Int. J. Communications, Network and System Sciences, **2012, 5, 671-677**
http://dx.doi.org/10.4236/ijcns.2012.510069 Published Online October 2012 (http://www.SciRP.org/journal/ijcns)

Structural Properties of Optimal Scheduling Policies for Wireless Data Transmission

Nomesh Bolia[1], Vidyadhar Kulkarni[2]
[1]Department of Mechanical Engineering, Indian Institute of Technology, Delhi, India
[2]Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, USA
Email: nomesh@mech.iitd.ac.in, vkulkarn@email.unc.edu

Received August 15, 2012; revised September 12, 2012; accepted October 8, 2012

ABSTRACT

We analyze a cell with a fixed number of users in a time period network. The base station schedules to serve at most one user in a given time period based on information about the available data rates and other parameter(s) for all the users in the cell. We consider infinitely backlogged queues and model the system as a Markov Decision Process (MDP) and prove the monotonicity of the optimal policy with respect to the "starvation age" and the available data rate. For this, we consider both the discounted as well as the long-run average criterion. The proofs of the monotonicity properties serve as good illustrations of analyzing MDPs with respect to their optimal solutions.

Keywords: MDP; Scheduling; Structural Properties

1. Introduction

We consider a fixed set of N mobile data users in a wireless cell served by a single base station and focus on the downlink channel. The base station maintains a separate queue of data for each user. Time is slotted and in each slot (time period in the standard MDP terminology) the base station can transmit data to exactly one user. Let $R_u^n\,(u=1,2,\cdots,N;n=1,\cdots)$ be the channel rate of user u during time period n, *i.e.*, the amount of data that can be transmitted to user u during time period n by the base station. We assume that the base station knows at all time periods n the vector $R^n=R_1^n,R_2^n,\cdots,R_N^n$. How this information is gathered depends on the system in use. An example of a resource allocation system widely known and used in practice is the CDMA2000 1xEV-DO system [1]. A good description of how this information is generated is also provided in [1]. A good framework for resource allocation and related issues in this (and more general) setting can be found in [2].

There are two objectives to be fulfilled while scheduling the data transfer. The first is to obtain a high data transfer rate. This can be achieved by serving a user u in period n whose channel rate R_u^n is the highest, *i.e.*, following a myopic policy. However if we follow the myopic policy, we run the risk of severely starving users whose channel rate is low for a long time. The second objective is to ensure that none of the users is severely starved. Thus these are conflicting objectives and any good algorithm tries to achieve a "good" balance between the two. We have proposed MDP based scheduling policies in [3] to achieve this balance. In this paper, for the sake of completeness we first describe the MDP framework (and our heuristic policies) and then analyze the important monotonicity properties of the (MDP-) optimal and our recommended policies.

Literature Survey

This problem of scheduling users for data transmission in a wireless cell has been considered in the literature mostly in the last decade and a half. One of the most widely used algorithms that takes advantage of multiuser diversity (users having different and time-varying rates at which they can be served data) while at the same time being fair to all users is the Proportional Fair Algorithm (PFA) of Tse [4]. When each user always has data to be served waiting at the base station (infinitely backlogged queues), the PFA performs well and makes good use of the multiuser diversity. However, it has been proven to be unstable when data isn't always available to be served to each user, and instead, there is external data arrival [5]. Most of the algorithms in this setting are not necessarily outcomes of any optimization framework. In our earlier publication [3], we take a novel approach to solving this problem. This approach develops the scheduling algorithm as an outcome of a systematic optimization framework. Therefore, Bolia and Kulkarni [3] develop MDP and policy improvement based scheduling policies. These policies are easy to implement, and shown to per-

form better [3] than existing policies.

However, while our recommended policies despite being sub-optimal exhibit better results than existing policies [3], our past work lacks any results about structural properties of both the recommended as well as the optimal policies. We believe it is important to establish some such properties to either gain further insight into the problem. Therefore, in this correspondence we prove some *monotonicity properties* of the optimal policies and the policies proposed in [3]. We first define and describe these monotonicity properties in the paper. Our contributions are two fold: A rigorous analysis of these properties along with the observation that our recommended policy in [3] is also monotone (thus being in line with the optimal policy at least with respect to some basic properties) and an illustration of analysis of optimal policies in the MDP framework. These ideas can serve as a good starting point and provide broad guidelines to analyze structural properties of MDPs.

The rest of the article is organized as follows: Section 2 describes the model and index policy and Section 3 proves monotonicity of the optimal policy in this setting. Section 3.3 extends the results for these properties to the long-run average criterion. We conclude the paper with remarks on possible extensions in Section 4.

2. The Model

In this section, for the sake of completeness, we start with a description of the stochastic model [3] for the multidimensional stochastic process $\{R^n, n \geq 1\}$ that represents the channel rates of all users in the cell. Let X_u^n be the channel state of user u at time n. This represents various factors such as the position of the user in the cell, the propagation conditions, etc. and determines the channel rate R_u^n of user u as described below. We assume that $\{X_u^n, n \geq 1\}$ is an irreducible Discrete Time Markov chain (DTMC) on state space $\Omega = \{1, 2, \cdots, M\}$ with Transition Probability Matrix (TPM) $P^u = \left[p^u_{i_u, j_u} \right]$. Note that the TPM can in general be different for different users. Further, as indicated in [1], a set of $M = 11$ fixed data rates is what is available to users in an actual system. For each $u = 1, 2, \cdots, N$, let r_k be the fixed data rate (or channel rate) associated with state $k \in \Omega$ of the DTMC $\{X_u^n, n \geq 1\}$. Thus, when $X_u^n = k$, the user u can receive data from the base station at rate $R_u^n = r_k$ if it is chosen to be served. Thus $\{R_u^n : n \geq 1\}$ is a Markov Chain with state space $r = [r_1, r_2, \cdots, r_M]$, *i.e.*, the vector of all fixed data rates. We assume, without loss of generality, that $r_1 \leq r_2 \leq \cdots \leq r_M$. Let $X^n = [X_1^n, \cdots, X_N^n]$ be the state vector of all the users. We assume the users behave independently of each other and that each user has ample data to ber served. This setting where each user always has ample data to be served is referred to as the "infinitely backlogged queues setting". Since each component of $\{X^n, n \geq 1\}$ is an independent DTMC on Ω, it is clear that $\{X^n, n \geq 1\}$ itself is a DTMC on Ω^N.

Let Y_u^n be the "starvation age" (or simply "age") of the user u at time n, defined as the time elapsed (in number of periods) since the user u was served most recently. Thus, the age of the user is zero at time $n+1$ if it is served in the n^{th} time period. Furthermore, for $m \geq 1$, if the user was served in time period n and it is not served for the next m time periods, its age at time $n+m$ is $m-1$. Let $Y^n = [Y_1^n, \cdots, Y_N^n]$ be the age vector (vector of ages of all users) at time n. The base station serves exactly one user in each time period. Let $v(n)$ be the user served in the n^{th} time period. The age process evolves according to

$$Y_u^{n+1} = \begin{cases} Y_u^n + 1 & \text{if } u \neq v(n), \\ 0 & \text{if } u = v(n). \end{cases} \quad (1)$$

The "state of the system" at time n is given by $[X^n, Y^n] \in \Omega^N \times Z^N$, where $Z = \{0, 1, 2, \cdots\}$. The "state" is thus a vector of $2N$ components and we assume that it is known at the base station in each time period. After observing $[X^n, Y^n]$ the base station decides to serve one of the N users in the time period n. We need a reward structure in order to make this decision optimally. We describe such a structure below. If we serve user u in the n^{th} time period, we earn a reward equal to $R_u^n = r_{X_u^n}$ for this user and none for the others. In addition, there is a cost of $D_l(y)$ if user l of age y is not served in period n. This cost corresponds to the penalty incurred due to "starvation" of the user(s) not served in a given time period. Clearly, we can assume $D_l(0) = 0$ since there is no starvation at age zero. Thus the net reward of serving user u at time n is

$$R_u^n - \sum_{l \neq u} D_l\left(Y_l^n\right). \quad (2)$$

We assume that there is no cost in switching from one user to another from period to period. This is not entirely true in practice, but including switching costs in the model will make the analysis intractable. For convenience we use the notation $W_u^n = \sum_{l \neq u} D_l\left(Y_l^n\right)$. The problem of scheduling a user in a given time period can now be formulated as a Markov Decision Process (MDP). The decision epochs are $\{1, 2, \cdots\}$. The state at time n is $[X^n, Y^n]$ with Markovian evolution as described above. The action space in every state is $A = \{1, 2, \cdots, N\}$ where action u corresponds to serving the user u. The reward in state $[X^n, Y^n]$ corresponding to action u is $R_u^n - W_u^n$.

For the sake of notational convenience, let t^u and $W_u(t)$ be defined as follows:

$$t^u = (t_1+1,\cdots,t_{u-1}+1,0,t_{u+1}+1,\cdots,t_N+1) \quad (3)$$

and,

$$W_u(t) = \sum_{l\neq u} D_l(t_l). \quad (4)$$

Let α be the discounting rate for the MDP [6]. Then, the standard Bellman equation for the discounted reward model is

$$V(i,t) = \max_{u=1,2,\cdots,N}\left[r_{i_u} - W_u(t) + \alpha h\left(i,t^u\right)\right], \quad (5)$$

where

$$h(i,t) = \sum_{j\in\Omega^N} p_{ij} V(j,t) \quad (6)$$

Let $dec(i,t)\in A$ be the optimal decision made (*i.e.*, the user served) in state $[i,t]$. Then,

$$dec(i,t) = \arg\max_{u=1,2,\cdots,N}\left[r_{i_u} - W_u(t) + \alpha h\left(i,t^u\right)\right].$$

Further, let

$$dec_k(i,t) = \arg\max_{u=1,2,\cdots,N}\left[r_{i_u} - W_u(t) + \alpha h_k\left(i,t^u\right)\right] \quad (7)$$

be the optimal decision at the k^{th} step of the value iteration scheme given by (8).

We use the following notation: For any real valued function $f(i,t)$ defined on $\Omega^N\times Z^N$, $f\downarrow t$ denotes that f decreases in every component of t.

3. Monotonicity of Optimal Policy

Although solving Equation (5) to optimality is infeasible, we can derive some important characteristics of the optimal policy. In this section, we consider two monotonicity properties of the optimal policy. We first consider monotonicity in age.

3.1. Monotonicity in Age

The intuition behind monotonicity is as follows. The penalty accrued for each user in a given time period is an increasing function of its current age. Hence we expect the propensity of the optimal policy serving any given user to increase with its age, *i.e.*, if the optimal policy serves a user u in the state $[i,t]$, it will serve user u in state $[i,t+e_u]$ as well, where e_u denotes an N-dimensional vector with the u^{th} component 1 and all other components 0.

Theorem 3.2 states and proves this monotonicity property of the optimal policy for discounted reward. Then we show that standard MDP theory [6] implies the result holds in the case of long-run average reward as well.

We will need the following result to prove theorem 3.2.

Theorem 3.1 $V(i,t)\downarrow t$.

Proof. The standard value iteration equations of (5) are given by

$$V_{k+1}(i,t) = \max_{u=1,2,\cdots,N}\left[r_{i_u} - W_u(t) + \alpha h_k\left(i,t^u\right)\right], \quad k\geq 0. \quad (8)$$

where

$$h_k(i,t) = \sum_j p_{ij} V_k(j,t), \quad (9)$$

and $V_0(i,t) = 0$. We have

$$\lim_{k\to\infty} V_k(i,t) = V(i,t), \quad [i,t]\in\Omega^N\times Z^N. \quad (10)$$

We will prove $V_k(i,t)\downarrow t$ using induction on k. Then the theorem follows from the above equation.

Note that $V_k(i,t)\downarrow t$ holds at $k=0$ since $V_0(i,t)=0$. Assume $V_k(i,t)\downarrow t$ for some $k\geq 0$. We prove $V_{k+1}(i,t)\downarrow t$. It is enough to prove that

$$V_{k+1}(i,t) - V_{k+1}(i,t+e_1)\geq 0, \quad (11)$$

since the proof for all components other than 1 follows similarly. Note that $V_k\downarrow t \Rightarrow h_k\downarrow t$. We consider four cases:

***Case* 1**: $dec_k(i,t)=1$ and $dec_k(i,t+e_1)=1$. From (8),

$$\begin{aligned} &V_{k+1}(i,t) - V_{k+1}(i,t+e_1)\\ &= \left[r_{i_1} - W_1(t) + \alpha h_k\left(i,t^1\right)\right]\\ &-\left[r_{i_1} - W_1(t+e_1) + \alpha h_k\left(i,(t+e_1)^1\right)\right]\geq 0, \end{aligned} \quad (12)$$

since $W_v(t+e_v) = W_v(t)$ and $(t+e_v)^v = t^v$ using Equations (3) and (4).

***Case* 2**: $dec_k(i,t)=1$ and $dec_k(i,t+e_1)=u\neq 1$. From (8), and using $W_u(t+e_1)\geq W_u(t)$ and $(t+e_1)^u = t^u + e_1$, we have

$$\begin{aligned} &V_{k+1}(i,t) - V_{k+1}(i,t+e_1)\\ &= \left[r_{i_1} - W_1(t) + \alpha h_k\left(i,t^1\right)\right]\\ &-\left[r_{i_u} - W_u(t+e_1) + \alpha h_k\left(i,(t+e_1)^u\right)\right]\\ &\geq \left[r_{i_1} - r_{i_u}\right] + \left[W_u(t) - W_1(t)\right]\\ &+\alpha\left[h_k\left(i,t^1\right) - h_k\left(i,(t^u+e_1)\right)\right]\\ &\geq \left[r_{i_1} - r_{i_u}\right] + \left[W_u(t) - W_1(t)\right]\\ &+\alpha\left[h_k\left(i,t^1\right) - h_k\left(i,t^u\right)\right]\geq 0. \end{aligned} \quad (13)$$

The second inequality holds because $h_k\downarrow t$ and the last inequality holds because $dec_k(i,t)=1$.

***Case* 3**: $dec_k(i,t)=u\neq 1$ and $dec_k(i,t+e_1)=u$. From (8),

$$\begin{aligned}
&V_{k+1}(i,t)-V_{k+1}(i,t+e_1)\\
&=\left[r_{i_u}-W_u(t)+\alpha h_k(i,t^u)\right]\\
&-\left[r_{i_u}-W_u(t+e_1)+\alpha h_k\left(i,(t+e_1)^u\right)\right]\\
&\geq\left[W_u(t+e_1)-W_u(t)\right]\\
&+\alpha\left[h_k(i,t^u)-h_k\left(i,(t^u+e_1)\right)\right]\geq 0,
\end{aligned}\tag{14}$$

using the same arguments as in Case 2.

***Case* 4**: $dec_k(i,t)=u\neq 1$ and $dec_k(i,t+e_1)=u$. From (8), and using $W_v(t+e_1)\geq W_v(t)$ and $(t+e_1)^v=t^v+e_1$, we have

$$\begin{aligned}
&V_{k+1}((i,t)-V_{k+1}(i,t+e_1)\\
&=\left[r_{i_u}-W_u(t)+\alpha h_k(i,t^u)\right]\\
&-\left[r_{i_v}-W_v(t+e_1)+\alpha h_k\left(i,(t+e_1)^v\right)\right]\\
&\geq\left[r_{i_u}-r_{i_v}\right]+\left[W_v(t)-W_u(t)\right]\\
&+\alpha\left[h_k(i,t^u)-h_k\left(i,(t^v+e_1)\right)\right]\\
&\geq[r_{i_u}-r_{i_v}]+\left[W_v(t)-W_u(t)\right]]\\
&+\alpha\left[h_k(i,t^u)-h_k(i,t^v)\right]\\
&\geq 0.
\end{aligned}\tag{15}$$

The last inequality holds because $dec_k(i,t)=u$.

Clearly, Cases 1 - 4 are exhaustive and thus Equations (12) through (15) prove that $V_{k+1}(i,t)\downarrow t$, thus completing our induction argument. Hence $V_k\downarrow t$ for all k. This completes the proof.

Now we move on to the main theorem of this section that says that the decision to serve a user in any time period is monotone in age.

Theorem 3.2 $dec(i,t)=v\Rightarrow dec(i,t+e_v)=v$.

Proof. Since $dec(i,t)=v$ we have,

$$\begin{aligned}
&r_{i_v}-W_v(t)+\alpha h(i,t^v)\\
&\geq r_{i_u}-W_u(t)+\alpha h(i,t^u),\quad u\in A.
\end{aligned}\tag{16}$$

To prove $dec(i,t+e_v)=v$, we need to prove

$$\begin{aligned}
&\left[r_{i_v}-r_{i_u}\right]+\left[W_u(t+e_v)-W_v(t+e_v)\right]\\
&+\alpha\left[h\left(i,(t+e_v)^v\right)-h\left(i,(t+e_v)^u\right)\right]\geq 0,
\end{aligned}\tag{17}$$

which follows from (16) (and $t^v=(t+e_v)^v$), and the results that $W_v(t+e_v)=W_v(t)$, $W_u(t+e_v)\geq W_u(t)$ [using Equation (4)] and $h\left(i,(t+e_v)^u\right)<h(i,t^u)$ [using theorem 3.1 and Equation (6)].

This theorem implies that if it is optimal to serve a given user, say v, in a given state $[i,t]$ of the system, it is optimal to serve the same user when everything is identical except the age of the same user (v) increases by one. Thus, everything else remaining constant, the optimal policy is monotone in the age of the users, a result we expect intuitively for any reasonable scheduling policy, but now proved rigorously for the optimal policy. Similarly, the rest of the theorems in the paper provide rigor to intuitively expected monotonocity in different settings.

3.2. Monotonicity in Rate

The MDP model has been formulated to maximize the infinite horizon expected total discounted net reward. The net reward over one time period in a given state $[i,t]$ equals the data rate of the user that is chosen to serve minus the penalty accrued by all other users. We expect the optimal policy to be monotone in the rate that can be potentially available to the users. In particular, we expect that if the optimal policy serves user v in state $[i,t]$, then it will serve v in state $[i+e_v,t]$ as well. We prove this in theorem 3.3. The proof of theorem 3.3 is similar (but more tedious) to the proof of theorem 3.2. However, it needs the additional condition of stochastic monotonicity of DTMCs, see [7].

Theorem 3.3 *If the Markov chain* $\{X^n,n\geq 1\}$ *is stochastically monotone, then*

$$dec(i,t)=v\Rightarrow dec(i+e_v,t)=v.\tag{18}$$

Proof. Since $dec(i,t)=v$ we have,

$$\begin{aligned}
&r_{i_v}-W_v(t)+\alpha h(i,t^v)\\
&\geq r_{i_u}-W_u(t)+\alpha h(i,t^u),\quad u\in A.
\end{aligned}\tag{19}$$

To prove $dec(i^+,t)=v$, we need to prove

$$\begin{aligned}
&\left[r_{(i+e_v)_v}-r_{(i+e_v)_u}\right]+\left[W_u(t)-W_v(t)\right]\\
&+\alpha\left[h(i+e_v,t^v)-h(i+e_v,t^u)\right]\geq 0.
\end{aligned}\tag{20}$$

To establish this, we can first prove

$$V(i+e_v,t^v)-V(i+e_v,t^u)\geq V(i,t^v)-V(i,t^u)\tag{21}$$

by considering the set of exhaustive cases similar to the proof of theorem 3.1. Stochastic monotonicity then implies

$$h(i+e_v,t^v)-h(i+e_v,t^u)\geq h(i,t^v)-h(i,t^u),\tag{22}$$

which yields 20, as required.

3.3. Long-Run Average Reward Criterion

In this section we extend the results of Section 3 to the long-run average reward criterion. As is well known, the objective in long-run average reward models is to maximize the long-run average reward, instead of the expected total discounted reward as considered in (5). If

$\mathrm{NR}_\pi(n)$ denotes the net reward at time n under policy π, then the objective of discounted reward models is to find a π that maximizes $\sum_n \alpha^n \mathrm{NR}_\pi(n)$ for a given $0 \le \alpha < 1$. The objective of long-run average reward models for the same dynamics, on the other hand, is to determine the policy π that maximizes

$$\lim_{N\to\infty} \sum_{n=1}^{N} \frac{\mathrm{NR}_\pi(n)}{N}.$$

It is well known [8] that a long-run average reward optimal policy $\{u(i,t):(i,t)\in \Omega^N x Z^N\}$ exists if there is a constant g (also called the gain) and a bias function $w(i,t)$ satisfying

$$\begin{aligned} & g + w(i,t) \\ & = \max_u \left\{ r_{i_u} - W_u(t) + \sum_j p_{ij} w\left(j, t^u\right) \right\}. \end{aligned} \tag{23}$$

The intuitive explanation of g and the bias function can be found in [3]. Here we end with the result that any u that maximizes $r_{i_u} - W_u(t) + \sum_j p_{ij} w\left(j, t^u\right)$ over all $u \in \{1, \cdots, N\}$ is an optimal action $u(i,t)$ in state (i,t).

Define a subset S of the state space $\Omega^N \times Z^N$ by

$$\begin{aligned} S = \{ & [i,t] \in \Omega^N \times Z^N : t_u = 0 \\ & \textit{for exactly one } u; \textit{ and for } u \ne v, t_u \ne t_v \}, \end{aligned} \tag{24}$$

i.e., a collection of states (i,t) such that no two users have the same starvation age and exactly one user has a starvation age of 0. Consider any stationary policy $\{f(i,t): \Omega^N \times Z^N \mapsto A\}$ of the original MDP introduced in the beginning of Section 2. Let $\{(X^n, Y^n), n \ge 1\}$ be the DTMC induced by f. Then we have the following lemma.

Lemma 3.4 *S is a closed communicating class of $\{(X^n, Y^n), n \ge 1\}$.*

Proof. Let $(X^n, Y^n) \in S$ for some $n \ge 1$. Since $\{Y^n, n \ge 1\}$ evolves according to (1) and we serve exactly one user in every time period, $[X^{n+1}, Y^{n+1}] \in S$. It is straightforward to show that the states in S communicate. Further, since $\{X^n, n \ge 1\}$ is a finite and irreducible DTMC, S is closed and communicating, as required.

We note that as a result of lemma 3.4 and the evolution of the age vector $\{Y^n, n \ge 1\}$, any state $[i,t] \in (\Omega^N \times Z^N) \setminus S$ is transient. Therefore, we restrict ourselves to proving monotonicity of the optimal policy on *S*. Let $[w(i,t):(i,t)\in S]$ be the bias vector satisfying (23). To prove that the monotonicity in age is valid (over *S*) for the long-run average reward criterion, we need to prove that for $[i,t] \in S$,

$$\begin{aligned} & r_{i_v} - W_v(t) + \sum_j p_{ij} w\left(j, t^v\right) \ge r_{i_u} \\ & - W_u(t) + \sum_j p_{ij} w\left(j, t^u\right) \Rightarrow \\ & r_{i_v} - W_v(t + e_v) + \sum_j p_{ij} w\left(j, (t+e_v)^v\right) \\ & \ge r_{i_u} - W_u(t + e_v) + \sum_j p_{ij} w\left(j, (t+e_v)^u\right). \end{aligned} \tag{25}$$

To do this, we choose a fixed integer T and for each $u \in A$ set

$$D_u(t) = \infty, \; t > T, \; u \in A. \tag{26}$$

Now, consider two systems:

- ***System A***: The MDP model described in the beginning of Section 2 with state space restricted to S and with the extra condition (26).
- ***System A′***: Identical to System *A* except that any user with age *T* has to be served. Therefore, the state space of this system is finite and is given by

$$S' = \{[i,t], \in S : t_u \le T, u \in A\}, \tag{27}$$

and the transition probabilities, reward structure are the same as that of System *A*. Clearly, as $T \uparrow \infty$, $S' \uparrow S$.

Our goal is to prove that for the long-run average reward criterion, the optimal policy is monotone in age in System *A*. We will show in theorem 3.5 that the monotonicity in age for the long-run average reward criterion holds for all fixed *T* in System A'. Further, since Systems *A* and A' are equivalent in the total optimal discounted reward sense of (33), we will conclude that monotonicity in age for the long-run average reward criterion holds for System *A*. Note that $dec(i,t)$ refers to the decision in state (i,t). For the long-run average reward criterion, it is obtained using Equation (23) in a way similar to the discounted reward criterion, *i.e.*, for the long-run average reward criterion,

$$dec(i,t) = \arg\max_u \left\{ r_{i_u} - W_u(t) + \sum_j p_{ij} w\left(j, t^u\right) \right\}.$$

Theorem 3.5 *The optimal policy for the long-run average reward criterion is monotone in age in System A', i.e. for $[i,t] \in S'$*

$$dec(i,t) = v \Rightarrow dec(i, t + e_v) = v. \tag{28}$$

Proof. Consider System A'. The state space S' is finite and using (2), the one step reward is bounded below by $C_L = r_1 - ND(T)$ and above by $C_U = r_N$. Thus the absolute value of the one step reward is bounded by $F = \max\{|C_L|, C_U\}$. Let $V'_\alpha(i,t)$ be the optimal expected total discounted reward of System A' starting in state $[i,t] \in S'$. Then $V'_\alpha(i,t)$ satisfies the standard Bellman equation given by (5). Using results in chapter 3 of [6], for a fixed $[k,m] \in S'$,

$$\left|V'_\alpha(i,t)-V'_\alpha(k,m)\right|<C<\infty \quad \text{for}[i,t]\in S', \tag{29}$$

where C is a positive constant. Then from Ross [9], there exists a constant g' and bias function $w'(i,t)$ satisfying (23) and given by

$$\begin{aligned} g' &= \lim_{\alpha\to 1}\left[V'_\alpha(k,m)(1-\alpha)\right] \\ w'(i,t) &= \lim_{\alpha\to 1}\left[V'_\alpha(i,t)-V'_\alpha(k,m)\right]. \end{aligned} \tag{30}$$

Theorem 3.2 implies that

$$\begin{aligned} & r_{i_v}-W_v(t)+\alpha\sum_j p_{ij}V'_\alpha\left(j,t^v\right)\ge r_{i_u} \\ & -W_u(t)+\alpha\sum_j p_{ij}V'_\alpha\left(j,t^u\right)\Rightarrow \\ & r_{i_v}-W_v(t+e_v)+\alpha\sum_j p_{ij}V'_\alpha\left(j,(t+e_v)^v\right) \\ & \ge r_{i_u}-W_u(t+e_v)+\alpha\sum_j p_{ij}V'_\alpha\left(j,(t+e_v)^u\right). \end{aligned} \tag{31}$$

Subtracting $V'_\alpha(k,m)=\sum_j p_{ij}V'_\alpha(k,m)$ on both sides of both the inequalities in (31) and taking the limit as $\alpha\to 1$, we get

$$\begin{aligned} & r_{i_v}-W_v(t)+\sum_j p_{ij}w'\left(j,t^v\right)\ge r_{i_u} \\ & -W_u(t)+\sum_j p_{ij}w'\left(j,t^u\right)\Rightarrow \\ & r_{i_v}-W_v(t+e_v)+\sum_j p_{ij}w'\left(j,(t+e_v)^v\right) \\ & \ge r_{i_u}-W_u(t+e_v)+\sum_j p_{ij}w'\left(j,(t+e_v)^u\right), \end{aligned} \tag{32}$$

using (??). Equation (32) implies (28), as required.

Thus the optimal policy of System A' is monotone in age for every T. Let $V_\alpha(i,t)$ be the optimal expected total discounted reward of System A starting in state $[i,t]\in S$. From the definition of Systems A and A', it is clear [8] that

$$V'_\alpha(i,t)=V_\alpha(i,t) \text{ for}[i,t]\in S'. \tag{33}$$

From Equations (33) and (29) through (32) it is clear that System A is monotone in age over S' constructed using any fixed T. Since $S'\uparrow S$ as $T\uparrow\infty$, we can conclude that the optimal policy of the MDP introduced in the beginning of Section 2 is monotone in age over S for the long-run average reward criterion.

Theorem 3.3 can be shown to hold in the long-run average reward case similarly and we omit the details for the sake of brevity expected in a correspondence.

3.4. Index Policy and Its Monotonicity

Now we consider the index policy proposed by Bolia and Kulkarni in [3]. It is described here for completeness. The decision $v(i,t)\in A$ in state $\left[X^n,Y^n\right]=[i,t]$ according to the index policy is given as follows:

$$\begin{aligned} I_u(i_u,t_u) &= r_{i_u}+K_u t_u\left(1+\frac{1}{q_u}\right)+\frac{K_u}{q_u}, \\ v(i,t) &= \arg\max_{u\in A} I_u(i_u,t_u). \end{aligned} \tag{34}$$

Here K_u and A_u are user dependent parameters that do not change with the state of the system (and as defined in [3], $K_u\ge 0$, $0\le q_u\le 1$). We prove the monotonicity of the index policy in age and rate below.

Theorem 3.6 *The Index Policy is monotone in age and rate, i.e.,*

$$v(i,t)=w\Rightarrow v(i,t+e_w)=w, \tag{35}$$

$$v(i,t)=w\Rightarrow v(i+e_w,t)=w, \tag{36}$$

Proof. The left hand side of (35) implies that, for $u\in A$,

$$r_{i_w}+K_w t_w\left(1+\frac{1}{q_w}\right)+\frac{K_w}{q_w}\ge r_{i_u}+K_u t_u\left(1+\frac{1}{q_u}\right)+\frac{K_u}{q_u}. \tag{37}$$

Therefore,

$$\begin{aligned} & r_{i_w}+K_w(t_w+1)\left(1+\frac{1}{q_w}\right)+\frac{K_w}{q_w} \\ & \ge r_{i_u}+K_u t_u\left(1+\frac{1}{q_u}\right)+\frac{K_u}{q_u}. \end{aligned} \tag{38}$$

yielding $v(i,t+e_w)=w$, which proves (35). Similarly, the left hand side of (37) clearly implies,

$$\begin{aligned} & r_{i_w+1}+K_w t_w\left(1+\frac{1}{q_w}\right)+\frac{K_w}{q_w} \\ & \ge r_{i_u}+K_u t_u\left(1+\frac{1}{q_u}\right)+\frac{K_u}{q_u}, \end{aligned}$$

yielding $v(i+e_w,t)=w$, which proves (36).

4. Conclusion

We considered a cellular data network, *i.e.* a system with a fixed number of buffers having time slotted Markov modulated departures and arrivals. The scheduling problem was modeled as an MDP and several structural (monotonicity) properties of its optimal policy proven. Although the entire analysis was carried out in the context of scheduling for wireless cellular data transfer, we emphasize that the structural properties hold true for any system with infinitely backlogged queues.

REFERENCES

[1] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhus-

hayana and A. Viterbi, "CDMA/HDR: A Bandwidth Efficient High Speed Wireless Data Service for Nomadic Users," *IEEE Communications Magazine*, Vol. 38, No. 7, 2000, pp. 70-77. doi:10.1109/35.852034

[2] L. Georgiadis, M. J. Neely and L. Tassiulas, "Resource Allocation and Cross-Layer Control in Wireless Networks," *Foundations and Trends in Networking*, Vol. 1, No. 1, 2006, pp. 1-144. doi:10.1561/1300000001

[3] N. Bolia and V. Kulkarni, "Index Policies for Resource Allocation in Wireless Networks," *IEEE Transactions on Vehicular Technology*, Vol. 58, No. 4, 2009, pp. 1823-1835. doi:10.1109/TVT.2008.2005101

[4] D. Tse, "Multiuser Diversity in Wireless Networks," 2011. http://www.eecs.berkeley.edu/~dtse/stanford416.ps

[5] M. Andrews, "Instability of the Proportional Fair Scheduling Algorithm for HDR," *IEEE Transactions on Wireless Communications*, Vol. 3, No. 5, 2002, p. 2004.

[6] Q. Hu and W. Yue, "Markov Decision Processes with Their Applications," Springer, New York, 2008.

[7] I. Kadi, N. Pekergin and J. M. Vincent, "Analytical and Stochastic Modeling Techniques and Applications," Springer, New York, 2009.

[8] M. Puterman, "Markov Decision Processes: Discrete Stochastic Dynamic Programming," John Wiley & Sons, Inc, New York, 1994.

[9] S. M. Ross, "Introduction to Stochastic Dynamic Programming," Academic Press, Inc., New York, 1983.

