Power Constrained and Delay Optimal Policies for Scheduling Transmission over a Fading Channel

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

I. Abstract

We consider an optimal power and rate scheduling problem for a single user transmitting to a base station on a fading wireless link with the objective of minimizing the mean delay subject to an average power constraint. The base station acts as a controller which, depending upon the transmitter buffer lengths and the signal power to interference ratio (SIR) on the uplink pilot channel, allocates transmission rate and power to the user. We provide structural results for an average cost optimal stationary policy under a long run average transmitter power constraint. We obtain a closed form expression relating the optimal policy when the SIR is the best, to the optimal policy for any other SIR value. We also obtain lower and upper bounds for the optimal policy.
Keywords: Power and rate control in wireless networks, Quality of service in wireless networks

II. Introduction

In communication systems, many fundamental problems involve the optimal allocation of resources subject to performance objectives. In a wired network, the crucial resources are the transmission data rates available on the link. Techniques such as flow control, routing and admission control are all centered around allocating these resources. We consider a resource allocation problem that arises in mobile wireless communication systems. Several challenging analytical problems arise because of the special limitations of a wireless link. One is the time varying nature of the multipath channel, and another is the limited battery power available at a typical wireless handset. It is desirable to allocate transmission rates to a user such that the energy used to transmit the information is minimized while keeping errors under control. Most applications, however, also have quality of service (QoS) objectives such as mean delay, delay jitter, and throughput. Thus there is a need for optimal allocation of wireless resources which provides such QoS guarantees subject to the above said error and energy constraints. Various methods for allocating transmission resources are part of most third generation cellular standards. They include adjusting the transmission power, changing the coding rate and varying the spreading gain in a CDMA based system.

The system model in our work is given in Fig. 1 and is explained below. We assume a slotted system where the higher layer presents the data, that arrives over a slot, to the link layer at the end of each slot. The link layer is assumed to have an infinite capacity buffer to hold the data. We assume that the channel gain and any other interference to the system remain fixed over a slot and vary independently from slot to slot. Over a mini-slot (shown as shaded in Fig. 1), the buffer length information is communicated to the receiver/controller, and the user transmits pilot bits at a fixed power level which we refer to as a pilot channel. The receiver estimates the signal to interference ratio (SIR) on the pilot channel. We assume that the estimates are perfect. Depending on the SIR estimates and the buffer length information, the receiver evaluates the optimal transmission rate and power for the current slot and communicates it back to the transmitter. In practice, there are some restrictions on how much these controls can vary. In this paper we assume that the transmitter can transmit at any arbitrary rate and power level. The transmitter removes that much amount of data from the buffer and encodes it at the allocated rate. All this exchange of information and the encoding is assumed to be completed within the time slot shown as shaded in the Fig. 1. After this the transmitter starts to transmit the encoded data.

Goldsmith and Varaiya [3] are probably the first to obtain the optimal power allocation policy for a single link fading wireless channel. Their emphasis was on the optimal physical layer performance, while ignoring the network layer performance such as queueing delay. In recent work, Berry and Gallager [1] have considered a problem similar to ours. They obtained structural results exhibiting a tradeoff between the network layer and physical layer performance, i.e. the optimal power and mean delay. They show that the optimal power vs; 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 this work, we improve upon the results obtained in [1]. We prove the existence of a stationary average optimal policy, and give a closed form expression for the optimal policy for any SIR value in terms of the optimal policy when the SIR is one, i.e., the best SIR. We also provide lower and upper bounds for the optimal rate allocation policy, not obtained by Berry and Gallager.

This paper is organized as follows. In Section III, we give the model of the system under consideration and formulate

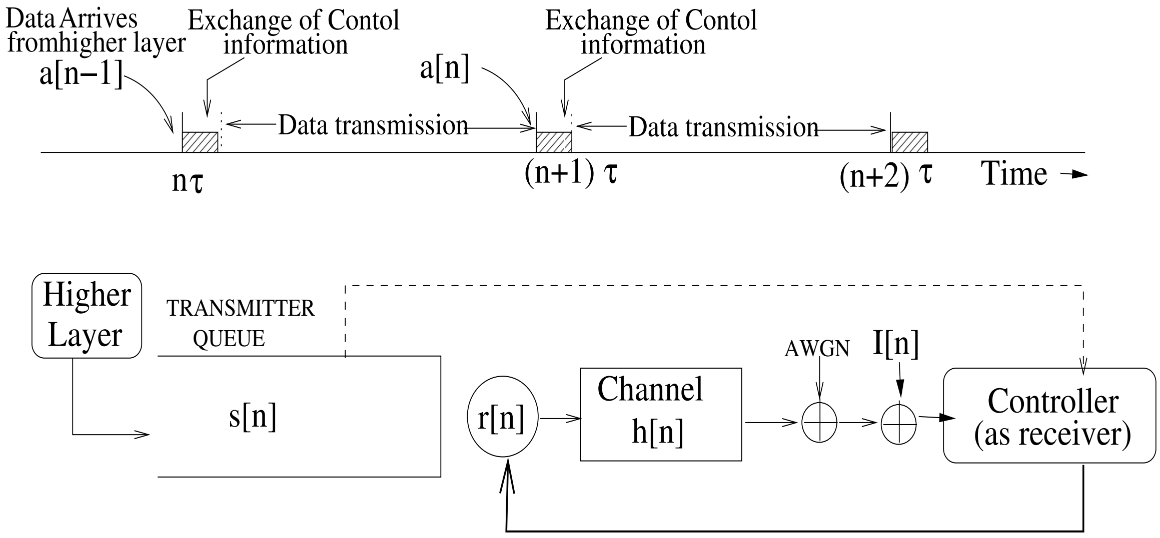

Fig. 1. System model

the controller objective as a constrained optimization problem. Then using the result from [4], we convert it 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 which can be obtained as a limit of discounted cost optimal policies, in Section IV. In Section V, we obtain structural results for the discounted cost optimal policies. We obtain structural results for the average optimal policy in Section VI. Finally, in Section VII, we find conditions under which the hypothesis of the Theorem stated in Section III, holds and hence the existence of a Lagrange multiplier, and the corresponding optimal policy which is also optimal for the original constrained MDP.

III. System Model and Problem Formulation

We consider a discrete time (slotted) fluid model for our analysis, and will later comment on how our results may be used for a packetized model. The slot length is τ (time units), and the n^{th} slot is the interval $[n\tau,(n+1)\tau), n \geq 0$. The data to be transmitted arrives into the system from a higher layer at the end of each slot and is placed into a buffer of infinite capacity (See Figure 1). The arrival process $A[n]$ is assumed to be an independent and identically distributed (iid) sequence; let $F(a)$ be its distribution function. The channel power gain process $H[n]$ is assumed to remain fixed over a slot and vary independently from slot to slot. Any other interference to the system is modelled by the process $I[n]$ which stays constant over a slot and is iid from slot to slot. We further assume that the receiver can correctly estimate the signal to interference ratio (SIR) γ on the "uplink" using a pilot channel. Let σ^2 be the receiver noise power. Without loss of generality, we assume the pilot transmitter power is fixed to σ^2 units. During the n^{th} slot, the SIR $\gamma[n]$ can be written in terms of the current channel gain ($h[n] \in (0,1]$), the receiver noise power, and the current other interference ($i[n] \in [0,\infty)$), as $\gamma[n] = \frac{\sigma^2 h[n]}{\sigma^2 + i[n]}$. Thus the process $\Gamma[n]$ is iid with the distribution function $G(\gamma)$. Further, from the above definition $\gamma[n] \in (0,1]$.

The receiver acts as a controller which, given the buffer state and the SIR, obtains an optimum transmission schedule that minimizes the mean buffer delay subject to a long run average transmitter power constraint $\bar{P}$. The buffer state is available to the receiver at the beginning of each slot. The measurement of SIR and the control decision are taken within a mini-slot shown as shaded and conveyed back to the transmitter also in the same mini-slot. Based on these decisions, the transmitter removes the data from the buffer, encodes it and transmits the encoded data over the channel.

According to our model, if in a frame n the user transmits a signal $y_s[n]$, then the receiver gets

$$y_r[n] = \sqrt{h[n]}y_s[n] + \zeta[n],$$

where $\zeta[n]$ constitutes the additive white Gaussian noise and the other users' interference signal. In this model we assume the external interference to be independent of the system being modelled.

Let, for $n \in \{0,1,2,\cdots\}$, $s[n]$ be the amount of fluid in the buffer at the n^{th} decision epoch and $\gamma[n]$ be the SIR in the n^{th} slot (i.e., the interval $[n\tau,(n+1)\tau)$). Let the state of the system be represented as $x[n] := (s[n],\gamma[n])$. At the n^{th} decision instant, the controller decides upon the amount of fluid $r[n]$ to be transmitted in the current slot depending on the entire history of state evolution, i.e., $x[k]$ for $k = \{0,1,2,\cdots,n\}$. Let $a[n], n \in \{0,1,2,\cdots\}$ be the amount of fluid arriving in the n^{th} slot. Since the amount of fluid transmitted in a slot should be less than the amount in the buffer, i.e., $r[n] \leq s[n]$, for all n, the evolution equation for the buffer can be written as

$$s[n+1] = s[n] - r[n] + a[n].$$

Denote by $X[n]$, $S[n]$, $R[n], n \geq 0$, the corresponding random processes.

The cost of serving r units of fluid in a slot is the total amount of energy required for transmission. We assume N channel symbols in a slot; N is related to the channel bandwidth via Nyquist's theorem. When N is sufficiently large, the power P, required to transmit reliably (i.e., with zero probability of decoding error), is related to the transmission of r units of data in N channel symbols, when the SIR as defined above is γ, by Shannon's formula [2] for the information theoretic capacity, i.e.,

$$r = \frac{1}{\theta} \ln\left(1 + \frac{\gamma P}{\sigma^2}\right),$$

where $\theta = \frac{2\ln(2)}{N}$. Thus when the system state is x, the power required to transmit r units of fluid is

$$P(x,r) = \frac{\sigma^2}{\gamma}(e^{\theta r} - 1).$$

Since, in practice, N is finite, there is positive a probability of decoding error. In section VIII-A, we will comment on how the problem gets modified by incorporating this error probability.

Since delay is related to the amount of data in the buffer by Little's formula [7], the objective is to minimize the mean buffer length. Given $x[0] = x$, the controller's problem is thus to obtain the optimal $r(\cdot)$ that minimizes

$$\limsup_n \tfrac{1}{n} E \textstyle\sum_{k=0}^{n} S[k],$$

subject to,

$$\limsup_n \tfrac{1}{n} E \textstyle\sum_{k=0}^{n} p(X[k], R[k]) \leq \bar{P}.$$

It can be seen from the above objective that the problem has the structure of a constrained Markov decision problem (MDP) [4], which we proceed to formulate in the next section.

A. Formulation as a MDP

Let $\{X[n], n \in \{0,1,2,\cdots\}\}$ denote a controlled Markov chain, with state space $\mathcal{X} = \mathcal{R}^+ \times (0,1]$, and action space $\mathcal{R}^+$, where $\mathcal{R}^+$ denotes the positive real half line. The set of feasible actions in state $x = (s,\gamma)$ is $[0,s]$. 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 Q, where

$$Q(y \in (\mathbf{S}', \Gamma') \subset \mathcal{X} | (x,r)) = \int_{\mathbf{S}'-s+r} dF(a) \int_{\Gamma'} dG(z).$$

Define the mapping $p : \mathcal{K} \to \mathcal{R}^+$ by $p(x,r) = \frac{\sigma^2}{\gamma}(e^{\theta r} - 1)$.

A policy π generates at time n 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 $[0, s[n]]$. Let Π be the space of all such policies. A stationary policy $f \in \Pi$ is a measurable mapping from $\mathcal{X}$ to $[0,s]$. 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,

$$B_x^\pi = \limsup_n \frac{1}{n} E_x^\pi \sum_{k=0}^{n} S[k].$$

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

Given $\bar{P} > 0$, 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,

$$(CP): \quad Minimize \;\; B_x^\pi \;\; subject\ to \;\; \pi \in \Pi_{\bar{P}} \tag{1}$$

The problem (CP) can be converted into a family of unconstrained optimization problem through a Lagrangian approach [4]. For every $\beta > 0$, the Lagrange multiplier, define a mapping $c_\beta : \mathcal{K} \to \mathcal{R}^+$ by,

$$c_\beta(x,r) = s + \beta p(x,r).$$

Define a corresponding Lagrangian functional for any policy $\pi \in \Pi$ by,

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

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: [4] Let, for some $\beta > 0$, $\pi^* \in \Pi$ be the policy that solves the following unconstrained problem (UP_β) defined as,

$$(UP_\beta): \quad Minimize \;\; J_\beta^\pi(x) \;\; subject\ to \;\; \pi \in \Pi \tag{2}$$

Further, if π^* yields the expressions B^{π^*} and K^{π^*} as limits for all $x \in \mathcal{X}$ and $K^{\pi^*} = \bar{P}, \forall x$, then the policy π^* is optimal for the constrained problem (CP).

Proof: (*See [4]*) Note that even though the result is stated in [4] for the countable state space case, the result holds also for the more general situation in our paper so long as we can provide a solution to UP_β with the requisite properties stated in the Theorem. $\triangleleft$

In the subsequent sections, we solve the problem (UP_β) and show in Section VII that the solution satisfies the hypothesis of the Theorem 3.1. The problem (UP_β) is a standard Markov decision problem with an average cost criterion. For ease of notation, we suppress the dependence on the parameter β.

IV. Existence of a Stationary Average Cost Optimal Policy

We consider the average cost problem (UP_β) and define a corresponding discounted cost MDP with discount factor α. We intend to study the average cost problem as a limit of discounted cost problems when the discount factor α goes to one. For initial state x, define

$$V_\alpha(x) = \min_{\pi \in \Pi} E_x^\pi \Big[\sum_{k=0}^{\infty} \alpha^k c_\beta(X[k], R[k]) \Big]$$

as the optimal total expected discounted cost for discount factor α, $0 < \alpha < 1$. $V_\alpha(x)$ is called the value function for the discounted cost MDP.

The following lemma [6] proves the existence of stationary discounted cost optimal policies. We will need Conditions $\mathbf{W}$.

W1. $\mathcal{X}$ is a locally compact space with a countable base.

W2. $R(x)$, the set of feasible actions in state x, is a compact subset of R (the action space), and the multifunction $x \rightarrow R(x)$ is upper semi continuous.

W3. Q is continuous in r with respect to weak convergence in the space of probability measures.

W4. $c(x, r)$ is lower semi-continuous.

Lemma 4.1: [[6], Proposition 2.1] Under Conditions $\mathbf{W}$, 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 4.1: [[6],Theorem 3.8] Suppose there exists a policy Ψ and an initial state $x \in \mathcal{X}$ such that the average cost $J^\Psi(x) < \infty$. Let $\sup_{\alpha<1} w_\alpha(x) < \infty$ for all $x \in \mathcal{X}$ and the Conditions W hold, 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 $x_m \rightarrow x$ such that $f_1(x) = \lim_{m \rightarrow \infty} f_{\alpha_m}(x_m)$. ◁

Remark: In Theorem 4.1, the subsequence of discount factors depends upon the choice of x.

First we verify the Conditions $\mathbf{W}$. Conditions **W1** holds true since the state space is a subset of $\mathcal{R}^2$ which is locally compact with a countable base. The set $R(x) = [0, s]$ is compact and the mapping $x(= (s, \gamma)) \rightarrow [0, s]$ is continuous, thus the condition **W2** holds. Condition **W3** follows from the definition of the transition kernel $Q(\cdot)$ since for distributions on $\mathcal{R}^2$, weak convergence is just convergence in distribution. As the function c is continuous, the condition **W4** follows. This implies the existence of stationary discounted cost optimal policies f_α.

The first hypothesis of Theorem 4.1 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 $x \in \mathcal{X}, x = (s, \gamma)$ we write the discounted cost optimality equation (DCOE) as

$$V_\alpha(x) = \min_{0 \le r \le s} \{c_\beta(x, r) + \alpha \int_{\mathcal{X}} V_\alpha(y) Q(dy|(x, r))\} \quad (3)$$

Given γ, $V_\alpha(s, \gamma)$ is clearly increasing in s since the larger is the initial buffer the larger will be the cost to go. Thus $arg \inf_{x \in \mathcal{X}} V_\alpha(x) = (0, 1) =: x_0$, i.e., the infimum is achieved when the system starts with an empty buffer and the best SIR. Also when the buffer is empty, the set of feasible actions is $\{0\}$. Thus as $c_\beta(x_0, 0) = 0$, we have,

$$V_\alpha(x_0) = \alpha \int_{\mathcal{X}} V_\alpha(y) Q(dy|(x_0, 0)).$$

In addition, considering the policy $r(x) = s$ for all $x \in \mathcal{X}$, we get,

$$V_\alpha(x) \le s + \frac{\beta\sigma^2}{\gamma}(e^{\theta s} - 1) + \alpha \int_{\mathcal{X}} V_\alpha(y) Q(dy|(x, s))$$

But from the definition of Q, it follows that $Q(dy|(x, s)) = Q(dy|(x_0, 0))$. Thus we get,

$$V_\alpha(x) \le s + \frac{\beta\sigma^2}{\gamma}(e^{\theta s} - 1) + V_\alpha(x_0).$$

By the definition of $w_\alpha(x)$ we have $w_\alpha(x) = V_\alpha(x) - V_\alpha(x_0)$ and hence

$$w_\alpha(x) \le s + \frac{\beta\sigma^2}{\gamma}(e^{\theta s} - 1) < \infty \text{ for } x \in \mathcal{X}.$$

Thus all the conditions of Theorem 4.1 are satisfied. *Hence 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 4.1.*

V. Analysis of the Discounted Cost MDP

In this section, we obtain some structural results for the α-discounted optimal policy for each $\alpha \in (0, 1)$. As α is fixed in the analysis that follows in this section, *we suppress the explicit dependence on* α. For a state-action pair $(x = (s, \gamma), r)$ define $u := s - r$, i.e., $u \in [0, s)$ is the amount of data not served when the system state is x. It follows from the definition of Q that $Q(dy|(x, r)) = Q(dy|u)$. Thus we can rewrite the DCOE (Equation 3) in terms of u as,

$$V(x) = \min_{u \in [0,s]} \{s + \frac{\beta\sigma^2}{\gamma}(e^{\theta(s-u)} - 1) + \alpha H(u)\}, \quad (4)$$

where the function $H(u)$ is defined as,

$$H(u) := \int_{0^-}^{\infty} \int_0^1 V(u + a, \gamma) dF(a) dG(\gamma). \quad (5)$$

Lemma 5.1:

(i)) $H(u)$ is convex, and hence $H'(u)$ is increasing.

(ii)) $1 \le H'(u) \le 1 + \theta e \sigma^2 \beta \eta e^{\theta u}$, where

$$\eta = \int_{0^-}^{\infty} \int_0^1 \frac{e^{\theta a}}{\gamma} dF(a) dG(\gamma)$$

.

Proof: See the Appendix. ◁

A. Structure of the optimal policy

Now it can be seen that for each x, the right hand side of the DCOE is a convex programming problem. Using standard techniques for solving a constrained convex programming problem, we obtained the following result for the $u(x)$ that achieves the minimum in Equation 4.

- $u(x) = 0$ for $\{x \in \mathcal{X} : \frac{\beta\theta\sigma^2}{\alpha\gamma} e^{\theta s} \leq H'(0)\}$.
- $u(x) = s$ for $\{x \in \mathcal{X} : H'(s) \leq \frac{\beta\theta\sigma^2}{\alpha\gamma}\}$
- Else $u(x)$ is the solution of $\frac{\sigma^2}{\gamma} e^{\theta s} = e^{\theta u} \frac{\alpha H'(u)}{\beta\theta}$, and $0 < u(x) < s$.

This solution is depicted in Figure 2.

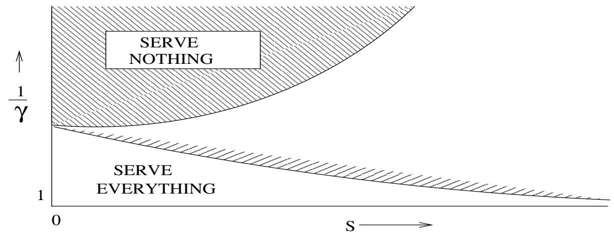

Fig. 2. Characterization of the optimal policy for the discounted cost MDP

Observations:

1) It is optimal not to serve anything when the SIR estimated is low (i.e., $\frac{1}{\gamma}$ is large).
2) When the SIR estimated at the receiver is high, it is optimal to serve everything until a value of the buffer size that increases with r.
3) In the low SIR region, as s increases it becomes optimal to serve data as the delay cost then exceeds the power cost.

B. A state space reduction

Now we show that the optimal policy for any SIR γ can be calculated by knowing the optimal policy when the SIR is 1. Note that $H(u)$ is a function that does not depend upon the policy, and β, θ, α are given constants. Consider two states x_1 and x_2. It follows from the result in Section V-A that except for the case when $u(x) = s$, the controls $u(x_1)$ and $u(x_2)$ are the same if

$$\frac{1}{\gamma_1} e^{\theta s_1} = \frac{1}{\gamma_2} e^{\theta s_2}$$

Thus we can compute the optimal policy $u(x)$ for any $x \in \mathcal{X}$ by knowing the optimal policy for the case when γ is fixed to one and only s is allowed to vary. In order to compute $u(s, \gamma)$, we first obtain s_1 such that $e^{\theta s_1} = e^{\theta s} \frac{1}{\gamma}$ and then the optimal policy $u(s, \gamma)$ can be written in terms of $u(\cdot, 1)$ as,

$$u(s, \gamma) = \min \Big\{ s, u\Big(s + \frac{1}{\theta} \ln\Big(\frac{1}{\gamma}\Big), 1\Big) \Big\}.$$

Thus for subsequent analysis, we can concentrate only on the evaluation of $u(s, 1)$ *and we henceforth call it the optimal policy.* For the notational convenience, we write $u(s, 1)$ as $u(s)$, $r(s, 1)$ as $r(s)$, and $V(s, 1)$ as $V(s)$. Thus from Equation 4, we have,

$$V(s) = \min_{u \in [0,s]} \big\{ s + \beta\sigma^2 (e^{\theta(s-u)} - 1) + \alpha H(u) \big\}. \quad (6)$$

The objective function, being sum of a strictly convex and a convex function, is strictly convex. Thus it has a unique minimizer for each s. Now we obtain some structural results for $u(s)$. Note that $r(s) = s - u(s)$. The following theorems give structural results for the discounted cost optimal policy $u(s)$ obtained as a solution to Equation 6.

Theorem 5.1: The optimal rate allocation policy $r(s) = s - u(s)$ is nondecreasing in s.

Proof: We show this by contradiction. Let there be s_1 and s_2 such that $s_1 < s_2$ but $r(s_1) > r(s_2)$. Thus $r(s_2) > r(s_1) \leq s_1 < s_2$ and hence a policy which uses $r(s_2)$ in state s_1 and $r(s_1)$ in state s_2 is feasible. Since $r(\cdot)$ is optimal, it follows that

$$\begin{aligned} s_1 + \beta\sigma^2(e^{\theta r(s_1)} - 1) + \alpha H(s_1 - r(s_1)) &< \\ s_1 + \beta\sigma^2(e^{\theta r(s_2)} - 1) &+ \alpha H(s_1 - r(s_2)) \\ s_2 + \beta\sigma^2(e^{\theta r(s_2)} - 1) + \alpha H(s_2 - r(s_2)) &< \\ s_2 + \beta\sigma^2(e^{\theta r(s_1)} - 1) &+ \alpha H(s_2 - r(s_1)), \end{aligned}$$

where the strict inequality holds due to uniqueness of the minimizer. Now by adding the two equations we get,

$$\begin{aligned} H(s_1 - r(s_2)) - H(s_1 - r(s_1)) &> \\ H(s_2 - r(s_2)) &- H(s_2 - r(s_1)) \end{aligned}$$

which contradicts the convexity of $H(\cdot)$ as proved in Lemma 5.1. $\triangleleft$

Observe that Theorem 5.1 implies that for any pair (s_1, s_2) satisfying $s_1 < s_2$, we have $s_1 - u(s_1) \leq s_2 - u(s_2)$, i.e., $u(s_2) - u(s_1) \leq s_2 - s_1$.

Theorem 5.2: The optimal policy $u(s) := s - r(s)$ is nondecreasing and

$$\frac{s}{2} - \frac{1}{2\theta} \ln(\kappa_2) \leq u(s) \leq s - \frac{1}{\theta} \ln(\kappa_1),$$

where κ_1 and κ_2 are constants.

Remark: We have $u(s) + r(s) = s$, and we see that both $r(s)$ and $u(s)$ are nondecreasing in s.

Proof: We argue by contradiction. Let there be s_1 and s_2 such that $s_1 < s_2$ but $u(s_1) > u(s_2)$. Thus a policy which uses $u(s_2)$ in state s_1 and $u(s_1)$ in state s_2 is feasible. Since $u(\cdot)$ is optimal, it follows from the uniqueness of the minimizer that

$$\begin{aligned} s_1 + \beta\sigma^2(e^{\theta(s_1 - u(s_1))} - 1) + \alpha H(u(s_1)) &< \\ s_1 + \beta\sigma^2(e^{\theta(s_1 - u(s_2))} - 1) &+ \alpha H(u(s_2)) \\ s_2 + \beta\sigma^2(e^{\theta(s_2 - u(s_2))} - 1) + \alpha H(u(s_2)) &< \\ s_2 + \beta\sigma^2(e^{\theta(s_2 - u(s_1)))} - 1) &+ \alpha H(u(s_1)). \end{aligned}$$

Adding the two equations we get,

$$e^{\theta s_1}(e^{-\theta u(s_1)} - e^{-\theta u(s_2)}) < e^{\theta s_2}(e^{-\theta u(s_1)} - e^{-\theta u(s_2)})$$

But since $u(s_1) > u(s_2)$, it implies $s_1 > s_2$, which is a contradiction. Thus $u(s_1) \leq u(s_2)$.

Now we obtain the bounds on $u(s)$. Using Lemma 5.1(ii) and the results for the optimal policy we get,

$$\kappa_1 e^{\theta u} \leq \sigma^2 e^{\theta s} \leq \kappa_2 e^{2\theta u},$$

where κ_1 and κ_2 take care of all the constants involved. From this the result follows. $\triangleleft$

Corollary 5.1: The sequence $\{u_\alpha(\cdot)\}, \alpha \in (0,1)$ of α-discount optimal policies is globally equi-Lipschitz with parameter 1, i.e., for each pair (s_1, s_2) of the state space and each $\alpha \in (0,1)$ we have,

$$|u_\alpha(s_2) - u_\alpha(s_1)| \leq |s_2 - s_1|.$$

This implies $u(\cdot)$ is differentiable almost everywhere and $u^{'}(s) \leq 1$. $\triangleleft$

VI. The Average Cost Optimal Policy

In this section we provide structural results for the average cost optimal policy. We reintroduce the dependence on the discount factor α of the related $\alpha-$discount optimal policies. Let $u_1(s)$ be the average cost optimal policy and $u_\alpha(s)$ be the $\alpha-$discount optimal policy for $\alpha \in (0,1)$. Consider the result stated in Theorem 4.1. The average cost optimal policy $u_1(s)$ is the limit of discount optimal policies which might be optimal for some close neighbour of s rather than s itself.

Lemma 6.1: Given $s \in [0, \infty)$, let α_n be the subsequence of discount factors as in Theorem 4.1. The average cost optimal policy $u_1(s)$ for any s can be obtained as a pointwise limit of discount optimal policies, i.e., $u_1(s) = \lim_n u_{\alpha_n}(s)$,

Proof: Given $s \in [0, \infty)$ and a sequence of discount factor converging to one, let α_n be the subsequence and s_n be the sequence converging to s such that $u_{\alpha_n}(s_n) \to u_1(s)$ as $n \to \infty$. Note that this holds as a result of Theorem 4.1. Then

$$\begin{aligned} &|u_1(s) - u_{\alpha_k}(s)| \\ &\quad \leq \quad |u_1(s) - u_{\alpha_k}(s_k)| + |u_{\alpha_k}(s_k) - u_{\alpha_k}(s)| \\ &\quad \leq^* \quad |u_1(s) - u_{\alpha_k}(s_k)| + |s_k - s| \to 0 \end{aligned}$$

where $(*)$ follows from the Corollary 5.1. $\triangleleft$

Lemma 6.2: The average cost optimal policy $u_1(s)$ is monotonically nondecreasing.

Proof: Consider $s_1 < s_2$. Using Lemma 6.1, let α_n be the subsequence of discount factors such that $u_{\alpha_n}(s_1) \to u_1(s_1)$. Considering α_n as the original sequence, we can again find a subsequence say α_{n_k} such that $u_{\alpha_{n_k}}(s_2) \to u_1(s_2)$. As $u_\alpha(s)$ is monotonic nondecreasing, we have $u_{\alpha_{n_k}}(s_2) - u_{\alpha_{n_k}}(s_1) \geq 0$ for all k. Now taking the limit as k tends to infinity, we get $u_1(s_2) \geq u_1(s_1)$. Since this is true for any s_1 and s_2, we have the result. $\triangleleft$

The following theorem gives the structural properties of the average cost optimal policy and further strengthen the results of Theorem 4.1.

Theorem 6.1: For the discount factor α_n, let $u_{\alpha_n}(s)$ be the minimizer of the right hand side of Equation 6,

$$V(s) = \min_{u \in [0,s]} \{s + \beta\sigma^2(e^{\theta(s-u)} - 1) + \alpha_n H(u)\}.$$

where $H(u)$ is as in Equation 5.

(i)) Given any sequence of discount factors converging to 1, there exists a subsequence $\{\alpha_n\}$ such that for any $s \in \mathcal{R}^+$, the average cost optimal policy $u_1(s) = \lim_n u_{\alpha_n}(s)$, i.e., the choice of subsequence does not depend upon the choice of s.

(ii)) The optimal policy $u_1(s)$ is monotonic nondecreasing and Lipschitz with parameter one. Also, we have bounds on $u_1(s)$, i.e.,

$$\frac{s}{2} - \frac{1}{2\theta}\ln(\kappa_2) \leq u_1(s) \leq s - \frac{1}{\theta}\ln(\kappa_1).$$

(iii)) Given any $x = (s, \gamma) \in \mathcal{X}$, the average optimal policy $u_1(x)$ representing the amount of data not served when in state x, is

$$u_1(x) = \min\left\{s, u_1\left(s + \frac{1}{\theta}\ln\left(\frac{1}{\gamma}\right)\right)\right\}.$$

Proof:

(i)) Let D_1 be a countable dense subset of $\mathcal{R}^+$. Since u_1 is monotonic, it can at most have countably many discontinuities. Let D_2 be the set of discontinuities. Define a countable set $D := D_1 \cup D_2$. Given $s_1 \in D$, let $\{\alpha_{1i}\}$ be the subsequence such that $u_{\alpha_{1i}}(s_1) \to u_1(s_1)$. Take $s_2 \in D$ and find a subsequence $\{\alpha_{2i}\} \subset \{\alpha_{1i}\}$ such that $u_{\alpha_{2i}}(s_2) \to u_1(s_2)$. Also we have $u_{\alpha_{2i}}(s_1) \to u_1(s_1)$. We keep on doing this till D is exhausted. By Cantor diagonalization procedure, we get a sequence $\{\alpha_n\}$ such that $u_{\alpha_n}(s) \to u_1(s)$ for all $s \in D$. Now take any $s \in \mathcal{R}^+ \setminus D$. Since D is dense in $\mathcal{R}^+$ and $u_1(\cdot)$ is continuous at $s \in \mathcal{R}^+ \setminus D$, given $\epsilon > 0$, there $\exists s_1 \in D$ such that $|s - s_1| < \frac{\epsilon}{3}$ and $|u_1(s) - u_1(s_1)| < \frac{\epsilon}{3}$. Choose N such that $|u_1(s_1) - u_{\alpha_n}(s_1)| < \frac{\epsilon}{3}$ for all $n > N$. Now for all $n > N$, we have,

$$\begin{aligned} |u_1(s) - u_{\alpha_n}(s)| &\leq |u_1(s) - u_1(s_1)| \\ &+ |u_1(s_1) - u_{\alpha_n}(s_1)| + |u_{\alpha_n}(s_1) - u_{\alpha_n}(s)| \\ &\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + |s_1 - s| \qquad \leq \epsilon \end{aligned}$$

Since ϵ is arbitrary, we have a result that given any sequence of discount factors converging to one, there exists a subsequence $\{\alpha_n\}$ such that $u_{\alpha_n}(s) \to u_1(s)$ for $\forall s \in \mathcal{R}^+$.

(ii)) The monotonic nondecreasing property of $u_1(s)$ is shown in Lemma 6.2. The bounds on $u_1(s)$ are obvious from the corresponding bounds on $u_\alpha(s)$ in Theorem 5.2. To prove Lipschitz continuity of $u_1(s)$, let α_n

be the subsequence converging to one as in the proof of (i). Given $\epsilon > 0$ and $s_1, s_2 \in S$, find N_1 and N_2 such that $|u_1(s_1) - u_{\alpha_n}(s_1)| < \frac{\epsilon}{2}$ for all $n > N_1$ and $|u_1(s_2) - u_{\alpha_n}(s_2)| < \frac{\epsilon}{2}$ for all $n > N_2$. Let $N = \max(N_1, N_2)$. Now for $n > N$, we have

$$\begin{aligned}|u_1(s_1) - u_1(s_2)| &\le |u_1(s_1) - u_{\alpha_n}(s_1)| \\ &+|u_{\alpha_n}(s_1) - u_{\alpha_n}(s_2)| + |u_{\alpha_n}(s_2) - u_1(s_2)| \\ &\le \frac{\epsilon}{2} + |s_1 - s_2| + \frac{\epsilon}{2}\end{aligned}$$

Since ϵ is arbitrary, we have the result. ◁

We have given structural results for the optimal policy for the unconstrained problem (UP_β). Now we show that there exists a $\beta > 0$ for which the optimal policy obtained above is also optimal for the constrained problem CP.

VII. The Optimal Policy under a Power Constraint

We reintroduce the dependence on the Lagrange multiplier β. Recollect that the solution to the problem (UP_β) is $r_1^\beta(s,\gamma) = s - u_1^\beta(s,\gamma)$, where $u_1^\beta(s,\gamma)$ is

$$\min\Big\{s, u_1^\beta\Big(s + \frac{1}{\theta}\ln\Big(\frac{1}{\gamma}\Big), 1\Big)\Big\}.$$

We find conditions under which the hypothesis of Theorem 3.1 holds. First we show the existence of a $\beta_0 > 0$ such that the average power cost is equal to the power constraint, i.e., $K^{u_1^{\beta_0}} = \bar{P}$.

As the parameter β is increased, power becomes more expensive and hence the average amount of fluid in the buffer increases. It has earlier been shown in [1] that as the delay increases, the power required decreases and the power-delay tradeoff curve is convex. Thus $K^{u_1^\beta}$ decreases as β increases. Moreover, $K^{u_1^\beta} \to 0$ as $\beta \to \infty$ and tends to infinity as $\beta \to 0$. Thus, each point on the curve can be obtained with a particular β, i.e., there exists a $\beta_0 > 0$ such that $K^{u_1^{\beta_0}} = \bar{P}$. But we have this result only in the $\limsup$ sense. If we can show that for this choice of β, the $\limsup$ and the $\liminf$ are equal then we will satisfy the hypothesis of Theorem 3.1. Since we have a stationary policy $u_1^{\beta_0}(s,\gamma)$, and $\{\Gamma_n\}$ is iid, it is clear that $\{S_n\}$ is a Markov chain on $\mathcal{R}^+$. To show that $u_1^{\beta_0}(s,\gamma)$ yields the expressions $B^{u_1^\beta}$ and $K^{u_1^\beta}$ as limits, it is thus sufficient to show that the controlled chain $\{S_n\}$ is ergodic under this policy.

Using the negative drift argument [5], the following drift condition is sufficient for the ergodicity of the controlled chain $\{S_n\}$ under the policy $u_1^{\beta_0}(s,\gamma)$.

Drift Condition: Given $\epsilon > 0$, there exists an $s_0 < \infty$ such that for all $s > s_0$ the following holds,

$$E\Big(S_{n+1} - S_n | S_n = s\Big) < -\epsilon \tag{7}$$

Theorem 7.1: Suppose there exists $0 < s_0 < \infty$ such that

$$r_1^{\beta_0}(s_0, 1) = s_0 - u_1^{\beta_0}(s_0, 1) > Q,$$

where Q is a constant satisfying

$$E\Big(Q - \frac{1}{\theta}\ln\Big(\frac{1}{\Gamma}\Big)\Big)^+ > E(A) + \epsilon.$$

where A is the arrival random variable. This particular choice of s_0 satisfies the drift condition and thus S_n is ergodic.

Remark: If γ only assumes values close to zero then it is possible that one may not get any finite value for Q.

Proof: By the monotonic nondecreasing nature of $r_1^{\beta_0}(s,\gamma)$, it follows from the hypothesis that for all $s > s_0$,

$$r_1^{\beta_0}(s, 1) > Q$$

As in Theorem 6.1, we have for all $s > s_0$ and all $\gamma \in (0, 1]$,

$$\begin{aligned}&r_1^{\beta_0}(s,\gamma) \\ &= s - \min\Big\{s, u_1^{\beta_0}\Big(s + \frac{1}{\theta}\ln\Big(\frac{1}{\gamma}\Big), 1\Big)\Big\} \\ &= \Big(s - u_1^{\beta_0}\Big(s + \frac{1}{\theta}\ln\Big(\frac{1}{\gamma}\Big), 1\Big)\Big)^+ \\ &= \Big(r_1^{\beta_0}\Big(s + \frac{1}{\theta}\ln\Big(\frac{1}{\gamma}\Big), 1\Big) - \frac{1}{\theta}\ln\Big(\frac{1}{\gamma}\Big)\Big)^+ \\ &> \Big(Q - \frac{1}{\theta}\ln\Big(\frac{1}{\gamma}\Big)\Big)^+\end{aligned}$$

Now we intend to use this lower bound to get an upper bound for the left hand side (LHS) of Equation 7. From Equation 7,

$$\begin{aligned}&E\Big(S_{n+1} - S_n | S_n = s\Big) \\ &= E[A] - \int_0^1 r_1^{\beta_0}(s,\gamma) dG(\gamma) \\ &\le^{(a)} E[A] - E\Big(Q - \frac{1}{\theta}\ln\Big(\frac{1}{\Gamma}\Big)\Big)^+ \\ &< -\epsilon\end{aligned}$$

where the inequality (a) follows from the lower bound on $r_1^{\beta_0}(s,\gamma)$ for all $s > s_0$. Hence proved. ◁

Lemma 7.1: If $Q < \infty$, the hypothesis of Theorem 7.1 is satisfied.

Remark: We will show that the average cost optimal rate $r_1^{\beta_0}(s, 1) \to \infty$ as s goes to ∞.

Proof: See Appendix for the detailed proof. ◁

Lemma 7.1 along with the Theorem 7.1 implies the existence of a stationary measure under the policy $u_1^{\beta_0}$ and that the expressions $B^{u_1^{\beta_0}}$ and $K^{u_1^{\beta_0}}$ are obtained as limits. Thus, in summary, if Q as defined in Theorem 7.1 is finite, there exists a Lagrange multiplier $\beta_0 > 0$ such that $u_1^{\beta_0}(\cdot)$ is the solution to the problem CP.

Now we show how the results obtained using the fluid model can be used for a packetized model. In a packetized model, we assume that a number of packets arrive in each slot and they cannot be fragmented during transmission. Thus we should look for integer solution of optimal rates. One way to do this is to use the floor of the optimal allocated rate. But with this algorithm the power available will be underutilized. Thus we suggest the following algorithm. Let $r(s,\gamma)$ be the optimal policy for the buffer state of s packets. With probability p we use the policy $\lfloor r(s,\gamma)\rfloor$ and with probability $1-p$ we use the policy $\lceil r(s,\gamma)\rceil$. Let K_1 be the average power required when the policy used is $\lfloor r(s,\gamma)\rfloor$ and K_2 when the policy is $\lceil r(s,\gamma)\rceil$. Thus the probability p can be obtained from $pK_1 + (1-p)K_2 = \bar{P}$.

VIII. The Main Result

In this section we give all the results obtained in this work along with all the assumptions.

Theorem 8.1: We assume that η, Q and $L(\alpha,1)$ as defined below are all finite.
Define $\eta = E\left[\frac{e^{\theta A}}{\Gamma}\right]$. Define $L(\alpha,1) = \frac{1}{\theta}\ln\left(\frac{\alpha}{(1-\alpha)\beta\theta\sigma^2}\right)$ for $0<\alpha<1$. Define Q as a solution to

$$E\Big(Q - \frac{1}{\theta}\ln\Big(\frac{1}{\Gamma}\Big)\Big)^+ > E(A) + \epsilon,$$

where ϵ is a positive given number. Let $u_{\alpha_n}(s)$ be the minimizer of the right hand side of the following equation,

$$V(s) = \min_{u\in[0,s]}\big\{s + \beta\sigma^2(e^{\theta(s-u)} - 1) + \alpha_n H(u)\big\}.$$

where $H(u)$ is as in Equation 5.

(i)) Given any sequence of discount factor converging to one, there exists a subsequence $\{\alpha_n\}$ such that for any $s \in S$, the average optimal policy $u_1(s) = \lim_n u_{\alpha_n}(s)$, i.e., the choice of subsequence does not depend upon the choice of s.

(ii)) The optimal policy $u_1(s)$ is monotonic nondecreasing and Lipschitz with parameter one. Also, we have bounds on $u_1(s)$, i.e.,

$$\frac{s}{2} - \frac{1}{2\theta}\ln(\kappa_2) \le u_1(s) \le s - \frac{1}{\theta}\ln(\kappa_1).$$

(iii)) Given any $x = (s,\gamma) \in \mathcal{X}$, the average optimal policy $u_1(x)$ representing the amount of data not served when in state x, is

$$u_1(x) = \min(s, u_1(s + \frac{1}{\theta}\ln\Big(\frac{1}{\gamma}\Big))).$$

(iv)) There exists a Lagrange multiplier $\beta > 0$ such that the optimal policy for (UP_β) is also optimal for the constrained problem (CP) $\triangleleft$

Remark A nearly optimal policy for the packetized model is obtained via the randomization between two fluid optimal policies; Section VII.

A. Decoding Error Possibility

Since the codewords used for transmission are of finite length (N), the formula for $P(x,r)$ used earlier is only a lower bound on the power required to transmit reliably at rate r. Thus if one transmits at $P(x,r)$, there will be a positive probability of decoding error. The bound on the probability of error event is given by the random coding bound. Let P_e be the probability of the error event. We assume that in case of a decoding error, the transmitted data is lost and needs to be retransmitted. The DCOE Equation 4 gets modified to

$$\begin{aligned} V(x) &= \min_{u\in[0,s]}\big\{s + \beta\sigma^2(e^{\theta(s-u)} - 1) \\ &\qquad + \alpha((1-P_e)H(u) + P_e H(s))\big\}. \end{aligned}$$

IX. Conclusion

In this work, we formulated the problem of scheduling communication resources in a fading wireless link in a Markov decision framework. The objective is to minimize the delay in the transmitter buffer subject to an average transmitter power constraint. We showed the existence of stationary average optimal policy and obtained some structural results.

In ongoing work, we are extending the problem to the multiuser case. We also intend to numerically compute the average cost optimal policy and compare the performance with some simple heuristic policies.

X. Acknowledgments

We are thankful to the reviewers for their insightful comments that have helped improve the presentation of this paper, and to Prof. V.S. Borkar for useful discussions.

References

[1] 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.
[2] R.G.Gallager, "Information Theory and Reliable Communication," New York: Wiley, 1968.
[3] 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.
[4] D J. Ma, A.M.Makowski, A.Shwartz, "Estimation and Optimal Control for Constrained Markov Chains," *IEEE Conf. on Decision and Control*, Dec 1986.
[5] S. P. Meyn, R. L. Tweedie, "Markov Chains and Stochastic Stability," Springer-Verlag, London, 1993.
[6] Manfred Schal, "Average Optimality in Dynamic Programming with General State Space," *Mathematics of Operations Research*, vol. 18, no. 1, 163-172, Feb 1993.
[7] R.W.Wolff, "Stochastic Modeling and the Theory of Queues," Prentice Hall, New Jersey, 1988.

APPENDIX

Proof of Theorem 5.1:

(i)) Since $H(u)$ is a convex combination of $V(u+a,\gamma)$, it suffices to show that $V(s,\gamma)$ is convex in s for each γ. We show it through induction on the following value iteration algorithm,

$$V_n(s,\gamma) = \min_{u\in[0,s]}\Big\{s + \frac{\beta\sigma^2}{\gamma}(e^{\theta(s-u)} - 1) \tag{8}$$
$$+\alpha\int_0^\infty\int_0^1 V_{n-1}(u+a,\gamma')dF(a)dG(\gamma')\Big\}. \tag{9}$$

For $n=0$, $V_0(s,\gamma)=0$ hence convex. Assume $V_{n-1}(s,\gamma)$ is convex in s for each γ. Now we fix γ. Let $u(s)$ be the optimal policy in state $x=(s,\gamma)$ for n^{th} iteration. Define $1-\lambda = \bar{\lambda}$ and $s = \lambda s_1 + \bar{\lambda}s_2$.

Let operator E represents averaging with respect to the random variables a and γ. Thus,

$$\begin{aligned}
&\lambda V_n(s_1,\gamma)+\bar{\lambda}V_n(s_2,\gamma)\\
&= s+\frac{\beta\sigma^2}{\gamma}(\lambda e^{\theta(s_1-u(s_1))}+\bar{\lambda}e^{\theta(s_2-u(s_2))})-\frac{\beta\sigma^2}{\gamma}\\
&\quad +\alpha E\Big[\lambda V_{n-1}(u(s_1)+a,\gamma')\\
&\qquad +\bar{\lambda}V_{n-1}(u(s_2)+a,\gamma')\Big]\\
&\geq s+\frac{\beta\sigma^2}{\gamma}(e^{\theta(\lambda(s_1-u(s_1))+\bar{\lambda}(s_2-u(s_2)))}-1)\\
&\quad +\alpha E\Big[V_{n-1}(\lambda u(s_1)+\bar{\lambda}u(s_2)+a,\gamma')\Big]\\
&\geq^* s+\frac{\beta\sigma^2}{\gamma}(e^{\theta(s-u(s))}-1)\\
&\quad +\alpha E\Big[V_{n-1}(u(s)+a,\gamma')\Big]\\
&= V_n(\lambda s_1+(1-\lambda)s_2),\gamma)
\end{aligned}$$

where the inequality $(*)$ follows from the fact that the policy $u(s)$ is optimal while $\lambda u(s_1)+(1-\lambda)u(s_2)$ is feasible for the state $s=\lambda s_1+(1-\lambda)s_2$. The last equality follows from the definition. Thus we have shown that the value function $V(s,\gamma)$ is convex in s for each γ.

(ii)) To prove the upper bound on $H'(u)$, consider a feasible policy that serves everything , i.e., $u(x)=0$ for all $x\in\mathcal{X}$. For this policy we have,

$$V(x)\leq s+\frac{\beta\sigma^2}{\gamma}(e^{\theta s}-1)+\alpha H(0).$$

Since $H(u)$ is increasing, we also have $V(x)\geq s+\alpha H(0)$ independent of the choice of the policy. Define

$$\eta:=\int_{0^-}^{\infty}\int_0^1\frac{e^{\theta a}}{\gamma}dF(a)dG(\gamma).$$

Now using the bounds on $V(x)$ and the Equation 5 we get,

$$H(u)\leq u+E[A]+\eta\beta\sigma^2e^{\theta u}+\alpha H(0)$$

and

$$H(u)\geq u+E[A]+\alpha H(0),$$

Since $H(u)$ is convex, it has a line of support at each point. Let $H(\cdot)$ be differentiable at u, thus for any $u_1>u$, we have

$$H(u_1)-H(u)\geq H'(u)(u_1-u)$$

. Using the above said upper and lower bounds on $H(\cdot)$, we have

$$H'(u)(u_1-u)\leq u_1-u+\beta\eta\sigma^2e^{\theta u_1}.$$

Since u_1 can take any value greater than u, we can minimize the right hand side to get an even tighter bound. Thus we have

$$H'(u)\leq 1+\beta\eta\sigma^2e^{\theta u}\min_{z>0}\{\frac{e^{\theta z}}{z}\}.$$

Solving this we get

$$H'(u)\leq 1+\theta\beta e\sigma^2\eta e^{\theta u}.$$

Now we show that $H'(u)\geq 1$. When the buffer is empty, the optimal policy is $u(0,\gamma)=0$ for all γ. Thus $V(0,\gamma)=\alpha H(0)$. Also $V(s,\gamma)\geq s+\alpha H(0)$. Thus we have

$$V'(0,\gamma)=\lim_{\epsilon\to 0}\frac{V(\epsilon,\gamma)-V(0,\gamma)}{\epsilon}\geq 1.$$

Since $V(s,\gamma)$ is convex in s, $V'(s,\gamma)$ is nondecreasing in s. Thus $V'(s,\gamma)\geq 1$ for all (s,γ). Also as $\eta<\infty$, using the dominated convergence theorem we can take the differentiation inside of the integral sign in Equation 5. Thus $H'(u)\geq 1$. ◁

Proof of Lemma 7.1: Let $r_\alpha(\cdot)$ be the $\alpha-$discounted cost optimal policy. We consider Equations 4, 5 and 8 and rewrite them here in value iteration form for convenience.

$$V_n(s,\gamma)=\min_{u\in[0,s]}\Big\{s+\frac{\beta\sigma^2}{\gamma}(e^{\theta(s-u(s,\gamma))}-1)+\alpha H_{n-1}(u)\Big\}.$$

$$H_{n-1}(u)=\int_0^{\infty}\int_0^1V_{n-1}(u+a,\gamma')dF(a)dG(\gamma').$$

We first show that given $\epsilon>0$, there exists an $s_\infty<\infty$ such that the optimal rate $r_\alpha(s,1)$ satisfies the following,

$$r_\alpha(s,1)>\frac{1}{\theta}\ln\Big(\frac{\alpha}{(1-\alpha)\beta\theta\sigma^2}\Big)-\epsilon,\quad\forall\quad s>s_\infty.$$

Define

$$\frac{1}{\theta}\ln\Big(\frac{\alpha\gamma}{(1-\alpha)\beta\theta\sigma^2}\Big)=:L(\alpha,\gamma).$$

We show this using the induction procedure on the value iteration algorithm. We initialize the algorithm with $H_0(u)=\left(\frac{1}{1-\alpha}\right)u$. Thus using the result for the optimal policy we get,

$$r_{\alpha,1}(s,\gamma)=L(\alpha,\gamma),\quad\forall\quad s>L(\alpha,\gamma).$$

Since $L(\alpha,\gamma)$ is increasing in γ, the above is also true for all $s>L(\alpha,1)$. The new value function for all $s>L(\alpha,1)$ is,

$$\begin{aligned}
&V_1(s,\gamma)\\
&= s+\frac{\beta\sigma^2}{\gamma}\Big(\frac{\alpha\gamma}{(1-\alpha)\beta\theta\sigma^2}-1\Big)\\
&\quad +\Big(\frac{\alpha}{1-\alpha}\Big)\Big(s-L(\alpha,\gamma)\Big).\\
&= \Big(\frac{1}{1-\alpha}\Big)s+\rho_1(\alpha,\gamma)
\end{aligned}$$

where $\rho_1(\alpha,\gamma)$ takes care of all the terms not involving s. Now from the definition of $H_{n-1}(\cdot)$, the value of $H_1(u)$ for all $u>L(\alpha,1)$ is,

$$H_1(u)=\Big(\frac{1}{1-\alpha}\Big)u+s_1,$$

where s_1 takes care of all the constants terms. Using the induction argument, let for all $u > (n-1)L(\alpha, 1)$,

$$H_{n-1}(u) = \Big(\frac{1}{1-\alpha}\Big)u + s_{n-1}.$$

Then the optimal rate for the n^{th} iteration is

$$r_{\alpha,n}(s,\gamma) = L(\alpha,\gamma), \quad \forall \quad s > nL(\alpha, 1).$$

Then for all $s > nL(\alpha, 1)$ we have,

$$\begin{aligned} &V_n(s,\gamma) \\ &= s + \frac{\beta\sigma^2}{\gamma}\Big(\frac{\alpha\gamma}{(1-\alpha)\beta\theta\sigma^2} - 1\Big) \\ &\quad + \Big(\frac{\alpha}{1-\alpha}\Big)\Big(s - L(\alpha,\gamma)\Big) + \alpha s_{n-1}. \\ &= \Big(\frac{1}{1-\alpha}\Big)s + \rho_n(\alpha,\gamma) \end{aligned}$$

Now the value of $H_n(u)$ for all $u > nL(\alpha, 1)$ is,

$$H_1(u) = \Big(\frac{1}{1-\alpha}\Big)u + s_n,$$

where s_1 takes care of all the constants terms.

Thus the induction procedure show that for each finite n there exists a a finite real number $nL(\alpha, 1)$ such that the rate equals $L(\alpha,\gamma)$ for all $s > nL(\alpha, 1)$. But we can't say this as n tends to infinity. Since we have pointwise convergence of $r_{\alpha,n}(s,\gamma)$ to the optimal policy for each α, given $\epsilon > 0$, there exists a finite n and a finite real number s_∞ such that for all $m > n$,

$$r_{\alpha,m}(s_\infty,\gamma) > L(\alpha,\gamma) - \epsilon.$$

Now as we know that the policies $r_{\alpha,m}$ are monotonic nondecreasing in s, we have the result that $r_\alpha(s, 1) > L(\alpha, 1) - \epsilon$ for all $s > s_\infty$.

Now as we let α tend to one, we get the average optimal policy. But as α goes to one, the function $L(\alpha, 1)$ goes to infinity. A similar argument as above proves the Lemma because if Q is finite, we can always find a real number s_0 such that $r_1(s, 1) > Q$ for all $s > s_0$. Simple contradiction argument can also be used to show the result. $\triangleleft$