A state action frequency approach to throughput maximization over uncertain wireless channels

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1109/INFCOM.2011.5935211">http://dx.doi.org/10.1109/INFCOM.2011.5935211</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>Version</td>
<td>Author's final manuscript</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/67313">http://hdl.handle.net/1721.1/67313</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike 3.0</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/3.0/">http://creativecommons.org/licenses/by-nc-sa/3.0/</a></td>
</tr>
</tbody>
</table>
A State Action Frequency Approach to Throughput Maximization over Uncertain Wireless Channels

Krishna Jagannathan, Shie Mannor, Ishai Menache, Eytan Modiano

Abstract—We consider scheduling over a wireless system, where the channel state information is not available a priori to the scheduler, but can be inferred from the past. 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 ON/OFF Markov chain. The scheduler, which is aware of the channel state, is oblivious to the current channel states, but can estimate them by using the acknowledgment history.

We first characterize the capacity region of the system using tools from Markov Decision Processes (MDP) theory. Specifically, we prove that the capacity region boundary is the uniform limit of a sequence of Linear Programming (LP) solutions. Next, we consider 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.

I. INTRODUCTION

In this paper, we consider the scheduling problem in a wireless uplink or downlink system, when there is no explicit instantaneous Channel State Information (CSI) available to the scheduler. The lack of CSI may arise in practice due to 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. We assume that only a single queue can be chosen at each time slot by the server for transmitting its data. The 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 ON.

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 which acknowledges whether the transmission was successful or not, 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.

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 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 of Tassiulas and Ephremides [6] considered the case where both channel states and queue lengths are fully available to the scheduler. It was shown in [6] that the max-weight algorithm, which serves the longest connected queue, is throughput optimal.

Following this paper, several variants of imperfect and delayed CSI scenarios have been considered in the literature [2], [5], [7], [8]. However, our scheduling problem fundamentally differs from the models considered in these 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 where the CSI is obtained through an acknowledgment process, as in our model. In [1], 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 [4], the authors propose a randomized round-robin scheduling policy for the system, which is inspired by the myopic sensing results in [1]. That policy is shown to stabilize arrivals that lie within an inner-bound to the rate region. However, the 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 frame-based policy we propose can stabilize arrival rates that lie arbitrarily close to the capacity region boundary, with a corresponding tradeoff in the computational complexity. Our proof of throughput optimality combines tools from Markov decision theory within a Lyapunov stability framework. We also provide a characterization of the capacity region boundary, as the uniform limit of a sequence of LP solutions.

This paper is organized as follows. The model is presented in Section II. In Section III, we formulate a linear program
which leads to the characterization of the capacity region.
In Section IV, we suggest the frame-based policy, which we prove
to be throughput optimal. Proofs are omitted throughout
due to space constraints, and can be found in [3].

II. SYSTEM DESCRIPTION

The network model. We model the wireless system as
consisting of $N$ parallel queues (see Fig. 1). Time is slotted
($t = 1, 2, \ldots$). Packets arrive to each queue $i \in \{1, 2, \ldots, N\}$
according to an independent stochastic process with rate $\lambda_i$. 
We assume that the arrival processes are independent of each
other, and independent and identically distributed (i.i.d.) from
slot-to-slot.

Due to the shared wireless medium, only a single trans-
mission is allowed at a given time. In our queuing model,
this is equivalent to having the queues connected to a single
server, which 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 packet has to be retransmitted. If it is ON and chosen
by the scheduler, a single packet is properly transmitted, and
an ACK is received by the scheduler.

We denote the channel state of the $i$-th link at time $t$ by
$C_i(t) \in \{ON, OFF\}$, $i = 1, \ldots, N$. 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 Fig. 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.
We further assume that $\epsilon < 0.5$, so that each channel is
positively correlated in time.

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 or not, 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.

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 $\lambda = (\lambda_1, \ldots, \lambda_N)$ is said to be supportable,
if there exists some scheduling policy under which the queue lengths
are finite (almost surely). The capacity region $\Gamma$ 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 $\Gamma$.

III. OPTIMAL POLICIES FOR A FULLY BACKLOGGED

SYSTEM

In the interest of simplicity of notation and exposition, we
restrict attention to the case of $N = 2$ queues in the rest of
the paper, although our methodology extends naturally to more
queues. In this section, we assume that the queues are fully
backlogged, i.e., the queues never empty. As we shall see,
our analysis of the fully backlogged system gives us insights
about the optimal scheduling policy for the dynamic system
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 $s(t) = [k_1(t), b_1(t), k_2(t), b_2(t)]$, where $k_i(t)$
is the number of slots since the queue $i$ was served, and $b_i(t) \in
\{0, 1\}$ is the state of the channel the last time it was observed.

Since the channels are Markovian, $s(t)$ is a sufficient statistic
for the fully backlogged system. Note that $\min(k_1(t), k_2(t)) =
1, \forall t$, and $\max(k_1(t), k_2(t)) \geq 2 \forall t$. Let $S$ denote the
(countably infinite) set of all possible states $s(t)$.

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

$$p_{11}^{(l)} = p_{10}^{(l)} = \frac{1 - (1 - 2\epsilon)^l}{2},\quad p_{11}^{(l)} = p_{00}^{(l)} = \frac{1 + (1 - 2\epsilon)^l}{2}.$$

Next, define the belief vector corresponding to state $s \in S$ as
$[\omega_1(s), \omega_2(s)]$, where $\omega_i(s)$, $i = 1, 2$ is the con-
tional probability that the channel $i$ is ON. For example,
if $s = [1, ON, 3, OFF]$, the corresponding belief vector is

1Throughout, 0 is used interchangeably to denote the channel state OFF,
and 1 is used to denote ON.
whereas a randomized stationary channels, every time a packet is successfully transmitted on that channel, i.e., when the server is assigned to a particular state according to a fixed distribution $P$. We have abused notation to write $\omega_i(t) = \omega_i(s(t))$.

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

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 $s(t)$ 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 average reward 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 $(\lambda_1, \lambda_2)$ is achievable in the fully backlogged system, if there exists some policy for which the infinite horizon time average reward vector equals $(\lambda_1, \lambda_2)$. The closure of the set of all achievable rate pairs is called the rate region $\Lambda$ 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 random arrivals. Thus, the capacity region $\Gamma$ of the queueing system is contained in the rate region $\Lambda$ of the fully backlogged system. In fact, we show in Section IV 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 $\Lambda$. We now proceed to obtain an implicit characterization of the rate region boundary.

### A. MDP formulation and state action frequencies

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

It is easy to show that the rate region $\Lambda$ is convex. Indeed, given two points in the rate region, each attainable by some policy, we can obtain any 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, any 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 $\Lambda$, then

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

for some weight vector $w = [w_1, w_2]$, with $w_1 + w_2 = 1$. The following proposition shows that if the rate pair $(\lambda_1, \lambda_2)$ is in $\Lambda$, then there must necessarily exist state action frequencies that satisfy a set of balance equations.

**Proposition 1:** Let $(\lambda_1, \lambda_2) \in \Lambda$. Then, for each state $s \in S$ and action $a \in \{1, 2\}$, there exists state action frequencies $x(s;a)$, that satisfy

$$ 0 \leq x(s;a) \leq 1, $$

the balance equations (3)-(6) (next page), the normalization condition

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

and the rate constraints

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

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

Let us now provide an intuitive explanation of the balance equations. Equations (3)-(6) simply equate the steady-state probability of being in a particular state, to the total probability of entering that state from all possible states. For example, the left side of (3) equals the steady-state probability of entering state $s_1$ at time $t$, while the right side equals the total probability of getting to the above state from other states, and similarly for the other balance equations. Equation (7) equates the total steady-state probability to unity. Finally, in Equation (8), the term $x(s;i) \omega_i(s)$ equals the probability that the state is $s$, the action $i$ is chosen, and the transmission succeeds. Thus, the right-side of (8) equals the total expected rate on channel $i$.

We now return to the characterization of the rate region boundary. In light of Proposition 1, Equation (1) can be rewritten as follows.

**Problem INFINITE($w$):**

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

subject to (2)-(8).

Since the state-space of this MDP is countably infinite, the optimization in (9) involves an infinite number of variables. In order to make this problem tractable, we now introduce an LP approximation.

### B. 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 an LP.
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 0.5 if it was OFF the last time it was scheduled. Similarly, the belief decreases monotonically to 0.5 if the channel was ON the last time it was scheduled. The key idea now is to construct a finite MDP whose states are the same as 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, 0.5. 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 0.5. The action space and the reward structure are exactly as before. We show that this truncated finite MDP closely approximates the original MDP when τ gets large.

Let us now specify the states and state action frequencies for this finite MDP. There are 4(τ − 2) states of the form \([1, b_1, k_2, b_2]\), \(2 \leq k_2 \leq τ - 1\), \(b_1, b_2 \in \{\text{ON, OFF}\}\) that correspond to the first channel being scheduled in the previous slot, and the second channel being scheduled less than \(τ\) time slots ago. In a symmetric fashion, there are 4(τ − 2) states of the form \([k_1, b_1, 1, b_2]\), \(2 \leq k_1 \leq τ - 1\), \(b_1, b_2 \in \{\text{ON, OFF}\}\) that correspond to the second channel being scheduled in the previous slot. Finally, there are 4 states \([1, b_1, \phi, \phi]\), \(b_1 \in \{\text{ON, OFF}\}\) and \([\phi, \phi, 1, b_2]\), \(b_2 \in \{\text{ON, OFF}\}\) in which one of the channels has not been seen for at least \(τ\) slots, and its belief reset to 0.5. Let us denote by \(\mathcal{S}\) the above set of states for the finite MDP, and let \(\hat{x}(s; a), s \in \mathcal{S}, a \in \{1, 2\}\) denote the state action frequencies for the finite MDP. These state action frequencies satisfy normalization, balance equations, and rate constraints, analogous to (2)-(8).

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

Problem \(\text{FINITE}(\tau, \mathbf{w})\):

\[
\begin{align*}
\hat{r}_1, \hat{r}_2 &= \arg \max_{\lambda_1, \lambda_2} w_1 \lambda_1 + w_2 \lambda_2 \\
\text{subject to normalization, balance equations and rate constraints, analogous to (2)-(8).}
\end{align*}
\]

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

Proposition 2: For a given \(\mathbf{w}\) with \(w_1 + w_2 = 1\), and \(τ\), let \(\hat{\eta}(\tau, \mathbf{w})\) denote the solution to the problem \(\text{FINITE}(\tau, \mathbf{w})\), and let \(\mathbf{r}^*(\mathbf{w})\) denote the solution to \(\text{INFINITE}(\mathbf{w})\). Then, \(\hat{\eta}(\tau, \mathbf{w})\) converges uniformly to \(\mathbf{r}^*(\mathbf{w})\), as \(τ \to \infty\).

We next show a result that asserts that using the state action frequencies obtained from the finite MDP in the backlogged system entails only a negligible sub-optimality, 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 served more than \(τ\) time slots ago, the finite MDP would 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 \(k_2 > τ\), we schedule according to the state action frequencies of the state \([1, b_1, \phi, \phi]\) in the finite MDP, and on so.

Proposition 3: Suppose that the optimal state action frequencies obtained by solving the problem \(\text{FINITE}(\tau, \mathbf{w})\) are used to perform scheduling in a fully backlogged system, as detailed above. Let \(\mathbf{r}(\tau, \mathbf{w})\) denote the average reward vector so obtained. Then for every \(\mathbf{w}\) with \(w_1 + w_2 = 1\), we have that \(\mathbf{r}(\tau, \mathbf{w})\) converges uniformly to the optimal reward \(\mathbf{r}^*(\mathbf{w})\), as \(τ \to \infty\).

We pause briefly to emphasize the subtle difference between Propositions 2 and 3. Proposition 2 asserts that 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, 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, Propositions 2 yields a characterization of the rate region, while Proposition 3 plays a key role in the throughput optimality proof of the frame-based policy in Section IV.

C. An Outer Bound

We now derive an outer bound to the rate region \(\Lambda\), 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
finite queues and stochastic arrivals. We propose a throughput optimal queue length based policy that operates over long frames consisting of \( T \) frames.

In this section, we use the finite LP approximation obtained in Section III-B to numerically compute and plot the capacity region for a two queue system. Specifically, we use the solution to the problem \( \text{FINITE}(\tau, w) \) with large enough \( \tau \), which, according to Proposition 2, uniformly approximates the rate region boundary for all \( w \). We also plot the genie-aided outer bound obtained above, and compare the rate region and our outer bound to the inner and outer bounds derived in [4].

Fig. 3 shows the numerically obtained rate region, the genie-aided outer bound, and the inner and outer bounds derived in [4] for our symmetric two queue system with \( \epsilon = 0.2 \). The capacity region, shown with a dark solid line, was obtained by solving the LP approximation \( \text{FINITE}(\tau, w) \) for all weight vectors, and large enough \( \tau \). The dash-dot curve in the figure is our genie-aided outer bound. The achievable region of the randomized round-robin policy proposed in [4], is shown by a dashed line. Finally, the outer most region in the figure is the outer bound derived in [4].

Interestingly, we observe that the genie-aided outer bound is tight at the symmetric rate point; see [3] for details.

D. Numerical Example

In this section, we use the infinite LP approximation obtained in Section III-B to numerically compute and plot the capacity region for a two queue system. Specifically, we use the solution to the problem \( \text{FINITE}(\tau, w) \) with large enough \( \tau \), which, according to Proposition 2, uniformly approximates the rate region boundary for all \( w \). We also plot the genie-aided outer bound obtained above, and compare the rate region and our outer bound to the inner and outer bounds derived in [4].

The main result of this paper is the throughput optimality of the frame-based policy, for large enough values of \( T \) and \( \tau \). Specifically, our frame-based policy can stabilize all arrival rates within a \( \delta \)-stripped region of \( \Lambda \), for any \( \delta > 0 \). As we shall see, a small \( \delta \) could require large values of \( T \) and \( \tau \), which increases the dimensionality of the LP (depends on \( \tau \)) and the average 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 frame.

Theorem 1: Given any \( \delta > 0 \), there exist large enough \( \tau \) and \( T \) such that the frame-based policy stabilizes all arrival rates in the \( \delta \)-stripped rate region \( \Lambda - \delta \mathbf{1} \).

REFERENCES