Thus from optimality of q^*, we have,

$$\alpha(W_{n-1}(q^* - 1) - W_{n-1}(q^* - 1 - r)) \leq C < W_{n-1}(q^*) - W_{n-1}(q^* - r).$$

Also,

- For $q \geq q^*$, we have $G_n(q+r) - G_n(q) = \alpha(W_{n-1}(q) - W_{n-1}(q - r))$
- For $q < q^*$ but $q + r \geq q^*$, we have $G_n(q+r) - G_n(q) = C$
- For $q + r < q^*$, we have $G_n(q + r) - G_n(q) = \alpha(W_{n-1}(q + r) - W_{n-1}(q))$

Thus we just need to show that

$$W_{n-1}(q^* - 1) - W_{n-1}(q^* - 1 - r) \leq C \leq W_{n-1}(q^*) - W_{n-1}(q^* - r).$$

But this is true by the very definition of optimality of q^*. Thus the result follows by induction. ■

A. Proof of Theorem 4.5

Proof: We first divide the optimality equation (Eq. 23) by ζ throughout (ζ is positive).

$$V_{n+1}(q, \zeta, a) = \frac{q}{\zeta} + \min\{(e^{\theta r} - 1) + \alpha W_n((q - r)^+, \zeta, a), \alpha W_n(q, \zeta, a)\}.$$

where $W_n(q, \zeta, a)$ is as defined. As a function of ζ, we need to show that the function $W_n(q+r, \zeta) - W_n(q, \zeta)$ is nonincreasing in ζ for all n, then we are done. We use induction. We have $W_1(q, \zeta) = \frac{q}{\zeta}$. The desired property holds for $n = 1$. Let W_{n-1} has the desired property. Thus the threshold $q^*_{n-1}(\zeta)$ for $n - 1$ stage problem is nondecreasing in ζ. Let ζ_1 be the largest value of ζ such that $q \geq q^*(\zeta_1)$ and ζ_2 be the smallest value of ζ such that $q + r < q^*(\zeta_2)$. Thus $\zeta_1 < \zeta_2$.

- For $\zeta \geq \zeta_2$, we have $G_n(q+r, \zeta) - G_n(q, \zeta) = W_{n-1}(q + r, \zeta) - W_{n-1}(q, \zeta)$
- For $\zeta_1 \leq \zeta < \zeta_2$, we have $G_n(q + r, \zeta) - G_n(q, \zeta) = e^{\theta r} - 1$
- For $\zeta < \zeta_1$, we have $G_n(q + r, \zeta) - G_n(q, \zeta) = W_{n-1}(q, \zeta) - W_{n-1}(q - r, \zeta)$

The result follows since for $\zeta > \zeta_2$, $W_{n-1}(q + r, \zeta) - W_{n-1}(q, \zeta) < (e^{\theta r} - 1)$ and for $\zeta < \zeta_1$ we have $W_{n-1}(q, \zeta) - W_{n-1}(q - r, \zeta) > (e^{\theta r} - 1)$. This implies $G_n(q + r, \zeta) - G_n(q, \zeta)$ is nonincreasing in ζ and $V_n(\cdot)$ has the same property. Now monotone nature of the transition probability matrix implies that $W_n(q + r, \zeta) - W_n(q, \zeta)$ is nonincreasing in ζ. Thus by induction hypothesis, $W(q+r, \zeta) - W(q, \zeta)$ is nonincreasing in ζ and hence the threshold $q^*(\zeta)$ is nondecreasing in ζ. ■