A Whittle's Index Based Approach for QoE Optimization in Wireless Networks

ARJUN ANAND, The University of Texas at Austin, USA
GUSTAVO DE VECIANA, The University of Texas at Austin, USA

The design of schedulers to optimize heterogeneous users' Quality of Experience (QoE) remains a challenging and important problem for wireless systems. This paper explores three inter-related aspects of this problem: 1) non-linear relationships between a user's QoE and flow delays; 2) managing load dependent QoE trade-offs among heterogeneous application classes; and 3), striking a good balance between opportunistic scheduling and greedy QoE optimization. To that end we study downlink schedulers which minimize the expected cost modeled by convex functions of flow delays for users with heterogeneous channel rate variations. The essential features of this challenging problem are modeled as a Markov Decision Process to which we apply Whittle's relaxation, which in turn is shown to be indexable. Based on the Whittle's relaxation we develop a new scheduling policy, Opportunistic Delay Based Index Policy (ODIP). We then prove various structural properties for ODIP which result in closed form expressions for Whittle's indices under different scheduler scenarios. Using extensive simulations we show that ODIP scheduler provides a robust means to realize complex QoE trade-offs for a range of system loads.

CCS Concepts: • **Mathematics of computing → Stochastic control and optimization**; • **Networks → Mobile networks**; *Wireless access points, base stations and infrastructure*; *Network algorithms*;

Additional Key Words and Phrases: Wireless networks, Opportunistic scheduling, Whittle's Index

ACM Reference Format:
Arjun Anand and Gustavo de Veciana. 2018. A Whittle's Index Based Approach for QoE Optimization in Wireless Networks. *Proc. ACM Meas. Anal. Comput. Syst.* 2, 1, Article 15 (March 2018), 39 pages. https://doi.org/10.1145/3179418

1 INTRODUCTION

Cellular networks will have to support a heterogeneous collection of applications ranging from mobile broadband to machine-to-machine type communications. The allocation of Base Station (BS) resources among heterogeneous traffic classes with possibly diverse Quality of Experience (QoE) metrics remains a challenging and central problem in wireless system design and is the focus of this paper.

Several studies have shown that the QoE for varous applications is dependent on flow transfer delays, e.g., associated with the delay to view a web page or downloading a file, see e.g., [7, 20, 25, 28]. In this paper we focus on optimizing QoE metrics which are based on file-level delays in a downlink setting. Traditional work on delay minimization, see e.g. [19, 22], has not simultaneously addressed the following aspects of user experience and resource allocation in wireless networks:

Authors' addresses: Arjun Anand, The University of Texas at Austin, 2501 Speedway EER Level 6, Austin, Texas, 78712, USA, arjun_anand@utexas.edu; Gustavo de Veciana, The University of Texas at Austin, 2501 Speedway EER 6.874, Austin, Texas, 78712, USA, gustavo@ece.utexas.edu.

2476-1249/2018/3-ART15 $15
https://doi.org/10.1145/3179418

(1) QoE of a user may be a non-linear function of the delay to download a file. For example, for many applications users can tolerate delays up to a certain threshold and beyond that the user experience deteriorates gradually [27].
(2) Applications may have different sensitivities to delay. Some applications could be more delay tolerant than others, e.g., a simple file download vs interactive web browsing, thus a scheduler can exploit this heterogeneity in delay sensitivity to realize appropriate QoE trade-offs among applications for a range of system loads.
(3) User service rates may change with time due to variations in wireless channel characteristics and different users may have different service rates at any given time.

In this paper we explore addressing the above mentioned issues simultaneously. To that end, we consider a setting in which each user in the system has a job to be served by the BS and it has an associated cost function which is a non-decreasing function of the delay to complete its service. Our aim is to study how to minimize the total expected cost in serving all types of jobs in the system.

The cost function models the QoE of a user as a function of the delay it experiences. The larger the cost, the poorer the QoE perceived by the user. Since the cost function could be non-linear and possibly be different for different jobs this approach takes into account both the non-linearity and the heterogeneity in users' QoE with respect to the delay experienced. Using this approach we can model several useful cost functions, for example, one could consider polynomial functions of delay to model the user's QoE[27] for applications like web browsing and FTP. QoE for stored video streaming (DASH framework) is slightly more complex as it is a function of several parameters like the amount of re-buffering, initial delay and variations in quality of video segments [25]. However, our notion of flow is flexible to accommodate this setting. Indeed current video streaming protocols essentially transfer a sequence of flows associated with video segments. The QoE can then be tied to the delays of these flows/files and/or variability associated with transferring them to the receiver. Cost functions can be obtained through offline studies which collect Mean Opinion Scores (MOS) from users, see for e.g. [7] and [25]. Henceforth, we shall use cost as a measure of a user's QoE.

An important challenge which is specific to systems with time-varying service rates is realizing the right trade-off between *opportunism and minimizing cost*. If we schedule the user with the highest service rate at all times, then we may increase the overall rate at which the jobs are served. However, this *opportunistic* selection of jobs for service may not be cost optimal, as delay critical jobs with low service rates may see poor cost performance. At the other extreme, if we schedule jobs solely based on their current marginal costs, then we may schedule users when their service rates are low and hence the overall rate at which jobs are processed goes down and overall jobs are delayed, resulting in poor overall cost. Therefore, one needs to find the right balance between being opportunistic and giving priority based on cost. This is explored in this paper by studying directly how to minimize the expected system cost.

1.1 Related Work

We classify the related work into two categories based on the underlying model for job arrivals to the system, namely: 1) *Dynamic* system in which jobs arrive according to a stochastic process (typically a Poisson process) and leave once they are serviced; and 2) *Transient* system in which there is a finite number of jobs at the beginning and no additional arrivals enter the system. We will make further classifications based on the information on job sizes available to the scheduler, for example, some works assume that the job sizes are known to the scheduler whereas others assume that there is perfect or partial knowledge of job size distributions. Another characteristic which distinguishes various works in the literature is whether they consider a system with time-varying service rates.

1.1.1 Dynamic Systems. Many authors have considered mean delay minimization in dynamic systems which process jobs at a constant service rate, see for e.g., [1, 2, 10, 22, 37]. If the job sizes are known to the scheduler, it has been shown that the Shortest Remaining Processing Time (SRPT) scheduler is the mean delay optimal scheduler [31]. Under the SRPT policy, the job with the least remaining processing time is scheduled for service at all times. If only the job size distributions are known and the job arrivals form a Poisson process, then it has been proved that the Gittins index scheduler is mean delay optimal [17]. Gittins index schedulers assign a priority to jobs depending on the service received to date and job size distributions. Properties of Gittins index based schedulers for different job size distributions have been studied extensively, see [1–3, 6]. There are few works which consider, however, time-varying service rates in a dynamic system, see [15, 16, 33]. These works either focus on establishing system stability rather than delay-based performance metrics, or propose heuristics which are based on schedulers developed for constant service rate systems.

An interesting line of work which focuses on non-linear cost functions of the *mean* file/job delay in multi-class systems is explored in [8, 23]. However, these works deal with cost functions of expected delays rather than expectation of cost functions of the delays experienced by users. This difference is crucial since minimizing the expectation of the cost functions of delay accounts for higher moments of the delay distribution, whereas, minimizing a metric based on functions of expected delays only accounts for the first moments. Our approach therefore, can model scenarios where the users are sensitive to both the mean and the variability in delay distributions seen by the users. Also, [8, 23] do not consider time-varying job service rates which are typical in wireless settings.

Another line of work which focuses on optimizing non-linear cost functions of delay and queue lengths in multi-class systems includes [14, 18, 24, 29, 32, 35]. They consider generalizations of $c\mu$ rule and prove its optimality in heavy traffic regime for various settings. They differ from our work in the following ways.

(1) The above works except [32] do not consider time-varying service rates.
(2) They do not use the job size information for scheduling, instead, use only the average job size of each class. Using knowledge of job sizes or distribution of actual size is beneficial as it helps us further discriminate jobs based on their sizes.
(3) They allow preemption among jobs of different classes but do not allow preemption among jobs of the same class. In wireless systems the jobs sizes could have large variations in their size. Therefore, if we do not allow preemption among jobs of the same class, then the system might suffer from high delays due to a big Head-of-the-line (HOL) job. Also in systems with time-varying service rates one should be able to switch between jobs quickly to opportunistically schedule users. In our work, we allow both preemption within a class and across classes.

In [9], the authors consider optimization of average cost under convex holding costs functions of the number of users in the system. This is different from our setting where we associate a cost with the delay experienced by each user.

1.1.2 Transient Systems. Unfortunately, many problems are analytically intractable in the dynamic setting. In particular there is no known optimal solution to the problem of minimizing mean delay in a dynamic system with time-varying service rates [4]. Therefore, many authors have focused on scheduling policies which optimize the relevant metrics in transient systems and propose such solutions as a heuristic for dynamic systems. The effectiveness of these policies are then studied through simulation. Our problem is also analytically intractable in a dynamic system and hence, we shall also consider transient systems. Next we will discuss related work focused on transient systems.

The authors of [5, 30] have considered minimizing mean delay in the transient setting where they assume that there is a time-scale separation between service-rate variations and job service times. This means that service rate variations occur at a time-scale which is much smaller than the overall time taken to serve a job. They also assume that the service rate fluctuations are statistically identical and independent across users. These assumptions are valid in situations where the job sizes are large and/or when the service rate variations are due to fast fading. Under this assumption, they have combined opportunistic scheduling with a SRPT like policy to minimize mean delay. The main issue with this approach is that the assumption of statistically identical service rate variations across users may not be valid in scenarios where there are users with heterogeneous mobility patterns. Also, the assumption of a time-scale separation may not hold when there are many short files to be transmitted.

Minimizing delay based metrics in a transient system with time-varying rates for jobs and without the assumption of time-scale separation between service-rate variations and job service times is unfortunately still analytically intractable due to the associated large state spaces. Recently there have been many works which leverage Whittle's indices to explore the optimization of delay performance in wireless networks in a transient setting [4, 11, 21, 34]. However, this line of work has focused only on minimizing weighted linear functions of delay and does not address non-linear cost functions of delay. In [11], the authors have shown that the problem of minimizing mean delay is indexable and derived the Whittle's index when job sizes are geometrically distributed with i.i.d. service rate variations across time. This result was extended to the case with Markovian service rate variations in [21], however, they do not show whether the problem is indexable. In [34], the authors consider a system model where the job sizes are not known but only the job size distributions are known. They derived index policies based on solving a Markov Decision Process, however, they consider only ON-OFF channel model. The approach used in [4] is closely related to our work. They approximate job sizes using shifted Pascal distributions, i.e., a phase-type distribution where each phase has an i.i.d. geometric distribution. They have also derived Whittle's indices when users have heterogeneous two-state i.i.d. channel variations.

1.2 Our Contributions

In this paper we focus on resource allocation strategies to minimize the expectation of possibly non-linear cost functions of job delays in a transient setting with time-varying service rates. To the best of our knowledge, this is the first paper which simultaneously addresses the challenges of 1) non-linearity and heterogeneity in users' experiences as a function of delay, and 2), time-varying service rates for jobs in a non-heavy traffic regime. To that end, we develop a Whittle's index based scheduling policy, which we denote as Opportunistic Delay Based Index Policy (ODIP), for a transient system. ODIP is simple and easy to implement. At any given time, each user has an index based on its residual file size, service rate and its cost function. In any slot we schedule a user based on the indices. The main results of this paper are as follows:

1) *Indexability:* We show that our delay/cost minimization problem is *indexable*. This means that we can associate a well-defined index with each possible state. These indices can then be used to assign priorities to active users.

2) *Opportunistic Delay Based Index Policy:* We derive structural properties of the ODIP index for the case of phase-type job size distributions, convex cost functions of delay, and i.i.d. (possibly heterogeneous) two-state service rates for each user. In particular we show that when a user's instantaneous channel has the best possible rate, then the user has a higher priority than users whose channels are not currently in their respective best possible rates. We then show the following structural properties of the Whittle's index:

Users' States	Parameter of Interest	Description
User i in its best possible rate and user j in its lowest possible rate	Service rate	Priority to user i
Both users in their lowest possible rate	Residual file sizes	Priority to user with the largest residual file size
Both users in their best possible rate	Residual file sizes	Priority order depends on the cost function
Both users in their best possible rate	Probability of best possible rate	Priority to user with the lower probability
Users i and j are in lowest possible rate or best possible rate	$c_i(t) \leq c_j(t) \quad \forall t$	Priority to user j

Table 1. Summary of structural properties of ODIP.

(1) Given two users with the same holding cost function and identical and independent channel statistics. If both the users are in their respective lowest possible rates, then the user with the *longest remaining service* time gets higher priority. However, if both users are in their respective best possible channel rates, then the priority order between the two users depends on both the cost function and their respective residual file sizes. These properties should be contrasted with the SRPT scheduling policy which gives the highest priority to the user with the smallest residual file size.
(2) If there are two users which differ only in the probability of their channel being in the best possible rate, then the user with the lowest probability of being at the best rate gets a higher priority. Therefore, ODIP is opportunistic and gives a higher priority to users likely to be in good rates.
(3) If there are two users which differ only in their cost functions and the cost function of one user strictly dominates the other, then the ODIP gives a higher priority to the user with the higher cost function.

These properties are summarized in Table 1 where we have characterized the priority order between two users when we vary one parameter of interest while the other parameters are kept the same.

Leveraging these structural properties, we derive expressions for the Whittle's index for a few special cases. Each case is characterized by two elements of the system model: 1) information on job size distribution available to the scheduler; and 2) service rate model. The cases considered in this paper are summarized in Table 2. In the scenario where job sizes are known to the scheduler, we shall approximate job sizes using an appropriate phase-type distribution. In all the scenarios, we assume that service rates are independent across users, however, they may not have to be statistically identical.

3) *Simulation Study:* For dynamic systems, we use the results from [12] to show that ODIP is maximally stable, i.e., ODIP ensures system stability if there exists a policy which stabilizes the system for the given system load. We then compare the performance of applying ODIP in a dynamic setting with other policies through simulation. We establish that ODIP makes trade-offs which

	Information on Jobs	Service Rate
1	Sizes known	Fixed across time slots
2	Geometric distribution and mean job size known	i.i.d. across time, two states
3	Sizes known	i.i.d. across time, multiple states

Table 2. Various scenarios for which Whittle's indices are obtained.

cannot be achieved by policies which do not take into account the non-linearity of users' QoE in file/job delays. We also show that simple priority based policies perform poorly as compared to ODIP when we consider higher moments of delays in the cost function.

1.3 Organization

The remainder of the paper is organized as follows. In Sec. 2, we describe our system model. In Sec. 3, we develop our Whittle's index based approach. In Sec. 4 we derive the structural properties of ODIP. Expressions for Whittle's index are provided in Sec. 5. Performance evaluation results based on simulation are presented in Sec. 6.

2 SYSTEM MODEL

We consider a transient setting where N users are present in the system at time $t = 0$, each with a single job to be served. Since there is a one-to-one correspondence between a user and a job, we shall use the terms user and job inter-changeably. Time is assumed to be slotted and is indexed by $t = 0, 1, 2, \ldots$. For simplicity we assume that the scheduler can schedule only one user in a given slot and this decision has to be made at the beginning of the slot. Users leave the system after their jobs are served to completion, and there are no further arrivals.

If a user i is scheduled at time t, then it is served at its current service/channel rate $R_i(t)$ measured in bits/slot. We shall assume that the service rate processes $(R_i(t), t \in \mathbb{Z}^+)$, $i = 1, 2, \ldots, N$ are

(1) i.i.d. across time slots and independent across users
(2) We assume that $R_i(t) \in \left\{r_{i,1}, r_{i,2}, \ldots, r_{i,L}\right\}$, and $R_i(t)$ can take the value $r_{i,l}$ with probability $q_{i,l}$. Without loss of generality we assume that $r_{i,1} > r_{i,2} > \ldots r_{i,L}$ and for all l, $q_{i,l} \neq 0$. Let R_i denote an r.v. with the above distribution. We call it as multi-state channel model. A restriction of this model to the case with $L = 2$ is called as a *two-state* channel model.

Independence of service rate across users is a reasonable assumption as the user mobilities are generally independent of each other, and hence, they experience independent and heterogeneous wireless channel variations. We can also account for the heterogeneity in long term channel variations like shadowing and path loss variations by selecting different mean service rates for different users. Small time-scale fast fading experienced by mobile users are taken care by the i.i.d. service rate variations across slots.

Further we assume that the job sizes are drawn from a phase-type distribution as in [4]. Thus the job size of user i is modeled by a random variable S_i given by:

$$S_i = \sum_{j=1}^{j_i} S_{i,j}, \tag{1}$$

where j_i is the number of phases, and $S_{i,j}, j = 1, 2, \ldots, j_i$ are i.i.d. geometric random variables with mean $1/\mu_i$ bits. We use such phase-type distributions to model the following two cases:

(1) If $j_i = 1$, then the phase-type distribution reduces to a geometric distribution. We consider geometric distributions in the second case in Table 2.

(2) If j_i is large we can model known deterministic file sizes by phase type distributions. For example, if the job size of user i is known to be s_i bits, then one can choose μ_i and j_i such that

$$s_i = j_i/\mu_i. \tag{2}$$

For a given value of s_i, as j_i increases, the phase-type approximation of a deterministic/known job size is more accurate. We will use this approximation to study the first and third cases in Table 2.

Next we explain how we model the effect of time varying service rates on the service time of a user. Let us first consider an example where the service rate of user i has a constant value of $r_{i,l}$ bits/slot. If $\mu_i r_{i,l} \leq 1$, then the average number of slots to complete the transmission of a phase of user i can be approximated by $1/\mu_i r_{i,l}$. Therefore, if the service rate is fixed at $r_{i,l}$, then the average number of slots to complete a phase has a geometric distribution with parameter $\mu_i r_{i,l}$. From (2), we require that $j_i \leq s_i/r_{i,l}$ for the condition $\mu_i r_{i,l} \leq 1$ to be true. To ensure that for all j we have $\mu_i r_{i,l} \leq 1$, we assume that for a given value of s_i, we choose j_i and μ_i such that (2) is satisfied and $j_i \leq s_i/r_{i,1}$. We shall assume that s_i is much larger than the number of bits that can be transmitted in a slot, and hence, j_i is large enough to closely approximate s_i with j_i phases.

This idea has a natural extension to time-varying service rates. If the current service rate of user i is $r_{i,l}$, and user i is scheduled for transmission in the current slot, then the probability that its current phase completes in this slot is given by $\mu_i r_{i,l}$. Therefore, the service rate of a user in a given slot modulates the probability of successful completion of the current phase. When all the phases of a user are serviced, then the user leaves the system.

In summary, we shall assume that the scheduler either has knowledge of the exact job sizes or the job size distribution, depending on the case being considered, see Table 2. When we assume that the scheduler has the knowledge of job sizes, we will use phase-type distributions to approximate job sizes. In this setting knowledge of job sizes would imply that the scheduler knows the parameters $\mu_i, i = 1, 2, \ldots, N$ and the number of remaining phases for each user. By contrast when we consider job sizes with geometric distributions, we will assume that the scheduler knows only the parameters of the distributions which are memoryless. We shall also assume that the scheduler knows the service rates of all the users in the next time slot for which a scheduling decision has to be made, and the service rate statistics of all the users.

Let us now introduce the objective function to be optimized:

$$\mathcal{OP}_1 : \quad \min_{\pi \in \Pi} : \ \mathbb{E}^{\pi}\left[\sum_{t=1}^{\infty}\sum_{i=1}^{N} c_i(t)\mathbf{1}\left\{Y_i^{\pi}(t) > 0\right\}\right], \tag{3}$$

where Π is the set of causal and feasible scheduling policies. Here $Y_i^{\pi}(t)$ is a random variable corresponding to the residual file size of user i at time t under policy π and $c_i(t)$ is the holding cost at slot t. A policy is said to be *causal* if it does not assume knowledge of future service rate realizations. A policy is feasible if only one user is scheduled per slot. For a feasible policy π we have that for all i and t:

$$\sum_{i=1}^{N} A_i^{\pi}(t) = 1, \quad A_i^{\pi}(t) \in \{0, 1\}, \quad \text{a.s.}, \tag{4}$$

where $A_i^{\pi}(t)$ is a random variable which is equal to one if user i is scheduled for transmission in slot t and zero otherwise.

The holding cost function $c_i(\cdot)$, is a function of time, that captures the sensitivity of user i's QoE to the delay. Suppose the user leaves the system at time d, then the overall accumulated cost, which we denote by $C_i(\cdot)$, is given by $C_i(d) = \sum_{t=0}^{d} c_i(t)$. Therefore, $c_i(\cdot)$ can be viewed as the marginal cost for a job staying an additional t^{th} slot in the system. The following assumption will be made on these functions.

2.1 Assumption on holding cost functions

(1) *Monotonicity:* For any user i, $c_i(\cdot)$ is a positive, non-decreasing function of time.
(2) *Bounded by polynomials:* There exist real numbers $\delta > 0$, $\zeta > 0$, and $t' \in \mathbb{Z}^+$ such that for $t > t'$ and $i = 1, 2, \ldots, N$, $c_i(t) < \delta t^{\zeta}$.
(3) *Non-zero:* For any user i, $c_i(t)$ is not equal to zero for all t.

The monotonicity assumption ensures that a properly interpolated $C_i(\cdot)$ would be a convex function of the holding time. The boundedness assumption is a technical assumption to ensure finiteness of indices for the policy to be discussed in the sequel. The last assumption rules out trivial solutions to $\mathcal{OP}_1$. Note that if for all t and user i $c_i(t) = c$, then $\mathcal{OP}_1$ reduces to the minimization of the overall mean delay.

The remainder of this paper is focused on exploring resource allocation strategies to solve $\mathcal{OP}_1$.

3 PROBLEM FORMULATION

The minimization problem $\mathcal{OP}_1$ can be viewed as a Markov Decision Process (MDP) when the channel rate variations are Markovian or i.i.d. across time. However, due to the large state space, in general it is not analytically tractable. Therefore, we will consider the so called Whittle's relaxation of $\mathcal{OP}_1$ [36].

The main idea underlying Whittle's relaxation is to relax the constraint of scheduling exactly one user per slot. Instead we add a cost ν for scheduling a user on a given slot, and we minimize a new total cost function which is given by:

$$\mathcal{OP}_2: \quad \min_{\pi \in \tilde{\Pi}} : \mathbb{E}^{\pi} \left[\sum_{t=0}^{\infty} \sum_{i=1}^{N} c_i(t) \mathbf{1} \left\{ Y_i^{\pi}(t) > 0 \right\} + \nu \sum_{t=0}^{\infty} \sum_{i=1}^{N} A_i^{\pi}(t) \right], \tag{5}$$

where $\tilde{\Pi}$ is the set of causal policies, which may no longer satisfy (4). This relaxed problem can now be de-coupled into sub-problems associated with each user i as follows:

$$\mathcal{SP}_i(\nu): \quad \min_{\pi \in \tilde{\Pi}} : \mathbb{E}^{\pi} \left[\sum_{t=0}^{\infty} c_i(t) \mathbf{1} \left\{ Y_i^{\pi}(t) > 0 \right\} + \nu \sum_{t=0}^{\infty} A_i^{\pi}(t) \right]. \tag{6}$$

Using Whittle's relaxation one can obtain a feasible policy for $\mathcal{OP}_1$ based on the solutions to $\mathcal{SP}_i(\nu), i = 1, 2, \ldots, N$. To that end we first explore the solution to the MDP associated with $\mathcal{SP}_i(\nu)$.

Consider $\mathcal{SP}_i(\nu)$. User i's state is specified by three variables: j the number of remaining phases including the current phase; r the current service rate; and, t the current time. There are two possible actions in a state, to Transmit (T) or Not to Transmit (NT). Let $P\left((j, r, t), (j', r', t'); a\right)$ be the transition probability from the state (j, r, t) to (j', r', t') under the action a. The transition probabilities under the two possible actions are summarized in Table 3. Let us consider an example to illustrate how they are obtained: a transition from (j, r, t) to $\left(j, r_{i,1}, t+1\right)$ occurs under the action T, if the transmission does not succeed in completing a phase in slot t, which happens with probability $(1 - \mu_i r)$ and the service rate in slot $t+1$ is $r_{i,1}$, which happens with probability $q_{i,1}$. Since these are independent events, we have $P\left((j, r, t), \left(j, r_{i,1}, t+1\right); T\right) = q_{i,1}(1 - \mu_i r)$. One can similarly define other transition probabilities. The transition probabilities from (j, r, t) to states other than those specified in Table 3 are zero.

Based on standard results for MDPs, it can be shown that there exists a time-varying Markov policy which is optimal for $\mathcal{SP}_i(\nu)$, see [13]. Therefore, we shall restrict ourselves to Markov

Transition Probability	Expression
$P\left((j,r,t),\left(j,r_{i,l},t+1\right);T\right)$	$q_{i,l}(1-\mu_i r)$
$P\left((j,r,t),\left(j-1,r_{i,l},t+1\right);T\right)$	$q_{i,l}\mu_i r$
$P\left((j,r,t),\left(j,r_{i,l},t+1\right);NT\right)$	$q_{i,l}$

Table 3. Transition probabilities in state (j,r,t)

policies. Let $V_i^*(j,r,t;\nu)$ be the total cost under the optimal policy for $\mathcal{SP}_i(\nu)$ starting from the state (j,r,t) for a transmission cost of ν. From the Bellman equations for MDPs, we have that

$$V_i^*(j,r,t;\nu) = \min\left\{c_i(t) + \overline{V}_i^*(j,t+1;\nu), c_i(t) + \nu + \mu_i r\overline{V}_i^*(j-1,t;\nu) + (1-\mu_i r)\overline{V}_i^*(j,t+1;\nu)\right\}, \tag{7}$$

where $j \in \{1,2,\ldots j_i\}$, $t \in \{0,1,2,3,\ldots\}$, $r \in \left\{r_{i,1}, r_{i,2}\right\}$, and $\overline{V}_i^*(j,t+1;\nu)$ is defined as follows:

$$\overline{V}_i^*(j,t+1;\nu) := \mathbb{E}\left[V_i^*(j,R_i,t+1;\nu)\right]. \tag{8}$$

$\overline{V}_i^*(j,t+1;\nu)$ is the optimal value function averaged over the service rates. Note that a holding cost $c_i(t)$ is incurred for slot t irrespective of the action taken in slot t. From (7) and the definition of $\overline{V}_i^*(j,t+1;\nu)$, it is clear that the optimal policy will transmit in (j,r,t) if and only if the following inequality holds:

$$\nu \le \mu_i r\Delta_i^*(j,t+1,\nu), \tag{9}$$

where $\Delta_i^*(j,t,\nu)$ is defined as follows:

$$\Delta_i^*(j,t,\nu) := \begin{cases} \overline{V}_i^*(j,t;\nu) - \overline{V}_i^*(j-1,t;\nu), & \text{if } j > 1, \\ \overline{V}_i^*(j,t;\nu), & \text{if } j = 1. \end{cases} \tag{10}$$

Indeed this policy minimizes the value functions by choosing the function minimizing the R. H. S. in (7). The inequality (9) is central to the main results of this paper. It implies it is optimal to transmit in a given state if and only if the marginal decrease in the future cost due to the transmission in the given state is more than the cost ν of transmission.

To develop a feasible solution for $\mathcal{OP}_1$ from $\mathcal{SP}_i(\nu)$, for $i = 1, 2, \ldots, N$, we first show that the problem is indexable. The indexability property, defined in [36] is re-stated here:

Definition 3.1. The optimization problem $\mathcal{SP}_i(\nu)$ is indexable if for any $j \in \{1,2,\ldots,j_i\}$, $r \in \left\{r_{i,1}, r_{i,2}, \ldots, r_{i,L}\right\}$, and $t \in \{0,1,2,\ldots\}$, there exists a value $\nu_i^*(j,r,t)$ such that

(1) It is optimal to transmit in (j,r,t) if $\nu < \nu_i^*(j,r,t)$:
(2) It is optimal not to transmit in (j,r,t) if $\nu > \nu_i^*(j,r,t)$.
(3) It is optimal to either transmit or not to transmit in (j,r,t) if $\nu = \nu_i^*(j,r,t)$.

The value $\nu_i^*(j,r,t)$ is known as the *Whittle's* index.

The indexability property ensures that the optimal action in a given state has a threshold structure in ν. Note that some problems are not indexable, see [36] for examples. However, $\mathcal{SP}_i(\nu)$ is indexable and this result is stated next with a proof given in Appendix A.

THEOREM 3.2. *Under Assumption 2.1, phase-type distribution for file sizes and i.i.d multi-state channel model, $\mathcal{SP}_i(\nu)$ is indexable.*

To construct a feasible solution for $\mathcal{OP}_1$ based on $\mathcal{SP}_i(\nu)$, $i = 1, 2, \ldots, N$, we schedule the user with the highest Whittle's index in each slot. We can interpret the Whittle's index as the lowest price at which it is optimal not to transmit in a given state. A higher Whittle's index means that the state is better suited for transmission. This is a natural heuristic which arises from the relaxation of $\mathcal{OP}_2$. Whittle's index based policies are known to have good performance in practice, see [4, 36]. The remainder of this paper will focus on the derivation and characteristics of the Whittle's index for various scenarios mentioned in Table 2.

4 WHITTLE'S INDEX

In this section we will characterize key structural properties of the Whittle's Index for $\mathcal{SP}_i(v)$. The first main result is given in the following theorem, which is proved in Appendix D.

THEOREM 4.1. *Under Assumption 2.1, phase-type distribution for file sizes and i.i.d multi-state channel model, the Whittle's index for any user i in phase $j \in \{1, 2, \ldots, j_i\}$ is such that*

$$v_i^* \left(j, r_{i,1}, t\right) = \infty, \tag{11}$$

$$v_i^* \left(j, r_{i,l}, t\right) < \infty \quad l \neq 1. \tag{12}$$

Theorem 4.1 implies that for any finite value of v, it is optimal to transmit when the current rate is $r_{i,1}$. Since the lowest price at which it is optimal not to transmit in $\left(j, r_{i,1}, t\right)$ is ∞. Since the Whittle's index for users experiencing their lowest possible rate is finite, they will have a lower priority than users experiencing their best possible channel rate. A similar result was proved in [4] in the setting of constant holding costs. Theorem 4.1 is thus a generalization of that result to convex holding costs.

Since the Whittle's index is ∞ for all users currently experiencing their highest possible service rates, scheduling users based on the Whittle's index policy alone is not feasible. We require a further tie-breaking rule to obtain a feasible policy. We will refer to (11) and (12) as the *primary* indices and the tie-breaking rule which we will derive next will be based on *secondary* indices. The secondary index is defined based on the discounted version of the problem and determined as the asymptotic behavior of the Whittle's index as the discount factor approaches one. The discounted version of $\mathcal{OP}_2$ is given by:

$$\mathcal{OP}_2^\beta : \quad \min_{\pi \in \Pi} \mathbb{E}^\pi \left[\sum_{t=0}^{\infty} \beta^t \left(\sum_{i=1}^{N} c_i(t) \mathbf{1} \left\{ Y_i^\pi(t) > 0 \right\} \right) + v \sum_{t=0}^{\infty} \beta^t \left(\sum_{i=1}^{N} A_i^\pi(t) \right) \right], \tag{13}$$

where $\beta \in [0, 1)$ is the discount factor. The discounted sub-problem for user i is in turn given by:

$$\mathcal{SP}_i^\beta(v) : \quad \min_{\pi \in \Pi} \mathbb{E}^\pi \left[\sum_{t=0}^{\infty} \beta^t c_i(t) \mathbf{1} \left\{ Y_i^\pi(t) > 0 \right\} + v \sum_{t=0}^{\infty} \beta^t A_i^\pi(t) \right]. \tag{14}$$

We can define the Whittle's index for the discounted version of the problem as follows:

Definition 4.2. Let $\mathcal{P}(j, r, t)$ denote the set of prices such that for $v' \in \mathcal{P}(j, r, t)$ it is optimal not to transmit in (j, r, t) when $v > v'$. We let the Whittle's index for the discounted problem for a user i in state (j, r, t), denoted by $v_{i,\beta}^*(j, r, t)$, be $v_{i,\beta}^*(j, r, t) := \inf \{v' : v' \in P(j, r, t)\}$.

The above definition differs from that of the un-discounted case since we do not show or require that the discounted problem be indexable.

The tie-breaking rule for users in their respective best possible service rate is based on the observation that for any $j \in \{1, 2, \ldots, j_i\}$ and $r \in \left\{r_{i,1}, r_{i,2}, \ldots, r_{i,L}\right\}$

$$\lim_{\beta \to 1} v_{i,\beta}^* (j, r, t) = v_i^* (j, r, t). \tag{15}$$

The tie-breaking rule for user i is obtained by considering the asymptote of $v_{i,\beta}^* \left(j, r_{i,1}, t\right)$ as $\beta \to 1$ which we shall call the *secondary* index. This is the same terminology as used in [4]. We define the

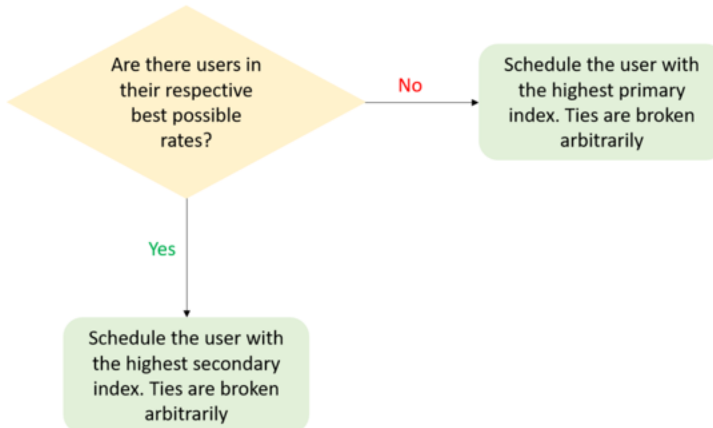

Fig. 1. Flow-chart for ODIP.

secondary index for state $\left(j, r_{i,1}, t\right)$ as given by

$$\xi_i^*\left(j, r_{i,1}, t\right) := \lim_{\beta \to 1} (1-\beta)\, v_{i,\beta}^*\left(j, r_{i,1}, t\right). \tag{16}$$

Since we have defined the secondary index in terms of a limit we have to show that the limit exists and it is finite. This is given by the next result which is proved in Appendix E.1.

THEOREM 4.3. *Under Assumption 2.1, phase-type distribution for job sizes and i.i.d. multi-state channel model, we have that for any $j \in \{1, 2, \ldots, j_i\}$ and $t \geq 0$, the secondary index $\xi_i^*\left(j, r_{i,1}, t\right)$ is finite and $\xi_i^*\left(j, r_{i,1}, t\right) < \infty$.*

With these in hand we can now describe our Whittle's index based policy, which we shall refer to as Opportunistic Delay Based Index Policy (ODIP).

4.1 Opportunistic Delay Based Index Policy (ODIP)

In any time-slot t, we will schedule a user based on the flow-chart exhibited in Fig. 1. We first check if there is any user whose current service rate is the best possible. If there is at least one such user, then we schedule the user with the highest secondary index for transmission. If there is no such user, then we will schedule the user with the highest primary index. The selected user in that case will have a finite primary index as guaranteed by Thm. 4.5.

The computation of indices in ODIP requires cost functions of various applications, channel statistics of users, and flow sizes. When a new user joins the network, there many not enough channel measurements to get reliable channel statistics. Hence, when a new user joins the system, one has to use the typical channel state distribution observed in the network. This can be obtained through offline data collection. As time evolves, one can then update the channel statistics from the channel measurements at the Base Station (BS). Below we develop some qualitative results on the primary and secondary indices, which characterize the scheduling policy.

4.2 Qualitative Results for Two-state Channel Model

In this section for simplicity we shall restrict ourselves to a two-state channel model, i.e., $L = 2$. First we compare the indices of two users where the cost function of one user dominates that of the other user. The proof of this result is given in Appendix G.3.

THEOREM 4.4. *Suppose users i and l have i.i.d. two-state service rate variations. If their holding cost functions are such that for all $t \geq 0$, $c_i(t) \leq c_l(t)$, then for any $j \in \{1, 2, \ldots, j_i\}$, $r \in \{r_{i,1}, r_{i,2}\}$ and $t \geq 0$, we have that $\Delta_i^*(j, t, v) \leq \Delta_l^*(j, t, v)$.*

The above theorem is used to prove the following two important corollaries.

COROLLARY 4.1. *Suppose users i and l have i.i.d. two-state service rate variations. If their holding cost functions are such that for all $t \geq 0$, $c_i(t) \leq c_l(t)$, then for any $j \in \{1, 2, \ldots, j_i\}$, $r \in \{r_{i,1}, r_{i,2}\}$ and $t \geq 0$, $v_i^*(j, r, t) \leq v_l^*(j, r, t)$ and $\xi_i^*(j, r, t) \leq \xi_l^*(j, r, t)$.*

COROLLARY 4.2. *For any user i and phase $j \in \{1, 2, \ldots, j_i\}$, and $t \geq 0$, $v_i^*(j, r_{i,2}, t) \leq v_i^*(j, r_{i,2}, t+1)$ and $\xi_i^*(j, r_{i,1}, t) \leq \xi_i^*(j, r_{i,1}, t+1)$.*

Corollary 4.1 implies that we will give priority to users with 'steeper' holding cost functions. Corollary 4.2 implies that the priority of a user increases with the time spent in the system. This is because of the non-decreasing property of $c_i(t)$, i.e., convex cumulative holding costs. Corollary 4.2 will be useful for studying the structural properties of the primary and secondary indices. The main result for the primary index is given below and it is proved in Appendix G.1.

THEOREM 4.5. *Under Assumption 2.1, phase-type file size distributions and i.i.d. two-state channel model, for any $(j', r_{i,2}, t')$ and $(j, r_{i,2}, t)$, if $j' \geq j$ and $j' + t' \geq j + t$, then $v_i^*(j, r_{i,2}, t) \leq v_i^*(j', r_{i,2}, t')$.*

The j' and t' which satisfy the condition in Thm. 4.5 for a given j and t are shown in Fig. 2. An important corollary to this theorem is given next

COROLLARY 4.3. *$v_i^*(j, r_{i,2}, t)$ is a non-decreasing function of both j and t.*

The above result implies that for any two identical users with the same i.i.d. service rate statistics, holding cost function if they both are in their lowest possible service rates, then the user with the *largest* number of phases remaining to be completed will have priority. This is similar to the Longest Remaining Time First (LRTF) scheduling policy. Intuitively, this is because a user with a large residual job size will have to transmit when service rates are low to reduce the overall holding cost, whereas, a user with a small residual job size can be served opportunistically, i.e., wait for a slot with higher service rate. Since $v_i^*(j, r_{i,2}, t)$ is a non-decreasing function of time, the priority for that user in the next slot is higher if we make a transition to $v_i^*(j, r_{i,2}, t+1)$, i.e., either if we do not transmit in state $(j, r_{i,2}, t)$ or we transmit and fail to complete a phase. However, if we transmit in $(j, r_{i,2}, t)$, complete a phase, and make a transition to $(j-1, r_{i,2}, t+1)$, then the priority may not necessarily increase.

Next we will consider the secondary index. We define the set of 'reachable' states from any state (j, r, t) as follows:

Definition 4.6. If there exists a Markov policy with non-zero transition probability from (j, r, t) to (j', r', t') (in one or more time slots), then (j', r', t') is said to be *reachable* from (j, r, t). The set of all reachable states from (j, r, t) is denoted by $\mathcal{R}(j, r, t)$.

Note that our system model permits only transitions from (j, r, t) to (j', r', t') such that $j' \leq j$ and $t' > t$. Also, we can complete at most one phase in a slot. Therefore, (j', r', t') is reachable from

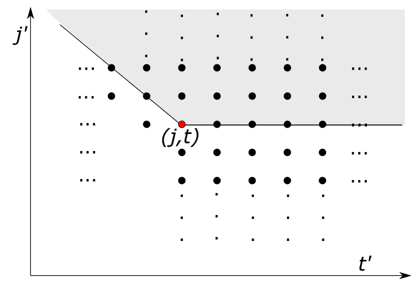

Fig. 2. The shaded region shows j' and t' which satisfy the conditions in Thm. 4.5.

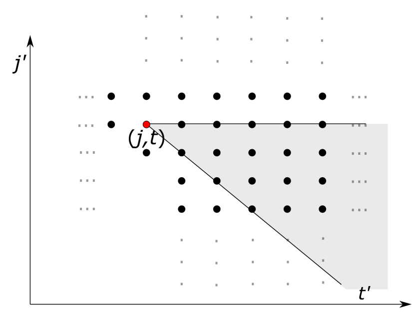

Fig. 3. The shaded region shows j' and t' such that (j', r, t') is reachable from (j, r, t) for any $r \in \{r_{i,1}, r_{i,2}\}$.

(j, r, t), if and only if 1) $j' \leq j$ and 2) $j' + t' \geq j + t$. The value of r' can be either $r_{i,1}$ or $r_{i,2}$, irrespective of the values of j, r, and t. This can be visualized with the help of Fig. 3. The states are exhibited as a two dimensional grid, with time represented in the x-axis and the residual number of phases on the y-axis. We do not explicitly show the channel rate in the representation but it can be understood from the context of the discussion. For $r \in \{r_{i,1}, r_{i,2}\}$, the shaded region represents $\mathcal{R}(j, r, t)$, i.e., if j' and t' are in the shaded region, then both $(j', r_{i,1}, t')$ and $(j', r_{i,2}, t')$ are in $\mathcal{R}(j, r, t)$. For the secondary index, we have the following result which is proved in Appendix G.2.

Theorem 4.7. *Under Assumption 2.1, phase-type file size distributions and i.i.d. two-state channel model, if $(j', r_{i,1}, t')$ is reachable from $(j, r_{i,1}, t)$, then $\xi_i^* (j', r_{i,1}, t') \geq \xi_i^* (j, r_{i,1}, t)$.*

The above theorem implies that for a given user i, the secondary index in slot $t + 1$ is higher than $\xi_i^* (j, r_{i,1}, t)$, whatever is the action taken in the state $(j, r_{i,1}, t)$. Therefore, similar to the primary index, the secondary index for the user in slot $t + 1$ is higher if we do not transmit in $(j, r_{i,1}, t)$ or if we transmit and fail to complete a phase. However, unlike the primary index, the secondary index also increases in slot $t + 1$ if we complete a phase in slot t.

The previous results do not help us characterize the ODIP when users have heterogeneous channels and/or cases where one cost function does not dominate the other. For this we have to find exact expressions for primary and secondary indices. These qualitative results, however, give basic insights and help us in further deriving exact expressions. We derive expressions for indices in the next section.

5 QUANTITATIVE RESULTS

We consider the three different cases mentioned in Table 2. Starting with the simplest case in which we schedule users with fixed service rates and where job sizes are known to the scheduler.

5.1 Fixed Service Rate, Known Deterministic File Sizes

As explained in Sec. 2, we model the job sizes using phase-type distributions. The fixed service rate is a special case of the two-level model described in the previous section where $q_{i,1} = 1$. Suppose user i is served at a fixed rate r_i bits/slot. In this case we shall assume that for all i, $\mu_i r_i = 1$. This would imply that if user i is scheduled in a given slot, then it will complete the phase with probability one. One can also visualize this as splitting the job into j_i equal parts where each part has a size of r_i bits and if user i is selected for transmission, then one part is serviced in that slot. Our main result for this setting is the following. A proof of this result is given in Appendix H.1.

Theorem 5.1. *Under Assumption 2.1, fixed service rate and phase type service requirement with $\mu_i r_i = 1$, ODIP reduces to scheduling a user with the highest secondary index. For a user i in state (j, r_i, t), the secondary index $\xi_i^*(j, r_i, t)$ is given by*

$$\xi_i^*(j, r_i, t) = \frac{1}{j} c_i(t + j). \tag{17}$$

The priority rule described above considers two factors– the residual service time and the cost function of the user. Recall that j, corresponds to the number of phases left to complete, i.e., the number of slots that will be required for that particular user to complete service. Therefore, on the R.H.S. of (17), the term $1/j$ gives more weight to a user with a smaller residual service time and the term $c_i(t + j)$ gives more weight to users with a steeper cost function. Note that $c_i(t + j)$ is the holding cost when the user i leaves the system if it is served without preemption till completion.

This policy can be viewed as a generalization of SRPT, which is known to be the mean delay optimal policy when the job sizes are known and service rate is fixed. If the holding cost function is constant and is same for all users, then (17) reduces to SRPT. With more general cost functions, the priority rule in (17) achieves a trade-off between accelerating short flows and giving priority to users with higher holding cost functions.

5.2 Two-state I.I.D. Service Rates, Geometric File Sizes

In this sub-section we consider the two-state channel model described in Sec. 2. We shall assume that file sizes are geometric. This is a special case of the phase-type distribution where each user has one phase. Since there is only one-phase for each user, we do not have to track the phase of active users. However, we shall explicitly represent this by $j = 1$ to maintain consistent notation as in other cases. We state the main result for this setting next which is proved in Appendix H.2.

Theorem 5.2. *Under Assumption 2.1 on $c_i(t)$, geometric file sizes and two-state i.i.d. service rate variations, the primary index for user i is given by*

$$\nu_i^*\left(1, r_{i,2}, t\right) = \frac{\mu_i \overline{r}_i r_{i,2}}{\overline{r}_i - r_{i,2}} \sum_{k=1}^{\infty} c_i(t + k)\left(1 - \mu_i \overline{r}_i\right)^{k-1} \tag{18}$$

where $\overline{r}_i := q_{i,1} r_{i,1} + (1 - q_{i,1}) r_{i,2}$. The secondary index in turn is given by

$$\xi_i^*\left(1, r_{i,1}, t\right) = q_i \left(\mu_i r_{i,1}\right)^2 \sum_{k=1}^{\infty} c_i(t + k)\left(1 - q_i \mu_i r_{i,1}\right)^{k-1}. \tag{19}$$

Let us now consider how the indices depend on the residual job size, cost functions and the service rates. Since the file sizes are geometric, and thus memoryless, the residual file size at any slot is given by $1/\mu_i$ bits. The larger the value of μ_i the smaller the residual file size. For a given $c_i(t)$, $r_{i,1}$, $r_{i,2}$, and $q_{i,1}$, it can be shown that $\nu_i^*\left(1, r_{i,2}, t\right)$ is a non-increasing function of μ_i. This means that among the users who have the same cost function and who are not in their best possible rates, the users with larger residual file sizes are given priority over users with smaller residual file sizes. The intuition behind this is similar to that underlying Corollary 4.3. However, unlike $\nu_i^*\left(1, r_{i,2}, t\right)$, the properties associated with the changes in $\xi_i^*\left(1, r_{i,1}, t\right)$ as function of μ_i depend on $c_i(t)$.

For a given $\overline{r}_i$ and $c_i(t)$, $\nu_i^*\left(1, r_{i,2}, t\right)$ and $\xi_i^*\left(1, r_{i,1}, t\right)$ are increasing functions of $r_{i,2}$ and $r_{i,1}$, respectively. This means that we give priority to users with better service rates when the other parameters are the same. It can be easily seen that a higher holding cost function results in a higher value for $\nu_i^*\left(1, r_{i,2}, t\right)$ and $\xi_i^*\left(1, r_{i,1}, t\right)$. Therefore, the primary and the secondary indices together achieve a trade-off between minimizing cost and opportunistically scheduling users. Note that if for all t we have $c_i(t) = c_i$, then ODIP reduces to the *Size-Aware Whittle's Index* SWA policy derived in [4]. Our results are thus the generalization of SWA.

5.3 Multi-state I.I.D. Service Rates, Known Deterministic File Sizes

The exact expressions for the primary indices are analytically intractable. Therefore, we will derive a lower bound. We state the main result for this setting which is proved in Appendix H.3.

Theorem 5.3. *Under Assumption 2.1, phase-type file size distributions, and i.i.d. multi-state channels for any $j \in \{1, 2, \ldots, j_i\}$, $t \geq 0$, and $l \in \{2, 3, \ldots, L\}$, the primary index for user i is lowered bounded by:*

$$\nu_i^*\left(j, r_{i,l}, t\right) \geq \frac{\mu_i\left(\sum_{n=1}^{l} q_{i,n} r_{i,n}\right) r_{i,l}}{\sum_{n=1}^{l} q_{i,n} r_{i,n} - r_{i,l}\left(\sum_{n=1}^{l} q_{i,n}\right)} \sum_{m=0}^{\infty} c_i(t+j-1+m)\left(1-\mu_i \sum_{n=1}^{L} q_{i,n} r_{i,n}\right)^m. \tag{20}$$

The secondary index for user i is given by the following equation.

$$\xi_i^*\left(j, r_{i,1}, t\right) = \frac{q_{i,1}\left(\mu_i r_{i,1}\right)^2}{j}\left[H_{i,1}^{\dagger}(j, t+1) - H_{i,1}^{\dagger}(j-1, t+1)\right], \tag{21}$$

where $H_{i,1}^{\dagger}(j,t)$ is the average total holding cost (transmission cost not included) incurred by the policy in which transmissions are done only when channel state $r = r_{i,1}$, when there are j remaining phases at time t. Its value is obtained by solving the following set of equations for all t:

$$H_{i,1}^{\dagger}(j,t) = c_i(t) + \left(1-\mu_i q_{i,1} r_{i,1}\right) H_{i,1}^{\dagger}(j,t) \tag{22}$$

$$+ \mu_i q_{i,1} r_{i,1} H_{i,1}^{\dagger}(j-1,t), \; j = 2, 3, \ldots, j_i, \tag{23}$$

$$H_{i,1}^{\dagger}(1,t) = \sum_{k=0}^{\infty} c_i(t)\left(1-\mu_i q_{i,1} r_{i,1}\right)^k, \tag{24}$$

$$H_{i,1}^{\dagger}(0,t) = 0. \tag{25}$$

The lower bound (20) retains the properties mentioned in Thm. 4.4 and 4.5. Therefore, it retains the priority ordering of various states for a given user as well as the priority ordering among states for two users when the cost function of one user dominates the other. However, it may affect the priority ordering between two users when cost functions do not dominate each other. This will not adversely affect the performance of our ODIP because at moderate to high system loads there would be a sufficient number of users in the system such that at least one user is in its best possible rate and therefore, the scheduling is primarily done based on secondary indices for which we can derive exact expressions.

Let us consider an example for the computation of $\xi_i^*\left(j, r_{i,1}, t\right)$. If $c_i(t) = t$, then we will get the following expression for $H_{i,1}^{\dagger}(j,t)$.

$$H_{i,1}^{\dagger}(j,t) = \frac{j}{\mu_i q_{i,1} r_{i,1}} t + \frac{j(j+1)}{2}\left[\frac{1-\mu_i q_{i,1} r_{i,1}}{\left(\mu_i q_{i,1} r_{i,1}\right)^2}\right]. \tag{26}$$

Substituting (26) in (21), we get the following equation for secondary index.

$$\xi_i^*\left(j, r_{i,1}, t\right) = \frac{\mu_i r_{i,1}(t+1)}{j} + \frac{1-\mu_i q_{i,1} r_{i,1}}{q_{i,1}}. \tag{27}$$

In the above example, the secondary index is a non-decreasing function of the remaining service requirement j for a given t. However, in general, for a given t, the manner in which $\xi_i^*\left(j, r_{i,1}, t\right)$ varies as a function of j depends on $c_i(t)$. Also for a given j, $\xi_i^*\left(j, r_{i,1}, t\right)$ is a non-decreasing function of time. From Corollary 4.2 this holds for any $c_i(t)$ which is a non-decreasing function of t. Another interesting property is that $\xi_i^*\left(j, r_{i,1}, t\right)$ is a non-increasing function of $q_{i,1}$, if all the other parameters are fixed. This can be proved using (21). A smaller $q_{i,1}$ implies that there is less chance of user i being in its best possible rate. Since it is a rare 'good' event, it is good to opportunistically use it to serve user i. Therefore, if all parameters except $q_{i,1}$ are the same for a set of users, then the user with the smallest $q_{i,1}$ gets the highest priority in this set. This is reminiscent of quantile based scheduling [26] and references therein.

6 DYNAMIC SYSTEM

In this section we discuss properties and performance of ODIP when applied to a dynamic setting. As we have stated previously, we propose to use ODIP as a heuristic for the dynamic setting. Instead of starting with a finite number of jobs at time $t = 0$, here we shall consider a system in which jobs arrive according to a Poisson process. Jobs are classified into K different classes based on their holding cost functions. All jobs in a class have the same cost function. Let λ_k be the arrival rate of jobs of class k. We shall assume the same channel model for jobs as in Sec. 2. We shall also assume that all jobs associated with a class have i.i.d. service rate distributions, both across time and between users. Therefore, with a slight abuse of notation, instead of specifying holding cost functions and the service rates of the individual jobs, we will specify them for an entire class. For example, $c_k(\cdot)$ is the cost function of class k and $r_{k,1}$ is the maximum service rate for a job of class k. Finally to specify the holding cost of job in given slot, it will be based on the sojourn time since its arrival to the system.

In a dynamic system, the first concern is whether the system is stable for a given set of arrival rates $\lambda_k, k = 1, 2, \ldots, K$. Let S_k be a r.v. denoting the job size (in bits) of a typical class k job. If the system stability is not maintained, then the delays experienced by jobs may grow unboundedly. From Theorem 5.2 in [12], we have the following result on the stability of the system under ODIP.

Corollary 6.1. *In a dynamic multi-class system with Poisson arrivals and multi-state i.i.d. service rates for jobs, ODIP is maximally stable and the arrival rates must satisfy:*

$$\sum_{k=1}^{K} \frac{\lambda_k \mathbb{E}\left[S_k\right]}{r_{k,1}} < 1. \tag{28}$$

Proof. A policy is said to be maximally stable if it can stabilize the system for any arrival rate for which a stabilizing policy exists. It has been shown in [12] that a class of policies called Best Rate (BR) policies are maximally stable. A BR policy serves a user whose current rare is best possible whenever such a user is present in the system. Our ODIP is also a BR policy and hence, maximally stable. □

We will evaluate the delay cost performance of ODIP for dynamic systems via simulation. In our simulations, we will classify the arriving jobs into two classes based on their QoE requirements. Let λ_1 and λ_2 be the average arrival rates of jobs of Class 1 and 2, respectively. We assume that we can make a scheduling decision every 0.01 sec, i.e., slot duration is 0.01 sec. A job of Class 1 has cost $C_1(d) = d^2$ for a delay of d seconds. We use the gradient of $C_1(\cdot)$ to obtain $c_1(\cdot)$, i.e., $c_i(t) = C_i(t) - C_i(t-1)$. Similarly, a job of Class 2 has cost $C_2(d) = \left(\frac{d}{1.5}\right)^2$. Therefore, Class 1 users are more sensitive to delays than Class 2 users. For Class 1 traffic the cost increases steeply after a delay of one second, whereas the Class 2 traffic can tolerate delays upto 1.5 seconds. We shall compare our scheme with the following three policies:

(1) *Size-Aware Whittle's Index Policy (SW):* This is a BR policy which considers the optimization of weighted mean delay in dynamic systems. It is a special case of ODIP which minimizes a weighted function of mean delays. The weight could be different for each user. This approach does not consider the non-linearity of user experience with respect to delay. In the sequel we will show that even if we optimize the weights for SW scheduling such that it has the least cost among all SW policies for a given set of arrival rates, the costs due to this policy are still higher than the costs under ODIP.

(2) *Proportional Fair (PF):* This is a commonly used rate-based policy in wireless networks in which at any time we schedule a user with the highest ratio of its current rate to the average rate allocated to the user previously. When the service rate is constant for each user, then this policy reduces to Processor Sharing. In a dynamic system with time-varying service rates, it has been shown in [16] that PF is maximally stable. We will compare our scheme with a weighted version of the PF algorithm where we assign a higher weight to the more delay sensitive class. We shall optimize the weight for each arrival rate vector so that the cost is least among all weighted Proportional Fair schedulers. We will show that even with optimized weights this policy cannot achieve good QoE.

(3) *Priority Based Policy:* We consider a simple priority based policy where we give absolute preemptive priority to the more delay sensitive Class 1 jobs over Class 2 jobs and within each class we will schedule users according to SW discipline with unit weights for all jobs. However, this policy is not a maximally stable and hence, we can compare with this policy only for smaller range of arrival rates.

In all the simulation scenarios considered, we shall generate jobs having Pareto file size distribution with c.c.d.f. $\overline{G}(x) = \left(\frac{4}{x+4}\right)^5$, where the size is measured in Mbits. This distribution has a mean of 1 Mbit. For practical systems, these parameters can be scaled appropriately. We now discuss the simulation results for two different settings based on the service rate model: fixed and time-variant service rates.

6.1 Fixed Service Rate

In this section we shall assume that all jobs can be processed at a constant rate of 1 Mbps. If we fix the arrival rate of a class and sweep the arrival rate of the other class, we will get two sets of simulation results. In Fig. 4, we compare the average cost of all the policies when λ_1 is fixed at 0.5 arrivals/sec. and λ_2 is swept. Similarly in Fig. 5, we have fixed λ_2 at 0.5 arrivals/sec. and have swept λ_1. In both the scenarios, ODIP performs better than the other policies. Note that we have optimized the weights of SWA and weighted PF for each data point. ODIP performs better than other policies because it takes into the non-linearity of cost functions. To understand this better, we have plotted the average cost per class when we sweep λ_1 and λ_2 in Figures 6 and 7, respectively. We have only plotted the comparisons with SW as it is the second best policy in terms of the average cost. In both

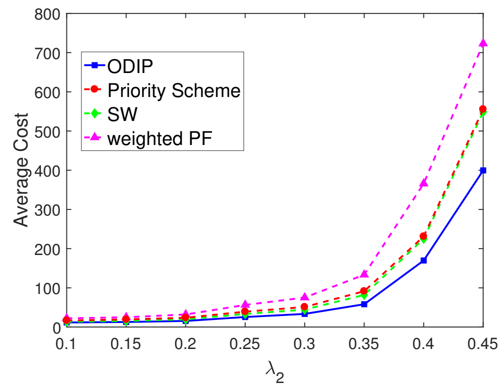

Fig. 4. Average cost as a function of λ_2 (λ_1 = 0.5 arrivals/sec.) in the system with fixed service rates for jobs (1 Mbps).

the scenarios, as the overall system load increases, ODIP protects the delay sensitive Class 1 at the expense of other class. SW which considers the minimization of weighted linear functions of delays does not have the required flexibility to make trade-offs as it can only give a higher weight to the more delay sensitive Class 1 jobs without considering the time spent by the jobs in the system. The priority scheme fully prioritizes Class 1 traffic and hence, jobs of Class 2 traffic have poor delay responses, which has resulted in higher overall cost.

6.2 Time-varying Service Rate

Next we compare ODIP with other policies in a system where users have time-varying service rates. We consider a two-state service rate for all jobs which is i.i.d. across time and users. The maximum rate is 1 Mbps and the minimum rate is 0.5 Mbps, and probability of being in the best possible rate is 0.5 for both the classes.

As in the fixed service case, we compare the average cost under different policies. Note that the priority based scheme is not maximally stable and hence, we cannot simulate it for the full range of arrival rates in the stability region. Figures 8and 9 exhibit the average cost versus λ_2 and λ_1 sweeps, respectively. In both scenarios, ODIP performs better than other policies. The priority scheme performs poorly because it does not fully exploit the opportunism in the system and becomes unstable. The weighted PF does not take into account the delay of jobs while scheduling. Therefore, it has a poor cost performance. SW and ODIP have similar costs at low loads, however, as load increases, ODIP performs better than SW. We have also compared the average cost per class in Fig 10 and 11. As the load increases, ODIP is able to balance the delays experienced by both the classes, wheres, SW can only give a higher weight to the more delay sensitive Class 1 at the expense of Class 2. This results in a better performance for Class 1, but the delays experienced by Class 2 traffic easily exceeds 1.5 seconds and hence, results in a larger cost.

7 CONCLUSIONS

In this paper we have explored the three inter-related problems in scheduling for wireless systems: 1) non-linear relationships between a user's QoE and flow delays; 2) managing load dependent QoE trade-offs among heterogeneous application classes; and 3) striking a good balance between opportunistic scheduling and greedy QoE optimization. We have used Whittle's relaxation to develop our proposed scheme ODIP and to study its structural properties. Simulations confirm the

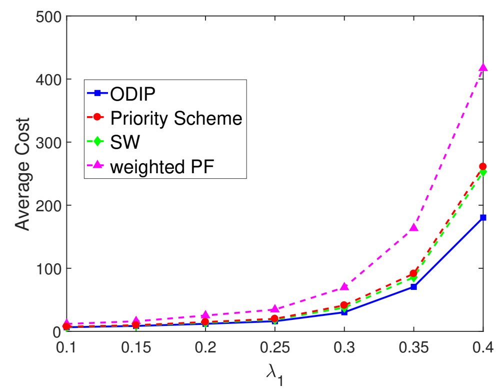

Fig. 5. Average cost as a function of λ_1 (λ_2 = 0.5 arrivals/sec.) in the system with fixed service rates for jobs (1 Mbps).

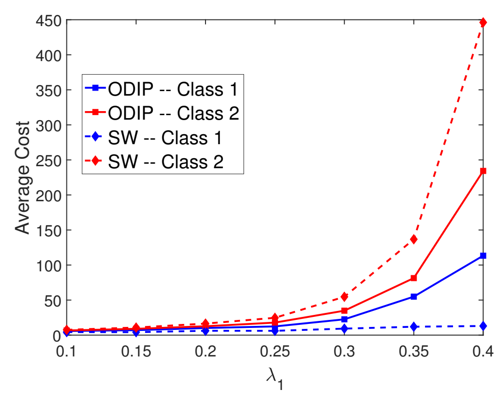

Fig. 6. Average cost as a function of λ_1 (λ_2 = 0.5 arrivals/sec.) in the system with fixed service rates for jobs (1 Mbps).

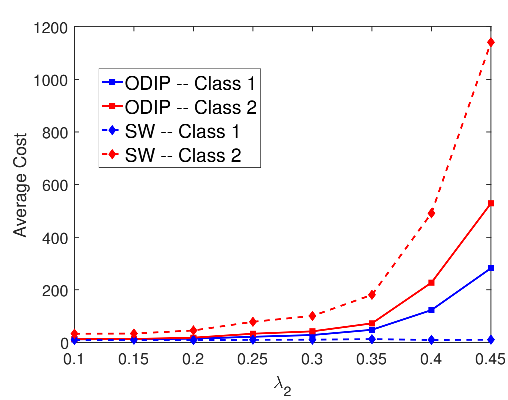

Fig. 7. Average cost as a function of λ_2 (λ_1 = 0.5 arrivals/sec.) in the system with fixed service rates for jobs (1 Mbps).

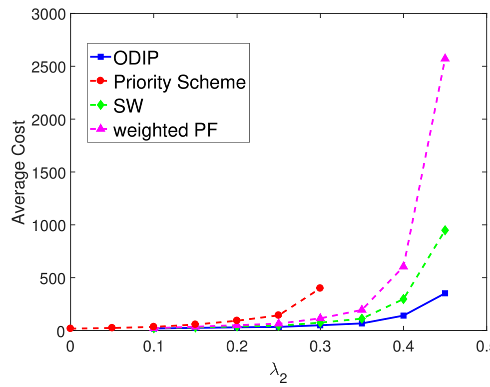

Fig. 8. Average cost as a function of λ_2 ($\lambda_1 = 0.5$ arrivals/sec.) in the system with time-varying service rates for jobs (peak rate 1 Mbps).

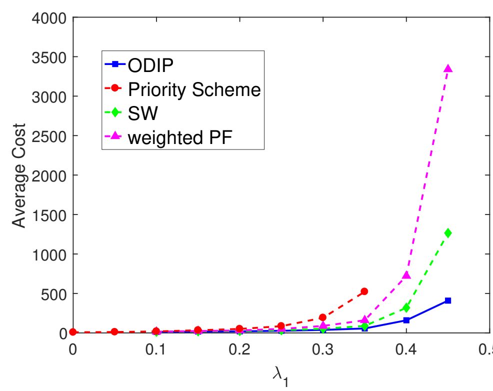

Fig. 9. Average cost as a function of λ_1 ($\lambda_2 = 0.5$ arrivals/sec.) in the system with time-varying service rates for jobs (peak rate 1 Mbps).

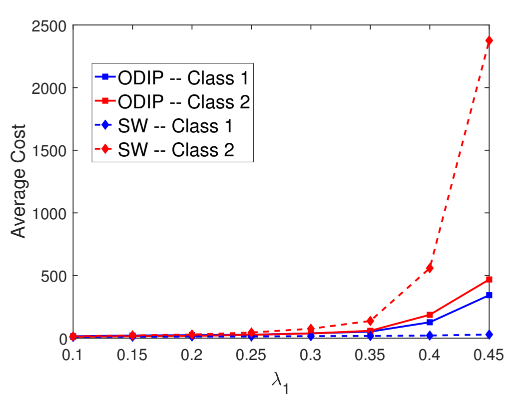

Fig. 10. Average cost as a function of λ_1 ($\lambda_2 = 0.5$ arrivals/sec.) in the system with time-varying service rates for jobs (peak rate 1 Mbps).

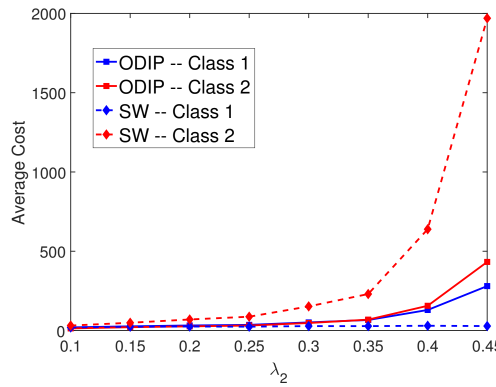

Fig. 11. Average cost as a function of λ_2 ($\lambda_1 = 0.5$ arrivals/sec.) in the system with time-varying service rates for jobs (peak rate 1 Mbps).

effectiveness of ODIP in achieving the complex QoE trade-offs among different traffic classes for a range of system loads.

8 ACKNOWLEDGEMENTS

The work of Arjun Anand was supported by FutureWei Technologies. Gustavo de Veciana was partially supported by NSF grant CNS-1731658 and FutureWei Technologies.

The authors would like to thank the anonymous referees and the shepherd for their valuable comments and suggestions.

REFERENCES

[1] Samuli Aalto and Urtzi Ayesta. 2008. Optimal Scheduling of Jobs with a DHR Tail in the M/G/1 Queue. In *Proc. Int. Conf. on Performance Evaluation Methodologies and Tools (ValueTools '08)*. Article 50, 8 pages.

[2] Samuli Aalto, Urtzi Ayesta, and Rhonda Righter. 2009. On the Gittins index in the M/G/1 queue. *Queueing Systems* 63, 1–4 (2009), 437–458.

[3] Samuli Aalto and Pasi Lassila. 2007. Impact of Size-Based Scheduling on Flow Level Performance in Wireless Downlink Data Channels. In *Managing Traffic Performance in Converged Networks*. Lecture Notes in Computer Science, Vol. 4516. Springer Berlin Heidelberg, 1096–1107.

[4] Samuli Aalto, Pasi Lassila, and Prajwal Osti. 2015. Whittle Index Approach to Size-aware Scheduling with Time-varying Channels. In *Proc. ACM SIGMETRICS Int. Conf. on Measurement and Modeling of Computer Systems (SIGMETRICS '15)*. ACM, 57–69.

[5] Samuli Aalto, Aleksi Penttinen, Pasi Lassila, and Prajwal Osti. 2011. On the Optimal Trade-off Between SRPT and Opportunistic Scheduling. In *Proc. ACM SIGMETRICS Joint Int. Conf. on Measurement and Modeling of Computer Systems (SIGMETRICS '11)*. 185–196.

[6] Samuli Aalto, Aleksi Penttinen, Pasi Lassila, and Prajwal Osti. 2012. Optimal Size-Based Opportunistic Scheduler for Wireless Systems. *Queueing Systems* 72, 1 (Oct. 2012), 5–30.

[7] Pablo Ameigeiras, Juan J. Ramos-Munoz, Jorge Navarro-Ortiz, Preben Mogensen, and Juan M. Lopez-Soler. 2010. QoE Oriented Cross-layer Design of a Resource Allocation Algorithm in Beyond 3G Systems. *Comput. Commun.* 33, 5 (Mar. 2010), 571–582.

[8] Arjun Anand and Gustavo de Veciana. 2017. Measurement-based Scheduler for Multi-class QoE Optimization in Wireless Networks. In *Proc. INFOCOM*. 1–9.

[9] P. S. Ansell, K. D. Glazebrook, J. Niño-Mora, and M. O'Keeffe. 2003. Whittle's Index Policy for a Multi-Class Queueing System with Convex Holding Costs. *Mathematical Methods of Operations Research* 57, 1 (01 Apr. 2003), 21–39.

[10] Konstantin Avrachenkov, Urtzi Ayesta, Patrick Brown, and Eeva Nyberg. 2004. Differentiation Between Short and Long TCP flows: Predictability of the Response Time. In *Proc. INFOCOM*, Vol. 2. 762–773.

[11] Urtzi Ayesta, Martin Erausquin, and Peter Jacko. 2010. A Modeling Framework for Optimizing the Flow-level Scheduling with Time-varying Channels. *Perform. Eval.* 67, 11 (Nov. 2010), 1014–1029.
[12] Urtzi Ayesta, Martin Erausquin, Matthieu Jonckheere, and Maaike Verloop. 2013. Scheduling in a Random Environment: Stability and Asymptotic Optimality. *IEEE/ACM Transactions on Networking* 21, 1 (Feb. 2013), 258–271.
[13] Dimitri P. Bertsekas. 2012. *Dynamic Programming and Optimal Control* (4 ed.). Vol. 2. Athena Scientific.
[14] Carlos F. Bispo. 2011. Single Server Scheduling Problem: Optimal policy for Convex Costs Depends on Arrival Rates. In *Proc. Multidisciplinary Int. Conf. on Scheduling : Theory and Applications (MISTA 2011)*. 275–296.
[15] Sem Borst. 2003. User-level Performance of Channel-Aware Scheduling Algorithms in Wireless Data Networks. In *Proc. INFOCOM*, Vol. 1. 321–331.
[16] Sem Borst and Matthieu Jonckheere. 2006. Flow-Level Stability of Channel-Aware Scheduling Algorithms. In *Proc. 4th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*. 1–6.
[17] John C. Gittins, Kevin D. Glazebrook, and Richard Weber. 2011. *Multi-armed Bandit Allocation Indices* (2nd ed.). Qiley.
[18] Itay Gurvich and Ward Whitt. 2009. Scheduling Flexible Servers with Convex Delay Costs in Many-Server Service Systems. *Manufacturing & Service Operations Management* 11, 2 (2009), 237–253.
[19] Mor Harchol-Balter. 2013. *Performance Modeling and Design of Computer Systems: Queueing Theory in Action*. Cambridge University Press.
[20] Tobias Hobfeld, Sebastian Biedermann, Raimund Schatz, Alexander Platzer, Sebastian Egger, and Markus Fiedler. 2011. The Memory Effect and its Implications on Web QoE modeling. In *Proc. Int. Teletraffic Cong.(ITC)*. 103–110.
[21] Peter Jacko. 2011. Value of Information in Optimal Flow-level Scheduling of Users with Markovian Time-varying Channels. *Perform. Eval.* 68, 11 (Nov. 2011), 1022–1036.
[22] Leonard Kleinrock. 1976. *Queueing Systems*. Vol. II: Computer Applications. Wiley Interscience.
[23] Chi-ping Li and Michael J. Neely. 2011. Delay and Power-Optimal Control in Multi-Class Queueing Systems. *ArXiv e-prints* (Jan. 2011). arXiv:math.OC/1101.2478
[24] Avishai Mandelbaum and Alexander L. Stolyar. 2004. Scheduling Flexible Servers with Convex Delay Costs: Heavy-Traffic Optimality of the Generalized c|mu rule. *Operations Research* 52, 6 (Dec. 2004).
[25] Ricky K. P. Mok, Edward W. W. Chan, and Rocky K. C. Chang. 2011. Measuring the Quality of Experience of HTTP Video Streaming. In *Proc. IFIP/IEEE International Symposium on Integrated Network Management (IM 2011) and Workshops*. 485–492.
[26] Shailesh Patil and Gustavo de Veciana. 2009. Measurement-based Opportunistic Scheduling for Heterogenous Wireless Systems. *IEEE Trans. Commun.* 57, 9 (Sep. 2009), 2745–2753.
[27] Magnus Proebster. 2014. Improving the Quality of Experience with Size-Based and Opportunistic Scheduling. In *Proc. International Symposium on Wireless Communications Systems (ISWCS)*. 443–448.
[28] Magnus Proebster, Matthias Kaschub, Thomas Werthmann, and Stefan Valentin. 2012. Context-Aware Resource Allocation for Cellular Wireless Networks. *EURASIP Journal on Wireless Communications and Networking* 2012, 1 (Jul. 2012), 1–19.
[29] Rhonda Righter and Susan H. Xu. 1991. Scheduling Jobs on Non-Identical IFR Processors to Minimize General Cost Functions. *Advances in Applied Probability* 23, 4 (Dec. 1991), 909–924.
[30] Bilal Sadiq and Gustavo de Veciana. 2010. Balancing SRPT Prioritization vs Opportunistic Gain in Wireless Systems with Flow Dynamics. In *Proc. Int. Teletraffic Cong. (ITC)*. 1–8.
[31] Linus Schrage. 1968. A Proof of the Optimality of the Shortest Remaining Processing Time Discipline. *Operations Research* 16, 3 (1968), 687–690.
[32] Alexander L. Stolyar. 2004. MaxWeight Scheduling in a Generalized Switch: State Space Collapse and Workload Minimization in Heavy Traffic. *Ann. Appl. Probab.* 14, 1 (Feb. 2004), 1–53.
[33] Ianire Taboada, Jose Oscar Fajardo, Fidel Liberal, and Bego Blanco. 2012. Size-Based and Channel-Aware Scheduling Algorithm Proposal for Mean Delay Optimization in Wireless Networks. In *Proc. ICC*. 6596–6600.
[34] Ianire Taboada, Peter Jacko, Urtzi Ayesta, and Fidel Liberal. 2014. Opportunistic Scheduling of Flows with General Size Distribution in Wireless Time-Varying Channels. In *Proc. Teletraffic Cong. (ITC)*. 1–9.
[35] Jan A. van Mieghem. 1995. Dynamic Scheduling with Convex Delay Costs: The Generalized c|mu Rule. *The Annals of Applied Probability* 5, 3 (1995), 809–833. http://www.jstor.org/stable/2245125
[36] Peter Whittle. 1988. Restless Bandits: Activity Allocation in a Changing World. *Journal of Applied Probability* 25 (1988), 287–298.
[37] S.F. Yashkov. 1992. Mathematical Problems in the Theory of Shared-processor Systems. *Journal of Soviet Mathematics* 58, 2 (1992), 101–147.

APPENDIX

A PROOF OF THEOREM 3.2

We will use the following definitions to explain the proofs:

$$\Delta_i^*(j,t,\nu) := \begin{cases} \overline{V}_i^*(j,t;\nu) - \overline{V}_i^*(j-1,t;\nu), & \text{if } j > 1, \\ \overline{V}_i^*(j,t;\nu), & \text{if } j = 1, \end{cases} \tag{29}$$

$$\Delta_{i,\beta}^*(j,t,\nu) := \begin{cases} \overline{V}_{i,\beta}^*(j,t;\nu) - \overline{V}_{i,\beta}^*(j-1,t;\nu), & \text{if } j > 1, \\ \overline{V}_{i,\beta}^*(j,t;\nu), & \text{if } j = 1. \end{cases} \tag{30}$$

We use the following two important lemmas which are proved in Sec. B to prove Thm. 4.1.

Lemma A.1. *For any user i, $j \in \{1, 2, \ldots, j_i\}$, $t \geq 0$, and $\nu > 0$, we have that $\Delta_i^*(j,t,\nu) > \frac{\nu}{\mu_i r_{i,1}}$.*

Lemma A.2. *For any user i, $j \in \{1, 2, \ldots, j_i\}$, and $t \geq 0$, we have that*

(1) $\Delta_i^*(j,t,\nu)$ is an non-decreasing concave function of ν and the following equation has a fixed point:

$$\mu_i r_{i,l} \Delta_i^*(j,t,\nu) = \nu \quad l \in \{2, 3, \ldots, L\} \tag{31}$$

(2) $\Delta_i^*(j,t,0) > 0$.

For $r \in \{r_{i,1}, r_{i,2}, \ldots, r_{i,L}\}$, and a fixed j and t, let us look at the fixed point of $\mu_i r \Delta_i^*(j,t,\nu)$, i.e., the solution to the following equation:

$$\nu = \mu_i r \Delta_i^*(j,t,\nu). \tag{32}$$

By Lemma A.1 and the fact that $\Delta_i^*(j,t,\nu)$ is continuous in ν, there does not exist a fixed point, i.e., solution to $\mu_i r \Delta_i^*(j,t,\nu) = \nu$ when $r = r_{i,1}$ and for any finite ν we have that $\nu < \mu_i r_{i,1} \Delta_i^*(j,t,\nu)$. From Bellman equation (7), this is implies that it is always optimal to transmit when $r = r_{i,1}$ for any $\nu < \infty$. Hence, $\nu_i^*(j, r_{i,1}, t) = \infty$.

Property 1 in Lemma A.2 shows that there exists a fixed point for $\mu_i r \Delta_i^*(j,t,\nu)$ when $r = r_{i,l}$, $l = 2, 3, \ldots, L$. Let us choose any such fixed point as the Whittle's index denoted by $\nu_i^*(j, r_{i,l}, t)$. For $\nu < \nu_i^*(j, r_{i,l}, t)$, from properties 1 and 2 in Lemma A.2, $\mu_i r_{i,l} \Delta_i^*(j,t,\nu) \geq \nu$. Therefore, from the Bellman equations (7) it is optimal to transmit in $(j, r_{i,l}, t)$. Similarly, for $\nu > \nu_i^*(j, r_{i,l}, t)$, it is optimal not to transmit in $(j, r_{i,l}, t)$. Thus we conclude that the problem is indexable for the multi-level i.i.d. service rate model.

B PROOF OF LEMMAS

B.1 Proof of Lemma A.1

We will prove this inequality by contradiction. Suppose that the inequality is not true. From the Bellman equations (7), this would imply that it is not optimal to transmit in states $(j, r_{i,1}, t)$ $l \in \{1, 2, \ldots, L\}$. From this we get the following:

$$\overline{V}_i^*(j,t;\nu) = c_i(t) + \overline{V}_i^*(j,t+1;\nu). \tag{33}$$

By Assumption 2.1 that for any t, there exists a t' such that $t' > t$ and $c_i(t') > 0$, it can be easily shown that $\overline{V}_i^*(j,t;\nu) < \overline{V}_i^*(j,t+1;\nu)$. However, (33) implies a contradiction. Hence, the inequality $\Delta_i^*(j,t,\nu) > \frac{\nu}{\mu_i r_{i,1}}$ must be true.

B.2 Proof of Lemma A.2

We will use the following two intermediate lemmas proved in Sec. C to prove Lemma A.2.

LEMMA B.1. *Let the truncated holding cost function for user i is defined as follows:*

$$c_i^{(k)}(t) := \begin{cases} c_i(t), & \text{if} \quad t \le k, \\ c_i(k), & \text{if} \quad t > k. \end{cases} \tag{34}$$

Let $\overline{V}_i^{*,(k)}(j,t;v)$ be the corresponding averaged optimal value function under the cost function $c_i^{(k)}(\cdot)$, then for all j, t, and v we have that

$$\lim_{k\to\infty} \overline{V}_i^{*,(k)}(j,t;v) = \overline{V}_i^{*}(j,t;v)\,. \tag{35}$$

LEMMA B.2. *If the cost function of user i is constant in time, i.e., $c_i(t) = c$, then under the multi-state channel model we have that $\Delta_i^*(j,t,v)$ is independent of j and t and is a concave, non-decreasing, piecewise linear function of v.*

The proof of Lemma A.2 is as follows. First we shall prove the non-decreasing property of $\Delta_i^*(j,t,v)$ with respect to v.

1) ***Non-decreasing:*** First we shall prove the non-decreasing property of $\Delta_i^*(j,t,v)$. To that end we will approximate $c(t)$ with a sequence of truncated holding cost functions $\{c_i^{(k)}(t), k = 1,2,3,\ldots\}$ as defined in (34). Let us define $\Delta_i^{*,(k)}(j,t,v) := \overline{V}_i^{*,(k)}(j,t;v) - \overline{V}_i^{*,(k)}(j-1,t;v)$. We will show that $\Delta_i^{*,(k)}(j,t,v)$ is a non-decreasing function of v and use Lemma B.1 to conclude that $\Delta_i^*(j,t,v)$ is also a non-decreasing function of v.

$c_i^{(k)}(\cdot)$ is a 'truncated' approximation of the holding cost function, in which the holding cost has a constant value of $c_i(k)$ after time k. Since the holding cost function is fixed after time k, the policy in the state (j,r,t') for any $t' > k$ is the same. Also, $\overline{V}_i^{*,(k)}(j,t;v)$ depends only on the actions in other states (j',r',t') such that $j' \le j$ and $t' \ge t$. Because of this we have to consider a finite number of feasible policies and the decisions that have to be made over time interval $[0,k]$.

Let $\pi^*\left(c_i^{(k)}(\cdot), v\right)$ be the optimal policy when the price is v and the holding cost function is $c_i^{(k)}(\cdot)$. If we fix a policy π, then the overall average cumulative holding cost from the state (j,r,t), denoted by $\overline{V}_i^{\pi,k}(j,t;v)$ is a linear function of v. Therefore, to find $\overline{V}_i^{*,(k)}(j,t;v)$, we are taking a minimum over a finite number of linear functions in v when the cost functions is $c_i^{(k)}(\cdot)$. This implies that $\overline{V}_i^{*,(k)}(j,t;v)$ is a piece-wise linear function in v and is concave. Therefore, for any v, there exists a *neighborhood* $N_\delta(v)$ where the policy $\pi^*\left(c_i^{(k)}(\cdot), v\right)$ is optimal. When we say neighborhood, we mean any of the three sets: $(v-\delta, v]$, $(v-\delta, v+\delta)$, or $[v, v+\delta)$, where $\delta > 0$. Next we state an important lemma which is proved in Sec. C.

LEMMA B.3. *$\Delta_i^{*,(k)}(j,t,v)$ is non-decreasing function of v in $N_\delta(v)$.*

Since $\Delta_i^{*,(k)}(j,t,v)$ is continuous in v and piece-wise linear function, the above lemma implies that $\Delta_i^{*,(k)}(j,t,v)$ is a non-decreasing function of v. Therefore, $\lim_{k\to\infty} \Delta_i^{*,(k)}(j,t,v) = \Delta_i^*(j,t,v)$ is also a non-decreasing function of v.

Concavity: Next we shall prove the concavity of $\Delta_i^*(j,t,v)$. We shall use truncated holding cost functions to prove this property. We shall prove that $\Delta_i^{*,(k)}(j,t,v)$ is concave in v. Using the fact that concavity is preserved on taking the limit $\lim_{k\to\infty} \Delta_i^{*,(k)}(j,t,v)$ we will conclude that

$\Delta_i^*(j,t,\nu)$ is concave in ν. We shall use prove the concavity of $\Delta_i^{*,(k)}(j,t,\nu)$ by induction. Let us assume that $t \le k$.

Base Case: For $t' \ge k$, we have that $\Delta_i^{*,(k)}(j,t',\nu)$ is a concave function of ν $\forall$ i and j. This is proved in Lemma B.2.

Induction Hypothesis: Let us assume that for any user i, $\Delta_i^{*,(k)}(j,t',\nu)$ is a concave function ν for $t+1 \le t' < k$.

We have to prove that $\Delta_i^{*,(k)}(j,t,\nu)$ is a concave function of ν $\forall$ j and k. We can re-write $\Delta_i^{*,(k)}(j,t,\nu)$ as follows:

$$\Delta_i^{*,(k)}(j,t,\nu) = \Delta_i^{*,(k)}(j,t+1,\nu) + \mathbb{E}\left[\min\left\{0, \nu - \mu_i R_i \Delta_i^{*,(k)}(j,t+1,\nu)\right\}\right] - \mathbb{E}\left[\min\left\{0, \nu - \mu_i R_i \Delta_i^{*,(k)}(j-1,t+1,\nu)\right\}\right], \quad (36)$$

where the expectation is computed with respect to R_i which is a r.v. with the same distribution as $R_i(t)$. Define

$$\tilde{l} := \max\left\{l : \nu \le \mu_i r_{i,l} \Delta_i^{*,(k)}(j,t+1,\nu)\right\}.$$

From Lemma A.1, $\tilde{l} \ge 1$. Therefore, the first two terms in the R.H.S. of (36) sum upto $\nu + \left(1 - \mu_i \sum_{l=1}^{\tilde{l}} q_{i,l} r_{i,l}\right) \Delta_i^{*,(k)}(j,t+1,\nu)$, which is a concave function of ν from the induction hypothesis. Similarly one can argue that the third term in the R.H.S. of (36) is also a concave function of ν. Since sum of concave functions is a concave function, $\Delta_i^{*,(k)}(j,t,\nu)$ is also a concave function. Therefore, from Lemma B.1, $\Delta_i^*(j,t,\nu)$ is also concave in ν.

To prove that (32) has a fixed point, we will have to show that curves $\mu_i r_{i,l} \Delta_i^*(j,t,\nu)$ as a function of ν and the linear function ν intersect when $l \ne 1$. For this we derive an upper bound on $\Delta_i^*(j,t,\nu)$. If we use the optimal policy when starting with $j-1$ stages at time t for the first $j-1$ phases when starting with j phases at time t, we will get an upper bound for $\overline{V}_i^*(j,t;\nu)$ which is given below:

$$\overline{V}_i^*(j,t;\nu) \le \mathbb{E}\left[\overline{V}_i^*(1, T(j-1,t,j-1;\nu);\nu)\right] + \overline{V}_i^*(j-1,t;\nu), \quad (37)$$

where $\mathbb{E}\left[\overline{V}_i^*(1, T(j-1,t,j-1;\nu);\nu)\right]$ is the average cumulative cost to finish one remaining phase if the time taken to finish the first $j-1$ phases is $T(j-1,t,j-1;\nu)$. Using this we can re-write $\Delta_i^*(j,t,\nu)$ as follows:

$$\Delta_i^*(j,t,\nu) = \overline{V}_i^*(j,t;\nu) - \overline{V}_i^*(j-1,t;\nu) \quad (38)$$
$$\le \mathbb{E}\left[\overline{V}_i^*(1, T(j-1,t,j-1;\nu);\nu)\right]. \quad (39)$$

We can bound the term the R.H.S. of the above equation with the average cumulative cost under the policy in which we transmit only when $R_i(t) = r_{i,1}$. We get the following:

$$\mathbb{E}\left[\overline{V}_i^*(1, T(j-1,t,j-1;\nu);\nu)\right] \le \mathbb{E}\left[H_i^\dagger(j, T(j-1,t,j-1;\nu))\right] + \frac{\nu}{\mu_i r_{i,1}}, \quad (40)$$

where $H_i^\dagger(j,t)$ is the cumulative average holding cost under the policy which transmits only when $R_i(t) = r_{i,1}$. Under this policy, the probability of success of completing a phase given that the user i transmits is given by $\mu_i r_{i,1}$. Hence, the average transmission cost is given by $\frac{\nu}{\mu_i r_{i,1}}$. The expectations in the above expression are all with the respect to the r.v. $T(j-1,t,j-1;\nu)$. So we have that

$$\Delta_i^*(j,t,\nu) \le \mathbb{E}\left[H_i^\dagger(j, T(j-1,t,j-1;\nu))\right] + \frac{\nu}{\mu_i r_{i,1}}. \quad (41)$$

Let $T_i^{\dagger}(j-1)$ be a r.v. denoting the time taken to finish $j-1$ stages under the policy in which transmits only when $R_i(t) = r_{i,1}$. Since it is optimal to transmit $R_i(t) = r_{i,1}$, we have that $T_i^{\dagger}(j-1) \overset{s.t.}{>} T(j-1, t, j-1; \nu)$. Since $\mathbb{E}\left[H_i^{\dagger}(j,t)\right]$ is a non-decreasing function of t, we have a further bound on $\Delta_i^*(j,t,\nu)$ and is given below:

$$\Delta_i^*(j,t,\nu) \le \mathbb{E}\left[H_i^{\dagger}\left(j, T_i^{\dagger}(j-1)\right)\right] + \frac{\nu}{\mu_i r_{i,1}}. \tag{42}$$

Therefore, $\Delta_i^*(j,t,\nu)$ is a concave, non-decreasing function of ν which is upper bounded by an affine function of ν with slope $1/\mu_i r_{i,1}$. This implies that for $l \neq 1$, $\mu_i r_{i,l} \Delta_i^*(j,t,\nu)$ is upper bounded by a function of ν with slope strictly less than one since $\frac{\mu_i r_{i,l}}{\mu_i r_{i,1}} < 1$. Hence, $\mu_i r_{i,1} \Delta_i^*(j,t,\nu)$ should intersect with ν and therefore, there exists a fixed point. Hence, this part of the lemma is proved.

2) When $\nu = 0$, it is optimal to transmit in all states. Therefore, the average cumulative cost includes only the holding cost component. $\Delta_i^*(j,t,0) = H_i^*(j,t,0) - H_i^*(j-1,t,0)$. The average cumulative cost to finish j phases is more than the cost to finish $j-1$ phases if we transmit in all states, and hence, $\Delta_i^*(j,t,0) > 0$.

C PROOF OF AUXILIARY LEMMAS: INDEXIBILTY

C.1 Proof of Lemma B.1

Let us consider $\left|\overline{V}_i^{*,(k)}(j,t;\nu) - \overline{V}_i^*(j,t;\nu)\right|$. Let us also consider $t \le k$. This is not a restrictive assumption as we would be taking the limit $k \longrightarrow \infty$ for a fixed t in the sequel. First we will find an upper bound on the term $\left|\overline{V}_i^{*,(k)}(j,t;\nu) - \overline{V}_i^*(j,t;\nu)\right|$. Let $\pi^*(c_i(\cdot),\nu)$ be the optimal policy when the cost function is $c_i(\cdot)$. Similarly, $\pi^*\left(c_i^{(k)}(\cdot),\nu\right)$ be the optimal policy when the cost function is $c_i^{(k)}(\cdot)$. To get an upper bound we shall use the following hybrid policy which combines both $\pi^*(c_i(\cdot),\nu)$ and $\pi^*\left(c_i^{(k)}(\cdot),\nu\right)$

- For $t \le k$, use $\pi^*(c_i(\cdot),\nu)$.
- For $t > k$, use $\pi^*\left(c_i^{(k)}(\cdot),\nu\right)$.

This policy is clearly sub-optimal for $c_i^{(k)}(\cdot)$ and hence, the average cumulative holding cost under this hybrid policy will be an upper bound on $\overline{V}_i^{*,(k)}(j,t;\nu)$. Let the total cost under this policy be denoted by $V_i^{h,(k)}(j,t;\nu)$.

We shall use a coupling argument next. Let us consider two systems, one which uses the hybrid policy with holding cost function $c_i^{(k)}(\cdot)$ and the other with $\pi^*(c_i(\cdot),\nu)$ and holding cost function $c_i(\cdot)$. Let us couple the job size random variables and the channel state process. Let us consider two mutually exclusive and exhaustive events 1) user i is served to completion before slot k 2) user i is served to completion after slot k. Conditioned on event 1, for any sample path, the difference between the cumulative cost of both the systems is zero. This is because, the policies are same and the holding are also the same for $t \le k$. Let us look at event 2. From lemma A.1 and Bellman equations (7), it is always to optimal to transmit when $R_i(t) = r_{i,1}$ $\forall$ t. Event 2 happens only if there less than j phases are successfully completed in $k-t$ slots. Therefore, probability of event 2 is upper bounded by $\sum_{j'=0}^{j} \binom{k-t}{j'} \left(q_{i,1}\mu_i r_{i,1}\right)^{j'} \left(1 - q_{i,1}\mu_i r_{i,1}\right)^{k-t-j'}$. If event 2 occurs, then there will be non-zero residual phases that has to be served after slot k. We can bound this cost by taking $\max_{j'' \le j} V_i^{h,(k)}\left(j'',k;\nu\right) - \overline{V}_i^*\left(j'',k;\nu\right)$. From the above discussion we have the following inequalities:

$$\overline{V}_i^{*,(k)}(j,t;\nu) - \overline{V}_i^{*}(j,t;\nu) \le V_i^{h,(k)}(j,t;\nu) - \overline{V}_i^{*}(j,t;\nu) \tag{43}$$

$$\le \sum_{j'=0}^{j} \binom{k-t}{j'} (\tilde{p}_i)^{j'} (1-\tilde{p}_i)^{k-t-j'} \tag{44}$$

$$\times \max_{j'' \le j} \left[V_i^{h,(k)}\left(j'',k;\nu\right) - \overline{V}_i^{*}\left(j'',k;\nu\right) \right], \tag{45}$$

where $\tilde{p}_i := q_{i,1}\mu_i r_{i,1}$. Since we have assumed that the holding cost functions are upper bounded by polynomials, the term $V_i^{h,(k)}(j'',k;\nu) - \overline{V}_i^{*}(j'',k;\nu)$ is a polynomial function of k. This is because the under $c_i^{(k)}(\cdot)$, holding cost is a constant $c_i(k)$ for $t \ge k$, and the average holding cost to complete any phase is just scaling an appropriate geometric random variable with $c_i(k)$. Note that this term is multiplied by an exponentially decaying function of k in (44). Therefore, on taking the limit $k \longrightarrow \infty$, the R. H. S. goes to zero. Hence, we have shown that the upper bound goes to zero. We can derive a lower bound for $\overline{V}_i^{*,(k)}(j,t;\nu) - \overline{V}_i^{*}(j,t;\nu)$ in a similar manner by interchanging the roles of $\pi^*(c_i(\cdot),\nu)$ and $\pi^*\left(c_i^{(k)}(\cdot),\nu\right)$ in the construction of hybrid policy and then using that to upper bound $\overline{V}_i^{*}(j,t;\nu)$. We shall skip the details in the interest of space. Therefore, we have that $\lim_{k\to\infty}\left|\overline{V}_i^{*,(k)}(j,t;\nu) - \overline{V}_i^{*}(j,t;\nu)\right| = 0$.

C.2 Proof of Lemma B.2

Suppose if we have that $\forall t\ c_i(t) = c$, then it should be clear that $\Delta_i^*(j,t,\nu)$ is independent of t. To study the effect of j, from the definition of $\Delta_i^*(j,t,\nu)$ we can write the following equation:

$$\Delta_i^*(j,t,\nu) = \Delta_i^*(j,t+1,\nu) + \mathbb{E}\left[\min\left\{0, \nu - \mu_i R_i \Delta_i^*(j,t+1,\nu)\right\}\right] - \mathbb{E}\left[\min\left\{0, \nu - \mu_i R_i \Delta_i^*(j-1,t+1,\nu)\right\}\right], \tag{46}$$

where R_i is a r.v. denoting the random service rate in a typical slot. Since $\Delta_i^*(j,t,\nu)$ is independent of t under constant holding cost assumption, we shall suppress the argument t in the sequel. Then the above equation simplifies to the following:

$$\mathbb{E}\left[\min\left\{0, \nu - \mu_i R_i \Delta_i^*(j,\nu)\right\}\right] = \mathbb{E}\left[\min\left\{0, \nu - \mu_i R_i \Delta_i^*(j-1,\nu)\right\}\right]. \tag{47}$$

Since the above equation holds for any service rate distribution, we have that $\Delta_i^*(j;\nu)$ must be independent of j. Therefore, we can re-write $\Delta_i^*(j;\nu)$ in the following manner:

$$\Delta_i^*(j,\nu) = \Delta_i^*(1,\nu) = \overline{V}_i^{*}(1;\nu). \tag{48}$$

From Bellman equations (7), if it is optimal to transmit in $R_i(t) = r_{i,l}$, then it is also optimal to transmit when $R_i(t) = r_{i,l'}$ for $l' < l$. We shall restrict ourselves to such policies. Let π be a policy where we transmit when $R_i(t) = r_{i,l'}$ for $l' = 1, 2, \ldots, l$. The average cumulative cost under such a policy is given by:

$$\overline{V}_i^{\pi}(1;\nu) = \frac{c}{\mu_i \sum_{l'=1}^{l} q_{i,l'} r_{i,l'}} + \frac{\nu \sum_{l'=1}^{l} q_{i,l'}}{\mu_i \sum_{l'=1}^{l} q_{i,l'} r_{i,l'}}. \tag{49}$$

This is because of probability of transmitting in a slot is $\mu_i \sum_{l'=1}^{l} q_{i,l'} r_{i,l'}$, and therefore the number of slots required to complete a phase on an average is $\frac{1}{\mu_i \sum_{l'=1}^{l} q_{i,l'} r_{i,l'}}$. Given that one transmits, probability of succeeding in completing a phase in a given slot is given by $\frac{\sum_{l'=1}^{l} q_{i,l'} r_{i,l'}}{\sum_{l'=1}^{l} q_{i,l'}}$. This is because the average rate conditioned on the fact that user i transmits is $\frac{\mu_i \sum_{l'=1}^{l} q_{i,l'} r_{i,l'}}{\sum_{l'=1}^{l} q_{i,l'}}$. Therefore, for any ν, to determine the optimal cost to go, we need only to take a minimum over a finite number of policies parametrized by $l = 1, 2, \ldots, L$. For each policy, the average cumulative cost is a non-decreasing linear function of ν. Therefore, from (48) $\Delta_i^*(j, \nu)$ is a non-decreasing, piecewise linear, concave function of ν.

C.3 Proof of Lemma B.3

Let $Y_i^{*,(k)}(t)$ be an r.v. denoting the residual number of phases of user i at time t. We can write $\overline{V}_i^*(j, t; \nu)$ as follows:

$$\overline{V}_i^{*,(k)}(j, t; \nu) = H_i^{*,(k)}(j, t, \nu) + \nu \mathbb{E}^{\pi^*\left(c_i^{(k)}(\cdot), \nu\right)}\left[\sum_{t'=t}^{\infty} A_i(t') | Y_i^{*,(k)}(t) = j\right], \tag{50}$$

where $H_i^{*,(k)}(j, t, \nu)$ is the average cumulative holding cost starting with j phases at time t and the second term is the average cumulative transmission cost incurred due to transmissions under the policy $\pi^*\left(c_i^{(k)}(\cdot), \nu\right)$. Therefore, we can re-write $\Delta_i^{*,(k)}(j, t, \nu)$ as follows:

$$\begin{aligned}\Delta_i^{*,(k)}(j, t, \nu) = H_i^{*,(k)}(j, t, \nu) - H_i^{*,(k)}(j-1, t, \nu) + \nu \Bigg(&\mathbb{E}^{\pi^*\left(c_i^{(k)}(\cdot), \nu\right)}\left[\sum_{t'=t}^{\infty} A_i(t') | Y_i^{*,(k)}(t) = j\right] \\ &- \mathbb{E}^{\pi^*\left(c_i^{(k)}(\cdot), \nu\right)}\left[\sum_{t'=t}^{\infty} A_i(t') | Y_i^{*,(k)}(t) = j-1\right]\Bigg).\end{aligned} \tag{51}$$

Since the optimal policy is same for all $\nu \in N_\delta(\nu)$, the term $H_i^{*,(k)}(j, t, \nu) - H_i^{*,(k)}(j-1, t, \nu)$ is independent of ν for $\nu \in N_\delta(\nu)$. If we can show that the slope of second term with respect to ν is greater than zero, then we can prove this lemma. To that end let us define $T(j, t, k; \nu)$ to be the random variable denoting the time to complete first k phases starting with j phases at time t, when the price is ν, under the optimal policy.

First we show that $T(j, t, j-1; \nu) \overset{s.t.}{\leq} T(j-1, t, j-1; \nu)$, i.e., the time to complete the first $j-1$ phases when starting with j phases at time t is stochastically less than the time to complete $j-1$ phases when starting with $j-1$ phases at time t. To see this, we can re-write $\overline{V}_i^{*,(k)}(j, t; \nu)$ as

$$\begin{aligned}\overline{V}_i^{*,(k)}(j, t; \nu) = &\text{ Average cumulative cost to finish first } j-1 \text{ phases} \\ &+ \text{Average cumulative cost to finish the last phase.}\end{aligned} \tag{52}$$

Individually each of the two terms on the R.H.S. above consists of a part due to the holding cost and a part due to the transmission cost ν. Also, note the two terms in the R.H.S. are not independent of each other. If the time to complete to first $j-1$ phases is longer, then the average cumulative holding cost in completing the last phase is also higher because the transmission of the last phase starts at a later time and the holding cost function is non-deceasing function of time. If $T(j, t, j-1; \nu) \overset{s.t.}{>} T(j-1, t, j-1; \nu)$, then we can replace the policy for the first $j-1$ phases when starting with j phases with the optimal policy for $j-1$ stages when starting with $j-1$ stages and therefore, we can obtain a better policy. Hence, $T(j, t, j-1; \nu) \overset{s.t.}{\leq} T(j-1, t, j-1; \nu)$.

Next observe that the average cumulative cost in completing $j-1$ phases starting with $j-1$ phases initially has to be less than the average cumulative cost in completing $j-1$ phases when starting with j phases. $T(j, t, j-1; \nu) \overset{s.t.}{\leq} T(j-1, t, j-1; \nu)$ would imply that the average cumulative holding cost in completing the first $j-1$ phases when starting with j phases is less than the average cumulative holding cost in completing $j-1$ phases when starting with $j-1$ phases. The only way that the average cumulative cost to complete the $j-1$ phases when starting with j phases is more than the average cumulative cost in completing $j-1$ phases when starting with $j-1$ phases is by having a larger average cumulative transmission cost. This would imply that the slope of the R.H.S. of (51) is positive with respect to ν. Hence, the Lemma B.3 is proven.

D PROOF OF THEOREM 4.1

In order to find the Whittle's index for any state (j,r,t), we have to find the fixed point of the following equation:

$$\nu = \mu_i r \Delta_i^* (j,t,\nu). \tag{53}$$

We have already shown in the Appendix A that when $r = r_{i,1}$ there does not exist a finite fixed point for the above equation and the $\nu_i^* \left(j, r_{i,1}, t\right) = \infty$. We have also shown that there exists a finite fixed point when $r \neq r_{i,1}$, and therefore, for $l \neq 1$, $\nu_i^* \left(j, r_{i,l}, t\right) < \infty$. Hence, proved.

E SECONDARY INDEX

E.1 Proof of Theorem 4.3

Consider the discounted sub-problem $\mathcal{SP}_i^{\beta}$. From the definition of Whittle's index for the discounted case, to find $\nu_{i,\beta}^* \left(j, r_{i,1}, t\right)$, we have to find the supremum of the fixed points of the following equation

$$\nu = \mu_i r_{i,1} \beta \Delta_{i,\beta}^* (j,t,\nu). \tag{54}$$

The supremum of the fixed points of the above equation is finite because of the following reasons

(1) When $\nu = 0$, it is optimal to transmit in all states and $\Delta_{i,\beta}^* (j,t,0) > 0$.
(2) When $\nu \to \infty$, it is optimal not to transmit in any of the states, and $\lim_{\nu\to\infty} \Delta_{i,\beta}^* (j,t,\nu) = 0$. This is because if it is not optimal to transmit in any of the states, then only average cumulative discounted holding cost is incurred. Therefore, $\overline{V}_{i,\beta}^* (j,t;\nu) = \sum_{k=t}^{\infty} c_i(k)\beta^k$, for any $j \in \{1,2,\ldots,j_i\}$. By our assumption that $c_i(t) < \delta t^{\zeta}$, we get that $\sum_{k=t}^{\infty} c_i(k)\beta^k < \infty$.
(3) We also know that $\overline{V}_{i,\beta}^* (j,t;\nu)$ is a continuous function of ν.

From the above observations and Intermediate Value Theorem, we can conclude that there exists at least a fixed point for (54), and we can find a supremum of the fixed points.

We know that $\lim_{\beta\to 1} \nu_{i,\beta}^* \left(j, r_{i,1}, t\right) = \nu_i^* \left(j, r_{i,1}, t\right) = \infty$. To the find the asymptote of $\nu_{i,\beta}^* \left(j, r_{i,1}, t\right)$ as $\beta \to 1$, we can use (54), since $\nu_{i,\beta}^* \left(j, r_{i,1}, t\right)$ is a fixed point of (54). To that end we will first study the characteristics of $\overline{V}_{i,\beta}^* (j,t;\nu)$ evaluated at $\nu = \nu_{i,\beta}^* \left(j, r_{i,1}, t\right)$ as $\beta \to 1$, which we denote by $\overline{V}_{i,\beta}^* \left(j,t;\nu_{i,\beta}^* \left(j, r_{i,1}, t\right)\right)$. We will show that the asymptote of $\overline{V}_{i,\beta}^* \left(j,t;\nu_{i,\beta}^* \left(j, r_{i,1}, t\right)\right)$ is same as that of a policy in which transmissions are always performed when $r = r_{i,1}$ and never performed otherwise. For any ν we can split the average cumulative cost into two, the average cumulative holding and transmission costs. Let $H_{i,\beta}^* (j,t,\nu)$ be the average cumulative holding cost under optimal policy starting from the phase j at time. Similarly, let the $N_{i,\beta}^* (j,t,\nu)$ be the cumulative discounted average number of transmissions under the optimal policy, i.e., $\mathbb{E}\left[\sum_{k=t}^{\infty} \beta^{k-t} A_{i,\beta}^*(k)\right]$, where $A_{i,\beta}^*(k) = 1$ if the optimal decision is to transmit in slot k and 0 otherwise. Therefore, the average cumulative cost is given by:

$$\overline{V}_{i,\beta}^* (j,t;\nu) = H_{i,\beta}^* (j,t,\nu) + \nu N_{i,\beta}^* (j,t,\nu). \tag{55}$$

Similarly we can define $N^{\dagger}_{i,\beta}(j,t)$ and $H^{\dagger}_{i,\beta}(j,t)$ for the policy in which transmission are done only if $r = r_{i,1}$ for all j and t. Note that $N^{\dagger}_{i,\beta}(j,t)$ and $H^{\dagger}_{i,\beta}(j,t)$ are independent of ν as the policy is fixed and does not change with ν. The average cumulative cost associated with this policy is thus given by:

$$\overline{V}^{\dagger}_{i,\beta}(j,t;\nu) = H^{\dagger}_{i,\beta}(j,t) + \nu N^{\dagger}_{i,\beta}(j,t). \tag{56}$$

The main result connecting the optimal policy for $\mathcal{SP}_i(\nu)$ and the policy with transmissions only in $r_{i,1}$ is given next. Proof of this lemma is given in Sec. F.1.

Lemma E.1. *Let $\overline{V}^{\dagger}_{i,\beta}(j,t;\nu)$ be the average cumulative cost starting from j and t for the policy in which transmissions are performed only when the channel is in the best possible state. We have that*

$$\frac{\lim_{\beta\to 1} H^{*}_{i,\beta}\left(j,t,\nu^{*}_{i,\beta}(j,r_{i,1},t)\right)}{\lim_{\beta\to 1} H^{\dagger}_{i,\beta}(j,t)} = 1, \tag{57}$$

$$\frac{\lim_{\beta\to 1} N^{*}_{i,\beta}\left(j,t,\nu^{*}_{i,\beta}(j,r_{i,1},t)\right)}{\lim_{\beta\to 1} N^{\dagger}_{i,\beta}(j,t)} = 1. \tag{58}$$

The above lemma proves that $\lim_{\beta\to 1} \overline{V}^{*}_{i,\beta}\left(j,t;\nu^{*}_{i,\beta}(j,r_{i,1},t)\right)$ and $\lim_{\beta\to 1} \overline{V}^{\dagger}_{i,\beta}\left(j,t;\nu^{*}_{i,\beta}(j,r_{i,1},t)\right)$ have same asymptotes, and hence, we can use the latter to find the asymptote of $\overline{V}^{*}_{i,\beta}\left(j,t;\nu^{*}_{i,\beta}(j,r_{i,1},t)\right)$. For $\overline{V}^{\dagger}_{i,\beta}\left(j,t;\nu^{*}_{i,\beta}(j,r_{i,1},t)\right)$ we can find closed form expressions as we know the structure of the policy.

First we will find an expression for $\nu^{*}_{i,\beta}(j,r_{i,1},t)$. Substituting (55) in (54) and noting that $\nu^{*}_{i,\beta}(j,r_{i,1},t)$ is a fixed point for (54), we get the following expression for $\nu^{*}_{i,\beta}(j,r_{i,1},t)$:

$$\nu^{*}_{i,\beta}(j,r_{i,1},t) = \frac{\mu_i r_{i,1}\beta\left[H^{*}_{i,\beta}\left(j,t+1,\nu^{*}_{i,\beta}(j,r_{i,1},t)\right) - H^{*}_{i,\beta}\left(j-1,t+1,\nu^{*}_{i,\beta}(j,r_{i,1},t)\right)\right]}{1-\mu_i r_{i,1}\beta\left[N^{*}_{i,\beta}\left(j,t+1,\nu^{*}_{i,\beta}(j,r_{i,1},t)\right) - N^{*}_{i,\beta}\left(j-1,t+1,\nu^{*}_{i,\beta}(j,r_{i,1},t)\right)\right]}. \tag{59}$$

Next we multiply both sides of (59) with $1-\beta$ and take the limit $\beta \to 1$ on both the sides. Using Lemma E.1, we can the replace the average cumulative costs related to the optimal policy with that of the policy in which transmissions are done only in $r = r_{i,1}$. Note that $N^{\dagger}_{i,\beta}(j,t)$ depends only on j and not on t. We have used this notation to maintain consistency. Further it can be shown that

$$\frac{(1-\beta)N^{\dagger}_{i,\beta}(j,t)}{q_{i,1}} = 1-\mu_i r_{i,1}\beta\left(N^{\dagger}_{i,\beta}(j,t) - N^{\dagger}_{i,\beta}(j-1,t)\right). \tag{60}$$

Substituting (60) in (59), re-arranging the terms, and using the fact that $\lim_{\beta\to 1} N^{\dagger}_{i,\beta}(j,t) = \frac{j}{\mu_i r_{i,1}}$, we get that

$$\xi^{*}_{i}(j,r_{i,1},t) = \lim_{\beta\to 1}(1-\beta)\nu^{*}_{i,\beta}(j,r_{i,1},t) = \frac{q_{i,1}\left(\mu_i r_{i,1}\right)^2}{j}\left[H^{\dagger}_{i,1}(j,t+1) - H^{\dagger}_{i,1}(j-1,t+1)\right]. \tag{61}$$

Due to Assumption 2.1 on $c_i(\cdot)$, it is bounded by a polynomial function of t. Therefore, the above expression is finite.

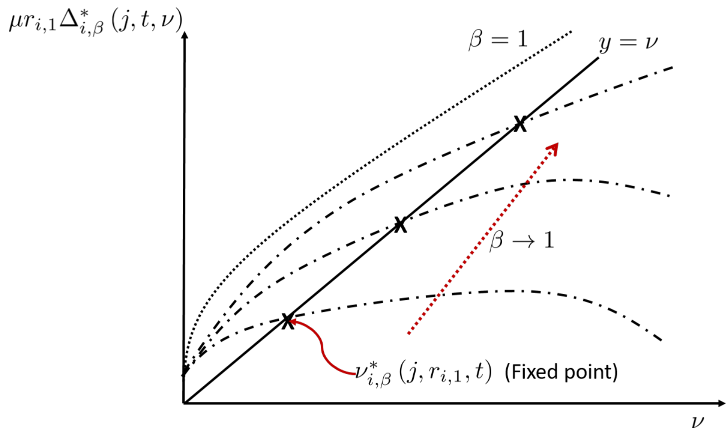

Fig. 12. Increasing β while setting $v = v^*_{i,\beta}\left(j, r_{i,1}, t\right)$ is illustrated here.

F PROOF OF AUXILIARY LEMMAS: SECONDARY INDEX

F.1 Proof of Lemma E.1

We have to the find the optimal policy when $\beta \longrightarrow 1$ while we set $v = v^*_{i,\beta}\left(j, r_{i,1}, t\right)$. This procedure is shown in the Fig. 12. In this proof, we shall show the following two properties of the optimal policy as $\beta \longrightarrow 1$, while $v = v^*_{i,\beta}\left(j, r_{i,1}, t\right)$:

(1) It is not optimal to transmit in $r = r_{i,l}$ when $l \neq 1$ for any j and t.
(2) It is always optimal to transmit in $r = r_{i,1}$ for j' and t' such that $\left(j', r_{i,1}, t'\right)$ is reachable from $\left(j, r_{i,1}, t\right)$

First we have the following result. Proof of the following lemma is given in Appendix F.2.

Lemma F.1. *For a given v, i, j, and t, $\Delta^*_{i,\beta}\left(j, t, v\right)$ is a non-decreasing function β.*

This would imply that $v^*_{i,\beta}\left(j, r_{i,l}, t\right)$ is a non-decreasing function of β. Hence, for any $\beta \in [0, 1]$ and $l \neq 1$, we have

$$v^*_{i,\beta}\left(j, r_{i,l}, t\right) \leq v^*_i\left(j, r_{i,l}, t\right) < \infty. \tag{62}$$

From the indexability property, if the price $v > v^*_{i,\beta}\left(j, r_{i,l}, t\right)$, it is not optimal to transmit in $\left(j, r_{i,l}, t\right)$. Let us take the limit $\beta \longrightarrow 1$ while $v = v^*_{i,\beta}\left(j, r_{i,1}, t\right)$. We know as $\beta \longrightarrow 1$, $v = v^*_{i,\beta}\left(j, r_{i,1}, t\right) \longrightarrow \infty$. We also know that as $\beta \longrightarrow 1$, $v^*_{i,\beta}\left(j, r_{i,l}, t\right) < \infty$. This implies that for any j and t there exists some $\beta'\left(j, r_{i,l}, t\right)$ such that for $\beta > \beta'\left(j, r_{i,l}, t\right)$, it is optimal not to transmit in $\left(j, r_{i,l}, t\right)$.

Now we have to show that it is optimal to transmit in when $r = r_{i,1}$ in all states *reachable* from $\left(j, r_{i,1}, t\right)$. We say that a state is reachable from $\left(j, r_{i,1}, t\right)$ if there exists a policy π such that there is a strictly positive probability of making a transition into that state in the future. The reachable states from $\left(j, r_{i,1}, t\right)$ is shown in the Fig. 3. Note that the transition probabilities permit only transition into states where $t > t'$, $j' \leq j$, and if it is in the region shown in the figure. This is because we can get only at most one successful transmission in a slot. The following lemma will help us characterize the optimal policy when β is increased to 1, such that $v = v^*_{i,\beta}\left(j, r_{i,1}, t\right)$. Proof of this lemma is given in the Appendix F.3.

Lemma F.2. *For large enough β, if it is optimal to transmit in $\left(j, r_{i,1}, t\right)$, then it is optimal to transmit in all states $\left(j', r_{i,1}, t'\right)$ such that $\left(j', r_{i,1}, t'\right)$ is reachable from $\left(j, r_{i,1}, t\right)$.*

The above lemma tells that if it is optimal to transmit when $r = r_{i,1}$ in any given time, then it is optimal to transmit in $r = r_{i,1}$ in all future times. If we choose $\nu = \nu^*_{i,\beta}\left(j, r_{i,1}, t\right)$, we know that it is optimal to transmit in $\left(j, r_{i,1}, t\right)$. Hence, it is optimal to transmit in all states in the future where $r = r_{i,1}$. Therefore, as $\beta \longrightarrow 1$ while $\nu = \nu^*_{i,\beta}\left(j, r_{i,1}, t\right)$, it is optimal to transmit when $r = r_{i,1}$ and not optimal to transmit when $r \neq r_{i,1}$. This completes the proof of this lemma.

F.2 Proof of Lemma F.1

We will show that this property holds for any $c_i^{(k)}(\cdot)$ and hence, in the limiting case too due to lemma B.1. We will first prove that $\Delta^{*,(k)}_{i,\beta}(j,t,\nu)$ is a non-decreasing function of β.

To prove the result for $c_i^{(k)}(\cdot)$, we will use induction over time which proceeds backwards from time k to t.

Base Case : We will first prove that $\Delta^{*,(k)}_{i,\beta}(j,t,\nu)$ is non-decreasing function of β for $t \geq k$. From the Bellman equations 7, we can re-write the value function as follows:

$$\overline{V}^{*,(k)}_{i,\beta}(j,t;\nu) \;=\; c_i^{(k)}(t) + \beta \overline{V}^{*,(k)}_{i,\beta}(j,t+1;\nu) + \mathbb{E}\left[\min\left\{0, \nu - \mu_i R_i \beta\left[\Delta^{*,(k)}_{i,\beta}(j,t+1,\nu)\right]\right\}\right], \quad (63)$$

where R_i has the same distribution as $R_i(t)$. Using the above form of $\overline{V}^{*}_{i,\beta}(j,t;\nu)$, we can re-write $\Delta^{*,(k)}_{i,\beta}(j,t,\nu)$ as follows:

$$\begin{aligned}\Delta^{*,(k)}_{i,\beta}(j,t,\nu) = \beta\Delta^{*,(k)}_{i,\beta}(j,t+1,\nu) + \mathbb{E}\left[\min\left\{0, \nu - \mu_i R_i \beta \Delta^{*,(k)}_{i,\beta}(j,t+1,\nu)\right\}\right] \\ - \mathbb{E}\left[\min\left\{0, \nu - \mu_i R_i \beta \Delta^{*,(k)}_{i,\beta}(j-1,t+1,\nu)\right\}\right]. \quad (64)\end{aligned}$$

We know that when the holding cost function $c_i^{(k)}(\cdot)$ has a constant value of $c_i(k)$ for $t \geq k$. Therefore, $\Delta^{*,(k)}_{i,\beta}(j,t,\nu) = \Delta^{*,(k)}_{i,\beta}(j,k,\nu)$ once $t \geq k$. Hence, substituting this in (64), we get that

$$\begin{aligned}(1-\beta)\,\Delta^{*,(k)}_{i,\beta}(j,k,\nu) - \mathbb{E}\left[\min\left\{0, \nu - \mu_i R_i \beta \Delta^{*,(k)}_{i,\beta}(j,k,\nu)\right\}\right] = \\ - \mathbb{E}\left[\min\left\{0, \nu - \mu_i R_i \beta \Delta^{*,(k)}_{i,\beta}(j-1,k,\nu)\right\}\right]. \quad (65)\end{aligned}$$

Using the above equation, we can argue that $\Delta^{*,(k)}_{i,\beta}(j,k,\nu)$ is an non-decreasing function of β. This is done via induction over j. If $j = 1$, then $\Delta^{*,(k)}_{i,\beta}(j,k,\nu) = \overline{V}^{*,(k)}_{i,\beta}(1,k;\nu)$. $\overline{V}^{*,(k)}_{i,\beta}(1,k;\nu)$ is an non-decreasing function of β because for any policy π, the average cumulative cost to complete (average cumulative holding cost + transmission cost) is a non-decreasing function of β and therefore, $\overline{V}^{*,(k)}_{i,\beta}(1,k;\nu)$, which is obtained by computing infemum of the cost under all policies, is also a non-decreasing function of β. If we assume the induction hypothesis that $\Delta^{*,(k)}_{i,\beta}(j,k,\nu)$ is a non-decreasing function of β till $j-1$, then from (65), it can be easily shown that $\Delta^{*,(k)}_{i,\beta}(j,k,\nu)$ is a non-decreasing function of β. Hence we have proved that $\Delta^{*,(k)}_{i,\beta}(j,k,\nu)$ is a non-decreasing function of β.

Induction Hypothesis: Assume that $\Delta^{*,(k)}_{i,\beta}(j,t',\nu)$ is a non-decreasing of β for any j and $t' \geq t+1$.

We have to show that $\Delta^{*,(k)}_{i,\beta}(j,t,\nu)$ is a non-decreasing function of β. Consider (64). Its R.H.S. is a non-decreasing function of β because of our induction assumption that $\Delta^{*,(k)}_{i,\beta}(j-1,t+1,\nu)$

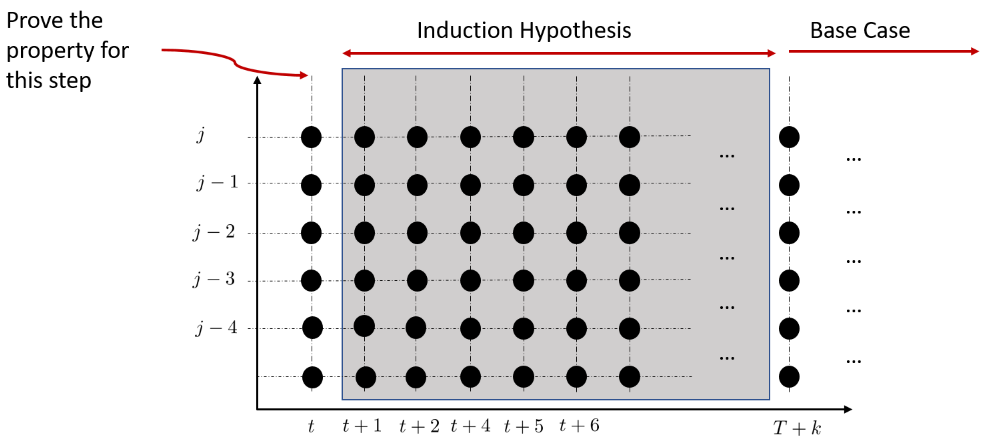

Fig. 13. Illustration of the induction procedure

and $\Delta^{*,(k)}_{i,\beta}(j-1,t+1,\nu)$ are non-decreasing functions of β. Therefore, $\Delta^{*,(k)}_{i,\beta}(j,t,\nu)$ is also a non-decreasing function of β. Hence, we have proved that $\Delta^{*,(k)}_{i,\beta}(j,t,\nu)$ is a non-decreasing function of β when the holding cost function is $c_i^{(k)}(\cdot)$. Therefore, on taking the limit as $k \longrightarrow \infty$, we get the result for $c_i(\cdot)$.

F.3 Proof of Lemma F.2

We will show that for large enough β, if it is optimal to transmit in $(j, r_{i,1}, t)$, then it is optimal to transmit in the states $(j, r_{i,1}, t+1)$ and $(j-1, r_{i,1}, t+1)$. This is enough to show that it is optimal to transmit in all states reachable from $(j, r_{i,1}, t)$ because we can iteratively use this result on the states $(j, r_{i,1}, t+1)$ and $(j-1, r_{i,1}, t+1)$ and its neighboring states and so on. We will prove this result for any $c_i^{(k)}(\cdot)$.

We have already argued that for large enough β (say $\beta > \beta'$), it is optimal not to transmit in $r_{i,l}$ $l \neq 1$ in all states reachable from $(j, r_{i,1}, t)$ if the price is scaled such that $\nu = \nu^*_{i,\beta}(j, r_{i,1}, t)$. Let us assume that β is large enough that it is optimal not to transmit in $r_{i,1}$ for all states reachable from $(j, r_{i,1}, t)$. Note that if we transmit in $(j, r_{i,1}, t)$, then it must be optimal to transmit in either $(j, r_{i,1}, t+1)$ or $(j-1, r_{i,1}, t+1)$. Else, it is optimal not to transmit in $(j, r_{i,1}, t)$, and instead transmit in the state $(j, r_{i,1}, t+1)$ incurring only the discounted cost $\beta\nu$. Next we have to show that it is optimal to transmit in both $(j, r_{i,1}, t+1)$ and $(j-1, r_{i,1}, t+1)$. We will prove this as two separate cases. The induction process is illustrated in the Fig. 13.

Base Case: We have to prove that if it is optimal to transmit in the state $(j, r_{i,1}, k)$, then it is optimal to transmit in the states $(j, r_{i,1}, k+1)$ and $(j-1, r_{i,1}, t+1)$. If it is optimal to transmit in $(j, r_{i,1}, k)$, then from Bellman equations, we know the following:

$$\nu \leq \mu_i r_{i,1} \beta \Delta^{*,(k)}_{i,\beta}(j, k+1, \nu)\,. \tag{66}$$

We know that $\Delta^{*,(k)}_{i,\beta}(j,k+1,\nu) = \Delta^{*,(k)}_{i,\beta}(j,k+2,\nu)$ as the holding cost function has a constant value for $t \geq k$. Therefore, $\nu \leq \mu_i r_{i,1} \beta \Delta^{*,(k)}_{i,\beta}(j,k+2,\nu)$. From Bellman equations, this would imply that it is optimal to transmit in $(j, r_{i,1}, k+1)$. Hence, base case is proved.

Induction Hypothesis: We shall assume that if $t+1 \leq t' \leq k$ and if it is optimal to transmit in $(j, r_{i,1}, t')$, then it is optimal to transmit in $(j, r_{i,1}, t'+1)$ and $(j-1, r_{i,1}, t'+1)$.

Using the induction hypothesis we will have to show that if it is optimal to transmit in $(j, r_{i,1}, t)$, then it is optimal to transmit in $(j, r_{i,1}, t+1)$ and $(j-1, r_{i,1}, t+1)$. We will consider two separate cases:

(1) If it is optimal to transmit in $(j, r_{i,1}, t)$ and $(j, r_{i,1}, t+1)$, then it is optimal to transmit in $(j-1, r_{i,1}, t+1)$.
(2) If it is optimal to transmit in $(j, r_{i,1}, t)$ and $(j-1, r_{i,1}, t+1)$, then it is optimal to transmit in $(j, r_{i,1}, t+1)$.

We will prove the above two cases separately via proof by contradiction.

Case 1. Suppose it is optimal to transmit in $(j, r_{i,1}, t)$ and $(j, r_{i,1}, t+1)$, and it is not optimal to transmit in $(j-1, r_{i,1}, t+1)$. Let us also assume that $j \geq 2$. From our induction hypothesis and Bellman equations the following is true for $t+1 \leq t' \leq k$:

$$\overline{V}^{*,(k)}_{i,\beta}(j, t'; \nu) - \overline{V}^{*,(k)}_{i,\beta}(j-1, t'; \nu) \leq \overline{V}^{*,(k)}_{i,\beta}(j-1, t'+1; \nu) - \overline{V}^{*}_{i,\beta}(j-2, t'+1; \nu). \tag{67}$$

The above equation is true because of the induction hypothesis that if it is optimal to transmit in $(j, r_{i,1}, t')$, then it is optimal to transmit in $(j-1, r_{i,1}, t'+1)$. First observe that if it is optimal to transmit in $(j, r_{i,1}, t)$, then from Bellman equations we get the following:

$$\nu \leq \mu_i r_{i,1} \beta \left(\overline{V}^{*,(k)}_{i,\beta}(j, t+1; \nu) - \overline{V}^{*,(k)}_{i,\beta}(j-1, t+1; \nu) \right). \tag{68}$$

Since we have assumed that it is optimal to transmit in $(j, r_{i,1}, t+1)$, we have the following:

$$\begin{aligned} \overline{V}^{*,(k)}_{i,\beta}(j, t+1; \nu) = c^{(k)}_i(t+1) + q_{i,1}\nu + \left(1 - \mu_i q_{i,1} r_{i,1}\right) \beta \overline{V}^{*,(k)}_{i,\beta}(j, t+2; \nu) \\ + \mu_i q_{i,1} r_{i,1} \beta \overline{V}^{*,(k)}_{i,\beta}(j-1, t+2; \nu). \end{aligned} \tag{69}$$

Similary, since it is not optimal to transmit in $(j-1, r_{i,1}, t+1)$, then we have that

$$\overline{V}^{*,(k)}_{i,\beta}(j-1, t+1; \nu) = c^{(k)}_i(t+1) + \beta \overline{V}^{*,(k)}_{i,\beta}(j-1, t+2; \nu). \tag{70}$$

Substituting (69) and (70) in (68), we get the following inequality:

$$\nu \leq \frac{\beta\left(1 - \mu_i q_{i,1} r_{i,1}\right)}{1 - \mu_i q_{i,1} r_{i,1}} \left[\mu_i r_{i,1} \beta \left(\overline{V}^{*,(k)}_{i,\beta}(j, t+2; \nu) - \overline{V}^{*,(k)}_{i,\beta}(j-1, t+2; \nu) \right) \right]. \tag{71}$$

Now let us look at the state $(j-1, r_{i,1}, t+1)$. Since it is not optimal to transmit in this state, from Bellman equations, we will get the following inequality:

$$\nu > \mu_i r_{i,1} \beta \left(\overline{V}^{*,(k)}_{i,\beta}(j-1, t+2; \nu) - \overline{V}^{*,(k)}_{i,\beta}(j-2, t+2; \nu) \right). \tag{72}$$

We will expand the terms in the R.H.S. of the above inequality. From our induction hypothesis, the states in which it is optimal transmit is shown in the Fig. 14. This includes all states reachable from $(j, r_{i,1}, t+1)$. This would imply that it is optimal to transmit in $(j-1, r_{i,1}, t+2)$. This will give us the following equation:

$$\begin{aligned} \overline{V}^{*,(k)}_{i,\beta}(j-1, t+2; \nu) = c^{(k)}_i(t+1) + q_{i,1}\nu + \left(1 - \mu_i q_{i,1} r_{i,1}\right) \beta \overline{V}^{*,(k)}_{i,\beta}(j-1, t+3; \nu) \\ + \mu_i q_{i,1} r_{i,1} \beta \overline{V}^{*,(k)}_{i,\beta}(j-2, t+3; \nu) \end{aligned} \tag{73}$$

Also, it has to be true that it is optimal not to transmit in $(j-2, r_{i,1}, t+2)$. This is because if it is optimal to transmit in both $(j, r_{i,1}, t+1)$ and $(j-2, r_{i,1}, t+2)$, then it must be optimal to

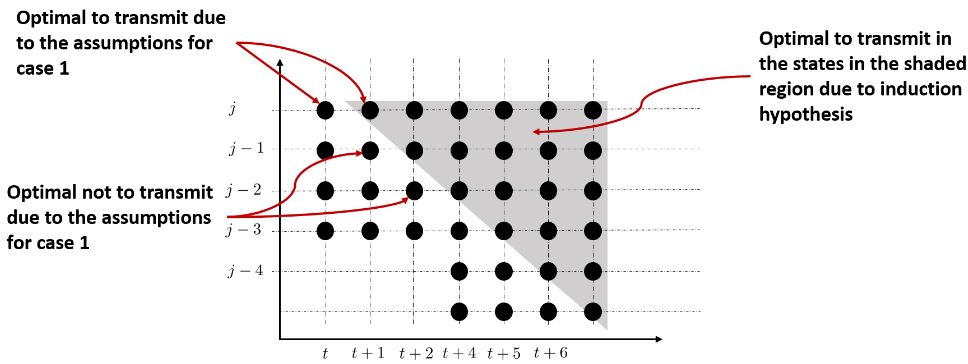

Fig. 14. Illustration of the induction steps

transmit in $\left(j-1, r_{i,1}, t+1\right)$. This is obtained directly from the Bellman equations and our induction hypothesis. Therefore, we have the following equation:

$$\overline{V}_{i,\beta}^{*,(k)}(j-2, t+2; \nu) = c_i^{(k)}(t+1) + \beta \overline{V}_{i,\beta}^{*,(k)}(j-2, t+3; \nu) \tag{74}$$

Substituting (73) and (74) in (72), we will get the following:

$$\nu > \frac{\beta\left(1-\mu_i q_{i,1} r_{i,1}\right)}{1-\mu_i q_{i,1} r_{i,1}}\left[\mu_i r_{i,1} \beta\left(\overline{V}_{i,\beta}^{*,(k)}(j-1, t+3; \nu) - \overline{V}_{i,\beta}^{*,(k)}(j-2, t+3; \nu)\right)\right]. \tag{75}$$

Using (71) and (75), we will get the following inequality:

$$\overline{V}_{i,\beta}^{*,(k)}(j-1, t+3; \nu) - \overline{V}_{i,\beta}^{*,(k)}(j-2, t+3; \nu) < \overline{V}_{i,\beta}^{*,(k)}(j, t+2; \nu) - \overline{V}_{i,\beta}^{*,(k)}(j-1, t+2; \nu). \tag{76}$$

However, this cannot be true due to (67). Therefore, we have proved the result via contradiction.

Case 2. Let us assume that it is optimal to transmit in both $\left(j, r_{i,1}, t\right)$ and $\left(j-1, r_{i,1}, t+1\right)$ and not optimal to transmit in $\left(j, r_{i,1}, t+1\right)$. We will prove that this is not possible by contradiction.

From our induction hypothesis if it is optimal to transmit in $\left(j-1, r_{i,1}, t+1\right)$, then it is optimal to transmit in the states shown in the Fig. 14. This would imply that if it is optimal not to transmit in $\left(j, r_{i,1}, t+1\right)$, then it is optimal not to transmit in any $\left(j, r_{i,1}, t'\right), \forall t' \geq t+2$. This is because if it was true for some t'', then using the fact that it is also optimal to transmit in $\left(j-1, r_{i,1}, t''\right)$, we can iteratively show that it is optimal to transmit in $\left(j, r_{i,1}, t'\right), \forall t' \geq t+1$. Therefore, if the transmission does not succeed in $\left(j, r_{i,1}, t\right)$, then there are no future transmissions. To derive analytic expressions for this property, let us first define the following term:

$$\widehat{H}_\beta(t) := \sum_{m=0}^{\infty} c_i^{(k)}(t+m)\beta^m. \tag{77}$$

$\widehat{H}_\beta(t)$ is the average cumulative cost if no transmission is performed after time t. This summation is guaranteed to be finite because of our assumption that $c_i(t) < \delta t^\zeta$ for large t. Therefore, in this setting, from our previous discussion $\overline{V}_{i,\beta}^{*,(k)}(j, t+1; \nu) = \widehat{H}_\beta(t)$. Since we have assumed that it is optimal to transmit in $\left(j, r_{i,1}, t\right)$, from Bellman equations, we have the following inequality:

$$\nu \leq \mu_i r_{i,1} \beta\left(\widehat{H}_\beta(t+1) - \overline{V}_{i,\beta}^{*,(k)}(j-1, t+1; \nu)\right). \tag{78}$$

Since it is not optimal to transmit in $\left(j, r_{i,1}, t+1\right)$, we can similarly write the following inequality:

$$\nu > \mu_i r_{i,1} \beta \left(\widehat{H}_\beta(t+2) - \overline{V}^{*,(k)}_{i,\beta}(j-1, t+2; \nu) \right). \tag{79}$$

Let us look at the term $\widehat{H}_\beta(t+1) - \overline{V}^{*,(k)}_{i,\beta}(j-1, t+1; \nu)$. We can re-write this term as follows:

$$\widehat{H}_\beta(t+1) - \overline{V}^{*,(k)}_{i,\beta}(j-1, t+1; \nu) = \mathbb{E}\left[\sum_{t'=t+2+T(j-1,t+1,j-1;\nu)}^{\infty} c_i^{(k)}(t')\beta^{t'} \right] - N^*_{i,\beta}(j-1, t+1, \nu) \tag{80}$$

The above equation is obtained by re-writing $\overline{V}^{*,(k)}_{i,\beta}(j-1, t+1; \nu)$ as follows:

$$\overline{V}^{*,(k)}_{i,\beta}(j-1, t+1; \nu) = \mathbb{E}\left[\sum_{t'=t+1}^{t+1+T(j-1,t+1,j-1;\nu)} c_i^{(k)}(t')\beta^{t'} \right] + N^*_{i,\beta}(j-1, t+1, \nu). \tag{81}$$

Similarly, we can re-write $\widehat{H}_\beta(t+2) - \overline{V}^{*,(k)}_{i,\beta}(j-1, t+2; \nu)$ as follows:

$$\widehat{H}_\beta(t+2) - \overline{V}^{*,(k)}_{i,\beta}(j-1, t+2; \nu) = \mathbb{E}\left[\sum_{t'=t+3+T(j-1,t+2,j-1;\nu)}^{\infty} c_i^{(k)}(t')\beta^{t'} \right] - N^*_{i,\beta}(j-1, t+2, \nu). \tag{82}$$

By our induction hypothesis and the fact that we are only transmitting when $r = r_{i,1}$, $T(j-1, t+1, j-1; \nu)$ and $T(j-1, t+2, j-1; \nu)$ are statistically identical. We also have that

$$N^*_{i,\beta}(j-1, t+1, \nu) = N^*_{i,\beta}(j-1, t+2, \nu). \tag{83}$$

Therefore, using the non-decreasing property of $c_i^{(k)}(t)$, we get the following inequality:

$$\widehat{H}_\beta(t+2) - \overline{V}^{*,(k)}_{i,\beta}(j-1, t+2; \nu) > \widehat{H}_\beta(t+1) - \overline{V}^{*,(k)}_{i,\beta}(j-1, t+1; \nu). \tag{84}$$

Therefore, a lower bound for ν is greater than its upper bound, which is a contradiction. Hence, proved.

G QUALITATIVE RESULTS

G.1 Proof of Theorem 4.5

First we will prove the following lemma which is useful to prove this theorem.

Lemma G.1. *If it is optimal to transmit in the state $\left(j, r_{i,2}, t\right)$, then it is optimal to transmit in any state $\left(j', r_{i,2}, t'\right)$ such that $j' \geq j$ and $t' \geq t$.*

Proof. We will show that this holds for the cost function $c_i^{(k)}(\cdot)$. For this we will use induction starting from time k and proceeding backwards to t as shown in Fig. 13.

Base Case: Note that for $t \geq k$ the holding cost function is a constant. For constant holding cost functions, $\Delta_i^{*,(k)}(j, t, \nu)$ is independent of j, see Proposition 1 in [4]. Therefore if it is optimal to transmit in $\left(j, r_{i,2}, t\right)$, then it is optimal to transmit in $\left(j', r_{i,2}, t\right)$ such that $j' \geq j$.

Induction Hypothesis: If it is optimal to transmit in $\left(j, r_{i,2}, t'\right)$ for any j and $t' \geq t+1$, then it is optimal to transmit in $\left(j', r_{i,2}, t''\right)$ for any $j' \geq j$ and $t'' \geq t'$.

Using the induction hypothesis, we will prove the result for any j at time t. First note that if it is optimal to transmit in $\left(j, r_{i,2}, t\right)$, then it is optimal to transmit in either $\left(j, r_{i,2}, t+1\right)$ or $\left(j-1, r_{i,2}, t+1\right)$. This can be proved using contradiction, i.e., we shall assume that it is optimal to transmit in $\left(j, r_{i,2}, t\right)$ and it is not optimal to transmit in $\left(j, r_{i,2}, t+1\right)$ and $\left(j-1, r_{i,2}, t+1\right)$. Now consider another policy in which we do not transmit in $\left(j, r_{i,2}, t\right)$ and we transmit in both

$\left(j, r_{i,2}, t+1\right)$ and $\left(j-1, r_{i,2}, t+1\right)$, while leaving the remaining actions unchanged with respect to an optimal policy. Starting with phase j at time t, the average cumulative cost with this policy is same as the average cumulative cost with the optimal policy, which is a contradiction as we had assumed that it is not optimal to transmit in $\left(j, r_{i,2}, t+1\right)$ and $\left(j-1, r_{i,2}, t+1\right)$. Therefore, it is optimal to transmit in either $\left(j, r_{i,2}, t+1\right)$ or $\left(j-1, r_{i,2}, t+1\right)$.

From induction hypothesis, if it is optimal to transmit in $\left(j, r_{i,2}, t+1\right)$ or $\left(j-1, r_{i,2}, t+1\right)$, it is also optimal to transmit in $\left(j, r_{i,2}, t+1\right)$ and $\left(j+1, r_{i,2}, t+1\right)$. If it is optimal to transmit in both $\left(j, r_{i,2}, t+1\right)$ and $\left(j+1, r_{i,2}, t+1\right)$, then it is optimal to transmit in $\left(j+1, r_{i,2}, t\right)$. To see this, let us re-write $\Delta_i^*(j+1, t, \nu)$ as follows:

$$\Delta_i^{*,(k)}(j+1, t, \nu) = \overline{V}_i^{*,(k)}(j+1, t+1; \nu) - \overline{V}_i^{*,(k)}(j, t+1; \nu), \tag{85}$$

$$= (1-\mu_i \bar{r}_i)\left(\Delta_i^{*,(k)}(j+1, t+1, \nu)\right) + \mu_i \bar{r}_i\left(\Delta_i^{*,(k)}(j, t+1, \nu)\right) \tag{86}$$

Note that in writing (86), we have used the fact that is optimal to transmit in $\left(j, r_{i,2}, t+1\right)$, $\left(j+1, r_{i,2}, t+1\right)$, $\left(j, r_{i,1}, t+1\right)$, and $\left(j+1, r_{i,1}, t+1\right)$. Since it is optimal to transmit in $\left(j, r_{i,2}, t+1\right)$ and $\left(j+1, r_{i,2}, t+1\right)$, from Bellman equations, we have that

$$\frac{\nu}{\mu_i r_{i,2}} \leq \Delta_i^{*,(k)}(j+1, t+1, \nu), \tag{87}$$

$$\frac{\nu}{\mu_i r_{i,2}} \leq \Delta_i^{*,(k)}(j, t+1, \nu). \tag{88}$$

From (86), this would imply that $\Delta_i^{*,(k)}(j+1, t, \nu) \geq \nu/\mu_i r_{i,2}$. This would mean that it is optimal to transmit in $\left(j+1, r_{i,2}, t\right)$. Since this was proved for any $c_i^{(k)}(\cdot)$, from lemma B.1 it holds for $c_i(t)$ too. □

The above lemma would imply that $\nu_i^*\left(j, r_{i,2}, t\right) \leq \nu_i^*\left(j', r_{i,2}, t'\right)$. Since j, j', t, and t' are arbitrarily chosen, this would imply that $\nu_i^*\left(j, r_{i,2}, t\right)$ is a non-decreasing function of j and t. To extend this result to the entire shaded region as shown in Fig. 2, from (64), one could show that if it is optimal to transmit in $\left(j, r_{i,2}, t+1\right)$ and $\left(j-1, r_{i,2}, t+1\right)$, then it is also optimal to transmit in $\left(j, r_{i,2}, t\right)$. If we use this property and the above lemma iteratively, then it can be shown that if it is optimal to transmit in $\left(j, r_{i,2}, t\right)$, then it is optimal to transmit in any state $\left(j', r_{i,1}, t'\right)$ such that $j' \geq j$ and $j' + t' \geq j + t$.

G.2 Proof of Theorem 4.7

We have already proved in Lemma F.2 that for large enough β if it is optimal to transmit in $\left(j, r_{i,1}, t\right)$, then it is optimal to transmit in all states reachable from $\left(j, r_{i,1}, t\right)$. This would also imply that it is optimal to transmit in all states $\left(j, r_{i,1}, t'\right)$ such that $t' \geq t$. Hence, $\nu_{i,\beta}^*\left(j, r_{i,1}, t\right) \leq \nu_{i,\beta}^*\left(j, r_{i,1}, t'\right)$. This would imply that

$$\lim_{\beta \to 1}(1-\beta)\,\nu_{i,\beta}^*\left(j, r_{i,1}, t\right) \leq \lim_{\beta \to 1}(1-\beta)\,\nu_{i,\beta}^*\left(j, r_{i,1}, t'\right). \tag{89}$$

Hence, the result is proved.

G.3 Proof of Theorem 4.4

We will show that this property holds for truncated holding cost functions $c_i^{(k)}(\cdot)$ and $c_l^{(k)}(\cdot)$. To prove this result for any k, we will use induction.

Base Case: By the definition of $c_i(t)$ and $c_l(t)$, we have that $c_i^{(k)}(t) \le c_l^{(k)}(t)$. This would also imply that $c_i(k) \le c_l(k)$. Using the result from [4] for constant holding costs, when the cost functions are $c_i^{(k)}(\cdot)$ and $c_l^{(k)}(\cdot)$, we get that $\Delta_i^{*,(k)}(j,k,\nu) \le \Delta_l^{*,(k)}(j,k,\nu)$. Hence, base case is true.

Induction Hypothesis: Assume that $\Delta_i^{*,(k)}(j,t',\nu) \le \Delta_l^{*,(k)}(j,t',\nu)$ for all $t+1 \le t' \le k$.

We will show that $\Delta_i^{*,(k)}(j,t,\nu) \le \Delta_l^{*,(k)}(j,t,\nu)$. Note that from (64) (with $\beta = 1$), $\Delta_i^{*,(k)}(j,t,\nu)$ is an increasing function of $\Delta_i^{*,(k)}(j,t+1,\nu)$ and $\Delta_i^{*,(k)}(j-1,t+1,\nu)$. Then from our induction hypothesis it follows that $\Delta_i^{*,(k)}(j,t,\nu) \le \Delta_l^{*,(k)}(j,t,\nu)$. Since we have proved it for truncated holding cost functions, from Lemma B.1 it follows that the result holds for $c_i(\cdot)$ and $c_l(\cdot)$.

H QUANTITATIVE RESULTS

H.1 Proof of Theorem 5.1

This is a special case with $q_{i,1} = 1$ and $\mu_i = 1$. We have already proved that $\nu_i^*(j,r_i,t) = \infty, \forall t$ and j. In the proof of Theorem 4.3, we have given a constructive proof to study the asymptote of $\nu_{i,\beta}^*(j,r_i,t)$ (with respect to β) in which we have shown that the optimal policy and the policy in which transmission is done only in $r_{i,1}$ have the same asymptote when we set $\nu = \nu_{i,\beta}^*(j,r_i,t)$. In this setting, we have $\mu_i r_i = 1$, i.e., all transmissions are successful in completing a phase with probability one. Substituting this in (61), we get

$$\xi_i^*\left(j, r_{i,1}, t\right) = \frac{1}{j} c_i(t+j). \tag{90}$$

Note that in writing the above equation, we have used the following expression for $H_{i,1}^{\dagger}(j,t)$, which was obtained because $\mu_i r_i = 1$:

$$H_{i,1}^{\dagger}(j,t) = \sum_{t'=t}^{t+j} c_i(t'). \tag{91}$$

H.2 Proof of Theorem 5.2

From Thm. 4.5, if it is optimal to transmit in $\left(1, r_{i,2}, t\right)$, then it is optimal to transmit in $\left(1, r_{i,2}, t'\right)$ $\forall t' \ge t$. From Bellman equations, if it is optimal to transmit in $\left(1, r_{i,2}, t'\right)$, then it is also optimal to transmit in $\left(1, r_{i,1}, t'\right)$. Therefore, if it is optimal to transmit in $\left(1, r_{i,2}, t\right)$, then it is optimal to transmit in all states in future. To find $\nu_i^*\left(1, r_{i,1}, t\right)$, we have to solve the following equation in ν:

$$\nu = \mu_i r_{i,2} \overline{V}_i^*(1, t+1; \nu). \tag{92}$$

Since it is optimal to transmit in all future states we can re-write $\overline{V}_i^*(1,t+1;\nu)$ as follows:

$$\overline{V}_i^*(1,t+1;\nu) = \sum_{j=1}^{\infty} c_i(t+j)(1-\mu_i \bar{r}_i)^{j-1} + \frac{\nu}{\mu_i \bar{r}_i}. \tag{93}$$

Substituting in (92), we get the expression for $\nu_i^*\left(1, r_{i,1}, t\right)$.

H.3 Proof of Theorem 5.3

Primary indices: It is difficult to find an exact expression for the primary index when $R(t) \neq r_{i,1}$ in a multi-state i.i.d. service rate setting with phase-type distribution for jobs sizes. In any state (j,r,t) we know from the Bellman equations (7) that if it is optimal to transmit in $r = r_{i,l}$, then it is optimal to transmit when $r = r_{i,l'}$, $l' = 1, 2, \ldots, l-1$. However, we do not know if it is optimal to transmit it when $R_i(t) = r_{i,l}$ for the future states.

We shall approximate $\nu_i^*\left(j, r_{i,l}, t\right)$ with a lower bound. Observe that if it is optimal to transmit in state $\left(1, r_{i,l}, t+j-1\right)$, then it is also optimal to transmit in state $\left(j, r_{i,l}, t\right)$. This directly follows from Thm. 4.5[1]. Therefore, $\nu_i^*\left(1, r_{i,l}, t+j-1\right)$ is a lower bound for $\nu_i^*\left(j, r_{i,l}, t\right)$.

Next we shall discuss computation of $\nu_i^*\left(1, r_{i,l}, t+j-1\right)$. For $j=1$, we have that

$$\Delta_i^*(1, t, \nu) = \overline{V}_i^*(1, t; \nu) = H_i^*(1, t, \nu) + \nu N_{i,1}^*(1, t, \nu). \tag{94}$$

If it is optimal to transmit in $\left(1, r_{i,l}, t\right)$, then it is optimal to transmit when the rate is greater than or equal to $r_{i,l}$ in all future states from Lemma G.1. However, we cannot say if it is optimal to transmit in future states with service rates strictly lower that $r_{i,l}$. Therefore, we shall find a lower bound for $\overline{V}_i^*(1, t+j-1; \nu)$ and use to find the fixed point of $\mu_i r_{i,l} \overline{V}_i^*(1, t+j-1; \nu)$. This fixed point using the lower bound of $\overline{V}_i^*(1, t+j-1; \nu)$ will be a lower bound for $\nu_i^*\left(1, r_{i,l}, t+j-1\right)$. First we shall derive a lower bound for $H_i^*(1, t, \nu)$. For any policy, the average holding cost is lower bounded by the cost under the policy in which transmission is always performed irrespective of the channel state. Therefore, we have that

$$H_i^*(1, t+j-1, \nu) \geq \sum_{m=0}^{\infty} c_i(t+j-1+m)\left(1-\mu_i \sum_{n=1}^{L} q_{i,n} r_{i,n}\right)^m. \tag{95}$$

Since we know that it is optimal to transmit when the rate is greater than or equal to $r_{i,l}$, we can lower bound the term $N_{i,1}^*(1, t, \nu)$ as follows:

$$N_{i,1}^*(1, t, \nu) \geq \frac{\sum_{l'=1}^{l} q_{i,l'}}{\mu_i \sum_{l'=1}^{l} q_{i,l'} r_{i,l'}}. \tag{96}$$

Solving the following equation in ν gives the required expression:

$$\nu = \mu_i r_{i,l}\left(\sum_{m=0}^{\infty} c_i(t+j-1+m)\left(1-\mu_i \sum_{n=1}^{L} q_{i,n} r_{i,n}\right)^m + \frac{\nu \sum_{l'=1}^{l} q_{i,l'}}{\mu_i \sum_{l'=1}^{l} q_{i,l'} r_{i,l'}}\right). \tag{97}$$

Secondary indices: We have computed the expression for secondary indices for i.i.d. multi-state channel in (61). This gives (21). If we transmit only when $R_i(t) = r_{i,1}$, then probability of completing a phase in any given slot is $q_{i,1}\mu_i r_{i,1}$. Using this and the definition of $H_{i,1}^\dagger(j, t)$ one could derive equations (22)–(25).

[1]One can easily extend the derivation for the two state i.i.d. channel to a multi-state i.i.d. channel setting and we state the result without proof.