Received 28 July 2024; accepted 16 August 2024. Date of publication 21 August 2024; date of current version 30 August 2024.

Digital Object Identifier 10.1109/OJCOMS.2024.3447157

User Association in Dense Millimeter Wave Networks With Multi-Channel Access Points Using the Whittle Index

RAVINDRA S. TOMAR, MANDAR R. NALAVADE, AND GAURAV S. KASBEKAR (Member, IEEE)

Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India

CORRESPONDING AUTHOR: M. R. NALAVADE (e-mail: 22d0531@iitb.ac.in)

This work was supported in part by the Project under Grant RD/0121-MEITY01-001.

ABSTRACT In dense millimeter wave (mmWave) networks, user association, i.e., the task of selecting the access point (AP) that each arriving user should join, significantly impacts the network performance. We consider a dense mmWave network in which each AP has multiple channels and can simultaneously serve different users using different channels. The different channels of an AP are susceptible to both blockage, which is common to all the channels of an AP, and frequency-selective fading, which is, in general, different for different channels. In each time slot, a user arrives with some probability. Our objective is to design a user association scheme for selecting the AP that each arriving user should join, so as to minimize the long-term total average holding cost incurred within the system, and thereby achieve low average delays experienced by users. This problem is an instance of the restless multi-armed bandit problem, and is provably hard to solve. We prove that the problem is Whittle indexable and present a method for calculating the Whittle indices corresponding to the different APs by solving linear systems of equations. We propose a user association policy under which, when a user arrives, it associates with the AP that has the lowest Whittle index in that time slot. Our extensive simulation results demonstrate that our proposed Whittle index-based policy outperforms user association policies proposed in prior research in terms of the average delay, average cost, as well as Jain's fairness index (JFI).

INDEX TERMS Millimeter wave networks, multiple channels, Markov decision process, user association, whittle index.

I. INTRODUCTION

THE EVER-INCREASING demand for next-generation wireless services necessitates innovative solutions to meet it. From smart cities to remote surgery, new networked applications are pushing the boundaries of wireless networks [1]. To handle the ever-growing data demand, these networks must undergo a major transformation, focusing on increased capacity, lower latency, and enhanced reliability. With the exponential increase of smart devices, cellular data traffic has been projected to surge by nearly 1,000 times between 2020 and 2030 [2]. The limited capacity of the sub-6 GHz spectrum imposes limitations on the ability to scale data rates sufficiently to meet future wireless communication demands. Future wireless requirements, e.g., supporting high-resolution real-time applications and providing ubiquitous connectivity for diverse Internet of Things (IoT) devices, make the utilization of high-frequency spectrum necessary. Millimeter wave (mmWave) frequencies unlock the potential for a large number of wide spectral channels, a key driver for achieving high data rates and network capacity in 5G networks and beyond. MmWave communication differs fundamentally from sub-6 GHz communication due to its high propagation loss, limited diffraction, high penetration losses, pronounced sensitivity to blockages, etc. [3], [4], [5]. To mitigate blockages and path loss, beamforming techniques are employed to create highly focused, high-gain directional beams [6].

The process of deciding as to which access point (AP) a user's device should be connected to in a wireless network is known as user association. User association algorithms significantly influence vital network performance metrics such as load balancing, spectral efficiency (SE),

and energy efficiency (EE), thereby significantly impacting the overall network performance [7]. In current networks, user-cell association is performed solely on the basis of the Reference Signal Received Power (RSRP) and Reference Signal Received Quality (RSRQ). RSRP depends on the received signal power, while RSRQ additionally considers noise power and interference power. An arriving user associates with the AP that provides the strongest signal based on a combination of these measurements [8], [9]. This approach, while simple and widely deployed, neglects crucial factors such as AP load and can lead to sub-optimal performance. Extensive research has been conducted on the design of user association policies in wireless networks, including both sub-6 GHz networks [10], [11], [12], [13], [14], [15], [16], [17], [18], [19] and mmWave networks [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30] (see Section XI for a review). However, with the exception of our prior work [31], [32], the powerful *Whittle index* technique [33] has so far not been employed for solving the user association problem. There are several differences between the system models and results in [31], [32] and those in this paper, which are explained in detail in Section XI. For example, the work in [31], [32] is limited to the simple model in which each AP operates on a single channel. This paper addresses the user association problem in dense mmWave networks, where each AP can operate on multiple channels. The proposed association policy is based on the theory of Whittle index. To the best of our knowledge, this paper is the first to study the user association problem in a mmWave network with multi-channel APs using the Whittle index; also, our model takes into account blockages and frequency-selective fading in such networks.

The seminal work of Whittle [33] introduced the Whittle index, a mathematical framework for decision-making in stochastic systems. The concept of the Whittle index has been utilized in solving problems in diverse domains, including real-time multicast scheduling in wireless broadcast networks [34], scheduling in processor-sharing systems [35], Quality of Experience (QoE) optimization in wireless networks [36], policy design for crawling ephemeral content [37], design of schemes for dynamic multi-channel opportunistic access [38], optimizing age of information in wireless networks [39], [40], [41], packet transmission scheduling in wireless networks [42], etc.

This study considers a dense wireless mmWave network with multiple APs, each of which operates on multiple channels. The different channels of an AP are susceptible to both blockage, which is common to all the channels of an AP, and frequency-selective fading, which is, in general, different for different channels. Each AP can serve multiple users simultaneously in a given time slot by allocating a different channel to each user. During each time slot, a user arrives with some probability. Our objective is to design a user association scheme, which assigns an AP to each user upon arrival, so as to minimize the long-term total average holding cost incurred within the system, and thereby achieve low average delays experienced by users. This problem belongs to the class of restless multi-armed bandit (RMAB) problems, and is provably hard to solve [43]. Employing the framework introduced by Whittle [33], the hard per-stage constraint that each user must be associated with a single AP in a time slot is relaxed to a long-term time-averaged constraint. Then, we employ the Lagrangian multiplier strategy to transform the problem into an unconstrained form and decompose it into separate Markov Decision Processes (MDPs) at different APs. We prove that the resulting problem is Whittle indexable and present a method for calculating the Whittle indices corresponding to different APs by solving linear systems of equations. We propose a user association policy under which, when a user arrives, it associates with the AP that has the lowest Whittle index in that time slot. Through simulations, we demonstrate that our proposed Whittle index-based user association policy performs better than various user association policies described in existing research in terms of average cost, average delay, as well as the Jain's fairness index [44]. In general, proving the Whittle indexability of a RMAB problem is challenging. Establishing the Whittle indexability of the user association problem in a mmWave network with multi-channel APs through a rigorous proof is the main contribution of this paper.

The remainder of this paper is structured as follows. In Section II, we describe the system model and formulate the problem. A stability analysis of the Markov chain associated with the problem is provided in Section III. In Section IV, we provide the dynamic programming equation (DPE) for the formulated problem. In Section V, we prove some structural properties of the value function. In Section VI, we prove the threshold behavior of the optimal policy. We prove that the problem is Whittle indexable in Section VII and describe our proposed policy for the calculation of the Whittle indices of different APs in Section VIII. Section IX presents a review of some user association schemes proposed in prior work and Section X demonstrates the superiority of the proposed Whittle index-based association policy over the schemes briefly described in Section IX using simulations. A review of related prior work is presented in Section XI. Finally, Section XII concludes the paper along with a discussion on directions for future research.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a dense mmWave network with K APs serving users in a small area such as a bus stop, conference hall, etc. Time is divided into slots of equal duration, denoted by $n \in \{0, 1, 2, \ldots\}$. In each time slot, a user arrives with probability (w.p.) $p \in (0, 1)$ and no user arrives w.p. $1 - p$. Fig. 1 illustrates the system model for an example network with $K = 4$ APs. Each AP has multiple channels, and a channel can be allocated to at most one user at a time. In particular, AP $i \in \{1, \ldots, K\}$ has N_i channels and hence can serve up to N_i associated users within a single time slot. Considering the inherent susceptibility of mmWave communications to blockages from obstacles such

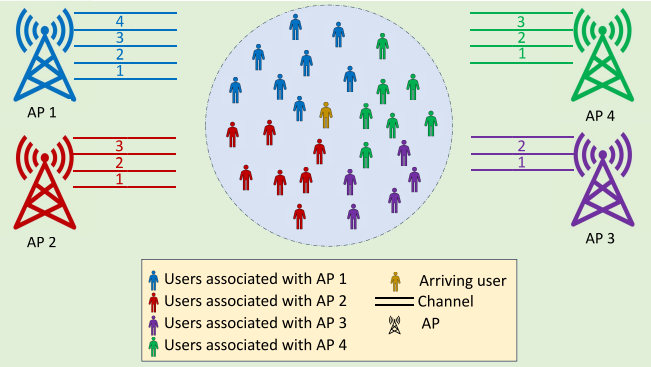

FIGURE 1. The figure shows a mmWave network with four APs, each serving its associated users. The numbers of channels with different APs are $N_1 = 4$, $N_2 = N_4 = 3$, and $N_3 = 2$.

as walls, humans, etc., we let s_i (respectively, $1 - s_i$) be the probability that the channels from AP i to the users are not blocked (respectively, are blocked) in a time slot. Note that since the users are located in a small area, we assume that blockage from an AP i occurs for either all the users or for none of them. Apart from blocking or its absence, channels also experience frequency-selective fading. In particular, different channels of an AP i experience independent and identically distributed frequency-selective fading due to the separation of the channels in the frequency-domain. Let $h_i \in (0, 1)$ (respectively, $1 - h_i$) denote the probability that the frequency-selective fading on a given channel of AP i is mild (respectively, severe). The user to which a channel of AP i is assigned departs in a time slot iff the channel is not blocked and the frequency-selective fading on the channel is mild in the time slot.

At the beginning of time slot n, let X_n^i be the number of users associated with AP i. Let D_n^i be the number of departures from AP i in time slot n. If $X_n^i = 0$, then $D_n^i = 0$. If $X_n^i \geq 1$, then D_n^i has the probability mass function $P_d = P(D_n^i = d)$ given in (1). We now explain why (1) holds. Since each of the N_i channels can serve at most one user in a time slot, the maximum number of departures is $\min(X_n^i, N_i)$. The number of departures is $d = 0$ if either the channels of AP i are blocked (which happens w.p. $1 - s_i$) or the channels are not blocked and all the N_i channels experience severe frequency-selective fading (which happens w.p. $s_i(1 - h_i)^{N_i}$). The number of departures is d, where $0 < d < \min(X_n^i, N_i)$, if the channels are not blocked (which happens w.p. s_i) and exactly d (respectively, $N_i - d$) of the N_i channels experience mild (respectively, severe) frequency-selective fading (which happens w.p. $\binom{N_i}{d}(h_i)^d(1 - h_i)^{N_i - d}$). The number of departures is $d = \min(X_n^i, N_i)$ if the channels are not blocked (which happens w.p. s_i) and l (respectively, $N_i - l$) of the channels experience mild (respectively, severe) fading (which happens w.p. $\binom{N_i}{l}(h_i)^l(1 - h_i)^{N_i - l}$), where $l \in \{d, d + 1, \ldots, N_i\}$.

$$P_d = \begin{cases} s_i(1 - h_i)^{N_i} + (1 - s_i), & \text{if } d = 0, \\ s_i\binom{N_i}{d}(h_i)^d(1 - h_i)^{N_i - d}, & \text{if } 0 < d < \min\left(X_n^i, N_i\right), \\ \sum_{l=d}^{N_i} s_i\binom{N_i}{l}(h_i)^l(1 - h_i)^{N_i - l}, & \text{if } d = \min\left(X_n^i, N_i\right). \end{cases} \quad (1)$$

The number of users associated with AP i at the beginning of time slot $n + 1$, X_{n+1}^i, is given by:

$$X_{n+1}^i = X_n^i - D_n^i + \mu_n^i\, \zeta_{n+1}, \quad (2)$$

where ζ_{n+1} denotes the number of user arrivals at the end of slot n:

$$\zeta_{n+1} = \begin{cases} 1, \text{ if a user arrives at the end of slot } n, \\ 0, \text{ otherwise.} \end{cases}$$

Note that the random variable ζ_{n+1} has a Bernoulli distribution and:

$$P(\zeta_{n+1} = 1) = p. \quad (3)$$

The admission control variable, μ_n^i, denotes the admission decision for the arriving user at AP i during time slot n. This binary parameter takes the value 1 if the user is admitted for service at AP i, and 0 otherwise. That is:

$$\mu_n^i = \begin{cases} 1, \text{ if AP } i \text{ admits the arriving user in time slot } n, \\ 0, \text{ otherwise.} \end{cases}$$

Each arriving user is associated with only one AP; thus, $\sum_{i=1}^K \mu_n^i = 1$.

Let $C_i > 0$ be the holding cost incurred at AP i per user in each time slot. The cumulative cost experienced by all APs in the network in time slot n is $\sum_{i=1}^K C_i X_n^i$. Our aim is to design a non-anticipating admission policy [45], which selects μ_n^i, $i \in \{1, \ldots, K\}$, in each time slot n with the objective of minimization of the long-term expected average cost incurred at the APs in the network. The objective is as follows:

$$\min \limsup_{\top \uparrow \infty} E\left[\frac{1}{\top}\sum_{n=0}^{\top-1}\sum_{i=1}^{K} C_i X_n^i\right],$$
$$\text{s.t.} \sum_{i=1}^{K} \mu_n^i = 1, \forall n. \quad (4)$$

Note that minimizing the average cost in (4) leads to a low average delay experienced by the users in the network.

The exact per-stage constraint $\sum_{i=1}^K \mu_n^i = 1$ makes it provably difficult to solve the above constrained optimization problem [43]. Hence, we employ a relaxation technique proposed by Whittle [33] and relax this constraint to the following time-averaged constraint:

$$\limsup_{\top \uparrow \infty} \frac{1}{\top}\sum_{n=0}^{\top-1}\sum_{i=1}^{K} E\left[\mu_n^i\right] = 1. \quad (5)$$

Utilizing the Lagrangian multiplier approach [46], the original problem in (4) is transformed into the following unconstrained optimization problem:

$$\min \limsup_{\top \uparrow \infty} \frac{1}{\top}\sum_{n=0}^{\top-1}\sum_{i=1}^{K} E\left[\mathcal{G}_i\left(X_n^i, \mu_n^i\right)\right],$$
$$\text{where } \mathcal{G}_i(x, \mu) = C_i x + \lambda(1 - \mu), \quad (6)$$

and λ is the Lagrangian multiplier. As in the analysis of Whittle [33], and since the above problem is a cost

minimization problem, λ can be treated as a tax or penalty. In the context of Whittle indexability, this implies that when an AP i declines an arriving user (i.e., when $\mu_n^i = 0$), a tax or penalty of λ is added to the AP's incurred cost. The minimization problem formulated in (6) decomposes into individual control problems for each AP when λ is fixed. The control problem corresponding to each AP i can be modeled as a MDP, where the objective is to minimize $\limsup_{\top\uparrow\infty} \frac{1}{\top}\sum_{n=0}^{\top-1} E[\mathcal{G}_i(X_n^i, \mu_n^i)]$, and the state represents the total number of users, X_n^i, associated with AP i. The action space consists of accepting or rejecting an arriving user at AP i ($\mu_n^i = 1$ or $\mu_n^i = 0$, respectively) in each time slot. This MDP framework is said to exhibit Whittle indexability if, for all feasible parameter sets $(C_i, s_i, h_i, p, N_i) \in (0,\infty)\times(0,1)\times(0,1)\times(0,1)\times\{1,\ldots,\infty\}$, and for each AP i, the collection of states where the arriving user is not admitted (hereafter termed "passive states") exhibits a monotonic decrease, transitioning from being the entire state space to becoming the empty set as we increase λ from $-\infty$ to ∞. For each AP, the Whittle index associated with state x represents the specific λ value at which the controller is indifferent between accepting and rejecting an arriving user. According to the Whittle index policy, during each time slot, the AP with the lowest Whittle index grants admission to the arriving user.

Without loss of generality, let us assume that $1 > h_1 \geq h_2 \geq \ldots \geq h_K > 0$. Additionally, to ensure stability of the system at each AP, we assume that $p < \min_{i\in\{1,2,\ldots,K\}}(N_i s_i h_i)$. In Section III, we present an analysis of the stability of the systems at individual APs.

Remark 1: The model we consider in this paper is particularly appropriate for mmWave networks, wherein APs are generally deployed densely. Since there are many APs with which an arriving user can potentially associate, the user association problem mentioned above becomes non-trivial. However, the results we provide in this paper are not limited to mmWave networks; all the results also apply to other wireless networks, such as sub-6 GHz networks, in which there is a dense deployment of APs.

Remark 2: All the results in our paper readily generalize to the case in which, in a time slot, if the channel from a user to its AP, say i, is not blocked and the frequency-selective fading on the channel is mild, then the user departs w.p. $e_i \in (0,1)$, where e_i models any factors related with the AP that affect the departures, e.g., the presence of noise, errors in transmission/ reception, etc.

Our results also readily generalize to the case in which, after a user arrives into the network, the number of packets that it exchanges with its AP before it departs is a random variable that is geometrically distributed with parameter g. That is, in each time slot in which a user receives error-free service, the packet that it exchanges is its last one w.p. g, and is not its last one w.p. $1-g$.

With the above two generalizations, the user to which a channel of AP i is assigned departs in a time slot iff the channel is not blocked, the frequency-selective fading on the channel is mild, *the user receives error-free service, and the packet that the user exchanges with AP i is its last one.*

Our results can be generalized in the above two ways by replacing the quantity h_i by $h_i e_i g$ throughout the analysis. In particular, (1) is replaced with the following equation:

$$P_d = \begin{cases} s_i(1-h_i e_i g)^{N_i} + (1-s_i), & \text{if } d = 0, \\ s_i\binom{N_i}{d}(h_i e_i g)^d (1-h_i e_i g)^{N_i-d}, & \text{if } 0 < d < \min\left(X_n^i, N_i\right), \\ \sum_{l=d}^{N_i} s_i\binom{N_i}{l}(h_i e_i g)^l (1-h_i e_i g)^{N_i-l}, & \text{if } d = \min\left(X_n^i, N_i\right). \end{cases}$$

Note that, for simplicity, we assume that in a time slot, at most one user's data can be processed using one channel. The case in which the data of $k \geq 2$ users can be processed using a channel in a time slot can be handled using our model by dividing a channel into k parts and treating each part as a separate channel.

Also note that for tractability, we assume that the number of packets that a user exchanges with its AP before it departs is geometrically distributed, since this distribution has the memoryless property. Recall that this assumption, as well as its counterpart in continuous-time, viz., that the size of the file transferred by a user is an exponentially distributed random variable, have been extensively made in prior work in the networking research literature for tractability. For example, the number of packets exchanged between an AP and a user is assumed to follow the geometric distribution in [47], [48], [49], [50], [51], [52], [53], [54]. Also, the sojourn time, packet transmission time, service time, session duration, etc., are assumed to follow the exponential distribution in [55], [56], [57], [58], [59], [60].

Remark 3: In this paper, the term "user association" means the selection of an AP to which an arriving user is assigned. This process does not include the assignment of bandwidth to the user. Several prior works, e.g., [7], [9], [10], [11], [12], [13], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], have also used the term "user association" in the above sense.

In our system model, in each time slot, out of the users associated with an AP i, up to N_i users are assigned a channel each for communication with the AP. This is done using an appropriate resource allocation algorithm, e.g., taking into account the urgency of the traffic to be exchanged with different users, fairness in the allocation of bandwidth to different users, etc. Also, different channels may be assigned to a given user in different time slots, i.e., channel switching is allowed under our system model.

Note that modern wireless transceivers are capable of rapidly switching across different channels. Channel switching times in modern wireless communication networks are of the order of a few tens of microseconds, e.g., the channel switching time is approximately 25 μs for the transceiver in [61] and 14 μs for the transceiver in [62].

In time slots in which a user has not been assigned any channel, it may exchange a small amount of control information with its AP on common control channel resources; this information includes scheduling decisions

by the AP for downlink and uplink transmissions, e.g., an indication of the channel number, if any, that is assigned to each user in the following time slot. These control channel resources may be similar to the Physical Downlink Control Channels (PDCCH) used in 5G cellular networks [7].

Our system model and all of our results are also applicable to the case where each AP has a single channel containing multiple subchannels, i.e., AP i has N_i subchannels. In this case, the above system model is similar to that in Orthogonal Frequency Division Multiple Access (OFDMA) based Wi-Fi and cellular networks, wherein each user is assigned an AP (or base station) when it arrives, and is dynamically assigned different subchannels in different time slots to communicate with its assigned AP.

III. STABILITY ANALYSIS

For a given control policy, the MDP of an individual AP reduces to a Discrete Time Markov Chain (DTMC). In the resulting induced DTMC, the state represents the number of users associated with the AP. Therefore, the corresponding state space $S = \mathbb{W}$, the set of non-negative integers.

Theorem 1: If $p < \min_{i\in\{1,2,\ldots,K\}}(N_i s_i h_i)$, then the induced DTMC for each AP i is positive recurrent.

The term p represents the average rate of user arrivals and $\min_{i\in\{1,2,\ldots,K\}}(N_i s_i h_i)$ is the minimum average departure rate from any AP. The condition $p < \min_{i\in\{1,2,\ldots,K\}}(N_i s_i h_i)$ signifies that the average number of arrivals is strictly less than the minimum average number of departures at any AP in the network in a time slot. Under this condition, regardless of the initial state, the induced DTMC will almost surely return to state zero in the future; hence, this condition ensures the stability of the controlled queue.

Proof: For system stability analysis, consider a Lyapunov function defined as $\Theta(x) = x$, where x represents the number of users associated with AP i. Using [63, Proposition 5.3] (p. 21), we can analyze the positive recurrence of the individual DTMC. The DTMC is positive recurrent if the following conditions hold for some constant $\epsilon > 0$:

$$\inf_{x\in S}\Theta(x) > -\infty, \tag{7}$$

$$\sum_{y\in S} p_{xy}\Theta(y) < \infty,\ \forall x \in S_0, \tag{8}$$

$$\Delta\Theta(x) = \sum_{y\in S} p_{xy}(\Theta(y) - \Theta(x)) \le -\epsilon,\ \forall x \notin S_0, \tag{9}$$

where S denotes the state space of the DTMC, $S_0 \subset S$ is a finite subset, the transition probability from state x to state y is denoted by p_{xy}, and $\Theta(x)$ denotes the Lyapunov function s.t. $\Theta : S \to \mathbb{R}$.

As $x \ge 0$ for all states x, the minimum value of the Lyapunov function is zero, i.e., $\min_{x\in S}\Theta(x) = 0$. This implies $\inf_{x\in S}\Theta(x) > -\infty$, satisfying the condition (7).

Consider a subset $S_0 = \{0, 1, \ldots, N_i - 1\} \subset S$, and state $x \in S_0$. If x is the present state of the DTMC, then the largest reachable next state is $x+1$, since at most one arrival is possible in a time slot. The finite transition probabilities, finite number of arrivals in a time slot, and the non-negativity of the Lyapunov function make $\sum_{y\in S} p_{xy}\Theta(y)$ finite. So condition (8) is satisfied.

Now, let $x \notin S_0$. Under the assumption $\min_{i\in\{1,2,\ldots,K\}}(N_i s_i h_i) > p$, there exists a positive constant ϵ s.t. $\min_{i\in\{1,2,\ldots,K\}}(N_i s_i h_i) - p \ge \epsilon$. Also, the drift term in (9) can be equivalently written as: $\Delta\Theta(x) = \mathbb{E}[\Theta(X^i_{n+1}) - \Theta(X^i_n)|X^i_n]$, where X^i_{n+1} and X^i_n represent the numbers of users associated with AP i in time slots $n+1$ and n, respectively. So for proving the condition (9), it is sufficient to show the following:

$$\Delta\Theta(x) = \mathbb{E}\Big[\Theta\Big(X^i_{n+1}\Big) - \Theta\Big(X^i_n\Big)|X^i_n\Big] \le -\epsilon.$$

If D^i_n (respectively, ζ_{n+1}) is the departure (respectively, arrival) random variable, then by using (2), the drift can be bounded as follows:

$$\Delta\Theta(x) \le \mathbb{E}\Big[-D^i_n|x\Big] + \mathbb{E}\Big[\zeta_{n+1}|x\Big] \le -(N_i s_i h_i) + p. \tag{10}$$

Equation (10) is true $\forall i \in \{1, 2, \ldots, K\}$. Under the assumption that $\min_{i\in\{1,2,\ldots,K\}}(N_i s_i h_i) - p \ge \epsilon$, condition (9) follows from (10). This completes the proof. ■

IV. DYNAMIC PROGRAMMING EQUATION

Recall that the problem (6) can be decomposed into separate MDPs at different APs for a given value of λ. Since the proof of Whittle indexability is identical for the decoupled problem associated with each AP i, henceforth, we simplify the notation by dropping the index i from all notation used in the proof. In any given time slot n, the maximum number of departing users for a given AP is $\min(N, X_n)$, where N denotes the number of available channels at the AP and X_n represents the number of users associated with the AP at the start of slot n. At the AP, let $X_n = x$ denote the current state of the system. We first formulate the Dynamic Programming Equation (DPE) for the individual MDP at the AP, and then prove its validity. The DPE for the value function, $V(\cdot)$, of the individual MDP is given by (see (1), (2), and (3)):

$$\begin{aligned} V(x) = Cx - \rho + \min\Big(&P_0\{(1-p)V(x) + pV(x+1)\} \\ &+ \sum_{d=1}^{\min(x-1,N-1)} P_d\{(1-p)V(x-d) + pV(x+1-d)\} \\ &+ P_{\min(x,N)}\Big[(1-p)V(\max(x,N) - N) \\ &+ pV(\max(x,N) - N + 1)\Big];\ \ \lambda + P_0 V(x) \\ &+ \sum_{d=1}^{\min(x-1,N-1)} P_d V(x-d) + P_{\min(x,N)} \times \\ &V(\max(x,N) - N)\Big), \end{aligned} \tag{11}$$

where P_d is given by (1). In (11), ρ is a constant, whose value we specify later. We prove (11) in the rest of this section. For any stationary policy denoted as π, the discounted cost over an infinite horizon with a discount factor $\theta \in (0, 1)$ for the

controlled process with the initial state x can be expressed as follows:

$$W^{\theta}(x,\pi) = E\left[\sum_{n=0}^{\infty}\theta^{n}((1-\mu_n)\lambda + CX_n)|X_0 = x\right].$$

The value function of the discounted problem is simply the minimum over all stationary control policies, and it is given by:

$$V^{\theta}(x) = \min_{\pi} W^{\theta}(x,\pi).$$

Given the transition probability function $p_{.|.}(\mu)$ of the controlled DTMC, the value function $V^{\theta}(x)$ can be characterized by the following DPE:

$$V^{\theta}(x) = \min_{\mu\in\{0,1\}}\left[Cx + (1-\mu)\lambda + \theta\sum_{y} p_{y|x}(\mu)V^{\theta}(y)\right].$$

Consider $\bar{V}^{\theta}(\cdot) = V^{\theta}(\cdot) - V^{\theta}(0)$. The above equation can be written as follows:

$$\bar{V}^{\theta}(x) = \min_{\mu\in\{0,1\}}\Bigg[Cx + (1-\mu)\lambda - (1-\theta)V^{\theta}(0) + \theta\sum_{y} p_{y|x}(\mu)\bar{V}^{\theta}(y)\Bigg]. \tag{12}$$

To calculate the value function $V(\cdot)$ and constant ρ in (11), we can use the value function of the θ discounted problem given in (12). In order to determine the above quantities, we can use the following lemma.

Lemma 1: The quantities $V(\cdot)$ and ρ satisfying (11) can be derived as: $\lim_{\theta\uparrow 1}\bar{V}^{\theta}(\cdot) = V(\cdot)$ and $\lim_{\theta\uparrow 1}(1-\theta)V^{\theta}(0) = \rho$. The constant ρ in (11) is unique and is equal to the optimal long-term expected average cost of the individual MDP at the AP. Under an optimal policy and the additional constraint $V(0) = 0$, $V(\cdot)$ also maintains uniqueness in states that are positive recurrent. For a given state x, the optimal choice of μ is obtained by finding the argmin of the RHS of (11).

Proof: The proof of this lemma is analogous to that of Lemma 4 in [35], with the difference being that we utilize the Lyapunov function introduced in Section III, which differs from the one used in [35]. For conciseness, the details of the proof are omitted. ■

V. STRUCTURAL PROPERTIES OF VALUE FUNCTION

In this section, we prove some structural properties of the value function, which are used later for proving the threshold nature of the optimal policy and the Whittle indexability of the problem.

Lemma 2: The function $V(\cdot)$ in (11) is an increasing function.

The proof of Lemma 2 is provided in Appendix A.

Lemma 3: The function $V(\cdot)$ in (11) has increasing differences, i.e., if:

$$z > 0,\ x_1 > x_2,\ \text{then } V(x_1+z) - V(x_1) \geq V(x_2+z) - V(x_2).$$

Here, z is a positive integer, and x_1 and x_2 are non-negative integers.

The proof of Lemma 3 is provided in Appendix B.

VI. THRESHOLD BEHAVIOR OF OPTIMAL POLICY

In this section, we show that the optimal policy is of threshold type. That is, under the optimal policy, there exists a state that serves as a threshold, determining whether an arrival is accepted or not for a given value of λ. For states at or below this threshold, user arrivals are accepted, whereas, for states above the threshold, they are rejected under the optimal policy. The value of this threshold state is determined by λ. The following lemma proves the threshold nature of the optimal policy.

Lemma 4: The optimal policy is a threshold policy.

Proof: Consider the function $f(x) = \mathbb{E}[V(x-D+\zeta)] - \mathbb{E}[V(x-D)]$, where x represents the number of users associated with the AP, D is the number of user departures with the probability mass function given in (1), and ζ is the number of user arrivals (see (3)). It follows from the definition of threshold policy that an optimal policy is a threshold policy if $f(x+w) \geq f(x)$ for $w > 0$ [35]. Hence, proving that $f(x+1) - f(x) \geq 0$ is sufficient. We prove this in the rest of this proof. We have:

$$\begin{aligned}&\mathbb{E}[V(x-D+\zeta)]\\ &= \sum_{\zeta=0}^{1} P_0 p_{\zeta} V(x+\zeta) + \sum_{d=1}^{\min(x-1,N-1)} P_d \sum_{\zeta=0}^{1} p_{\zeta} V(x+\zeta-d)\\ &\quad + P_{\min(x,N)}\sum_{\zeta=0}^{1} p_{\zeta} V(\max(N,x) - N + \zeta).\end{aligned} \tag{13}$$

Using the fact that $\sum_{d=0}^{\min(x,N)} P_d = 1$, we get:

$$P_{\min(x,N)} = 1 - P_0 - \sum_{d=1}^{\min(N-1,x-1)} P_d. \tag{14}$$

Using (14), (13) becomes:

$$\begin{aligned}&\mathbb{E}[V(x-D+\zeta)]\\ &= P_0\left[(1-p)V(x) + pV(x+1) - \sum_{\zeta=0}^{1} p_{\zeta} V(\max(x,N) - N + \zeta)\right]\\ &\quad + \sum_{d=1}^{\min(N-1,x-1)} P_d \sum_{\zeta=0}^{1} p_{\zeta}[V(x+\zeta-d) - V(\max(x,N) - N + \zeta)]\\ &\quad + \sum_{\zeta=0}^{1} p_{\zeta} V(\max(N,x) - N + \zeta).\end{aligned}$$

Similarly:

$$\begin{aligned}&\mathbb{E}[V(x-D)]\\ &= P_0[V(x) - V(\max(x,N) - N)] + V(\max(N,x) - N)\\ &\quad + \sum_{d=1}^{\min(x-1,N-1)} P_d[V(x-d) - V(\max(x,N) - N)].\end{aligned}$$

Therefore, $f(x)$ becomes:

$$\begin{aligned}f(x) &= \mathbb{E}[V(x-D+\zeta)] - \mathbb{E}[V(x-D)]\\ &= P_0\left[\sum_{\zeta=0}^{1} p_{\zeta}\{V(x+\zeta) - V(\max(x,N) - N + \zeta)\}\right.\end{aligned}$$

$$- (V(x) - V(\max(x,N)-N))\Big] + \sum_{d=1}^{\min(x-1,N-1)} P_d \times \Big[\sum_{\zeta=0}^{1} p_\zeta \{V(x+\zeta-d) - V(\max(x,N)-N+\zeta)\} - \{V(x-d) - V(\max(x,N)-N)\}\Big] + \sum_{\zeta=0}^{1} p_\zeta V(\max(x,N)-N+\zeta) - V(\max(x,N)-N)$$
$$= P_0\big[pV(x+1) - pV(x) + pV(\max(x,N)-N) - pV(\max(x,N)-N+1)\big] + \sum_{d=1}^{\min(x-1,N-1)} P_d \times \big[pV(x-d+1) - pV(x-d) - pV(\max(x,N)-N+1) + pV(\max(x,N)-N)\big] + pV(\max(x,N)-N+1) - pV(\max(x,N)-N). \quad (15)$$

Similarly:

$$f(x+1) = P_0\big[p\{V(x+2) - V(x+1) - (V(\max(x+1,N)-N+1) - V(\max(x+1,N)-N))\}] + \sum_{d=1}^{\min(x,N-1)} P_d \times \big[p\{V(x-d+2) - V(x-d+1) - V(\max(x+1,N)-N+1) + V(\max(x+1,N)-N)\}] + p\{V(\max(x+1,N)-N+1) - V(\max(x+1,N)-N)\}. \quad (16)$$

The maximum number of departures can be either x or N. Therefore, we have two cases.

Case 1: $\min(x+1,N) = x+1$, i.e., $\min(x,N-1) = x$ *or* $\min(x,N) = x$.
In this case:

$$f(x) = P_0\, p[V(x+1) - V(x) - (V(1) - V(0))] + \sum_{d=1}^{x-1} P_d\, p[V(x-d+1) - V(x-d) - (V(1) - V(0))] + p[V(1) - V(0)]. \quad (17a)$$

$$f(x+1) = P_0\, p[V(x+2) - V(x+1) - (V(1) - V(0))] + \sum_{d=1}^{x-1} P_d\, p[V(x-d+2 - V(x-d+1) - (V(1) - V(0))] + P_x\, p[V(2) - V(1) - (V(1) - V(0))] + p[V(1) - V(0)]. \quad (17b)$$

Subtracting (17a) from (17b):

$$f(x+1) - f(x) = P_0\, p[V(x+2) - V(x+1) - (V(x+1) - V(x))] + \sum_{d=1}^{x-1} P_d\, p[V(x-d+2) - V(x-d+1) - (V(x-d+1) - V(x-d))] + P_x\, p \times [V(2) - V(1) - (V(1) - V(0))].$$

From Lemma 3, it follows that $V(x+2-d) - V(x+1-d)$ is always greater than or equal to $V(x+1-d) - V(x-d)$. Similarly, $V(x+2) - V(x+1) \geq V(x+1) - V(x)$ and $V(2) - V(1) \geq V(1) - V(0)$. Thus, $f(x+1) - f(x) \geq 0$.

Case 2: $\min(x+1,N) = N$ *and* $\min(x,N) = N$.
As in Case 1, we can write:

$$f(x) = P_0\, p[V(x+1) - V(x)] + \sum_{d=1}^{N} P_d\, p[V(x+2-d) - V(x+1-d)]. \quad (18a)$$

$$f(x+1) = P_0\, p[V(x+2) - V(x+1)] + \sum_{d=1}^{N} P_d\, p[V(x+2-d) - V(x+1-d)]. \quad (18b)$$

Taking the difference between (18b) and (18a), we get:

$$f(x+1) - f(x) = P_0\, p[(V(x+2) - V(x+1)) - (V(x+1) - V(x))] + \sum_{d=1}^{N} P_d\, p[(V(x+2-d) - V(x+1-d)) - (V(x+1-d) - V(x-d))]. \quad (19)$$

From Lemma 3, it follows that $V(x+2-d) - V(x+1-d)$ is always greater than or equal to $V(x+1-d) - V(x-d)$. Similarly, $V(x+2) - V(x+1) \geq V(x+1) - V(x)$. This implies that the entire expression in (19) is always greater than or equal to zero; it follows that the optimal policy is a threshold policy in both cases. ■

Lemma 5: For the induced DTMC, let v_t be the stationary distribution under the threshold policy with threshold t. Then $\sum_{q=0}^{t} v_t(q)$ is an increasing function of t.

Proof: This result can be shown by using Lemma 4 and following steps similar to those in the proof of Lemma 8 in [35]. We omit the details for brevity. ■

VII. WHITTLE INDEXABILITY

Whittle indexability of the problem requires that as the value of λ decreases from ∞ to $-\infty$, the set of passive states (i.e., the states for which the AP rejects the arriving user, if any) monotonically increases from the empty set to the set of all possible states. To establish Whittle indexability, we need to introduce some supporting lemmas.

Lemma 6: Let $\mathfrak{g} : \mathbb{R} \times \mathbb{N} \to \mathbb{R}$ be submodular, i.e., $\forall \lambda_1 > \lambda_2$ and $x_1 > x_2$:

$$\mathfrak{g}(\lambda_2, x_1) + \mathfrak{g}(\lambda_1, x_2) \geq \mathfrak{g}(\lambda_1, x_1) + \mathfrak{g}(\lambda_2, x_2).$$

Also, let $x(\lambda) := \inf\{x^* : \mathfrak{g}(\lambda, x^*) \leq \mathfrak{g}(\lambda, x),\ \forall x\}$. Then $x(\lambda)$ is a non-decreasing function of λ.

Proof: The proof of this Lemma is presented in Section X-B of [64]. ∎

Lemma 7: Let the average cost of the threshold policy with threshold t, under the tax λ, be:

$$\mathfrak{g}(\lambda, t) = C\sum_{j=0}^{\infty} jv_t(j) \ + \lambda \sum_{j=t+1}^{\infty} v_t(j).$$

Then the function $\mathfrak{g}$ is submodular.

Proof: The proof is similar to that of Lemma 7 in [32] and is omitted for brevity. ∎

Theorem 2: The problem is Whittle indexable.

Proof: Under the unichain property [65], for any given stationary policy, say π, there exists a unique stationary distribution, denoted by v. Let Z represent the set of states at which the AP rejects an arriving user under the stationary policy π with a given λ. The optimal expected average cost under any stationary policy can be expressed as:

$$\rho(\lambda) = \inf_{\pi} \left\{ C\sum_{j=0}^{\infty} jv(j) + \lambda \sum_{j\in Z} v(j) \right\}.$$

By Lemma 4, $\rho(\lambda)$ is the infimum over all threshold policies. Therefore, $\rho(\lambda)$ can be expressed as:

$$\begin{aligned}\rho(\lambda) \ &= C\sum_{j=0}^{\infty} jv_{t(\lambda)}(j) + \lambda \sum_{j=t(\lambda)+1}^{\infty} v_{t(\lambda)}(j) \\ &:= w(\lambda, t(\lambda)),\end{aligned}$$

where $t(\lambda)$ is the optimal threshold for the given tax λ. By Lemma 7, w is submodular. So by Lemma 6, the threshold $t(\lambda)$ increases monotonically with λ. Under the optimal threshold policy, the set of passive states, Z, i.e., the set of states for which the AP does not admit the arriving user, if any, is $Z = \{t(\lambda) + 1, t(\lambda) + 2, \ldots, \infty\}$. Therefore, as λ increases from $-\infty$ to $+\infty$, Z monotonically decreases from the whole state space to the empty set. This proves the Whittle indexability of the problem. ∎

VIII. WHITTLE INDEX COMPUTATION

For a state x, the Whittle index λ is computed iteratively. Specifically, with $\tau \geq 0$ being the iteration number, the value of λ is updated according to the following equation:

$$\lambda_{\tau+1} = \lambda_\tau + \gamma \left(\sum_j p_a(j|x) V_{\lambda_\tau}(j) - \sum_j p_b(j|x) V_{\lambda_\tau}(j) - \lambda_\tau \right), \tag{20}$$

where $\gamma > 0$ represents a small step size. $V_{\lambda_\tau}(\cdot)$ is the value function, which appears in the DPE (11), for the tax λ_τ. $p_a(\cdot|x)$ and $p_b(\cdot|x)$ denote the transition probabilities under the events that the AP accepts an arriving user and does not accept an arriving user, respectively, in the current slot, given the present state x. The above iteration (20) for λ reduces the difference between the two arguments in the minimization function described in (11). So from (20), we observe that λ converges to a value at which the AP obtains equal expected utilities by accepting and not accepting the arriving user.

We find V_λ using a linear system of equations, which can be solved using the present value of λ after each iteration of (20). In particular, we solve the following system of equations for $V = V_{\lambda_\tau}$ and $\rho = \rho(\lambda_\tau)$, using the value of the parameter $\lambda = \lambda_\tau$:

$$V(y) = Cy - \rho + \sum_z p_a(z|y)V(z), y \leq x,$$

$$V(y) = Cy + \lambda - \rho + \sum_z p_b(z|y)V(z), y > x,$$

$$V(0) = 0.$$

For a given state x, the value of λ obtained after convergence of the iteration (20) is the Whittle index. To reduce the computational cost, we execute the iteration (20) only for a subset of the states x and then, to find the Whittle indices for the remaining states, we use interpolation. Under the proposed Whittle index based policy, the AP that has the smallest Whittle index admits the user in each time slot.

IX. POLICIES FROM PRIOR WORK FOR COMPARISON

This section reviews different user association policies described in [66] and [32]. In Section X, we compare the performance of the proposed Whittle index-based policy with that of the user association policies reviewed in this section through simulations.

A. RANDOM POLICY

Under the random policy, the arriving user, if any, in each time slot is associated with an AP selected uniformly at random from among the K APs.

B. LOAD BASED POLICY

Under the load based policy, the arriving user, if any, in a slot is associated with the AP with the minimum queue length at the slot's beginning, with ties being resolved randomly. That is, in slot n, the arriving user is admitted to AP $argmin_{i\in\{1,\ldots,K\}} X_n^i$, where X_n^i denotes the number of users associated with AP i at the beginning of slot n.

C. SIGNAL-TO-NOISE RATIO (SNR) BASED POLICY

Under the SNR policy, the arriving user, if any, in a time slot is admitted by the AP offering the highest SNR, with ties being broken randomly. That is, if an arrival occurs in slot n, then the AP chosen is $argmax_{i\in\{1,\ldots,K\}} s_i h_i$.

D. THROUGHPUT BASED POLICY

Under the throughput based policy, in any slot with a user arrival, the user is admitted to the AP that maximizes the throughput (ratio of the average data rate to number of associated users), which the user will get after association, with ties being broken randomly. That is, the chosen AP at the beginning of slot n is $argmax_{i\in\{1,\ldots,K\}} \frac{s_i h_i}{X_n^i + 1}$.

E. MIXED POLICY

The mixed policy associates the arriving user, if any, based on a weighted sum of the average data rate ($s_i h_i$) and the throughput ($\frac{s_i h_i}{X_n^i+1}$) that the user will get upon arrival, with ties being broken at random. Specifically, the chosen AP in slot n is $argmax_{i\in\{1,\ldots,K\}}(0.2 \cdot s_i h_i + \frac{s_i h_i}{X_n^i+1})$. Note that the throughput is scaled by a factor 0.2 since it was empirically found to result in the best performance of the policy in [67].

F. WHITTLE INDEX-BASED POLICY FROM PRIOR WORK

Reference [32] this is a Whittle index-based policy proposed in our prior work [32] for the user association problem formulated in that paper. A detailed comparison between the system model and results in [32] and those in this paper is provided in Section XI. The Whittle index-based policy proposed in [32] is adapted to address the problem in this paper by setting $r_i = s_i h_i$ and $L = N_i$ (see Section XI for the definitions of r_i and L) while computing the Whittle index of AP i and selecting the AP with the smallest Whittle index for user association in every time slot in which a user arrives.

X. SIMULATIONS

We use MATLAB simulations to assess how the proposed Whittle index-based user association policy performs relative to those briefly described in the previous section. The metrics that are considered for the performance comparison are as follows: long-term average cost, average delay, and Jain's fairness index (JFI) [44]. The average number of time slots for which a user stays in the system after its arrival is considered as its average delay. The JFI is defined as: $\frac{(\sum_{j=1}^{\mathbf{T}} D_j)^2}{\mathbf{T}\sum_{j=1}^{\mathbf{T}} D_j^2}$, where D_j is the average delay of the j^{th} user and $\mathbf{T}$ is the total number of users. The value of the JFI lies between 0 and 1. As the fairness in the distribution of the average delays of different users improves, the JFI increases [44].

We assume that initially, there are no users associated with any AP in the network, i.e., the initial state of each AP is zero. The simulations run for a period (T) of 20, 000 time slots. We analyze the average cost incurred under the seven policies during the final 10, 000 time slots and the results are plotted. This is done in order to be able to compare the long-term costs, and ignore the short-term cost variations. Let $\mathbf{h} = [h_1, \ldots, h_K]$, $\mathbf{C} = [C_1, \ldots, C_K]$, $\mathbf{N} = [N_1, \ldots, N_K]$ and $\mathbf{s} = [s_1, \ldots, s_K]$, where h_i, C_i, N_i, and s_i are as defined in Section II. Let *MAX_USERS* be the maximum number of users that can be associated with a given AP. Table 1 shows the time-averaged signal-to-interference-plus-noise ratio (SINR) values, of the link from an AP to its associated user, corresponding to some representative values of s_i and h_i used in the simulations.

Figs. 2-6 and Table 2 show that the proposed Whittle index-based policy gives the lowest average cost relative to every other policy for all the parameter values considered. Also, the SNR-based policy is the least effective of all the policies considered in terms of the average cost, which is consistent with intuition, since it does not balance the load across different APs. Fig. 7 shows that the average delay is the minimum for the proposed Whittle index based policy among all the user association policies considered. Finally, Fig. 8 shows that the proposed Whittle index-based policy achieves a higher JFI as compared to all the other policies. Our results show that in several scenarios, the performance of the Mixed policy and Load based policy is quite good– in some cases, they perform close to our proposed Whittle index based policy. Intuitively, this is because the Load based policy selects an AP to assign an arriving user to by taking into account the current load at every AP. It assigns each arriving user to the currently least loaded AP, and hence effectively balances the load across different APs. Also, the Mixed policy assigns each arriving user to an AP with which it will get a high throughput and/ or average data rate; hence, users tend to be served at high throughput and average data rates under this policy. In addition, note that

TABLE 1. The table shows the time-averaged SINR values corresponding to some representative values of s_i and h_i used in the simulations. Path loss models from Millimeter-Wave Based Mobile Radio Access Network for 5G Integrated Communications (mmMAGIC) [68] and the following parameter values were used for the calculation of these average SINR values: transmitted power = 24 dBm, Tx/Rx gain = 20 dBi, carrier frequency = 28 GHz, bandwidth = 200 MHz, average interference-plus-noise power = −91 dBm, transmitter-receiver separation = 100 m, AP height = 10 m, and height of user equipment (UE) = 1.5 m. Also, Line of Sight (LOS) and Non-Line of Sight (NLOS) path loss values were calculated from Table IV, p. 6220, [69], and these values are 101.43 dB and 150 dB, respectively. As in the model [68], the worst case attenuation values due to mild fading and severe fading are 6 dB and 20 dB, respectively.

s_i	h_i	SINR (dB)
0.23	0.25	32.74
0.23	0.132	29.96
0.23	0.19	31.54
0.18	0.25	31.67
0.18	0.132	28.88
0.18	0.19	30.48
0.14	0.25	30.58
0.14	0.132	27.30
0.14	0.19	29.39

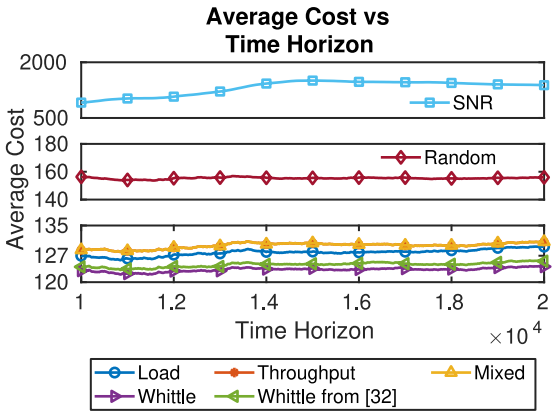

FIGURE 2. The figure shows a comparison of the average costs incurred under the seven association policies. The parameter values used are as follows: $K = 5$, $p = 0.25$, *MAX_USERS* = 150, N = [7, 6, 5, 7, 8], h = [0.25, 0.24, 0.23, 0.22, 0.21], C = [80, 75, 70, 65, 60], and s = [0.18, 0.17, 0.16, 0.17, 0.18].

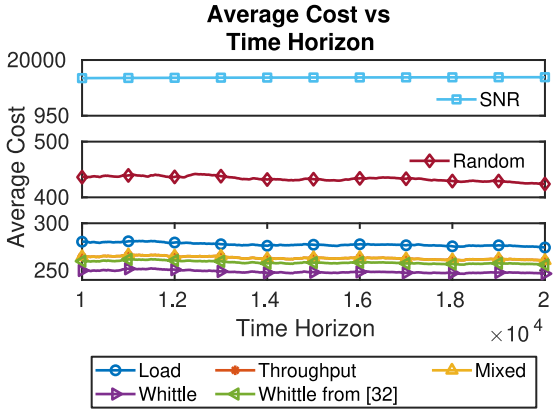

FIGURE 3. The figure shows a comparison of the average costs incurred under the seven association policies. The parameter values used are as follows: $K = 8$, $p = 0.45$, $MAX\_USERS = 200$, h = [0.16, 0.156, 0.152, 0.148, 0.144, 0.14, 0.136, 0.132], C = [81, 77, 73, 69, 65, 61, 57, 53], N = [5, 7, 7, 5, 4, 6, 5, 7], and s = [0.17, 0.21, 0.23, 0.22, 0.19, 0.18, 0.17, 0.16].

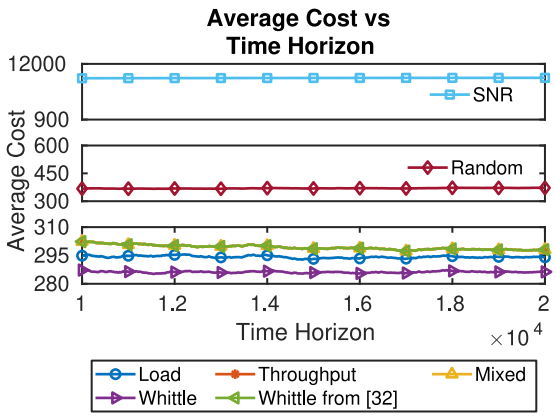

FIGURE 4. The figure shows a comparison of the average costs incurred under the seven association policies. The parameter values used are as follows: $K = 10$, $p = 0.4$, $MAX\_USERS = 100$, N = [6, 6, 6, 6, 6, 6, 6, 6, 6, 6], h = [0.18, 0.176, 0.172, 0.168, 0.164, 0.160, 0.156, 0.152, 0.148, 0.144], C = [95, 92, 89, 86, 83, 80, 77, 74, 71, 68], and s = [0.18, 0.178, 0.176, 0.174, 0.172, 0.17, 0.168, 0.166, 0.164, 0.162].

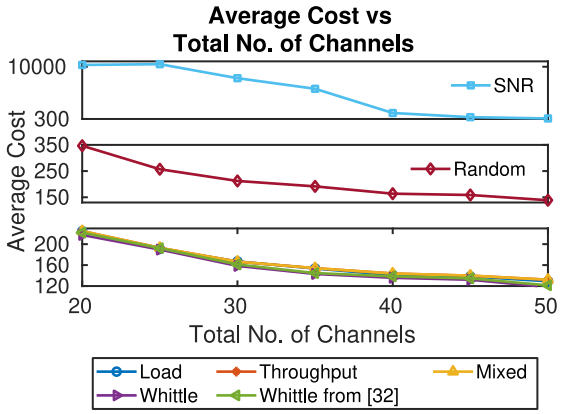

FIGURE 5. The figure shows a comparison of the average costs incurred under the seven association policies. The parameter values used are as follows: $K = 5$, $p = 0.2$, $MAX\_USERS = 150$, h = [0.18, 0.175, 0.17, 0.165, 0.16], C = [90, 85, 80, 75, 70], s = [0.14, 0.15, 0.14, 0.15, 0.14], and different total numbers of channels, N_{total}, varying from 20 to 50. For each value of N_{total}, every AP has the same number of channels.

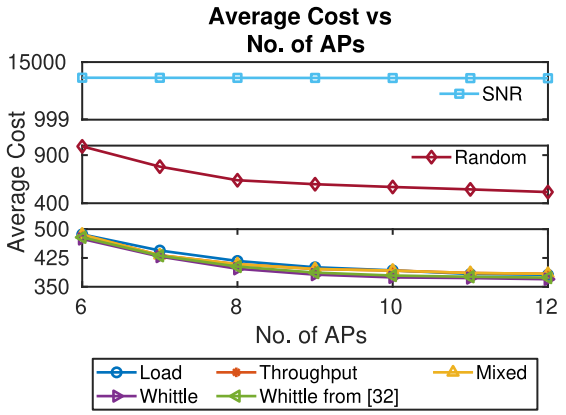

FIGURE 6. The figure shows a comparison of the average costs incurred under the seven association policies. The parameter values used are as follows: $p = 0.6$, $MAX\_USERS = 150$, and K varies from 6 to 12. For $K = 6$, the parameter values used are as follows: h = [0.18, 0.176, 0.172, 0.168, 0.164, 0.16], C = [79, 78, 77, 76, 75, 74], N = [7, 6, 5, 6, 7, 7], and s = [0.17, 0.16, 0.15, 0.16, 0.17, 0.16]. For every subsequent addition of the i^{th} AP, the values of h_i and C_i are selected as $0.174 - 0.04i$ and $80 - i$, respectively, and the values of N_i and s_i are 7 and 0.17, respectively, for every $i \in \{6, 7, \ldots, 15\}$.

the Mixed policy is a heuristic carefully designed based on insights obtained from the study of the optimal solution of an MDP model of a simple network with two APs in [70]. The good performance of the Mixed and Load based policies can be explained by the above reasons. However, note that in several scenarios, the proposed Whittle index based policy significantly outperforms both the Mixed policy and the Load based policy. For example, in Figs. 2, 3 and 4, the average costs incurred by the proposed Whittle based policy are significantly less than those incurred by the Mixed and Load based policies. In summary, our proposed Whittle index-based policy achieves the best performance among all the policies in terms of the average cost, average delay, as well as JFI.

TABLE 2. The table shows the average costs under the seven association policies vs the arrival probability (p). The following parameter values are used: $K = 6$, $MAX\_USERS = 50$, h = [0.2, 0.195, 0.19, 0.185, 0.18, 0.175], C = [79, 78.5, 78, 77.5, 77, 76.5], s = [0.16, 0.15, 0.14, 0.18, 0.18, 0.16], $N = [7, 6, 5, 6, 8, 7]$, and different values of p varying from 0.1 to 0.9.

Arrival Probability (p)	Average Cost						
	Load	SNR	Throughput	Random	Mixed	Whittle	Whittle [32]
0.1	66	106.7	59.5	70.6	59.5	57.3	58.6
0.2	132.6	1682.8	122	157.4	122	118.9	121.5
0.3	199.8	3395.2	187.7	267.8	187.7	184.7	186.2
0.4	270.7	3609.6	256.9	387.7	256.9	253.8	256.4
0.5	343.7	3650.8	330.6	555.8	330.6	326.8	330.6
0.6	441.9	3679.8	423.9	895.6	423.9	417.7	421.8
0.7	542.3	3714.7	526.4	1346.3	526.8	520.3	524.7
0.8	690.5	3727.6	673.9	2888.7	672.9	662.1	669.3
0.9	861.8	3738.5	838.4	4366.5	840.1	827	835.5

XI. RELATED WORK

This section provides an overview of the existing research on user association schemes in different types of wireless networks.

In [10], the user association problem in massive multiple-input and multiple-output (MIMO) enabled heterogeneous networks (HetNets) was formulated as a network logarithmic utility maximization problem with the objective of jointly

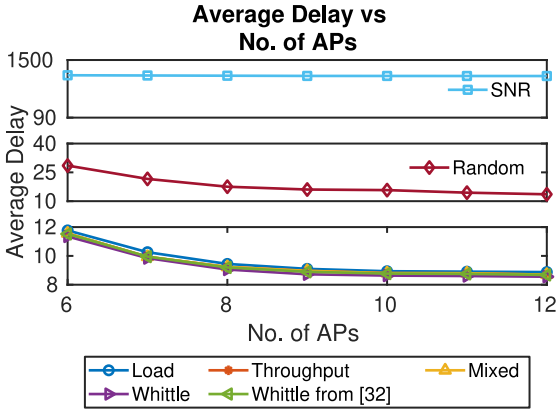

FIGURE 7. The figure shows a comparison of the average delays incurred under the seven association policies. The parameter values used are as follows: $p = 0.6$, $MAX\_USERS = 150$, and K varies from 6 to 12. For $K = 6$, the parameter values used are as follows: h = [0.18, 0.176, 0.172, 0.168, 0.164, 0.16], C = [79, 78, 77, 76, 75, 74], N = [7, 6, 5, 6, 7, 7], and s = [0.17, 0.16, 0.15, 0.16, 0.17, 0.16]. For every subsequent addition of the i^{th} AP, the values of h_i and C_i are selected as $0.174 - 0.04i$ and $80 - i$, respectively, and the values of N_i and s_i are 7 and 0.17, respectively, for every $i \in \{6, 7, \ldots, 12\}$.

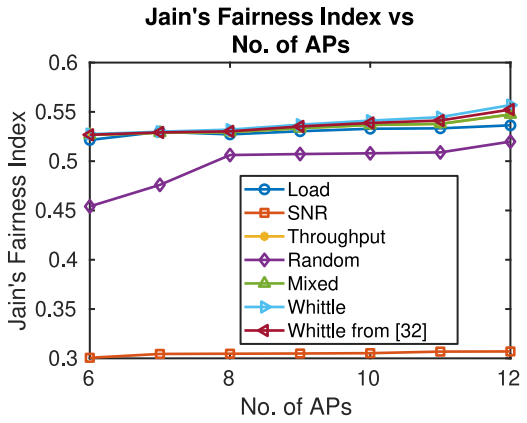

FIGURE 8. The figure shows a comparison of the Jain's fairness indices under the seven association policies. The parameter values used are as follows: $p = 0.6$, $MAX\_USERS = 150$, and K varies from 6 to 12. For $K = 6$, the parameter values used are as follows: h = [0.18, 0.176, 0.172, 0.168, 0.164, 0.16], C = [79, 78, 77, 76, 75, 74], N = [7, 6, 5, 6, 7, 7], and s = [0.17, 0.16, 0.15, 0.16, 0.17, 0.16]. For every subsequent addition of the i^{th} AP, the values of h_i and C_i are selected as $0.174 - 0.04i$ and $80 - i$, respectively, and the values of N_i and s_i are 7 and 0.17, respectively, for every $i \in \{6, 7, \ldots, 12\}$.

achieving EE and user fairness. The joint power allocation and user association problem, with the objective of minimizing the total transmit power, in a multi-cell massive MIMO downlink system was addressed in [11]. Lower bounds on the ergodic channel capacity for general channels and precoding, as well as for Rayleigh fading channels and zero-forcing precoders, were derived. Using these bounds, a power minimization problem was formulated. The problem was proved to be solvable by linear programming. In [12], the EE-SE trade-off in massive MIMO networks for the proportional rate fairness problem was addressed. The authors formulated a multi-objective problem, transformed it into a single-objective problem, and proposed an effective decomposition algorithm for an efficient solution. In [13], an association strategy to maximize the weighted sum EE in MIMO-enabled HetNets was proposed. The proposed approach utilizes a multi-layer iterative method to address the formulated problem. The outer layer employs a Newton-like method to optimize EE parameters and signal-to-interference-and-noise ratio (SINR) constraints, while the inner layer leverages the Lagrange multiplier method for association index optimization. In [14], the AP to device association problem in a cell-free massive MIMO system was formulated with the objective of maximizing the EE subject to minimum rate constraints for all the devices. The formulated non-convex problem was solved by model-free deep reinforcement learning (DRL) methods. In [15], the joint user association and power control problem in HetNets was formulated as a mixed-integer programming problem. To handle the non-convexity, the authors leveraged Lagrange duality, decomposing the problem into sub-problems for user association and power control, which were solved iteratively. The results show that the proposed algorithm outperforms an association algorithm that assumes equal power distribution. In [16], the user association problem in HetNets was formulated as a local optimization problem at the macro base station (BS), using limited channel state information (CSI) feedback, where users report synchronization signal power instead of the full CSI. This enables the design of a low-complexity successive offloading scheme with near-optimal performance. The proposed scheme improves upon conventional approaches that use equal power allocation. In [17], a framework for user association in multi-tier HetNets was proposed. The authors proposed two online algorithms: a centralized and a semi-distributed approach. Both algorithms leverage the reinforcement learning (RL) multi-armed bandit (MAB) technique to achieve efficient user association while ensuring a balanced load distribution across BSs. In [18], a joint user association and power allocation scheme for cell-free visible light communication (VLC) networks was proposed. This problem was formulated as a non-convex network utility maximization problem, aiming to simultaneously improve network performance metrics such as user fairness, load balancing, and power control. In [19], a novel user association and resource block (RB) allocation scheme for non-orthogonal multiple access (NOMA) based HetNets, aiming for both fairness and SE, was proposed. By decomposing the optimization problem into manageable sub-problems, it determines the optimal RB allocation for requests at small base stations (SBSs), followed by NOMA group formation and RB allocation within each SBS for maximized SE. Simulations show significant improvements in the Jain's fairness index and SE compared to conventional approaches. However, unlike the work presented in this paper, none of the works [10], [11], [12], [13], [14], [15], [16], [17], [18], [19] investigates the user association problem in a mmWave network with multi-channel APs.

In [25], a multi-objective method for joint user association and beamforming in mmWave networks was proposed. The

considered approach was based on a Fuzzy Interference System (FIS) and aimed to optimize the system throughput, latency, and remote radio unit (RRU) load balancing. In [26], a reconfigurable intelligent surface (RIS)-user association framework that maximizes user data rates while balancing the load across the system was proposed. This formulation utilizes a centralized multi-player MAB algorithm, where mmWave users, RIS boards, and achievable user rates act as players, arms, and rewards, respectively. The proposed algorithms are based on the upper confidence bound (UCB) algorithm, and numerical results show the superiority of the proposed algorithms over the maximum received power-based RIS-user association scheme. In [27], to tackle the issue of blockage in mmWave networks, a hybrid association scheme within a user-centric architecture was introduced. The hybrid association scheme performs mmWave BS (mBS)-user association and mBS-RIS-user association. To mitigate blockage, a user is allowed to connect to multiple mBSs simultaneously and line-of-sight (LOS) links are constructed using RISs. The aim is to jointly optimize user association and power allocation across multiple links to maximize the user's sum rate, effectively mitigating the blocking effect using a low-complexity alternating iteration algorithm based on Lagrangian duality and fractional programming. This iterative algorithm is used to solve the sub-problems of power allocation, mBS-user association, and mBS-RIS-user association. In [28], a decentralized user association technique for mmWave and terahertz (THz) Ultra Dense Networks (UDNs) that uses multi-agent actor-critic (AC) RL to optimize the EE was provided. In the proposed approach, the actor-network determines the user association, while the critic network provides feedback on the EE of the user association decisions to the actor-network. Using this interaction, DRL agents in each SBS learn user association strategies that maximize the overall network's EE. The EE gain of the proposed scheme is higher than that of conventional user association schemes. The joint user association and transmission scheduling problem in mmWave train-ground communication systems was addressed in [29]. The system employed mobile relays (MR) and these relays operated in full-duplex (FD) mode. The problem was formulated as a non-linear programming (NLP) problem, with the aim of maximizing system throughput and maintaining the Quality of Service (QoS) requirements of the users. A novel algorithm was proposed, which was based on coalitional game theory. In [30], a user association algorithm in a multi-cell mmWave network assisted with multiple RISs was addressed. The proposed association algorithm considers cell interference. The problem was formulated as a non-convex mixed-integer optimization problem, with the objective of jointly optimizing user-mBS association, active beamforming in mBSs, and passive beamforming at the RISs to maximize the network SE. The proposed iterative algorithm is based on alternating optimization (AO) and solves the problem by decoupling it into separate steps of optimizing active beamforming, passive beamforming, and user association. However, unlike the work presented in this paper, none of the above papers [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30] addresses the user association problem in mmWave networks using the theory of Whittle index.

The closest to this paper are our prior works [31], [32], which used the Whittle index to solve the user association problem in mmWave networks, and [35], which addressed job allocation in processor sharing systems using the Whittle index. Although some proofs in this work are similar to those in [31], [32], and [35], there are significant differences. In particular, unlike [35], which explores job allocation in egalitarian processor-sharing systems, this work addresses user association in mmWave networks, due to which the system model in this paper is substantially different from that in [35].

The system model in this paper differs from that in [32] in several respects: First, in the system model in [32], each mBS has a single channel. In contrast, in the system model in this paper, an AP can have multiple channels, and can serve different users simultaneously on different channels. Also, the different channels of an AP are susceptible to both blockage, which is common to all the channels of an AP, and frequency-selective fading, which is, in general, different for different channels. Neither blockage nor frequency-selective fading are considered in the system model in [32]. Second, in the system model in [32], each time slot is divided into L mini-slots. Also, in each mini-slot, a packet departs w.p. r_i from mBS i. So the potential number of departures from an mBS i in a time slot has a simple distribution– it is Binomially distributed. In contrast, in the system model in this paper, the potential number of departures from an AP i in a time slot has a much more complicated distribution, which is non-standard, and is given in (1). Third, in the system model in [32], the maximum number of departures from an mBS in a time slot is the same for every mBS and equals L, which is independent of i. In contrast, the maximum number of departures from AP i in the system model in this paper is N_i, which depends on i. Fourth, in the generalized version of the system model in this paper (see Remark 2 in Section II), in each time slot, a user served by AP i receives error-free service w.p. $e_i \in (0, 1)$. Also, after a user arrives into the network, the number of packets that it exchanges with its AP before it departs is a random variable that is geometrically distributed with parameter g. The statistics of the departure process from an AP i are affected by these facts and are functions of the parameters e_i and g. In contrast, the statistics of the departure process in the system model in [32] do not depend on any such parameters. Fifth, the number of departures in a time slot in the system model in this paper is a function of five parameters– g and the following four parameters, which are all different for different APs in general– (i) s_i, (ii) h_i, (iii) N_i, and (iv) e_i. In contrast, in the system model in [32], apart from L, which is the same for all mBSs, the number of departures from an mBS i is a function of only one parameter, viz., r_i. Due to the above differences between the system models in this

paper and [32], there are several differences between the results in this paper and those in [32]: The system evolution equation in this paper (see (2)) is expressed in terms of the number of *users*, X_n^i, associated with AP i at the beginning of slot n. In contrast, the system evolution equation in [32], in which each arriving user brings a file of random size, is expressed in terms of the total number of *packets* at the users associated with mBS i at the beginning of slot n. The DPE in this paper (see (11)) is significantly different from that in [32] since the arrival and departure processes in the system model in this paper are very different from those in [32] as explained above. Note that the methodologies for proving Whittle indexability in this paper, in [32], and in a large number of existing papers, e.g., [35], [36], [37], [38], [41], [42], on this topic are broadly similar, e.g., proving that the induced DTMC corresponding to each bandit is stable, the value function is monotone and has monotone differences, the optimal policy is threshold type, etc. However, the details of the proofs of the above results in this paper differ significantly from those in [32] due to the above differences in the system models. The techniques used to prove several of the results in this paper significantly differ from those used to prove the corresponding results in [32]. For instance, the fact, stated above, that the potential number of departures from an mBS in a time slot in the system model in [32] has a Binomial distribution, is exploited in the analysis in that paper, e.g., by using standard identities about Binomial coefficients. However, since the potential number of departures from an AP in the system model in this paper does not have a Binomial distribution, we cannot use such standard identities. In [32], it is proved that the value function has non-decreasing differences by relaxing the state and action spaces to allow them to take real values and then proving the convexity of the value function for the relaxed problem. On the other hand, for the system model in this paper, it is difficult to use such a relaxation due to the complicated departure process, and instead, it is proved (Lemma 3) that the value function has increasing differences by using a direct approach and a detailed case-by-case analysis. The specific stability condition assumed in this paper is totally different from that assumed in [32]. In particular, in this paper, the stability condition states that p should be less than a certain function of the parameters s_i, h_i, and N_i of different APs i (see Theorem 1), whereas that in [32] states that the mean number of arrivals should be less than a certain function of the parameter L and the smallest serving rate, r_K. The simulation setup and results in this paper are very different from those in [32]. For example, in this paper, the JFI is calculated with respect to the average delays experienced by individual users, whereas in [32], the JFI is calculated with respect to the average throughput achieved by individual users. The parameters h_i, s_i, and N_i are considered in the simulations in this paper, whereas no such parameters are considered in the simulations in [32]. In [32], the SNR-based, load-based, throughput-based, and mixed association policies, whose performance is compared with that of the proposed policy via simulations, take decisions based on the values of X_n^i and/ or r_i of different mBSs i. On the other hand, in the simulations in this paper, these association policies take decisions based on the values of X_n^i, s_i, and/ or h_i of different APs i. Also, the number of simulation runs over which the results in this paper are averaged are comparatively higher than those in [32]. Finally, in this paper, we have compared the performance of the proposed Whittle index-based policy with that of the Whittle index-based policy proposed in [32] via extensive simulations. Our simulation results show that the proposed Whittle index-based policy outperforms the Whittle index-based policy proposed in [32] in terms of the average cost, average delay, as well as JFI. Note that the system model in [31] is a special case of that in [32] in which the number of mini-slots, L, equals 1 and each arriving user brings exactly one packet. Hence, all the above differences also hold between the system model and results in [31] and those in this paper. Due to the above differences, the system model and results in this paper differ considerably from those presented in [31], [32], and [35].

XII. CONCLUSION AND FUTURE WORK

We studied the user association problem in a dense mmWave network, in which each AP has multiple channels, and the different channels of an AP are susceptible to both blockage, which is common to all the channels of an AP, and frequency-selective fading, which is, in general, different for different channels. We formulated the problem as a RMAB problem and rigorously proved its Whittle indexability. Moreover, we proposed a policy that associates an arriving user in a slot with the AP having the lowest Whittle index. Through detailed simulations, we demonstrated that our proposed Whittle index-based user association policy outperforms several existing policies proposed in prior work in terms of average cost, average delay, as well as JFI. The results presented in this paper can be generalized as future work to encompass scenarios where each AP has multiple channels and time slots are subdivided into mini-slots, in each of which a user may be served.

APPENDIX

A. PROOF OF LEMMA 2

We use mathematical induction for proving the lemma. Consider the DPE with a finite horizon of l steps and a discount factor θ; then with $0 \leq y < l$:

$$V_y^\theta(x) = \min_{\mu\in\{0,1\}} \left[c(x,\mu) + \theta \sum_{\zeta=0}^{1} \left\{ \sum_{d=0}^{\min(x,N)} V_{y-1}^\theta(x-d+\mu\zeta) P_d(x) \right\} p_\zeta \right], \tag{21}$$

where $c(x,\mu) = Cx + (1-\mu)\lambda$ and $V_0^\theta(x) = Cx, x \geq 0$. In (21), $P_d(x)$ represents the probability of d users departing from the AP when there are currently x users associated with

the AP, and is given in (1), and p_ζ is given by the following expression:

$$p_\zeta = \begin{cases} p, & \text{if a user arrives at the end of slot } n, \\ 1-p, & \text{otherwise.} \end{cases} \quad (22)$$

Clearly, $V_0^\theta(x_1) > V_0^\theta(x_2)$ for all $x_1 > x_2$ and $x_1, x_2 \geq 0$. Assume that:

$$V_{l-1}^\theta(x_1) > V_{l-1}^\theta(x_2), \ \forall x_1 > x_2 \text{ and } x_1, x_2 \geq 0. \quad (23)$$

x_2 can be greater than or equal to N, or it can be less than N. In each case, we will show that $V_l^\theta(x_1) > V_l^\theta(x_2)$ for all $x_1 > x_2$ and $x_1, x_2 \geq 0$. From (21), we can write:

$$V_l^\theta(x) = \min_{\mu\in\{0,1\}} \Bigg[c(x,\mu) + \theta \sum_{\zeta=0}^{1} \Bigg\{ P_0 V_{l-1}^\theta(x+\mu\zeta) + \sum_{d=1}^{\min(x,N)} V_{l-1}^\theta(x-d+\mu\zeta) P_d(x) \Bigg\} p_\zeta \Bigg], \quad (24)$$

where $P_0 = s(1-h)^N + (1-s)$ represents the probability that there are no departures. When the current state is x_1 (respectively, x_2), let the minimum value in the RHS of (24) be obtained at μ_1 (respectively, μ_2), i.e.,:

$$V_l^\theta(x_1) = c(x_1,\mu_1) + \theta \sum_{\zeta=0}^{1} \Bigg[P_0 V_{l-1}^\theta(x_1+\mu_1\zeta) + \sum_{d=1}^{\min(x_1,N)} V_{l-1}^\theta(x_1-d+\mu_1\zeta) P_d(x_1) \Bigg] p_\zeta, \quad (25a)$$

$$V_l^\theta(x_2) = c(x_2,\mu_2) + \theta \sum_{\zeta=0}^{1} \Bigg\{ P_0 V_{l-1}^\theta(x_2+\mu_2\zeta) + \sum_{d=1}^{\min(x_2,N)} V_{l-1}^\theta(x_2-d+\mu_2\zeta) P_d(x_2) \Bigg\} p_\zeta. \quad (25b)$$

x_2 can be greater than or equal to N, or it can be less than N. Therefore, we have two cases.

Case 1: $0 \leq x_2 < N$ *and* $x_1 > x_2$.

For any $x_1 > x_2$, we can write:

$$\sum_{d=1}^{\min(x_1,N)} (\cdot) P_d(x_1) = \sum_{d=1}^{x_2-1} (\cdot) P_d(x_1) + \sum_{d=x_2}^{\min(x_1,N)} (\cdot) P_d(x_1). \quad (26)$$

The admission controls, μ_1 and μ_2, are constrained to binary values. In the case where both controls are identical, i.e., $\mu_1 = \mu_2 = \mu$ (say), μ can have a value of either 0 or 1. Then, using (26), we can write the difference between (25a) and (25b) as:

$$\begin{aligned} &V_l^\theta(x_1) - V_l^\theta(x_2) \\ &= \theta \sum_{\zeta=0}^{1} P_0 \Big[V_{l-1}^\theta(x_1+\mu\zeta) - V_{l-1}^\theta(x_2+\mu\zeta) \Big] p_\zeta + \theta \times \\ &\quad \sum_{\zeta=0}^{1} \sum_{d=1}^{x_2-1} \Big[V_{l-1}^\theta(x_1-d+\mu\zeta) - V_{l-1}^\theta(x_2-d+\mu\zeta) \Big] \bar{P} p_\zeta \\ &\quad + \theta \sum_{\zeta=0}^{1} \Bigg[\sum_{d=x_2}^{\min(x_1,N)} V_{l-1}^\theta(x_1-d+\mu\zeta) P_d(x_1) \\ &\quad - V_{l-1}^\theta(\mu\zeta) P_{x_2}(x_2) \Bigg] p_\zeta + c(x_1-x_2), \end{aligned} \quad (27)$$

where $\bar{P} = P_d(x_1) = P_d(x_2) = s\binom{N}{d}(h)^d(1-h)^{N-d}$, which is valid for the case when, in a time slot, the number of departing users, d, is strictly lower than the lowest current state, x_2. When the current state is $x \leq N$, then $P_x(x) = \sum_{d=x}^{N} \bar{P}$. Since $x_1 > x_2$ and using (23), we can say that the first, second, and fourth terms of the RHS of (27) are always non-negative. To determine whether the third term of the RHS of (27) is positive or not, we need to consider two cases. If $x_1 > N$, then the third term of the RHS of (27) is equivalent to:

$$\theta \sum_{\zeta=0}^{1} \Bigg[\sum_{d=x_2}^{N} \big\{ V_{l-1}^\theta(x_1-d+\mu\zeta) - V_{l-1}^\theta(\mu\zeta) \big\} \bar{P} \Bigg] p_\zeta,$$

where $\bar{P}$ is the same as stated before. As the number of departures, d, is always less than x_1, using (23) leads to the conclusion that the third term of the RHS of (27) is positive; hence, $V_l^\theta(x_1) > V_l^\theta(x_2)$ holds for all $x_1 > x_2$.

If $x_1 \leq N$, then the third term of the RHS of (27) is equivalent to:

$$\theta \sum_{\zeta=0}^{1} \Bigg\{ \sum_{d=x_2}^{x_1-1} \big[V_{l-1}^\theta(x_1-d+\mu\zeta) - V_{l-1}^\theta(\mu\zeta) \big] \bar{P} \Bigg\} p_\zeta.$$

As $d < x_1$ and by (23), the above quantity is also non-negative. Hence we can conclude that $V_l^\theta(x_1) > V_l^\theta(x_2)$ for all $x_1 > x_2$. So far, we have assumed that $\mu_1 = \mu_2$. Let us consider the remaining cases. Suppose $\mu_1 = 1$ and $\mu_2 = 0$. Then, from the above results for the case where $\mu_1 = \mu_2$, it follows that:

$$V_l^\theta(x_1)\Big|_{\mu_1=0} \geq V_l^\theta(x_1)\Big|_{\mu_1=1} > V_l^\theta(x_2)\Big|_{\mu_2=1} \geq V_l^\theta(x_2)\Big|_{\mu_2=0}.$$

Thus, we can say that $V_l^\theta(x_1) > V_l^\theta(x_2)$. Now, suppose $\mu_1 = 0$ and $\mu_2 = 1$. Similar to the above, we get that:

$$V_l^\theta(x_1)\Big|_{\mu_1=1} \geq V_l^\theta(x_1)\Big|_{\mu_1=0} > V_l^\theta(x_2)\Big|_{\mu_2=0} \geq V_l^\theta(x_2)\Big|_{\mu_2=1}.$$

Hence, we can conclude that $V_l^\theta(x_1) > V_l^\theta(x_2), \forall \mu_1, \mu_2$.

Case 2: $x_2 \geq N$ *and* $x_1 > x_2$.

For this case, (25) becomes:

$$V_l^\theta(x_1) = c(x_1,\mu_1) + \theta \sum_{\zeta=0}^{1} \Bigg\{ P_0 V_{l-1}^\theta(x_1+\mu_1\zeta) + \sum_{d=1}^{N} V_{l-1}^\theta(x_1-d+\mu_1\zeta) \bar{P} \Bigg\} p_\zeta, \quad (28a)$$

$$V_l^\theta(x_2) = c(x_2,\mu_2) + \theta \sum_{\zeta=0}^{1} \Bigg\{ P_0 V_{l-1}^\theta(x_2+\mu_2\zeta) + \sum_{d=1}^{N} V_{l-1}^\theta(x_2-d+\mu_2\zeta)\bar{P} \Bigg\} p_\zeta, \quad (28b)$$

where $\bar{P}$ is the same as stated above. Similar to Case 1, consider all possible values of μ_1 and μ_2. For $\mu_1 = \mu_2 = \mu$ (say), the difference between (28a) and (28b) becomes:

$$\begin{aligned} &V_l^\theta(x_1) - V_l^\theta(x_2) \\ &= \theta \sum_{\zeta=0}^{1} P_0 \big[V_{l-1}^\theta(x_1+\mu\zeta) - V_{l-1}^\theta(x_2+\mu\zeta) \big] p_\zeta \\ &\quad + c(x_1,x_2) + \theta \sum_{\zeta=0}^{1} \sum_{d=1}^{N} \big[V_{l-1}^\theta(x_1-d+\mu\zeta) \\ &\qquad - V_{l-1}^\theta(x_2-d+\mu\zeta) \big] \bar{P} \big\} p_\zeta. \end{aligned}$$

As $x_1 > x_2$, from (23), we can say that $V_l^\theta(x_1) > V_l^\theta(x_2)$. As in Case 1, we can prove the validity of the inequality for different cases of μ_1 and μ_2.

Taking limits as $l \uparrow \infty$, the above inequality remains valid for the infinite horizon, i.e., $V^\theta(x_1) \geq V^\theta(x_2)$. Further, by taking limits as $\theta \uparrow 1$, we can conclude that:

$$V(x_1) \geq V(x_2) \ \forall x_1 > x_2 \text{ and } x_1 x_2 \geq 0.$$

The result follows.

B. PROOF OF LEMMA 3

Consider the DPE in (21) with the same parameters as before. With $x_1 > x_2$ and $z > 0$, let us prove the following inequality using induction:

$$V_y^\theta(x_1+z) - V_y^\theta(x_1) - V_y^\theta(x_2+z) + V_y^\theta(x_2) \geq 0. \quad (29)$$

It is clearly valid for $y = 0$; let us assume that it is also valid for $y = l-1$. Thus:

$$V_{l-1}^\theta(x_1+z) - V_{l-1}^\theta(x_1) - V_{l-1}^\theta(x_2+z) + V_{l-1}^\theta(x_2) \geq 0. \quad (30)$$

Define $W^1 = V_l^\theta(x_1+z) - V_l^\theta(x_1)$ and $W^2 = V_l^\theta(x_2+z) - V_l^\theta(x_2)$. Thus, our objective is to show that $W^1 - W^2 \geq 0$. Let the minimum value of $V_l^\theta(.)$ at points x_1+z, x_1, x_2+z and x_2 be obtained at the controls μ_1, μ_2, μ_3 and μ_4, respectively. Note that these controls can take only binary values. By (21), we get:

$$\begin{aligned} &W^1 - W^2 \\ &= (\mu_2 - \mu_1 - \mu_4 + \mu_3)\lambda \\ &\quad + \theta \sum_{\zeta=0}^{1} \big[P_0 \big\{ V_{l-1}^\theta(x_1+z+\mu_1\zeta) - V_{l-1}^\theta(x_1+\mu_2\zeta) \\ &\quad - V_{l-1}^\theta(x_2+z+\mu_3\zeta) + V_{l-1}^\theta(x_2+\mu_4\zeta) \big\} \big] p_\zeta \\ &\quad + \theta \sum_{\zeta=0}^{1} \Bigg[\sum_{d=1}^{\min(x_1+z,N)} V_{l-1}^\theta(x_1+z-d+\mu_1\zeta) P_{d,1} \\ &\quad - \sum_{d=1}^{\min(x_1,N)} V_{l-1}^\theta(x_1-d+\mu_2\zeta) P_{d,2} \\ &\quad - \sum_{d=1}^{\min(x_2+z,N)} V_{l-1}^\theta(x_2+z-d+\mu_3\zeta) P_{d,3} \\ &\quad + \sum_{d=1}^{\min(x_2,N)} V_{l-1}^\theta(x_2-d+\mu_4\zeta) P_{d,4} \Bigg] p_\zeta, \end{aligned} \quad (31)$$

where $P_{d,1}, P_{d,2}, P_{d,3}$, and $P_{d,4}$ represent the probabilities of d departures when the current states are x_1+z, x_1, x_2+z, and x_2, respectively, with $d \geq 1$, and p_ζ is as in (22). Consider $\mu_1 = \mu_3 = \mu^1$ and $\mu_2 = \mu_4 = \mu^2$. We will consider other possible values of μ_1, μ_2, μ_3, and μ_4 later. Then, (31) can be written as:

$$\begin{aligned} &W^1 - W^2 \\ &= \theta \sum_{\zeta=0}^{1} \Big[P_0 \Big\{ V_{l-1}^\theta\Big(x_1+z+\mu^1\zeta\Big) - V_{l-1}^\theta\Big(x_1+\mu^2\zeta\Big) \\ &\quad - V_{l-1}^\theta\Big(x_2+z+\mu^1\zeta\Big) + V_{l-1}^\theta\Big(x_2+\mu^2\zeta\Big) \Big\} \Big] p_\zeta \\ &\quad + \theta \sum_{\zeta=0}^{1} \Bigg[\sum_{d=1}^{\min(x_1+z,N)} V_{l-1}^\theta\Big(x_1+z-d+\mu^1\zeta\Big) P_{d,1} \\ &\quad - \sum_{d=1}^{\min(x_1,N)} V_{l-1}^\theta\Big(x_1-d+\mu^2\zeta\Big) P_{d,2} \\ &\quad - \sum_{d=1}^{\min(x_2+z,N)} V_{l-1}^\theta\Big(x_2+z-d+\mu^1\zeta\Big) P_{d,3} \\ &\quad + \sum_{d=1}^{\min(x_2,N)} V_{l-1}^\theta\Big(x_2-d+\mu^2\zeta\Big) P_{d,4} \Bigg] p_\zeta. \end{aligned} \quad (32)$$

Consider different cases based on the values of x_1, x_2, z and N and let us analyze (32) for each case.

Case 1: $x_2 \geq N$ *and* $x_1 > x_2$.

This implies that $x_2 + z > N$, $x_1 > N$ and $x_1 + z > N$. Thus, (32) can be written as:

$$\begin{aligned} &W^1 - W^2 \\ &= \theta \sum_{\zeta=0}^{1} \Big[P_0 \Big\{ V_{l-1}^\theta\Big(x_1+z+\mu^1\zeta\Big) - V_{l-1}^\theta\Big(x_1+\mu^2\zeta\Big) \\ &\quad - V_{l-1}^\theta\Big(x_2+z+\mu^1\zeta\Big) + V_{l-1}^\theta\Big(x_2+\mu^2\zeta\Big) \Big\} \Big] p_\zeta \\ &\quad + \theta \sum_{\zeta=0}^{1} \Bigg[\sum_{d=1}^{N} \Big\{ V_{l-1}^\theta\Big(x_1+z-d+\mu^1\zeta\Big) \\ &\quad - V_{l-1}^\theta\Big(x_1-d+\mu^2\zeta\Big) - V_{l-1}^\theta\Big(x_2+z-d+\mu^1\zeta\Big) \\ &\quad + V_{l-1}^\theta\Big(x_2-d+\mu^2\zeta\Big) \Big\} \widetilde{P} \Bigg] p_\zeta, \end{aligned}$$

where $\widetilde{P} = P_{d,1} = P_{d,2} = P_{d,3} = P_{d,4}$ and is equal to $s\binom{N}{d}(h)^d(1-h)^{N-d}$. If $\mu^1 = \mu^2$, then $W^1 - W^2$ becomes:

$$\begin{aligned} &W^1 - W^2 \\ &= \theta \sum_{\zeta=0}^{1} \Big[P_0 \Big\{ V_{l-1}^\theta\Big(x_1+z+\mu^1\zeta\Big) - V_{l-1}^\theta\Big(x_1+\mu^1\zeta\Big) \end{aligned}$$

$$-V_{l-1}^{\theta}\left(x_2+z+\mu^1\zeta\right)+V_{l-1}^{\theta}\left(x_2+\mu^1\zeta\right)\Big\}\Big]p_\zeta$$
$$+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=1}^{N}\Big\{V_{l-1}^{\theta}\left(x_1+z-d+\mu^1\zeta\right)$$
$$-V_{l-1}^{\theta}\left(x_1-d+\mu^1\zeta\right)-V_{l-1}^{\theta}\left(x_2+z-d+\mu^1\zeta\right)$$
$$+V_{l-1}^{\theta}\left(x_2-d+\mu^1\zeta\right)\Big\}\widetilde{P}\Bigg]p_\zeta.$$

Since z is integer-valued and $z>0$, it follows that $z\geq 1$; also, since $\zeta\in\{0,1\}$, and $x_1>x_2$ and using (30), we can conclude that $W^1-W^2\geq 0$. If $\mu^1=1$ and $\mu^2=0$, then W^1-W^2 becomes:

$$W^1-W^2$$
$$=\theta\sum_{\zeta=0}^{1}\big[P_0\big\{V_{l-1}^{\theta}(x_1+z+\zeta)-V_{l-1}^{\theta}(x_2+z+\zeta)$$
$$-V_{l-1}^{\theta}(x_1)+V_{l-1}^{\theta}(x_2)\big\}\big]p_\zeta$$
$$+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=1}^{N}\big\{V_{l-1}^{\theta}(x_1+z-d+\zeta)-V_{l-1}^{\theta}(x_1-d)$$
$$-V_{l-1}^{\theta}(x_2+z-d+\zeta)+V_{l-1}^{\theta}(x_2-d)\big\}\widetilde{P}\Bigg]p_\zeta.$$

Similar to the previous case, we can prove that $W^1-W^2\geq 0$. If $\mu^1=0$ and $\mu^2=1$, then W^1-W^2 becomes:

$$W^1-W^2$$
$$=\theta\sum_{\zeta=0}^{1}\big[P_0\big\{V_{l-1}^{\theta}(x_1+z)-V_{l-1}^{\theta}(x_1+\zeta)-V_{l-1}^{\theta}(x_2$$
$$+z)+V_{l-1}^{\theta}(x_2+\zeta)\big\}\big]p_\zeta+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=1}^{N}\big\{V_{l-1}^{\theta}(x_1+z-d)$$
$$-V_{l-1}^{\theta}(x_1-d+\zeta)-V_{l-1}^{\theta}(x_2+z-d)$$
$$+V_{l-1}^{\theta}(x_2-d+\zeta)\Big\}\widetilde{P}\Bigg]p_\zeta.$$

Similar to the previous case, it follows that $W^1-W^2\geq 0$.

Case 2: $x_2<N$, $x_2+z\geq N$, *and* $x_1\geq N$.

In this case:

$$W^1-W^2$$
$$=\theta\sum_{\zeta=0}^{1}\Big[P_0\Big\{V_{l-1}^{\theta}\left(x_1+z+\mu^1\zeta\right)-V_{l-1}^{\theta}\left(x_1+\mu^2\zeta\right)$$
$$-V_{l-1}^{\theta}\left(x_2+z+\mu^1\zeta\right)+V_{l-1}^{\theta}\left(x_2+\mu^2\zeta\right)\Big\}\Big]p_\zeta$$
$$+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=1}^{N}V_{l-1}^{\theta}\left(x_1+z-d+\mu^1\zeta\right)P_{d,1}$$
$$-\sum_{d=1}^{N}V_{l-1}^{\theta}\left(x_1-d+\mu^2\zeta\right)P_{d,2}$$
$$-\sum_{d=1}^{N}V_{l-1}^{\theta}(x_2+z-d+\mu^1\zeta)$$
$$P_{d,3}+\sum_{d=1}^{x_2}V_{l-1}^{\theta}\left(x_2-d+\mu^2\zeta\right)P_{d,4}\Bigg]p_\zeta. \tag{33}$$

As $z\geq 1$, $x_1>x_2$, μ^1, μ^2, ζ belong to $\{0,1\}$, and from (30), the first term of the RHS of (39) is always positive. The second term of the RHS of (39) can be written as:

$$\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=1}^{x_2-1}\Big\{V_{l-1}^{\theta}\left(x_1+z-d+\mu^1\zeta\right)-V_{l-1}^{\theta}(x_1-d+\mu^2\zeta)$$
$$-V_{l-1}^{\theta}\left(x_2+z-d+\mu^1\zeta\right)+V_{l-1}^{\theta}(x_2-d+\mu^2\zeta)\Big\}\widetilde{P}\Bigg]p_\zeta$$
$$+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=x_2}^{N}\Big\{V_{l-1}^{\theta}\left(x_1+z-d+\mu^1\zeta\right)$$
$$-V_{l-1}^{\theta}\left(x_1-d+\mu^2\zeta\right)-V_{l-1}^{\theta}\left(x_2+z-d+\mu^1\zeta\right)$$
$$+V_{l-1}^{\theta}\left(\mu^2\zeta\right)\Big\}\widetilde{P}\Bigg]p_\zeta, \tag{34}$$

where $\widetilde{P}$ is the same as stated earlier. Due to reasons similar to those stated after (39), the first term in (34) is non-negative. We can re-write the second term of (34) as:

$$\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=x_2}^{N}\Big\{\Big(V_{l-1}^{\theta}\left(x_1+z-d+\mu^1\zeta\right)$$
$$-V_{l-1}^{\theta}\left(x_1-d+\mu^2\zeta\right)-V_{l-1}^{\theta}\left(z+\mu^1\zeta\right)+V_{l-1}^{\theta}\left(\mu^2\zeta\right)\Big)$$
$$+\Big(V_{l-1}^{\theta}\left(z+\mu^1\zeta\right)-V_{l-1}^{\theta}\left(x_2+z-d+\mu^1\zeta\right)\Big)\Big\}\widetilde{P}\Bigg]p_\zeta. \tag{35}$$

Let $X_1+Z=x_1+z-d+\mu^1\zeta$ and $X_1=x_1-d+\mu^2\zeta$. This implies $Z=z+(\mu^1-\mu^2)\zeta$; since $z\geq 1$ and $\min(\mu^1-\mu^2)\zeta=-1$, it follows that $Z\geq 0$. Similarly, $X_2+Z=z+\mu^1\zeta$ and $X_2=\mu^2\zeta$. Therefore, by using (35), (34), (39), (30) and Lemma 2, we can conclude that $W^1-W^2\geq 0$. Similar to Case 1, we can show that $W^1-W^2\geq 0$ $\forall\mu^1,\mu^2$.

Case 3: $x_2<N$, $x_2+z\geq N$ *and* $x_1<N$.

This implies $x_1+z>N$. Thus, (32) can be written as:

$$W^1-W^2$$
$$=\theta\sum_{\zeta=0}^{1}\Big[P_0\Big\{V_{l-1}^{\theta}\left(x_1+z+\mu^1\zeta\right)-V_{l-1}^{\theta}\left(x_1+\mu^2\zeta\right)$$
$$-V_{l-1}^{\theta}\left(x_2+z+\mu^1\zeta\right)+V_{l-1}^{\theta}\left(x_2+\mu^2\zeta\right)\Big\}\Big]p_\zeta$$
$$+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=1}^{N}V_{l-1}^{\theta}\left(x_1+z-d+\mu^1\zeta\right)P_{d,1}$$
$$-\sum_{d=1}^{x_1}V_{l-1}^{\theta}\left(x_1-d+\mu^2\zeta\right)P_{d,2}-\sum_{d=1}^{N}V_{l-1}^{\theta}(x_2+z$$

$$
\begin{aligned}
&- d+\mu^1\zeta)P_{d,3}+\sum_{d=1}^{x_2} V_{l-1}^{\theta}\Big(x_2-d+\mu^2\zeta\Big)P_{d,4}\Bigg]p_\zeta\\
&=\theta\sum_{\zeta=0}^{1}\Bigg[P_0\Big\{V_{l-1}^{\theta}\Big(x_1+z+\mu^1\zeta\Big)-V_{l-1}^{\theta}\Big(x_1+\mu^2\zeta\Big)\\
&-V_{l-1}^{\theta}\Big(x_2+z+\mu^1\zeta\Big)+V_{l-1}^{\theta}\Big(x_2+\mu^2\zeta\Big)\Big\}\Bigg]p_\zeta\\
&+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=1}^{x_2-1}\Big\{V_{l-1}^{\theta}\Big(x_1+z-d+\mu^1\zeta\Big)-V_{l-1}^{\theta}(x_1\\
&-d+\mu^2\zeta)-V_{l-1}^{\theta}\Big(x_2+z-d+\mu^1\zeta\Big)+V_{l-1}^{\theta}(x_2\\
&-d+\mu^2\zeta)\Big\}\widetilde{P}\Bigg]p_\zeta+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=x_2}^{N} V_{l-1}^{\theta}(x_1+z-d\\
&+\mu^1\zeta)P_{d,1}-\sum_{d=x_2}^{x_1} V_{l-1}^{\theta}\Big(x_1-d+\mu^2\zeta\Big)P_{d,2}\\
&-\sum_{d=x_2}^{N} V_{l-1}^{\theta}\Big(x_2+z-d+\mu^1\zeta\Big)P_{d,3}\\
&+\sum_{d=x_2}^{N} V_{l-1}^{\theta}\Big(\mu^2\zeta\Big)\widetilde{P}\Bigg]p_\zeta\\
&=\theta\sum_{\zeta=0}^{1}\Bigg[P_0\Big\{V_{l-1}^{\theta}\Big(x_1+z+\mu^1\zeta\Big)-V_{l-1}^{\theta}\Big(x_1+\mu^2\zeta\Big)\\
&-V_{l-1}^{\theta}\Big(x_2+z+\mu^1\zeta\Big)+V_{l-1}^{\theta}\Big(x_2+\mu^2\zeta\Big)\Big\}\Bigg]p_\zeta\\
&+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=1}^{x_2-1}\Big\{V_{l-1}^{\theta}\Big(x_1+z-d+\mu^1\zeta\Big)\\
&-V_{l-1}^{\theta}\Big(x_1-d+\mu^2\zeta\Big)-V_{l-1}^{\theta}\Big(x_2+z-d+\mu^1\zeta\Big)\\
&+V_{l-1}^{\theta}\Big(x_2-d+\mu^2\zeta\Big)\Big\}\widetilde{P}\Bigg]p_\zeta\\
&+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=x_2}^{x_1-1}\Big\{\Big(V_{l-1}^{\theta}\Big(x_1+z-d+\mu^1\zeta\Big)-V_{l-1}^{\theta}(x_1\\
&-d+\mu^2\zeta)-V_{l-1}^{\theta}\Big(z+\mu^1\zeta\Big)+V_{l-1}^{\theta}\Big(\mu^2\zeta\Big)\Big)\\
&+\Big(V_{l-1}^{\theta}\Big(z+\mu^1\zeta\Big)-V_{l-1}^{\theta}\Big(x_2+z-d+\mu^1\zeta\Big)\Big)\Big\}\widetilde{P}\Bigg]p_\zeta\\
&+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=x_1}^{N}\Big\{V_{l-1}^{\theta}\Big(x_1+z-d+\mu^1\zeta\Big)\\
&-V_{l-1}^{\theta}\Big(x_2+z-d+\mu^1\zeta\Big)\Big\}\widetilde{P}\Bigg]p_\zeta. \qquad (36)
\end{aligned}
$$

In (36), $z\geq 1$ and $x_1>x_2$; so from (30), we can conclude that the first and second terms of (36) are non-negative. In the third term of (36), let $X_1+Z=x_1+z-d+\mu^1\zeta$ and $X_1=x_1-d+\mu^2\zeta$. This implies $Z=z+(\mu^1-\mu^2)\zeta$; since $z\geq 1$ and $\min(\mu^1-\mu^2)\zeta=-1$, it follows that $Z\geq 0$. Similarly, $X_2+Z=z+\mu^1\zeta$ and $X_2=\mu^2\zeta$. Thus, using (30) and Lemma 2, we can conclude that this term is also non-negative. From Lemma 2, we can conclude that the fourth term of (36) is also non-negative. Hence, it follows that $W^1-W^2\geq 0$. Similar to Case 1, we can show that $W^1-W^2\geq 0$ $\forall\mu^1,\mu^2$.

Case 4: $x_2<N, x_2+z<N$ *and* $x_1\geq N$.
In this case, (32) becomes:

$$
\begin{aligned}
&W^1-W^2\\
&=\theta\sum_{\zeta=0}^{1}P_0\Bigg[\Big\{V_{l-1}^{\theta}\Big(x_1+z+\mu^1\zeta\Big)-V_{l-1}^{\theta}\Big(x_1+\mu^2\zeta\Big)\\
&-V_{l-1}^{\theta}\Big(x_2+z+\mu^1\zeta\Big)+V_{l-1}^{\theta}\Big(x_2+\mu^2\zeta\Big)\Big\}\Bigg]p_\zeta\\
&+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=1}^{N} V_{l-1}^{\theta}\Big(x_1+z-d+\mu^1\zeta\Big)P_{d,1}\\
&-\sum_{d=1}^{N} V_{l-1}^{\theta}\Big(x_1-d+\mu^2\zeta\Big)P_{d,2}-\sum_{d=1}^{x_2+z} V_{l-1}^{\theta}(x_2+z\\
&-d+\mu^1\zeta)P_{d,3}+\sum_{d=1}^{x_2} V_{l-1}^{\theta}\Big(x_2-d+\mu^2\zeta\Big)P_{d,4}\Bigg]p_\zeta\\
&=\theta\sum_{\zeta=0}^{1}P_0\Bigg[\Big\{V_{l-1}^{\theta}\Big(x_1+z+\mu^1\zeta\Big)-V_{l-1}^{\theta}\Big(x_1+\mu^2\zeta\Big)\\
&-V_{l-1}^{\theta}\Big(x_2+z+\mu^1\zeta\Big)+V_{l-1}^{\theta}\Big(x_2+\mu^2\zeta\Big)\Big\}\Bigg]p_\zeta\\
&+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=1}^{x_2-1}\Big\{V_{l-1}^{\theta}\Big(x_1+z-d+\mu^1\zeta\Big)\\
&-V_{l-1}^{\theta}\Big(x_1-d+\mu^2\zeta\Big)-V_{l-1}^{\theta}\Big(x_2+z-d+\mu^1\zeta\Big)\\
&+V_{l-1}^{\theta}\Big(x_2-d+\mu^2\zeta\Big)\Big\}\widetilde{P}\Bigg]p_\zeta\\
&+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=x_2}^{x_2+z-1}\Big\{V_{l-1}^{\theta}\Big(x_1+z-d+\mu^1\zeta\Big)\\
&-V_{l-1}^{\theta}\Big(x_1-d+\mu^2\zeta\Big)+V_{l-1}^{\theta}\Big(\mu^2\zeta\Big)-V_{l-1}^{\theta}\Big(z+\mu^1\zeta\Big)\\
&+\Big(V_{l-1}^{\theta}\Big(z+\mu^1\zeta\Big)-V_{l-1}^{\theta}\Big(x_2+z-d+\mu^1\zeta\Big)\Big)\Big\}\widetilde{P}\Bigg]p_\zeta\\
&+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=x_2+z}^{N}\Big\{V_{l-1}^{\theta}\Big(x_1+z-d+\mu^1\zeta\Big)\\
&-V_{l-1}^{\theta}\Big(x_1-d+\mu^2\zeta\Big)-V_{l-1}^{\theta}\Big(z+\mu^1\zeta\Big)+V_{l-1}^{\theta}\Big(\mu^2\zeta\Big)\\
&+\Big(V_{l-1}^{\theta}\Big(z+\mu^1\zeta\Big)-V_{l-1}^{\theta}\Big(\mu^1\zeta\Big)\Big)\Big\}\widetilde{P}\Bigg]p_\zeta. \qquad (37)
\end{aligned}
$$

As in Case 1 and Case 3, we can prove the non-negativity of W^1-W^2 given in (38) $\forall\mu^1,\mu^2$.

Case 5: $x_2<N, x_2+z<N, x_1<N$ *and* $x_1+z\geq N$.

In this case, (32) becomes:

$$\begin{aligned}
&W^1 - W^2 \\
&= \theta \sum_{\zeta=0}^{1} P_0 \Big[\Big\{ V_{l-1}^{\theta}\Big(x_1 + z + \mu^1\zeta\Big) - V_{l-1}^{\theta}\Big(x_1 + \mu^2\zeta\Big) \\
&\quad - V_{l-1}^{\theta}\Big(x_2 + z + \mu^1\zeta\Big) + V_{l-1}^{\theta}\Big(x_2 + \mu^2\zeta\Big) \Big\} \Big] p_\zeta \\
&\quad + \theta \sum_{\zeta=0}^{1} \Bigg[\sum_{d=1}^{N} V_{l-1}^{\theta}\Big(x_1 + z - d + \mu^1\zeta\Big) P_{d,1} \\
&\quad - \sum_{d=1}^{x_1} V_{l-1}^{\theta}\Big(x_1 - d + \mu^2\zeta\Big) P_{d,2} - \sum_{d=1}^{x_2+z} V_{l-1}^{\theta}(x_2 + z \\
&\quad - d + \mu^1\zeta) P_{d,3} + \sum_{d=1}^{x_2} V_{l-1}^{\theta}\Big(x_2 - d + \mu^2\zeta\Big) P_{d,4} \Bigg] p_\zeta \\
&= \theta \sum_{\zeta=0}^{1} P_0 \Big[\Big\{ V_{l-1}^{\theta}\Big(x_1 + z + \mu^1\zeta\Big) - V_{l-1}^{\theta}\Big(x_1 + \mu^2\zeta\Big) \\
&\quad - V_{l-1}^{\theta}\Big(x_2 + z + \mu^1\zeta\Big) + V_{l-1}^{\theta}\Big(x_2 + \mu^2\zeta\Big) \Big\} \Big] p_\zeta \\
&\quad + \theta \sum_{\zeta=0}^{1} \Bigg[\sum_{d=1}^{x_2-1} \Big\{ V_{l-1}^{\theta}\Big(x_1 + z - d + \mu^1\zeta\Big) \\
&\quad - V_{l-1}^{\theta}\Big(x_1 - d + \mu^2\zeta\Big) - V_{l-1}^{\theta}\Big(x_2 + z - d + \mu^1\zeta\Big) \\
&\quad + V_{l-1}^{\theta}\Big(x_2 - d + \mu^2\zeta\Big) \Big\} \widetilde{P} \Bigg] p_\zeta + \theta \sum_{\zeta=0}^{1} \Bigg[\sum_{d=x_2}^{N} V_{l-1}^{\theta}(x_1 \\
&\quad + z - d + \mu^1\theta) P_{d,1} - \sum_{d=x_2}^{x_1} V_{l-1}^{\theta}\Big(x_1 - d + \mu^2\zeta\Big) P_{d,2} \\
&\quad - \sum_{d=x_2}^{x_2+z} V_{l-1}^{\theta}\Big(x_2 + z - d + \mu^1\zeta\Big) P_{d,3} \\
&\quad + \sum_{d=x_2}^{N_i} V_{l-1}^{\theta}\Big(\mu^2\zeta\Big) \widetilde{P} \Bigg] p_\zeta .
\end{aligned}$$

If $x_1 \geq x_2 + z$, then we get:

$$\begin{aligned}
&W^1 - W^2 \\
&= \theta \sum_{\zeta=0}^{1} P_0 \Big[\Big\{ V_{l-1}^{\theta}\Big(x_1 + z + \mu^1\zeta\Big) - V_{l-1}^{\theta}\Big(x_1 + \mu^2\zeta\Big) \\
&\quad - V_{l-1}^{\theta}\Big(x_2 + z + \mu^1\zeta\Big) + V_{l-1}^{\theta}\Big(x_2 + \mu^2\zeta\Big) \Big\} \Big] p_\zeta \\
&\quad + \theta \sum_{\zeta=0}^{1} \Bigg[\sum_{d=1}^{x_2-1} \Big\{ V_{l-1}^{\theta}\Big(x_1 + z - d + \mu^1\zeta\Big) \\
&\quad - V_{l-1}^{\theta}\Big(x_1 - d + \mu^2\zeta\Big) - V_{l-1}^{\theta}\Big(x_2 + z - d + \mu^1\zeta\Big) \\
&\quad + V_{l-1}^{\theta}\Big(x_2 - d + \mu^2\zeta\Big) \Big\} \widetilde{P} \Bigg] p_\zeta \\
&\quad + \theta \sum_{\zeta=0}^{1} \Bigg[\sum_{d=x_2}^{x_2+z-1} \Big\{ V_{l-1}^{\theta}\Big(x_1 + z - d + \mu^1\zeta\Big) - V_{l-1}^{\theta}(x_1 \\
&\quad - d + \mu^2\zeta) + V_{l-1}^{\theta}\Big(\mu^2\zeta\Big) - V_{l-1}^{\theta}\Big(z + \mu^1\zeta\Big) \\
&\quad + \Big(V_{l-1}^{\theta}\Big(z + \mu^1\zeta\Big) - V_{l-1}^{\theta}\Big(x_2 + z - d + \mu^1\zeta\Big) \Big) \Big\} \widetilde{P} \Bigg] p_\zeta \\
&\quad + \theta \sum_{\zeta=0}^{1} \Bigg[\sum_{d=x_2+z}^{x_1-1} \Big\{ V_{l-1}^{\theta}\Big(x_1 + z - d + \mu^1\zeta\Big) \\
&\quad - V_{l-1}^{\theta}\Big(x_1 - d + \mu^2\zeta\Big) - V_{l-1}^{\theta}\Big(z + \mu^1\zeta\Big) + V_{l-1}^{\theta}\Big(\mu^2\zeta\Big) \\
&\quad + \Big(V_{l-1}^{\theta}\Big(z + \mu^1\zeta\Big) - V_{l-1}^{\theta}\Big(\mu^1\zeta\Big) \Big) \Big\} \widetilde{P} \Bigg] p_\zeta \\
&\quad + \theta \sum_{\zeta=0}^{1} \sum_{d=x_1}^{N} \Big\{ V_{l-1}^{\theta}\Big(x_1 + z - d + \mu^1\zeta\Big) - V_{l-1}^{\theta}\Big(\mu^1\zeta\Big) \Big\} \widetilde{P} p_\zeta .
\end{aligned} \tag{38}$$

As in Case 1 and Case 3, we can prove the non-negativity of $W^1 - W^2$ given in (37) $\forall \mu^1, \mu^2$.

If $x_1 < x_2 + z$, then we get:

$$\begin{aligned}
&W^1 - W^2 \\
&= \theta \sum_{\zeta=0}^{1} P_0 \Big[\Big\{ V_{l-1}^{\theta}\Big(x_1 + z + \mu^1\zeta\Big) - V_{l-1}^{\theta}\Big(x_1 + \mu^2\zeta\Big) \\
&\quad - V_{l-1}^{\theta}\Big(x_2 + z + \mu^1\zeta\Big) + V_{l-1}^{\theta}\Big(x_2 + \mu^2\zeta\Big) \Big\} \Big] p_\zeta \\
&\quad + \theta \sum_{\zeta=0}^{1} \Bigg[\sum_{d=1}^{x_2-1} \Big\{ V_{l-1}^{\theta}\Big(x_1 + z - d + \mu^1\zeta\Big) - V_{l-1}^{\theta}(x_1 \\
&\quad - d + \mu^2\zeta) - V_{l-1}^{\theta}\Big(x_2 + z - d + \mu^1\zeta\Big) + V_{l-1}^{\theta}(x_2 \\
&\quad - d + \mu^2\zeta) \Big\} \widetilde{P} \Bigg] p_\zeta + \theta \sum_{\zeta=0}^{1} \Bigg[\sum_{d=x_2}^{x_1-1} \{ V_{l-1}^{\theta}(x_1 + z \\
&\quad - d + \mu^1\zeta) - V_{l-1}^{\theta}\Big(x_1 - d + \mu^2\zeta\Big) + V_{l-1}^{\theta}\Big(\mu^2\zeta\Big) \\
&\quad - V_{l-1}^{\theta}\Big(z + \mu^1\zeta\Big) + \Big(V_{l-1}^{\theta}\Big(z + \mu^1\zeta\Big) - V_{l-1}^{\theta}(x_2 \\
&\quad + z - d + \mu^1\zeta) \Big) \Big\} \widetilde{P} \Bigg] p_\zeta + \theta \sum_{\zeta=0}^{1} \Bigg[\sum_{d=x_1}^{x_2+z-1} \{ V_{l-1}^{\theta}(x_1 \\
&\quad + z - d + \mu^1\zeta) - V_{l-1}^{\theta}\Big(x_2 + z - d + \mu^1\zeta\Big) \Big\} \widetilde{P} \Big] p_\zeta \\
&\quad + \theta \sum_{\zeta=0}^{1} \Bigg[\sum_{d=x_2+z}^{N} \Big\{ V_{l-1}^{\theta}\Big(x_1 + z - d + \mu^1\zeta\Big) \\
&\quad - V_{l-1}^{\theta}\Big(\mu^1\ \zeta\Big) \Big\} \widetilde{P} \Bigg] p_\zeta .
\end{aligned} \tag{39}$$

As in Case 1 and Case 3, we can prove the non-negativity of $W^1 - W^2$ given in (39) $\forall \mu^1, \mu^2$.

Case 6: $x_1 + z < N$.

This implies $x_2 < N$, $x_2 + z < N$, and $x_1 < N$. Then (32) becomes:

$$\begin{aligned}
&W^1 - W^2 \\
&= \theta \sum_{\zeta=0}^{1} P_0 \Big[\Big\{ V_{l-1}^{\theta}\Big(x_1 + z + \mu^1\zeta\Big) - V_{l-1}^{\theta}\Big(x_1 + \mu^2\zeta\Big) \Big\}
\end{aligned}$$

$$
\begin{aligned}
&-V_{l-1}^{\theta}\left(x_2+z+\mu^1\zeta\right)+V_{l-1}^{\theta}\left(x_2+\mu^2\zeta\right)\Big\}\Big]p_\zeta\\
&+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=1}^{x_1+z}V_{l-1}^{\theta}\left(x_1+z-d+\mu^1\zeta\right)P_{d,1}\\
&-\sum_{d=1}^{x_1}V_{l-1}^{\theta}\left(x_1-d+\mu^2\zeta\right)P_{d,2}-\sum_{d=1}^{x_2+z}V_{l-1}^{\theta}(x_2+z\\
&-d+\mu^1\zeta)P_{d,3}+\sum_{d=1}^{x_2}V_{l-1}^{\theta}\left(x_2-d+\mu^2\zeta\right)P_{d,4}\Bigg]p_\zeta\\
=\theta&\sum_{\zeta=0}^{1}P_0\Big[\Big\{V_{l-1}^{\theta}\left(x_1+z+\mu^1\zeta\right)-V_{l-1}^{\theta}\left(x_1+\mu^2\zeta\right)\\
&-V_{l-1}^{\theta}\left(x_2+z+\mu^1\zeta\right)+V_{l-1}^{\theta}\left(x_2+\mu^2\zeta\right)\Big\}\Big]p_\zeta\\
&+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=1}^{x_2-1}\Big\{V_{l-1}^{\theta}\left(x_1+z-d+\mu^1\zeta\right)-V_{l-1}^{\theta}(x_1\\
&-d+\mu^2\zeta)-V_{l-1}^{\theta}\left(x_2+z-d+\mu^1\zeta\right)+V_{l-1}^{\theta}(x_2\\
&-d+\mu^2\zeta)\Big\}\widetilde{P}\Bigg]p_\zeta+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=x_2}^{x_1+z}V_{l-1}^{\theta}(x_1+z-d\\
&+\mu^1\zeta)P_{d,1}-\sum_{d=x_2}^{x_1}V_{l-1}^{\theta}\left(x_1-d+\mu^2\zeta\right)P_{d,2}\\
&-\sum_{d=x_2}^{x_2+z}V_{l-1}^{\theta}\left(x_2+z-d+\mu^1\zeta\right)P_{d,3}\\
&+\sum_{d=x_2}^{N}V_{l-1}^{\theta}\left(\mu^2\zeta\right)\widetilde{P}\Bigg]p_\zeta.
\end{aligned}
$$

If $x_1 < x_2+z$, then we get:

$$
\begin{aligned}
&W^1-W^2\\
&=\theta\sum_{\zeta=0}^{1}P_0\Big[\Big\{V_{l-1}^{\theta}\left(x_1+z+\mu^1\zeta\right)-V_{l-1}^{\theta}\left(x_1+\mu^2\zeta\right)\\
&-V_{l-1}^{\theta}\left(x_2+z+\mu^1\zeta\right)+V_{l-1}^{\theta}\left(x_2+\mu^2\zeta\right)\Big\}\Big]p_\zeta\\
&+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=1}^{x_2-1}\Big\{V_{l-1}^{\theta}\left(x_1+z-d+\mu^1\zeta\right)-V_{l-1}^{\theta}(x_1-d\\
&+\mu^2\zeta)-V_{l-1}^{\theta}\left(x_2+z-d+\mu^1\zeta\right)+V_{l-1}^{\theta}(x_2-d\\
&+\mu^2\zeta)\Big\}\widetilde{P}\Bigg]p_\zeta+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=x_2}^{x_1-1}\Big\{V_{l-1}^{\theta}\left(x_1+z-d+\mu^1\zeta\right)\\
&-V_{l-1}^{\theta}\left(x_1-d+\mu^2\zeta\right)+V_{l-1}^{\theta}\left(\mu^2\zeta\right)-V_{l-1}^{\theta}\left(z+\mu^1\zeta\right)\\
&+\Big(V_{l-1}^{\theta}\left(z+\mu^1\zeta\right)-V_{l-1}^{\theta}\left(x_2+z-d+\mu^1\zeta\right)\Big)\Big\}\widetilde{P}\Bigg]p_\zeta\\
&+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=x_1}^{x_2+z-1}\Big\{V_{l-1}^{\theta}\left(x_1+z-d+\mu^1\zeta\right)-V_{l-1}^{\theta}(x_2\\
&+z-d+\mu^1\zeta)\Big\}\widetilde{P}\Bigg]p_\zeta+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=x_2+z}^{x_1+z-1}\{V_{l-1}^{\theta}(x_1\\
&+z-d+\mu^1\zeta)-V_{l-1}^{\theta}\left(\mu^1\zeta\right)\Big\}\widetilde{P}\Bigg]p_\zeta. \qquad (40)
\end{aligned}
$$

As in Case 1 and Case 3, we can prove the non-negativity of W^1-W^2 given in (40) $\forall\mu^1,\mu^2$.

If $x_1 \geq x_2+z$, then we get:

$$
\begin{aligned}
&W^1-W^2\\
&=\theta\sum_{\zeta=0}^{1}P_0\Big[\Big\{V_{l-1}^{\theta}\left(x_1+z+\mu^1\zeta\right)-V_{l-1}^{\theta}\left(x_1+\mu^2\zeta\right)\\
&-V_{l-1}^{\theta}\left(x_2+z+\mu^1\zeta\right)+V_{l-1}^{\theta}\left(x_2+\mu^2\zeta\right)\Big\}\Big]p_\zeta\\
&+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=1}^{x_2-1}\Big\{V_{l-1}^{\theta}\left(x_1+z-d+\mu^1\zeta\right)-V_{l-1}^{\theta}(x_1\\
&-d+\mu^2\zeta)-V_{l-1}^{\theta}\left(x_2+z-d+\mu^1\zeta\right)+V_{l-1}^{\theta}(x_2\\
&-d+\mu^2\zeta)\Big\}\widetilde{P}\Bigg]p_\zeta+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=x_2}^{x_2+z-1}\{V_{l-1}^{\theta}(x_1+z-d\\
&+\mu^1\zeta)-V_{l-1}^{\theta}\left(x_1-d+\mu^2\zeta\right)-V_{l-1}^{\theta}\left(z+\mu^1\zeta\right)\\
&+V_{l-1}^{\theta}\left(\mu^2\zeta\right)+\Big(V_{l-1}^{\theta}\left(z+\mu^1\zeta\right)-V_{l-1}^{\theta}(x_2+z-d\\
&+\mu^1\zeta)\Big)\Big\}\widetilde{P}\Bigg]p_\zeta+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=x_2+z}^{x_1-1}\{V_{l-1}^{\theta}(x_1+z-d\\
&+\mu^1\zeta)-V_{l-1}^{\theta}\left(x_1-d+\mu^2\zeta\right)-V_{l-1}^{\theta}(z+\mu^1\zeta)\\
&+V_{l-1}^{\theta}\left(\mu^2\zeta\right)+\Big(V_{l-1}^{\theta}\left(z+\mu^1\zeta\right)-V_{l-1}^{\theta}\left(\mu^1\zeta\right)\Big)\Big\}\widetilde{P}\Bigg]p_\zeta\\
&+\theta\sum_{\zeta=0}^{1}\Bigg[\sum_{d=x_1}^{x_1+z-1}\Big\{V_{l-1}^{\theta}\left(x_1+z-d+\mu^1\zeta\right)\\
&-V_{l-1}^{\theta}\left(\mu^1\zeta\right)\Big\}\widetilde{P}\Bigg]p_\zeta. \qquad (41)
\end{aligned}
$$

As in Case 1 and Case 3, we can prove the non-negativity of W^1-W^2 given in (41) $\forall\mu^1,\mu^2$. For all possible cases under the assumption that $\mu_1=\mu_3$ and $\mu_2=\mu_4$ (where μ_1,μ_2,μ_3 and μ_4 are the admission controls at which the minimum value of $V_l^{\theta}(\cdot)$ is obtained for the points x_1+z, x_1, x_2+z and x_2, respectively), we have proved that (31) is non-negative. Specifically, $.W^1|_{\mu_1=a,\mu_2=b} \geq .W^2|_{\mu_3=a,\mu_4=b}$, where $a,b\in\{0,1\}$. Now suppose $a=1$ and $b=0$; then we get:

$$
W^1\Big|_{\mu_1=1,\mu_2=0} \geq W^2\Big|_{\mu_3=1,\mu_4=0}.
$$

If the minimum values of $V_l^{\theta}(\cdot)$ at points x_1+z and x_1 are attained at $\mu_1=1$ and $\mu_2=0$, respectively, then

the minimum of W^1 is achieved when $V_l^\theta(x_1+z)$ is at its minimum and $V_l^\theta(x_1)$ is at its maximum, i.e., at $\mu_1=1$ and $\mu_2=1$. Similarly, the maximum of W^1 is attained at $\mu_1=0$ and $\mu_2=0$. Thus, we can write:

$$W^1\Big|_{\mu_1=0,\mu_2=0} \geq W^1\Big|_{\mu_1=1,\mu_2=0}, W^1\Big|_{\mu_1=0,\mu_2=1} \geq W^1\Big|_{\mu_1=1,\mu_2=1}. \quad (42)$$

If the minimum values of $V_l^\theta(\cdot)$ at points x_2+z and x_2 are attained at $\mu_3=0$ and $\mu_4=1$, respectively, then the minimum of W^2 is achieved when $V_l^\theta(x_2+z)$ is at its minimum and $V_l^\theta(x_2)$ is at its maximum, i.e., at $\mu_3=0$ and $\mu_4=0$. Similarly, the maximum of W^2 is attained at $\mu_3=1$ and $\mu_4=1$. Thus, we can write:

$$W^2\Big|_{\mu_3=1,\mu_4=1} \geq W^2\Big|_{\mu_3=1,\mu_4=0}, W^2\Big|_{\mu_3=0,\mu_4=1} \geq W^2\Big|_{\mu_3=0,\mu_4=0}. \quad (43)$$

However, we know that $.W^1|_{\mu_1=1,\mu_2=1} \geq .W^2|_{\mu_3=1,\mu_4=1}$. So from (42) and (43), we can conclude that $.W^1|_{\mu_1=1,\mu_2=0} \geq .W^2|_{\mu_3=0,\mu_4=1}$. Similarly, we can show increasing differences for all possible values of μ_1, μ_2, μ_3 and μ_4. This completes the proof of (29).

Next, taking limits as $y \uparrow \infty$ in (29), the infinite horizon value function, $V^\theta(\cdot)$, also satisfies the increasing differences property; again, by taking limits as $\theta \uparrow 1$, we can conclude that the value function $V(\cdot)$ has increasing differences. The result follows.

REFERENCES

[1] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: Opportunities and challenges," *Wireless Netw.*, vol. 21, pp. 2657–2676, Nov. 2015.

[2] S. Sharma and B. Singh, "5G networks: The next gen evolution," in *Proc. Int. Conf. Signal Process. Commun. (ICSC)*, 2016, pp. 55–60.

[3] T. S. Rappaport et al., "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[4] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[5] M. Xiao et al., "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.

[6] S. Kutty and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 949–973, 2nd Quart., 2015.

[7] D. Liu et al., "User association in 5G networks: A survey and an outlook," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1018–1044, 2nd Quart., 2016.

[8] "Technical specification group radio access network; study on new radio (NR) to support non-terrestrial networks," 3GPP, Sophia Antipolis, France, Rep. TR 36.872, Dec. 2021.

[9] H. Ramazanali, A. Mesodiakaki, A. Vinel, and C. Verikoukis, "Survey of user association in 5G HetNets," in *Proc. 8th IEEE Latin-Am. Conf. Commun. (LATINCOM)*, 2016, pp. 1–6.

[10] D. Liu, L. Wang, Y. Chen, T. Zhang, K. K. Chai, and M. Elkashlan, "Distributed energy efficient fair user association in massive MIMO enabled HetNets," *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1770–1773, Oct. 2015.

[11] T. Van Chien, E. Björnson, and E. G. Larsson, "Joint power allocation and user association optimization for massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6384–6399, Sep. 2016.

[12] Y. Hao, Q. Ni, H. Li, and S. Hou, "Energy and spectral efficiency tradeoff with user association and power coordination in massive MIMO enabled HetNets," *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2091–2094, Oct. 2016.

[13] T. Zhou, Z. Liu, D. Qin, N. Jiang, and C. Li, "User association with maximizing weighted sum energy efficiency for massive MIMO-enabled heterogeneous cellular networks," *IEEE Commun. Lett.*, vol. 21, no. 10, pp. 2250–2253, Oct. 2017.

[14] N. Ghiasi, S. Mashhadi, S. Farahmand, S. M. Razavizadeh, and I. Lee, "Energy efficient AP selection for cell-free massive MIMO systems: Deep reinforcement learning approach," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 1, pp. 29–41, Mar. 2023.

[15] S. Aboagye, A. Ibrahim, T. M. Ngatched, and O. A. Dobre, "Joint user association and power control for area spectral efficiency maximization in HetNets," in *Proc. IEEE 92nd Veh. Technol. Conf.*, 2020, pp. 1–6.

[16] P. Han, Z. Zhou, and Z. Wang, "User association for load balance in heterogeneous networks with limited CSI feedback," *IEEE Commun. Lett.*, vol. 24, no. 5, pp. 1095–1099, 2020.

[17] A. Alizadeh and M. Vu, "Multi-armed bandit load balancing user association in 5G cellular HetNets," in *Proc. IEEE Glob. Commun. Conf.*, 2020, pp. 1–6.

[18] R. Jiang, Q. Wang, H. Haas, and Z. Wang, "Joint user association and power allocation for cell-free visible light communication networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 1, pp. 136–148, Jan. 2018.

[19] S. Sobhi-Givi, M. G. Shayesteh, H. Kalbkhani, and N. Rajatheva, "Resource allocation and user association for load balancing in NOMA-based cellular heterogeneous networks," in *Proc. Iran Workshop Commun. Inf. Theory (IWCIT)*, 2020, pp. 1–6.

[20] Z. Su et al., "User association and wireless backhaul bandwidth allocation for 5G heterogeneous networks in the millimeter-wave band," *China Commun.*, vol. 15, no. 4, pp. 1–13, 2018.

[21] R. Liu, Q. Chen, G. Yu, and G. Y. Li, "Joint user association and resource allocation for multi-band millimeter-wave heterogeneous networks," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8502–8516, Dec. 2019.

[22] C. Chaieb, Z. Mlika, F. Abdelkefi, and W. Ajib, "On the optimization of user association and resource allocation in HetNets with mm-Wave base stations," *IEEE Syst. J.*, vol. 14, no. 3, pp. 3957–3967, Sep. 2020.

[23] R. Liu, M. Lee, G. Yu, and G. Y. Li, "User association for millimeter-wave networks: A machine learning approach," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4162–4174, Jul. 2020.

[24] Y. Zhang, L. Dai, and E. W. Wong, "Optimal BS deployment and user association for 5G millimeter wave communication networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 5, pp. 2776–2791, May 2021.

[25] P. Zhu, Z. Zhang, Y. Qian, C.-K. Wen, and S. Jin, "Joint user association and beamforming multi-objective optimization for distributed mmWave networks," *IEEE Commun. Lett.*, vol. 27, no. 2, pp. 620–624, Feb. 2023.

[26] E. M. Mohamed, S. Hashima, K. Hatano, E. Takimoto, and M. Abdel-Nasser, "Load balancing multi-player MAB approaches for RIS-aided mmWave user association," *IEEE Access*, vol. 11, pp. 15816–15830, 2023.

[27] Q. Xue, H. Xia, J. Mu, Y. Xu, L. Yan, and S. Ma, "User-centric association for dense mmWave communication systems with multi-connectivity," *IEEE Trans. Green Commun. Netw.*, vol. 8, no. 1, pp. 177–189, Mar. 2024.

[28] J. Moon, S. Kim, H. Ju, and B. Shim, "Energy-efficient user association in mmWave/THz ultra-dense network via multi-agent deep reinforcement learning," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 2, pp. 692–706, 2023.

[29] X. Zhang et al., "QoS-aware user association and transmission scheduling for Millimeter-wave train-ground communications," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 9, pp. 9532–9545, Sep. 2023.

[30] E. M. Taghavi, R. Hashemi, A. Alizadeh, N. Rajatheva, M. Vu, and M. Latva-aho, "Joint active-passive beamforming and user association in IRS-assisted mmWave cellular networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 8, pp. 10448–10461, Aug. 2023.

[31] S. K. Singh, V. S. Borkar, and G. S. Kasbekar, "User association in dense mmWave networks as restless bandits," *IEEE Trans. Veh. Technol.*, vol. 71, no. 7, pp. 7919–7929, Jul. 2022.

[32] M. R. Nalavade, G. S. Kasbekar, and V. S. Borkar, "Whittle index based user association in dense millimeter wave networks," 2024, *arXiv:2403.09279*.
[33] P. Whittle, "Restless bandits: Activity allocation in a changing world," *J. Appl. Probabil.*, vol. 25, no. A, pp. 287–298, 1988.
[34] V. Raghunathan, V. Borkar, M. Cao, and P. R. Kumar, "Index policies for real-time multicast scheduling for wireless broadcast systems," in *Proc. 27th Conf. Comput. Commun.*, 2008, pp. 1570–1578.
[35] V. S. Borkar and S. Pattathil, "Whittle indexability in egalitarian processor sharing systems," *Ann. Oper. Res.*, vol. 317, no. 2, pp. 417–437, 2022.
[36] A. Anand and G. de Veciana, "A Whittle's index based approach for QoE optimization in wireless networks," in *Proc. ACM Meas. Anal. Comput. Syst.*, 2018, pp. 1–39.
[37] K. E. Avrachenkov and V. S. Borkar, "Whittle index policy for crawling ephemeral content," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 1, pp. 446–455, Mar. 2018.
[38] K. Liu and Q. Zhao, "Indexability of restless bandit problems and optimality of Whittle index for dynamic multichannel access," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5547–5567, Nov. 2010.
[39] I. Kadota, A. Sinha, and E. Modiano, "Optimizing age of information in wireless networks with throughput constraints," in *Proc. IEEE Conf. Comput. Commun.*, 2018, pp. 1844–1852.
[40] I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Scheduling policies for minimizing age of information in broadcast wireless networks," *IEEE/ACM Trans. Netw.*, vol. 26, no. 6, pp. 2637–2650, Dec. 2018.
[41] M. Chen, K. Wu, and L. Song, "A whittle index approach to minimizing age of multi-packet information in IoT network," *IEEE Access*, vol. 9, pp. 31467–31480, 2021.
[42] G. Karthik, V. S. Borkar, and G. S. Kasbekar, "Scheduling in wireless networks using Whittle index theory," in *Proc. Nat. Conf. Commun. (NCC)*, 2022, pp. 367–372.
[43] C. H. Papadimitriou and J. N. Tsitsiklis, "The complexity of optimal queueing network control," in *Proc. IEEE 9th Annu. Conf. Struct. Complex. Theory*, 1994, pp. 318–322.
[44] R. K. Jain, D.-M. W. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination," Eastern Res. Lab., Digit. Equip. Corp., Hudson, MA, USA, Rep. DEC-TR-301, 1984.
[45] A. Hordijk and F. A. V. D. D. Schouten, "Average optimal policies in Markov decision drift processes with applications to a queueing and a replacement model," *Adv. Appl. Probabil.*, vol. 15, no. 2, pp. 274–303, 1983.
[46] V. S. Borkar, "Convex analytic methods in Markov decision processes," in *Handbook of Markov Decision Processes: Methods and Applications*. Boston, MA, USA: Springer, 2002, pp. 347–375.
[47] B. Shrader and T. C. Royster, "Cooperative multicast strategies under heterogeneous link loss rates," in *Proc. IEEE Glob. Telecommun. Conf.*, 2011, pp. 1–5.
[48] R. Framjee and J. Bredow, "Packet wait times in voice over IP wireless reverse links," in *Proc. IEEE 10th Consum. Commun. Netw. Conf. (CCNC)*, 2013, pp. 239–240.
[49] A. Razi, F. Afghah, and A. Abedi, "Channel-adaptive packetization policy for minimal latency and maximal energy efficiency," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2407–2420, Mar. 2016.
[50] P. Baziana, G. Fragkouli, and E. Sykas, "Analytical receiver collisions performance modeling of a multi-channel network," in *Proc. IEEE Conf. Russ. Young Res. Electr. Electron. Eng. (EIConRus)*, 2017, pp. 115–121.
[51] S. Pandey, K. Shandilya, and S. Agarwal, "Prioritized S-ALOHA for URLLC," in *Proc. Int. Wireless Commun. Mobile Comput. (IWCMC)*, 2020, pp. 1842–1847.
[52] V. Mancuso, P. Castagno, M. Sereno, and M. A. Marsan, "Serving HTC and critical MTC in a RAN slice," in *Proc. IEEE 22nd Int. Symp. World Wireless, Mobile Multimedia Netw. (WoWMoM)*, 2021, pp. 189–198.
[53] H. Taramit, L. O. Barbosa, and A. Haqiq, "Energy efficiency framework for time-limited contention in the IEEE 802.11 ah standard," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2021, pp. 1–6.
[54] Y. Li, Y. Dong, J. Wang, and B. Shim, "On-time communications over fading channels," *IEEE Trans. Veh. Technol.*, vol. 73, no. 1, pp. 605–619, 2024.
[55] N. Rastegardoost and B. Jabbari, "Minimizing Wi-Fi latency with unlicensed LTE opportunistic white-space utilization," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 1914–1926, Mar. 2019.
[56] A. M. Bedewy, Y. Sun, and N. B. Shroff, "The age of information in multihop networks," *IEEE/ACM Trans. Netw.*, vol. 27, no. 3, pp. 1248–1257, Jun. 2019.
[57] X. Zheng, S. Zhou, Z. Jiang, and Z. Niu, "Closed-form analysis of non-linear age of information in status updates with an energy harvesting transmitter," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4129–4142, Aug. 2019.
[58] M. Moltafet, M. Leinonen, and M. Codreanu, "On the age of information in multi-source queueing models," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 5003–5017, Aug. 2020.
[59] R. Liu and G. Yu, "User association for millimeter-wave ultra-reliable low-latency communications," *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 315–319, Feb. 2021.
[60] L. Sharma, J.-M. Liang, and S.-L. Wu, "Discontinuous reception based energy-efficient user association for 5G heterogeneous networks," *IEEE Access*, vol. 12, pp. 13634–13647, 2024.
[61] "2.4 GHz to 2.5 GHz 802.11G/b RF transceiver, PA, and Rx/Tx/antenna diversity switch," Data Sheet MAX2830, Analog devices, Wilmington, MA, USA, Mar. 2011. [Online]. Available: http://datasheets.maximintegrated.com/en/ds/MAX2830.pdf.
[62] W. Lou, X. Liu, P. Feng, and N. Wu, "An integrated 0.38–6 GHz, 9–12 GHz fully differential fractional-N frequency synthesizer for multi-standard reconfigurable MIMO communication application," *Analog Integr. Circuits Signal Process.*, vol. 78, pp. 807–817, Jan. 2014.
[63] S. Asmussen, *Applied Probability and Queues*, 2nd ed. New York, NY, USA: Springer, 2003.
[64] R. K. Sundaram, *A First Course in Optimization Theory*. Cambridge, U.K.: Cambridge Univer. Press, 1996.
[65] S. M. Ross, *Introduction to Stochastic Dynamic Programming*. Cambridge, MA, USA: Academic, 2014.
[66] V. K. Gupta, S. K. Singh, and G. S. Kasbekar, "Stability analysis of simple and online user association policies for millimeter wave networks," *IEEE Access*, vol. 9, pp. 62405–62429, 2021.
[67] G. S. Kasbekar, J. Kuri, and P. Nuggehalli, "Online association policies in IEEE 802.11 WLANs," in *Proc. 4th Int. Symp. Model. Optim. Mobile, Ad Hoc Wireless Netw.*, 2006, pp. 1–10.
[68] "Measurement results and final mmMAGIC channel models," Millimeter-Wave Based Mobile Radio Access Netw. Fifth Gener. Integr. Commun. (mmMAGIC), 5G-PPP, Heidelberg, Germany, Rep. H2020-ICT-671650-mmMAGIC/D2.2, 2017.
[69] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of millimeter wave communications for fifth-generation (5G) wireless networks—With a focus on propagation models," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 6213–6230, Dec. 2017.
[70] G. S. Kasbekar, P. Nuggehalli, and J. Kuri, "Online client-AP association in WLANs," in *Proc. RAWNET*, 2006, pp. 1–8.

RAVINDRA S. TOMAR received the master's degree in signal processing, communication and networks engineering from the Indian Institute of Technology Kanpur, India, in 2022. He is currently working as a Project Research Associate with the Indian Institute of Technology Bombay since 2023. From 2017 to 2019, he worked as a Junior Telecom Officer with Bharat Sanchar Nigam Limited, Gujarat, India. His research interests include signal processing for wireless communication systems, designing and analyzing resource allocation strategies to optimize performance in millimeter-wave communication systems.

MANDAR R. NALAVADE received the B.E. degree in electronics and telecommunication engineering from the Vivekanand Education Society's Institute of Technology, Mumbai, India, in 2011, the M.Tech. degree in electrical engineering from Veermata Jijabai Technological Institute, Mumbai, in 2014. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai. He worked as an Assistant Professor with MIT College, Pune, India, from 2015 to 2022. His research interests include design and analysis of resource allocation algorithms in wireless communication networks.

GAURAV S. KASBEKAR (Member, IEEE) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology (IIT) Bombay, Mumbai, India, in 2004, the M.Tech. degree in electronics design and technology from the Indian Institute of Science, Bengaluru, India, in 2006, and the Ph.D. degree from the University of Pennsylvania, Philadelphia, PA, USA, in 2011. He is currently an Associate Professor with the Department of Electrical Engineering, IIT Bombay. His research interests include communication networking and network security. He received the CEDT Design Medal for being adjudged the Best Masters Student in EDT at IISc.