Internet Mathematics Vol. 9, Nos. 2–3: 136–160

A State Action Frequency Approach to Throughput Maximization over Uncertain Wireless Channels

Krishna Jagannathan, Shie Mannor, Ishai Menache, and Eytan Modiano

Abstract. We consider scheduling over a wireless system in which the channel state information is not available a priori to the scheduler but can be inferred from past history. Specifically, the wireless system is modeled as a network of parallel queues. We assume that the channel state of each queue evolves stochastically as an independent ON/OFF Markov chain. The scheduler, which is aware of the queue lengths but is ignorant of the channel states, has to choose at most one queue at a time for transmission. The scheduler has no information regarding the current channel states but can estimate them from the acknowledgment history.

We first characterize the capacity region of the system using tools from the theory of Markov decision processes (MDPs). Specifically, we prove that the capacity region boundary is the uniform limit of a sequence of linear programming (LP) solutions. Next, we combine the LP solution with a queue-length-based scheduling mechanism that operates over long frames to obtain a throughput optimal policy for the system. By incorporating results from MDP theory within the Lyapunov-stability framework, we show that our frame-based policy stabilizes the system for all arrival rates that lie in the interior of the capacity region.

ISSN: 1542-7951 print

1. Introduction

In this paper, we consider a scheduling problem in a wireless uplink or downlink system in which there is no explicit instantaneous channel state information (CSI) available to the scheduler. The lack of CSI may arise in practice for several reasons. For example, the control overheads, as well as the delay and energy costs associated with channel probing, might make instantaneous CSI too costly or impractical to obtain.

Our system consists of N wireless links, which are modeled as N parallel queues that are fed by stochastic traffic. Due to the shared wireless medium, only a single queue can be chosen at each time slot for transmitting its data. The channel quality (or state) of each wireless link is time-varying, evolving as an independent ON/OFF Markov chain. A given transmission is successful only if the underlying channel is currently in the ON state.

Our basic assumption in this paper is that the scheduler cannot observe the current state of any of the wireless links. Nonetheless, when the scheduler serves one of the queues in a given time slot t, there is an ACK feedback mechanism that acknowledges whether the transmission was successful, thereby revealing the channel state a posteriori. Since the channels are correlated across time by the Markovian assumption, this a posteriori CSI can be used for predicting the channel state of the chosen queue in future time slots. We emphasize that the ACK mechanism is the only means by which CSI is made available to the scheduler. From a practical viewpoint, this ACK mechanism is natural and should be available as part of the underlying LLC/MAC protocol.

The *capacity region* (or the rate region) of the system described above is the set of all arrival-rate vectors that are stably supportable by some behavioristic scheduling policy. Our aim is to characterize the capacity region of the system and to design a throughput-optimal scheduling policy.

The general problem of scheduling parallel queues with time-varying connectivity has been widely studied for almost two decades. The seminal paper [Tassiulas and Ephremides 93] considered the case in which both channel states and queue lengths are fully available to the scheduler. It was shown in that paper that the *max-weight algorithm*, which serves the longest connected queue, is throughput optimal. Notably, the algorithm stabilizes all rates in the capacity region without requiring any a priori knowledge of the arrival rates.

Following that paper, several variants of imperfect and delayed CSI scenarios have been considered in the literature; see, e.g., [Pantelidou et al. 09, Ying and Shakkottai 08, Ying and Shakkottai 09, Gopalan et al. 12] and references therein. However, our scheduling problem differs fundamentally from the models considered in those references. Specifically, no explicit CSI is ever made available

to the scheduler, and acquiring channel state information is a part of the scheduling decision made at each time instant. This adds significant difficulties to the scheduling problem.

Two recent papers consider the scheduling problem in which the CSI is obtained through an acknowledgment process, as in our model. In [Ahmad et al. 09], the authors consider the objective of maximizing the *sum rate* of the system, under the assumption that the queues are *fully backlogged* (i.e., there is always data to send in each queue). It is shown that a simple *myopic policy* is sum rate optimal. The suggested policy keeps scheduling the channel that is being served as long as it remains ON, and switches to the least recently served channel when the current channel goes OFF.

In [Li and Neely 11], the authors propose a randomized round-robin scheduling policy for the system, which is inspired by the myopic sensing results in [Ahmad et al. 09]. Their policy is shown to stabilize arrivals that lie within an inner bound to the rate region. However, their policy is not throughput optimal, and their method cannot be used to characterize the capacity region.

In this paper, we propose a throughput-optimal scheduling policy for the system. In particular, the policy we propose can stabilize arrival rates that lie arbitrarily close to the capacity region boundary, with a corresponding tradeoff in the computational complexity. We also provide a characterization of the capacity region boundary as the limit of a sequence of LP solutions.

The scheduling problem we consider is related to the celebrated restless bandits problem [Whittle 88], which is known to be computationally difficult in general. In fact, every point on the boundary of the capacity region can be implicitly expressed as the optimal solution to a restless bandits problem. Such a solution involves solving an MDP with a countably infinite state space. Since obtaining this solution may be computationally and analytically prohibitive, we approximate the original MDP by a finite-state MDP with a "tunable" number of states. We then employ a linear programming approach to solve the resulting finite-state MDP [Puterman 94].

We prove that the solution to the LP approximates the boundary of the capacity region arbitrarily closely, where the accuracy of the approximation improves with the number of states in the underlying finite MDP. Thus, there is a tradeoff between the accuracy of the approximation and the dimensionality of the LP.

Next, we combine the LP solution with a queue-length-based scheduling mechanism that operates over long time frames to obtain a dynamic scheduling policy for the system. Our main result establishes that this "frame-based" policy is *throughput optimal*, i.e., can stably support all arrival rates in the interior of the capacity region. Our proof of throughput optimality combines tools from Markov decision theory within a Lyapunov-stability framework.

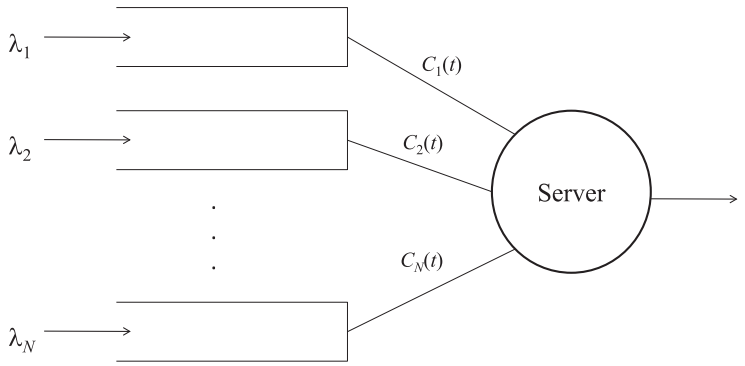

Figure 1. A system of parallel queues served by a single server. The channels connecting the queues to the server are randomly time-varying.

The remainder of this paper is organized as follows. The model is presented in Section 2. In Section 3, we formulate a linear program that leads to the characterization of the capacity region. In Section 4, we suggest the frame-based policy, which we prove to be throughput optimal. We conclude the paper in Section 5.

2. System Description

2.1. The Network Model

We model the wireless system as consisting of N parallel queues (see Figure 1). Time is slotted ($t = 1, 2, \dots$). Packets arrive at each queue $i \in \{1, 2, \dots, N\}$ according to an independent stochastic process with rate λ_i. We assume that the arrival processes are independent of each other, and independent and identically distributed (i.i.d.) from slot to slot. We further assume that the number of arrivals in a slot at each of the queues has a finite variance.

Due to the shared wireless medium, only a single transmission is allowed at a given time. In our queuing model, this is equivalent to having the queues connected to a single server belonging to one of the queues that is capable of serving only a single packet per slot. Each queue is connected to the server by an ON/OFF channel, which models the time-varying channel quality of the underlying wireless link. If a particular channel is OFF and the queue is chosen by the scheduler, the packet has to be retransmitted to avoid transmission failure. If it is ON and chosen by the scheduler, a single packet is properly transmitted, and an ACK is received by the scheduler.

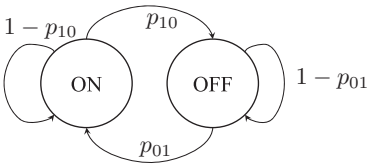

Figure 2. The Markov chain governing the time evolution of each channel's state $C_i(t)$.

We denote the channel state of the ith link at time t by $C_i(t) \in \{\text{ON}, \text{OFF}\}$, $i = 1, \ldots, N$. We assume that the states of different channels are statistically independent of each other. The time evolution of each of the channels is given by a two-state ON/OFF Markov chain (see Figure 2). Although our methodology allows for different Markov chains for different channels, we shall assume for ease of notation and exposition that the Markov chains are identically distributed across users, with the structure shown in Figure 2. We further assume that $p_{01} + p_{10} < 1$, so that each channel is positively correlated in time. The steady-state probability of a channel being in the ON state is given by

$$\pi_{\text{ON}} = \frac{p_{01}}{p_{01} + p_{10}}. \tag{2.1}$$

2.2. Information Structure

At each time t, we assume that the scheduler knows the current queue lengths $Q_i(t)$ prior to making the scheduling decision. Yet no information about the current channel conditions is made available to the scheduler. Only after scheduling a particular queue does the scheduler get to know whether the transmission succeeded, by virtue of the ACK mechanism. The scheduler thus has access to the entire history of transmission successes and failures. However, due to the Markovian nature of the channels, it is sufficient to record how long ago each channel was served, and the state of the channel (ON/OFF) when it was last served. In addition to the above, the scheduler also knows precisely the statistical properties of each of the channels (i.e., the Markov chain of Figure 2).

2.3. Scheduling Objective

Given the above information structure, our objective is to design a scheduling policy that can support the largest possible set of input rates. More precisely, an arrival-rate vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$ is said to be *supportable* if there exists some scheduling policy under which the queue lengths are finite (almost surely). The *capacity region* Γ of the system is the closure of all supportable rate vectors. A

policy is said to be *throughput optimal* if it can support all arrival rates in the interior of Γ.

3. Optimal Policies for a Fully Backlogged System

In the interest of simplicity of notation and exposition, we restrict attention hereinafter to the case of $N = 2$ queues, although our methodology extends naturally to more queues. In this section, we assume that the queues are fully backlogged, i.e., the queues are never empty. As we shall see, our analysis of the fully backlogged system gives us insights into the optimal scheduling policy for dynamic systems with finite queues.

Since the queues are assumed to be infinitely backlogged in this section, the state of the system is completely specified by the state of each channel the last time it was served and how long ago each channel was served. In a system with two fully backlogged queues, the *information state* during slot t has the form $\mathbf{s}(t) = [k_1(t), b_1(t), k_2(t), b_2(t)]$, where $k_i(t)$ is the number of slots since queue i was served, and $b_i(t) \in \{0, 1\}$ is the state of the channel the last time it was observed.[1] Since the channels are Markovian, $\mathbf{s}(t)$ is a sufficient statistic for the fully backlogged system. Note that $\min(k_1(t), k_2(t)) = 1$ for all t and $\max(k_1(t), k_2(t)) \geq 2$ for all t. Let $\mathcal{S}$ denote the (countably infinite) set of all possible states $\mathbf{s}(t)$.

Denote the l-step transition probabilities of the channel Markov chain in Figure 2 by $p_{11}^{(l)}$, $p_{01}^{(l)}$, $p_{10}^{(l)}$, and $p_{00}^{(l)}$. It can be shown by explicit computation that for $l \geq 1$,

$$p_{01}^{(l)} = \frac{p_{01}[1-(1-q)^l]}{q}, \quad p_{10}^{(l)} = \frac{p_{10}[1-(1-q)^l]}{q},$$
$$p_{00}^{(l)} = \frac{p_{10}+p_{01}(1-q)^l}{q}, \quad p_{11}^{(l)} = \frac{p_{01}+p_{10}(1-q)^l}{q},$$

where $q = p_{01} + p_{10}$. Next, define the *belief vector* corresponding to state $\mathbf{s} \in \mathcal{S}$ as $[\omega_1(\mathbf{s}), \omega_2(\mathbf{s})]$, where $\omega_i(\mathbf{s})$, $i = 1, 2$, is the conditional probability that channel i is ON. For example, if $\mathbf{s} = [1, \text{ON}, 3, \text{OFF}]$, then the corresponding belief vector is $\left[p_{11}, p_{01}^{(3)}\right]$. It can be shown that the belief vector has a one-to-one mapping to the information state and is therefore also a sufficient statistic for the fully backlogged problem.

[1] Throughout, 0 is used interchangeably to denote the channel state OFF, and 1 is used to denote ON.

In each slot, there are two possible actions $a \in \{1, 2\}$, corresponding to serving one of the two queues. Given a state and an action at a particular time, the belief for the next slot is updated according to the following equation:

$$\omega_i(t+1) = \begin{cases} p_{11}\omega_i(t) + p_{01}(1 - \omega_i(t)), & \text{if } a(t) \neq i, \\ p_{11}, & \text{if } a(t) = i,\ C_{a(t)}(t) = 1, \\ p_{01}, & \text{if } a(t) = i,\ C_{a(t)}(t) = 0, \end{cases}$$

where we have abused notation to write $\omega_i(t) = \omega_i(\mathbf{s}(t))$.

A *policy* for the fully backlogged system is a rule that associates an action $a(t) \in \{1, 2\}$ to the state $\mathbf{s}(t)$ for each t. A *deterministic stationary* policy is a map from $\mathcal{S}$ to $\{1, 2\}$, whereas a *randomized stationary* policy picks an action given the state according to a fixed distribution $\mathbb{P}\{a \mid \mathbf{s}(\cdot)\}$.

Suppose that a unit reward is accrued from each of the two channels every time a packet is successfully transmitted on that channel, i.e., when the server is assigned to a particular channel and the channel is ON. Given a state $\mathbf{s}(t)$ at a particular time and an action $a(t)$, the probability that a unit reward is accrued in that time slot is simply equal to the belief of the channel that was chosen. We are interested in the long-term time-average rate achieved on each of the channels under a given policy. From the viewpoint of the reward defined above, the average rate translates to the infinite-horizon time-average reward obtained on each channel under a given policy.

We say that rate pair (λ_1, λ_2) is *achievable* in the fully backlogged system if there exists some policy for which the infinite-horizon time-average reward vector equals (λ_1, λ_2). The closure of the set of all achievable rate pairs is called the *rate region* Λ of the fully backlogged system. It should be evident that a rate pair that is not achievable in the fully backlogged system cannot be supportable in the dynamic system with finite queues. Thus, the capacity region Γ of the queuing system is contained in the rate region Λ of the fully backlogged system. In fact, we show in Section 4 that the two rate regions have the same interior, by deriving a queue-length-based policy for the original system that can stabilize any arrival rate in the interior of Λ. We now proceed to obtain an implicit characterization of the rate region boundary.

3.1. An MDP Formulation and State Action Frequencies

Let us consider a Markov decision process (MDP) formulation of the belief space for characterizing the rate region boundary.

It is easy to show that the rate region Λ is convex. Indeed, given two points in the rate region each attainable by some policy, we can obtain every convex

combination of the rate points by time-sharing the policies over sufficiently long intervals. Further, the rate region is also closed by definition. Therefore, every point on its boundary maximizes a weighted sum rate expression. That is, if (r_1^*, r_2^*) is a rate pair on the boundary of Λ, then

$$(r_1^*, r_2^*) = \arg\max_{(\lambda_1,\lambda_2)\in\Lambda} w_1\lambda_1 + w_2\lambda_2 \tag{3.1}$$

for some weight vector $\mathbf{w} = [w_1,\ w_2]$, with $w_1 + w_2 = 1$. The following proposition shows that if the rate pair (λ_1, λ_2) is in Λ, then there necessarily exists a *state action frequency* vector whose entries satisfy a set of balance equations.

Proposition 3.1. *Let $(\lambda_1, \lambda_2) \in \Lambda$. Then for each state $\boldsymbol{s} \in \mathcal{S}$ and action $a \in \{1, 2\}$, there exists a state action frequency $x(\boldsymbol{s}; a)$ that satisfies*

$$0 \le x(\boldsymbol{s}; a) \le 1, \tag{3.2}$$

the balance equations

$$\begin{aligned}
&x([1, \text{ON}, k, b_2]; 1) + x([1, \text{ON}, k, b_2]; 2) && (3.3)\\
&\quad = x([1, \text{ON}, k-1, b_2]; 1)p_{11} + x([1, \text{OFF}, k-1, b_2]; 1)p_{01},\ \ k > 2,\\
&x([1, \text{OFF}, k, b_2]; 1) + x([1, \text{OFF}, k, b_2]; 2) && (3.4)\\
&\quad = x([1, \text{OFF}, k-1, b_2]; 1)p_{00} + x([1, \text{ON}, k-1, b_2]; 1)p_{10}, \quad k > 2,\\
&x([1, \text{ON}, 2, b_2]; 1) + x([1, \text{ON}, 2, b_2]; 2) && (3.5)\\
&\quad = \sum_{l\ge 2} \left(x([l, \text{ON}, 1, b_2]; 1)p_{11}^{(l)} + x([l, \text{OFF}, 1, b_2]; 1)p_{01}^{(l)} \right),\\
&x([1, \text{OFF}, 2, b_2]; 1) + x([1, \text{OFF}, 2, b_2]; 2) && (3.6)\\
&\quad = \sum_{l\ge 2} \left(x([l, \text{OFF}, 1, b_2]; 1)p_{00}^{(l)} + x([l, \text{ON}, 1, b_2]; 1)p_{10}^{(l)} \right),\\
&x([k, b_1, 1, \text{ON}]; 1) + x([k, b_1, 1, \text{ON}]; 2) && (3.7)\\
&\quad = x([k-1, b_1, 1, \text{ON}]; 2)p_{11} + x([k-1, b_1, 1, \text{OFF}]; 2)p_{01}, \quad k > 2,\\
&x([k, b_1, 1, \text{OFF}]; 1) + x([k, b_1, 1, \text{OFF}]; 2) && (3.8)\\
&\quad = x([k-1, b_1, 1, \text{OFF}]; 2)p_{00} + x([k-1, b_1, 1, \text{ON}]; 2)p_{10}, \quad k > 2,\\
&x([2, b_1, 1, \text{ON}]; 1) + x([2, b_1, 1, \text{ON}]; 2) && (3.9)\\
&\quad = \sum_{l\ge 2} \left(x([1, b_1, l, \text{ON}]; 2)p_{11}^{(l)} + x([1, b_1, l, \text{OFF}]; 2)p_{01}^{(l)} \right),\\
&x([2, b_1, 1, \text{OFF}]; 1) + x([2, b_1, 1, \text{OFF}]; 2) && (3.10)\\
&\quad = \sum_{l\ge 2} \left(x([1, b_1, l, \text{OFF}]; 2)p_{00}^{(l)} + x([1, b_1, l, \text{ON}]; 2)p_{10}^{(l)} \right),
\end{aligned}$$

where $b_1, b_2 \in \{\text{ON}, \text{OFF}\}$*, the normalization condition*

$$\sum_{s \in \mathcal{S}} (x(\boldsymbol{s}; 1) + x(\boldsymbol{s}; 2)) = 1, \tag{3.11}$$

and the rate constraints

$$\lambda_i \leq \sum_{\boldsymbol{s} \in \mathcal{S}} x(\boldsymbol{s}; i)\omega_i(\boldsymbol{s}), \quad i = 1, 2. \tag{3.12}$$

Proof. The result follows from the linear programming formulation of countable MDPs; see [Altman 99]. □

Intuitively, a state action frequency vector corresponds to a stationary randomized policy such that $x(\mathbf{s}; a)$ equals the steady-state probability that in a given time slot, the state is $\mathbf{s}$ and the action is a. Further, conditioned on being in state $\mathbf{s}$, the action a is chosen with probability $x(\mathbf{s}; a)/\mathbb{P}\{\mathbf{s}\}$, where $\mathbb{P}\{\mathbf{s}\} = x(\mathbf{s}; 1) + x(\mathbf{s}; 2)$. (If $\mathbb{P}\{\mathbf{s}\} = 0$, the policy prescribes actions arbitrarily.)

Let us now provide an intuitive explanation of the balance equations. Equations (3.4)–(3.10) simply equate the steady-state probability of being in a particular state with the total probability of entering that state from all possible states. For example, the left-hand side of (3.4) equals the steady-state probability of being in the state $[1, \text{ON}, k, b_2]$, $k > 2$, while the right-hand side equals the total probability of getting to the above state from other states, and similarly for the other balance equations. Equation (3.11) equates the total steady-state probability with unity. Finally, in (3.12), the term $x(\mathbf{s}; i)\omega_i(\mathbf{s})$ equals the steady-state probability that the state is $\mathbf{s}$, the action i is chosen, and the transmission succeeds. Thus, the right-hand side of (3.12) equals the total expected rate on channel i.

We now return to the characterization of the rate region boundary. In light of Proposition 3.1, (3.1) can be rewritten as follows.

Problem 3.2. (INFINITE($\mathbf{w}$).)

$$(r_1^*, r_2^*) = \arg\max_{(\lambda_1, \lambda_2)} w_1\lambda_1 + w_2\lambda_2 \tag{3.13}$$

subject to (3.2)–(3.12).

Since the number of state spaces of the MDP is countably infinite, the optimization in (3.13) involves an infinite number of variables. In order to make this problem tractable, we now introduce an LP approximation.

3.2. LP Approximation Using a Finite MDP

In this section, we introduce an MDP with a finite state space, which, as we show, approximates the original MDP arbitrarily closely. The state action frequencies corresponding to the finite MDP approximation can then be solved as a linear program.

First note that the belief of a channel that has not been observed for a long time increases monotonically toward the steady-state value of π_{ON} if it was OFF the last time it was scheduled. Similarly, the belief decreases monotonically to π_{ON} if the channel was ON the last time it was scheduled. These observations follow from the l-step transition probabilities given in Section 3. The key idea now is to construct a finite MDP whose states are the same as those of the original MDP, with the exception that the belief of a channel that remains unobserved for a long time is clamped to the steady-state ON probability π_{ON}. Specifically, when a channel has not been scheduled for τ or more time slots, its observation history is entirely forgotten, and the belief on it is assumed to be π_{ON}. The action space and the reward structure are exactly as before. We show that this truncated finite MDP approximates the original MDP better and better as τ gets large.

Let us now specify the states and state action frequencies for this finite MDP. There are $4(\tau-2)$ states of the form $[1, b_1, k_2, b_2],\ 2 \le k_2 \le \tau-1,\ b_1, b_2 \in \{\text{ON}, \text{OFF}\}$, which correspond to the first channel being scheduled in the previous slot and the second channel being scheduled fewer than τ time slots ago. In a symmetric fashion, there are $4(\tau-2)$ states of the form $[k_1, b_1, 1, b_2],\ 2 \le k_1 \le \tau-1,\ b_1, b_2 \in \{\text{ON}, \text{OFF}\}$, which correspond to the second channel being scheduled in the previous slot. Finally, there are four states $[1, b_1, \phi, \phi]$, $b_1 \in \{\text{ON}, \text{OFF}\}$ and $[\phi, \phi, 1, b_2]$, $b_2 \in \{\text{ON}, \text{OFF}\}$ in which one of the channels has not been seen for at least τ slots and its belief has been reset to π_{ON}. Let us denote by $\hat{\mathcal{S}}$ the above set of states for the finite MDP, and let $\hat{x}(\mathbf{s}; a),\ \mathbf{s} \in \hat{\mathcal{S}},\ a \in \{1, 2\}$ denote the state action frequencies for the finite MDP. These state action frequencies satisfy

$$0 \le \hat{x}(\mathbf{s}; a) \le 1, \tag{3.14}$$

$$\sum_{s \in \hat{\mathcal{S}}} \hat{x}(\mathbf{s}; 1) + \hat{x}(\mathbf{s}; 2) = 1, \tag{3.15}$$

$$\hat{\lambda_i} \le \sum_{\mathbf{s} \in \hat{\mathcal{S}}} \hat{x}(\mathbf{s}; i)\omega_i(\mathbf{s}), \quad i = 1, 2, \tag{3.16}$$

and a set of balance equations analogous to (3.4)–(3.10).

For a fixed $\mathbf{w}$ and τ, let us now consider the following LP.

Problem 3.3. (FINITE($\tau, \mathbf{w}$).)

$$(\hat{r}_1, \hat{r}_2) = \arg\max_{(\hat{\lambda}_1, \hat{\lambda}_2)} w_1 \hat{\lambda}_1 + w_2 \hat{\lambda}_2 \tag{3.17}$$

subject to (3.14)–(3.16) and the balance equations.

The main result of this section shows that the solution to this LP approximates the boundary point specified by the problem INFINITE($\mathbf{w}$) for every $\mathbf{w}$ when τ is large.

Proposition 3.4. *For a given $\boldsymbol{w}$ with $w_1 + w_2 = 1$, and τ, let $\hat{\boldsymbol{r}}(\tau, \boldsymbol{w})$ denote the solution to the problem* FINITE($\tau, \boldsymbol{w}$)*, and let $\boldsymbol{r}^*(\boldsymbol{w})$ denote the solution to* INFINITE($\boldsymbol{w}$)*. Then $\hat{\boldsymbol{r}}(\tau, \boldsymbol{w})$ converges uniformly to $\boldsymbol{r}^*(\boldsymbol{w})$ as $\tau \to \infty$. In other words, given any $\kappa > 0$ and any $\boldsymbol{w}$, there exists $\tau_0 > 0$ that depends on κ but not on $\boldsymbol{w}$ such that for all $\tau > \tau_0$, we have*

$$|\hat{\boldsymbol{r}}(\tau, \boldsymbol{w}) - \boldsymbol{r}^*(\boldsymbol{w})| < \kappa.$$

Proof. The convergence of $\hat{\mathbf{r}}(\tau, \mathbf{w})$ to $\mathbf{r}^*(\mathbf{w})$ for a fixed $\mathbf{w}$ follows from the classical work in [Whitt 78, Whitt 79]. The difficulty is in proving that the convergence is uniform across all $\mathbf{w}$. Without loss of generality, we assume that $\mathbf{w} = (x, 1-x)$ for $x \in [0, 1]$. The main observation here is that the function $f_\tau : [0, 1] \to \mathbb{R}$ that takes an element x and returns $\hat{\mathbf{r}}(\tau, (x, 1-x))$ is a convex function for every τ, since it is the solution of a parametric linear program [Bertsimas and Tsitsiklis 97]. It also follows that $f_\tau(0)$ and $f_\tau(1)$ are the same for all τ (since these are the cases in which only one of the channels matters). Let us define the function $f_\infty(x) : [0, 1] \to \mathbb{R}$ to be the function that takes x and returns $\mathbf{r}^*((x, 1-x))$. Take a finite grid of points on $[0, 1]$ denoted by G. We have convergence for every $g \in G$ of $f_\tau(g)$ to $f_\infty(g)$ [Whitt 78, Whitt 79]. Since these are all convex functions, the uniform convergence for all values of x follows; see [Rockafellar 70]. □

We next prove a result that asserts that using the state action frequencies obtained from a finite MDP in a backlogged system entails only a negligible suboptimality when τ is large. The finite-MDP solution is applied to the backlogged system as follows. If the state in the backlogged system is such that both channels were served no more than τ time slots ago, then we schedule according to the state action frequencies of that particular state in the finite MDP. On the other hand, if one of the channels was last served more than τ time slots ago, the finite MDP will not have a corresponding state and state action frequencies. In such a case, we schedule according to the state action frequencies of one of the

four states in the finite MDP in which the belief is clamped to the steady-state value. For example, if the system state is $[1, b_1, k_2, b_2]$, with $k_2 > \tau$, we schedule according to the state action frequencies of the state $[1, b_1, \phi, \phi]$ in the finite MDP, and so on.

Proposition 3.5. *Suppose the optimal state action frequencies obtained by solving the problem* FINITE$(\tau, \boldsymbol{w})$ *are used to perform scheduling in a fully backlogged system, as detailed above. Let* $\bar{\boldsymbol{r}}(\tau, \boldsymbol{w})$ *denote the average reward vector so obtained. Then for every* $\boldsymbol{w}$ *with* $w_1 + w_2 = 1$*, we have that* $\bar{\boldsymbol{r}}(\tau, \boldsymbol{w})$ *converges uniformly to the optimal reward* $\boldsymbol{r}^*(\boldsymbol{w})$ *as* $\tau \to \infty$*.*

Proof outline. Proposition 3.4 asserts that $\hat{\mathbf{r}}(\tau, \mathbf{w})$ converges to $\mathbf{r}^*(\mathbf{w})$ uniformly. It therefore suffices to prove that $\bar{\mathbf{r}}(\tau, \mathbf{w})$ converges uniformly to $\hat{\mathbf{r}}(\tau, \mathbf{w})$. In words, we need to prove that the evaluation of the optimal policy of the truncated MDP that is evaluated on the truncated MDP ($\hat{\mathbf{r}}(\tau, \mathbf{w})$) converges to the evaluation of this policy on the infinite MDP ($\bar{\mathbf{r}}(\tau, \mathbf{w})$) uniformly with respect to $\mathbf{w}$. Indeed, we will prove a stronger result claiming that this holds for every stationary policy for the finite MDP and not just for optimal policies under some $\mathbf{w}$.

Suppose that we are given a stationary policy Π defined on the truncated MDP with a "memory" of τ, and let Π_∞ be the extension of Π to the infinite state space as discussed above. To proceed with the proof, we imitate the methodology of [Whitt 78, Whitt 79]. While the details are lengthy and technical, the main observation that is required to obtain uniform convergence is that the reward that is obtained in the finite MDP for Π is obtained in the same states as is obtained for Π_∞ for the infinite MDP (and this is true for all $\mathbf{w}$). The difference between the finite and infinite MDPs in terms of transitions is only in the transitions out of the four additional states $[1, b_1, \phi, \phi]$, $b_1 \in \{\text{ON}, \text{OFF}\}$ and $[\phi, \phi, 1, b_2]$, $b_2 \in \{\text{ON}, \text{OFF}\}$ that have the same policy as the appropriate states where one of the queues was not visited for τ steps (by construction). As long as the transition is within these four states or within the other states that are identical for the truncated and infinite MDPs, the rewards are the same. Once there is a transition out of these states, the conditional transition probability becomes close as τ increases (i.e., exiting each of the four states has a conditional probability that becomes closer to the conditional probability on exiting the matching states in the infinite MDP). The fact that the transitions are becoming closer makes the values of the policies similar uniformly over all policies. □

We pause momentarily to emphasize the subtle difference between Propositions 3.4 and 3.5. Proposition 3.4 asserts that the optimal reward obtained from the finite MDP is close to the optimal reward of the infinite MDP. In this case,

the optimal solution to the finite MDP is applied to the finite state space. On the other hand, in Proposition 3.5, the optimal policy obtained from the finite MDP is used on the original *infinite* state space, and the ensuing reward is shown to be close to the optimal reward. From a practical perspective, Proposition 3.4 is useful in obtaining a characterization of the rate region, while Proposition 3.5 plays a key role in the throughput optimality proof of the frame-based policy.

3.3. An Outer Bound

We now derive an outer bound to the rate region Λ, using "genie-aided" channel information. Although the bound is not used in deriving our optimal policy, it is of interest to compare the outer bound we obtain to existing bounds in the literature.

Consider a fictitious fully backlogged system in which the channel processes follow the same sample paths as in the original system. However, after a channel is served in a particular time slot, a genie reveals the states of all the channels in the system. Therefore, at the beginning of a time slot in the fictitious system, the scheduler has access to all the channel states in the previous slot, and not just the channel that was served. Clearly, the rate region boundary for the genie-aided system is an outer bound to the rate region of the original system.

Let us compute the above outer bound for our two-user system. Indeed, there are only four possibilities for the channel states in the previous slots: $\{\text{ON}, \text{ON}\}$, $\{\text{OFF}, \text{ON}\}$, $\{\text{ON}, \text{OFF}\}$, and $\{\text{OFF}, \text{OFF}\}$. Furthermore, since the two channels are independent, the states above occur with probabilities π_{ON}^2, $\pi_{\text{ON}}(1-\pi_{\text{ON}})$, $\pi_{\text{ON}}(1-\pi_{\text{ON}})$, and $(1-\pi_{\text{ON}})^2$ respectively, in steady state. Using these facts, we can obtain the rate region for the genie-aided fictitious system.

Indeed, let Λ_{00} be the convex hull of the vectors $(p_{01}, 0)$ and $(0, p_{01})$. Intuitively, Λ_{00} is the set of all rate vectors that are achievable exclusively in the time slots with $\{\text{OFF}, \text{OFF}\}$ as the channel states in the previous slot. Similarly, let $\Lambda_{01} = \mathcal{C}\{(p_{01}, 0), (0, p_{11})\}$, $\Lambda_{10} = \mathcal{C}\{(p_{11}, 0), (0, p_{01})\}$, and $\Lambda_{11} = \mathcal{C}\{(p_{11}, 0), (0, p_{11})\}$, where $\mathcal{C}$ stands for convex hull. Then the rate region of the fictitious system is given by

$$\overline{\Lambda} = \left\{\boldsymbol{\lambda} \geq 0 \mid \boldsymbol{\lambda} = \left((1-\pi_{\text{ON}})^2 \boldsymbol{\lambda}_{00} + \pi_{\text{ON}}(1-\pi_{\text{ON}})(\boldsymbol{\lambda}_{01} + \boldsymbol{\lambda}_{10}) + \pi_{\text{ON}}^2 \boldsymbol{\lambda}_{11}\right)\right\},$$

where $\boldsymbol{\lambda_{00}} \in \Lambda_{00}$, etc.

3.4. A Numerical Example

In this section, we use the finite LP approximation obtained in Section 3.2 to numerically compute and plot the capacity region for a two-user system.

Specifically, we use the solution to the problem FINITE$(\tau, \mathbf{w})$ with large enough τ, which, according to Proposition 3.4, uniformly approximates the rate region boundary for all $\mathbf{w}$. We also plot the genie-aided outer bound obtained above, and compare our rate region and outer bound to the inner and outer bounds derived in [Li and Neely 11]. We assume in this section that the channel Markov chains have a symmetric structure with $p_{10} = p_{01} = \epsilon$, so that $\pi_{\text{ON}} = 0.5$.

Figures 3 and 4 show the numerically obtained rate region, the genie-aided outer bound, and the inner and outer bounds derived in [Li and Neely 11] for a symmetric two-user system. Figure 3 is for the case $\epsilon = 0.2$ (higher correlation in time), while Figure 4 is for $\epsilon = 0.4$ (lower correlation in time). The rate region, shown with a dark solid line, was obtained by solving the LP approximation FINITE$(\tau, \mathbf{w})$ for all weight vectors and large enough τ. We observed that $\tau \approx 30$ and $\tau \approx 10$ were sufficient for the cases $\epsilon = 0.2$ and $\epsilon = 0.4$, respectively. The dash-dotted curve in each figure is the genie-aided outer bound, derived in Section 3.3. The achievable region of the randomized round-robin policy proposed

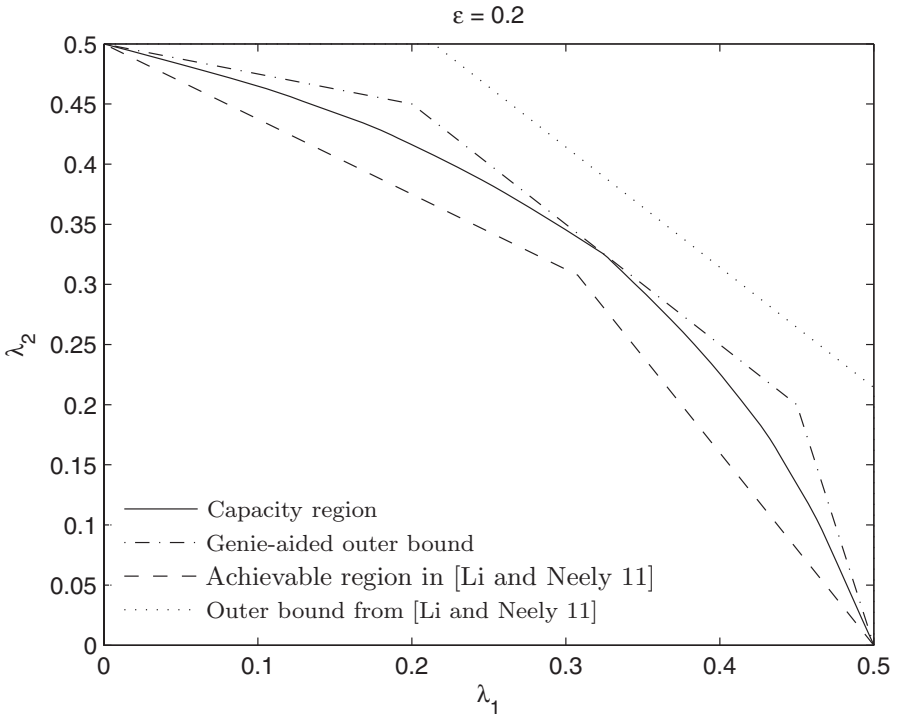

Figure 3. The rate region, our outer bound, and the inner and outer bounds derived in [Li and Neely 11], for $\epsilon = 0.2$.

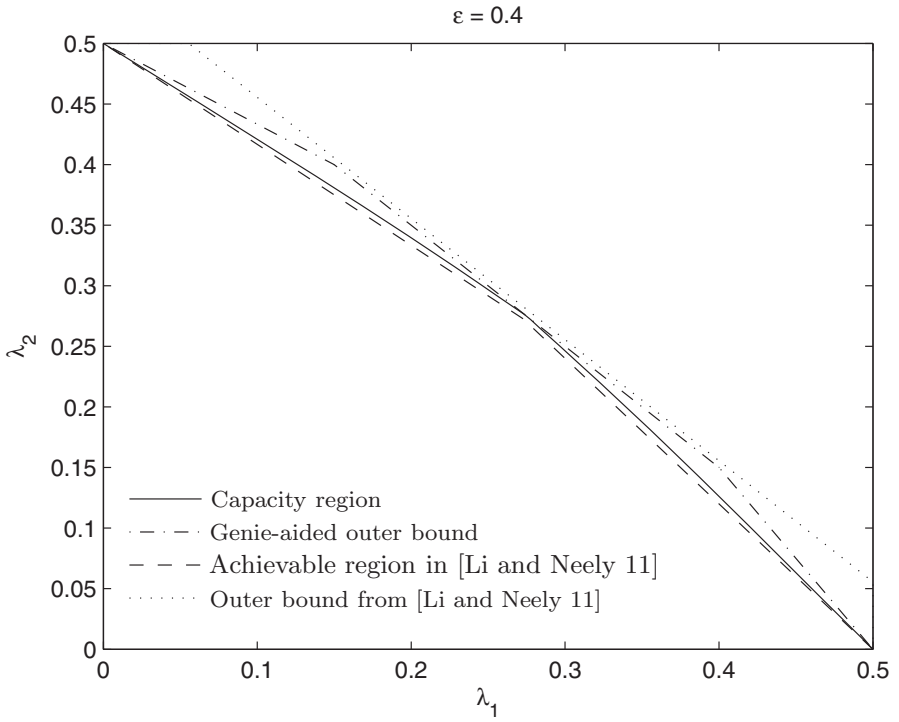

Figure 4. The rate region, our outer bound, and the inner and outer bounds derived in [Li and Neely 11], for $\epsilon = 0.4$.

in [Li and Neely 11] is shown by a dashed line. Finally, the outermost region in the figure is the outer bound derived in [Li and Neely 11]. We observe that the genie-aided bound is uniformly better than the outer bound derived in [Li and Neely 11].

It is evident from Figures 3 and 4 that the genie-aided outer bound is achievable at the symmetric rate point, since the rate region boundary touches the outer bound. To see this analytically, we first determine that the symmetric rate point on the genie-aided outer bound is given by $(3/8 - \epsilon/4, 3/8 - \epsilon/4)$. Next, in the original fully backlogged system, consider a *myopic* policy that stays with a queue as long as its channel remains ON, and switches to the other queue when the channel goes OFF. The sum throughput of this policy can be shown by direct computation to be $3/4 - \epsilon/2$ (see [Ahmad et al. 09], for example). Since this sum throughput is equally shared between the two channels, it follows that the symmetric rate point on the outer bound is achievable. Interestingly, the above argument constitutes a simple optimality proof of myopic sensing for the case of

two symmetric channels. This is a special case of the general optimality result derived in [Ahmad et al. 09] for any N.

4. A Throughput-Optimal Frame-Based Policy

In this section, we return to the original problem, with finite queues and stochastic arrivals. We propose a throughput-optimal queue-length-based policy that operates over long frames.

In our frame-based policy, the time axis is divided into frames consisting of T slots each, and the queue lengths are updated at the beginning of each frame. Given the queue-length vector $\mathbf{Q}(kT)$ at the beginning of each frame, the idea is to maximize a weighted sum rate quantity over the frame, where the weight vector is the queue-length vector for that frame. The weighted rate maximization is, in turn, performed approximately by solving the finite MDP. Intuitively, the above procedure has the net effect of performing max-weight scheduling over each time frame, where MDP techniques are employed to compute each of the "optimal schedules." More precisely, our policy operates according to Algorithm 1.

Our main result in this section is the throughput optimality of the frame-based policy, for large enough values of T and τ. Specifically, our frame-based policy can stabilize all arrival rates within a δ-stripped region of Λ, for every $\delta > 0$. The δ-stripped region is defined as

$$\Lambda - \delta\mathbf{1} = \{\lambda \mid \lambda + \delta\mathbf{1} \in \Lambda\}.$$

As we shall see, a small δ could require large values of T and τ, which would increase the dimensionality of the LP (depends on τ) as well as the average

Algorithm 1. (Frame-based policy.)

1. At the beginning of time frame k, update the queue-length vector $\mathbf{Q}(kT)$.
2. Compute the normalized queue-length vector $\tilde{\mathbf{Q}}(kT)$, whose entries sum to 1.
3. Solve the problem $\text{FINITE}(\tau, \tilde{\mathbf{Q}}(kT))$ and obtain the state action frequencies $\hat{x}(\mathbf{s}, a)$, $\mathbf{s} \in \hat{\mathcal{S}}$, $a \in \{1, 2\}$.
4. Schedule according to the state action frequencies obtained in the previous step during each slot in the frame, even if it means scheduling an empty queue.

delay (depends on T). Thus our policy offers a tradeoff between computational complexity and delay on the one hand, and better throughput on the other. Our main theorem is stated below. Note also that our policy requires queue-length information only at the beginning of each time frame.

Theorem 4.1. *Given any $\delta > 0$, there exist large enough τ and T such that the frame-based policy stabilizes all arrival rates in the δ-stripped rate region $\Lambda - \delta\mathbf{1}$.*

A proof of the theorem is given in Section 6.

4.1. Simulations of the Frame-Based Policy

We now provide some basic simulation results for the frame-based policy. In Figures 5 and 6, we plot the average queue length of one of the queues, under the frame-based policy, as a function of the arrival rate. We take $\epsilon = 0.25$ and consider a symmetric rate scenario whereby independent Poisson traffic of equal rates feeds the two queues. Each simulation run was carried out over ten thousand frames, with frame sizes of $T = 10$ and $T = 50$ in Figures 5 and 6, respectively.

Under this symmetric traffic scenario, the theoretical boundary of the capacity region lies at $(\lambda_1, \lambda_2) = (0.3125, 0.3125)$. The first observation we make from the figure is that the frame-based policy easily stabilizes arrival rates up to 0.29

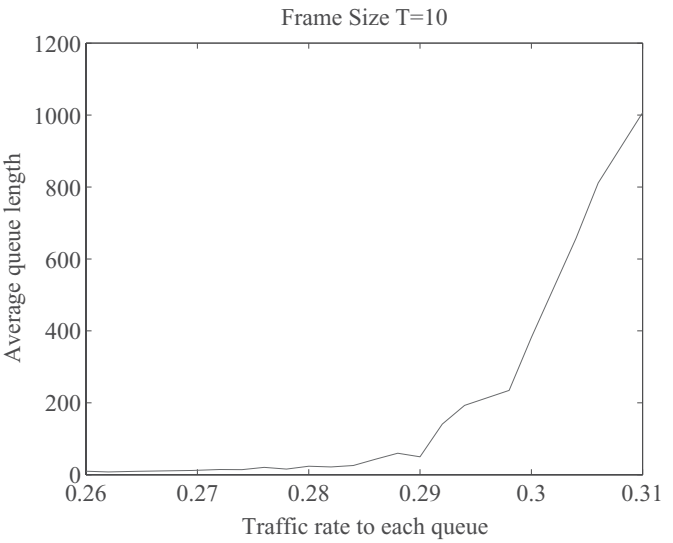

Figure 5. The average queue length as a function of the symmetric arrival rate under the frame-based policy, for $T = 10$.

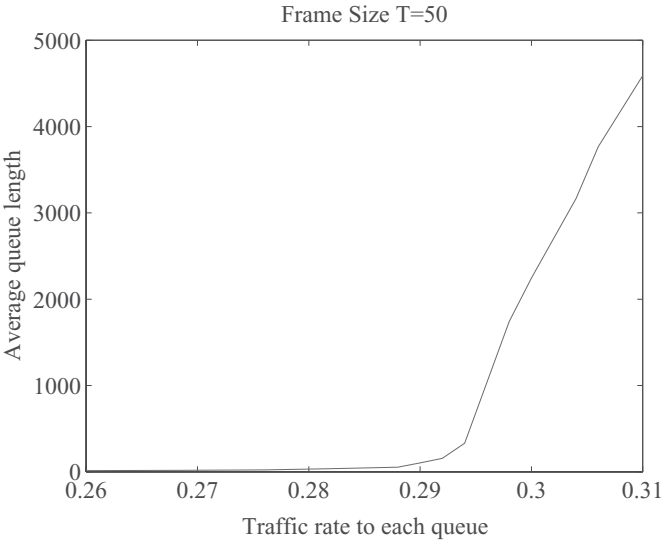

Figure 6. The average queue length as a function of the symmetric arrival rate under the frame-based policy, for $T = 50$.

even for small frame sizes such as $T = 10$. There is considerable queue buildup at $(\lambda_1, \lambda_2) = (0.3, 0.3)$, and very large buildup when the symmetric rate equals 0.31.

Another interesting point to note from the figure is that in heavy traffic, the average queue length when $T = 50$ is roughly a factor of five larger than when $T = 10$. This conforms to the theoretical prediction that the frame-based policy inherently suffers from an $O(T)$ average congestion level in the queues. This implies that although the frame-based policy is theoretically optimal for large T, it is possible that for a given traffic rate, a large frame size leads to considerable delay.

5. Conclusions

In this paper, we have studied the problem of scheduling over uncertain wireless channels, where channel state information can be only indirectly obtained, using past successes and failures of transmissions. We showed that the capacity region boundary for such a system can be approximated arbitrarily well by a sequence of LPs. We then incorporated the LP solution into a queue-length-based scheduling framework to obtain a throughput-optimal policy for the system.

Although we explicitly dealt with a two-user setting with statistically identical channels, our methodology extends naturally to more than two heterogeneous channels. However, when the number of channels becomes asymptotically large, the dimensionality of the LP approximation increases exponentially in the number of channels. In such a case, it may be more practical to resort to the suboptimal policy from [Li and Neely 11]. On the other hand, for relatively small system sizes (say $N = 10$), our method may entail solving an LP with a dimensionality of a few thousands (i.e., $\tau \cdot 2^{10}$), which is by no means prohibitive.

For future work, it would be interesting to obtain *structural properties* of optimal policies for the backlogged system. For example, we believe that threshold policies should be sufficient to achieve the rate region boundary. If this is indeed the case, we can use a simple threshold policy over long frames to obtain a throughput-optimal policy instead of solving a large LP in every frame. Finally, we believe that combining frame-based scheduling with Whittle's indexability [Liu and Zhao 10, Ouyang et al. 10] can lead to computationally simple algorithms that work well in practice.

6. Appendix: Proof of Theorem 4.1

We prove stability of the queuing system under the frame-based policy by calculating the expected Lyapunov drift over each frame. This multistep Lyapunov-drift analysis has been used in the past [Neely et al. 05] to show stability of max-weight scheduling. In our setting, the main challenge lies in establishing that the empirical service rates obtained under the frame-based policy are close to maximizing a weighted sum rate quantity, where the weights are the queue lengths at the beginning of the frame.

Let us define the Lyapunov function

$$L(\mathbf{Q}(t)) = \frac{1}{2}\sum_i Q_i^2(t)$$

and the corresponding conditional drift over a frame

$$\Delta_T(kT) = \mathbb{E}\left[L(\mathbf{Q}((k+1)T)) - L(\mathbf{Q}(kT)) \mid \mathbf{Q}(kT)\right].$$

Let $A_i(\cdot)$ and $D_i(\cdot)$, respectively, denote the arrival and departure processes from the ith queue. The evolution of queue i is given by

$$Q_i(t+1) = Q_i(t) + A_i(t) - D_i(t).$$

Next, define $\hat{D}_i(t)$ as the departure process from a fully backlogged system when our frame-based policy is used on it. That is, $\hat{D}_i(t)$ is the same as the departure

process $D_i(t)$, except there are no lost departures due to empty queues. For each queue i, we have

$$Q_i((k+1)T) \leq \max\left(Q_i(kT) - \sum_{\sigma=0}^{T-1} \hat{D}_i(kT+\sigma), 0\right) + \sum_{\sigma=0}^{T-1} A_i(kT+\sigma).$$

The above expression is an inequality because some arrivals during the frame may leave during the same frame. Squaring both sides, and noting that $\max^2(x,0) \leq x^2$, we have

$$\begin{aligned} Q_i^2((k+1)T) &\leq \left(Q_i(kT) - \sum_{\sigma=0}^{T-1} \hat{D}_i(kT+\sigma)\right)^2 + \left(\sum_{\sigma=0}^{T-1} A_i(kT+\sigma)\right)^2 \\ &\quad + 2\left(\sum_{\sigma=0}^{T-1} A_i(kT+\sigma)\right) Q_i(kT) \\ &= Q_i^2(kT) + \left(\sum_{\sigma=0}^{T-1} \hat{D}_i(kT+\sigma)\right)^2 + \left(\sum_{\sigma=0}^{T-1} A_i(kT+\sigma)\right)^2 \\ &\quad + 2Q_i(kT)\left(\sum_{\sigma=0}^{T-1} A_i(kT+\sigma) - \sum_{\sigma=0}^{T-1} \hat{D}_i(kT+\sigma)\right). \end{aligned}$$

Thus

$$\begin{aligned} &\frac{1}{2T}(Q_i^2((k+1)T) - Q_i^2(kT)) \\ &\quad \leq \frac{1}{2T}\left(\sum_{\sigma=0}^{T-1} \hat{D}_i(kT+\sigma)\right)^2 + \frac{1}{2T}\left(\sum_{\sigma=0}^{T-1} A_i(kT+\sigma)\right)^2 \\ &\qquad + Q_i(kT)\left(\frac{1}{T}\sum_{\sigma=0}^{T-1} A_i(kT+\sigma) - \frac{1}{T}\sum_{\sigma=0}^{T-1} \hat{D}_i(kT+\sigma)\right). \end{aligned}$$

Summing the above expression over all queues and taking conditional expectations, we arrive at the following bound on the T-step Lyapunov drift:

$$\Delta_T(kT)/T \leq B + \sum_i Q_i(kT)\lambda_i - \sum_i Q_i(kT)\mathbb{E}\left[\frac{1}{T}\sum_{\sigma=0}^{T-1} \hat{D}_i(kT+\sigma)|\mathbf{Q}(kT)\right], \tag{6.1}$$

where B is a constant that depends on the (finite) second moment of the arrival process. We have also used the i.i.d. nature of arrivals to obtain the second term on the right-hand side of (6.1).

We now pause to make some definitions. Let

$$\hat{D}_T(kT) = \sum_i Q_i(kT)\left(\frac{1}{T}\sum_{\sigma=0}^{T-1} \hat{D}_i(kT+\sigma)\right).$$

Given a weight vector $\mathbf{w}$, let $\mathbf{r}^*(\mathbf{w})$ denote the rate vector on the boundary of the original capacity region Λ that maximizes the $\mathbf{w}$-weighted sum of rates. Define

$$R^*(kT) = \sum_i Q_i(kT) r_i^*(\tilde{\mathbf{Q}}(kT)),$$

where $\tilde{\mathbf{Q}}(kT)$ is the normalized queue-length vector at time kT. Next, for a weight vector $\mathbf{w}$, let $\bar{\mathbf{r}}(\tau, \mathbf{w})$ be as defined in Proposition 3.5. Define

$$\overline{R}(kT) = \sum_i Q_i(kT) \bar{r}_i(\tau, \tilde{\mathbf{Q}}(kT)).$$

Observe that $\mathbf{r}^*(\cdot)$ and $\bar{\mathbf{r}}(\tau, \cdot)$ are deterministic vectors once the weight vector and the truncation threshold τ are fixed. On the other hand, $\hat{D}_i(\cdot)$ is a random variable, which is determined by the channel outcomes and the outcomes of the randomized actions dictated by the state action frequencies.

Next, we invoke a result stating that the mixing of the finite MDP is exponentially fast, so that the empirical average reward obtained over a long frame of length T is very close to the infinite-horizon average reward.

Lemma 6.1. *Regardless of the state at time kT, and for all $\kappa > 0$, there exists $\eta(\kappa) > 0$ such that*[2]

$$\mathbb{P}\left\{\left\|\frac{1}{T}\sum_{\sigma=0}^{T-1} \hat{\boldsymbol{D}}(kT+\sigma) - \bar{\boldsymbol{r}}(\tau, \tilde{\boldsymbol{Q}}(kT))\right\| > \kappa\right\} < ce^{-\eta(\kappa)T}.$$

Proof. The result follows from [Mannor and Tsitsiklis 05]. □

Let us return to the drift expression (6.1) and rewrite it as

$$\begin{aligned}
&\Delta_T(kT)/T \qquad (6.2)\\
&\quad \le B + \sum_i Q_i(kT)\lambda_i - \mathbb{E}\left[\hat{D}_T(kT) \mid \mathbf{Q}(kT)\right]\\
&\quad = B + \sum_i Q_i(kT)\left[\lambda_i - r_i^*(\tilde{\mathbf{Q}}(kT))\right] + \mathbb{E}\left[R^*(kT) - \hat{D}_T(kT) \mid \mathbf{Q}(kT)\right]\\
&\quad \le B + \sum_i Q_i(kT)\left[\lambda_i - r_i^*(\tilde{\mathbf{Q}}(kT))\right] + \mathbb{E}\left[\left|R^*(kT) - \overline{R}(kT)\right| \mid \mathbf{Q}(kT)\right]\\
&\qquad + \mathbb{E}\left[\left|\overline{R}(kT) - \hat{D}_T(kT)\right| \mid \mathbf{Q}(kT)\right].
\end{aligned}$$

The bound in (6.2) is due to the triangle inequality.

[2] Throughout this paper, $\|\cdot\|$ denotes the 2-norm.

We now bound the two expectation terms on the right-hand of (6.2). First, we have

$$\begin{aligned}
&\mathbb{E}\left[\left|R^*(kT)-\overline{R}(kT)\right| \mid \mathbf{Q}(kT)\right] \\
&\quad=\mathbb{E}\left[\left|\left\langle \mathbf{Q}(kT), \mathbf{r}^*(\tilde{\mathbf{Q}}(kT))-\bar{\mathbf{r}}(\tilde{\mathbf{Q}}(kT),\tau)\right\rangle\right| \mid \mathbf{Q}(kT)\right] \\
&\quad\le \mathbb{E}\left[\|\mathbf{Q}(kT)\|\left\|\mathbf{r}^*(\tilde{\mathbf{Q}}(kT))-\bar{\mathbf{r}}(\tau,\tilde{\mathbf{Q}}(kT))\right\| \mid \mathbf{Q}(kT)\right] \qquad (6.3)\\
&\quad\le \kappa\|\mathbf{Q}(kT)\|, \qquad (6.4)
\end{aligned}$$

where (6.3) follows from the Cauchy–Schwarz inequality, and (6.4) is due to Proposition 3.5 for large enough τ. Next, we bound the second expectation term in (6.2):

$$\begin{aligned}
&\mathbb{E}\left[\left|\overline{R}(kT)-\hat{D}_T(kT)\right| \mid \mathbf{Q}(kT)\right] \qquad (6.5)\\
&\quad=\mathbb{E}\left[\left|\overline{R}(kT)-\hat{D}_T(kT)\right| \mid \mathbf{Q}(kT), \left|\overline{R}(kT)-\hat{D}_T(kT)\right| \le \kappa\|\mathbf{Q}(kT)\|\right] \\
&\qquad\times \mathbb{P}\left\{\left|\overline{R}(kT)-\hat{D}_T(kT)\right| \le \kappa\|\mathbf{Q}(kT)\| \mid \mathbf{Q}(kT)\right\} \\
&\qquad+\mathbb{E}\left[\left|\overline{R}(kT)-\hat{D}_T(kT)\right| \mid \mathbf{Q}(kT), \left|\overline{R}(kT)-\hat{D}_T(kT)\right| > \kappa\|\mathbf{Q}(kT)\|\right] \\
&\qquad\times \mathbb{P}\left\{\left|\overline{R}(kT)-\hat{D}_T(kT)\right| > \kappa\|\mathbf{Q}(kT)\| \mid \mathbf{Q}(kT)\right\} \\
&\quad\le \kappa\|\mathbf{Q}(kT)\| \\
&\qquad+\left(\sum_i Q_i(kT)\right)\mathbb{P}\left\{\left|\overline{R}(kT)-\hat{D}_T(kT)\right| > \kappa\|\mathbf{Q}(kT)\| \mid \mathbf{Q}(kT)\right\}.
\end{aligned}$$

In arriving at the bound in (6.5), we have used

$$\begin{aligned}
&\mathbb{E}\left[\left|\overline{R}(kT)-\hat{D}_T(kT)\right| \mid \mathbf{Q}(kT), \left|\overline{R}(kT)-\hat{D}_T(kT)\right| > \kappa\|\mathbf{Q}(kT)\|\right] \\
&\quad\le \mathbb{E}\left[\left|\hat{D}_T(kT)\right| \mid \mathbf{Q}(kT), \left|\overline{R}(kT)-W_T(kT)\right| > \kappa\|\mathbf{Q}(kT)\|\right] \le \sum_i Q_i(kT).
\end{aligned}$$

Let us next bound the probability term in (6.5) using the Cauchy–Schwarz inequality:

$$\begin{aligned}
&\mathbb{P}\left\{\left|\overline{R}(kT)-\hat{D}_T(kT)\right| > \kappa\|\mathbf{Q}(kT)\| \mid \mathbf{Q}(kT)\right\} \\
&\quad\le \mathbb{P}\left\{\left\|\bar{\mathbf{r}}(\tau,\tilde{\mathbf{Q}}(kT))-\frac{1}{T}\sum_{\sigma=0}^{T-1}\hat{\mathbf{D}}(kT+\sigma)\right\| > \kappa \mid \mathbf{Q}(kT)\right\}. \qquad (6.6)
\end{aligned}$$

The right-hand side of (6.6) is bounded, by Lemma 6.1. Returning to bounding (6.5), we have

$$\mathbb{E}\left[\left|\overline{R}(kT) - \hat{D}_T(kT)\right| \mid \mathbf{Q}(kT)\right] \leq \kappa \left\|\mathbf{Q}(kT)\right\| + \left(\sum_i Q_i(kT)\right)\left(ce^{-\eta(\kappa)T}\right). \tag{6.7}$$

We can now use (6.7) together with (6.4) to bound the drift in (6.2) from above:

$$\frac{\Delta_T(kT)}{T} \leq B + \sum_i Q_i(kT)\left[\lambda_i - r_i^*(\mathbf{Q}(kT))\right] + \left(\sum_i Q_i(kT)\right)\left(2\kappa + ce^{-\eta(\kappa)T}\right). \tag{6.8}$$

Let $\delta = 2\kappa + ce^{-\eta(\kappa)T}$. Assume now that the input-rate vector λ lies in the interior of the δ-stripped region $\Lambda - \delta\mathbf{1}$. That is, there exists $\xi > 0$ such that $\lambda + \xi\mathbf{1} = r - \delta\mathbf{1}$, for $r \in \Lambda$. Thus,

$$\frac{\Delta_T(kT)}{T} \leq B + \sum_i Q_i(kT)\left[r_i - r_i^*(\mathbf{Q}(kT))\right] - \left(\sum_i Q_i(kT)\right)\xi.$$

Finally, noting that $\sum_i Q_i(kT)\left[r_i - r_i^*(\mathbf{Q}(kT))\right] \leq 0$, by the definition of $r_i^*(\mathbf{Q}(kT))$, we get

$$\frac{\Delta_T(kT)}{T} \leq B - \left(\sum_i Q_i(kT)\right)\xi. \tag{6.9}$$

According to [Neely et al. 05, Theorem 3], the bound in (6.9) shows that the queuing system is stable under our frame-based policy for arrival rates in the interior of the δ-stripped region $\Lambda - \delta\mathbf{1}$. Since δ can be made arbitrarily small by choosing sufficiently large values for T and τ, our policy can support rates arbitrarily close to the capacity region boundary, with a corresponding tradeoff in delay and computational complexity.

Acknowledgments. This work was partly supported by NSF grant CNS-0915988 and by ARO Muri grant W911NF-08-1-0238. Shie Mannor was partially supported by the ISF under contract 890015. Ishai Menache was supported by a Marie Curie International Fellowship within the Seventh European Community Framework Programme.

References

[Ahmad et al. 09] S. H. A. Ahmad, M. Liu, T. Javidi, Q. Zhao, and B. Krishnamachari. "Optimality of Myopic Sensing in Multichannel Opportunistic Access." *IEEE Transactions on Information Theory* 55:9 (2009), 4040–4050.

[Altman 99] E. Altman. *Constrained Markov Decision Processes*. Chapman and Hall, 1999.

[Bertsimas and Tsitsiklis 97] D. Bertsimas and J. N. Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, 1997.

[Gopalan et al. 12] A. Gopalan, C. Caramanis, and S. Shakkottai. "On Wireless Scheduling with Partial Channel-State Information." *IEEE Transactions on Information Theory* 58:1 (2012), 403–420.

[Li and Neely 11] C.-p. Li and M. J. Neely. "Exploiting Channel Memory for Multiuser Wireless Scheduling without Channel Measurement: Capacity Regions and Algorithms." *Performance Evaluation*, DOI: 10.1016/j.peva.2011.01.007, 2011.

[Liu and Zhao 10] K. Liu and Q. Zhao. "Indexability of Restless Bandit Problems and Optimality of Whittle Index for Dynamic Multichannel Access." *IEEE Transactions on Information Theory* 56:11 (2010), 5547–5567.

[Mannor and Tsitsiklis 05] S. Mannor and J. N. Tsitsiklis. "On the Empirical State-Action Frequencies in Markov Decision Processes under General Policies." *Mathematics of Operations Research.* 30:3 (2005), 545.

[Neely et al. 05] M. J. Neely, E. Modiano, and C. E. Rohrs. "Dynamic Power Allocation and Routing for Time-Varying Wireless Networks." *IEEE Journal on Selected Areas in Communications* 23:1 (2005), 89–103.

[Ouyang et al. 10] W. Ouyang, S. Murugesan, A. Eryilmaz, and N. B. Shroff. "Exploiting Channel Memory for Joint Estimation and Scheduling in Downlink Networks." arXiv:1009.3959, 2010.

[Pantelidou et al. 09] A. Pantelidou, A. Ephremides, and A. L. Tits. "A Cross-Layer Approach for Stable Throughput Maximization under Channel State Uncertainty." *ACM/Kluwer Journal of Wireless Networks* 15:5 (2009), 555–569.

[Puterman 94] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, 1994.

[Rockafellar 70] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[Tassiulas and Ephremides 93] L. Tassiulas and A. Ephremides. "Dynamic Server Allocation to Parallel Queues with Randomly Varying Connectivity." *IEEE Transactions on Information Theory* 39:2 (1993), 466–478.

[Whitt 78] W. Whitt. "Approximation of Dynamic Programs I." *Mathematics of Operations Research* 3 (1978), 231–243.

[Whitt 79] W. Whitt. "Approximation of Dynamic Programs II." *Mathematics of Operations Research* 4 (1979), 179–185.

[Whittle 88] P. Whittle. "Restless Bandits: Activity Allocation in a Changing World." *Journal of Applied Probability* 25 (1988), 287–298.

[Ying and Shakkottai 08] L. Ying and S. Shakkottai. "On Throughput-Optimal Scheduling with Delayed Channel State Feedback." In *Information Theory and Applications Workshop*, pp. 339–344. IEEE, 2008.

[Ying and Shakkottai 09] L. Ying and S. Shakkottai. "Scheduling in Mobile Ad Hoc Networks with Topology and Channel-State Uncertainty." *IEEE INFOCOM, Rio de Janeiro, Brazil*, pp. 2347–2355. IEEE, 2009.

Krishna Jagannathan, Department of Electrical Engineering, IIT Madras, Chennai 600036, India (krishnaj@ee.iitm.ac.in)

Shie Mannor, The Technion, Faculty of Electrical Engineering, Fishbach Building, Room 456, Haifa 32000, Israel (shie@ee.technion.ac.il)

Ishai Menache, eXtreme Computing Group, Microsoft, Building 115, 14855 NE 36th Street, Redmond, WA 98052-5388, USA (ishai@microsoft.com)

Eytan Modiano, Massachusetts Institute of Technology, Room 33-412A, Cambridge, MA 02139, USA (modiano@mit.edu)