

# Low-Complexity Optimal Scheduling over Time-Correlated Fading Channels with ARQ Feedback

Wenzhuo Ouyang, Atilla Eryilmaz, and Ness B. Shroff

**Abstract**—We investigate the downlink scheduling problem under Markovian ON/OFF fading channels, where the instantaneous channel state information is not directly accessible, but is revealed via ARQ-type feedback. The scheduler can exploit the temporal correlation/channel memory inherent in the Markovian channels to improve network performance. However, designing low-complexity and throughput-optimal algorithms under temporal correlation is a challenging problem. In this paper, we find that under an average number of transmissions constraint, a low-complexity index policy is throughput-optimal. The policy uses Whittle’s index value, which was previously used to capture opportunistic scheduling under temporally correlated channels. Our results build on the interesting finding that, under the intricate queue length and channel memory evolutions, the importance of scheduling a user is captured by a simple multiplication of its queue length and Whittle’s index value. The proposed queue-based index policy has provably low complexity. Numerical results show that significant throughput gains can be realized by exploiting the channel memory using the proposed low-complexity policy.

## I. INTRODUCTION

In wireless networks with randomly fluctuating channels, intelligently scheduling users is critical for achieving high network efficiency. Under the assumption that the scheduler possesses accurate instantaneous Channel State Information (CSI), many sophisticated scheduling algorithms have been proposed and extensively studied (e.g., [2]-[5]).

In practice, accurate instantaneous CSI is difficult to obtain at the scheduler. Hence, in this work we consider the important scenario where the instantaneous CSI is not directly accessible to the scheduler, but is instead revealed through ARQ-type feedback only *after* each scheduled data transmission. Many works have focused on scheduling algorithms design with imperfect CSI, where the channel state is considered independent and identically distributed (i.i.d.) processes across time (e.g., [10]-[13]). On the other hand, although the i.i.d. channel model facilitates more tractable analysis, it does not capture the time-correlation of the fading channels. ARQ-based protocols over time-correlated channels are studied in [6]-[9] under the scenarios where user scheduling is not required.

The time-correlation or channel memory inherent in the fading channels can be exploited by the scheduler for more informed decisions, and hence to obtain large throughput/utility

gains (e.g., [14]-[26]). Under imperfect CSI, channel memory, and limited network resources, designing efficient scheduling schemes is highly challenging. This is because the scheduler needs to optimally balance the intricate ‘exploitation-exploration tradeoff’, i.e., to decide whether to exploit the channels with more up-to-date CSI, or to explore the channels with outdated CSI.

In this work, we study downlink scheduling with imperfect CSI and time correlated channels where, differing from works [14]-[18] in this domain, the packets destined to each user randomly arrive in time, and are stored in a corresponding observable data queue before transmission. As a result, the queue lengths randomly evolve with time. Our goal is to design scheduling algorithm that is throughput optimal, i.e., no scheduling policy can ensure system stability for arrival rates that are not supportable by the proposed scheduler. Considering queue lengths along with imperfect CSI and time correlation is highly challenging because to develop throughput-optimal scheduler requires a complex characterization of the interplay between user scheduling, channel memory evolution and queue evolution. Traditional techniques, which assume known service rate (e.g.[19][20]), or assume i.i.d. channel state process and are based on minimizing instantaneous Lyapunov drift in each slot (e.g., scheduling user with maximal instantaneous product of queue length and transmission rate [2][3]), does not apply in this context.

Under this model, because of the aforementioned complications, traditional Dynamic Programming based approaches can be used for designing scheduling schemes, but are intractable due to the well-known ‘curse of dimensionality’. In [21][22], simple round-robin based scheduling policies are shown to possess the throughput-optimality property. The optimality of greedy scheduling algorithm are proven in [23][24]. However, these schemes [21]-[24] are only optimal in the regime where users have *identical* ON/OFF Markovian channel statistics. In [25][26], throughput-optimal frame-based policies are proposed. These policies rely on solving a Linear Programming in each frame, which is hindered by the curse of dimensionality where the computational complexity grows exponentially with the network size.

In this work, we study throughput-optimal downlink scheduling under imperfect CSI over heterogeneous Markovian fading channels. We consider time-correlation by modeling the fading channel as an ‘ON/OFF’ Markov chain. Differing from the previous works [21]-[26] that consider scheduling problems under strict interference constraints (e.g., only one user can be scheduled at each time slot), we assume that each user occupies a dedicated channel, i.e., all users can

Wenzhuo Ouyang is with the Department of ECE, Rice University (e-mail: wenzhuo.ouyang@rice.edu). Atilla Eryilmaz is with the Department of ECE, The Ohio State University (e-mail: eryilmaz@ece.osu.edu). Ness B. Shroff holds a joint appointment in both the Department of ECE and the Department of CSE at The Ohio State University (e-mail: shroff@ece.osu.edu).

A preliminary version of this paper appeared in WiOpt 2012 [1].

This work is partly supported by NSF grants CNS-0721434, CNS-0831919, CNS-0953515, CCF-0916664, DTRA grant HDTRA 1-08-1-0016, Army Research Office MURI Awards W911NF-08-1-0238 and W911NF-07-1-0376.

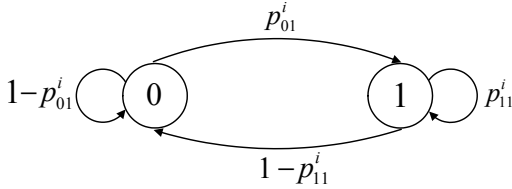


Fig. 1: Two state Markov Chain model.

transmit simultaneously, but the *long-term average* number of transmissions is limited. In this setup, we show that a low complexity scheduling policy is throughput optimal. Such a constraint on long-term average number of transmissions can be used to limit the long-term energy consumption. An example to limit the energy consumption is the *green cellular networks* (e.g., [27]-[29]). It is estimated that the cellular base stations consume 4.5 GW of power globally, which corresponds to more than 40 million metric tons of CO<sub>2</sub> emission and over \$10 billion electricity bill annually [27][28]. With energy expenditure rising by 15-20% each year, an important objective in green cellular networks design is to reduce the long-run average number of data transmissions to decrease energy consumption [28]. Therefore, it is of great interest to understand the relationship between the achievable throughput region and the constraint on the long-term average number of transmissions. The results proposed in this work can be applied to green cellular networks for throughput-optimal scheduling under imperfect CSI and the long-term average energy constraint.

Our contributions are as follows:

- Under the constraint on the long-term average number of transmissions, we propose a low-complexity *throughput-optimal* policy. The policy operates over separate time frames and, in each time frame, tries to maximize a queue-weighted average sum-throughput. We are able to conduct a *frame-based Lyapunov analysis* to this policy and prove its optimality by showing that it minimizes the average Lyapunov drift over each frame. Compared to the traditional approaches for i.i.d. channels based on minimizing *instantaneous* Lyapunov drift each slot, the frame-based approach is useful for analysis in scenarios with time-correlated channels. The per-frame computational complexity is at most  $O((2\tau + 1)N \log(2\tau + 1)N)$  with the number of users  $N$ , where  $\tau$  is a control parameter independent of  $N$ . Therefore, the policy does not suffer from the curse of dimensionality.
- The proposed policy builds on Whittle's index analysis of Restless Multi-armed Bandit Problem (RMBP) [31], where Whittle's index value is used to measure the importance of scheduling a user under the time-correlated channel [16]. Whittle's index policies are known to have optimality properties in various RMBP processes and have been shown to have low-complexity (e.g., [15][19][20]). We find that, interestingly, under the coupled queue length and channel memory evolution, the importance of scheduling a user is measured by a simple *multiplication* of the queue length and Whittle's index value that is given in closed-form. This property is essential for the low-complexity nature of our policy.

## II. SYSTEM MODEL

### A. Downlink Scheduling Problem

We consider a time-slotted wireless downlink network with one base station and  $N$  users, where each user  $i$  occupies a dedicated wireless channel. The channel state of user  $i$ , denoted by  $C_i[t]$  at slot  $t$ , evolves according to an ON/OFF Markov chain across time slots within the state space  $\mathcal{S} = \{0, 1\}$ , independently across channels. When the channel is in state '1', one packet can be successfully transmitted, otherwise no packet can be delivered. As shown in Fig. 1, the channel state evolution is represented by the transition probabilities

$$p_{11}^i := \Pr(C_i[t]=1|C_i[t-1]=1),$$

$$p_{01}^i := \Pr(C_i[t]=1|C_i[t-1]=0).$$

We assume that the Markovian channels are positively correlated, i.e.,  $p_{11}^i > p_{01}^i$  for  $i=1, 2, \dots, N$ . This assumption is commonly made in this field (e.g., [16][21][25][32]), which means that auto-correlation of the channel state process is non-negative [17]. This means, roughly speaking, that the Markov channel is more likely to stay in its state than changing to another state, which captures the typical slow fading or fast transmission scenarios. For ease of presentation, we ignore the trivial case when  $p_{11}^i = 1$  or  $p_{01}^i = 0$ ,  $i \in \{1, \dots, N\}$ .

At the beginning of each time slot, the scheduler chooses users for data transmission. The scheduling decisions are made without the exact knowledge of the channel state in the current slot. Instead, the accurate ON/OFF channel state of a scheduled user is revealed via ACK/NACK feedback from the receiver, only at the end of each slot following data transmission.

We consider the class  $\Phi$  of (possibly non-stationary) scheduling policies that make scheduling decisions based on the history of observed channel states, arrival processes, and scheduling decisions. Under the aforementioned restrictions on average energy consumption, the scheduling schemes are subject to the constraint that the long-term average number of scheduled transmissions is under  $M$ ,

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{i=1}^N a_i^\phi[t] \right] \leq M, \quad (1)$$

where  $a_i^\phi[t] \in \{0, 1\}$  indicates whether user  $i$  is scheduled at slot  $t$  under policy  $\phi \in \Phi$ , and  $M \leq N$ .

Data packets destined for different users are stored in separate queues before transmission. The queue length for user  $i$  is denoted by  $q_i[t]$  at slot  $t$ . We assume that the packet arrivals for the  $i$ -th user form an *i.i.d.* process  $A_i[t]$  with mean  $\lambda_i$  and a bounded second moment. Hence, the  $i$ -th data queue evolves as  $q_i[t+1] = \max\{0, q_i[t] - a_i[t] \cdot C_i[t]\} + A_i[t]$ .

### B. Belief Value Evolution

The scheduler maintains a belief value  $\pi_i[t]$  for each channel  $i$ , defined as the probability of channel  $i$  being in state 1 at the beginning of  $t$ -th slot conditioned on the past channel state observations. The belief values are hence updated according to

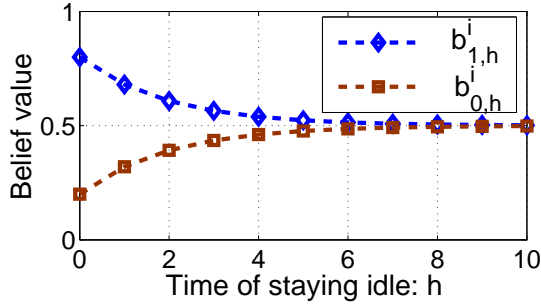


Fig. 2: Belief value evolution,  $p_{11}^i = 0.8$ ,  $p_{01}^i = 0.2$ ,  $b_s^i = 0.5$ . the scheduling decisions and accurate channel state feedbacks,

$$\pi_i[t+1] = \begin{cases} p_{11}^i & \text{if } a_i[t] = 1 \text{ and } C_i[t] = 1, \\ p_{01}^i & \text{if } a_i[t] = 1 \text{ and } C_i[t] = 0, \\ Q_i(\pi_i[t]) & \text{if } a_i[t] = 0, \end{cases} \quad (2)$$

where  $Q_i(x) = xp_{11}^i + (1-x)p_{01}^i$  is the belief evolution operator when user  $i$  is not scheduled in the current slot. In our setup, the belief values are known to be sufficient statistics to represent the past scheduling decisions and channel state feedback [33]. In the meanwhile, the belief value  $\pi_i[t]$  is the expected throughput for user  $i$  if it is scheduled in slot  $t$ .

For the  $i$ -th user, we use  $b_{c,h}^i$  to denote the state of its belief value when the most recent channel state was observed  $h$  time slots ago and was in state  $c \in \{0, 1\}$ . The closed form expression of  $b_{c,h}^i$  can be calculated from (2) and is given as

$$b_{0,h}^i = \frac{p_{01}^i - (p_{11}^i - p_{01}^i)^h p_{01}^i}{1 + p_{01}^i - p_{11}^i}, b_{1,h}^i = \frac{p_{01}^i + (1 - p_{11}^i)(p_{11}^i - p_{01}^i)^h}{1 + p_{01}^i - p_{11}^i}.$$

As depicted in Fig. 2, if the scheduler is never informed of the  $i$ -th user's channel state, the belief value monotonically converges to the stationary probability  $b_s^i := p_{01}^i / (1 + p_{01}^i - p_{11}^i)$  of the channel being in state 1. We assume that the belief values of all channels are initially set to their stationary values. It is then clear that, based on (2), each belief value  $\pi_i[t]$  evolves over a countable state space, denoted by  $\mathcal{B}_i = \{b_{c,h}^i : c \in \{0, 1\}, h \in \mathbb{Z}^+\}$ .

### C. Network Stability Region and Achievable Rate Region

We adopt the following definition of queue stability [3]: queue  $i$  is stable if there exists a limiting stationary distribution  $F_i$  such that  $\lim_{t \rightarrow \infty} P(q_i[t] \leq q) = F_i(q)$ . The *network stability region*  $\Lambda$  is defined as the closure of the set of arrival rate vectors supported by all policies in class  $\Phi$  that does not lead to system instability while abiding by the constraint (1). A policy is called *throughput optimal* if, for any arrival rate vector  $\lambda$  within arbitrary  $\epsilon$  interior of  $\Lambda$ , i.e.,  $\lambda + \epsilon \mathbf{1} \in \Lambda$ , all queues are stable under the policy and constraint (1) is satisfied.

In the meanwhile, we define the *achievable rate region*  $\Gamma$  as the closure of the set of service rate vectors  $\gamma$  that can be achieved by all policies, i.e.,

$$\Gamma = \text{Cl}\{\gamma : \exists \phi \in \Phi \text{ with } \gamma_i = \liminf_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \pi_i[t] \cdot a_i^\phi[t] \right], \quad i = 1, \dots, N, \text{ subject to constraint (1)}\}, \quad (3)$$

where  $\text{Cl}\{\cdot\}$  denotes the closure of the set. The rate region is

convex since randomization can be performed among different policies. The achievable rate region  $\Gamma$  contains the set of the expected service rate vectors that can be achieved with all the policies in  $\Phi$ , in the system with *infinitely backlogged queues*.

### III. OPTIMAL POLICY FOR WEIGHTED SUM-THROUGHPUT MAXIMIZATION

In this section, we postpone discussion on queue evolution and consider a simplified problem with infinitely backlogged queues, and derive the corresponding optimal policy for weighted sum-throughput maximization. The policy introduced here, which is based on scaling the Whittle's index values, is useful to characterize the boundary point of the achievable rate region  $\Gamma$ , and is also an important part in the throughput-optimal policy in the next section that stabilizes all arrival rates within the system stability region  $\Lambda$  – the main result of the paper.

#### A. Weighted Sum-throughput Maximization Problem

Consider the following weighted sum-throughput maximization problem  $\Psi(\mathbf{r}, M)$  for a given vector  $\mathbf{r} = (r_i)_{i=1}^N$ , where the expected service rate for each user  $i$  is scaled by a non-negative factor  $r_i$ ,

$$\max_{\phi \in \Phi} \liminf_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{i=1}^N r_i \cdot \pi_i[t] \cdot a_i^\phi[t] \right] \quad (4)$$

$$\text{s.t. } \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{i=1}^N a_i^\phi[t] \right] \leq M. \quad (5)$$

The above problem  $\Psi(\mathbf{r}, M)$  is hence a constrained Partially Observable Markov Decision Process (CPOMDP) [34][35].

#### B. Whittle's Index for Restless Multi-armed Bandit Problem

The problem (4)-(5) appears difficult because of the complex 'exploitation - exploration' tradeoff. To tackle this problem, we study it in the framework of the Restless Multiarmed Bandit Problem (RMBP) [31] and make use of the associated Whittle's indexability analysis. We next give a brief review of the Whittle's indices for RMBP.

RMBPs refer to a collection of sequential dynamic resource allocation problems where several independently evolving projects compete for service. In each slot, a subset of these competing projects is served. The state of each project stochastically evolves over time, based on the current state of the project and on whether the project is served in the slot. Serving a project brings a reward whose value depends on its state. Hence, in RMBPs, the controller needs to consider the fundamental tradeoff between decisions that bring high instantaneous rewards, versus those decisions that bring better future rewards but sacrifices the instantaneous rewards. Solving RMBPs are known to be PSPACE-hard [30] in general.

Whittle's index analysis [31] for RMBPs considers the following *virtual system*: in each slot, the controller makes one of the two decisions for each project  $P$ : (1) Serve project  $P$  and accrue an immediate reward as a function of its state which is the same as in the original RMBP. (2) Do not serve project  $P$  and obtain an immediate reward  $\omega$  for passivity. The state evolution of the project  $P$  is the same as in the original RMBP, depending on its current state and current action. In



this virtual system, the design goal is to maximize the long-term expected reward by balancing the ‘reward for serving’ and the ‘subsidy for passivity’ in each slot.

Letting  $\mathcal{I}(\omega)$  denote the set of states of project  $P$  in which the optimal action is to stay passive, the Whittle’s indexability condition is defined as follows.

*Project  $P$  is Whittle indexable if the set  $\mathcal{I}(\omega)$  monotonically increases from  $\emptyset$  to the state space  $\mathcal{S}$  of project  $P$ , as  $\omega$  increases from  $-\infty$  to  $\infty$ . The RMBP is Whittle indexable if every project is Whittle indexable.*

If Indexability holds, for each state  $s$  of a project, the Whittle’s index  $W(s)$  is defined as the infimum of  $\omega$  in which it is optimal to stay idle in the  $\omega$ -subsidized system, i.e.,

$$W(s) = \inf\{\omega : s \in \mathcal{I}(\omega)\}.$$

Under an average constraint on the number of projects scheduled per slot, it is known that, upon the satisfaction of the Indexability condition, an optimal algorithm exists based on the ‘Whittle’s indices’: activate the projects with large Whittle’s index value [31].

The RMBP theories and the associated Whittle’s indices can be used in our downlink scheduling problem. Here, each downlink user corresponds to a project in the RMBP, with the associated state being the belief value of its channel. Correspondingly, the project is considered served if the user is scheduled for data transmission at a slot. Hence the Whittle’s index policy is very attractive to provide optimal solutions to our problem, as we shall elaborate in the rest of the paper.

### C. Optimal Policy for Weighted Sum-throughput Maximization

It was shown in that our downlink scheduling problem is Whittle indexable [18], and, under uniform weight vector  $\mathbf{r}=\mathbf{1}$ , an optimal policy for problem  $\Psi(\mathbf{1}, M)$  exists based on Whittle’s indexability analysis of Restless Multi-armed Bandit Problem [16]. Specifically, for channel  $i$ , a closed form Whittle’s index value  $W_i^1(\pi)$  is assigned to each belief state  $\pi \in \mathcal{B}_i$ . These indices intelligently capture the exploitation-exploration value to be gained from scheduling the user at the corresponding belief state [16]. The closed form expression of the Whittle’s index value  $W_i^1(\pi)$ ,  $\pi \in \mathcal{B}_i$ , is given as follows [16][18],

$$W_i^1(\pi) = \begin{cases} \frac{(\pi - Q_i(\pi))(h+1) + Q_i(\pi)}{1 - p_{11}^i + (\pi - Q_i(\pi))h + Q_i(\pi)} & \text{if } p_{01}^i \leq \pi = b_{0,h}^i < b_s^i \\ \frac{p_{01}^i}{(1 - p_{11}^i)(1 + p_{01}^i - p_{11}^i) + p_{11}^i} & \text{if } b_s^i \leq \pi \leq p_{11}^i \end{cases} \quad (6)$$

It was shown that  $W_i^1(\pi)$  monotonically increases with  $\pi$  and satisfies  $W_i^1(\pi) \in [0, 1]$  [16][18]. In the following lemma, we give the optimal algorithm to the problem  $\Psi(\mathbf{r}, M)$  with arbitrary non-negative weight vector  $\mathbf{r}$ . The proof of the lemma follows the line of [31] and is re-proven in Appendix A.

**Lemma 1.** *There exists an optimal stationary policy  $\phi^*(\mathbf{r}, M)$  for problem  $\Psi(\mathbf{r}, M)$  (cf. (4)-(5)), parameterized by a user index  $i^*$ , a threshold  $\omega^*$  and a randomization factor  $\rho^*$ , such that*

(i) *The scheduler maintains an  $\mathbf{r}$ -weighted index value  $W_i^{\mathbf{r}}(\pi_i[t]) = r_i \cdot W_i^1(\pi_i[t])$  for user  $i$ .*

(ii) *User  $i$  is scheduled if  $W_i^{\mathbf{r}}(\pi_i[t]) > \omega^*$ , or if  $W_i^{\mathbf{r}}(\pi_i[t]) = \omega^*$  with  $i > i^*$ . User  $i$  stays idle if  $W_i^{\mathbf{r}}(\pi_i[t]) < \omega^*$ , or if  $W_i^{\mathbf{r}}(\pi_i[t]) = \omega^*$  with  $i < i^*$ . If  $W_i^{\mathbf{r}}(\pi_i[t]) = \omega^*$  with  $i = i^*$ , user  $i$  is scheduled with probability  $\rho^*$ .*

(iii) *The parameters  $i^*$ ,  $\omega^*$  and  $\rho^*$  are such that the long-term average number of transmissions equals  $M$ .*

**Remarks:** Interestingly, by multiplying the Whittle’s index values  $W_i^1(\pi_i[t])$  with  $r_i$ , the optimal policy  $\phi^*(\mathbf{1}, M)$  extends to more general problem  $\Psi(\mathbf{r}, M)$ . This property is important for designing the throughput-optimal policy in Section IV.

### D. Approximate $i^*$ , $\omega^*$ and $\rho^*$ using State Space Truncation

Note that the parameters  $i^*$ ,  $\omega^*$  and  $\rho^*$  need to be carefully chosen to satisfy the complementary slackness condition, i.e., Lemma 1(iii). While directly finding these parameters may be difficult, we next introduce an algorithm to derive approximate values of  $i^*$ ,  $\omega^*$  and  $\rho^*$  based on a fictitious model over *truncated belief state space*. This fictitious model facilitates more tractable design and analysis. More importantly, we shall show that, when implementing these approximate values over the original untruncated system, the performance will get arbitrary close to the optimality.

Recall that the belief value  $\pi_i[t]$  evolves over a countable state space  $\mathcal{B}_i$  for user  $i$  and approaches the stationary value if the channel is not active for a long time. This motivates us to consider the following fictitious belief evolution model over the truncated state space: the belief value of a user is set to its steady state (i.e., its channel state history is entirely forgotten) if the corresponding channel has not been scheduled for a long time, say  $\tau$  slots. We use  $\pi_i^\tau[t]$  to denote this ‘heuristic belief value’. The evolution of  $\pi_i^\tau[t]$  is hence,

$$\pi_i^\tau[t+1] = \begin{cases} p_{11}^i & \text{if } a_i[t] = 1 \text{ and } C_i[t] = 1, \\ p_{01}^i & \text{if } a_i[t] = 1 \text{ and } C_i[t] = 0, \\ Q_i(\pi_i[t]) & \text{if } a_i[t] = 0, \prod_{k=1}^{\tau-1} (1 - a_i[t-k]) = 0, \\ b_s^i & \text{if } \prod_{k=0}^{\tau-1} (1 - a_i[t-k]) = 1. \end{cases} \quad (7)$$

We let  $\mathcal{B}_i^\tau$  denote the truncated state space for the  $i$ -th user, i.e.,  $\mathcal{B}_i^\tau = \{b_s^i, b_{c,l}^i : c \in \{0, 1\}, l = 1, 2, \dots, \tau\}$  and let  $\mathbf{B}^\tau = [\mathcal{B}_1^\tau, \dots, \mathcal{B}_N^\tau]$ . Over the fictitious truncated state space, we consider the following policy  $\phi_{j,\omega,\rho}^{trunc}$ :

**Policy  $\phi_{j,\omega,\rho}^{trunc}$  over the truncated state space:** *User  $i$  is scheduled if  $W_i^{\mathbf{r}}(\pi_i^\tau[t]) > \omega$ , or if  $W_i^{\mathbf{r}}(\pi_i^\tau[t]) = \omega^*$  with  $i > j$ . User  $i$  stays idle if  $W_i^{\mathbf{r}}(\pi_i^\tau[t]) < \omega$ , or if  $W_i^{\mathbf{r}}(\pi_i^\tau[t]) = \omega^*$  with  $i < j$ . If  $W_i^{\mathbf{r}}(\pi_i^\tau[t]) = \omega$  with  $i = j$ , it is scheduled with probability  $\rho$ .*

Under this setup, we let the parameter  $\alpha_i^\tau(j, \omega, \rho)$  denote the long-term expected fraction of time transmitting to user  $i$ , i.e.,

$$\alpha_i^\tau(j, \omega, \rho) = \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} a_i^{\phi_{j,\omega,\rho}^{trunc}}[t] \right], \quad (8)$$

where  $a_i^{\phi_{j,\omega,\rho}^{trunc}}[t] \in \{0, 1\}$  indicates whether user  $i$  is scheduled at time  $t$  under policy  $\phi_{j,\omega,\rho}^{trunc}$ . The closed-form expression of

$\alpha_i^\tau(j, \omega, \rho)$  is given by the following lemma. The proof of the lemma is given in Appendix B.

**Lemma 2.** *Let the value  $\tau_0$  be*

$$\tau_0 = \left\lceil 4 \max \left\{ \frac{1}{-\log(p_{11}^i - p_{01}^i)}, \frac{1}{\log^2(p_{11}^i - p_{01}^i)}, i=1, \dots, N \right\} \right\rceil. \quad (9)$$

*Over the truncated state space and under policy  $\phi_{j, \omega, \rho}^{trunc}$ , if  $\tau > \tau_0$ , the following hold for  $\alpha_i^\tau(j, \omega, \rho)$ ,*

(i) *The closed-form expression of  $\alpha_j^\tau(j, \omega, \rho)$  is given by*

$$\alpha_j^\tau(j, \omega, \rho) = \begin{cases} \frac{\rho(b_{0,h}^j - b_{0,h+1}^j) + 1 - p_{11}^j + b_{0,h+1}^j}{\rho b_{0,h}^j + (1-\rho)b_{0,h+1}^j + (1-p_{11}^j)(h+1-\rho)} & \text{if } \omega = W_j^F(b_{0,h}^j), h < \tau \\ \frac{\rho(b_{0,\tau}^j - b_s^j) + 1 - p_{11}^j + b_s^j}{\rho b_{0,\tau}^j + (1-\rho)b_s^j + (1-p_{11}^j)(\tau+1-\rho)} & \text{if } \omega = W_j^F(b_{0,\tau}^j) \\ \frac{\rho(1-p_{11}^j + b_s^j)}{(1+\tau\rho)(1-p_{11}^j) + \rho b_s^j} & \text{if } \omega = W_j^F(b_s^j) \\ 0 & \text{if } \omega > W_j^F(b_s^j). \end{cases}$$

*The closed-form expression of  $\alpha_i^\tau(j, \omega, \rho)$ ,  $i \neq j$  is given by*

$$\alpha_i^\tau(j, \omega, \rho) = \begin{cases} \frac{1-p_{11}^i + b_{0,h+1}^i}{b_{0,h+1}^i + (1-p_{11}^i)(h+1)} & \text{if } h < \tau, \omega = W_i^F(b_{0,h}^i), i < j \\ \frac{1-p_{11}^i + b_{0,h}^i}{b_{0,h}^i + (1-p_{11}^i)h} & \text{if } h \leq \tau, \omega = W_i^F(b_{0,h}^i), i > j; \\ & \text{or if } h \leq \tau, W_i^F(b_{0,h-1}^i) < \omega < W_i^F(b_{0,h}^i) \\ \frac{1-p_{11}^i + b_s^i}{b_s^i + (1-p_{11}^i)(\tau+1)} & \text{if } \omega = W_i^F(b_{0,\tau}^i), i < j; \\ & \text{or if } \omega = W_i^F(b_s^i), i > j \\ 0 & \text{if } \omega = W_i^F(b_s^i), i < j; \\ & \text{or if } \omega > W_i^F(b_s^i). \end{cases}$$

(ii) *For fixed  $\pi_j \in \{b_{0,1}^j, b_{0,2}^j, \dots, b_{0,\tau}^j, b_s^j\}$ ,  $\alpha_j^\tau(j, W_j^F(\pi_j), \rho)$  strictly increases with  $\rho$ . For fixed  $\rho$ ,  $\alpha_i^\tau(j, W_i^F(\pi_i), \rho)$  strictly decreases with  $\pi_i$  for  $\pi_i \in \{b_{0,1}^i, b_{0,2}^i, \dots, b_{0,\tau}^i, b_s^i\}$  and all  $i$ .*

We approximate the optimal values  $i^*$ ,  $\omega^*$  and  $\rho^*$  (defined in Lemma 1) using the fictitious truncated state space model. The approximate value  $i_\tau$ ,  $\omega_\tau$  and  $\rho_\tau$  are such that, under policy  $\phi_{i_\tau, \omega_\tau, \rho_\tau}^{trunc}$  over the truncated state space, the long-term average number of transmissions equals  $M$ , i.e.,

$$\sum_{i=1}^N \alpha_i^\tau(i_\tau, \omega_\tau, \rho_\tau) = M. \quad (10)$$

Note that, equation (10) is the truncated-state-space correspondence of Lemma 1(iii). We next design an algorithm, denoted by  $G^\tau(\mathbf{r}, M)$ , to calculate  $i_\tau$ ,  $\omega_\tau$  and  $\rho_\tau$ , described to the right and explained next.

- The algorithm first calculates the  $\mathbf{r}$ -weighted index values  $W_i^F(\pi_i)$  by scaling  $W_i^1(\pi_i)$  by  $r_i$ , and stores the value and the corresponding user in vector  $\mathbf{I}$  (line 7-15).

- The algorithm then sorts all the  $\mathbf{r}$ -weighted indices of each belief state of all users to a  $(2\tau+1)N$ -dimensional vector  $\mathbf{w}$  in increasing order (line 16).

- The algorithm then calculates  $\omega_\tau$  and  $\rho_\tau$  based on the monotonicity property in Lemma 2(ii). Hence, fixing the randomization factor  $\rho=1$ , it increases the threshold  $\omega$  by going through the indices in  $\mathbf{w}$  and calculates the long-term average number of transmission when threshold  $\omega$  equals

---

**Algorithm  $G^\tau(\mathbf{r}, M)$ :** Calculation of  $i_\tau$ ,  $\omega_\tau$  and  $\rho_\tau$

---

```

1: TxTime[i] = 1 for all  $i \in \{1, \dots, N\}$ 
2: TotalTime = N
3: struct Index
4: { float value
5:   int user
6: }  $\mathbf{I}[(2\tau+1)N]$ ,  $\mathbf{w}[(2\tau+1)N]$ 
7:  $j = 0$ 
8: for  $i = 1$  to  $N$  do
9:   for each  $\pi_i \in \mathcal{B}_i^\tau$  do
10:     $W_i^F(\pi_i) = r_i \cdot W_i^1(\pi_i)$ 
11:     $\mathbf{I}[j].value = W_i^F(\pi_i)$ 
12:     $\mathbf{I}[j].user = i$ 
13:     $j \leftarrow j + 1$ 
14:   end for
15: end for
16:  $\mathbf{w} = \text{sort}(\mathbf{I})$   $\triangleright$  Sort the elements in  $\mathbf{I}$  in increasing order
    of the index value and outputs to vector  $\mathbf{w}$ .
    For index values that are equal, they are ordered
    in increasing order of the associated
    user index.
17: for  $k = 1$  to  $\text{size}(\mathbf{w})$  do
18:    $\text{NewTime}[\mathbf{w}[k].user] = \alpha_{\mathbf{w}[k].user}^\tau(\mathbf{w}[k].value, 1)$ 
19:    $\text{TimeDiff} = \text{TxTime}[\mathbf{w}[k].user] - \text{NewTime}[\mathbf{w}[k].user]$ 
20:    $\text{TotalTime} = \text{TotalTime} - \text{TimeDiff}$ 
21:   if  $\text{TotalTime} < M$  then
22:      $i_\tau = \mathbf{w}[k-1].user$ 
23:      $\omega_\tau = \mathbf{w}[k-1].value$ 
24:      $\text{TxTime}[\mathbf{w}[k-1].user] = M - \sum_{i \neq \mathbf{w}[k-1].user} \text{TxTime}[i]$ 
25:      $\rho_\tau = \beta_{\mathbf{w}[k-1].user}(\omega_\tau, \text{TxTime}[\mathbf{w}[k-1].user])$ 
26:     Break
27:   end if
28:    $\text{TxTime}[\mathbf{w}[k].user] = \text{NewTime}[\mathbf{w}[k].user]$ 
29: end for
30: return  $\omega_\tau, \rho_\tau$ 

```

---

to that index. For each element of  $\mathbf{w}$ , it first calculates the long-term expected fraction of time  $\text{NewTime}[\mathbf{w}[k].user]$  transmitting to the corresponding user  $\mathbf{w}[k].user$  in line 18, and hence the decreased amount, denoted by  $\text{TimeDiff}$ , as compared with previous value  $\text{TxTime}[\mathbf{w}[k].user]$  in line 19. Note that, in each iteration, only the user corresponding to  $\mathbf{w}[k]$  will have an updated expected fraction of transmission time. The total expected number of transmission, denoted by  $\text{TotalTime}$ , is then updated by decreasing the same amount (line 20). The threshold  $\omega$  keeps increasing until the total expected number of transmission is below  $M$  (line 21). Noting that  $\alpha_i^\tau(\omega, 1)$  decreases with  $\omega$ , we then set  $i_\tau = \mathbf{w}[k-1].user$  and  $\omega_\tau = \mathbf{w}[k-1].value$  (line 21-22). Then we calculate the expected transmission time to the user that corresponds to  $\mathbf{w}[k-1]$  (line 23) and select the randomization factor  $\rho_\tau$  so that the constraint (10) is satisfied (line 24), where the function  $\beta_i : (\omega, \alpha) \rightarrow \rho$  calculates the randomization factor  $\rho$  required to achieve the long-term expected fraction of time

$\alpha$  transmitting to user  $i$  at threshold  $\omega$ , and is derived from lemma 2(i) as,

$$\beta_i(\omega, \alpha) = \begin{cases} \frac{(1-\alpha)(1-p_{11}^i + b_{0,h+1}^i) - \alpha h(1-p_{11}^i)}{(1-\alpha)(b_{0,h+1}^i - b_{0,h}^i) - \alpha(1-p_{11}^i)} & \text{if } \omega = W_i^{\mathbf{r}}(b_{0,h}^i), h < \tau; \\ \frac{(1-\alpha)(1-p_{11}^i + b_s^i) - \alpha\tau(1-p_{11}^i)}{(1-\alpha)(b_s^i - b_{0,\tau}^i) - \alpha(1-p_{11}^i)} & \text{if } \omega = W_i^{\mathbf{r}}(b_{0,\tau}^i); \\ \frac{\alpha(1-p_{11}^i)}{(1-\alpha\tau)(1-p_{11}^i) + (1-\alpha)b_s^i} & \text{if } \omega = W_i^{\mathbf{r}}(b_s^i); \\ 0 & \text{if } \omega > W_i^{\mathbf{r}}(b_s^i). \end{cases}$$

*E. Performance of policy over untruncated state space with approximate parameters  $\omega_\tau, \rho_\tau$*

We next examine, over the *original untruncated model*, the policy that uses the approximated parameters  $i_\tau, \omega_\tau$  and  $\rho_\tau$ . We denote such policy as  $\phi_\tau(\mathbf{r}, M)$  and present it next.

---

**Algorithm  $\phi_\tau(\mathbf{r}, M)$ :  $\mathbf{r}$ -weighted Index Policy**

---

- 1: **Initialization phase:** The parameters  $i_\tau, \omega_\tau$  and  $\rho_\tau$  are calculated by algorithm  $G^\tau(\mathbf{r}, M)$ .
  - 2: **At slot  $t$ :** user  $i$  is scheduled if the  $\mathbf{r}$ -weighted index value  $W_i^{\mathbf{r}}(\pi_i[t]) > \omega_\tau$ , or if  $W_i^{\mathbf{r}}(\pi_i[t]) = \omega_\tau$  with  $i > i_\tau$ . User  $i$  stays passive if  $W_i^{\mathbf{r}}(\pi_i[t]) < \omega_\tau$ , or if  $W_i^{\mathbf{r}}(\pi_i[t]) = \omega_\tau$  with  $i < i_\tau$ . If  $W_i^{\mathbf{r}}(\pi_i[t]) = \omega_\tau$  with  $i = i_\tau$ , user  $i$  is scheduled with probability  $\rho_\tau$ .
- 

**Remark:** The computational complexity of the initialization phase of algorithm  $\phi_\tau(\mathbf{r}, M)$  is dominated by sorting the index values in Algorithm  $G^\tau(\mathbf{r}, M)$  (line 16), which has complexity  $O((2\tau + 1)N \cdot \log((2\tau + 1)N))$ . After initialization, the  $\mathbf{r}$ -weighted Index Policy  $\phi_\tau(\mathbf{r}, M)$  takes a very simple threshold-form with per-slot computational complexity  $O(N)$ .

We let  $V^*(\mathbf{r}, M)$  be the weighted sum-throughput under the optimal policy  $\phi^*(\mathbf{r}, M)$  defined in lemma 1, and let  $V_\tau(\mathbf{r}, M)$  be that under the afore-mentioned policy  $\phi_\tau(\mathbf{r}, M)$ , i.e.,

$$V^*(\mathbf{r}, M) = \liminf_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{i=1}^N r_i \cdot \pi_i[t] \cdot a_i^{\phi^*(\mathbf{r}, M)}[t] \right]. \quad (11)$$

$$V_\tau(\mathbf{r}, M) = \liminf_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{i=1}^N r_i \cdot \pi_i[t] \cdot a_i^{\phi_\tau(\mathbf{r}, M)}[t] \right]. \quad (12)$$

Since we also require the long-term average number of transmissions of the policy  $\phi_\tau(\mathbf{r}, M)$  to satisfy the constraint (1), we denote  $Z_\tau(\mathbf{r}, M)$  as the time-average expected number of transmissions under this policy, i.e.,

$$Z_\tau(\mathbf{r}, M) = \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{i=1}^N a_i^{\phi_\tau(\mathbf{r}, M)}[t] \right].$$

Recall that  $\tau_0$  is defined in Lemma 2. The next lemma shows that the policy  $\phi_\tau(\mathbf{r}, M)$  asymptotically achieves the maximum weighted sum-throughput of (4)(5) as the truncation size increases, while abiding the long-term average number of transmissions constrain (1). The proof is given in Appendix C.

**Lemma 3.** *For  $\tau \geq \tau_0$ , we have*

(i) *The weighted sum-throughput performance difference between the policies  $\phi^*(\mathbf{r}, M)$  and  $\phi_\tau(\mathbf{r}, M)$  is bounded by*

$$|V^*(\mathbf{r}, M) - V_\tau(\mathbf{r}, M)| \leq f(\tau) \sum_{i=1}^N r_i, \quad (13)$$

where  $f(\tau) = \sum_{i=1}^N f_i(\tau)$ , which satisfies  $f(\tau) \rightarrow 0$  as  $\tau \rightarrow \infty$  with

$$f_i(\tau) = \frac{\rho(b_{0,\tau}^i - b_{0,\tau+1}^i) + 1 - p_{11}^i + b_{0,\tau+1}^i}{\rho b_{0,\tau}^i + (1-\rho)b_{0,\tau+1}^i + (1-p_{11}^i)(\tau+1-\rho)}. \quad (14)$$

(ii) *The long-term average number of transmissions under policy  $\phi_\tau(\mathbf{r}, M)$  satisfies the constraint (1), i.e.,  $Z_\tau(\mathbf{r}, M) \leq M$ .*

**Remark:** Note that the truncation size  $\tau$  needs to be sufficiently large (i.e.,  $\tau \geq \tau_0$ ) to prove the Lemma. This is because sufficiently large truncation size can provide enough level of approximation that facilitates analytical characterization. Specifically, in the proof,  $\tau_0$  is used in Lemma 4.

#### IV. QUEUE-BASED INDEX POLICY OVER TIME FRAMES

Note that the Index Policy in the last section, as well as the associated Whittle's index value, is for the system with infinitely backlogged queues and the corresponding weighted sum-throughput maximization problem (4)-(5). In this section, we consider scheduler design under random arrival of data packets and the associated queue evolution in the time-correlated downlink. The objective here is to not only obtain maximum weighted sum-throughput, but also maintain queue stability. In the presence of queue evolution, the problem get much more complicated. Note that, in the weighted sum-throughput maximization problem, the reward of scheduling a user is captured by the Whittle's index value. Under the additional consideration of queue stability, the queue lengths need to be jointly taken into account for scheduling, i.e., a user is scheduled for transmission not only because it has a high index value, but may also because it has a large queue length.

Next, we propose a throughput-optimal scheduling policy based on scaling the Whittle's index by the queue length. The policy is implemented over separate time-frames and has low-complexity.

We divide the time slots  $\{0, 1, 2, \dots\}$  into separate *time frames* of length  $T$ , i.e., the  $k$ -th frame,  $k \in \{0, 1, 2, \dots\}$ , includes time slots  $kT, \dots, (k+1)T-1$ . The scheduling decisions in the  $k$ -th frame are made based on the queue length information  $\mathbf{q}[kT]$  at the beginning of that frame. During the  $k$ -th frame, the policy  $\phi_\tau(\mathbf{q}[kT], M)$ , developed in the last section, is implemented. Formally, the  $T$ -frame queue-based index policy, denoted by  $\text{Q-Index}_\tau(\mathbf{T}, M)$ , is introduced next.



---

**Algorithm Q-Index $_{\tau}(T, M)$ :  $T$ -Frame Queue-based Index Policy**


---

- 1: The time slots are divided into frames of length  $T$ . Slot  $t$  is in the  $k^{\text{th}}$  frame if  $kT \leq t < (k+1)T$ ,  $k \in \{0, 1, \dots\}$ .
  - 2: **At the beginning of the  $k^{\text{th}}$  frame:** At the beginning of slot  $kT$ , implement the algorithm  $G^{\tau}(\mathbf{q}[kT], M)$  that outputs  $\omega_{\tau}$  and  $\rho_{\tau}$ .
  - 3: **In each slot  $t$  of the  $k^{\text{th}}$  frame:**
    - User scheduling:** user  $i$  is scheduled if the  $\mathbf{q}[kT]$ -weighted index value  $W_i^{\mathbf{q}[kT]}(\pi_i[t]) > \omega_{\tau}$ , or if  $W_i^{\mathbf{q}[kT]}(\pi_i[t]) = \omega_{\tau}$  with  $i > i_{\tau}$ . User  $i$  stays passive if  $W_i^{\mathbf{q}[kT]}(\pi_i[t]) < \omega_{\tau}$ , or if  $W_i^{\mathbf{q}[kT]}(\pi_i[t]) = \omega_{\tau}$  with  $i < i_{\tau}$ . If  $W_i^{\mathbf{q}[kT]}(\pi_i[t]) = \omega_{\tau}$  with  $i = i_{\tau}$ , user  $i$  is scheduled with probability  $\rho_{\tau}$ . If a user with empty queue is scheduled, then a dummy packet is transmitted to the user.
    - ARQ feedback:** At the end of each slot, the scheduled users send ARQ feedback to the BS. The belief values are updated according to the feedback at the scheduler.
- 

**Remarks:** We next describe the intuition behind designing the above algorithm.

- (1) Note that, for queue stability, instead of using queue length information in every slot, it is sufficient only to consider the sampled queue length information at the periodic slots, i.e.,  $\mathbf{q}[kT]$ ,  $k = 0, 1, \dots$ . The queue is stable if and only if the periodically sampled queue length evolution process is stable.
- (2) Within each frame, we wish to maximize the weighted sum-throughput, where each user's throughput is weighted by its queue length sample value at the beginning of the time frame. Hence, in step 2-3, we implement the Index policy  $\phi^{\tau}(\mathbf{q}[kT], M)$  developed in the previous section. The rationale is because, first, we would like to schedule the users to achieve the higher throughput promised by the Index policy that exploits the temporal correlated channels. Moreover, for system stability, we would like to choose users with large queue-lengths. Hence, by considering the queue weighted throughput and using the Index policy  $\phi^{\tau}(\mathbf{q}[kT], M)$  in frame  $T$ , an overloaded queue can get served with potentially higher rate. As a direct result, a user  $i$ 's index is scaled by its queue length  $q[kT]$ .
- (3) An intuitive explanation of the multiplication of index and queue length is as follows. We schedule a user not only because of its longer queues, but also when its underlying 'channel quality' is favorable (in terms of both exploitation and exploration values). Consider the example where a user's channel is strongly correlated and is observed '0' state in the previous slot. Hence it is highly likely to stay in '0' state for a while. Hence scheduling it can result in wasted system resource since packets are unlikely to be successfully delivered. Correspondingly, this 'quality' of a channel is reflected in the close-to-zero Whittle's index value. The multiplication of queue length and the Whittle's index value is able to capture both the queue length and the channel's 'quality' for scheduling. Summation of the index and queue length, on the other hand, fails capture both of these properties.
- (4) Dividing the time slots into different frames brings us advantages in the realm of large frame length (i.e.,  $T$ ). Since we implement the Index policy within each finite-horizon

frame, if the frame length is small, we lose from exploiting the channel correlation because the Index policy is optimal only in the infinite horizon. As the frame length scales, the (per-slot) loss of exploiting the channel correlation diminishes.

(5) Note that a dummy packet is transmitted to a scheduled user with empty queue. The dummy packet is known to the users and contains no new information and hence does not bring throughput gains if it is transmitted. However, the scheduler will still receive channel state update from the corresponding scheduled users. This mechanism is useful to establish our results.

The next proposition and corollary establish throughput-optimality of the queue-based index policy over time frames, where, recall that,  $f(\tau)$  is given in Lemma 3. The proof is given in Appendix D.

**Proposition 1.** *If  $\tau \geq \tau_0$ , then there exist  $T_0$  and function  $g(\tau) = 3f(\tau)$  such that the following holds whenever  $T > T_0$ : If the arrival rate  $\lambda$  satisfies  $\lambda + g(\tau)\mathbf{1} \in \Gamma$  and the  $T$ -frame queue-based index policy  $Q\text{-Index}_{\tau}(T, M - g(\tau)/2)$  is implemented, then all queues are stable and constraint (1) on the average number of transmissions is satisfied. The function  $g(\tau)$  satisfies  $\lim_{\tau \rightarrow \infty} g(\tau) = 0$ .*

**Corollary 1.** *The achievable rate region  $\Gamma$ , expressed in (3), is equal to the stability region  $\Lambda$ .*

**Proof:** Recall that the achievable rate region  $\Gamma$  corresponds to the expected service rate vectors that can be achieved in the system with infinitely backlogged queues, by any policy in  $\Phi$ . Now consider all the arrival rates within the interior of the stability region  $\Lambda$ . For each arrival vector  $\lambda \in \Lambda$ , there exists a certain policy in  $\Phi$  that stabilizes it, i.e., provides a service rate not below  $\lambda$ . Therefore, the achievable rate region  $\Gamma$  provides an upper bound on the stability region  $\Lambda$ . Since the previous proposition states that the queue-based index policy stabilizes arrival rates arbitrarily close to the boundary of the achievable rate region  $\Gamma$ , the achievable rate region  $\Gamma$  and the stability region  $\Lambda$  share the same interior. Because both regions  $\Gamma$  and  $\Lambda$  are defined over closure of sets, we have  $\Gamma = \Lambda$ . ■

Proposition 1 and Corollary 1 together establish the throughput optimality of the proposed policy. With sufficiently large  $\tau$  and  $T$ , the proposed policy  $Q\text{-Index}_{\tau}(T, M - g(\tau)/2)$  can support arrival rate  $\lambda$  within arbitrary  $\epsilon$  interior of the stability region, i.e.,  $\lambda + \epsilon\mathbf{1} \in \Lambda$  and satisfy constraint (1).

**Remarks:**

- (1) Note that, in Proposition 1, the parameter  $M$  in the queue-based index policy is scaled down by  $g(\tau)/2$ . This mechanism is needed to guarantee the constraint on the long-term average number of transmission. The details are given in the proof.
- (2) In the queue-based index policy, a user is scheduled based on its  $\mathbf{q}[kT]$ -weighted Whittle's index value. The Whittle's index value is necessary for the results because it measures the importance of a wireless channel for scheduling, considering jointly the instantaneous throughput and future throughput (e.g., see [18][31], Lemma 1). It is interesting to note that a simple multiplication of queue length and Whittle's index

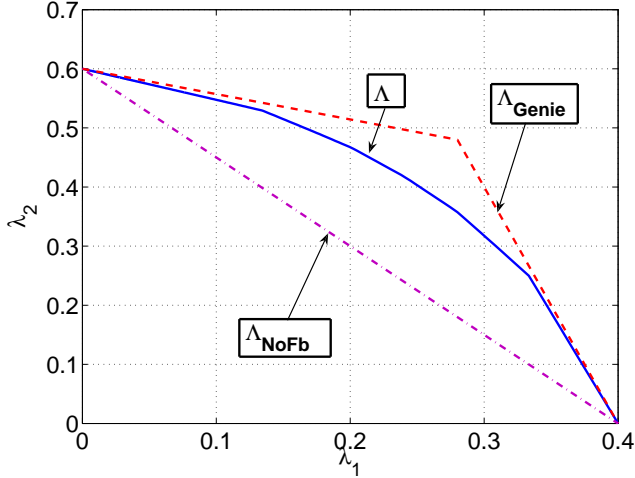


Fig. 3: Comparison of stability regions. Parameters used:  $p_{11}^1 = 0.7$ ,  $p_{01}^1 = 0.2$ ;  $p_{11}^2 = 0.8$ ,  $p_{01}^2 = 0.3$

value captures the importance of scheduling a user under two sophisticated system features – the queue evolution and the fundamental exploration-exploitation tradeoff.

(3) Calculation of  $\mathbf{q}[kT]$ -weighted index value is very simple, which only requires scaling the *pre-calculated* Whittle’s index value. Under the queue-based index policy, in each frame, implementation of  $G^\tau(\mathbf{q}[kT], M-g(\tau))$  in step 2 of policy  $\text{Q-Index}_\tau(T, M-g(\tau))$  has computational complexity  $O((2\tau+1)N \log(2\tau+1)N)$ , while implementing step 3 of policy  $\text{Q-Index}_\tau(T, M-g(\tau))$  over the frame has complexity  $O(TN)$  (see the remark in Section III-E). Accordingly, the *per-frame* complexity is  $O((2\tau+1)N \log(2\tau+1)N + TN)$ . Therefore, as the frame length  $T$  scales up, the *per-slot* complexity decreases toward  $O(N)$ .

(4) The scheduling decisions are made by comparing each user’s own index value to a threshold, independently from other users. Hence our policy is also applicable for *distributed implementation* in uplink scenarios.

## V. NUMERICAL RESULTS

### A. Illustration of Stability Region

In Fig. 3, we compute the stability region  $\Lambda$  and compare it with other regions of interest. We consider the scenario with two users and with the scheduling constraint on the long-term average number of scheduled transmissions  $M = 1$ . The Markov transition statistics are selected as  $(p_{11}^1, p_{01}^1) = (0.7, 0.2)$ ,  $(p_{11}^2, p_{01}^2) = (0.8, 0.3)$ . For comparison, in the same system, we consider another scenario where the scheduler throws away the ARQ feedback from the scheduled user. We denote the corresponding stability region by  $\Lambda_{NoFb}$ , expressed as  $\Lambda_{NoFb} = \{\lambda : \lambda_1/b_s^1 + \lambda_2/b_s^2 \leq 1\}$  [36]. As can be observed in the figure, by exploiting the channel memory from ARQ feedback, our policy achieves significant throughput gain (as high as 30%) over the policy that ignores the channel memory. We also compare the stability region  $\Lambda$  with that of a ‘genie-aided’ system, denoted by  $\Lambda_{Genie}$ . In the ‘genie-aided’ system, the same scheduling constraint (1) is imposed, while a genie reveals channel states of *all users* in the current

slot to the scheduler at the end of the slot. The region  $\Lambda_{Genie}$  is expressed as

$$\Lambda_{Genie} = b_s^1 b_s^2 \lambda_{00} + (1 - b_s^1) b_s^2 \lambda_{01} + b_s^1 (1 - b_s^2) \lambda_{10} + (1 - b_s^1) (1 - b_s^2) \lambda_{11}$$

with  $\lambda_{ij} \in \Lambda_{ij}$  where  $\Lambda_{ij} = \mathcal{CH}\{(p_{i1}^1, 0), (0, p_{j1}^2)\}$ ,  $i, j = 0, 1$  with  $\mathcal{CH}\{\cdot\}$  denoting the convex hull of the set [25]. Because the genie facilitates more informed decisions at the scheduler, the resultant stability region  $\Lambda_{Genie}$  provides an outer bound on region  $\Lambda$ , as demonstrated in Fig. 3.

### B. Delay Performance Analysis

In this section, we numerically evaluate the delay performances of the proposed policy. We consider a two users system with the long-term average number of transmission constraint  $M = 1$ , i.e., one user can be scheduled on average. The channel states of both users evolve as the ‘ON/OFF’ Markov chain with transition statistics  $(p_{11}^1, p_{01}^1) = (0.7, 0.2)$ ,  $(p_{11}^2, p_{01}^2) = (0.8, 0.3)$ , i.e., which can be typical situations where both users have moderate degree of correlation across time.

Over this system, we implement the proposed  $T$ -frame queue-based index policy  $\text{Q-Index}_\tau(T, M-g(\tau)/2)$ , defined in section IV with  $\tau=20$ . We first consider fixed arrival rates  $\lambda_1 = \lambda_2 = 0.25$  and implement the policies  $\text{Q-Index}_\tau(T, M-g(\tau)/2)$  with frame lengths  $T=10$  and  $T=100$ , respectively. The sample paths of the average queue length, i.e.,  $(Q_1[t] + Q_2[t])/2$ , are plotted in Fig. 4. It can be observed that, while the queues in both scenarios are stable, the variation of the queue evolution is notably higher when the frame size changes from 10 to 100. This is because, as the frame size increases, the frame-based algorithm obtains less frequent updates of the queue sizes. Therefore, within a frame, the algorithm can continue to serve a user even if its current queue length becomes small while neglecting the other user that has accumulated a large queue size, leading to a higher degree of queue length variation as well as average queue size. Correspondingly, higher delay and delay variation are expected as the frame size increases. For example, suppose the initial queue length of user 1 is empty, while the initial queue length for user 2 is nonempty. Then user 1 in the first frame will not be scheduled. Now after the first frame, the expected queue length of user 1 will be significantly larger for the case when  $T = 100$  compared with the case when  $T = 10$ . Hence, at the second frame, the scheduler dedicates most of the resources to user 1. As a result, the expected queue length of the user 1 will go down after second frame, and the expected queue length of user 2 will grow. Both the expected change of queue lengths of user 1 and 2 will be much more significant when  $T = 100$  compared with when  $T = 10$ . The process repeats in time and results in a higher degree of queue length variation when  $T = 100$  as compared to  $T = 10$ .

We next implement the aforementioned policy  $\text{Q-Index}_\tau(T, M-g(\tau)/2)$  and evaluate the average queueing delay experienced by users as the arrival rates scale toward the boundary of the stability region, with varying frame length  $T$ . For the two user system previously discussed,



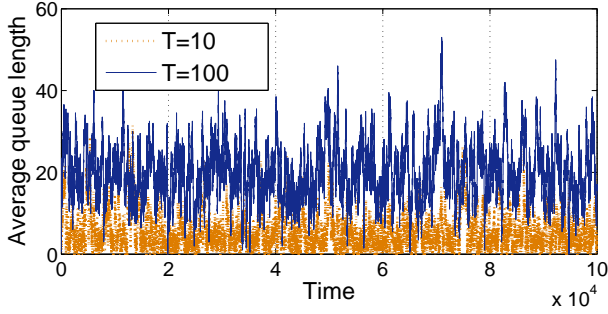


Fig. 4: Sample paths of queue evolution.

Fig. 5 examines the average queueing delay when the arrival rate vector  $(\lambda_1, \lambda_2)$  increases with  $\lambda_1 = \lambda_2 = \lambda$ . As can be observed in the figure, as the arrival rates grow toward the boundary of stability region, the queue length quickly blows up, resulting in steep increase of delay. The steep increase is because, as the arrival rates grow toward the boundary of stability region, the queue lengths quickly blow up because they are becoming unstable, resulting in steep increase of average delay. Fig. 5 also show that, as the frame length grows, the average delay in the downlink network increases. This is, again, a consequence of infrequent update of queue length information at the scheduler.

Another interesting observation can be observed from Fig. 5. When we implement the proposed policy  $Q\text{-Index}_r(T, M - g(\tau)/2)$  with the frame lengths  $T$  growing from 9 to 100, the system delay curves for different values of  $T$  start to build up significantly at *around the same value* (i.e., around 0.29 which is on the boundary of the stability region). Note that we needed the frame size to be large enough to prove Proposition 1. However, in practice, the frame size  $T$  may not need to be as large to guarantee queue stability. This numerical result, along with many other numerical evaluations we have conducted, indicates that the queues are stable under only moderate value of frame size in the proposed queue-based index policy.

Fig. 5 also plots the delay performance of a policy  $\phi^{NoFb}$  that ignores the channel memory, i.e., not using the channel state feedback. In each slot of this policy, a user  $i$  with the largest multiplication of steady state transmission rate (i.e.,  $b_s^i$ ) and queue length  $q_i[t]$  is scheduled. The delay performance of maximum weight matching policy  $\phi^{MWM}$  is also plotted, where, in each slot  $t$ , a user  $i$  with the largest multiplication of belief value  $\pi^i[t]$  and queue length  $q_i[t]$  is scheduled. Fig. 5 further plots the delay performance of a naive policy  $\phi^{NaiveInd}$  where a user  $i$  with the largest multiplication of index value  $W_i^1(\pi^i[t])$  and queue length  $q_i[t]$  is scheduled. For all of these policies, the values of arrival rate  $\lambda$  where the queueing delay increases steeply are at a smaller value than our proposed policy, implying the sub-optimality of these policies. This is partly because these policies only schedule strictly  $M$  users per slot, but our work is in the domain of a relaxed constraint of average number of scheduled users. The sub-optimality of policy  $\phi^{MWM}$  is also because it only exploits the channel condition in the instantaneous slot, i.e.,  $\pi_i[t]$ , but it does not consider exploring outdated channels. It is interesting

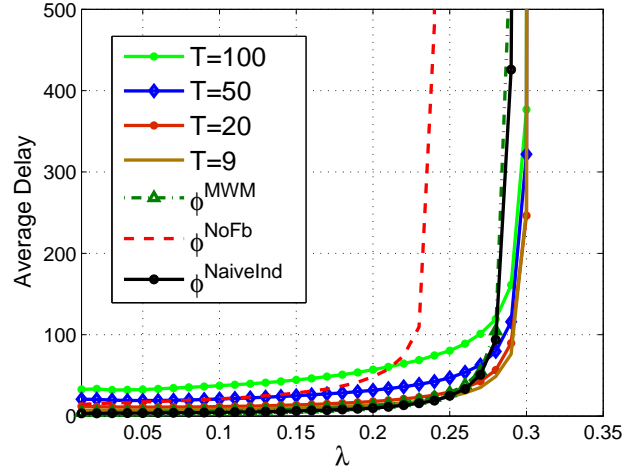


Fig. 5: Delay performance comparison when  $N = 2$ .

to note that policy  $\phi^{MWM}$  and  $\phi^{NaiveInd}$  performs better than the policy that ignores channel state feedback, as the value of  $\lambda$  where steep increase of queueing delay occurs is much larger as compared to  $\phi^{NoFb}$ . This observation illustrates the throughput gains that can be achieved by using the channel state feedback.

## VI. CONCLUSION

In this work, we have studied downlink scheduling problem over Markovian evolving ON/OFF fading channels and imperfect instantaneous channel state information. The scheduling decisions are made based on the single-bit ARQ-type feedback and the channel memory inherent in the Markovian channels. We propose a throughput-optimal policy that operates over time frames. In the proposed policy, the importance of scheduling a user is measured by a simple multiplication of the queue length and Whittle's index value. Because of this property, the proposed policy has low-complexity per frame in the network size and the truncation level of the belief state space. Most notably, our policy does not suffer from the curse of dimensionality that is observed in earlier works in this context. Numerical evaluations show that significant throughput performance gains can be achieved by exploiting the channel memory, via the frame-based low-complexity queue-based index policy with moderate frame size. Future directions include considering larger state space model, and considering feedback mechanisms that collects CSI from unscheduled users, as well as more stringent instantaneous scheduling constraints. Another open direction is to consider adaptive power allocation with hybrid ARQ protocols (e.g., [9]), where the index value not only implies the attractiveness of scheduling a user, but also guides the power allocation across time.

## APPENDIX A PROOF OF LEMMA 1

The proof of the lemma is an extension of the proof of Proposition 1 in [16]. Consider the problem  $\Psi(\mathbf{r}, M)$  with weight vector  $\mathbf{r}$ . The constraint (1) can be written in an

equivalent form that requires at least  $N - M$  channels to be *passive* on average, i.e.,

$$\liminf_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{i=1}^N (1 - a_i^\phi[t]) \right] \geq N - M. \quad (15)$$

Associating a Lagrange multiplier  $\omega$  to the constraint (15), we have the following Lagrangian function  $L(\phi, \omega)$  for problem  $\Psi(\mathbf{r}, M)$ ,

$$\begin{aligned} L(\phi, \omega) = & \liminf_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{i=1}^N r_i \cdot \pi_i[t] \cdot a_i^\phi[t] \right] \\ & + \omega \cdot \liminf_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{i=1}^N (1 - a_i^\phi[t]) \right] - \omega \cdot (N - M). \end{aligned} \quad (16)$$

The dual function  $D(\omega)$  is defined as  $D(\omega) = \max_{\phi \in \Phi} L(\phi, \omega)$ . Following the lines of proof in [16] we have

$$D(\omega) = \sum_{i=1}^N U_i^{r_i}(\omega) + \omega(N - M).$$

in which  $U_i^{r_i}(\omega)$  is a  $\omega$ -subsidy problem under weight  $r_i$ ,

$$\begin{aligned} U_i^{r_i}(\omega) = & \max_{\phi \in \Phi_i} \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} [r_i \cdot \pi_i[t] \cdot a_i^\phi[t] \right. \\ & \left. + \omega \cdot (1 - a_i^\phi[t]) \right], \end{aligned} \quad (17)$$

where  $\Phi_i$  denotes the set of scheduling policies that activate and idle the user  $i$  according to the observed channel history. In the above problem (17), for each channel  $i$  at belief state  $\pi_i$ , it will receive a reward  $r_i \pi_i$  when it activates, otherwise it will receive a subsidy  $\omega$  for passivity. We let  $\mathcal{I}_i^{r_i}(\omega) \subseteq \mathcal{B}_i$  be the set of belief states for which it is optimal to stay idle.

Under the unit weight  $r_i = 1$ , it was shown in [18] that the problem is Whittle indexable, i.e.,  $\mathcal{I}_i^1(\omega)$  monotonically increases from  $\emptyset$  to  $\mathcal{B}_i$  as  $\omega$  increase from 0 to  $\infty$  for each user  $i$ . The Whittle's index value  $W_i^1(\pi)$  is defined as the infimum subsidy value for which the belief state  $\pi$  is at the boundary of  $\mathcal{I}_i^1(\omega)$ , i.e.,

$$W_i^1(\pi) = \inf \{ \omega : \pi \in \mathcal{I}_i^1(\omega) \}.$$

It follows from [16] that, for the  $\omega$ -subsidy problem under unit weight  $r_i = 1$ , the optimal policy is to activate the user at time slot  $t$  if  $W_i^r(\pi) > \omega$ , and to stay idle if  $W_i^r(\pi) < \omega$ , with tie breaking arbitrarily if  $W_i^r(\pi) = \omega$ .

We next extend the optimal algorithm for the  $\omega$ -subsidy problem under unit weight to the general case with arbitrary non-negative weight  $r_i$ . An equivalent form of  $U_i^{r_i}(\omega)$  is as follows,

$$\begin{aligned} & U_i^{r_i}(\omega) \\ = & r_i \max_{\phi \in \Phi_i} \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} [\pi_i[t] a_i^\phi[t] + \frac{\omega}{r_i} (1 - a_i^\phi[t])] \right]. \end{aligned} \quad (18)$$

Therefore, the optimal solution for the  $\omega$ -subsidy problem (17) with weight  $r_i$  takes the same form as the optimal solution for the  $\omega/r_i$ -subsidy problem with weight 1. Accordingly, the optimal solution takes the following form: a user  $i$  is scheduled

at slot  $t$  if  $W_i^r(\pi_i[t]) > \omega/r_i$ , and stay idle if  $W_i^r(\pi) < \omega/r_i$ , with tie breaking arbitrarily if  $W_i^r(\pi) = \omega/r_i$ .

We define the  $\mathbf{r}$ -weighted index value as  $W_i^r(\pi) = r_i \cdot W_i^1(\pi)$ ,  $\pi \in \mathcal{B}_i$ ,  $i \in \{1, \dots, N\}$ . The optimal policy for the reward maximization problem in (18) is then to activate the user  $i$  if  $W_i^r(\pi) > \omega$ , and to stay idle if  $W_i^r(\pi) < \omega$ , with tie breaking arbitrarily if  $W_i^r(\pi) = \omega$ . Because of this threshold-based policy and arbitrary tie-breaking at the threshold, the dual function value  $D(\omega)$  can be achieved by the following threshold-based policy implemented over the  $\mathbf{r}$ -weighted index values  $W_i^r(\pi)$ : User  $i$  is scheduled if  $W_i^r(\pi_i) > \omega$ , or if  $W_i^r(\pi_i) = \omega$  with  $i > j$ . User  $i$  stays idle if  $W_i^r(\pi_i) < \omega$ , or if  $W_i^r(\pi_i) = \omega$  with  $i < j$ . If  $W_i^r(\pi_i) = \omega$  with  $i = j$ , user  $i$  is scheduled with probability  $\rho$ .

Following the similar proof techniques of Lemma 11 in [16], by appropriately choosing the aforementioned parameters  $(j, \omega, \rho)$  to be  $(i^*, \omega^*, \rho^*)$  such that the constraint (1) on the average number of transmissions is strictly satisfied with equality, the corresponding policy is optimal for the problem  $\Psi(\mathbf{r}, M)$ . Denoting such a policy as  $\phi^*(\mathbf{r}, M)$ , the proposition is proven.

## APPENDIX B PROOF OF LEMMA 2

We next prove the Lemma for  $\alpha_j^r(j, \omega, \rho)$ .

Case (1). First consider  $\alpha_j^r(j, W_j^r(b_{0,h}^j), \rho)$  with  $h < \tau$ . Hence user  $j$  is scheduled if its belief value is above  $b_{0,h}^j$ , or is scheduled with probability  $\rho$  at belief value  $b_{0,h}^j$ . According to the belief value evolution rule (2), in the next slot, its belief value will either be  $p_{11}^j$  or  $p_{01}^j$ , depending on the whether the revealed channel state is '0' or '1' at the end of the current slot. If the user's belief value is below  $b_{0,h}^j$ , it will not be scheduled and its belief value will move one step toward  $b_{0,h+1}^j$ . Hence, in this case, the belief value evolution for user  $j$  follows a Markov Chain over  $\mathcal{B}_j^r$ , as depicted in Fig. 6.

From Fig. 6, one can observe that the belief Markov chain is ergodic and the recurrent states are  $\{b_{1,1}^j, b_{0,l}^j, l = 1, \dots, h+1\}$ . We denote the stationary probability of belief value being  $\pi_j$  as  $\zeta_j(\pi_j)$ ,  $\pi_j \in \mathcal{B}_j^r$ . The global balance equations are

$$\begin{aligned} & \rho(1 - b_{0,h}^j) \zeta_j(b_{0,h}^j) + \zeta_j(b_{0,h+1}^j)(1 - b_{0,h+1}^j) \\ & \quad + b_{1,1}^j(1 - p_{11}^j) = \zeta_j(b_{0,1}^j) \\ & \zeta_j(b_{0,1}^j) = \zeta_j(b_{0,2}^j) = \dots = \zeta_j(b_{0,h}^j) \\ & (1 - \rho) \zeta_j(b_{0,h}^j) = \zeta_j(b_{0,h+1}^j) \\ & \rho \zeta_j(b_{0,h}^j) + \zeta_j(b_{0,h+1}^j) = (1 - p_{11}^j) \zeta_j(b_{1,1}^j) \end{aligned}$$

From the balance equations, we can calculate the expression of the stationary probability as follows,

$$\zeta_j(\pi_j) = \begin{cases} \frac{1 - p_{11}^j}{\rho b_{0,h}^j + (1 - \rho) b_{0,h+1}^j + (1 - p_{11}^j)(h + 1 - \rho)} & \text{if } \pi_j = b_{0,k}^j, k \leq h; \\ \frac{(1 - \rho)(1 - p_{11}^j)}{\rho b_{0,h}^j + (1 - \rho) b_{0,h+1}^j + (1 - p_{11}^j)(h + 1 - \rho)} & \text{if } \pi_j = b_{0,h+1}^j; \\ \frac{b_{0,h+1}^j + \rho(b_{0,h}^j - b_{0,h+1}^j)}{\rho b_{0,h}^j + (1 - \rho) b_{0,h+1}^j + (1 - p_{11}^j)(h + 1 - \rho)} & \text{if } \pi_j = b_{1,1}^j; \\ 0 & \text{otherwise.} \end{cases}$$

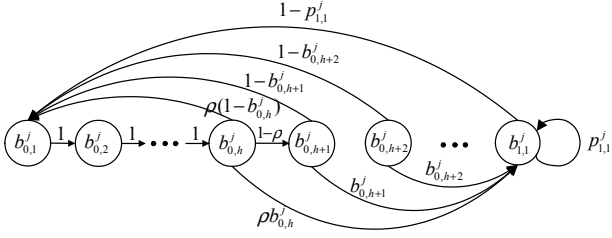


Fig. 6: Belief value transition in steady state when  $\omega = W_j^F(b_{0,h}^j)$ .

Hence, the expected fraction of time transmitting to user  $j$  is

$$\begin{aligned} \alpha_j^\tau(j, W_j^F(b_{0,h}^j), \rho) &= \rho \zeta_j(b_{0,h}^j) + \zeta_j(b_{0,h+1}^j) + \zeta_j(b_{0,1}^j) \\ &= \frac{\rho(b_{0,h}^j - b_{0,h+1}^j) + 1 - p_{11}^j + b_{0,h+1}^j}{\rho b_{0,h}^j + (1-\rho)b_{0,h+1}^j + (1-p_{11}^j)(h+1-\rho)}, \end{aligned}$$

as given in Lemma 2(i). To prove part (ii), we consider its reciprocal, i.e.,

$$\begin{aligned} &[\alpha_j^\tau(j, W_j^F(b_{0,h}^j), \rho)]^{-1} \\ &= 1 + \frac{(1-p_{11}^j)(h-\rho)}{\rho(b_{0,h}^j - b_{0,h+1}^j) + 1 - p_{11}^j + b_{0,h+1}^j} \\ &= 1 + \frac{1-p_{11}^j}{b_{0,h+1}^j - b_{0,h}^j} \left[ 1 - \frac{1-p_{11}^j + b_{0,h+1}^j + h(b_{0,h}^j - b_{0,h+1}^j)}{\rho(b_{0,h}^j - b_{0,h+1}^j) + b_{0,h+1}^j + (1-p_{11}^j)} \right]. \end{aligned} \quad (19)$$

Considering the numerator inside the parenthesis of (19), we have

$$\begin{aligned} &1 - p_{11}^j + b_{0,h+1}^j + h(b_{0,h}^j - b_{0,h+1}^j) \\ &\geq 1 - p_{11}^j + b_{0,h+1}^j + (h+1)(b_{0,h}^j - b_{0,h+1}^j) \geq 0, \end{aligned}$$

where the last inequality is from (51). Noting that the denominator inside the parenthesis of (19) strictly decreases with  $\rho$ , hence  $[\alpha_j^\tau(j, W_j^F(b_{0,h}^j), \rho)]^{-1}$  strictly decreases with  $\rho$ . Therefore  $\alpha_j^\tau(j, W_j^F(\pi_j), \rho)$  strictly decreases with  $\rho$  for  $\pi_j$  for  $\pi_j \in \{b_{0,1}^j, b_{0,2}^j, \dots, b_{0,\tau-1}^j\}$

Since for  $h+1 < \tau$ ,  $\alpha_j^\tau(j, W_j^F(b_{0,h}^j), 0) = \alpha_j^\tau(j, W_j^F(b_{0,h+1}^j), 1)$ , we have,

$$\begin{aligned} \alpha_j^\tau(j, W_j^F(b_{0,h+1}^j), \rho) &\leq \alpha_j^\tau(j, W_j^F(b_{0,h+1}^j), 1) \\ &= \alpha_j^\tau(j, W_j^F(b_{0,h}^j), 0) \\ &\leq \alpha_j^\tau(j, W_j^F(b_{0,h}^j), \rho). \end{aligned}$$

Therefore, for fixed  $\rho$ ,  $\alpha_j^\tau(j, W_j^F(\pi_j), \rho)$  strictly decreases with  $\pi_j$  for  $\pi_j \in \{b_{0,1}^j, b_{0,2}^j, \dots, b_{0,\tau-1}^j\}$ .

Case (2). Next consider  $\alpha_j^\tau(j, W_j^F(b_{0,\tau}^j), \rho)$ . We can perform

a similar analysis as in case (1) to obtain

$$\begin{aligned} \alpha_j^\tau(j, W_j^F(b_{0,\tau}^j), \rho) &= \frac{\rho(b_{0,\tau}^j - b_s^j) + 1 - p_{11}^j + b_s^j}{\rho b_{0,\tau}^j + (1-\rho)b_s^j + (1-p_{11}^j)(\tau+1-\rho)}, \\ &[\alpha_j^\tau(j, W_j^F(b_{0,\tau}^j), \rho)]^{-1} \\ &= 1 + \frac{1-p_{11}^j}{b_s^j - b_{0,\tau}^j} \left[ 1 - \frac{1-p_{11}^j + b_s^j + \tau(b_{0,\tau}^j - b_s^j)}{\rho(b_{0,\tau}^j - b_s^j) + b_s^j + (1-p_{11}^j)} \right]. \end{aligned} \quad (20)$$

When  $\tau > \tau_0$ , it can be derived that the numerator  $1 - p_{11}^j + b_s^j + \tau(b_{0,\tau}^j - b_s^j)$  inside (20) is positive. Therefore  $\alpha_j^\tau(j, W_j^F(b_{0,\tau}^j), \rho)$  strictly increases with  $\rho$ . Similar to Case (1), we have

$$\begin{aligned} \alpha_j^\tau(j, W_j^F(b_{0,\tau}^j), \rho) &\leq \alpha_j^\tau(j, W_j^F(b_{0,\tau}^j), 1) = \alpha_j^\tau(j, W_j^F(b_{0,\tau-1}^j), 0) \\ &\leq \alpha_j^\tau(j, W_j^F(b_{0,\tau-1}^j), \rho). \end{aligned}$$

Case (3). Consider  $\alpha_j^\tau(j, W_j^F(b_s^j), \rho)$ . Similar to Case (1), we obtain,

$$\alpha_j^\tau(j, W_j^F(b_s^j), \rho) = \frac{\rho(1-p_{11}^j + b_s^j)}{(1+\tau\rho)(1-p_{11}^j) + \rho b_s^j}.$$

Taking the reciprocal we have

$$[\alpha_j^\tau(j, W_j^F(b_s^j), \rho)]^{-1} = \frac{1}{1-p_{11}^j + b_s^j} \left( (1-p_{11}^j)(\tau + \frac{1}{\rho}) + b_s^j \right),$$

which strictly decreases with  $\rho$ . Hence  $\alpha_j^\tau(j, W_j^F(b_s^j), \rho)$  strictly increases with  $\rho$ . We also have

$$\begin{aligned} \alpha_j^\tau(j, W_j^F(b_s^j), \rho) &\leq \alpha_j^\tau(j, W_j^F(b_s^j), 1) = \alpha_j^\tau(j, W_j^F(b_{0,\tau}^j), 0) \\ &\leq \alpha_j^\tau(j, W_j^F(b_{0,\tau}^j), \rho). \end{aligned}$$

From Case (1)-(3), the lemma is established for  $\alpha_j^\tau(j, W_j^F(b_{0,h}^j), \rho)$ . Noting that for user  $i \neq j$ , there is no randomization associated with scheduling. Hence, the above derivation for  $\alpha_j^\tau(j, W_j^F(b_{0,h}^j), \rho)$  naturally extends to  $\alpha_i^\tau(j, W_i^F(b_{0,h}^j), \rho)$ . The only change is there is no longer randomization involved. Details are hence neglected here.

## APPENDIX C PROOF OF LEMMA 3

### A. Proof outline

We establish the proof by first proving lemma 4 that bounds the difference of weighted sum-throughput between policies with different threshold parameters, with respect to the difference between expected fraction of transmission time to each user. We then prove the lemma under two cases, i.e., whether  $\omega^* < W_i^F(b_{0,\tau}^j)$  for all user  $i$ . The first case is uncomplicated to prove. For the second case, we first prove a useful fact that only one of the three cases holds:  $\omega_\tau > \omega^*$ , or  $\omega_\tau = \omega^*$  with  $\rho_\tau < \rho^*$  and  $i_\tau = i^*$ , or  $\omega_\tau = \omega^*$  with  $i_\tau > i^*$ . Based on these cases, we can bound the difference between expected fraction of time transmitting to different users. We then use Lemma 4 to finish the proof.

### B. Notations

Recall that, in the untruncated state space, the optimal policy  $\phi^*(\mathbf{r}, M)$  corresponds to the parameters  $(i^*, \omega^*, \rho^*)$ . Also



recall that, in the truncated state space, the policy  $\phi_\tau(\mathbf{r}, M)$  corresponds to the parameter  $(i_\tau, \omega_\tau, \rho_\tau)$ .

Over the actual *untruncated* model, consider the following policy denoted as  $\phi_{j,\omega,\rho}^{untrunc}$  with the parameters  $(j, \omega, \rho)$ : User  $i$  is scheduled if  $W_i^r(\pi_i[t]) > \omega$ , or if  $W_i^r(\pi_i[t]) = \omega^*$  with  $i > j$ . User  $i$  stays idle if  $W_i^r(\pi_i[t]) < \omega$ , or if  $W_i^r(\pi_i[t]) = \omega^*$  with  $i < j$ . If  $W_i^r(\pi_i[t]) = \omega$  with  $i = j$ , it is scheduled with probability  $\rho$ . In this model, similar to (8), we let  $\alpha_i(j, \omega, \rho)$  denote the long-term expected fraction of time transmitting to user  $i$  under policy  $\phi_{j,\omega,\rho}^{untrunc}$ , i.e.,

$$\alpha_i(j, \omega, \rho) = \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} a_i^{\phi_{j,\omega,\rho}^{untrunc}} [t] \right]. \quad (21)$$

The closed-form expression of  $\alpha_i(j, \omega, \rho)$  can be calculated from the same technique we used to prove Lemma 2 as follows.

$$\alpha_i(j, \omega, \rho) = \begin{cases} \frac{\rho(b_{0,h}^i - b_{0,h+1}^i) + 1 - p_{11}^i + b_{0,h+1}^i}{\rho b_{0,h}^i + (1-\rho)b_{0,h+1}^i + (1-p_{11}^i)(h+1-\rho)} & \text{if } \omega = W_i^r(b_{0,h}^i), i=j; \\ \frac{1-p_{11}^i + b_{0,h+1}^i}{b_{0,h+1}^i + (1-p_{11}^i)(h+1)} & \text{if } \omega = W_i^r(b_{0,h}^i), i < j \\ \frac{1-p_{11}^i + b_{0,h}^i}{b_{0,h}^i + (1-p_{11}^i)h} & \text{if } \omega = W_i^r(b_{0,h}^i), i > j; \\ 0 & \text{or if } W_i^r(b_{0,h-1}^i) < \omega < W_i^r(b_{0,h}^i), i \neq j \\ & \text{if } \omega \geq W_i^r(b_{0,h}^i). \end{cases} \quad (22)$$

We also let  $v_i(j, \omega, \rho)$  denote the long-term expected transmission rate to user  $i$ , i.e.,

$$v_i(j, \omega, \rho) = \liminf_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} r_i \cdot \pi_i[t] \cdot a_i^{\phi_{j,\omega,\rho}^{untrunc}} [t] \right], \quad (23)$$

Over the *truncated* model, correspondingly, we let  $v_i^\tau(j, \omega, \rho)$  denote the long-term expected transmission rate to user  $i$  under policy  $\phi_{j,\omega,\rho}^{trunc}$  defined in section III-D, i.e.,

$$v_i^\tau(j, \omega, \rho) = \liminf_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} r_i \cdot \pi_i^\tau[t] \cdot a_i^{\phi_{j,\omega,\rho}^{trunc}} [t] \right]. \quad (24)$$

Using techniques similar to the proof of Lemma 2, we can derive the analytical expressions of  $v_i(j, \omega, \rho)$  and  $v_i^\tau(j, \omega, \rho)$  as follows,

$$v_i(j, \omega, \rho) = \begin{cases} r_i \cdot \frac{\rho b_{0,h}^i + (1-\rho)b_{0,h+1}^i}{\rho b_{0,h}^i + (1-\rho)b_{0,h+1}^i + (1-p_{11}^i)(h+1-\rho)} & \text{if } \omega = W_i^r(b_{0,h}^i), i=j \\ r_i \cdot \frac{b_{0,h+1}^i}{b_{0,h+1}^i + (1-p_{11}^i)(h+1)} & \text{if } \omega = W_i^r(b_{0,h}^i), i < j \\ r_i \cdot \frac{b_{0,h}^i}{b_{0,h}^i + (1-p_{11}^i)h} & \text{if } \omega = W_i^r(b_{0,h}^i), i > j \\ 0 & \text{or if } W_i^r(b_{0,h-1}^i) < \omega < W_i^r(b_{0,h}^i), i \neq j \\ & \text{if } \omega \geq W_i^r(b_{0,h}^i). \end{cases} \quad (25)$$

The expression of  $v_j^\tau(j, \omega, \rho)$  is given as follows,

$$v_j^\tau(j, \omega, \rho) = \begin{cases} r_j \cdot \frac{\rho b_{0,h}^j + (1-\rho)b_{0,h+1}^j}{\rho b_{0,h}^j + (1-\rho)b_{0,h+1}^j + (1-p_{11}^j)(h+1-\rho)} & \text{if } h < \tau, \omega = W_j^r(b_{0,h}^j); \\ r_j \cdot \frac{\rho b_{0,\tau}^j + (1-\rho)b_s^j}{\rho b_{0,\tau}^j + (1-\rho)b_s^j + (1-p_{11}^j)(\tau+1-\rho)} & \text{if } \omega = W_j^r(b_{0,\tau}^j); \\ r_j \cdot \frac{\rho b_s^j}{(1+\tau\rho)(1-p_{11}^j) + \rho b_s^j} & \text{if } \omega = W_j^r(b_s^j); \\ 0 & \text{if } \omega > W_j^r(b_s^j). \end{cases} \quad (26)$$

The expression of  $v_i^\tau(j, \omega, \rho)$ ,  $i \neq j$  is expressed as follows.

$$v_i^\tau(j, \omega, \rho) = \begin{cases} r_i \cdot \frac{b_{0,h+1}^i}{b_{0,h+1}^i + (1-p_{11}^i)(h+1)} & \text{if } h < \tau, \omega = W_i^r(b_{0,h}^i), i < j \\ r_i \cdot \frac{b_{0,h}^i}{b_{0,h}^i + (1-p_{11}^i)h} & \text{if } h \leq \tau, \omega = W_i^r(b_{0,h}^i), i > j; \\ & \text{or if } h \leq \tau, W_i^r(b_{0,h-1}^i) < \omega < W_i^r(b_{0,h}^i) \\ r_i \cdot \frac{b_s^i}{(1+\tau)(1-p_{11}^i) + b_s^i} & \text{if } \omega = W_i^r(b_{0,\tau}^i), i < j; \\ & \text{or if } \omega = W_i^r(b_s^i), i > j \\ 0 & \text{if } \omega = W_i^r(b_s^i), i < j \\ & \text{or if } \omega > W_i^r(b_s^i). \end{cases} \quad (27)$$

### C. Proof of Lemma 3

We first prove the following lemma that provides properties of  $\alpha_i^\tau(j, \omega, \rho)$  and  $v_i^\tau(j, \omega, \rho)$ .

**Lemma 4.** For a user  $i$ , if  $\tau \geq \tau_0$ , we have

- (i) For fixed  $\pi_j \in \{b_{0,1}^j, b_{0,2}^j, \dots, b_{0,\tau}^j, b_s^j\}$ ,  $v_j^\tau(j, W_j^r(\pi_i), \rho)$  strictly increases with  $\rho$ . For fixed  $\rho$ ,  $v_i^\tau(j, W_i^r(\pi_i), \rho)$  strictly decreases with  $\pi_i$  for  $\pi_i \in \{b_{0,1}^i, b_{0,2}^i, \dots, b_{0,\tau}^i, b_s^i\}$  and all  $i$ ;
- (ii) for any two sets of parameter  $\{j_1, \omega_1, \rho_1\}$  and  $\{j_2, \omega_2, \rho_2\}$ ,

$$\begin{aligned} & \left| v_i^\tau(j_1, \omega_1, \rho_1) - v_i^\tau(j_2, \omega_2, \rho_2) \right| \\ & \leq r_i \cdot \left| \alpha_i^\tau(j_1, \omega_1, \rho_1) - \alpha_i^\tau(j_2, \omega_2, \rho_2) \right|. \end{aligned}$$

**Proof:** See Appendix E. ■

Note that we need  $\tau \geq \tau_0$  for the proof to hold. Since the untruncated state space is in the asymptotic regime of the truncated scenario when  $\tau \rightarrow \infty$ , a straightforward extension of properties of  $\alpha_i^\tau(j, \omega, \rho)$  and  $v_i^\tau(j, \omega, \rho)$  in Lemma 2 and Lemma 4 to  $\alpha_i(j, \omega, \rho)$  and  $v_i(j, \omega, \rho)$  in the untruncated scenario leads to the next Lemma.

**Lemma 5.** For a user  $i$ , if  $\tau \geq \tau_0$ , we have

- (i) For fixed  $\pi_j \in \{b_{0,1}^j, b_{0,2}^j, \dots, b_{0,\tau}^j, b_s^j\}$ ,  $v_j(j, W_j^r(\pi_i), \rho)$  and  $\alpha_i(j, W_i^r(\pi_i), \rho)$  strictly increase with  $\rho$ . For fixed  $\rho$ ,  $v_i(j, W_i^r(\pi_i), \rho)$  and  $\alpha_i(j, W_i^r(\pi_i), \rho)$  strictly decrease with  $\pi_i$  for  $\pi_i \in \{b_{0,1}^i, b_{0,2}^i, \dots, b_{0,\tau}^i, b_s^i\}$ ;
- (ii) for any two sets of parameters  $\{j_1, \omega_1, \rho_1\}$  and

$\{j_2, \omega_2, \rho_2\}$ ,

$$\begin{aligned} & \left| v_i(j_1, \omega_1, \rho_1) - v_i(j_2, \omega_2, \rho_2) \right| \\ & \leq r_i \left| \alpha_i(j_1, \omega_1, \rho_1) - \alpha_i(j_2, \omega_2, \rho_2) \right|. \end{aligned}$$

We proceed to prove Lemma 3 under two cases.

Case (1). If the threshold  $\omega^*$  satisfies  $\omega^* < W_i^{\mathbf{r}}(b_{0,\tau}^i)$  for all user  $i$ , then the approximation parameters  $i_\tau = i^*$ ,  $\omega_\tau = \omega^*$  and  $\rho_\tau = \rho^*$ . This is because, if  $\omega^* < W_i^{\mathbf{r}}(b_{0,\tau}^i)$  for all user  $i$ , no user will stay idle for more than  $\tau$  slots under the optimal policy  $\phi^*(\mathbf{r}, M)$ . To see this in more detail, the expected amount of transmissions equals to  $M$ , i.e.,  $\sum_{i=1}^N \alpha_i^{\mathbf{r}}(j, \omega, \rho) = M$ , when  $j = i^*$ ,  $\omega = \omega^*$ ,  $\rho = \rho^*$ , which meets the constraint (10). Therefore, thanks to the strict monotonicity property in Lemma 2(ii), the algorithm  $G^{\mathbf{r}}(\mathbf{r}, M)$  outputs  $i_\tau = i^*$ ,  $\omega_\tau = \omega^*$  and  $\rho_\tau = \rho^*$ , and hence policy  $\phi_\tau(\mathbf{r}, M)$  is equivalent to the policy  $\phi^*(\mathbf{r}, M)$ . We hence have  $|V^*(\mathbf{r}, M) - V_\tau(\mathbf{r}, M)| = 0$  and  $Z_\tau(\mathbf{r}, M) = M$ .

Case (2). If there exists a user  $i$  with  $\omega^* \geq W_i^{\mathbf{r}}(b_{0,\tau}^i)$ , we let  $\Theta$  denote the corresponding set of users, i.e.,  $\Theta = \{i : W_i^{\mathbf{r}}(b_{0,\tau}^i) \leq \omega^*\}$ . Therefore,

$$\begin{aligned} & |V^*(\mathbf{r}, M) - V_\tau(\mathbf{r}, M)| \\ & = \left| \sum_{i=1}^N v_i(i^*, \omega^*, \rho^*) - \sum_{i=1}^N v_i(i_\tau, \omega_\tau, \rho_\tau) \right| \\ & \leq \sum_{i \in \Theta} \left| v_i(i^*, \omega^*, \rho^*) - v_i(i_\tau, \omega_\tau, \rho_\tau) \right| \\ & \quad + \sum_{i \notin \Theta} \left| v_i(i^*, \omega^*, \rho^*) - v_i(i_\tau, \omega_\tau, \rho_\tau) \right|. \quad (28) \end{aligned}$$

Before bounding (28), we first show that, for this case, we have only one of the three cases:  $\omega_\tau > \omega^*$ , or  $\omega_\tau = \omega^*$  with  $\rho_\tau < \rho^*$  and  $i_\tau = i^*$ , or  $\omega_\tau = \omega^*$  with  $i_\tau > i^*$ .

We prove the above statement by first showing that  $\sum_{i=1}^N \alpha_i^{\mathbf{r}}(i^*, \omega^*, \rho^*) \geq \sum_{i=1}^N \alpha_i(i_\tau, \omega_\tau, \rho_\tau) = M$ : For any user  $i \notin \Theta$ , we have  $\alpha_i^{\mathbf{r}}(i^*, \omega^*, \rho^*) = \alpha_i(i^*, \omega^*, \rho^*)$  since  $(i^*, \omega^*, \rho^*)$  does not exceed the truncation level. For user  $i \in \Theta$ , 1) if  $\omega^* \geq W_i^{\mathbf{r}}(b_s^i)$ , we have  $\alpha_i^{\mathbf{r}}(i^*, \omega^*, \rho^*) \geq \alpha_i(i^*, \omega^*, \rho^*)$  since  $\alpha_i(i^*, \omega^*, \rho^*) = 0$ . 2) If  $W_i^{\mathbf{r}}(b_{0,\tau}^i) < \omega^* < W_i^{\mathbf{r}}(b_s^i)$  for  $i \in \Theta$ , we have

$$\begin{aligned} \alpha_i^{\mathbf{r}}(i^*, \omega^*, \rho^*) & = \alpha_i^{\mathbf{r}}(i^*, W_i^{\mathbf{r}}(b_s^i), 1) = \frac{1 - p_{11}^i + b_s^i}{(1 + \tau)(1 - p_{11}^i) + b_s^i} \\ & > \frac{1 - p_{11}^i + b_{0,\tau+1}^i}{(1 + \tau)(1 - p_{11}^i) + b_{0,\tau+1}^i} \\ & = \alpha_i(i^*, W_i^{\mathbf{r}}(b_{0,\tau}^i), 0) \geq \alpha_i(i^*, \omega^*, \rho^*), \end{aligned}$$

where the first equality holds because, when  $W_i^{\mathbf{r}}(b_{0,\tau}^i) < \omega^* < W_i^{\mathbf{r}}(b_s^i)$ , the user is scheduled when its belief value is not below  $b_s^i$  and stays idle otherwise. Because of the truncation, the next belief value above  $b_{0,\tau}^i$  is  $b_s^i$ . Since user  $i$ 's index value will not be exactly  $\omega^*$ , the randomization factor  $\rho^*$  at the threshold does not play a role. Hence the expected fraction of transmission time  $\alpha_i^{\mathbf{r}}(i^*, \omega^*, \rho^*)$  equals  $\alpha_i^{\mathbf{r}}(i^*, W_i^{\mathbf{r}}(b_s^i), 1)$ , i.e., transmit to user  $i$  when its belief value

is not below  $b_s^i$  with probability 1. The second and the third equality are from lemma 2(i) and (22), respectively. The first inequality holds since  $b_s^i > b_{0,\tau+1}^i$ . The last inequality holds because  $W_i^{\mathbf{r}}(b_{0,\tau}^i) < \omega^* < W_i^{\mathbf{r}}(b_s^i)$ , hence from (22) and the monotonicity property in Lemma 5(i),

$$\begin{aligned} & \alpha_i(i^*, W_i^{\mathbf{r}}(b_{0,\tau}^i), 0) = \alpha_i(i^*, W_i^{\mathbf{r}}(b_{0,\tau+1}^i), 1) \\ & = \alpha_i(i^*, \omega^*, \rho^*) \quad \text{if } W_i^{\mathbf{r}}(b_{0,\tau}^i) < \omega^* < W_i^{\mathbf{r}}(b_{0,\tau+1}^i), \\ & \alpha_i(i^*, W_i^{\mathbf{r}}(b_{0,\tau}^i), 0) = \alpha_i(i^*, W_i^{\mathbf{r}}(b_{0,\tau+1}^i), 1) \\ & \geq \alpha_i(i^*, \omega^*, 1) \geq \alpha_i(i^*, \omega^*, \rho^*) \quad \text{if } W_i^{\mathbf{r}}(b_{0,\tau+1}^i) \leq \omega^* < W_i^{\mathbf{r}}(b_s^i). \end{aligned}$$

3) If  $\omega^* = W_i^{\mathbf{r}}(b_{0,\tau}^i)$ , similarly, for  $i \in \Theta$ ,

$$\alpha_i^{\mathbf{r}}(i^*, \omega^*, \rho^*) > \alpha_i(i^*, \omega^*, \rho^*).$$

Hence from 1)-3) we have  $\alpha_i^{\mathbf{r}}(i^*, \omega^*, \rho^*) \geq \alpha_i(i^*, \omega^*, \rho^*)$  for  $i \in \Theta$ . Also noting that, for  $i \notin \Theta$ ,  $\alpha_i(i^*, \omega^*, \rho^*) = \alpha_i^{\mathbf{r}}(i^*, \omega^*, \rho^*)$ , we hence have

$$\begin{aligned} & \sum_{i=1}^N \alpha_i^{\mathbf{r}}(i^*, \omega^*, \rho^*) = \sum_{i \in \Theta} \alpha_i^{\mathbf{r}}(i^*, \omega^*, \rho^*) + \sum_{i \notin \Theta} \alpha_i^{\mathbf{r}}(i^*, \omega^*, \rho^*) \\ & = \sum_{i \in \Theta} \alpha_i^{\mathbf{r}}(i^*, \omega^*, \rho^*) + \sum_{i \notin \Theta} \alpha_i(i^*, \omega^*, \rho^*) \\ & \geq \sum_{i \in \Theta} \alpha_i(i^*, \omega^*, \rho^*) + \sum_{i \notin \Theta} \alpha_i(i^*, \omega^*, \rho^*) \\ & = \sum_{i=1}^N \alpha_i(i^*, \omega^*, \rho^*) = M. \end{aligned}$$

Hence if we implement the policy with threshold parameters  $(i^*, \omega^*, \rho^*)$  over the fictitious truncated belief space, the expected number of transmissions will equal to or exceed the constraint. Therefore, from the monotonicity property in Lemma 2, to ensure the constraint (10) on the long-term expected number of transmissions over the truncated state space, it must be one of the following three cases  $\omega_\tau > \omega^*$ , or  $\omega_\tau = \omega^*$  with  $\rho_\tau < \rho^*$  and  $i_\tau = i^*$ , or  $\omega_\tau = \omega^*$  with  $i_\tau > i^*$ . From this property as well as Lemma 5(i), we have,

$$\alpha_i(i_\tau, \omega_\tau, \rho_\tau) \leq \alpha_i(i^*, \omega^*, \rho^*) \quad \text{for all } i, \quad (29)$$

and, because  $i \in \Theta$ ,

$$v_i(i_\tau, \omega_\tau, \rho_\tau) \leq v_i(i^*, \omega^*, \rho^*) \leq v_i(i, W_i^{\mathbf{r}}(b_{0,\tau}^i), 1), \quad \text{for } i \in \Theta. \quad (30)$$

Hence, for  $i \in \Theta$ ,

$$\begin{aligned} & \left| v_i(i^*, \omega^*, \rho^*) - v_i(i_\tau, \omega_\tau, \rho_\tau) \right| \leq v_i(i, W_i^{\mathbf{r}}(b_{0,\tau}^i), 1) \\ & \leq r_i \cdot \alpha_i(i, W_i^{\mathbf{r}}(b_{0,\tau}^i), 1), \quad (31) \end{aligned}$$

where the first inequality is from (30) and the last equality holds because instantaneous reward is upper bounded by  $r_i$ .

Similar to (30), from the monotonicity properties of  $\alpha_i^{\mathbf{r}}(j, \omega, \rho)$  and  $\alpha_i(j, \omega, \rho)$  and because  $i \in \Theta$ ,

$$\alpha_i^{\mathbf{r}}(i_\tau, \omega_\tau, \rho_\tau) \leq \alpha_i^{\mathbf{r}}(i^*, \omega^*, \rho^*) \leq \alpha_i^{\mathbf{r}}(i, W_i^{\mathbf{r}}(b_{0,\tau}^i), 1), \quad i \in \Theta, \quad (32)$$

$$\alpha_i(i^*, \omega^*, \rho^*) \leq \alpha_i(i^*, \omega^*, 1) \leq \alpha_i(i, W_i^{\mathbf{r}}(b_{0,\tau}^i), 1), \quad i \in \Theta. \quad (33)$$

For  $i \notin \Theta$ , we have  $\alpha_i^\tau(i^*, \omega^*, \rho^*) = \alpha_i(i^*, \omega^*, \rho^*)$ . Hence,

$$\begin{aligned}
& \sum_{i \notin \Theta} |v_i(i^*, \omega^*, \rho^*) - v_i(i_\tau, \omega_\tau, \rho_\tau)| \\
& \leq \sum_{i \notin \Theta} r_i \cdot |\alpha_i(i^*, \omega^*, \rho^*) - \alpha_i(i_\tau, \omega_\tau, \rho_\tau)| \\
& = \sum_{i \notin \Theta} r_i \cdot [\alpha_i(i^*, \omega^*, \rho^*) - \alpha_i(i_\tau, \omega_\tau, \rho_\tau)] \\
& \leq \sum_{i \notin \Theta} r_i \cdot \sum_{i \notin \Theta} [\alpha_i(i^*, \omega^*, \rho^*) - \alpha_i(i_\tau, \omega_\tau, \rho_\tau)] \\
& \leq \sum_{i \notin \Theta} r_i \cdot \left[ \sum_{i \notin \Theta} [\alpha_i(i^*, \omega^*, \rho^*) - \alpha_i^\tau(i_\tau, \omega_\tau, \rho_\tau)] + \right. \\
& \quad \left. \sum_{i \notin \Theta} [\alpha_i^\tau(i_\tau, \omega_\tau, \rho_\tau) - \alpha_i(i_\tau, \omega_\tau, \rho_\tau)] \right], \quad (34)
\end{aligned}$$

where the first inequality is from Lemma 5(ii) and the first equality holds from (29).

Consider the first summand inside the parenthesis of (34). Since  $\sum_{i=1}^N \alpha_i(i^*, \omega^*, \rho^*) = \sum_{i=1}^N \alpha_i^\tau(i_\tau, \omega_\tau, \rho_\tau) = M$ , subtracting both sides by  $\sum_{i \notin \Theta} \alpha_i^\tau(i_\tau, \omega_\tau, \rho_\tau) + \sum_{i \in \Theta} \alpha_i(i^*, \omega^*, \rho^*)$  we have

$$\begin{aligned}
& \sum_{i \notin \Theta} [\alpha_i(i^*, \omega^*, \rho^*) - \alpha_i^\tau(i_\tau, \omega_\tau, \rho_\tau)] \\
& = \sum_{i \in \Theta} [\alpha_i^\tau(i_\tau, \omega_\tau, \rho_\tau) - \alpha_i(i^*, \omega^*, \rho^*)] \\
& \leq \sum_{i \in \Theta} |\alpha_i^\tau(i_\tau, \omega_\tau, \rho_\tau) - \alpha_i(i^*, \omega^*, \rho^*)|. \quad (35)
\end{aligned}$$

Note that, for  $i \in \Theta$ , from (32)-(33),

$$|\alpha_i^\tau(i_\tau, \omega_\tau, \rho_\tau) - \alpha_i(i^*, \omega^*, \rho^*)| \leq \alpha_i(i, W_i^\Gamma(b_{0,\tau}^i), 1). \quad (36)$$

Substituting (36) back to (35), we have

$$\sum_{i \notin \Theta} [\alpha_i(i^*, \omega^*, \rho^*) - \alpha_i^\tau(i_\tau, \omega_\tau, \rho_\tau)] \leq \sum_{i \in \Theta} \alpha_i(i, W_i^\Gamma(b_{0,\tau}^i), 1). \quad (37)$$

Now consider the second summand inside (34), we have, for  $i \notin \Theta$ ,

$$\alpha_i^\tau(i_\tau, \omega_\tau, \rho_\tau) - \alpha_i(i_\tau, \omega_\tau, \rho_\tau) = 0, \text{ if } \omega_\tau < W_i^\Gamma(b_{0,\tau}^i), \quad (38)$$

$$\begin{aligned} & \alpha_i^\tau(i_\tau, \omega_\tau, \rho_\tau) - \alpha_i(i_\tau, \omega_\tau, \rho_\tau) \\ & \leq \alpha_i(i, W_i^\Gamma(b_{0,\tau}^i), 1), \text{ if } \omega_\tau = W_i^\Gamma(b_{0,\tau}^i), \end{aligned} \quad (39)$$

where (39) holds because both  $\alpha_i^\tau(i_\tau, \omega_\tau, \rho_\tau) \leq \alpha_i(i, W_i^\Gamma(b_{0,\tau}^i), 1)$  and  $\alpha_i(i_\tau, \omega_\tau, \rho_\tau) \leq \alpha_i(i, W_i^\Gamma(b_{0,\tau}^i), 1)$ . Therefore,

$$\begin{aligned}
& \sum_{i \notin \Theta} [\alpha_i^\tau(i_\tau, \omega_\tau, \rho_\tau) - \alpha_i(i_\tau, \omega_\tau, \rho_\tau)] \\
& \leq \sum_{i \notin \Theta} \alpha_i(i, W_i^\Gamma(b_{0,\tau}^i), 1). \quad (40)
\end{aligned}$$

Substituting (37) and (40) in (34),

$$\sum_{i \notin \Theta} |v_i(i^*, \omega^*, \rho^*) - v_i(i_\tau, \omega_\tau, \rho_\tau)| \leq \sum_{i \notin \Theta} r_i \sum_{i=1}^N \alpha_i(i, W_i^\Gamma(b_{0,\tau}^i), 1). \quad (41)$$

From (31) and (41), the difference in (28) can be bounded as follows,

$$\begin{aligned}
& |V^*(\mathbf{r}, M) - V_\tau^*(\mathbf{r}, M)| \\
& \leq \sum_{i \in \Theta} r_i \cdot \alpha_i(i, W_i^\Gamma(b_{0,\tau}^i), 1) + \sum_{i \notin \Theta} r_i \sum_{i=1}^N \alpha_i(i, W_i^\Gamma(b_{0,\tau}^i), 1) \\
& \leq \sum_{i=1}^N r_i \cdot \sum_{i=1}^N \alpha_i(i, W_i^\Gamma(b_{0,\tau}^i), 1).
\end{aligned}$$

We let  $f_i(\tau) = \alpha_i(i, W_i^\Gamma(b_{0,\tau}^i), 1)$  and  $f(\tau) = \sum_{i=1}^N f_i(\tau)$ . Since  $\alpha_i(i, W_i^\Gamma(b_{0,\tau}^i), 1) \rightarrow 0$  as  $\tau \rightarrow \infty$ , part (i) of the lemma is established. From (29), we have

$$Z_\tau(\mathbf{q}, M) = \sum_{i=1}^N \alpha_i(i_\tau, \omega_\tau, \rho_\tau) \leq \sum_{i=1}^N \alpha_i(i^*, \omega^*, \rho^*) = M,$$

which proves part (ii).  $\blacksquare$

## APPENDIX D PROOF OF PROPOSITION 1

Define Lyapunov function  $L(\mathbf{q}) = \frac{1}{2} \sum_{i=1}^N q_i^2$ . We consider the  $T$ -frame average Lyapunov drift  $\Delta L(\mathbf{q}[kT])$  over the  $k$ -th frame, expressed as,

$$\begin{aligned}
& \Delta L(\mathbf{q}[kT])/T \\
& = \frac{1}{T} \mathbb{E} \left[ L(\mathbf{q}[(k+1)T]) - L(\mathbf{q}[kT]) \mid \mathbf{q}[kT], \boldsymbol{\pi}[kT] \right] \\
& \leq BT + \sum_{i=1}^N q_i[kT] \cdot \lambda_i - \sum_{i=1}^N q_i[kT] \cdot \frac{1}{T} \\
& \quad \cdot \mathbb{E} \left[ \sum_{t=0}^{T-1} \pi_i[kT+t] \cdot a_i^{\phi_\tau}(\mathbf{q}[kT], M-g(\tau)/2)[kT+t] \mid \boldsymbol{\pi}[kT] \right], \quad (42)
\end{aligned}$$

where  $B$  is a constant whose value is determined by the second moment of the arrival process [37]. Because  $\boldsymbol{\lambda} + g(\tau)\mathbf{1} \in \Gamma$ , for any non-negative vector  $\mathbf{q}$ , we have

$$\sum_{i=1}^N q_i \cdot (\lambda_i + g(\tau)) \leq V^*(\mathbf{q}, M),$$

where  $V^*(\mathbf{q}, M)$  is defined in (11). The Lyapunov drift (42) now becomes,

$$\begin{aligned}
& \Delta L(\mathbf{q}[kT])/T \leq BT - g(\tau) \sum_{i=1}^N q_i[kT] + \\
& \quad V^*(\mathbf{q}[kT], M) - V_\tau^T(\mathbf{q}[kT], M - g(\tau)/2)
\end{aligned}$$



$$\begin{aligned}
&= BT - g(\tau) \sum_{i=1}^N q_i [kT] + V^*(\mathbf{q}[kT], M) - V_\tau(\mathbf{q}[kT], M) \\
&+ V_\tau(\mathbf{q}[kT], M) - V_\tau(\mathbf{q}[kT], M - g(\tau)/2) \\
&+ V_\tau(\mathbf{q}[kT], M - g(\tau)/2) - V_\tau^T(\mathbf{q}[kT], M - g(\tau)/2). \quad (43)
\end{aligned}$$

where  $V_\tau(\mathbf{q}[kT], M)$  is defined in (12), and  $V_\tau^T(\mathbf{q}[kT], M)$  is the  $T$ -horizon expected transmission rate achieved under the policy  $\phi_\tau(\mathbf{q}[kT], M)$ , i.e.,

$$\begin{aligned}
&V_\tau^T(\mathbf{q}[kT], M) \\
&= \sum_{i=1}^N q_i [kT] \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \pi_i [kT+t] \cdot a_i^{\phi_\tau(\mathbf{q}[kT], M)} [kT+t] \middle| \boldsymbol{\pi} [kT] \right].
\end{aligned}$$

Note that, in (43), the difference  $V^*(\mathbf{q}[kT], M) - V_\tau(\mathbf{q}[kT], M)$  is bounded in Lemma 3. We proceed to bound the rest of the terms in (43). Specifically, the difference  $V_\tau(\mathbf{q}[kT], M - g(\tau)/2) - V_\tau^T(\mathbf{q}[kT], M - g(\tau)/2)$  is bounded in Lemma 6, and the difference  $V_\tau(\mathbf{q}[kT], M) - V_\tau(\mathbf{q}[kT], M - g(\tau)/2)$  is bounded in Lemma 7. These bounds help us to bound the Lyapunov drift  $\Delta L(\mathbf{q}[kT])/T$  and later to establish the proof using Lyapunov stability theory.

We denote  $Z_\tau^T(\mathbf{q}, M)$  as the finite  $T$ -horizon expected number of transmissions, under the policy  $\phi_\tau(\mathbf{q}[kT], M)$ , i.e.,

$$Z_\tau^T(\mathbf{q}, M) = \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{i=1}^N a_i^{\phi_\tau(\mathbf{q}, M)} [t] \right].$$

The next lemma states that, as the length of the time horizon tends to infinity, the expected achieved rate in finite horizon asymptotically converges to infinite horizon achievable rate, and the expected number of transmissions converges to the value  $M$ .

**Lemma 6.** *For any  $M$  and  $\kappa > 0$ , we have, uniformly over  $\mathbf{q}$ ,  $M$ , and the initial state  $\boldsymbol{\pi} [kT]$ ,*

(a) *there exist positive constants  $c_1$  and  $c_2$  such that*

$$\left| V_\tau(\mathbf{q}, M) - V_\tau^T(\mathbf{q}, M) \right| < (\kappa + c_1 \exp(-c_2 T)) \sum_{i=1}^N q_i.$$

(b) *there exist positive constants  $d_1$  and  $d_2$  such that*

$$\left| Z_\tau^T(\mathbf{q}, M) - M \right| < (\kappa + d_1 \exp(-d_2 T)).$$

**Proof:** We first prove part (a). We define the random variable  $\mu_\tau^T(\mathbf{q}, M)$  as

$$\mu_\tau^T(\mathbf{q}, M) = \sum_{i=1}^N q_i \frac{1}{T} \sum_{t=0}^{T-1} \pi_i [kT+t] \cdot a_i^{\phi_\tau(\mathbf{q}, M)} [kT+t].$$

Therefore,  $V_\tau^T(\mathbf{q}, M) = \mathbb{E}[\mu_\tau^T(\mathbf{q}, M)]$ . We denote event

$\Omega := \{ |\mu_\tau^T(\mathbf{q}, M) - V_\tau(\mathbf{q}, M)| \leq \kappa \sum_{i=1}^N q_i \}$ , then

$$\begin{aligned}
&\mathbb{E} \left[ \left| \mu_\tau^T(\mathbf{q}, M) - V_\tau(\mathbf{q}, M) \right| \right] \\
&\leq \mathbb{E} \left[ \left| \mu_\tau^T(\mathbf{q}, M) - V_\tau(\mathbf{q}, M) \right| \middle| \Omega \right] \cdot \Pr(\Omega) \\
&\quad + \mathbb{E} \left[ \left| \mu_\tau^T(\mathbf{q}, M) - V_\tau(\mathbf{q}, M) \right| \middle| \bar{\Omega} \right] \cdot \Pr(\bar{\Omega}) \\
&\leq \kappa \sum_{i=1}^N q_i + \sum_{i=1}^N q_i \cdot \Pr \left( \left| \mu_\tau^T(\mathbf{q}, M) - V_\tau(\mathbf{q}, M) \right| > \kappa \sum_{i=1}^N q_i \right). \quad (44)
\end{aligned}$$

Note that

$$\begin{aligned}
&\left| \mu_\tau^T(\mathbf{q}, M) - V_\tau(\mathbf{q}, M) \right| \\
&= \left| \sum_{i=1}^N q_i \cdot \left[ \frac{1}{T} \sum_{t=0}^{T-1} \pi_i [kT+t] \cdot a_i^{\phi_\tau(\mathbf{q}, M)} [kT+t] \right. \right. \\
&\quad \left. \left. - \lim_{\mathbb{T} \rightarrow \infty} \frac{1}{\mathbb{T}} \sum_{t=0}^{\mathbb{T}-1} \pi_i [kT+t] \cdot a_i^{\phi_\tau(\mathbf{q}, M)} [kT+t] \right] \right| \\
&\leq \sum_{i=1}^N q_i \cdot \left[ \sum_{i=1}^N \left[ \frac{1}{T} \sum_{t=0}^{T-1} \pi_i [kT+t] \cdot a_i^{\phi_\tau(\mathbf{q}, M)} [kT+t] \right. \right. \\
&\quad \left. \left. - \lim_{\mathbb{T} \rightarrow \infty} \frac{1}{\mathbb{T}} \sum_{t=0}^{\mathbb{T}-1} \pi_i [kT+t] \cdot a_i^{\phi_\tau(\mathbf{q}, M)} [kT+t] \right]^2 \right]^{\frac{1}{2}} \\
&:= \sum_{i=1}^N q_i \cdot \left\| \boldsymbol{\eta}^\tau(\mathbf{q}, M) - \boldsymbol{\eta}_T^\tau(\mathbf{q}, M) \right\|.
\end{aligned}$$

where the inequality follows from Cauchy-Schwarz inequality and  $\boldsymbol{\eta}^\tau(\mathbf{q}, M)$  and  $\boldsymbol{\eta}_T^\tau(\mathbf{q}, M)$  are vectors with

$$\boldsymbol{\eta}_i^\tau(\mathbf{q}, M) = \lim_{\mathbb{T} \rightarrow \infty} \frac{1}{\mathbb{T}} \sum_{t=0}^{\mathbb{T}-1} \pi_i [kT+t] \cdot a_i^{\phi_\tau(\mathbf{q}, M)} [kT+t], \quad (45)$$

$$\boldsymbol{\eta}_{T,i}^\tau(\mathbf{q}, M) = \frac{1}{T} \sum_{t=0}^{T-1} \pi_i [t] \cdot a_i^{\phi_\tau(\mathbf{q}, M)} [kT+t]. \quad (46)$$

Therefore,

$$\begin{aligned}
&\Pr \left( \left| \mu_\tau^T(\mathbf{q}, M) - V_\tau(\mathbf{q}, M) \right| > \kappa \sum_{i=1}^N q_i \right) \\
&\leq \Pr \left( \left\| \boldsymbol{\eta}^\tau(\mathbf{q}, M) - \boldsymbol{\eta}_T^\tau(\mathbf{q}, M) \right\| > \kappa \right) \\
&\leq \Pr \left( \bigcup_{i=1}^N \left\{ \left| \boldsymbol{\eta}_{T,i}^\tau(\mathbf{q}, M) - \boldsymbol{\eta}_i^\tau(\mathbf{q}, M) \right| > \kappa/N \right\} \right) \\
&\leq \sum_{i=1}^N \Pr \left( \left| \boldsymbol{\eta}_{T,i}^\tau(\mathbf{q}, M) - \boldsymbol{\eta}_i^\tau(\mathbf{q}, M) \right| > \kappa/N \right). \quad (47)
\end{aligned}$$

Recall that, under the policy  $\phi^\tau(\mathbf{q}, M)$ , the belief states of different users, i.e.,  $\{\mathcal{B}_i^\tau, i = 1, \dots, N\}$ , are sorted, in the initialization phase given by algorithm  $G^\tau(\mathbf{q}, M)$ , in the vector  $\mathbf{w}$  according to their  $\mathbf{q}$ -weighted index values. Consider another vector  $\boldsymbol{\varsigma}$  where each element  $\varsigma_i$  corresponds to the unique belief state the  $i^{\text{th}}$  element  $w_i$  represents. So each weighing vector  $\mathbf{q}$  corresponds to a vector  $\mathbf{w}$  and hence  $\boldsymbol{\varsigma}$ . Note that, the activation/passive scheduling decision to a user depends on the the location of the threshold for transmission, i.e., above which belief value the user is scheduled and with how much randomization. From the implementation of algorithm  $G^\tau(\mathbf{q}, M)$ , as long as different policies correspond

to the same  $\varsigma$ , for each user, the transmission/idle action (at each belief state) is the same function of belief state, and hence the belief state of each user evolves as the same finite-state space ergodic Markov chain. Therefore, for a policy, denoted by  $\phi^\varsigma$ , that corresponds to a vector  $\varsigma$ , there exist constants  $c_1^\varsigma$  and  $c_2^\varsigma$  such that, for each user  $i$  uniform over the initial belief state and  $\mathbf{q}$  [39],

$$\Pr \left( \left| \frac{1}{T} \sum_{t=0}^{T-1} \pi_i[t] \cdot a_i^{\phi^\varsigma}[t] - \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \pi_i[t] \cdot a_i^{\phi^\varsigma}[t] \right| > \kappa/N \right) < c_1^{\phi^\varsigma} \exp(-c_2^{\phi^\varsigma} T). \quad (48)$$

Note that the number of users, as well as the number of vectors  $\varsigma$ , are finite. From (45)-(48), there exist constants  $c_1$  and  $c_2$  such that, regardless of  $\mathbf{q}$  and the initial belief state,

$$\Pr \left( \left| \mu_\tau^T(\mathbf{q}, M) - V_\tau(\mathbf{q}, M) \right| > \kappa \sum_{i=1}^N q_i \right) < c_1 \exp(-c_2 T).$$

Substituting the above inequality in (44), part(a) thus holds.

The proof of part (b) follows a similar approach as part (a). Here, the immediate reward is  $a_i^{\phi^\tau(\mathbf{q}, M)}[kT + t]$  instead of  $\pi_i[kT + t] \cdot a_i^{\phi^\tau(\mathbf{q}, M)}[kT + t]$ . ■

**Lemma 7.** *When  $\tau > \tau_0$ , for any  $\epsilon > 0$ , the difference between the expected transmission rate achieved under policy  $\phi_\tau(\mathbf{q}, M)$  and  $\phi_\tau(\mathbf{q}, M - \epsilon)$  satisfies the following bound,*

$$\left| V_\tau(\mathbf{q}, M) - V_\tau(\mathbf{q}, M - \epsilon) \right| \leq \epsilon \sum_{i=1}^N q_i.$$

**Proof:** Suppose, under the weight  $\mathbf{q}$ , the policies  $\phi_\tau(\mathbf{q}, M)$  and  $\phi_\tau(\mathbf{q}, M - \epsilon)$  correspond to parameter set  $\{i_M^\tau, \omega_M^\tau, \rho_M^\tau\}$  and  $\{i_{M-\epsilon}^\tau, \omega_{M-\epsilon}^\tau, \rho_{M-\epsilon}^\tau\}$ , respectively. For user  $i$ , we let  $y_i(\epsilon)$  denote be the difference between activation time under policy  $\phi_\tau(\mathbf{q}, M - \epsilon)$  and  $\phi_\tau(\mathbf{q}, M)$ , i.e.,  $y_i(\epsilon) = \alpha_i(i_M^\tau, \omega_M^\tau, \rho_M^\tau) - \alpha_i(i_{M-\epsilon}^\tau, \omega_{M-\epsilon}^\tau, \rho_{M-\epsilon}^\tau)$ , where, recall that,  $\alpha_i(j, \omega, \rho)$  is defined in (22). From Lemma 5(i), we have  $y_i(\epsilon) \geq 0, \forall i$ . Since the difference of the total expected number of transmissions between the two policies is  $\epsilon$ , we have  $\sum_{i=1}^N y_i(\epsilon) = \epsilon$ . From Lemma 5(ii), we have,

$$\begin{aligned} & \left| V_\tau(\mathbf{q}, M) - V_\tau(\mathbf{q}, M - \epsilon) \right| \\ &= \left| \sum_{i=1}^N v_i(i_M^\tau, \omega_M^\tau, \rho_M^\tau) - \sum_{i=1}^N v_i(i_{M-\epsilon}^\tau, \omega_{M-\epsilon}^\tau, \rho_{M-\epsilon}^\tau) \right| \\ &\leq \sum_{i=1}^N \left| v_i(i_M^\tau, \omega_M^\tau, \rho_M^\tau) - v_i(i_{M-\epsilon}^\tau, \omega_{M-\epsilon}^\tau, \rho_{M-\epsilon}^\tau) \right| \\ &\leq \sum_{i=1}^N q_i \cdot \left| \alpha_i(i_M^\tau, \omega_M^\tau, \rho_M^\tau) - \alpha_i(i_{M-\epsilon}^\tau, \omega_{M-\epsilon}^\tau, \rho_{M-\epsilon}^\tau) \right| \\ &= \sum_{i=1}^N q_i \cdot y_i(\epsilon) \leq \sum_{i=1}^N q_i \left[ \sum_{j=1}^N y_j(\epsilon) \right] = \epsilon \sum_{i=1}^N q_i. \end{aligned}$$

We hence have proved the lemma. ■

From Lemma 3 and Lemma 6-7, the Lyapunov drift (43)

can be further bounded as follows,

$$\begin{aligned} & \Delta L(\mathbf{q}[kT])/T \\ &\leq BT + \\ & \quad \left[ -g(\tau) + f(\tau) + \frac{g(\tau)}{2} + (\kappa + c_1 \exp(-c_2 T)) \right] \cdot \sum_{i=1}^N q_i[kT] \\ &= BT + \left[ -\frac{g(\tau)}{2} + f(\tau) + (\kappa + c_1 \exp(-c_2 T)) \right] \sum_{i=1}^N q_i[kT] \\ &= BT + \left[ -f(\tau)/2 + [\kappa + c_1 \exp(-c_2 T)] \right] \sum_{i=1}^N q_i[kT] \end{aligned} \quad (49)$$

where the last equality holds because we let  $g(\tau) = 3f(\tau)$ . For fixed  $\tau$ , by choosing  $\kappa$  sufficiently small and  $T$  sufficiently large, say  $T > T_1$ , the Lyapunov drift is negative whenever the sum of the queue lengths gets sufficiently large. Therefore, the queues are stable according to the Foster-Lyapunov criterion.

Note that, under the policy Q-Index $_\tau(T, M - g(\tau)/2)$ , the expected number of transmissions in the  $k$ -th frame,  $Z_\tau^T(\mathbf{q}[kT], M - g(\tau)/2)$ , is bounded by Lemma 6 as,

$$\left| Z_\tau^T(\mathbf{q}[kT], M - g(\tau)/2) - (M - g(\tau)/2) \right| < (\kappa + d_1 \exp(-d_2 T)),$$

for some constant  $d_1$  and  $d_2$ . Therefore, there exists  $T_2$  such that  $Z_\tau^T(\mathbf{q}[kT], M - g(\tau)/2) < M$  for  $T > T_2$ . Hence, the long term constraint on the average number of transmissions is satisfied. From Lemma 3, we have  $\lim_{\tau \rightarrow \infty} g(\tau) = 0$ . Letting  $T' = \max\{T_1, T_2\}$ , the proposition is then established.

## APPENDIX E PROOF OF LEMMA 4

(i) We first prove part (i) of the lemma with  $i = j$ .

Case (1). If  $\pi_j = b_{0,h}^j$  and  $h < \tau$ , we consider the reciprocal of  $v_j^\tau(W_j^\tau(b_{0,h}^j), \rho)$ ,

$$\begin{aligned} & r_j \cdot [v_j^\tau(j, W_j^\tau(b_{0,h}^j), \rho)]^{-1} = 1 + \frac{(1 - p_{11}^j)(h + 1 - \rho)}{\rho(b_{0,h}^j - b_{0,h+1}^j) + b_{0,h+1}^j} \\ &= 1 + \frac{1 - p_{11}^j}{b_{0,h+1}^j - b_{0,h}^j} \left[ 1 + \frac{b_{0,h+1}^j - (h+1)(b_{0,h+1}^j - b_{0,h}^j)}{\rho(b_{0,h+1}^j - b_{0,h}^j) - b_{0,h+1}^j} \right] \end{aligned} \quad (50)$$

Consider the numerator in the parenthesis of (50)

$$\begin{aligned} & b_{0,h+1}^j - (h+1)(b_{0,h+1}^j - b_{0,h}^j) = (h+1)b_{0,h}^j - hb_{0,h+1}^j \\ &= [1 + (1 - p_{11}^j + p_{01}^j)h]b_{0,h}^j - hp_{01}^j \\ &= \frac{p_{01}^j[1 - (p_{11}^j - p_{01}^j)^h]}{1 - (p_{11}^j - p_{01}^j)} - hp_{01}^j(p_{11}^j - p_{01}^j)^h \\ &= p_{01}^j[1 + (p_{11}^j - p_{01}^j) + \dots + (p_{11}^j - p_{01}^j)^{h-1}] - hp_{01}^j(p_{11}^j - p_{01}^j)^h \\ &> hp_{01}^j(p_{11}^j - p_{01}^j)^h - hp_{01}^j(p_{11}^j - p_{01}^j)^h = 0. \end{aligned} \quad (51)$$

Since the denominator in the parenthesis of (50) strictly increases with  $\rho$ ,  $[v_j^\tau(j, W_j^\tau(b_{0,h}^j), \rho)]^{-1}$  strictly decreases with  $\rho$  and hence  $v_j^\tau(W_j^\tau(j, b_{0,h}^j), \rho)$  strictly increases with  $\rho$  in this case.

Case (2). If  $\pi_j = b_{0,\tau}^j$ , we have

$$\begin{aligned} r_j \cdot [v_j^\tau(j, W_j^\tau(b_{0,\tau}^j), \rho)]^{-1} &= 1 + \frac{(1-p_{11}^j)(\tau+1-\rho)}{\rho(b_{0,\tau}^j - b_s^j) + b_s^j} \\ &= 1 + \frac{1-p_{11}^j}{b_s^j - b_{0,\tau}^j} \left[ 1 + \frac{b_s^j - (\tau+1)(b_s^j - b_{0,\tau}^j)}{\rho(b_s^j - b_{0,\tau}^j) - b_s^j} \right]. \end{aligned} \quad (52)$$

When  $\tau > \tau_0$ , it can be derived that the numerator  $b_s^j - (\tau+1)(b_s^j - b_{0,\tau}^j)$  inside (52) is positive. Therefore,  $v_j^\tau(j, W_j^\tau(b_{0,\tau}^j), \rho)$  strictly increases with  $\rho$  in this case.

Case (3). If  $\pi_j = b_s^j$ ,

$$r_j \cdot [v_j^\tau(j, W_j^\tau(b_s^j), \rho)]^{-1} = \frac{1}{b_s^j} \left( \tau(1-p_{11}^j) + \frac{1-p_{11}^j}{\rho} + b_s^j \right)$$

It is then clear from the above expression that  $v_j^\tau(j, W_j^\tau(b_s^j), \rho)$  strictly increases with  $\rho$  in this case.

Now consider fixed  $\rho$ . For  $v_j^\tau(W_j^\tau(b_{0,h}^j), \rho)$  and  $v_j^\tau(j, W_j^\tau(b_{0,h+1}^j), \rho)$  with  $h+1 \leq \tau$ , we have

$$\begin{aligned} v_j^\tau(j, W_j^\tau(b_{0,h}^j), \rho) &\geq v_j^\tau(j, W_j^\tau(b_{0,h}^j), 0) = v_j^\tau(j, W_j^\tau(b_{0,h+1}^j), 1) \\ &\geq v_j^\tau(j, W_j^\tau(b_{0,h+1}^j), \rho), \end{aligned}$$

where the first and last inequality is from case (1) we have just proven. The first equality is from expression (27). Since  $v_j^\tau(j, W_j^\tau(b_{0,h}^j), \rho) = v_j^\tau(j, W_j^\tau(b_{0,h}^j), 0)$  only if  $\rho = 0$ , and  $v_j^\tau(j, W_j^\tau(b_{0,h+1}^j), 1) = v_j^\tau(j, W_j^\tau(b_{0,h+1}^j), \rho)$  only if  $\rho = 1$ . We hence have  $v_j^\tau(j, W_j^\tau(b_{0,h}^j), \rho) > v_j^\tau(j, W_j^\tau(b_{0,h+1}^j), \rho)$  strictly. Following a similar derivation, we have  $v_j^\tau(j, W_j^\tau(b_{0,\tau}^j), \rho) > v_j^\tau(j, W_j^\tau(b_s^j), \rho)$ . Therefore the monotonicity property in part (i) holds for user  $j$  with randomized transmission. The monotonicity result easily extends to user  $i \neq j$  where there is no longer randomization in scheduling user  $i$ .

(ii) We proceed to prove part (ii) by first establishing the statement when  $j_1 = j_2 = j$ ,  $\omega_1 = \omega_2 = \omega$ .

Case (1). If  $\omega = W_j^\tau(b_{0,h}^j)$  and  $h < \tau$ , from Lemma 2(i) and (27) we have that

$$\begin{aligned} v_j^\tau(j, \omega, \rho) &= r_j \left[ \alpha_j^\tau(j, b_{0,h}^j, \rho) + \frac{-(1-p_{11}^j)}{\rho b_{0,h}^j + (1-\rho)b_{0,h+1}^j + (1-p_{11}^j)(h+1-\rho)} \right]. \end{aligned} \quad (53)$$

Case (2). If  $\omega = W_j^\tau(b_{0,\tau}^j)$ , we have

$$\begin{aligned} v_j^\tau(j, \omega, \rho) &= r_j \left[ \frac{\rho b_{0,\tau}^j + (1-\rho)b_s^j}{\rho b_{0,\tau}^j + (1-\rho)b_s^j + (1-p_{11}^j)(\tau+1-\rho)} \right] \\ &= r_j \left[ \alpha_j^\tau(j, \omega, \rho) + \frac{-(1-p_{11}^j)}{\rho b_{0,\tau}^j + (1-\rho)b_s^j + (1-p_{11}^j)(\tau+1-\rho)} \right]. \end{aligned} \quad (54)$$

Case (3) If  $\omega = W_j^\tau(b_s^j)$ , we have

$$\begin{aligned} v_j^\tau(j, \omega, \rho) &= r_j \left[ \frac{b_s^j \rho}{\tau \rho (1-p_{11}^j) + (1-p_{11}^j) + \rho b_s^j} \right] \\ &= r_j \left[ \alpha_j^\tau(j, \omega, \rho) + \frac{-\rho(1-p_{11}^j)}{\tau \rho (1-p_{11}^j) + (1-p_{11}^j) + \rho b_s^j} \right]. \end{aligned} \quad (55)$$

Case (4). If  $\omega > W_j^\tau(b_s^j)$ , since  $v_j^\tau(j, \omega, \rho) = \alpha_j^\tau(j, \omega, \rho) = 0$ , the statement holds trivially.

Note that, in the above Case (1)-(3), the second summand in (53)-(55) decreases with the randomization parameter  $\rho$ . Since, from Lemma 2(ii) and part (i), both  $\alpha_j^\tau(j, \omega, \rho)$  and  $v_j^\tau(j, \omega, \rho)$  increase with  $\rho$ , we have for any  $\rho_1 > \rho_2$ ,

$$0 \leq v_j^\tau(j, \omega, \rho_1) - v_j^\tau(j, \omega, \rho_2) \leq r_j \left[ \alpha_j^\tau(j, \omega, \rho_1) - \alpha_j^\tau(j, \omega, \rho_2) \right].$$

We also have  $v_i^\tau(j, \omega, \rho_1) = v_i^\tau(j, \omega, \rho_1)$  and  $\alpha_i^\tau(j, \omega, \rho_1) = \alpha_i^\tau(j, \omega, \rho_1)$  for  $i \neq j$  since there is no randomization associated with user  $i$ . Therefore, for all user  $i$ ,

$$0 \leq v_i^\tau(j, \omega, \rho_1) - v_i^\tau(j, \omega, \rho_2) \leq r_i \left[ \alpha_i^\tau(j, \omega, \rho_1) - \alpha_i^\tau(j, \omega, \rho_2) \right]. \quad (56)$$

Next consider when  $i < j$ ,

$$\begin{aligned} 0 &= v_i^\tau(j, \omega, \rho) - v_i^\tau(i, \omega, 0) \\ &= r_i \cdot \left[ \alpha_i^\tau(i, \omega, 1) - \alpha_i^\tau(i, \omega, 0) \right] \\ &= r_i \cdot \left[ \alpha_i^\tau(i, \omega, \rho) - \alpha_i^\tau(i, \omega, 0) \right]. \end{aligned} \quad (57)$$

When  $i = j$ , from (56) we have

$$v_i^\tau(j, \omega, \rho) - v_i^\tau(i, \omega, 0) \leq r_i \left[ \alpha_i^\tau(j, \omega, \rho) - \alpha_i^\tau(i, \omega, 0) \right]. \quad (58)$$

When  $i > j$ , from (56) we have

$$\begin{aligned} &v_i^\tau(j, \omega, \rho) - v_i^\tau(i, \omega, 0) \\ &= v_i^\tau(i, \omega, 1) - v_i^\tau(i, \omega, 0) \\ &\leq r_i \left[ \alpha_i^\tau(i, \omega, 1) - \alpha_i^\tau(i, \omega, 0) \right] \\ &\leq r_i \left[ \alpha_i^\tau(i, \omega, \rho) - \alpha_i^\tau(i, \omega, 0) \right]. \end{aligned} \quad (59)$$

Therefore, from (57)-(59), we have

$$v_i^\tau(j, \omega, \rho_1) - v_i^\tau(i, \omega, 0) \leq r_i \left[ \alpha_i^\tau(i, \omega, 1) - \alpha_i^\tau(i, \omega, 0) \right] \quad (60)$$

Similarly, we have

$$v_i^\tau(i, \omega, 1) - v_i^\tau(j, \omega, \rho) \leq r_i \left[ \alpha_i^\tau(i, \omega, 1) - \alpha_i^\tau(j, \omega, \rho) \right]. \quad (61)$$

Now consider the case when  $\omega_1 \neq \omega_2$ . Suppose  $\omega_1 = W_i^\tau(b_{0,h_1}^i)$  and  $\omega_2 = W_i^\tau(b_{0,h_2}^i)$  with  $h_1 < h_2 \leq \tau$ .

$$\begin{aligned} &\left| v_i^\tau(j_1, W_i^\tau(b_{0,h_1}^i), \rho_1) - v_i^\tau(j_2, W_i^\tau(b_{0,h_2}^i), \rho_2) \right| \\ &\leq \left| v_i^\tau(j_1, W_i^\tau(b_{0,h_1}^i), \rho_1) - v_i^\tau(i, W_i^\tau(b_{0,h_1}^i), 0) \right| \\ &\quad + \sum_{h_1 < h < h_2} \left[ v_i^\tau(i, W_i^\tau(b_{0,h}^i), 1) - v_i^\tau(i, W_i^\tau(b_{0,h}^i), 0) \right] \\ &\quad + v_i^\tau(i, W_i^\tau(b_{0,h_2}^i), 1) - v_i^\tau(j_2, W_i^\tau(b_{0,h_2}^i), \rho_2) \Big| \\ &\leq r_i \left| \alpha_i^\tau(j_1, W_i^\tau(b_{0,h_1}^i), \rho_1) - \alpha_i^\tau(i, W_i^\tau(b_{0,h_1}^i), 0) \right| \\ &\quad + \sum_{h_1 < h < h_2} \left[ \alpha_i^\tau(i, W_i^\tau(b_{0,h}^i), 1) - \alpha_i^\tau(i, W_i^\tau(b_{0,h}^i), 0) \right] \\ &\quad + \alpha_i^\tau(i, W_i^\tau(b_{0,h_2}^i), 1) - \alpha_i^\tau(j_2, W_i^\tau(b_{0,h_2}^i), \rho_2) \Big| \\ &= r_i \left| \alpha_i^\tau(j_1, W_i^\tau(b_{0,h_1}^i), \rho_1) - \alpha_i^\tau(j_2, W_i^\tau(b_{0,h_2}^i), \rho_2) \right| \\ &= r_i \left| \alpha_i^\tau(j_1, \omega_1, \rho_1) - \alpha_i^\tau(j_2, \omega_2, \rho_2) \right|, \end{aligned}$$

where the first inequality is because  $v_i^\tau(i, W_i^\tau(b_{0,h}^i), 0) =$



$v_i^\tau(i, W_i^f(b_{0,h+1}^i), 1)$  and  $\alpha_i^\tau(i, W_i^f(b_{0,h}^i), 0) = \alpha_i^\tau(i, W_i^f(b_{0,h+1}^i), 1)$ , which can be observed from (27) and Lemma 2(i). The second equality is from (60)-(61). For other combinations of  $\omega_1$  and  $\omega_2$ , the proof holds similarly. Part (ii) thus holds.

## REFERENCES

- [1] W. Ouyang, A. Eryilmaz, N. B. Shroff, "Low-complexity Optimal Scheduling Over Correlated Fading Channels with ARQ Feedback," *IEEE WiOpt 2012*, Paderborn, Germany.
- [2] L. Tassiulas, A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Transactions on Automatic Control*, vol. 37, no. 12, pp. 1936-1948, Dec. 1992.
- [3] L. Tassiulas, A. Ephremides, "Dynamic server allocation to parallel queues with randomly varying connectivity," *IEEE Transactions on Information Theory*, vol. 39, pp. 466-478, 1993.
- [4] X. Lin, N. B. Shroff, "Joint rate control and scheduling in multihop wireless networks," *43rd IEEE Conference on Decision and Control*, Atlantis, Bahamas, Dec. 2004.
- [5] A. Eryilmaz, R. Srikant, "Fair resource allocation in wireless networks using queue-length based scheduling and congestion control," *IEEE/ACM transactions on networking*, vol. 15, no. 6, pp. 1333-1344, Dec. 2007.
- [6] J. S. Harsini and F. Lahouti and M. Levorato and M. Zorzi, "Analysis of non-cooperative and cooperative type II hybrid arq protocols with amc over correlated fading channels," *IEEE Transactions on Wireless Communications*, vol. 10, no. 3, pp. 877-889, 2011.
- [7] C. Li and X. Wang, "Throughput analysis for parallel arq over correlated MIMO channels", *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1322-1332, 2007.
- [8] S. M. Kim, W. Choi, T. W. Ban and D. K. Sung, "Optimal rate adaptation for hybrid arq in time-correlated rayleigh fading channels", *IEEE Transactions on Wireless Communications*, vol. 10, no. 3, pp. 968-979, 2011.
- [9] B. Makki, A. Gaell i Amat and T. Eriksson, "Green Communication via Power-Optimized HARQ Protocols", *IEEE Transactions on Vehicular Technology*, vol. 63, no. 1, pp. 161-177, 2014.
- [10] W. Ouyang, S. Murugesan, A. Eryilmaz, N. B. Shroff, "Scheduling with rate adaptation under incomplete knowledge of channel/estimator statistics," in *Allerton Conference*, 2010.
- [11] J. Huang, R. A. Berry, and M. L. Honig, "Wireless scheduling with hybrid ARQ", *IEEE transactions on wireless communications*, vol. 4, no. 6, 2005.
- [12] R. Aggarwal, M. Assaad, C. E. Koksai, and P. Schniter, "Joint scheduling and resource allocation in the ofdma downlink: utility maximization under imperfect channel-state information," *IEEE transactions on signal processing*, vol. 59, no. 11, Nov 2011.
- [13] W. Ouyang, N. Prasad, S. Rangarajan, "Exploiting Hybrid Channel Information for Downlink Multi-User MIMO Scheduling," *IEEE WiOpt 2013*, Tsukuba Science City, Japan.
- [14] W. Ouyang, S. Murugesan, A. Eryilmaz, N. Shroff, "Exploiting channel memory for joint estimation and scheduling in downlink networks," *IEEE INFOCOM*, Shanghai, China, Apr. 2011.
- [15] P. Jacko, S. S. Villary, "Opportunistic Schedulers for Optimal Scheduling of Flows in Wireless Systems with ARQ Feedback," *24th International Teletraffic Congress*, Sept. 2012.
- [16] W. Ouyang, A. Eryilmaz, N. B. Shroff, "Asymptotically optimal downlink scheduling over markovian fading channels," *IEEE INFOCOM 2012*, Orlando, Florida (ArXiv Preprint: 1108.3768).
- [17] S.H. Ahmad, M. Liu, T. Javidi, Q. Zhao and B. Krishnamachari, "Optimality of myopic sensing in multi-Channel opportunistic access," *IEEE Trans. on Info. Theory*, 2009
- [18] K. Liu, Q. Zhao, "Indexability of restless bandit problems and optimality of whittle's index for dynamic multichannel access," *IEEE Transactions on Information Theory*, vol. 56, pp. 5547-5567, 2008.
- [19] P. S. Ansell, K. D. Glazebrook, J. Nino-Mora, M. O'Keefe "Whittle's index policy for a multi-class queueing system with convex holding costs," *Mathematical Methods of Operations Research*, vol. 57, pp. 21-39, 2003.
- [20] K.D. Glazebrook, D.J. Hodge, and C. Kirkbride. "General notions of indexability for queueing control and asset management," *The Annals of Applied Probability*, vol.21, no.3, pp. 876-907, 2011.
- [21] C. Li, M. J. Neely, "Exploiting channel memory for multiuser wireless scheduling without channel measurement: capacity regions and algorithms," *Elsevier Performance Evaluation*, vol 68, no. 8, pp. 631-657, Aug. 2011.
- [22] C. Li, M. J. Neely, "Network utility maximization over partially observable markovian channels," *IEEE 10th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, May 2011.
- [23] S. H. Ahmad, M. Liu, T. Javidi, Q. Zhao, B. Krishnamachari, "Optimality of myopic sensing in multi-channel opportunistic access," *IEEE Transactions on Information Theory*, vol. 55, pp. 4040-4050, 2009.
- [24] S. Murugesan, P. Schniter, N. B. Shroff, "Multiuser scheduling in a Markov-modeled downlink using randomly delayed ARQ feedback," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 4040-4050, 2012.
- [25] K. Jagannathan, S. Mannor, I. Menache, E. Modiano, "A state action frequency approach to throughput maximization over uncertain wireless channels," *IEEE INFOCOM*, Shanghai, China, Apr. 2011.
- [26] G. Celik, E. Modiano, "Scheduling in networks with time-varying channels and reconfiguration delay," *IEEE INFOCOM*, Orlando, FL, Mar. 2012.
- [27] H. Bogucka, A. Conti, "Degrees of freedom for energy savings in practical adaptive wireless systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 38-45, 2011.
- [28] E. Oh, B. Krishnamachari, X. Liu, Z. Niu, "Toward dynamic energy-efficient operation of cellular network infrastructure," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 56 -61, 2011.
- [29] K. Son, H. Kim, Y. Yi, B. Krishnamachari, "Base station operation and user association mechanisms for energy-delay tradeoffs in green cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 8, pp. 1525 - 1536, 2011.
- [30] C. Papadimitriou, J.N. Tsitsiklis "The complexity of optimal queueing network control," *Mathematics of Operation Research*, vol. 24, pp. 293-305, 1999.
- [31] P. Whittle, "Restless bandits: activity allocation in a changing world," *Journal of Applied Probability*, vol. 25, pp. 287-298, 1988.
- [32] S. Murugesan, P. Schniter, N. B. Shroff, "Opportunistic Scheduling using ARQ feedback in Multi-Cell Downlink," Asilomar, 2010.
- [33] E. J. Sondik, "The optimal control of partially observable Markov Decision Processes," PhD thesis, Stanford University, 1971.
- [34] Eitan Altman, "Constrained Markov Decision Processes", Chapman & Hall, 1999.
- [35] J. D. Isom, S. Meyn, R. D. Braatz, "Piecewise linear dynamic programming for constrained POMDPs," *National Conference on Artificial Intelligence*, pp. 291-296, 2008.
- [36] C. Li and M. J. Neely, "Energy-Optimal Scheduling with Dynamic Channel Acquisition in Wireless Downlinks," *IEEE Transactions on Mobile Computing*, vol. 9, pp. 527-539, 2010.
- [37] L. Georgiadis, M. Neely, L. Tassiulas, "Resource allocation and cross-Layer control in wireless networks," NOW Publishers Inc., 2006
- [38] S. Meyn, "Control Techniques for Complex Networks," Cambridge University Press, 2007.
- [39] P. W. Glynn, D. Ormoneit, "Hoeffding's inequality for uniformly ergodic Markov chains," *Statistics and probability letters*, vol. 56, no. 2, pp. 143-146, 2002.