Markov-modeled Downlink Environment: Opportunistic Multiuser Scheduling and the Stability Region

Journal:	*IEEE/ACM Transactions on Networking*
Manuscript ID:	draft
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Murugesan, Sugumar; The Ohio State University, ECE Schniter, Philip; The Ohio State University, ECE Shroff, Ness; The Ohio State University, ECE and CSE
Keywords:	Markov channel, Downlink, Multiuser Scheduling, Greedy Policy, Stability Region

Markov-modeled Downlink Environment: Opportunistic Multiuser Scheduling and the Stability Region

Sugumar Murugesan, Philip Schniter, *Senior Member, IEEE* and Ness B. Shroff, *Fellow, IEEE*

***Abstract*— In this paper, we focus on the downlink of a cellular system, which corresponds to the bulk of the data transfer in such communication systems. We address the problem of multiuser scheduling with partial channel information. In our setting, the channel of each user is modeled by a two-state Markov chain. The scheduler indirectly estimates the channel via accumulated Automatic Repeat Request (ARQ) feedback from the scheduled users and uses this information in future scheduling decisions. This problem falls under the *restless multi-armed bandit* processes that have been shown to be PSPACE-hard to solve in general. Using a Partially Observable Markov Decision Process (POMDP), we formulate a throughput maximization problem and show that, despite the visible complexity of this problem, a simple round-robin fashioned scheduling policy optimizes the system for the special case of three or less users in the system. We study the structure of this policy for an arbitrary number of users and establish a sufficient condition for the optimality of this policy. Drawing equivalence with a genie-aided system and assuming random arrivals of packets at the scheduler, we study the stability region of the downlink system and derive explicit expressions for the sum capacity of the downlink.**

***Index Terms* – Markov channel, downlink, multiuser scheduling, greedy policy, stability region.**

I. Introduction

With the ever increasing demand for high data rates, opportunistic multiuser scheduling, introduced by Knopp and Humblet in [1] and defined as *allocating the resources to the user experiencing the most favorable channel conditions* has gained immense popularity among wireless network designers. Opportunistic multiuser scheduling essentially taps the multiuser diversity in the system and has motivated several researchers (e.g., [2]–[6]) to study the performance gains obtained by opportunistic scheduling under various scenarios. While the i.i.d flat fading model is popularly used by these researchers in modeling time varying channels (for a general treatment on opportunistic scheduling with minimal assumptions on the channel, see [7]), it fails to capture the memory in the channel observed in realistic scenarios. This motivated the researchers to consider the Gilbert Elliott model [8] that represents the channel by a two state Markov chain. Specifically, a user experiences error-free transmission when it observes a “good” channel, and unsuccessful transmission in a “bad” channel. Several works have been done on opportunistic multiuser scheduling under the assumption of this Markov modeled channel (e.g., [9]–[13]). In these works, the channel state information that is crucial for the success of any opportunistic scheduling scheme is assumed to be readily available at the scheduler. This is a simplifying assumption that does not hold in reality, where a non-trivial amount of resource must be spent in gathering the information on the channel state. A new line of work (e.g., [14], [15]) attempts to exploit the memory in the Markov-modeled channels to gather this information. Specifically, Automatic Repeat reQuest (ARQ) feedback, that is traditionally used for error control (e.g., [16]–[19]) at the data link layer, is used to estimate the state of the Markov-modeled channels.

In this paper, we combine these two lines of work: exploit multiuser diversity in the Markov-modeled channels (e.g., [9]–[13]) and use ARQ feedback to estimate the state of these Markov-modeled channels (e.g., [14], [15]). Specifically, we consider a Markov-modeled downlink system with an ARQ feedback provision. Using a Partially Observable Markov Decision Process (POMDP) formulation ([20]–[23]), we show that, for $N \leq 3$ users, a simple greedy policy that maximizes the current reward is optimal in terms of the sum throughput. The greedy policy can be implemented via a simple round-robin based solution that does not require the statistics of the underlying Markov chain, so that it is easily amenable for practical implementation. Then, for the general N user case, by exploiting the round-robin structure of the greedy policy, we conjecture a sufficient condition for the optimality of the greedy policy. We provide extensive simulations that suggest that the greedy policy indeed satisfies this sufficient condition and is likely to be optimal for an arbitrary number of users in the system. After establishing an equivalence with a genie-aided system, we next derive a simple expression for the sum capacity of the Markov-modeled downlink system, for the two user case. Assuming that the base station maintains separate queues for the data meant for each user, we consider random data arrivals. Under this assumption, we study the stabilizable and the unstabilizable rate regions of the downlink system.

The paper is organized as follows. The problem setup is described in Section II and followed by the proof of the optimality of the greedy policy for the $N = 2$ case in Section III. Section IV discusses the round-robin structure of

This work was supported by National Science Foundation grants CCR-0237037, 0721434-CNS, 0635202-CCF, and Office of Naval Research grant N00014-07-1-0209.

Murugesan and Schniter are with the Department of Electrical and Computer Engineering, and Shroff holds a joint appointment in both the Departments of Electrical and Computer Engineering, and Computer Science and Engineering at The Ohio State University, Columbus, OH 43210, USA. (E-mail: {murugess, schniter, shroff}@ece.osu.edu)

the greedy policy. The sufficient condition for the optimality of the greedy policy for the general case of N users is derived in Section V. In the same section, we prove that the greedy policy is optimal for $N = 3$ and make a conjecture about the $N > 3$ case. In Section VI, we derive the sum capacity of the Markov-modeled downlink and obtain results on the stabilizable and the unstabilizable rate regions of the downlink. Conclusions are provided in Section VII.

II. Problem Setup

A. Channel Model

We consider downlink transmissions with N users. For each user, there is an associated queue at the base station that accumulates packets intended for that user. We assume an infinite backlog at each queue in Sections III - V and relax this condition in Section VI. The channel between the base station and each user is modeled by an i.i.d two-state Markov chain. We call this the ON-OFF channel with the ON state allowing the successful transmission of a fixed length packet. Time is slotted and the channel of each user remains fixed for a slot and evolves into another state in the next slot according to the Markov chain statistics. The time slots of all users are synchronized. The two-state Markov channel is characterized by a 2×2 probability transition matrix

$$P \quad = \quad \begin{bmatrix} p & q \\ r & s \end{bmatrix}, \tag{1}$$

where

- p := prob(channel is ON in the current slot | channel was ON in the previous slot)
- $q := 1 - p$
- r := prob(channel is ON in the current slot | channel was OFF in the previous slot)
- $s := 1 - r$.

We assume that $p \geq r$ throughout this work. This assumption implies that, for any user, the channel state is positively correlated between adjacent time slots.

B. Scheduling Problem

The base station is the central controller that controls the transmission to the users in each slot. In any time slot, the base station does not know the *exact* channel state of the users and it must schedule the transmission of the head of line packet of exactly one user. Thus, a TDMA styled scheduling is performed here. The power spent in each transmission is fixed, and a traditional ARQ based transmission is deployed. Here, at the beginning of a time slot, the head of line packet of the scheduled user is transmitted. If the packet does not go through, i.e, it cannot be decoded by the user (when the channel is in the OFF state), a NACK is sent back from the user at the end of the slot, and the packet is retained at the head of the queue. If the packet goes through (when the channel is in the ON state), an ACK is sent back and the packet is removed from the queue. Note that both ACKs and NACKs are assumed to be transmitted over a dedicated error-free channel. This ARQ information, along with the label of the time slot in which it is acquired, will be used in future scheduling decisions. The performance metric that the base station aims to maximize is the sum throughput of the system. Details will be discussed in the next section.

C. Formal Problem Definition

Since the base station must make scheduling decisions based only on a partial observation[1] of the underlying Markov chain, we employ techniques from partially observable Markov decision process (POMDP) theory in this work. See [20] for an overview of POMDP. We now proceed to introduce the terms/entities that we use in this work, many of which are borrowed from the POMDP literature. The key quantities used throughout this paper are summarized in Appendix II.

Control interval k: Each time slot in our problem setup will henceforth be called a control interval. The "end" of the POMDP is fixed. A control interval is indexed by k if there are $k - 1$ more intervals until the end of the process.

Action a_k: Indicates the index of the user scheduled in control interval k and hence takes on values from $1 \dots N$.

Belief vector at the k^{th} *control interval* π_k: The i^{th} element of π_k represents the probability that the channel of user $i \in 1 \dots N$ in the k^{th} control interval is in the ON state, given all the past information about the channel. Let f_k denote the ARQ feedback at the end of the control interval k from the scheduled user a_k. Let $f_k = 1$ indicate an ACK and $f_k = 0$ indicate a NACK. The belief vector evolves from control interval k to $k - 1$[2], $\forall k > 1$, as follows:

$$\pi_{k-1}(i) = \begin{cases} p, & \text{if } i = a_k, f_k = 1 \\ r, & \text{if } i = a_k, f_k = 0 \\ p\pi_k(i) + r(1 - \pi_k(i)), & \text{if } i \neq a_k. \end{cases} \tag{2}$$

where the first case indicates that user i is scheduled in control interval k and an ACK feedback was received. Thus, according to the Markov chain statistics in (1), $\pi_{k-1}(i) = p$. The second case is explained similarly when a NACK feedback is received. The last case indicates that user i was not scheduled for transmission in control interval k and hence the base station must estimate the belief value at the current control interval ($\pi_{k-1}(i)$) from the belief value at the previous control interval ($\pi_k(i)$) and the Markov chain statistics in (1). It has been proven in [20] that the belief vector π_k is a sufficient statistic to any information about the channels in the past control interval, in our case, the scheduling decisions and the ARQ information from the past. Thus the scheduling decision in any control interval can be solely based on the belief vector for that interval and not on the past ARQ or schedule information.

Scheduling Policy $\mathfrak{A}_k$: A scheduling policy $\mathfrak{A}_k$ in the control interval k is a mapping from the belief vector and the control interval index to an action as follows:

$$\mathfrak{A}_k : (\pi_k, k) \rightarrow a_k \quad \forall k \geq 1, \pi_k \in [0,1]^N.$$

Note that the scheduling policy can, in general, be time-variant.

[1] In this case, the set of time-stamped ARQ feedback on the channels.

[2] Note that the index of the control intervals decrease with time, consistent with the POMDP theory.

Reward Structure: In any control interval k, a reward of 1 is accrued when the transmission is successful, i.e, when $f_k = 1$, and no reward is accrued when $f_k = 0$. Note that this reward structure is defined to be consistent with our performance metric, the sum throughput (to be discussed shortly).

Net Expected Reward in the control interval m, V_m: With the belief vector, π_m, and the scheduling policy, $\{\mathfrak{A}_k\}_{k\leq m}$, fixed, the net expected reward, V_m, is the sum of the reward, $R_m(\pi_m, a_m)$, expected in the current control interval m and $\mathrm{E}[V_{m-1}]$, the net reward expected in the future control intervals conditioned on the belief vector and the scheduling decision in the current control interval. Formally,

$$\begin{aligned} & V_m(\pi_m, \{\mathfrak{A}_k\}_{k\leq m}) \\ &= R_m(\pi_m, a_m) + \mathrm{E}[V_{m-1}(\pi_{m-1}, \{\mathfrak{A}_k\}_{k\leq m-1})|\pi_m, a_m], \end{aligned} \tag{3}$$

where the expectation is over the belief vector π_{m-1}. Since the reward in each control interval is either 1 or 0, the expected current reward can be written as

$$R_m(\pi_m, a_m) = \pi_m(a_m).$$

Performance Metric- the Sum Throughput, η_{sum}: For a given scheduling policy, $\{\mathfrak{A}_k\}_{k\geq 1}$, the sum throughput is given by

$$\eta_{\text{sum}}(\{\mathfrak{A}_k\}_{k\geq 1}) = \lim_{m\to\infty} \frac{V_m(\pi_{\text{ss}}, \{\mathfrak{A}_k\}_{k\geq 1})}{m}, \tag{4}$$

where $\pi_{\text{ss}}(i), i \in 1\ldots N$ is the steady state probability that the channel of user i is ON in the underlying Markov chain.

Optimal Scheduling Policy, $\{\mathfrak{A}_k^*\}_{k\geq 1}$:

$$\{\mathfrak{A}_k^*\}_{k\geq 1} := \arg \max_{\{\mathfrak{A}_k\}_{k\geq 1}} \eta_{\text{sum}}(\{\mathfrak{A}_k\}_{k\geq 1}). \tag{5}$$

III. Optimal Scheduling Policy for Two Users

Consider the following policy:

$$\begin{aligned} \widehat{\mathfrak{A}}_k : (\pi_k, k) \to a_k &= \arg\max_{a_k} R_k(\pi_k, a_k) \\ &= \arg\max_{i} \pi_k(i) \quad \forall k \geq 1, \pi_k \in [0,1]^N. \end{aligned}$$

Since the above given policy attempts to maximize the expected current reward, without any regard to the expected future reward, it follows an approach that is fundamentally *greedy* in nature. For this reason, we henceforth call $\{\widehat{\mathfrak{A}}_k\}_{k\geq 1}$ the Greedy Policy.

Proposition 1: The sum throughput, η_{sum}, of the system is maximized by the greedy policy $\{\widehat{\mathfrak{A}}_k\}_{k\geq 1}$ for the case when $N = 2$, i.e.,

$$\mathfrak{A}_k^*|_{N=2} = \widehat{\mathfrak{A}}_k \quad \forall k \geq 1.$$

Proof: From Subsection II-C, to prove the optimality of the greedy policy, it is sufficient to prove that the greedy policy maximizes the net expected reward in any control interval, i.e., $\forall m \geq 1, \pi_m \in [0,1]^2$,

$$\{\widehat{\mathfrak{A}}_k\}_{k\leq m} = \arg \max_{\{\mathfrak{A}_k\}_{k\leq m}} V_m(\pi_m, \{\mathfrak{A}_k\}_{k\leq m}).$$

We proceed to prove the above statement for $N = 2$, using induction. We first prove the following statement:

(P) If, for a fixed $m > 1$, $\forall \pi_{m-1} \in [0,1]^2$,

$$\{\widehat{\mathfrak{A}}_k\}_{k\leq m-1} = \arg \max_{\{\mathfrak{A}_k\}_{k\leq m-1}} V_{m-1}(\pi_{m-1}, \{\mathfrak{A}_k\}_{k\leq m-1}),$$

then, $\forall \pi_m \in [0,1]^2$,

$$\{\widehat{\mathfrak{A}}_k\}_{k\leq m} = \arg \max_{\{\mathfrak{A}_k\}_{k\leq m}} V_m(\pi_m, \{\mathfrak{A}_k\}_{k\leq m}).$$

In words, (P) states that, for any $m > 1$, if the greedy policy maximizes the net expected reward in control interval $m-1$, then it maximizes the net expected reward in control interval m as well. We prove (P) as follows. Let π_m, a_m be fixed. The net expected reward, V_m, under the policy $\{a_m, \{\widehat{\mathfrak{A}}_k\}_{k\leq m-1}\}$, is given by

$$\begin{aligned} & V_m(\pi_m, \{a_m, \{\widehat{\mathfrak{A}}_k\}_{k\leq m-1}\}) \\ &= \pi_m(a_m) + \mathrm{E}[V_{m-1}(\pi_{m-1}, \{\widehat{\mathfrak{A}}_k\}_{k\leq m-1})|\pi_m, a_m]. \end{aligned} \tag{6}$$

We now focus on the expected future reward,

$$\begin{aligned} & \mathrm{E}[V_{m-1}(\pi_{m-1}, \{\widehat{\mathfrak{A}}_k\}_{k\leq m-1})|\pi_m, a_m] \\ &= \sum_{k=m-1}^{1} \mathrm{E}_{\pi_k|\pi_m, a_m}\left[R_k(\pi_k, \hat{a}_k)\right] \\ &= \sum_{k=m-1}^{1} \mathrm{E}_{\pi_{k+1}|\pi_m, a_m}\left[\mathrm{E}_{\pi_k|\pi_{k+1}, \pi_m, a_m}\left[R_k(\pi_k, \hat{a}_k)\right]\right], \end{aligned} \tag{7}$$

where we have used $\hat{a}_k$ to emphasize the fact that, in every control interval $k \leq m-1$, the action is chosen under the greedy policy $\widehat{\mathfrak{A}}_k$. $\mathrm{E}_{\pi_k|\pi_m, a_m}[.]$ indicates the average over π_k conditioned on π_m and a_m. Now consider the expected current reward for the control interval $k \leq m-1$, i.e., R_k, conditioned on the belief vector of the previous control interval π_{k+1} and the initial conditions π_m, a_m. Let $\bar{a}_{k+1}$ indicate the index of the user that is NOT scheduled in control interval $k+1$ and a_{k+1} indicate the scheduled user. We have the following two cases.

1. *When the ARQ feedback $f_{k+1} = 1$, (occurs with probability $\pi_{k+1}(a_{k+1})$):*

$$\begin{aligned} \pi_k(a_{k+1}) &= p \\ \pi_k(\bar{a}_{k+1}) &= p\pi_{k+1}(\bar{a}_{k+1}) + r(1 - \pi_{k+1}(\bar{a}_{k+1})). \end{aligned}$$

Since $p \geq r$, we have, $\pi_k(a_{k+1}) \geq \pi_k(\bar{a}_{k+1})$. Hence the greedy policy chooses the action $a_k = a_{k+1}$ with the expected current reward $R_k = p$.

2. *When $f_{k+1} = 0$, (occurs with probability $1 - \pi_{k+1}(a_{k+1})$):*

$$\begin{aligned} \pi_k(a_{k+1}) &= r \\ \pi_k(\bar{a}_{k+1}) &= p\pi_{k+1}(\bar{a}_{k+1}) + r(1 - \pi_{k+1}(\bar{a}_{k+1})). \end{aligned}$$

Since $p \geq r$, we have, $\pi_k(a_{k+1}) \leq \pi_k(\bar{a}_{k+1})$. Thus the greedy policy chooses $a_k = \bar{a}_{k+1}$ with the expected current reward $R_k = \pi_k(\bar{a}_{k+1}) = p\pi_{k+1}(\bar{a}_{k+1}) + r(1 - \pi_{k+1}(\bar{a}_{k+1}))$.

Define the state vector S_l such that $S_l(i)$ indicates the state (1-ON/0-OFF) of the channel of user i in control interval l.

Now, averaging over the two cases discussed above, we have

$$\begin{aligned}
&\mathrm{E}_{\pi_k|\pi_{k+1},\pi_m,a_m} R_k(\pi_k,\hat{a}_k)\\
&\quad= \pi_{k+1}(a_{k+1})p + (1-\pi_{k+1}(a_{k+1}))(p\pi_{k+1}(\bar{a}_{k+1})\\
&\qquad +r(1-\pi_{k+1}(\bar{a}_{k+1})))\\
&\quad= \Big(\pi_{k+1}(a_{k+1}) + (1-\pi_{k+1}(a_{k+1}))\pi_{k+1}(\bar{a}_{k+1})\Big)p\\
&\qquad +\Big((1-\pi_{k+1}(a_{k+1}))(1-\pi_{k+1}(\bar{a}_{k+1}))\Big)r\\
&\quad= P\big(\{S_{k+1}(1)=1 \cup S_{k+1}(2)=1\}|\pi_{k+1}\big)p\\
&\qquad +P\big(\{S_{k+1}(1)=0 \cap S_{k+1}(2)=0\}|\pi_{k+1}\big)r.
\end{aligned}$$

Note that, in the last equation, the events $\{S_{k+1}(1) = 1 \cup S_{k+1}(2) = 1\}$ and $\{S_{k+1}(1) = 0 \cap S_{k+1}(2) = 0\}$ are independent of π_m and a_m given π_{k+1}, since π_{k+1} is a sufficient statistic to all the information prior to $k+1$. The reward expected in any future control interval, $k \leq m-1$, is now given by

$$\begin{aligned}
&\mathrm{E}_{\pi_k|\pi_m,a_m}\left[R_k(\pi_k,\hat{a}_k)\right]\\
&\quad= \mathrm{E}_{\pi_{k+1}|\pi_m,a_m}\Big[P\big(\{S_{k+1}(1)=1 \cup S_{k+1}(2)=1\}|\pi_{k+1}\big)p\\
&\qquad +P\big(\{S_{k+1}(1)=0 \cap S_{k+1}(2)=0\}|\pi_{k+1}\big)r\Big]\\
&\quad= P\big(\{S_{k+1}(1)=1 \cup S_{k+1}(2)=1\}|\pi_m,a_m\big)p\\
&\qquad +P\big(\{S_{k+1}(1)=0 \cap S_{k+1}(2)=0\}|\pi_m,a_m\big)r\\
&\quad= P\big(\{S_{k+1}(1)=1 \cup S_{k+1}(2)=1\}|\pi_m\big)p\\
&\qquad +P\big(\{S_{k+1}(1)=0 \cap S_{k+1}(2)=0\}|\pi_m\big)r, \qquad (8)
\end{aligned}$$

where the last equation follows from the fact that $S_l(i)$, for $l \leq m, i \in \{1,2\}$ is independent of the action a_m, given the belief vector π_m. Thus, the reward expected in any future control interval is independent of the current scheduling decision a_m. Returning to (7), we have

$$\begin{aligned}
&\mathrm{E}[V_{m-1}(\pi_{m-1},\{\widehat{\mathfrak{A}}_k\}_{k\leq m-1})|\pi_m, a_m=1]\\
&\quad= \mathrm{E}[V_{m-1}(\pi_{m-1},\{\widehat{\mathfrak{A}}_k\}_{k\leq m-1})|\pi_m, a_m=2]. \qquad (9)
\end{aligned}$$

Thus, from (6),

$$\begin{aligned}
&V_m(\pi_m,\{a_m=1,\{\widehat{\mathfrak{A}}_k\}_{k\leq m-1}\})\\
&\qquad -V_m(\pi_m,\{a_m=2,\{\widehat{\mathfrak{A}}_k\}_{k\leq m-1}\})\\
&\quad= \pi_m(1)-\pi_m(2). \qquad (10)
\end{aligned}$$

(P) follows from the above equation. Since

$$\widehat{\mathfrak{A}}_1 = \arg\max_{\mathfrak{A}_1} V_1(\pi_1,\mathfrak{A}_1) \quad \forall \pi_1 \in [0,1]^2,$$

Proposition 1 follows from (P) by induction. ■

It has to be mentioned that a parallel work, [24], by Qing Zhao et al., addresses a similar problem in a cognitive radio setting where a single user attempts to opportunistically access one of the several radio channels. Due to the fundamental difference in the application areas targeted, the overlap between our paper and [24] is limited to the result on the optimality of the greedy policy when $N = 2$ users. The proof technique they have used involves evaluating the net expected reward by averaging over the channel states of both the users in the current control interval. Whereas, in our case, we average over the belief vector. Moreover, we explicitly establish that the reward expected to be accrued in *any* future control interval is independent of the current scheduling decision. This is unlike [24], where the independence result is obtained only for the net future reward. Our independence result (summarized in Corollary 2) will be pivotal in obtaining a simple, closed form expression for the sum capacity of the Markov-modeled downlink in Section VI.

Corollary 2: The reward expected to be accrued in any future control interval $k \leq m-1$ is independent of the scheduling decision a_m, as long as the greedy policy is implemented in control interval k. Formally, $\forall k \leq m-1$,

$$\mathrm{E}_{\pi_k|\pi_m,a_m=1} R_k(\pi_k,\hat{a}_k) = \mathrm{E}_{\pi_k|\pi_m,a_m=2} R_k(\pi_k,\hat{a}_k),$$

where $\hat{a}_k$ indicates the use of the greedy policy in control interval k.

Corollary 2 follows form (8). The significance of the above observation will be discussed in Section VI.

IV. Structure of the Greedy Policy for N Users

Having established the optimality of the greedy policy for the $N = 2$ users case, we now take a closer look at the structure of the greedy policy. We begin by defining the following quantity.

Schedule order vector, O_k: The ordered arrangement of the index of the users in decreasing order of $\pi_k(i)$, i.e.,

$$\begin{aligned}
O_k(1) &= \arg\max_i \pi_k(i)\\
&\;\;\vdots\\
O_k(N) &= \arg\min_i \pi_k(i).
\end{aligned}$$

Thus, under the greedy policy in k, $a_k = O_k(1)$.

We now discuss the evolution of O_k to O_{k-1}. Consider any two users that are not scheduled in control interval k, i.e, consider users $i \neq a_k$ and $j \neq a_k$. Thus from (2), the belief value of user i evolves from control interval k to $k-1$ as follows: $\pi_{k-1}(i) = p\pi_k(i) + r(1-\pi_k(i)) = (p-r)\pi_k(i)+r$. Similarly, considering user j, $\pi_{k-1}(j) = (p-r)\pi_k(j)+r$. Thus, since $p \geq r$, $\pi_{k-1}(i) \geq \pi_{k-1}(j)$ if $\pi_k(i) \geq \pi_k(j), \forall i \neq a_k, j \neq a_k$, i.e., the order of the belief values of any two users whose channels are not observed in the current control interval is retained in the next control interval. This follows from the *positive correlation in time* property of the underlying Markov chain, facilitated by the assumption $p \geq r$. Now consider the user scheduled in control interval k, i.e., user a_k. If the ARQ feedback $f_k = 1$, then $\pi_{k-1}(a_k) = p$. Since for any user $i \neq a_k$, $\pi_{k-1}(i) = p\pi_k(i) + r(1-\pi_k(i))$ and since $p \geq r$, we have, $\pi_{k-1}(a_k) \geq \pi_{k-1}(i), \forall i \neq a_k$. The preceding statement can be interpreted as follows: the channel that was ON with probability 1 in the previous control interval is more likely to be ON in the current control interval than the channel that was ON with probability less than 1 in the previous control interval. When $f_k = 0$, $\pi_{k-1}(a_k) = r \leq \pi_{k-1}(i), \forall i \neq a_k$. From the preceding observations,

$$O_{k-1} = \begin{cases} [a_k\ \{O_k - a_k\}], & \text{if } f_k = 1\\ [\{O_k - a_k\}\ a_k], & \text{if } f_k = 0, \end{cases} \qquad (11)$$

where $\{O_k - a_k\}$ is the schedule order vector O_k with the element valued a_k removed. For instance, $\{[x\ y\ z] - y\} = [x\ z]$.

As a special case, when the greedy policy is employed in control interval k, i.e., when $a_k = O_k(1)$,

$$O_{k-1} = \begin{cases} O_k, & \text{if } f_k = 1 \\ [O_k(2)\ O_k(3)\ \ldots O_k(N)\ O_k(1)], & \text{if } f_k = 0. \end{cases} \quad (12)$$

The evolution of the schedule order vector under the greedy policy is illustrated in Fig. 1 assuming $\pi_m(1) \geq \pi_m(2) \ldots \geq \pi_m(N)$.

We are now in a position to make the following important observation:

Let the greedy policy be implemented from control interval m. Let the schedule order vector, O_m, be available to the base station. The scheduling algorithm is implemented as follows: *Schedule the user positioned at the top of the schedule order vector (i.e., $a_m = O_m(1)$). If an ACK is received, schedule the same user again in the next control interval. Otherwise, schedule the next user in the schedule order vector O_m. Repeat the same procedure in all the future control intervals. If the bottom of the schedule order vector is reached, repeat from the top.* Formally, the algorithm is implemented in the following simple steps

- Step 1: Initialize the control interval index $k \leftarrow m$ and the position of the scheduled user in the schedule order vector as $i \leftarrow 1$.
- Step 2: Schedule user $O_m(i)$ in control interval k, i.e., $a_k = O_m(i)$.
- Step 3: If $f_k = 0$ and $i < N$, then $i \leftarrow i + 1$. If $f_k = 0$ and $i = N$, then $i \leftarrow 1$.
- Step 4: $k \leftarrow k - 1$. If $k > 0$, then repeat steps 2-4.

Thus, the scheduling algorithm, under the greedy policy, boils down to a simple round-robin algorithm with a change in scheduling decision stimulated by a NACK feedback (illustrated in Fig. 2). The schedule order vector, O_m, provides the order of this round-robin approach. The greedy policy that aims to maximize the expected current reward takes the form of the round-robin algorithm due to the following properties of the Markov channels in our problem:

- The shorter the time since a channel is observed to be in the ON state, the more likely it is that the channel is currently ON. This explains why the greedy policy, via the round-robin algorithm, retains the scheduling decision on receiving an ACK feedback.
- The longer the time since a channel is observed to be in the OFF state, the more likely it is that the channel is currently ON. This explains why the greedy policy, on receiving a NACK feedback, schedules the user next in the schedule order vector. Note that this user, assuming the round robin algorithm has completed at least one full cycle, has spent the most amount of time since being observed to have an OFF channel.

Note that the round-robin algorithm does not involve evaluating the belief vector in every control interval. Hence the Markov transition matrix information is not required. This structure makes the greedy policy particularly attractive from an implementation point of view. Motivated by this development, we proceed to examine the optimality of the greedy policy in a general N user setting.

V. On the Optimality of the Greedy Policy for N Users

A. Sufficient Condition for the Optimality of the Greedy Policy

Consider a control interval $m > 1$ with belief vector π_m and action a_m. Let the users be indexed in the order of their belief values in control interval m, i.e, $O_m = [1 \ldots N]$. Assuming $\{\mathfrak{A}_k\}_{k \leq m-1} = \{\widehat{\mathfrak{A}}_k\}_{k \leq m-1}$ and recalling the definition of state vector S_k from Section III, we rewrite the net expected reward from (3) as follows

$$\begin{aligned} & V_m(\pi_m, \{a_m, \{\widehat{\mathfrak{A}}_k\}_{k \leq m-1}\}) \\ & \quad = \quad \pi_m(a_m) + \sum_{S_m} P_{S_m|\pi_m}(S_m|\pi_m)\hat{V}_{m-1}(S_m, O_{m-1}), \end{aligned}$$

where $\hat{V}_{m-1}$ is the expected future reward conditioned on the state vector in control interval m. The *hat* on this quantity emphasizes the use of the greedy policy in all $k \leq m - 1$. $P_{S_m|\pi_m}(S_m|\pi_m)$ is the conditional probability of the current state vector S_m given the belief vector π_m. Note that the schedule order vector O_{m-1} is only a function of O_m and the state $S_m(a_m)$, thus maintaining consistency with the amount of information available for scheduling decision in the actual problem setup. We now proceed to compare the net expected reward when $a_m = n$ and $a_m = n+1$ where $n \in \{1 \ldots N-1\}$. Let Y and X be random binary vectors of lengths $n - 1$ and $N-n-1$ (empty when the length is non-positive) respectively. Then,

$$\begin{aligned} & V_m(\pi_m, \{a_m = n, \{\widehat{\mathfrak{A}}_k\}_{k \leq m-1}\}) \\ & \quad = \quad \pi_m(n) + \sum_{Y,X} P_{S_m|\pi_m}([Y\ 0\ 0\ X]|\pi_m) \times \\ & \qquad\qquad\qquad \hat{V}_{m-1}\big([Y\ 0\ 0\ X], [\{O_m - n\}\ n]\big) \\ & \qquad + \sum_{Y,X} P_{S_m|\pi_m}([Y\ 0\ 1\ X]|\pi_m) \times \\ & \qquad\qquad\qquad \hat{V}_{m-1}\big([Y\ 0\ 1\ X], [\{O_m - n\}\ n]\big) \\ & \qquad + \sum_{Y,X} P_{S_m|\pi_m}([Y\ 1\ 0\ X]|\pi_m) \times \\ & \qquad\qquad\qquad \hat{V}_{m-1}\big([Y\ 1\ 0\ X], [n\ \{O_m - n\}]\big) \\ & \qquad + \sum_{Y,X} P_{S_m|\pi_m}([Y\ 1\ 1\ X]|\pi_m) \times \\ & \qquad\qquad\qquad \hat{V}_{m-1}\big([Y\ 1\ 1\ X], [n\ \{O_m - n\}]\big), \end{aligned} \quad (13)$$

where $O_m \rightarrow O_{m-1}$, the evolution of the schedule order vector, follows (11).

Since the Markov channel statistics are identical across the users, we have the following symmetry property: for any $k \geq 1$,

$$\begin{aligned} \hat{V}_k(S_{k+1}, O_k) &= \hat{V}_k(\tilde{S}_{k+1}, \tilde{O}_k) \\ \text{if} \quad S_{k+1}(O_k(i)) &= \tilde{S}_{k+1}(\tilde{O}_k(i)) \ \ \forall\, i \in \{1 \ldots N\}. \end{aligned} \quad (14)$$

Thus, for instance,

$$\begin{aligned} & \hat{V}_{m-1}\big([Y\ 0\ 1\ X],[\{O_m-n\}\ n]\big) \\ = \ & \hat{V}_{m-1}\big([Y\ 1\ 0\ X],[\{O_m-(n+1)\}\ (n+1)]\big) \\ = \ & \hat{V}_{m-1}\big([Y\ 1\ X\ 0],[1\ldots N]\big). \end{aligned}$$

Expanding $V_m(\pi_m,\{a_m = n+1,\{\widehat{\mathfrak{A}}_k\}_{k\le m-1}\})$ along the lines of (13), and using the symmetry property, with further mathematical simplification, we can evaluate the difference in the net expected reward as follows,

$$\begin{aligned} & V_m(\pi_m,\{a_m=n,\{\widehat{\mathfrak{A}}_k\}_{k\le m-1}\}) \\ & \quad -V_m(\pi_m,\{a_m=n+1,\{\widehat{\mathfrak{A}}_k\}_{k\le m-1}\}) \\ = \ & \Big(\pi_m(n)-\pi_m(n+1)\Big)\times \\ & \Big(1-\sum_{Y,X}\Big[\big[\hat{V}_{m-1}\big([Y\ 1\ X\ 0],[1\ldots N]\big) \\ & -\hat{V}_{m-1}\big([1\ Y\ 0\ X],[1\ldots N]\big)\big]\times \\ & P_{S_m|\pi_m}\big([S_m(1)\ldots S_m(n-1)]=Y|\pi_m\big)\times \\ & P_{S_m|\pi_m}\big([S_m(n+2)\ldots S_m(N)]=X|\pi_m\big)\Big]\Big). \end{aligned} \tag{15}$$

Proposition 3: A sufficient condition for the optimality of the greedy policy is given as follows,

$$\begin{aligned} & \hat{V}_{m-1}\big([Y\ 1\ X\ 0],[1\ldots N]\big) \\ & \quad -\hat{V}_{m-1}\big([1\ Y\ 0\ X],[1\ldots N]\big)\le 1, \end{aligned} \tag{16}$$

$\forall\ m>1$, $n\in\{1\ldots N-1\}$, Y, X being random binary vectors of length $n-1$ and $N-n-1$ and $\{\mathfrak{A}_k\}_{k\le m-1} = \{\widehat{\mathfrak{A}}_k\}_{k\le m-1}$.

Proof: Let condition (16) be true. Let $m>1$ be fixed. Since, by assumption, $\pi_m(n)\ge\pi_m(n+1)\ \forall n\in\{1\ldots N-1\}$, we have from (15),

$$\begin{aligned} & V_m(\pi_m,\{a_m=n,\{\widehat{\mathfrak{A}}_k\}_{k\le m-1}\}) \\ \ge \ & V_m(\pi_m,\{a_m=n+1,\{\widehat{\mathfrak{A}}_k\}_{k\le m-1}\}). \end{aligned}$$

Therefore,

$$\begin{aligned} & V_m(\pi_m,\{a_m=\arg\max_i \pi_m(i)=1,\{\widehat{\mathfrak{A}}_k\}_{k\le m-1}\}) \\ \ge \ & V_m(\pi_m,\{a_m\in\{2\ldots N\},\{\widehat{\mathfrak{A}}_k\}_{k\le m-1}\}). \end{aligned}$$

We now have the following statement:

If $\forall\pi_{m-1}\in[0,1]^N$,

$$\{\widehat{\mathfrak{A}}_k\}_{k\le m-1} = \arg\max_{\{\mathfrak{A}_k\}_{k\le m-1}} V_{m-1}(\pi_{m-1},\{\mathfrak{A}_k\}_{k\le m-1}),$$

then $\forall\pi_m\in[0,1]^N$,

$$\{\widehat{\mathfrak{A}}_k\}_{k\le m} = \arg\max_{\{\mathfrak{A}_k\}_{k\le m}} V_m(\pi_m,\{\mathfrak{A}_k\}_{k\le m}). \tag{17}$$

Since $\widehat{\mathfrak{A}}_1 = \arg\max_{\mathfrak{A}_1} V_1(\pi_1,\mathfrak{A}_1), \forall\pi_1\in[0,1]^N$, using (17), by induction, we have

$$\begin{aligned} \{\widehat{\mathfrak{A}}_k\}_{k\le m} \ = \ & \arg\max_{\{\mathfrak{A}_k\}_{k\le m}} V_m(\pi_m,\{\mathfrak{A}_k\}_{k\le m}) \\ & \forall m\ge 1,\pi_m\in[0,1]^N. \end{aligned}$$

The proposition follows from (5). ■

B. Optimality of the Greedy Policy with $N=3$ Users

Proposition 4: When $N=3$ users, the greedy policy satisfies the sufficient condition in Proposition 3 and is hence optimal.

Proof: With $N=3$, $n\in\{1,2\}$. The sufficient condition in Proposition 3 becomes, $\forall\ m>1$,

$$\begin{aligned} & \Big[\hat{V}_{m-1}\big([1\ X\ 0],[1\ 2\ 3]\big)-\hat{V}_{m-1}\big([1\ 0\ X],[1\ 2\ 3]\big)\Big]\le 1 \\ & \qquad\qquad\qquad\qquad \text{when } n=1, \\ & \Big[\hat{V}_{m-1}\big([Y\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_{m-1}\big([1\ Y\ 0],[1\ 2\ 3]\big)\Big]\le 1 \\ & \qquad\qquad\qquad\qquad \text{when } n=2, \end{aligned} \tag{18}$$

where X, Y are binary numbers and $\{\mathfrak{A}_k\}_{k\le m-1} = \{\widehat{\mathfrak{A}}_k\}_{k\le m-1}$. The proof of (18) is tedious and hence is moved to the appendix to maintain continuity of the discussion. ■

Due to the complex relationship between the scheduling decision in a control interval and the reward expected in the future intervals, an analysis of the optimality of the greedy policy for the general N-user case appears difficult. But with support from simulation results, we conjecture that the sufficient condition in Proposition 3 is satisfied for any value of N, thus suggesting the optimality of the greedy policy for any N. Fig. 3 plots the values of $\hat{V}_{m-1}\big([Y\ 1\ X\ 0],[1\ldots N]\big) - \hat{V}_{m-1}\big([1\ Y\ 0\ X],[1\ldots N]\big)$ for various values of $n\in\{1\ldots N-1\}$ when $N=5$, $P=\begin{bmatrix}0.6324 & 0.3676\\ .0975 & .9025\end{bmatrix}$. In each of the four subplots, Y and X are allowed to take on every possible binary word of length $n-1$ and $N-n-1$, respectively.

VI. On the Stability Region of the Markov-Modeled Downlink with $N=2$ Users

So far, we have assumed that the queues for each user maintained at the base station are infinitely backlogged. This assumption is not always true in practice, where packets arrive at each queue according to arrival processes and arrival rates dictated by the applications serving the users. In this scenario, we are interested in the *complete* set of arrival rate vectors that can be supported by the downlink system without leading to the instability of any of the queues[3]. This complete set of arrival rate vectors is known as the *stability region* of the system. To examine the stability region, we first evaluate the sum capacity of the downlink system.

A. Sum Capacity of the Markov-modeled Downlink

With the greedy policy established as the sum-throughput maximizing scheduling policy for $N=2$ users, the sum capacity of the system is given by the sum throughput under the greedy policy. We now give the following result.

Proposition 5: When $N=2$, the sum capacity of the given Markov-modeled downlink equals that of a genie-aided Markov-modeled downlink where, at the end of every control interval, the base station learns the state of the channels of all the users in that control interval. The sum capacity is given as

$$\begin{aligned} C_{\text{sum}} \ &= \ p_s p + (1-p_s)p_s \\ \text{with } \ p_s \ &= \ \frac{r}{1-(p-r)}, \end{aligned}$$

[3] A queue with backlog B is defined to be stable iff $\lim_{b\to\infty} P(B>b)=0$.

where p_s is the probability that the channel of user 1 (or 2) is ON in steady state. The greedy policy achieves the sum capacity of both systems.

Proof: We begin with the derivation of the sum throughput of the greedy policy in the genie-aided system. From (4), with $\pi_{\text{ss}}(1) = \pi_{\text{ss}}(2) = p_s$,

$$\begin{aligned}
&\eta_{\text{sum}}^{\text{genie}}(\{\widehat{\mathfrak{A}}_k\}_{k\geq 1}) \\
&\quad = \lim_{m\to\infty} \frac{V_m^{\text{genie}}(\pi_{\text{ss}}, \{\widehat{\mathfrak{A}}_k\}_{k\geq 1})}{m} \\
&\quad = \lim_{m\to\infty} \frac{1}{m}\Big(p_s \\
&\quad\quad + \sum_{k=m-1}^{1} \mathrm{E}_{S_{k+1}(1),S_{k+1}(2)|\pi_m=\pi_{\text{ss}}} R_k(S_{k+1}(1), S_{k+1}(2))\Big),
\end{aligned}$$

where, for any k,

$$\begin{aligned}
&\mathrm{E}_{S_{k+1}(1),S_{k+1}(2)|\pi_m=\pi_{\text{ss}}} R_k(S_{k+1}(1), S_{k+1}(2)) \\
&\quad = \quad P\big(\{S_{k+1}(1) = 1 \cup S_{k+1}(2) = 1\}|\pi_m = \pi_{\text{ss}}\big)p \\
&\quad\quad\quad +P\big(\{S_{k+1}(1) = 0 \cap S_{k+1}(2) = 0\}|\pi_m = \pi_{\text{ss}}\big)r \\
&\quad = \quad (2p_s - p_s^2)p + (1-p_s)^2 r \\
&\quad = \quad p_s p + (1-p_s)p_s.
\end{aligned}$$

The last equation follows from the property of the steady state probability: $p_s = p_s p + (1-p_s)r$. The sum throughput is now given by,

$$\eta_{\text{sum}}^{\text{genie}}(\{\widehat{\mathfrak{A}}_k\}_{k\geq 1}) \quad = \quad p_s p + (1-p_s)p_s, \tag{19}$$

with p_s derived as below.

The Markov chain transition matrix $P = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$ can be written as $P = U\Lambda V$ where

$$\begin{aligned}
U &= \begin{bmatrix} 1 & 1 \\ 1 & \frac{-r}{q} \end{bmatrix} \\
\Lambda &= \begin{bmatrix} 1 & 0 \\ 0 & p+s-1 \end{bmatrix} \\
V &= \frac{1}{1+\frac{r}{q}} \begin{bmatrix} \frac{r}{q} & 1 \\ 1 & -1 \end{bmatrix}
\end{aligned}$$

with $VU = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Assuming[4] $p + s < 2$,

$$\begin{aligned}
\lim_{n\to\infty} P^n &= \begin{bmatrix} \frac{r}{1-(p-r)} & 1 - \frac{r}{1-(p-r)} \\ \frac{r}{1-(p-r)} & 1 - \frac{r}{1-(p-r)} \end{bmatrix} \\
\Rightarrow p_s &= \frac{r}{1-(p-r)}.
\end{aligned}$$

We now proceed to prove that the sum throughput of the greedy policy in the original system equals that of the greedy policy in the genie-aided systems. Consider the scheduling problem for the original system in control interval k under the greedy policy. When the user scheduled in the previous control interval a_{k+1} sends back an ACK, the scheduling decision is retained in the current interval, i.e., $a_k = a_{k+1}$. Otherwise, the other user is scheduled in k. This procedure is evident from the structure of the greedy policy discussed in Section IV. We can interpret this decision logic as follows:

When at least one of the users had an ON channel in the previous control interval, that user[5] is identified for scheduling in the current control interval k, leading to an expected current reward $R_k = p$. Reward $R_k = r$ is accrued only when both the channels were in the OFF state.

From this observation we see that, under the greedy policy, no improvement in sum throughput can be achieved even if the channel states of both the users in control interval $k+1$ were available for the scheduling decision in control interval k. This establishes the equivalence between the original system and the genie-aided system in terms of the sum throughput achieved by the greedy policy. We have already proved the sum throughput optimality of the greedy policy in the original system when $N = 2$, in Section III. Thus the sum capacity of the original system is given by (19).

We now proceed to prove that (19) is the sum capacity of the genie-aided system as well by examining the sum throughput optimality of the greedy policy in the genie-aided system. For any control interval m, we rewrite the net expected reward from (3) for the genie aided system below.

$$\begin{aligned}
&V_m^{\text{genie}}(\pi_m, \{\mathfrak{A}_k\}_{k\leq m}) \\
&\quad = \quad R_m(\pi_m, a_m) + \mathrm{E}[V_{m-1}^{\text{genie}}(\pi_{m-1}, \{\mathfrak{A}_k\}_{k\leq m-1})|\pi_m, a_m].
\end{aligned}$$

Note that since the current channel state of both the users ($S_m(1)$ and $S_m(2)$) are available at the base station at the end of the control interval m, the belief vector π_{m-1} and hence the expected future reward $\mathrm{E}[V_{m-1}^{\text{genie}}]$ are independent of the scheduling decision a_m. Therefore, using a proof technique similar to that of Proposition 1, it can be proved that in any control interval, the net expected reward is maximized by the greedy policy in the genie-aided system. This establishes the sum throughput optimality of the greedy policy in the genie-aided system as well. The proposition thus follows. ■

Insights on the result in Proposition 5 can be obtained by examining the fundamental trade-off involved in the scheduling decisions in the Markov-modeled downlink. Transmission to a scheduled user and eventually obtaining ARQ feedback from that user accomplishes the following two objectives:

- data transmission in the current slot, which influences the current reward R_k.
- probing the channel of a user for future scheduling decisions, which influences the expected reward in future control intervals.

The optimal schedule strikes a balance between these two objectives (that need not always contradict each other).

From the discussion in the proof of Proposition 5, we see that, in the original system, the choice of the user whose channel is probed becomes irrelevant as far as the future reward is concerned[6]. That explains the result in Corollary 2 and why the greedy policy is optimal in the sum throughput sense. Considering the genie-aided system, since the channel

[4] $p + s = 2$ leads to $P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, a trivial case with no steady state.

[5] User a_{k+1} is given higher priority if both channels were ON.

[6] As long as one of the users is probed.

state information of both the users are freely available to the scheduler at the end of the control interval, there is no need to probe the channel of any user to help with future scheduling decisions. This makes the greedy policy the sum throughput optimal for the genie-aided system as well.

B. Stabilizable and the Unstabilizable Rate Regions of the Markov-Modeled Downlink

The stabilizable rate region of a two user queuing system is defined such that if the arrival rate vector (λ_1, λ_2) belongs to this region, then there exists a scheduling policy with per-user throughput vector[7] (μ_1, μ_2) such that

$$(\lambda_1, \lambda_2) \preceq (\mu_1, \mu_2),$$

where $\preceq$ is the point wise inequality. Note that the above condition is necessary and sufficient to ensure the stability of the queues as defined in the beginning of this section. We now introduce the following proposition.

Proposition 6: The Markov-modeled downlink has a stabilizable rate region given by the set of points (λ_1, λ_2) such that

$$\begin{aligned} (\lambda_1, \lambda_2) &\in H_{\text{convex}}(O, A_1, S, A_2) \\ \text{where } O &= (0,0) \\ A_1 &= (p_s, 0) \\ S &= (\frac{C_{\text{sum}}}{2}, \frac{C_{\text{sum}}}{2}) \\ A_2 &= (0, p_s). \end{aligned} \tag{20}$$

with C_{sum} and p_s introduced in the previous subsection. $H_{\text{convex}}(X)$ is the convex hull of the set of points X defined as,

$$\begin{aligned} &H_{\text{convex}}(X) \\ &= \Big\{ \sum_{i=1}^{\text{size}(X)} \beta_i x_i \;\Big|\; x_i \in X, \beta_i \in \mathbb{R}, \beta_i \geq 0, \sum_{i=1}^{\text{size}(X)} \beta_i = 1 \Big\}. \end{aligned}$$

Proof: Refer to Fig. 4 for an idea on the relative positions of the points A_1, A_2, S and O. Due to the inherent symmetry between the users in the system, from the results of the previous subsection, the service rate vector S can be achieved by the greedy policy. The service rate vectors A_1 and A_2 can be achieved by scheduling transmission to only user 1 or 2, respectively, in every control interval. Any service rate on the edges A_1S or A_2S can be achieved by time sharing between the corresponding two policies. From the definition of the convex hull, for any arrival rate vector belonging to the region in (20), we can always find a stabilizing service rate vector on one of the edges A_1S or A_2S. The proposition thus follows. ■

The unstabilizable rate region of a two user queuing system is such that, if (λ_1, λ_2) belongs to this region, then for every scheduling policy with associated service rate vector (μ_1, μ_2),

$$(\lambda_1, \lambda_2) \npreceq (\mu_1, \mu_2).$$

With this definition we introduce the following proposition.

Proposition 7: The Markov-modeled downlink has an unstabilizable rate region given by the set of points (λ_1, λ_2) such that

$$\begin{aligned} (\lambda_1, \lambda_2) &\notin H_{\text{convex}}(O, A_1, B_1, B_2, A_2) \\ \text{where } O &= (0,0) \\ A_1 &= (p_s, 0) \\ B_1 &= \big(p_s p + (1-p_s)^2 r, (1-p_s) p_s p\big) \\ A_2 &= (0, p_s) \\ B_2 &= \big((1-p_s) p_s p, p_s p + (1-p_s)^2 r\big). \end{aligned} \tag{21}$$

Proof: The relative positions of the points A_1, A_2, B_1, B_2, S and O are illustrated in Fig. 4. We proceed by showing that (21) is an unstabilizable region of the genie-aided system. This can be established if we prove that any scheduling scheme in the genie-aided system achieves a service rate vector within the convex hull $H_{\text{convex}}(O, A_1, B_1, B_2, A_2)$. Consider a broad class of schedulers in the genie-aided system, with each member identified by the parameters $\alpha_i \in [0,1]$, $i \in \{1, \ldots, 4\}$. A member of this class follows the following decision logic in control interval k:

- If $\begin{bmatrix} S_{k+1}(1) \\ S_{k+1}(2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, schedule user 1 with probability α_1 and user 2 w. p. $1 - \alpha_1$.
- If $\begin{bmatrix} S_{k+1}(1) \\ S_{k+1}(2) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $a_k = \begin{cases} 1 \ w.p. \ \alpha_2 \\ 2 \ w.p. \ 1 - \alpha_2 \end{cases}$
- If $\begin{bmatrix} S_{k+1}(1) \\ S_{k+1}(2) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $a_k = \begin{cases} 1 \ w.p. \ \alpha_3 \\ 2 \ w.p. \ 1 - \alpha_3 \end{cases}$
- If $\begin{bmatrix} S_{k+1}(1) \\ S_{k+1}(2) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $a_k = \begin{cases} 1 \ w.p. \ \alpha_4 \\ 2 \ w.p. \ 1 - \alpha_4 \end{cases}$

Since $\begin{bmatrix} S_{k+1}(1) \\ S_{k+1}(2) \end{bmatrix}$ is a sufficient statistic for $\begin{bmatrix} S_k(1) \\ S_k(2) \end{bmatrix}$, any scheduling scheme in the genie-aided system falls under the above class of schedulers or will have a member of this class achieving the same service rate vector as itself. Thus it is now sufficient to prove that the service rate vector achieved by any member of this class belongs to $H_{\text{convex}}(O, A_1, B_1, B_2, A_2)$.

With $\alpha_1 \ldots \alpha_4 \in [0,1]$ fixed, the service rate for user 1 is given by

$$\begin{aligned} \mu_1 &= \sum_{i,j \in \{0,1\}} P\Big(\begin{bmatrix} S_{k+1}(1) \\ S_{k+1}(2) \end{bmatrix} = \begin{bmatrix} i \\ j \end{bmatrix} \Big) \times \\ &\quad P\Big(a_k = 1 \Big| \begin{bmatrix} S_{k+1}(1) \\ S_{k+1}(2) \end{bmatrix} = \begin{bmatrix} i \\ j \end{bmatrix} \Big) \times \\ &\quad P(S_k(1) = 1 | S_{k+1}(1) = i) \\ &= (1-p_s)^2 \alpha_1 r + (1-p_s) p_s \alpha_2 r + p_s (1-p_s) \alpha_3 p \\ &\quad + p_s^2 \alpha_4 p, \end{aligned} \tag{22}$$

with $p_s = \frac{r}{1-(p-r)}$. Similarly,

$$\begin{aligned} \mu_2 &= (1-p_s)^2 (1-\alpha_1) r + (1-p_s) p_s (1-\alpha_2) p \\ &\quad + p_s (1-p_s)(1-\alpha_3) r + p_s^2 (1-\alpha_4) p, \end{aligned} \tag{23}$$

and the sum throughput is given by,

$$\mu_1 + \mu_2 = p_s + (1-p_s) p_s (p-r)(\alpha_3 - \alpha_2).$$

[7] Will henceforth be referred to as service rate vector.

We see that the values of α_1 and α_4 are irrelevant from the sum throughput point of view. Consider the following two cases.

Case 1, when $\alpha_3 \le \alpha_2$:

$$0 \le \mu_1 + \mu_2 \le p_s.$$

Since $A_1(1) + A_1(2) = A_2(1) + A_2(2) = p_s$, we have

$$(\mu_1, \mu_2) \in H_{\text{convex}}(O, A_1, A_2). \qquad (24)$$

Case 2, when $\alpha_3 > \alpha_2$:

$$p_s < \mu_1 + \mu_2 \le p_s + (1-p_s)p_s(p-r) = p_sp + (1-p_s)p_s.$$

Since $B_1(1) + B_1(2) = B_2(1) + B_2(2) = p_sp + (1-p_s)^2r + (1-p_s)p_sp = p_sp + (1-p_s)p_s$, we can find points $E_{A_1B_1}$ and $E_{A_2B_2}$ on edges A_1B_1 and A_2B_2 respectively, such that $E_{A_1B_1}(1) + E_{A_1B_1}(2) = E_{A_2B_2}(1) + E_{A_2B_2}(2) = \mu_1 + \mu_2$. Any point $X_{A_1B_1}$ on the edge A_1B_1 can be written as a convex combination of points A_1 and B_1, i.e., $\exists\, \beta \in [0,1]$ such that

$$\begin{aligned} X_{A_1B_1} &= A_1\beta + B_1(1-\beta) \\ &= \Big(p_s\beta + (p_sp + (1-p_s)^2r)(1-\beta), \\ &\qquad (1-p_s)p_sp(1-\beta)\Big). \end{aligned}$$

With $\beta = 1 - (\alpha_3 - \alpha_2)$, we have $X_{A_1B_1}(1) + X_{A_1B_1}(2) = \mu_1 + \mu_2$. Thus

$$\begin{aligned} E_{A_1B_1} = &\Big(p_s(1-(\alpha_3-\alpha_2)) + (p_sp + (1-p_s)^2r)(\alpha_3-\alpha_2), \\ &\qquad (1-p_s)p_sp(\alpha_3-\alpha_2)\Big). \end{aligned}$$

Due to the symmetry between A_1, B_1 and A_2, B_2, we have $E_{A_2B_2} = (E_{A_1B_1}(2),\ E_{A_1B_1}(1))$. Using μ_1 from (22), it can be shown that, for any $\alpha_{i\in\{1\ldots4\}} \in [0,1]$ with $\alpha_3 > \alpha_2$,

$$E_{A_2B_2}(1) \le \mu_1 \le E_{A_1B_1}(1). \qquad (25)$$

Since $E_{A_1B_1}(1) + E_{A_1B_1}(2) = E_{A_2B_2}(1) + E_{A_2B_2}(2) = \mu_1 + \mu_2$, (25) translates to,

$$(\mu_1, \mu_2) \in H_{\text{convex}}(E_{A_1B_1}, E_{A_1B_1}).$$

The above relation along with the fact that $E_{A_1B_1} \in H_{\text{convex}}(A_1, B_1)$ and $E_{A_2B_2} \in H_{\text{convex}}(A_2, B_2)$ yields,

$$(\mu_1, \mu_2) \in H_{\text{convex}}(A_1, B_1, B_2, A_2). \qquad (26)$$

Combining the results in (24) and (26), we establish that the region in (21) is an unstabilizable region for the genie-aided system. The proposition thus follows. ■

The stabilizable and the unstabilizable rate region results are summarized in Fig. 5

Corollary 8: The stability region of the genie-aided system is given by the set of points (λ_1, λ_2) such that

$$(\lambda_1, \lambda_2) \in H_{\text{convex}}(O, A_1, B_1, B_2, A_2),$$

with O, A_1, B_1, B_2 and A_2 defined as before.

Proof: Consider a scheduler belonging to the class introduced in the previous discussion. With $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\} = \{1, 0, 1, 1\}$, from (22) and (23), it can be seen that the service rate vector $(\mu_1, \mu_2) = B_1$. With $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\} = \{0, 0, 1, 0\}$, a service rate vector $(\mu_1, \mu_2) = B_2$ can be achieved. Both these policies are fundamentally greedy in nature. However, the former gives priority to user 1 while the latter to user 2. We have already seen that points A_1 and points A_2 can be achieved by scheduling to only user 1 or 2 respectively. Points along the edges A_1B_1, B_1B_2 and B_2A_2 can be achieved by time sharing between the corresponding two policies. The preceding arguments establish $H_{\text{convex}}(O, A_1, B_1, B_2, A_2)$ as a stabilizable region for the genie-aided downlink. We have already established that $H_{\text{convex}}(O, A_1, B_1, B_2, A_2)$ is an unstabilizable region as well, hence proving the corollary. ■

VII. Conclusion

We have considered the problem of scheduling under partial channel state information in a Markov-modeled downlink with ARQ feedback. Using POMDP formulation, we have shown that, for $N \le 3$ users, a simple greedy policy that maximizes the current reward is optimal in terms of sum throughput. By developing a simple round-robin based implementation that does not require the statistics of the underlying Markov chain, we have shown that the greedy policy is attractive from a practical point of view. We have also derived a sufficient condition for the optimality of the greedy policy in the general N user case. We conjectured that the greedy policy satisfies this condition and is hence optimal for any number of users in the system. By establishing an equivalence with a genie-aided system, a simple expression for the sum capacity of the Markov-modeled downlink system has been derived when $N = 2$. Assuming random arrivals in the queues at the base station, we have studied the stabilizable and the unstabilizable rate regions of the downlink system for the two user case. Before we conclude, note that our problem is a special case of the *restless multi-armed bandit* problem [25]. This problem has been shown to be PSPACE-hard to solve in general [26]. Thus our result on the optimality of the greedy policy may be of importance in understanding the properties of the optimal policy in the general restless multi-armed bandit processes.

Appendix I
Proof of Proposition 4

We rewrite the sufficient condition from (18) below. For any $m > 1$,

$$\begin{aligned} \Big[\hat{V}_{m-1}\big([1\ X\ 0], [1\ 2\ 3]\big) - \hat{V}_{m-1}\big([1\ 0\ X], [1\ 2\ 3]\big)\Big] &\le 1 \\ \text{when } n &= 1, \\ \Big[\hat{V}_{m-1}\big([Y\ 1\ 0], [1\ 2\ 3]\big) - \hat{V}_{m-1}\big([1\ Y\ 0], [1\ 2\ 3]\big)\Big] &\le 1 \\ \text{when } n &= 2, \end{aligned}$$

where X, Y are binary numbers and $\{\mathfrak{A}_k\}_{k \le m-1} = \{\widehat{\mathfrak{A}}_k\}_{k \le m-1}$.

We first consider $n = 1$. $X = 0$ is a trivial case. Hence, we focus on $X = 1$. With $\hat{V}_0 = 0$ and $1 \le k \le m-1$, using an expansion along the lines of (13), we have

$$\begin{aligned}
&\hat{V}_k\big([1\ 1\ 0],[1\ 2\ 3]\big)\\
&\quad= p+P_{S_k|S_{k+1}=[110]}([0\ 0\ 0])\hat{V}_{k-1}\big([0\ 0\ 0],[2\ 3\ 1]\big)\\
&\qquad+P_{S_k|S_{k+1}=[110]}([0\ 0\ 1])\hat{V}_{k-1}\big([0\ 0\ 1],[2\ 3\ 1]\big)\\
&\qquad+P_{S_k|S_{k+1}=[110]}([0\ 1\ 0])\hat{V}_{k-1}\big([0\ 1\ 0],[2\ 3\ 1]\big)\\
&\qquad+P_{S_k|S_{k+1}=[110]}([0\ 1\ 1])\hat{V}_{k-1}\big([0\ 1\ 1],[2\ 3\ 1]\big)\\
&\qquad+P_{S_k|S_{k+1}=[110]}([1\ 0\ 0])\hat{V}_{k-1}\big([1\ 0\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[110]}([1\ 0\ 1])\hat{V}_{k-1}\big([1\ 0\ 1],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[110]}([1\ 1\ 0])\hat{V}_{k-1}\big([1\ 1\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[110]}([1\ 1\ 1])\hat{V}_{k-1}\big([1\ 1\ 1],[1\ 2\ 3]\big).
\end{aligned}$$

Note that the schedule order vector evolves according to (12). Using the symmetry property of (14), we rewrite

$$\begin{aligned}
&\hat{V}_k\big([1\ 1\ 0],[1\ 2\ 3]\big)\\
&\quad= p+P_{S_k|S_{k+1}=[110]}([0\ 0\ 0])\hat{V}_{k-1}\big([0\ 0\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[110]}([0\ 0\ 1])\hat{V}_{k-1}\big([0\ 1\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[110]}([0\ 1\ 0])\hat{V}_{k-1}\big([1\ 0\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[110]}([0\ 1\ 1])\hat{V}_{k-1}\big([1\ 1\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[110]}([1\ 0\ 0])\hat{V}_{k-1}\big([1\ 0\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[110]}([1\ 0\ 1])\hat{V}_{k-1}\big([1\ 0\ 1],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[110]}([1\ 1\ 0])\hat{V}_{k-1}\big([1\ 1\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[110]}([1\ 1\ 1])\hat{V}_{k-1}\big([1\ 1\ 1],[1\ 2\ 3]\big).
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\hat{V}_k\big([1\ 0\ 1],[1\ 2\ 3]\big)\\
&\quad= p+P_{S_k|S_{k+1}=[101]}([0\ 0\ 0])\hat{V}_{k-1}\big([0\ 0\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[101]}([0\ 0\ 1])\hat{V}_{k-1}\big([0\ 1\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[101]}([0\ 1\ 0])\hat{V}_{k-1}\big([1\ 0\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[101]}([0\ 1\ 1])\hat{V}_{k-1}\big([1\ 1\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[101]}([1\ 0\ 0])\hat{V}_{k-1}\big([1\ 0\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[101]}([1\ 0\ 1])\hat{V}_{k-1}\big([1\ 0\ 1],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[101]}([1\ 1\ 0])\hat{V}_{k-1}\big([1\ 1\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[101]}([1\ 1\ 1])\hat{V}_{k-1}\big([1\ 1\ 1],[1\ 2\ 3]\big).
\end{aligned}$$

From the preceding equations,

$$\begin{aligned}
&\hat{V}_k\big([1\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_k\big([1\ 0\ 1],[1\ 2\ 3]\big)\\
&\quad=pqr\Big(\hat{V}_{k-1}\big([1\ 0\ 1],[1\ 2\ 3]\big)-\hat{V}_{k-1}\big([1\ 1\ 0],[1\ 2\ 3]\big)\Big)\\
&\qquad+p^2s\Big(\hat{V}_{k-1}\big([1\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_{k-1}\big([1\ 0\ 1],[1\ 2\ 3]\big)\Big)\\
&\qquad+q^2r\Big(\hat{V}_{k-1}\big([0\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_{k-1}\big([1\ 0\ 0],[1\ 2\ 3]\big)\Big)\\
&\qquad+qps\Big(\hat{V}_{k-1}\big([1\ 0\ 0],[1\ 2\ 3]\big)-\hat{V}_{k-1}\big([0\ 1\ 0],[1\ 2\ 3]\big)\Big)\\
&\quad=\Big[p\Big(\hat{V}_{k-1}\big([1\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_{k-1}\big([1\ 0\ 1],[1\ 2\ 3]\big)\Big)\\
&\qquad+q\Big(\hat{V}_{k-1}\big([1\ 0\ 0],[1\ 2\ 3]\big)-\hat{V}_{k-1}\big([0\ 1\ 0],[1\ 2\ 3]\big)\Big)\Big]\times\\
&\qquad(p-r). \qquad (27)
\end{aligned}$$

Now we examine two key quantities in (27).

$$\begin{aligned}
&\hat{V}_k\big([1\ 0\ 0],[1\ 2\ 3]\big)\\
&\quad= p+P_{S_k|S_{k+1}=[100]}([0\ 0\ 0])\hat{V}_{k-1}\big([0\ 0\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[100]}([0\ 0\ 1])\hat{V}_{k-1}\big([0\ 1\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[100]}([0\ 1\ 0])\hat{V}_{k-1}\big([1\ 0\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[100]}([0\ 1\ 1])\hat{V}_{k-1}\big([1\ 1\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[100]}([1\ 0\ 0])\hat{V}_{k-1}\big([1\ 0\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[100]}([1\ 0\ 1])\hat{V}_{k-1}\big([1\ 0\ 1],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[100]}([1\ 1\ 0])\hat{V}_{k-1}\big([1\ 1\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[100]}([1\ 1\ 1])\hat{V}_{k-1}\big([1\ 1\ 1],[1\ 2\ 3]\big).
\end{aligned}$$

Note that the symmetry property of (14) is used in the above expansion. Similarly,

$$\begin{aligned}
&\hat{V}_k\big([0\ 1\ 0],[1\ 2\ 3]\big)\\
&\quad= r+P_{S_k|S_{k+1}=[010]}([0\ 0\ 0])\hat{V}_{k-1}\big([0\ 0\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[010]}([0\ 0\ 1])\hat{V}_{k-1}\big([0\ 1\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[010]}([0\ 1\ 0])\hat{V}_{k-1}\big([1\ 0\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[010]}([0\ 1\ 1])\hat{V}_{k-1}\big([1\ 1\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[010]}([1\ 0\ 0])\hat{V}_{k-1}\big([1\ 0\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[010]}([1\ 0\ 1])\hat{V}_{k-1}\big([1\ 0\ 1],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[010]}([1\ 1\ 0])\hat{V}_{k-1}\big([1\ 1\ 0],[1\ 2\ 3]\big)\\
&\qquad+P_{S_k|S_{k+1}=[010]}([1\ 1\ 1])\hat{V}_{k-1}\big([1\ 1\ 1],[1\ 2\ 3]\big).
\end{aligned}$$

Together we have

$$\begin{aligned}
&\hat{V}_k\big([1\ 0\ 0],[1\ 2\ 3]\big)-\hat{V}_k\big([0\ 1\ 0],[1\ 2\ 3]\big)\\
&\quad=\Big[1-r\Big(\hat{V}_{k-1}\big([1\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_{k-1}\big([1\ 0\ 1],[1\ 2\ 3]\big)\Big)\Big]\times\\
&\qquad(p-r). \qquad (28)
\end{aligned}$$

Thus, with $\hat{V}_0=\hat{V}_{-1}=0$, (27) becomes

$$\begin{aligned}
&\hat{V}_k\big([1\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_k\big([1\ 0\ 1],[1\ 2\ 3]\big)\\
&\quad=\Bigg[p\Big(\hat{V}_{k-1}\big([1\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_{k-1}\big([1\ 0\ 1],[1\ 2\ 3]\big)\Big)\\
&\qquad+\Big[1-r\Big(\hat{V}_{k-2}\big([1\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_{k-2}\big([1\ 0\ 1],[1\ 2\ 3]\big)\Big)\Big]\times\\
&\qquad q(p-r)\Bigg](p-r). \qquad (29)
\end{aligned}$$

For a fixed $k\le m-1$, if $\hat{V}_{k-2}\big([1\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_{k-2}\big([1\ 0\ 1],[1\ 2\ 3]\big)\in[0,1]$ and $\hat{V}_{k-1}\big([1\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_{k-1}\big([1\ 0\ 1],[1\ 2\ 3]\big)\in[0,1]$ then, since $p\ge r$, from (29), $\hat{V}_k\big([1\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_k\big([1\ 0\ 1],[1\ 2\ 3]\big)\ge 0$. To examine the upper boundedness, we expand

$$\begin{aligned}
&\hat{V}_k\big([1\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_k\big([1\ 0\ 1],[1\ 2\ 3]\big)\\
&= \Bigg[p\Big(\hat{V}_{k-1}\big([1\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_{k-1}\big([1\ 0\ 1],[1\ 2\ 3]\big)\Big)\\
&\quad+\Big[\Big(\hat{V}_{k-2}\big([1\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_{k-2}\big([1\ 0\ 1],[1\ 2\ 3]\big)\Big)\times\\
&\quad qr(r-p)\Big]+qp-qr\Bigg](p-r)\\
&\leq (p-r)\big[p+qr+q-qr\big] \text{ (under the given condition)}\\
&\leq 1.
\end{aligned}$$

Therefore, we have the following statement. For any $k \leq m-1$,

$$\begin{aligned}
&\text{If } \hat{V}_{k-2}\big([1\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_{k-2}\big([1\ 0\ 1],[1\ 2\ 3]\big)\in[0,1]\\
&\text{and } \hat{V}_{k-1}\big([1\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_{k-1}\big([1\ 0\ 1],[1\ 2\ 3]\big)\in[0,1]\\
&\text{then } \hat{V}_{k}\big([1\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_{k}\big([1\ 0\ 1],[1\ 2\ 3]\big)\in[0,1].
\end{aligned} \tag{30}$$

Note that $\hat{V}_1\big([1\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_1\big([1\ 0\ 1],[1\ 2\ 3]\big)=0$ and from (27), $\hat{V}_2\big([1\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_2\big([1\ 0\ 1],[1\ 2\ 3]\big)=q(p-r)^2\in[0,1]$. From the preceding observations along with statement (30), we have, by induction, $\forall\ k\leq m-1$,

$$0\leq\hat{V}_k\big([1\ X\ 0],[1\ 2\ 3]\big)-\hat{V}_k\big([1\ 0\ X],[1\ 2\ 3]\big)\leq 1. \tag{31}$$

This completes the first part of the proof, i.e, the $n=1$ case. The second part, $n=2$ case, is proved if we show

$$\hat{V}_{m-1}\big([Y\ 1\ 0],[1\ 2\ 3]\big)-\hat{V}_{m-1}\big([1\ Y\ 0],[1\ 2\ 3]\big)\leq 1.$$

$Y=1$ is a trivial case. When $Y=0$, from (28) and (31), $\forall\ k\leq m-1$,

$$0\leq\hat{V}_k\big([1\ 0\ 0],[1\ 2\ 3]\big)-\hat{V}_k\big([0\ 1\ 0],[1\ 2\ 3]\big)\leq 1.$$

This proves the second part.

Appendix II
Key Quantities

N	:	Number of users in the downlink environment
p	:	prob (channel is ON in the current slot | channel was ON in the previous slot)
q	:	prob (channel is OFF in the current slot | channel was ON in the previous slot)
r	:	prob (channel is ON in the current slot | channel was OFF in the previous slot)
s	:	prob (channel is OFF in the current slot | channel was OFF in the previous slot)
a_k	:	Index of the user scheduled in control interval k
π_k	:	Belief vector in the control interval k
R_k	:	Expected current reward in the control interval k
V_k	:	Net expected reward in the control interval k
η_{sum}	:	Sum throughput
$\mathfrak{A}_k$	:	Scheduling policy applied in the control interval k
$\widehat{\mathfrak{A}}_k$	:	Greedy scheduling policy applied in the control interval k
$\mathfrak{A}_k^*$	:	Optimal scheduling policy
S_k	:	State vector such that $S_k(i)$ indicates the state (1-ON/0-OFF) of the channel of user i in control interval k
O_k	:	Schedule order vector in the control interval k, the ordered arrangement of the index of the users in decreasing order of $\pi_k(i)$
C_{sum}	:	Sum capacity of the downlink environment
p_s	:	Steady state ON probability of the Markov channels

References

[1] R. Knopp and P. A. Humblet, “Information capacity and power control in single cell multiuser communications,” *Proc. IEEE International Conference on Communications,* (Seattle, WA), pp. 331-335, June 1995.

[2] P. Viswanath, D. Tse, and R. Laroia, “Opportunistic beamforming using dumb antennas,” *IEEE Transactions on Information Theory,* vol. 48, no. 6, pp. 1277-1294, Jun. 2002.

[3] R. W. Heath, M. Airy, and A. J. Paulraj, “Multiuser diversity for MIMO wireless systems with linear receivers,” *Proc. Asilomar Conf. Signals, Systems, and Computers,* (Pacific Grove, CA), pp. 1194-1199, Nov. 2001.

[4] A. Gyasi-Agyei, “Multiuser diversity based opportunistic scheduling for wireless data networks,” *IEEE Communications Letters,* vol. 9, issue 7, pp. 670-672, Jul. 2005.

[5] J. Chung, C. S. Hwang, K. Kim, and Y. K. Kim, “A random beamforming technique in MIMO systems exploiting multiuser diversity,” *IEEE Journal on Selected Areas in Communications,* vol. 21, pp. 848-855, Jun. 2003.

[6] S. Murugesan, E. Uysal-Biyikoglu and P. Schniter, “Optimization of Training and Scheduling in the Non-Coherent MIMO Multiple-Access Channel,” *IEEE Journal on Selected Areas in Communications,* vol. 25, no. 7, pp. 1446-1456, Sep. 2007.

[7] X. Liu, E. K. P. Chong, and N. B. Shroff, “Opportunistic Transmission Scheduling with Resource-Sharing Constraints in Wireless Networks,” *IEEE Journal on Selected Areas in Communications,* vol. 19, pp. 2053-2064, Oct. 2001.

[8] E. Gilbert, “Capacity of a burst-noise channel,” *Bell Systems Technical Journal,* vol. 39, pp. 1253-1266, 1960.

[9] S. Lu, V. Bharghavan, and R. Srikant, “Fair scheduling in wireless packet networks,” *IEEE/ACM Transactions on Networking,* vol. 7, no. 4, pp. 473-489, Aug. 1999.

[10] T. Nandagopal, S. Lu, and V. Bharghavan, “A unified architecture for the design and evaluation of wireless fair queueing algorithms,” *Proc. ACM Mobicom,* Aug. 1999.

[11] T. Ng, I. Stoica, and H. Zhang, “Packet fair queueing algorithms for wireless networks with location-dependent errors,” *Proc IEEE INFOCOM,* (New York), vol. 3, 1998.

[12] S. Shakkottai and R. Srikant, "Scheduling real-time traffic with deadlines over a wireless channel," *Proc. ACM Workshop on Wireless and Mobile Multimedia,* (Seattle, WA), Aug. 1999.

[13] Y. Cao and V. Li, "Scheduling algorithms in broadband wireless networks," *Proc. IEEE,* vol. 89, no. 1, pp. 76-87, Jan. 2001.

[14] M. Zorzi and R. Rao, "Error control and energy consumption in communications for nomadic computing," *IEEE Transactions on Computers,* vol. 46, pp. 279-289, Mar. 1997.

[15] L. A. Johnston and V. Krishnamurthy, "Opportunistic File Transfer over a Fading Channel: A POMDP Search Theory Formulation with Optimal Threshold Policies," *IEEE Transactions on Wireless Communications,* vol. 5, no. 2, Feb. 2006.

[16] S. Lin, D. Costello, and M. Miller, "Automatic-repeat-request error control schemes," *IEEE Communications Magazine,* vol. 22, pp. 5-17, Dec. 1984.

[17] D. L. Lu and J. F. Chang, "Performance of ARQ protocols in nonindependent channel errors," *IEEE Transactions on Communications,* vol. 41, pp. 721-730, May 1993.

[18] M. Zorzi, R. R. Rao, and L. B. Milstein, "ARQ error control on fading mobile radio channels," *IEEE Transactions on Vehicular Technology,* vol. 46, pp. 445-455, May 1997.

[19] Y. J. Cho and C. K. Un, "Performance analysis of ARQ error controls under Markovian block error pattern," *IEEE Transactions on Communications,* vol. 42, pp. 2051-2061, Feb.-Apr. 1994.

[20] R. D. Smallwood and E. J. Sondik, "The Optimal Control of Partially Observable Markov Processes Over a Finite Horizon," *Operations Research,* Sep. 1973.

[21] S. Christian Albright, "Structural Results for Partially Observable Markov Decision Processes," *Operations Research,* vol. 27, no. 5, pp. 1041-1053, Sep.-Oct. 1979.

[22] C. C. White and W. Scherer, "Solution procedures for partially observed Markov decision processes," *Operations Research,* pp. 791-797, 1985.

[23] G. E. Monahan, "A survey of partially observable Markov decision processes: Theory, Models, and Algorithms," *Management Science,* vol. 28, no. 1, pp. 1-16, Jan. 1982.

[24] Q. Zhao, B. Krishnamachari, and K. Liu, "On Myopic Sensing for Multi-Channel Opportunistic Access," submitted to *IEEE Transactions on Wireless Communications*, November, 2007. (`http://www.ece.ucdavis.edu/˜qzhao/RecentPublication.html`).

[25] P. Whittle, "Restless Bandits: Activity Allocation in a Changing World," *Journal of Applied Probability,* vol. 25, pp. 287-298, 1988.

[26] C. H. Papadimitriou, J. N. Tsitsiklis, "The complexity of optimal queueing network control," *Mathematics of Operations Research,* vol. 24, no. 2, pp. 293305, May 1999.

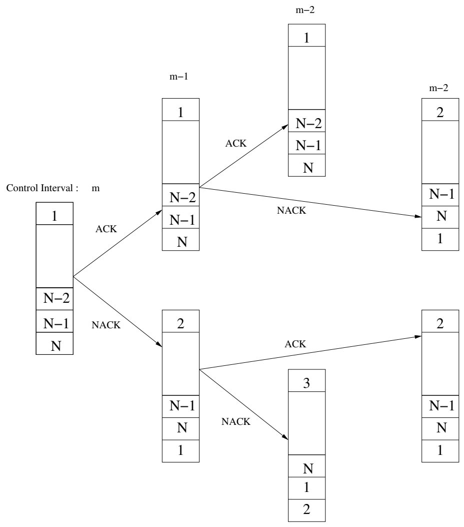

Fig. 1. Users are ordered according to their belief values in the current interval. The illustration shows how ARQ feedback rearranges the order of the users in a round robin fashion.

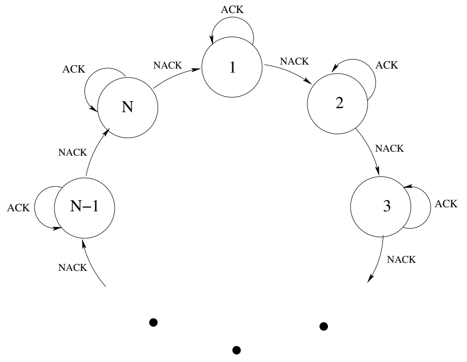

Fig. 2. Illustration showing how NACK feedback stimulates the scheduler to choose the next user in the queue.

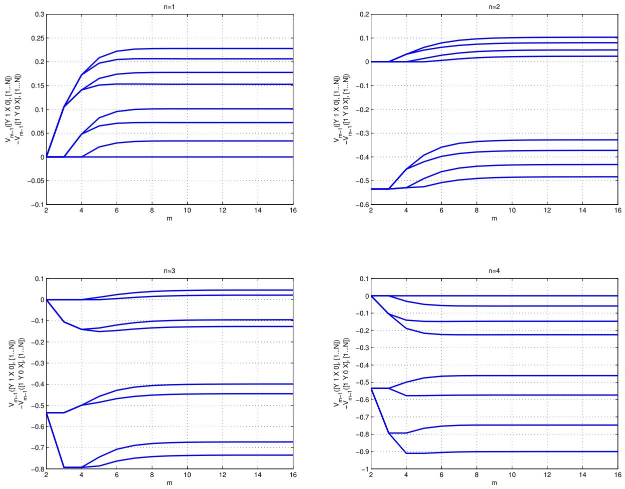

Fig. 3. Illustration showing $\hat{V}_{m-1}([Y \quad 1 \quad X \quad 0], [1 \ldots N]) - \hat{V}_{m-1}([1 \ Y \ 0 \ X], [1 \ldots N]) \leq 1 \ \forall n \in \{1 \ldots N-1\}$ with $N = 5$ users.

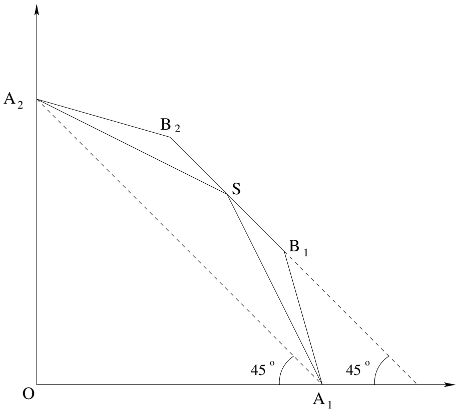

Fig. 4. Relative locations of points involved in Propositions 6 and 7.

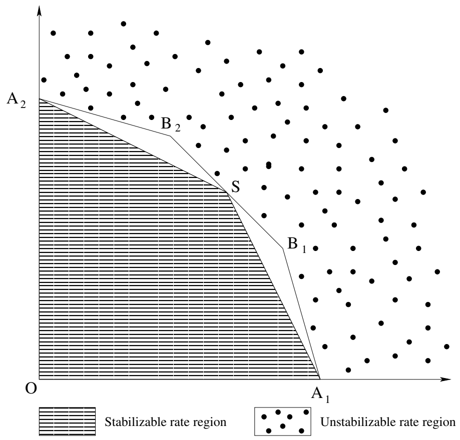

Fig. 5. Stabilizable and Unstabilizable rate regions for the Markov-modeled downlink.

PLACE PHOTO HERE

Sugumar Murugesan received the B.E. degree in Electronics and Communication Engineering from the College of Engineering, Anna University, India in 2004. He received the M.S. degree in Electrical and Computer Engineering from the Ohio State University (OSU), USA in 2006. He is a recipient of the OSU University Fellowship (2004-05). He is currently working towards the Ph.D. degree in Electrical and Computer Engineering at OSU. His current research interests include communication theory, queueing theory and wireless networks.

PLACE PHOTO HERE

Philip Schniter (S03M93SM05) received the B.S. and M.S. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 1992 and 1993, respectively. In 2000, he received the Ph.D. degree in electrical engineering from Cornell University, Ithaca, NY. From 1993 to 1996, he was with Tektronix Inc., Beaverton, OR, as a Systems Engineer. Subsequently, he joined the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH, where he is now an Associate Professor. His research interests include signal processing, communication theory, and wireless networks. Dr. Schniter received the National Science Foundation CAREER Award in 2003, and he currently serves on the IEEE Signal Processing for Communications Technical Committee.

PLACE PHOTO HERE

Ness B. Shroff (S91 / M93 / SM01/ F07) is currently the Ohio Eminent Scholar of Networking and Communications, and Professor of ECE and CSE at The Ohio State University. Previously, he was a Professor of ECE at Purdue University and the director of the Center for Wireless Systems and Applications (CWSA), a university-wide center on wireless systems and applications. His research interests span the areas of wireless and wireline communication networks, where he investigates fundamental problems in the design, performance, pricing, and security of these networks. Dr. Shroff has received numerous awards for his networking research, including the NSF CAREER award, the best paper awards for IEEE INFOCOM 06 and IEEE INFOCOM08, the best paper award for IEEE IWQoS06, the best paper of the year award for the Computer Networks journal, and the best paper of the year award for the Journal of Communications and Networks (JCN) (his IEEE INFOCOM05 paper was one of two runner-up papers).