User Scheduling and Beam Alignment in mmWave Networks With a Large Number of Mobile Users

Jihyun Lee, *Student Member, IEEE*, and Eylem Ekici, *Fellow, IEEE*

***Abstract*— In this paper, we study an optimal user scheduling with minimum beam alignment overhead in millimeter wave networks. The problem is posed as constrained Markov decision process (CMDP) with the goal of minimizing the average beam alignment overhead subject to the average rate constraint on each user. Under a certain assumption on the rate function of the users, by using a structural result derived from the Lagrangian formulation of the CMDP, we show that the optimal policy should keep scheduling the users that are scheduled in the previous time slot unless an abrupt change in the beam direction occurs. Using this result, the complexity of the problem decreases to polynomial in the number of users. In addition, we provide a heuristic deterministic algorithm that achieves $(1+\epsilon)$ approximation of the optimal solution, with smaller ϵ at the cost of longer transmission interval of each user. Lastly, to deal with the case where the assumption on the rate function does not hold due to beam conflicts between the users, we consider a system model that accounts for an angular channel information. A new CMDP is formulated for the problem and a heuristic algorithm based on the age information is proposed.**

***Index Terms*— Millimeter wave communications, beam alignment, constrained markov decision process, multi-user scheduling.**

I. Introduction

TO OVERCOME the high signal attenuation inherent at 30-300 GHz electromagnetic spectrum (which corresponds to 10mm to 1mm wavelength), millimeter wave (mmWave) networks must employ highly directional beamforming antennas. However, the use of narrow beams makes link establishment and maintenance much more challenging than traditional omnidirectional antennas as an mmWave link is established only when the transmit and receive antenna beams are steered in the correct directions. Moreover, even a slight misalignment of the beam directions or signal interruption can easily lead to complete link breakage and requires frequent beam re-alignments to maintain seamless connectivity especially under mobility.

To enable beamforming, a BS (base station) and a UE (user equipment) have to go through beam searching procedure, which typically incurs tens to hundreds of milliseconds overhead for the initial link establishment if exhaustive search over all possible combinations of transmission and reception directions is performed through a sequential pilot transmission [2]. To reduce the overhead, current standard activities [3], [4] suggest a two-stage beam search technique, in which a coarse grained sector level sweep is performed, followed by beam-level alignment phase. However, since the mmWave channel frequently varies over time in a mobile network, it may lead to unaffordable overhead to perform an exhaustive search from scratch every time. Hence, more efficient schemes exploiting the temporal correlation on the channel are preferred under mobility.

Fast beam search methods in mmWave networks under UE mobility have been extensively studied in the literature. In [5], a smart beam steering algorithm is proposed for fast directional link re-establishment under node mobility, which uses knowledge of the previous feasible antenna sector pair to narrow the sector search space. In [6] a priori aided (PA) channel tracking scheme is proposed to predict the support of beam space in the following time slots without a channel estimation under the assumption that there is no blockage. In [7], Kalman filter based tracking algorithm and an abrupt change detection method based on a threshold test are proposed and evaluated through simulation. In [8], based on linear dynamic model, the authors proposed probing protocols to identify the beam errors caused by link blockage and user movement. In [9], by leveraging a deep learning model that learns how to use an omni or quasi-omni uplink signal from the user to predict the best coordinated beamforming at the BSs, a low latency and low overhead beam training method is proposed for a high mobile user. Despite of the large volume, however, they only consider a single UE case.

With a large number of UEs, it is not affordable acquiring the channel information for all the links in the network. This makes the UE scheduling problem more complicated as the channel information is imperfect and the achievable data rate now depends on the beam alignment. Therefore, beam alignment and transmission scheduling in multi-user mmWave networks has drawn attention lately. In [6], the authors consider point to multipoint channel estimation, but it is assumed that the number of RF chains are equivalent to the number of UEs to guarantee the spatial multiplexing of all UEs, and thus the UE scheduling problem is not considered. In [10], transmission scheduling of multiple links is studied with an objective of optimizing network throughput, but the problem is defined for the multiple point-to-point links. In [11], energy efficient joint beam alignment protocols is addressed, with the goal to minimize the power consumption subject to rate constraints. However, the authors assume the time division

Manuscript received October 5, 2020; revised February 23, 2021; accepted April 9, 2021. Date of publication April 28, 2021; date of current version October 11, 2021. This work was supported in part by NSF under Grant CNS-1731698. This article was presented in part at the Proceedings of WiOpt 2019. The associate editor coordinating the review of this article and approving it for publication was J. Yang. *(Corresponding author: Jihyun Lee.)*

The authors are with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: lee.6432@osu.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2021.3074700.

Digital Object Identifier 10.1109/TWC.2021.3074700

multiplexing (TDM) of two users and the more general settings such as the spatial multiplexing of more than two users are not considered.

It is worth noting that none of the above considers long-term system performance of the beam alignment and UE transmission under multiple UE, multiple RF scenarios with temporally correlated channel. In this paper, we consider an mmWave network consisting of a fixed BS and multiple mobile UEs where more than one UE can be served at the same time communicating with directional antenna patterns. Our objective is to find a UE schedule that minimizes the average beam alignment overhead while satisfying the minimum average data transmission rate constraints for each UE. The main contributions of the paper are as follows:

- While most of the work on mmWave beam alignment algorithms focus on the link level performance improvement, we consider the problem of UE scheduling in an mmWave networks possibly involving a large number of UEs. In our system model, both the abrupt changes and slow variations in beam direction of each link due to UE mobility and environmental changes are taken into account. We formulate the scheduling problem with minimum data rate constraint as CMDP (constrained markov decision process) and its equivalent linear programming formulation is presented.
- To avoid exponential complexity in the number of UEs, we use Lagrangian multiplier method to convert the constrained problem to an unconstrained MDP and show that with the optimal policy, in each transmission schedule, the BS should continue to include the UE that is scheduled in the previous time unless an abrupt change happens to the UE. The optimal solution is a mixture of the solutions of unconstrained MDP and the problem reduces to finding the optimal mixture ratio. Using this structural result, the complexity of the problem decreases to polynomial in the number of UEs.
- For a practical use, a deterministic scheduling algorithm is proposed. With this algorithm, the BS schedules the UEs in a circular order but with different consecutive transmission times allocated to each UE. The length of the transmission time at each UE's turn is determined by the rate requirement of the UE. This algorithm ensures that every UE is given a transmission opportunity in a finite time. The algorithm achieves $(1+\epsilon)$ approximation of the optimal solution for cases where the minimum data rate requirement of individual user is sufficiently small compared to the channel capacity.
- To account for the case where the beams for the scheduled UEs are not resolvable, we discuss a channel model based on a random walk on the angular channel domain and formulate a CMDP model which incorporates the angular channel related information of each UE in addition to the age information. A heuristic algorithm which only considers the age information is proposed and shown to have smaller error as narrower mmWave beam is employed.

The rest of the paper is organized as follows. In Section II, we present the system model and formulate the problem as CMDP. In Section III, we design an optimal scheduling policy with an assumption on the cost and the reward functions. Based on the result of Section III, we provide a deterministic heuristic scheduling algorithm in Section IV. In Section V, a modified CMDP model that accounts for the beam collision between UEs is provided and a heuristic sub-optimal algorithm is proposed. Finally, we evaluate the performance of our proposed algorithms in Section VI and conclude the paper Section VII.

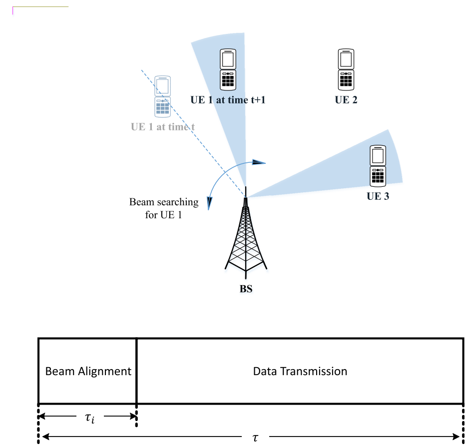


Fig. 1. An example of a multiuser mmWave network and its frame structure.

II. System Model and Problem Formulation

We consider a network consisting of a BS and N UEs labeled $n \in \{1, 2, \ldots, N\}$. At the BS, K RF chains are deployed so that it can serve K UEs at a time. As shown in Fig. 1, a time slot consists of two segments: beam alignment and data transmission. We assume that the BS decides which UEs to serve at the beginning of each time slot, based on the information it has on each link. When a UE is selected for data transmission for the time slot, the BS and the UE first decide which beam to use before its data transmission by searching over the possible combinations of beams and finding a best beam with highest SNR.

A. Beam Alignment (BA)

Beam alignment (BA) introduces overhead because it requires time and energy which can otherwise be used for data transmission. The overhead is proportional to the number of directions to be tested and thus depends on the prior information on the correct direction of the beam. For example, if a UE is static, the beam found in the previous time slot can be reused unless some abrupt changes in the environment happen. Let τ and τ_i^t denote the length of a time slot and the time consumed for beam alignment with UE i at time slot t respectively. We assume that τ is set to a fixed value such that it can support a seamless connectivity during the time slot under UE mobility (unless an abrupt change such as blockage happens in the environment). To maintain the connectivity, at the beginning of each time slot, beam search algorithm first finds which beam to use for the data transmission by sequentially checking the directions from the one with the highest probability of UE presence to minimize the searching overhead. After the beam alignment, for the remaining time $(\tau - \tau_i^t)$ of the time slot, the data is transmitted. We assume that at each time slot t, the channel gain (or path gain) of i-th UE is independent and identically distributed with the expectation of ξ_i. The data rate of UE i during the time slot t is then determined by the channel gain and the available transmission time in the slot. If blockage happens, the best beam direction in the next time slot can become completely independent to the previous one due to the loss of LOS path or change in the path of dominant reflection. In this case, a UE can disappear and re-appear in the following time slot uniform randomly on the entire angular search domain and we assume the search algorithm should scan the entire space. Note that as the time elapsed from the previous beam alignment increases, the information on the correct beam direction becomes more uncertain (i.e., less correlated to the previous direction) and the number of directions to be checked tends to increase and so does τ_i^t. We assume that the search algorithm should scan the entire space if the latest beam alignment becomes too outdated.

B. Problem Formulation: Weakly Coupled CMDP

When the BA procedure in Section II-A is employed, at time t, **the system state space** is defined by an N-tuple $\mathcal{X} = \{(x_1, x_2, \ldots, x_N) : x_n \in \{1, 2, \ldots, L\}\}$ and $x^t \in \mathcal{X}$ denote the state of time t, where x_n^t represents the amount of time that has been passed from the last beam alignment of UE n at the time of t. The state L is the state where a full beam search for the UE is performed, and the value of L can be determined by the system requirement. **The action space** is defined by an N-tuple $\mathcal{A} = \{(a_1, a_2, \ldots, a_N) : a_n \in \{0, 1\}, \sum_{n=1}^{N} \mathbb{1}(a_n = 1) \leq K\}$, where 0 stands for no transmission and 1 for transmission. $a^t \in \mathcal{A}$ is the action at time t. If UE n is scheduled for transmission at time t ($a_n^t = 1$), the BS first performs BA with the UE and then transmits data for the remaining period of the time slot. The constraint on the action space is due to the limited number of RF chains. At each time t, at most $K (\leq N)$ UEs can be selected for transmission.

1) Transition Probability: The transition probability defines the evolution of the system and reflects the natural independence of UE transitions, i.e. state and action taken for an UE don't influence the transition of the other UEs. Thus, the state of each user transit according to independent homogenous transition law, i.e. the probability that $x^{t+1} = j$ given $x^t = i$ and $a^t = a$ is: $P(x^{t+1} = j | x^t = i, a^t = a) = \prod_{i=1}^{N} P_{i_n, j_n}^{a_n}$, where

$$P_{i_n,j_n}^{a_n} = P(x_n^{t+1} = j_n | x_n^t = i_n, a_n^t = a_n) = \begin{cases} 1 & \text{if } a_n = 0, j_n = \min(L, i_n + 1), \\ q & \text{if } a_n = 1, j_n = 1, \\ p & \text{if } a_n = 1, j_n = L, \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

$q = 1 - p$ and $p \in (0, 1)$ is the probability of abrupt change (blocking). When a UE is scheduled for transmission, if a blockage does not happen, the state of the UE in the next time slot becomes 1 regardless of the state the UE is in at the current time slot. However, if blockage happens, the UE appears at a completely random direction in the next time slot independent of the current state, which means no prior information on the beam direction (i.e. no correlation with previous beam) and transition to state L. Transition probability graph of (1) is shown in Fig. 2. Since the transition probability is stationary, we can drop the time notation and use $P_{i,j}^a = \prod_{n=1}^{N} P(x_n^{t+1} = j_n | x_n^t = i_n, a_n^t = a_n)$.

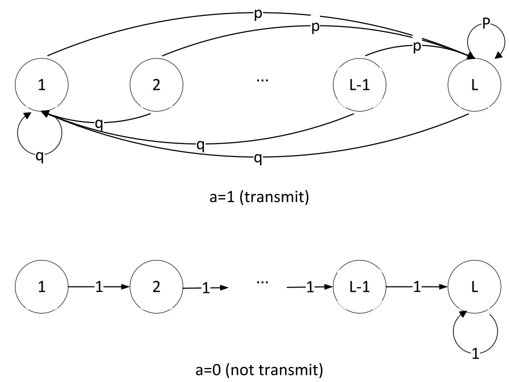


Fig. 2. Transition probability graph for each of the two possible actions (transmit or not transmit) for a single user.

2) Costs: We define data transmission and BA overhead with following assumptions.

(A1) Data transmission of UE n at state x_n and action a_n, $r_n(x_n, a_n)$ is non-increasing in x_n. BA overhead of UE n of state x_n and action a_n, $c_n(x_n, a_n)$ is non-decreasing in x_n.

(A2) The costs (or rewards) are dependent only on individual state and action and additive across UEs. More specifically,

- *Data transmission of n-th UE*: $r_n(x, a) = r_n(x_n, a_n)$.
- *BA overhead*: $c(x, a) = \sum_{n=1}^{N} c_n(x, a) = \sum_{n=1}^{N} c_n(x_n, a_n)$.

It is assumed that fixed per-antenna power and a fixed symmetric antenna configuration for each RF chains are used. (A1) is a reasonable assumption since as more time passes from the last beam alignment, the uncertainty on the correct beam direction increases.

Under the assumption, if $K < N$ the problem is in the form of weakly coupled CMDP [12] with a linkage constraint $\sum_{n=1}^{N} \mathbb{1}(a_n = 1) \leq K$. If $K = N$, the problem can be decomposed to K independent subproblems of a single UE case [13]. A mapping from a state $x \in \mathcal{X}$ to probabilities of selecting each possible action $a \in \mathcal{A}$ is called a policy and denoted u, where $u_x(a)$ is the probability that action a is taken at state x. Then, the expected average data transmission rate and power cost associated to policy u are given by [14]:

$$C(u) = \lim_{T \to \infty} \frac{1}{T} \mathbb{E}^u \sum_{t=1}^{T} c(x^t, a^t), \tag{2}$$

$$R_n(u) = \lim_{T \to \infty} \frac{1}{T} \mathbb{E}^u \sum_{t=1}^{T} r_n(x^t, a^t), \quad n = 1, 2, \ldots, N. \tag{3}$$

And the problem of optimizing a transmission policy is formally given by

$$C^* = \inf_u C(u) \quad \text{s.t. } R_n(u) \geq \gamma_n, \quad n = 1, 2, \ldots, N, \tag{4}$$

where γ_n is a constant minimum average data rate required by UE m.

C. Equivalent LP Formulation

We note that the MDP of our problem is unichain, i.e. under any deterministic policy, the corresponding Markov chain contains a single ergodic class. Thus, the problem of (4) is equivalent to the following linear programming (LP). For a detailed proof of the equivalence between the CMDP and its LP formulation, see Theorem 4.3 in [14]:

$$\textbf{(LP)} \quad \min_{v} \sum_{x\in\mathcal{X}}\sum_{a\in\mathcal{A}} c(x,a)v(x,a) \tag{5}$$

$$\text{s.t.} \sum_{x\in\mathcal{X}}\sum_{a\in\mathcal{A}} r_n(x,a)v(x,a) \geq \gamma_n,\ n=1,2,\ldots,N \tag{6}$$

$$v \in \mathcal{V}, \tag{7}$$

where

$$\mathcal{V} = \left\{ \begin{array}{l} v(x,a), x\in\mathcal{X}, a\ \in\mathcal{A}: \\ (C1)\displaystyle\sum_{x\in\mathcal{X}}\sum_{a\in\mathcal{A}} v(x,a)(\delta_y(x) - P^a_{x,y}) = 0, y\in\mathcal{X}, \\ (C2)\displaystyle\sum_{x\in\mathcal{X}}\sum_{a\in\mathcal{A}} v(x,a) = 1, \\ (C3) v(x,a)\geq 0, \forall x,a. \end{array} \right\}$$

$v(x,a) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}^u[\mathbb{1}(x^t = x, a^t = a)] = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}P^u(x^t=x,a^t=a)$, $a^t\in\mathcal{A}(x^t)$, which can be interpreted as the expected average number of times action a is executed in state x. δ_y is the Dirac probability measure concentrated on y and $P^u(E)$ is the probability of event E under policy u. The constraint set $\mathcal{V}$ can be interpreted as the conservation of flow through each of the states. An optimal policy can be computed from a solution to the LP as:

$$u_x(a) = \frac{v(x,a)}{\sum_{a\in\mathcal{A}} v(x,a)}, \tag{8}$$

where $u_x(a)$ is the probability that the controller executes action a when it encounters state x. For a detailed proof on the optimality of (8), see Theorem 4.2 in [14].

Curse of dimensionality In principle, the optimal policy can be found by solving the LP or dynamic programming (DP). However, the complexity of the CMDP is exacerbated for large number of UEs since the state and action spaces for the process typically consists of cross-product of those from individual UE processes, thus exponential in the number of UEs, i.e. $O(L^N)$ states. For heuristic techniques dealing with this problem, see [13] and [12]. In the following section, it will be shown that the complexity of our problem can be reduced to $O(N^K)$ by exploiting its structural property.

III. Optimal Scheduling Policy

In this section, an optimal UE scheduling algorithm with a polynomial complexity is presented.

A. Optimal Scheduling of Multiple UEs

We define a set $\mathcal{V}_I \subset \mathcal{V}$ by adding an additional constraint (C4) on the set $\mathcal{V}$ as follows.

$$\mathcal{V}_I = \left\{ \begin{array}{l} v(x,a), x\in\mathcal{X}, a\ \in\mathcal{A}: \\ (C1),(C2),(C3) \text{ and} \\ (C4) v(x,a) = 0, \forall (x,a)\notin\mathcal{G}, \end{array} \right\} \tag{9}$$

where

$$\mathcal{G} = \left\{ \begin{array}{l} (x,a), x\in\mathcal{X}, a\ \in\mathcal{A}: \\ \displaystyle\sum_{i=1}^{N}\mathbb{1}(x_i=1) = m, \sum_{i=1}^{N}\mathbb{1}(x_i=L) = N-m, \\ \quad 0\leq m\leq K, \\ \sum_{i=1}^{N}\mathbb{1}(x_i=1, a_i=0) = 0. \end{array} \right\}$$

Algorithm 1: Optimal Transmission Schedule

Data: $\mathcal{X}$, $\mathcal{A}$, $r(x,a)$ $c(x,a)$, $p^a_{i,j}$ for all $x\in\mathcal{X}$ and $a\in\mathcal{A}$, channel capacity ξ_i, minimum rate requirement γ_i for each UE i
Result: transmission scheduling $\{a^t\}_{t=1,2,\cdots}$
Initialization: find $v^*(x,a)$ for all $x\in X$ and $a\in A$ by solving (LP I). $t=1$.
while *true* **do**
  if $t=1$ **then**
    $x^t_i = L$ for all $i\in\mathcal{N}$
  end
  choose $a^t = a$ with probability $\frac{v^*(x^t,a)}{\sum_{u\in A} v^*(x^t,u)}$
  for $i=1$ *to* N **do**
    if $a^t_i = 1$ **then**
      if *an abrupt change occur to user* i **then**
        $x^{t+1}_i = L$
      else
        $x^{t+1}_i = 1$
      end
    else
      $x^{t+1}_i = x^t_i + 1$
    end
  end
  $t = t+1$
end

In words, $\mathcal{G}$ is a set of state and action pairs $(x,a)\in\mathcal{X}\times\mathcal{A}$ where less than or equal to K UEs are in state 1, the others are in state L and action 1 is assigned to at least one of the UEs in state 1. Note also that $\mathcal{V}_I\subset\mathcal{V}$ with most of the elements in $\mathcal{V}$ set to 0 and the number of unknowns in $\mathcal{V}_I$ is $O(N^K)$.

Theorem 1: If the assumptions (A1)-(A2) are satisfied for the CMDP problem (4), Algorithm 1 achieves an optimal solution.

Proof: See Appendix A. ■

Let us consider the following LP problem:

$$\textbf{(LP I)} \quad \min_{v} \sum_{x\in\mathcal{X}}\sum_{a\in\mathcal{A}} c(x,a)v(x,a) \tag{10}$$

$$\text{s.t.} \sum_{x\in\mathcal{X}}\sum_{a\in\mathcal{A}} r_m(x,a)v(x,a) \geq \gamma_m,\ m=1,2,\ldots,N \tag{11}$$

$$v \in \mathcal{V}_I, \tag{12}$$

Algorithm 1 uses the solution of (LP I) which is the same as (LP) except that the constraint (12) is replaced by $v \in \mathcal{V}_I$. From Theorem 1, the optimal solution of (LP I) is equivalent to the optimal solution of (LP). Therefore, the complexity of the algorithm is polynomial in N. In the proof of the theorem 1, we use a Lagrangian approach which converts a constrained control problem into an equivalent minmax non-constrained control problem. This approach solves the problem (4) by adding a Lagrangian multiplier per additional constraint while every Lagrangian multiplier results in a separate policy. Then the optimal randomized policy of a CMDP is computed as a mix policy of multiple optimal pure policies for all the Lagrangian multipliers. (See [14], [15] for comprehensive discussions about this topic). The theorem is proved by the structural property (38) showing that the condition (C4) holds for the pure policies of all the Lagrangian multipliers.

B. Application of Algorithm 1 to a Single UE Network

By Theorem 1, the optimal solution can be found by solving (LP I). For a single UE network, $N = 1$, $K = 1$ and $\mathcal{G} = \{(1,1),(L,0),(L,1)\}$. Therefore, there are only three unknowns in $\mathcal{V}_I$ and $v^* \in \mathcal{V}_I$ that minimizes the cost function can be found by a simple calculation. From (8), the optimal policy u is as follows:

$$u_x(1) = \begin{cases} 1 & \text{if } x = 1, \\ \dfrac{p\gamma}{pr(L,1) + qr(1,1) - q\gamma} & \text{if } x = L, \\ \text{arbitrary} & \text{else,} \end{cases} \tag{13}$$

and $u_x(0) = 1 - u_x(1)$.

The optimal scheduling of the UE is to keep transmitting as long as it succeeds and once a failure (blockage) happens, the BS flips a coin and transmits the data with probability $\frac{p\gamma}{pr(L,1)+qr(1,1)-q\gamma}$ and stays idle with probability $1-\frac{p\gamma}{pr(L,1)+qr(1,1)-q\gamma}$. Note that in the case of failure, it needs to search the beam from scratch while consecutive successes utilize the information of previous alignments. We note that the optimal solution (13) is a mix policy of u_1^* and u_2^* in Fig. 3.

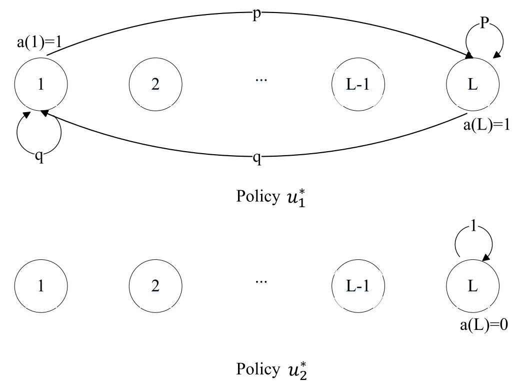


Fig. 3. Optimal solution of a single user case as a mixture of two deterministic policy. Note that the states $2, 3, \ldots, L-1$ are transient and the randomized decision occurs at state L.

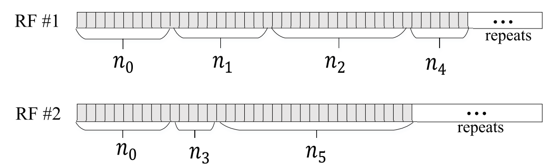


Fig. 4. An example transmission schedule of Algorithm 2, when there are five users ($N = 5$) and the number of RF chains are two ($K = 2$). In this example, the set of users $\mathcal{N} = \{1,2,3,4,5\}$ are partitioned into two groups such that $\mathcal{N}_1 = \{1,2,4\}$ and $\mathcal{N}_2 = \{3,5\}$.

IV. Deterministic Scheduling

Even though the optimal policy found in section III satisfies the constraints in expectation, depending on the realization, the throughput of a UE may in fact be less than the minimum constraint when the algorithm is executed in a finite time. Therefore, in practice, it is more desirable to use a deterministic policy to reduce the variance in constraints by ensuring every UE is given a transmission opportunity in a finite time [12]. In this section, we propose a heuristic deterministic policy, in which the UEs are first divided into K groups and the UEs in each group are scheduled in a circular order, but with different consecutive transmission time slots assigned to each of them depending on its channel and rate requirement. An example is shown in Fig. 4.

Note that $n^{(k)}$ in the intialization step is a vector and the ceiling function $ceil(\cdot)$ applies to each element. Let c^* denote the optimal cost function of the problem (4). Then, it is shown that for any $\epsilon > 0$, Algorithm 2 achieves an average cost function that is **at most** $(1+\epsilon)c^*$, by setting n_0 sufficiently large and inversely proportional to ϵ. For detailed analysis and discussion on the performance bound of Algorithm 2, see [1].

We also note that an assumption on the minimum required data rate for each user is necessary to guarantee the feasibility of user partition to K sets. Moreover, even though the feasibility is guaranteed, finding the partition is NP-complete since the problem reduces to a set partition problem [16]. In this paper, we assume that γ_i are sufficiently small compared to the channel capacity, thus a partition can be easily found. For example, if γ_i is given such that $\gamma_i \le \frac{q\bar{\xi}_i}{3(\lceil N/K \rceil - 1)}(\tau - qc(1,1) - pc(L,1))$, for all $i \in \mathcal{N}$, then any partition that allocate less than or equal to $\lceil N/K \rceil$ users to each RF chain will be a feasible solution. This issue is discussed more in section VI with simulation results.

V. Beam Collision and Age-Only Algorithms

In the previous sections, an optimal scheduling algorithm is presented under the assumption that for a given UE scheduling, there exist orthogonal beams to serve K UEs at the same time. However, it is possible that more than one UE are in the same angular bin of the transmission beams at some time slots. In the case, if the BS happens to schedule those UEs at the same time, ***the rate of a UE is no longer independent to the other UEs as the beams can be non- resolvable***. In this section, we discuss a case where the assumption (A1) and (A2) in section II does not hold. We first provide a new CMDP model which accounts for the non-orthogonal UEs in a

Algorithm 2: Fixed Transmission Time Schedule

Data: $\mathcal{X}$, $\mathcal{A}$, $r(x,a)$ $c(x,a)$, $p^a_{i,j}$ for all $x \in \mathcal{X}$ and $a \in \mathcal{A}$, channel capacity ξ_i, minimum rate requirement γ_i for each UE i

Result: transmission scheduling $\{a^t\}_{t=1,2,\ldots}$

initialization: find a partition $\{\mathcal{N}_k\}_{k=1}^K$, fix n_0 and set n_i, $\forall i \in \mathcal{N}$, as follows:

for all $k \in \mathcal{N}_k$, let $n^{(k)} = \{n_i\}_{i \in \mathcal{N}_k}$, $r^{(k)}(\cdot,\cdot) = \{r_i(\cdot,\cdot)\}_{i \in \mathcal{N}_k}$ and $\gamma^{(k)} = \{\gamma_i\}_{i \in \mathcal{N}_k}$.

$n^{(k)} = ceil([diag(pr^{(k)}(L,1) + qr^{(k)}(1,1)) - \gamma^{(k)}]^{-1}(qr^{(k)}(L,1) + qr^{(k)}(1,1) + \gamma^{(k)} n_0))$.

$t = 1$.

while *true* **do**

  $a_i^t = 0, \forall i \in \mathcal{N}$

  for $k = 1$ *to* K **do**

    $d(k,t) = t \bmod (\sum_{i \in \mathcal{N}_k} n_i + n_0)$

    if $1 \le d(k,t) \le n_1^{(k)}$ **then**

      $(a^{(k)})_1^t = 1$

    end

    for $j = 2$ *to* $|N_k|$ **do**

      if $\sum_{l=1}^{j-1} n_l^{(k)} < d(k,t) \le \sum_{l=1}^{j} n_l^{(k)}$ **then**

        $(a^{(k)})_j^t = 1$

      end

    end

  end

  $t = t + 1$

end

multi-user mmWave network and propose an algorithm based on the age information of the UEs.

A. Channel Model

We consider a transmitter with a ULA (Uniform Linear Array) of N_T antennas and receivers with a single antenna. The received signal at time $t = 1, 2, \ldots$ is then

$$y_t = \mathbf{h}_t \mathbf{x}_t + n_t, \tag{14}$$

where $\mathbf{h}_t$ is the $1 \times N_T$ MIMO channel matrix at time t, $\mathbf{x}_t$ is the $N_T \times 1$ transmit symbol vector at time t, n_t is the Gaussian noise at time t from $\mathcal{CN}(0, \sigma_n^2)$. $\mathbf{h}_t$ can be expressed as [17]

$$\mathbf{h}_t = \sqrt{N_T} \sum_{l=1}^{L_p} \xi_{t,l} \mathbf{a}_T^H(\theta_{t,l}^T), \tag{15}$$

where $\xi_{t,l} \sim \mathcal{CN}(0, \sigma^2)$ is the complex gain of l-th path at time t, $\theta_{t,l}^T$ is the angle of departure (AoD) of l-th path at time t. The direction θ and the physical direction $\phi \in [-\frac{\pi}{2}, +\frac{\pi}{2}]$ is $\theta = \frac{d\sin(\phi)}{\lambda}$, where d and λ are the spacings between two adjacent antenna elements and the signal wavelength. We assume $d = \frac{1}{2}\lambda$.

$$\mathbf{a}_\mathrm{T}(\theta^T) = \frac{1}{\sqrt{N_T}} \left[1, e^{-\iota 2\pi\theta^T}, \ldots, e^{-\iota(N_T-1)2\pi\theta^T}\right]^T. \tag{16}$$

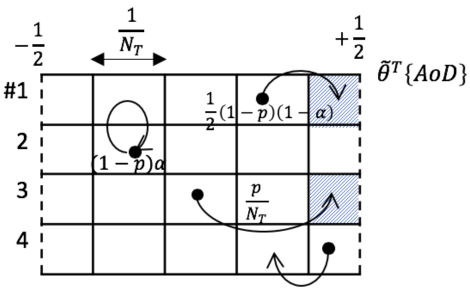


Fig. 5. An example of channel dynamics for a system of four UEs. Each UE has a single receiver antenna and the channel of each UE consists of a single path. Beam collision between UE 1 and UE 3 happens in a colored beam direction.

The following equation maps the channel $\mathbf{h}_t$ to virtual channel $\tilde{\mathbf{h}}_t$

$$\mathbf{h}_t = \tilde{\mathbf{h}}_t \mathbf{A}_T^H, \tag{17}$$

where $\mathbf{A}_T = [\mathbf{a}_\mathrm{T}(\tilde{\theta}_1^T), \ldots, \mathbf{a}_\mathrm{T}(\tilde{\theta}_{N_T}^T)]$ and $\tilde{\theta}_j^T = -\frac{1}{2} + \frac{j-1}{N_T}, \quad j = 1, \ldots, N_T$.

Channel dynamics Let us now model the temporal channel evolution. Firstly, we assume that the blockage occurs at each time slot after BA with the following Bernoulli random process B_t.

$$B_t = \begin{cases} 1 & \text{w.p. } p, \\ 0 & \text{w.p. } 1-p, \end{cases} \tag{18}$$

where p is the probability of blockage at time t.

The channel dynamics due to user mobility is modeled as follows. The angular component of the channel is modeled as Markovian where each path moves from the current column location i to another column location j in $\tilde{\mathbf{h}}_{t+1}$ in a Markovian manner. An example of a system with a single path, i.e., $L_p = 1$ is as follows:

$$\begin{aligned} &P\{\theta_{t+1,1}^T = \tilde{\theta}_i^T \mid B_t = 0\} \\ &= \begin{cases} \alpha & \text{if } \theta_{t,1}^T = \tilde{\theta}_i^T, \\ \dfrac{1-\alpha}{2} & \text{if } \theta_{t,1}^T = \tilde{\theta}_{(i-1) \bmod N_T}^T \text{ or } \tilde{\theta}_{(i+1) \bmod N_T}^T, \\ 0 & \text{otherwise}, \end{cases} \end{aligned} \tag{19}$$

$$\begin{aligned} &P\{\theta_{(t+1),1}^T \\ &= \tilde{\theta}_i^T \mid B_t = 1\} = \frac{1}{N_T}, \text{ for all } i \in \{1, 2, \ldots, N_T\}, \end{aligned} \tag{20}$$

where $\alpha \in [0, 1)$ indicates the mobility of the user. The path gain of l-th path $\xi_{t,l}$ is constant over a time slot and i.i.d. across time slots. Then, the dynamics of $\tilde{\mathbf{h}}_t$ can be seen as a discrete time random walk of non-zero elements on the $1 \times N_T$ grid matrix with the value of each non-zero element being drawn from an i.i.d. distribution at each time.

Beam alignment As shown in Figure 1, each time slot consists of τ symbol times. We assume that The transmitter sequentially picks an angular direction of $\mathbf{A}_T$ for a pilot transmission to search the corresponding column of $\tilde{\mathbf{h}}_t$ until it finds the UE.

Multi-user mmWave MIMO channel We now present the downlink channel model from a BS to multiple UEs. We consider a BS having N_T transmit antennas and N UEs with each UE having a single receive antenna. The channel from the BS to the i-th UE at time t is

$$\mathbf{h}_t^{(i)} = \sqrt{N_T}\sum_{l=1}^{L_p}\xi_{t,l}^{(i)}\mathbf{a}_T^H(\theta_{t,l}^{(i)}), \tag{21}$$

where $\xi_{t,l}^{(i)} \sim \mathcal{CN}(0,\sigma_i^2)$ is the complex gain of l-th path of UE i at time t, $\theta_{t,l}^{(i)}$ is the angle of departure (AoD) of l-th path to UE i at time t. Let $\mathbf{H}_t = [\mathbf{h}_t^{(1)},\dots,\mathbf{h}_t^{(N)}]^T$. Then, similar to (17), we can map the channel $\mathbf{H}_t$ to VCM $\widetilde{\mathbf{H}}_t$.

$$\mathbf{H}_t = \widetilde{\mathbf{H}}_t\mathbf{A}_T^H, \tag{22}$$

where $\mathbf{H}_t$ is the $N\times N_T$ MIMO channel matrix at time t. Note that the i-th row corresponds to the channel between the BS and the i-th UE. Each UE moves from the current column location to another column location in $\widetilde{\mathbf{H}}_{t+1}$ in a Markovian fashion with a transition probability defined by a UE mobility model as in (19). Note that this channel representation is especially useful in our problem setting because of the orthogonality between different columns in VCM. Beam conflicts among the UEs can be identified as the UEs being assigned in the same column of the VCM. Otherwise, the beams are resolvable at the UE side [18].

B. Age Based UE Scheduling

Unlike the MDP model we proposed in section II, ***representing the state of the UEs by its age information alone is not sufficient to identify the set of UEs that are in the same beam direction*** at each time slot. Therefore, in addition to the age information, we keep track of the latest angular location of each UE. The probability distribution of the UE location over the entire beam direction $[-\frac{1}{2},-\frac{1}{2}+\frac{1}{N_T},\dots,+\frac{1}{2}]$ then can be determined. Let us denote the state space as follows:

$$S_n = \{(x_n,\nu_n) : x_n\in\{0,1,2,\dots,\},\nu_n\in\{1,2,\dots,M\}\}, \tag{23}$$

where M is the number of angular bins (i.e., the number of columns of the VCM, for generalization, we used M instead of N_T), x_n denotes the age of UE n, i.e., the number of time slots that has been passed from the latest beam alignment with the UE, and ν_n denotes the latest beam direction of the UE n. Let $\tilde{P}$ denote the movement of the UE at the beginning of each time slot, where the (i,j)-th element of $\tilde{P}$ is the probability that the UE move to bin i from j. Then, the probability distribution of the UE's angular channel direction over M angular bins at state (x_n,ν_n), π_{x_n,ν_n} is as follows:

$$\pi_{x_n,\nu_n} = \begin{cases}\left((1-p)\tilde{P}+p\frac{1}{M}U\right)^{x_n-1}\tilde{P}\mathbf{1}_{\nu_n} & \text{if } x_n\in\{1,2,\dots\},\\ \lim_{k\to\infty}\left((1-p)\tilde{P}+p\frac{1}{M}U\right)^{k-1}\tilde{P}\mathbf{1}_{\nu_n} & \text{if } x_n=0,\end{cases} \tag{24}$$

where U is $M\times M$ all ones matrix. Let $\mathbf{1}_\nu$ denote a $M\times 1$ vector whose ν-th element is 1 and the others are zeros. (24) follows from that $\pi_{1,\nu} = \tilde{P}\mathbf{1}_\nu$ and $\pi_{x_n,\nu} = \big((1-p)\tilde{P} + p\frac{1}{M}U\big)\pi_{x_{n-1},\nu}$.

Given any current state $s_n = (x_n,\nu_n)$ and action a_n, the probability of each possible next state $x_n' = (s_n',\nu_n')$ is as follows.

$$P^{a_n}_{(x_n,\nu_n),(x_n',\nu_n')} = \begin{cases}(1-p)\pi_{(x_n,\nu_n)}(\nu_n') & \text{if } a_n=1, x_n'=1,\\ \frac{p}{M} & \text{if } a_n=1, x_n'=0,\\ 1 & \text{if } a_n=0, x_n'=x_n+1, \nu_n'=\nu_n,\\ 0 & \text{otherwise,}\end{cases} \tag{25}$$

where $\pi_{x_n,\nu_n}(\nu_n')$ denotes ν_n'-th element of π_{s_n,ν_n}. Now, let d denote the beam searching sequence, which is a permutation of $\{1,2,\dots,M\}$. Then, given any current state and action, $s_n = (x_n,\nu_n)$ and a_n and next state $s_n' = (x_n',\nu_n')$, the cost function of beam alignment for UE n is

$$c_n(s,a,s') = \begin{cases} d_{\nu_n}^{-1}(\nu_n') = \{i : d_{\nu_n}(i) = \nu_n'\} & \text{if } a_n = 1,\\ 0 & \text{otherwise,}\end{cases} \tag{26}$$

where $d_{v_n}(i)$ is the i-th element of d_{v_n}. Under a condition on $\tilde{P}$ such that $\tilde{P}_{i,j} > \tilde{P}_{k,j}$ if $|i-j|<|k-j|$ and $\tilde{P}_{i,j} = \tilde{P}_{k,j}$ if $|i-j| = |k-j|$ (i.e., the UE movement is symmetric and the probability of moving to the other direction gets smaller as it is further away from the current direction), $d_{\nu_n}^{-1}(\nu_n')$ can be written as follows.

$$d_{\nu_n}^{-1}(\nu_n') = 2[(\nu_n'-\nu_n) \bmod M] + \mathbf{1}(\nu_n'\geq\nu_n). \tag{27}$$

In addition to the cost, we also define a reward function, the data rate achievable for each UE i. Similar to section II,

$$r_n(s,a,s') = \begin{cases}\left[\tau - \frac{c_n(s_n,a_n,s_n')}{c_s}\right]r_s & \text{if } a_n=1, x_n'=1,\\ & \{j : \nu_n' = \nu_j', a_j = 1, j\neq n\} = \emptyset,\\ 0 & \text{otherwise,}\end{cases} \tag{28}$$

where r_s and c_s are the symbol rate and the symbol cost for UE n, respectively. So, $\frac{c_n(x_n,a_n,x_n')}{c_s}$ corresponds to the symbol times consumed for beam alignment. It is assumed that if the scheduled UEs are in the same beam direction, beam collision happens and the achievable data rates of them become zero. We note that the data rate of each UE now not only depends on its individual state but also the others. Now the problem is finding a UE schedule to minimize the average total cost subject to individual rate constraint:

$$C(u) = \lim_{T\to\infty}\frac{1}{T}\mathbb{E}^u\sum_{t=1}^{T}\sum_{n=1}^{N} c_n(s^t,a^t,s^{t+1}), \tag{29}$$

$$R_n(u) = \lim_{T\to\infty}\frac{1}{T}\mathbb{E}^u\sum_{t=1}^{T} r_i(s^t,a^t,s^{t+1}), \quad n=1,2,\dots,N. \tag{30}$$

$$C^* = \inf_u C(u) \quad \text{s.t. } R_n(u) \geq \gamma_n, \quad n = 1, 2, \ldots, N, \tag{31}$$

where γ_n is a constant minimum average data rate required by UE n. Note that the assumption (A1)-(A2) in Section II no longer hold due to the inter-dependency of the rate functions (28) between the UEs.

The problem is equivalent to the following linear programming:

$$\textbf{(LP II)} \quad \min_v \sum_{s\in\mathcal{S}} \sum_{a\in\mathcal{A}} c(s,a)v(s,a) \tag{32}$$

$$\text{s.t. } \sum_{s\in\mathcal{S}} \sum_{a\in\mathcal{A}} r_n(s,a)v(s,a) \geq \gamma_n, \tag{33}$$

$$n = 1, 2, \ldots, N$$

$$v \in \mathcal{V}_{II},$$

$$\mathcal{V}_{II} = \left\{ \begin{array}{l} v(s,a), s\in\mathcal{S}, a \in\mathcal{A}: \\ (C1) \quad \sum\limits_{s\in\mathcal{S}} \sum\limits_{a\in\mathcal{A}} v(s,a)(\delta'_s(s) \\ -P^a_{s,s'}) = 0, s'\in\mathcal{S}, \\ (C2) \quad \sum\limits_{s\in\mathcal{S}} \sum\limits_{a\in\mathcal{A}} v(s,a) = 1, \\ (C3) \quad v(s,a) \geq 0, \forall s, a, \end{array} \right\} \tag{34}$$

where $c(s,a) = \sum_{s'} P^a_{s,s'} c(s,a,s')$ and $r_i(s,a) = \sum_{s'} P^a_{s,s'} r_i(s,a,s')$. The following algorithm provides a sub-optimal solution to the problem (31) by scheduling the UEs based on the age information only. As the age-based algorithm works in the same way as Algorithm 1 and does not use the location information ν, beam collisions are not taken into account at the time of scheduling. The algorithm is designed to compensate for the possible rate loss due to beam collision by scheduling the users more often. Algorithm 3 is the same as Algorithm 1 with Data and Initialization step replaced as shown below in Algorithm 3. The rest of the algorithm follows the same steps in Algorithm 1 and are therefore omitted. However, in place of x_i^t in Algorithm 1, the age information field x_i^t of the state $s_i^t = (x_i^t, \nu_i^t)$ should be used.

Algorithm 3: Age-Based Transmission Schedule

Data: $\mathcal{S}$, $\mathcal{A}$, $r_i(s,a,s')$ $c_i(s,a,s')$, $p^a_{s,s'}$ for all $s\in\mathcal{S}$ and $a\in\mathcal{A}$, channel capacity $\tilde{\xi}_i$ and minimum rate requirement γ_i for each UE i

Result: transmission scheduling $\{a^t\}_{t=1,2,\ldots}$

Initialization: find $v^*(x,a)$ for all $x\in\mathcal{X}$ and $a\in\mathcal{A}$ by solving **(LP I)** with $\tilde{r}(x,a)$ defined in Lemma 2. Set $v(s,a) = \frac{\rho v^*(x,a)}{M^{|x|}}$. $t = 1$.

The following Lemma and Theorem show the performance bound of Algorithm 3. In Lemma 1 and Lemma 2, we first define a new reward function such that they satisfy (A1)-(A2) and show Algorithm 1 achieves the optimal solution for the new rate functions. Then, in the Theorem, the solution from Algorithm 1 is adjusted so that the rate constraints are satisfied when the new rate functions are replaced with the original rate functions. The adjustment is done by multiplying a constant ρ, which depends on K and M.z

Lemma 1: If $\sum_{1\leq i\leq M} \tilde{P}_{ij} = 1$ for all $1 \leq j \leq M$, $v(x,\nu,a) = \frac{1}{M^{|x|}} v^*(x,a)$ is feasible for (LP II), i.e., $v\in\mathcal{V}_{II}$, where $v^*(x,a)\in\mathcal{V}_{\mathcal{I}}$.

Proof: It is clear that $v(x,\nu,a) \geq 0$ for all (x,ν,a), and $\sum_{x,\nu,a} v(x,\nu,a) = 1$. It remains to show (C1), i.e., flow reservation for each state. Let $X(k) = \{s : |x| = k\}$, for $k = 0,1,\ldots,K$, and $|x| = \sum_{i=1}^N \mathbf{1}(x_i = 1)$. Without loss of generality, let $s'\in X(k)$.

$$\begin{aligned} &\sum_{s\in S}\sum_{a\in A} v(s,a)P^a_{s,s'} \\ &= \sum_{s\in S}\sum_{a\in A} v(x,\nu,a)P^a_{s,s'} \\ &= \sum_{j=1}^K \sum_{x\in X(j)} \sum_{a\in A} \frac{v^*(x,a)}{M^j} P^a_{x,x'} \sum_\nu P^a_{(x,x',\nu),\nu'} \\ &= \sum_{j=1}^K \sum_{x\in X(j)} \sum_{a\in A} \frac{v^*(x,a)}{M^j} P^a_{x,x'} \frac{1}{M^{k-j}} \\ &= \sum_{x\in S}\sum_{a\in A} \frac{v^*(x,a)}{M^k} P^a_{x,x'} = \sum_{a\in A} v(s',a). \end{aligned}$$

The third equality follows from the assumption that $\sum_{1\leq i\leq M} \tilde{P}_{ij} = 1$. ■

Lemma 2: Let $\tilde{r}_i(s,a,s') = \left[\tau - \frac{c_i(s_i,a_i,s_i')}{c_s}\right] r_s$ if $a_i = 1$, and 0 otherwise for all $i\in\mathcal{N}$. Then, $\tilde{r}_i(s_1,a) = \tilde{r}_i(s_2,a)$ if $x_1 = x_2$, i.e., $\tilde{r}_i(s,a)$ is independent of ν, given x. An optimal solution of (LP II) with the rate function $\tilde{r}(s,a,s')$ can be obtained by solving (LP I) with the rate function $\tilde{r}(x,a,x') = \sum_{\nu'} P^a_{(x,x',\nu),\nu'} \tilde{r}_i(s,a,s')$.

Proof: Note that the rate function of UE i, $\tilde{r}_i(s,a,s')$ does not depend on the other UEs, i.e., $\tilde{r}_i(s,a,s') = \tilde{r}_i(s_i,a_i,s_i')$.

$$\begin{aligned} \tilde{r}_i(s,a) &= \sum_{s'} P^a_{s,s'} \tilde{r}_i(s,a,s') \\ &= \sum_{x'} P^a_{x,x'} \sum_{\nu'} P^a_{(x,x',\nu),\nu'} \tilde{r}_i(s,a,s') \\ &= \sum_{x'} P^a_{x,x'} \tilde{r}_i(x,a,x') = \tilde{r}_i(x,a). \end{aligned}$$

From (27) and definition of $\tilde{r}$, $\sum_{\nu'} P^a_{(x,x',\nu),\nu'} \tilde{r}_i(s,a,s') = \sum_{\nu'} \pi_{x_i,\nu_i}(\nu_i')\left[\tau - \frac{d^{-1}_{\nu_i}(v_i')}{c_s}\right] r_s$. Since both $\pi_{x_i,\nu_i}(\nu_i')$ and $d^{-1}_{\nu_i}(v_i')$ depends only on $\nu - \nu'$ for given x and x', summing it over ν' is the same for any ν. We define it as $\tilde{r}_i(x,a,x')$. Then, it is easy to see that the optimal cost of (LP II) with $\tilde{r}_i(s,a,s')$ can be obtained by setting $v(s,a) = \frac{v(x,a)}{M^{|x|}}$, which is also feasible for (LP II) by Lemma 1. ■

Theorem 2: Let c^* denote the optimal cost function of the problem (31). Algorithm 3 achieves a cost function which is at most $(1+\epsilon)c^*$, where $\epsilon = \left(\prod_{i=0}^{K-1} \frac{M}{M-i}\right) - 1$. Note $\epsilon \to 0$ as $M \to \infty$ and $\epsilon = 0$ if $K = 1$.

Proof: For a given state set $X(k) = \{s : \sum_{i=1}^N (x_i = 1) = k, \sum_{i=1}^N (x_i = 0) = N-k\}$, there are M^k states. Among these, the number of state that has k distinctive elements in ν is $\frac{M!}{(M-k)!}$. Let $\rho = \left(\prod_{i=0}^{K-1} \frac{M-i}{M}\right)^{-1}$. Then, for a given $X(k)$, at least $\frac{1}{\rho}$ of the states have k distinctive elements in ν. Note that $\rho \to 1$ as $M \to \infty$ and $\rho - 1 = O(\frac{1}{M})$.

Now, we show the individual data rate R_i satisfies the minimum requirement.

$$\begin{aligned}
&R_i \\
&= \sum_{s,a} v(s,a) \sum_{s'} P^a_{s,s'} r(s,a,s') \\
&\geq \sum_{x,a} \frac{\rho v(x,a)}{M^{|x|}} \sum_{x'} P^a_{x,x'} \frac{1}{\rho} \sum_{\nu'} \sum_{\nu} P^a_{(x,x',\nu),\nu'} \tilde{r}(s,a,s') \\
&= \sum_{x,a} \frac{v(x,a)}{M^{|x|}} \sum_{x'} P^a_{x,x'} M^{|x|} \tilde{r}(x,a,x') \\
&= \sum_{x,a} v(x,a) \sum_{x'} P^a_{x,x'} \tilde{r}_i(x,a,x') = \sum_{x,a} v(x,a) \tilde{r}(x,a) \geq \gamma_i.
\end{aligned}$$

The first inequality follows from the fact that 1) $\sum_\nu P^a_{(x,x',\nu),\nu'} \tilde{r}(s,a,s')$ is the same for any ν' and 2) by setting $r_i(s,a,s') = 0$ for all $s' = (x', \nu')$ which do not have $|x'|$ distinctive element for a given x',

$$\begin{aligned}
&\sum_{\nu'} \sum_{\nu} P^a_{(x,x',\nu),\nu'} r(s,a,s') \\
&\qquad \geq \sum_{\{v': s' \in X(|x'|)\}} \sum_{\nu} P^a_{(x,x',\nu),\nu'} \tilde{r}(s,a,s') \\
&\qquad \geq \frac{M^{|x'|}}{\rho} \sum_{\nu} P^a_{(x,x',\nu),\nu'} \tilde{r}(s,a,s') \\
&\qquad = \frac{1}{\rho} \sum_{\nu'} \sum_{\nu} P^a_{(x,x',\nu),\nu'} \tilde{r}(s,a,s').
\end{aligned}$$

The second equality follows from exchanging the order of summation over v and v' and using the definition of $\tilde{r}(x,a,x')$ in Lemma 2. Let $\tilde{C}$ denote the cost of the optimal solution of LP I with $\tilde{r}$. Similarly, the cost function C is

$$\begin{aligned}
C &= \sum_{s,a} v(s,a) c(s,a) = \sum_{x,a} \frac{\rho v(x,a)}{M^{|x|}} \sum_{\nu} c(s,a) \\
&= \sum_{x,a} \rho v(x,a) c(x,a) = \rho \tilde{C}
\end{aligned}$$

Also, $C^* \geq \tilde{C}$ because the feasible set of $v(s,a)$ with $\tilde{r}_i(s,a,s')$ contains the feasible set with $r_i(s,a,s')$ and by Lemma 2, the optimal cost of $\tilde{r}_i(s,a,s')$ is the same as the optimal cost of (LP I) with $\tilde{r}_i(x,a)$. Therefore, $\frac{C}{C^*} \leq \frac{\rho \tilde{C}}{\tilde{C}} = \rho$, which completes the proof. ■

In Algorithm 3, we can also consider adding an additional step of selecting orthogonal UEs before the data transmission. Since the BS obtains the angular channel information of the scheduled UEs after the BA, in case of collision, the BS can choose UEs orthogonally in the angular domain, i.e., at most one UE for each angular bin, for the data transmission. However, this does not reduce the cost function as the cost is determined by the resource used for the BA, even though it achieves higher data rate than what is required by the constraints. Finding a tighter ρ to reduce the cost function using this method needs to be further studied.

VI. Performance Evaluation

In this section, the algorithms in Section III, IV and V are evaluated numerically. We consider an mmWave network consisting of a BS and multiple UEs. We assume the following probabilistic model for temporal changes in the beam orientation [17]:

- Slow change of beam orientation of each UE is modeled as an independent random walk on a polygon with M sides in angular domain. The change occurs at the beginning of each time slot. Due to UE mobility, with probability $\alpha \in (0,1)$ the beam direction does not change and with $\frac{(1-\alpha)}{2}$, it changes to either of the neighboring directions. See (19).
- An abrupt change in the beam direction occurs with probability $p > 0$. When the abrupt change occurs, the beam direction in the following time slot is determined by a uniform random selection on a polygon with M sides. See (20).
- At each time slot, beam alignment is performed independently for each scheduled user. The data transmission of UE n is $r_n(x_n, a_n) = (\tau - \tau_p N_p(x_n))\tilde{\xi}_n$ if UE n is scheduled, i.e., $a_n = 1$, and 0 otherwise. $\tilde{\xi}_n$ is a known constant data rate of UE n when beam-aligned, τ_p is the time required for a pilot transmission and $N_p(x_n)$ is the number of pilot transmission needed for a beam alignment when the UE is at state x_n. The BA overhead of UE n is $c_n(x_n, a_n) = c_p N_p(x_n)$ if $a_n = 1$, and 0 otherwise. c_p is the power consumption of a pilot transmission. Note that $N_p(x_n)$ is determined by the realisation of the probabilistic change in beam orientation of the UE.

Table I shows the average cost of the proposed algorithms for different number of UEs and rate constraints for a given α and p. The BS has 2 RF chains and can serve at most 2 UEs at each time slot. We assume the same data rate, i.e., $\tilde{\xi}_n = 1$ for every UE n when the UE is beam-aligned and 0 otherwise. The length of a time slot $\tau = 1$. For a pilot transmission, we set the cost $c_p = 0.1$ and $\tau_p = 0.05$. The quantities in Table I and Fig. 6 are averaged over 10^4 time slots and 100 different runs. We consider 5 UEs, i.e., $\mathcal{N} = \{1,2,3,4,5\}$ with different rate constraints. (B) is further divided into two cases in which different UE partition to each RF chain is used. For (C), we consider 2 UEs, i.e., $\mathcal{N} = \{1,2\}$. The minimum rate of UE 1 and UE 2 are set to be the same as the total rate of $\mathcal{N}_1$ and $\mathcal{N}_2$ of (B2), respectively.

(A) $\gamma =$ [0.3, 0.3 0.3, 0.3, 0.3],

(B1) $\gamma =$ [0.15, 0.15 0.15, 0.15, 0.15],, $\mathcal{N}_1 = \{1,2,3,4\}$, $\mathcal{N}_2 = \{5\}$,

(B2) $\gamma =$ [0.15, 0.15 0.15, 0.15, 0.15], $\mathcal{N}_1 = \{1,2,3\}$, $\mathcal{N}_2 = \{4,5\}$

(C) $\gamma =$ [0.45, 0.3], $\mathcal{N}_1 = \{1\}$, $\mathcal{N}_2 = \{2\}$.

Since there does not exist any partition of $\mathcal{N}$ into two which can support the rate requirement of (A), the γ of (A) is not feasible for Algorithm 2, whereas Algorithm 1 has the average cost very close to the LP solution. The variation from the optimal is due to finite sampling. As the number of runs and time slots increases, this will converge to the LP solution. We note that the action space of Algorithm 2 is a subset of Algorithm 1 since it uses a fixed assignment of UEs to each RF chain. However, when the rate requirement is sufficiently small as in (B), we can readily find a feasible partition $\mathcal{N}_1$

TABLE I

PERFORMANCE COMPARISON WITH DIFFERENT RATE REQUIREMENTS (γ) FOR GIVEN α AND p ($\alpha = 0.5$, $p = 0.1$)

		(A)	(B1)	(B2)	(C)
LP solution		1.0293	0.5147	0.5147	0.5147
Alg. 1 (95 % CI)		1.0297 (±0.0017)	0.5137 (±0.0021)	0.5147 (±0.0017)	0.5126 (±0.0057)
Alg. 2	n_0=11	-	0.8018	0.9198	0.6987
	n_0=44	-	0.5948	0.6327	0.5622

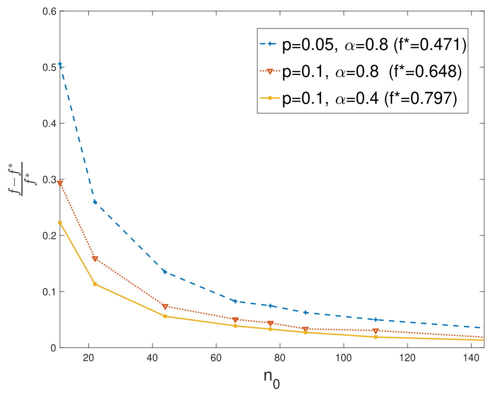
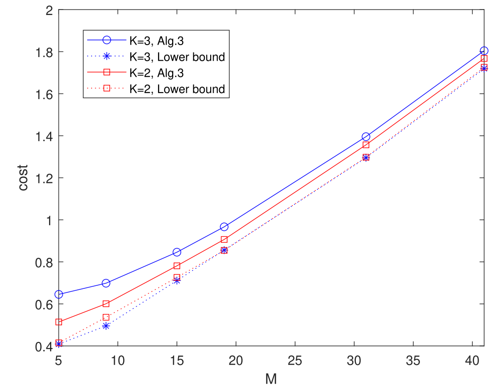


Fig. 6. Performance comparison with different settings of UE mobility (α), blocking probability (p) and n_0 for a given minimum rate requirement γ.

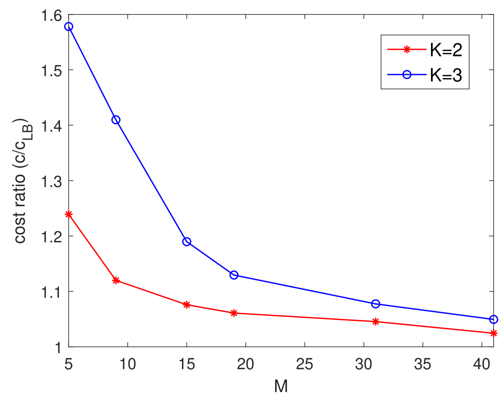
and $\mathcal{N}_2$. The average cost of the optimal policy does not depend on the partition as long as the given γ is feasible for partition. Therefore, the performance of Algorithm 1 is the same for (B1) and (B2). However, depending on the partition, the performance of Algorithm 2 can be different. This is because for a given performance error, different idle period n_0 is required [1]. It is shown in Table I. that the cost of (B2) is larger than (B1) for both $n_0 = 11$ and $n_0 = 44$. Regardless of which partition we use, however, the average cost decreases as n_0 increases. The result of (C) shows how the number of users affects the performance of the network. The BS serves UE 1 and UE 2 separately using different RF chain (and its corresponding antenna sets). For each RF, the rate constraint is the same as the total rate of (B2) and thus the optimal cost of (C) is the same as (B2). This is no surprise because for each $k \in \{1,2\}$, if we merge the states $\{(L\mathbf{1}_{\mathcal{N}_k} - (L-1)\mathbf{1}_i)\}_{i\in\mathcal{N}_k}$ of (B2) to a single state (state 1), we can obtain the same MDP as (C). However, the average cost of (C) is lower than (B2) when Algorithm 2 is used. This is because more users with smaller rate requirements causes larger rounding errors when the transmission interval n_i is determined in the initialization step.

Fig. 6 shows the performance of Algorithm 2 for different UE mobility (α) and probability of blockage (p) with a given rate requirement γ. Five UEs with $\gamma_n = 0.1$ are assigned to each RF chain. Obviously, the average cost increases as either the probability of blocking (abrupt change) or UE mobility increases for both the optimal solution and Algorithm 2. It is also shown that the normalized error $\frac{f-f^*}{f^*}$ of Algorithm 2 decreases as n_0 increases as expected in section IV.

Fig. 7 shows the cost function and an upper bound on the fractional errors of the algorithm 3 for different values of



(a) Comparison of cost



(b) Upper bound on the fractional error

Fig. 7. Performance of Algorithm 3 for different beam-width ($1/M$) and the number of RF chains (K) with N = 9 and $\alpha = 0.6$.

M and K. The simulation setting is the same as before but the rate of a UE using colliding beams with others becomes zero. The lower bound in Fig. 7(a) is computed by setting ρ to 1, which is an optimal solution of LP II with $r(s,a,a')$ replaced by $\tilde{r}(s,a,s')$. Since $r(s,a,a') \leq \tilde{r}(s,a,s')$ for all s, a and s', no other algorithm can do better. As proven in Theorem 3, the error bound decreases as M increases, i.e., as narrower beams (larger number of antennas) are used, due to less probability of a beam collision. However, as K increases, i.e., more UEs are scheduled at the same time, the performance of the algorithm degrades due to the rate loss caused by higher probability of beam collisions between the scheduled UEs. Even though the bound is loose and the cost ratio in Fig. 7(b) is very high for small M, the performance bound improves significantly around the usual M used in a practical mmWave system. For 15 degrees beamwidth ($M = 24$), the cost ratio in Fig. 7(b) is bounded by 0.61 for $K = 2$. The beam alignment cost increases in M as the expected beam search time increases accordingly. Fig. 8 shows how the algorithm performance changes depending on the number of UEs. The beam alignment cost increases linearly as $N(\alpha)$ increases with a given rate constraints. For a large N, and the total number of pilot overhead increases due to higher total transmission rate required. However, as the UE scheduling is performed before the beam alignment, the probability of beam conflict depends on K, not N. In fact, Algorithm 3 is designed to have ρ fractional error which only depends on M and K.

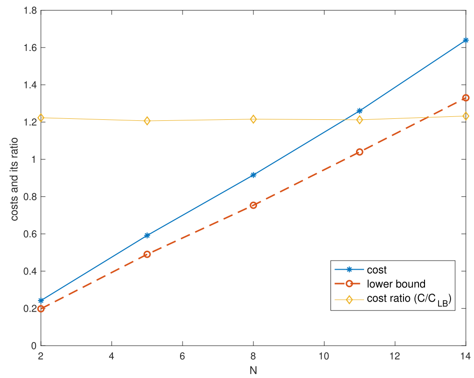


Fig. 8. Performance of Algorithm 3 for different number of UEs (N) with $K = 3$, $\alpha = 0.7$ and $M = 9$.

VII. Conclusion

This paper explores transmission scheduling algorithms for mmWave networks under user mobility, where the beam alignment is required before each transmission. The problem of minimizing beam alignment overhead under the minimum rate constraints is formulated as CMDP under certain assumptions on rate functions and cost function. From the structural result derived from the Lagrangian formulation of the MDP, it is shown that the complexity of the CMDP can be reduced from $O(L^N)$ to $O(N^K)$, and based on the result, a heuristic deterministic algorithm is proposed and shown to achieve $(1 + \epsilon)$ approximation of the optimal solution. Then we move on to investigate a case where the assumptions on rate functions do not hold. One example is where the beams for the scheduled UEs are not resolvable. To deal with the case, we discuss a channel model which is based on a random walk on the angular channel domain and formulate a new CMDP model that accounts for the beam collisions on the angular channel. A heuristic sub-optimal algorithm which only considers the age information is proposed and shown to have a smaller error as narrower beams are employed.

In our future work, a deterministic user scheduling algorithm which accounts for the beam collision will be studied based on the result of section V. In addition, an optimization problem with different power allocation and beam alignment methods will also be investigated such that it includes an option of beam tracking for non-scheduled users or reuse of the previous beam alignment for scheduled users.

Appendix

A. Proof of Theorem 1

Proof: In this proof, we use $1 \times N$ vector notation to present states and actions of multiple users. When S is a set of indices, $S \subseteq \{1, 2, \dots, N\} = \mathcal{N}$, we let $\mathbf{1}_S$ denote a $1 \times N$ vector whose components on S are 1 and 0 elsewhere, e.g. if $N = 5$ and $S = \{2, 3\}$, then $\mathbf{1}_S = (0, 1, 1, 0, 0)$. $\mathbf{1}_S$ multiplied by a constant c is $c\mathbf{1}_S$, e.g., $L\mathbf{1}_{\mathcal{N}} = \{L, L, \dots, L\}$. Similarly, for some $\mathbf{x} \in \mathcal{X}$, we let $\mathbf{x}_S$ denote $\mathbf{x} \circ \mathbf{1}_S$, where $\circ$ is entry-wise product such that $(\mathbf{x}_S)_j = x_j$ if $j \in S$ and 0 if $j \notin S$. Let $I_a = \{i : a_i = 1\}$. We reformulate the CMDP as a parameterized unconstrained MDP. For each Lagrangian multiplier $\lambda > 0$, define the instantaneous Lagrangian cost by

$$c(x, a; \lambda) = c(x, a) - \sum_{i=1}^{N} \lambda_i r_i(x, a). \tag{35}$$

We note that for any fixed $\lambda > 0$, $c(x, a; \lambda)$ is increasing in x since it is the sum of increasing functions. The Lagrangian average cost for a policy u is then defined as follows:

$$J(u; \lambda) = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}^u \sum_{t=1}^{T} c(x^t, a^t : \lambda). \tag{36}$$

The corresponding unconstrained MDP is to minimize the above Lagrangian average cost:

$$V = \inf_u J(u; \lambda) \quad u^*_\lambda = \arg\inf_u J(u; \lambda). \tag{37}$$

The proof proceeds in two steps.

STEP 1 (Pure Policy for a Given λ): Bellman Equation with cost function (35) for a discounted cost MDP is as follows.

$$V_\alpha(x) = \min_{a\in\mathcal{A}}\{c(x, a; \lambda) + \alpha \sum_{y\in\mathcal{X}} P^a_{x,y} V_\alpha(y)\} = \min_{a\in\mathcal{A}} Q_\alpha(x, a).$$

where $0 < \alpha < 1$ is the discount factor. This can be computed by the recursion, $V^{t+1}_\alpha = \min_{a\in\mathcal{A}} Q^{t+1}_\alpha(x, a)$, where $Q^{t+1}_\alpha(x, a) = c(x, a; \lambda) + \alpha \sum_{x\in\mathcal{X}} P^a_{x,y} V^t_\alpha(y)$. We used the Bellman equation for a discounted cost MDP since an average cost MDP inherits the properties of a discounted cost MDP [15]. From now on, we omit subscript α. we first show that for all $\mathbf{a} \in \mathcal{A}$ the following holds for all $I \subseteq I_a$, $\mathbf{b} \neq \mathbf{a}$.

$$\begin{aligned} &Q(L\mathbf{1}_{\mathcal{N}}, \mathbf{a}) - Q(L\mathbf{1}_{\mathcal{N}} - k\mathbf{1}_I, \mathbf{b}) \\ &\quad \le Q(L\mathbf{1} - k\mathbf{1}_I, \mathbf{a}) - Q(L\mathbf{1}_{\mathcal{N}} - k\mathbf{1}_I, \mathbf{b}), 1 \le k \le L - 1. \end{aligned} \tag{38}$$

(38) implies that if action $\mathbf{a}$ is optimal at state $L\mathbf{1}_{\mathcal{N}}$, so is it at states $L\mathbf{1}_{\mathcal{N}} - k\mathbf{1}_I$, for all $I \subseteq I_\mathbf{a}$ and $k = 1, 2, \dots, L - 1$.

$$\begin{aligned} &Q(L\mathbf{1}_{\mathcal{N}}, \mathbf{b}) - Q(L\mathbf{1}_{\mathcal{N}} - k\mathbf{1}_I, \mathbf{b}) \le Q(L\mathbf{1}_{\mathcal{N}}, \mathbf{a}) \\ &\qquad -Q(L\mathbf{1}_{\mathcal{N}} - k\mathbf{1}_I, \mathbf{a}) \\ &\Leftrightarrow \sum_{i\in I\cap I_b} [c_i(L, 1; \lambda) - c_i(L-k, 1; \lambda)] + \sum_{J\subseteq I_b} p^{|I_b|-|J|} q^{|J|} \\ &\qquad [V(L\mathbf{1}_{\mathcal{N}} - (L-1)\mathbf{1}_J) - V(L\mathbf{1}_{\mathcal{N}} \\ &\qquad -(L-1)\mathbf{1}_J - (k+1)\mathbf{1}_{I\cap\bar{I}_b})] \\ &\le \sum_{i\in I\cap I_b} [c_i(L, 1; \lambda_i) - c_i(L-k, 1; \lambda_i)] \\ &\qquad + \sum_{i\in I\cap\bar{I}_b} [c_i(L, 1; \lambda_i) - c_i(L-k, 1; \lambda_i)]. \end{aligned}$$

It suffices to show that for any $S \subset \mathcal{N}$ and $\mathbf{x} \in \mathcal{X}$, for $2 \le k \le L - 2$,

$$\begin{aligned} &V(L\mathbf{1}_S + \mathbf{x}_{\bar{S}}) - V((L-k)\mathbf{1}_S + \mathbf{x}_{\bar{S}}) \\ &\qquad \le \sum_{i\in S}\{c_i(L, 1; \lambda_i) - c_i(L-k-1, 1; \lambda_i)\}, \end{aligned} \tag{39}$$

where $\bar{S}$ is a complement set of S. Since V converges for any initial condition, we can select $V^1(\mathbf{x}) = 0$, for all $\mathbf{x} \in \mathcal{X}$.

Then (39) holds at $t=1$. Suppose that (39) holds for $t=s$. That is:

$$\begin{aligned}&V^s(L\mathbf{1}_S+\mathbf{x}_{\bar{S}}) - V^s((L-k)\mathbf{1}_S+\mathbf{x}_{\bar{S}})\\ &\le \sum_{i\in S}\{c_i(L,1;\lambda_i)-c_i(L-k-1,1;\lambda_i)\}, 2\le k\le L-2.\end{aligned}$$

Then, there exist some action $\mathbf{a}^{(1)}$ and $\mathbf{a}^{(2)}$ such that $V^{s+1}(L\mathbf{1}_S+\mathbf{x}_{\bar{S}}) = \min_{\mathbf{a}\in\mathcal{A}} Q^{s+1}(L\mathbf{1}_S+\mathbf{x}_{\bar{S}},\mathbf{a}) = Q^{s+1}(L\mathbf{1}_S+\mathbf{x}_{\bar{S}},\mathbf{a}^{(1)})$, and $V^{s+1}((L-k)\mathbf{1}_S+\mathbf{x}_{\bar{S}}) = \min_{\mathbf{a}\in\mathcal{A}} Q^{s+1}((L-k)\mathbf{1}_S+\mathbf{x}_{\bar{S}},\mathbf{a}) = Q^{s+1}((L-k)\mathbf{1}_S+\mathbf{x}_{\bar{S}},\mathbf{a}^{(2)})$. Then, at $t=s+1$,

$$\begin{aligned}&V^{s+1}(L\mathbf{1}_S+\mathbf{x}_{\bar{S}}) - V^{s+1}(L\mathbf{1}_S+\mathbf{x}_{\bar{S}})\\ &= \underbrace{Q^{s+1}(L1_S+\mathbf{x}_{\bar{S}},\mathbf{a}^{(1)}) - Q^{s+1}(L\mathbf{1}_S+\mathbf{x}_{\bar{S}},\mathbf{a}^{(2)})}_{\le 0 \text{ by optimality}}\\ &\quad +Q^{s+1}(L\mathbf{1}_S+\mathbf{x}_{\bar{S}},\mathbf{a}^{(2)}) - Q^{s+1}((L-k)\mathbf{1}_S+\mathbf{x}_{\bar{S}},\mathbf{a}^{(2)})\\ &\le \sum_{i\in S\cap S_0}[c_i(L,1;\lambda_i)-c_i(L-k,1;\lambda_i)]\\ &\quad + \sum_{J\subseteq S_0} p^{|S_0|-|J|}q^{|J|}\ \ [V^s(L\mathbf{1}_{\mathcal{N}}-(L-1)\mathbf{1}_J)\\ &\quad -V^s(L\mathbf{1}_{\mathcal{N}}-(L-1)\mathbf{1}_J-(k+1)\mathbf{1}_{S\cap\bar{S}_0})]\\ &\le \sum_{i\in S}[c_i(L,1;\lambda_i)-c_i(L-k,1;\lambda_i)] \qquad (40)\\ &\le \sum_{i\in S}[c_i(L,1;\lambda_i)-c_i(L-k-1,1;\lambda_i)], \qquad (41)\end{aligned}$$

where $S_0 = I_{a^{(2)}} = \{i : a_i^{(2)} = 1\}$. (40) follows from the assumption at $t=s$, and (41) follows since $c_i(x,1;\lambda)$ is increasing in x for $\lambda>0$.

STEP 2 (Randomized Mixture of Multiple Policies): Now, let $\Lambda_{\mathbf{a}} = \{\lambda : Q(L\mathbf{1}_{\mathcal{N}},\mathbf{a}) - Q(L\mathbf{1}_{\mathcal{N}},\mathbf{b}) < 0\}$, for all $\mathbf{b}\neq\mathbf{a}$, and let $\mathcal{X}_{\mathbf{a}} = \{x : p^{\mathbf{a}}_{L\mathbf{1}_{\mathcal{N}},x} > 0\}$ be the one-step reachable states from $L\mathbf{1}_{\mathcal{N}}$ by taking the action $\mathbf{a}$. Then by (38), for any $x\in\mathcal{X}_{\mathbf{a}}$, $Q(x,\mathbf{a}) < Q(x,\mathbf{b})$, for all $\mathbf{b}\neq\mathbf{a}$. Also, $\{y : p^{\mathbf{a}}_{x,y} > 0, \forall x\in\mathcal{X}_{\mathbf{a}}\}\subseteq\mathcal{X}_{\mathbf{a}}$. So, $v(x,a)=0$ for all $(x,a)\notin\mathcal{G}_{\mathbf{a}}$, where $\mathcal{G}_{\mathbf{a}} = \{(x,\mathbf{a}) : x\in\mathcal{X}_{\mathbf{a}}\}$. Since $\cup_{\mathbf{a}\in\mathcal{A}}\Lambda_{\mathbf{a}} = \{\lambda : \lambda > 0\}$ and $\cup_{\mathbf{a}\in A}\mathcal{G}_{\mathbf{a}} = \mathcal{G}$, this completes the proof. ∎

References

[1] J. Lee and E. Ekici, "Beam alignment and user scheduling in mmWave networks under mobility," in *Proc. Int. Symp. Modeling Optim. Mobile, Ad Hoc, Wireless Netw. (WiOPT)*, Jun. 2019, pp. 1–8.

[2] 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.

[3] *IEEE Standard for Information Technology–Telecommunications and Information Exchange Between Systems–Local and Metropolitan Area Networks–Specific Requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band*, IEEE Standard 802.11ad-2012 (Amendment to IEEE Std 802.11-2012, as amended by IEEE Std 802.11ae-2012 IEEE Std 802.11aa-2012), Dec. 2012, pp. 1–628.

[4] J. Wang *et al.*, "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 8, pp. 1390–1399, Oct. 2009.

[5] A. Patra, L. Simić, and P. Mähönen, "Smart mm-wave beam steering algorithm for fast link re-establishment under node mobility in 60 GHz indoor WLANs," in *Proc. 13th ACM Int. Symp. Mobility Manage. Wireless Access*, 2015, pp. 53–62.

[6] X. Gao, L. Dai, Y. Zhang, T. Xie, X. Dai, and Z. Wang, "Fast channel tracking for terahertz beamspace massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 5689–5696, Jul. 2017.

[7] C. Zhang, D. Guo, and P. Fan, "Tracking angles of departure and arrival in a mobile millimeter wave channel," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.

[8] Y. M. Tsang and A. S. Y. Poon, "Detecting human blockage and device movement in mmWave communication system," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2011, pp. 1–6.

[9] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, "Deep learning coordinated beamforming for highly-mobile millimeter wave systems," *IEEE Access*, vol. 6, pp. 37328–37348, 2018.

[10] H. Shokri-Ghadikolaei, L. Gkatzikis, and C. Fischione, "Beam-searching and transmission scheduling in millimeter wave communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 1292–1297.

[11] R. A. Hassan and N. Michelusi, "Multi-user beam-alignment for millimeter-wave networks," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2018, pp. 1–7.

[12] C. Boutilier and T. Lu, "Budget allocation using weakly coupled, constrained Markov decision processes," in *Proc. 32nd Conf. Uncertainty Artif. Intell. (UAI)*, 2016, pp. 52–61.

[13] D. Bertsimas and J. Niño-Mora, "Restless bandits, linear programming relaxations, and a primal-dual index heuristic," *Oper. Res.*, vol. 48, no. 1, pp. 80–90, Feb. 2000.

[14] E. Altman, *Constrained Markov Decision Processes*, vol. 7. Boca Raton, FL, USA: CRC Press, 1999.

[15] V. Krishnamurthy, *Partially Observed Markov Decision Processes*. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[16] J. Fairbrother and A. N. Letchford, "Projection results for the K-partition problem," *Discrete Optim.*, vol. 26, pp. 97–111, Nov. 2017.

[17] J. Seo, Y. Sung, G. Lee, and D. Kim, "Training beam sequence design for millimeter-wave MIMO systems: A POMDP framework," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1228–1242, Mar. 2016.

[18] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

Jihyun Lee (Student Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2007 and 2009, respectively, and the Ph.D. degree in electrical and computer engineering from The Ohio State University, Columbus, OH, USA, in 2020. Her research interests include cognitive radio, network optimization, and resource management in wireless networks.

Eylem Ekici (Fellow, IEEE) received the B.S. and M.S. degrees in computer engineering from Boğaziçi University, Turkey, in 1997 and 1998, respectively, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, in 2002. He is currently a Professor with the Department of Electrical and Computer Engineering, The Ohio State University. His current research interests include dynamic spectrum access, vehicular communication systems, and next generation wireless systems, with a focus on algorithm design and resource management. He was the General Co-Chair of ACM MobiCom 2012. He was also the TPC Co-Chair of the IEEE INFOCOM 2017. He is also an Associate Editor-in-Chief of the IEEE TRANSACTIONS ON MOBILE COMPUTING, and a former Associate Editor for the IEEE/ACM TRANSACTIONS ON NETWORKING, the IEEE TRANSACTIONS ON MOBILE COMPUTING, and *Computer Networks*.