Radar Enhanced Multi-Armed Bandit for Rapid Beam Selection in Millimeter Wave Communications

Akanksha Sneh[1], Sumit Darak[1], Shobha Sundar Ram[1] and Manjesh Hanawal[2]

[1]Indraprastha Institute of Information Technology Delhi, New Delhi 110020 India

[2]Indian Institute of Technology Bombay, Mumbai 400076 India

E-mail: {akankshas, sumit, shobha }@iiitd.ac.in; mhanawal@iitb.ac.in

Abstract—**Multi-arm bandit (MAB) algorithms have been used to learn optimal beams for millimeter wave communication systems. Here, the complexity of learning the optimal beam linearly scales with the number of beams, leading to high latency when there are a large number of beams. In this work, we propose to integrate radar with communication to enhance the MAB learning performance by searching only those beams where the radar detects a scatterer. Further, we use radar to distinguish the beams that show mobile targets from those which indicate the presence of static clutter, thereby reducing the number of beams to scan. Simulations show that our proposed radar-enhanced MAB reduces the exploration time by searching only the beams with distinct radar mobile targets resulting in improved throughput.**

Index Terms—**multi-armed bandit, joint radar communication, upper confidence bound, analog beamforming**

I. INTRODUCTION

Millimeter wave (mmW) unlicensed spectrum has been identified as a viable solution for realizing high data rate communications between connected vehicles [1]–[5]. The communication links are, however, characterized by high atmospheric absorption and hence can be operational only in short-range line-of-sight scenarios with highly directional beams realized through analog or digital beamforming at the transmitter/receiver. Digital beamforming allows for multiple simultaneous beams but is costly and complicated to implement since multiple phase and time-synchronized RF/mmW chains are required [6]. Analog beamforming is less costly since it involves a single beam at a time. But there is considerable overhead expended by the communication protocol and a long search/exploration time to scan the entire field of view and select the best beams for each mobile user (MU). This results in high latency and shorter service/exploitation time available for communication causing low throughput.

There have been several recent works that have applied multi-armed bandit (MAB) algorithms for reducing the exploration time of the best beams in order to increase the exploitation time for subsequent mmW communications [7]–[10]. MAB algorithms are a class of algorithms within the reinforcement learning framework which provides a basis for making decisions under uncertainty. MAB-based beam selection works, such as [8], [9], have relied on a strategy where the base station (BS) waits for the feedback (reward) over the uplink for the beam selection in subsequent time slots. In such time-slotted communication, the transmitter can switch the beam only once in a slot, and the duration of each slot depends on the time taken by the MU to process the downlink signal and share the feedback over the uplink.

The use of radar signals for detecting the presence of targets can potentially speed-up the beam search. However, there are certain challenges in the integration of radar with communication physical layer. The use of an auxiliary radar sensor for detecting MU, cannot be considered, as it would increase the cost and complexity of the system. Further, the radar and communication functionalities would have to be synchronized as well as managed for interference. Instead, we propose that an integrated sensing and communication system be utilized for mmW communication such as those proposed by [11], [12]. Here, a common waveform on a common spectrum is used for joint radar sensing and communications. Hence, no separate hardware/spectrum/synchronization or interference management is required to support both functionalities. Note that joint radar communication (JRC) systems have been explored over the last several decades to tackle spectral congestion issues [13]. While some works have studied how to manage mutual interference that arises from the coexistence of both systems on a common spectrum [14], others have exploited the communication signal as an opportunistic illumination for passive radar receivers [15]. We identify our work to belong to the third category of research that explores the collaborative design of JRC systems to improve the performance of each functionality [16], [17].

In this work, we propose incorporating a radar sensing mechanism into the MAB framework at the BS to overcome the limitations listed above. In the proposed framework, the radar at the BS is used to detect the presence of MU in the candidate beams based on the strength of the scattered signal (*amplitude gated radar enhanced MAB*) and the Doppler frequency shift (*Doppler gated radar enhanced MAB*) introduced to the radar signal. Only those beams that indicate the presence of a mobile radar target will be further scanned for the presence of a MU. Radar detection-based decision-making will be much faster than communication metric-based decision-making due to multiple reasons. First, the feedback for a radar signal is nearly instantaneous since it is based on the electromagnetic scattering of the signal by mobile targets. Second, the exploration time is substantially reduced by restricting the number of candidate beams that have to be scanned. Due to these factors, the overall exploration time will be reduced resulting in rapid beam alignment and improved overall communication throughput.

Notation: In our paper, scalar variables, vectors, and matrices are denoted with regular, and boldface lower and upper case characters respectively. Vector superscript T and symbol $\otimes$ denote transpose and convolution operations.

II. RAD-COM SIGNAL MODEL

We first present the signal models for the JRC transmitter and receiver based on the IEEE 802.11ad protocol where the Golay sequences in the communication frame are exploited for radar sensing [11], [12], [17]. The digital waveform $\mathbf{x}_q[m]$ corresponds to the Golay sequence in the qth packet transmitted at a pulse repetition interval of T_P with $m = 1, 2, \ldots, M$ samples. These digital packets are then converted into analog signals $\mathbf{x}_q(t)$ at the BS as follows:

$$\mathbf{x}_q(t) = \sum_{m=0}^{M} x_q[mT_s]\delta\left(t - mT_s - (q-1)T_P\right), \qquad (1)$$

where T_s is the sampling time. The signal is then amplified with energy E_s, convolved with a transmit shaping filter, $\mathbf{g}_T$, and then passed through analog upconversion to the mmW carrier frequency f_c as

$$\mathbf{x}_{q_{uc}}(t) = \sqrt{E_s}\left(\mathbf{x}_q(t) \otimes \mathbf{g}_T(t)\right) e^{+j2\pi f_c t}. \tag{2}$$

The upconverted signal is then transmitted via analog beamforming through a uniform linear array (ULA) of P_{BS} elements after applying complex antenna weight vector at BS transmitter (BS-TX), $\mathbf{w}_{BS_\theta} \in \mathcal{C}^{P_{BS} \times 1}$ for a given angle θ, resulting in

$$\mathbf{X}_{q_{uc}}(t) = \mathbf{w}_{BS_\theta}\mathbf{x_{q_{uc}}}^T(t). \tag{3}$$

Here, $\mathbf{w}_{BS_\theta} = \begin{bmatrix} 1 & e^{-jk_c d_{BS} \sin\theta} & \cdots & e^{-jk_c d_{BS}(P_{BS}-1)\sin\theta} \end{bmatrix}$ where k_c is the propagation constant and d_{BS} is the uniform element spacing. *It is important to note that the problem of searching for a new beam only arises for a MU that has changed its position and not a static user.*

Radar received signal: Along a pre-determined beam angle θ, we assume that there are B radar targets present in the channel including MU and other discrete clutter scatterers. Then the received signal at the P_{BS}-element ULA at the BS receiver (BS-RX) after being reflected from the targets, is

$$\hat{\hat{\mathbf{x}}}_q(t) = \sum_{b=1}^{B} \sigma_b \mathbf{w}_{BS_\theta} \mathbf{u}_\theta \mathbf{H_r}^2 \mathbf{u}_\theta^T \left[\mathbf{X}_{q_{uc}}(t - 2\tau_b)\right] + \rho(t), \tag{4}$$

where τ_b denotes the time delay caused by one-way propagation and σ_b is the strength of the reflection from each bth point target obtained from Frii's radar range equation. $\mathbf{H_r}^2$ is the $P_{BS} \times P_{BS}$ channel matrix that includes the direct path and multipath due to static clutter scatterers (SCS) present in the environment modeled in a manner described in [18] but for two-way propagation. The steering vector from the ULA corresponding to θ is given by $\mathbf{u}_\theta^T = \begin{bmatrix} 1 & e^{jk_c d_{BS} \sin\theta} & \cdots & e^{jk_c d_{BS}(P_{BS}-1)\sin\theta} \end{bmatrix}$ while ρ is the additive circular symmetric white Gaussian noise at the BS-RX. We assume that each radar target is moving with a constant radial velocity v_b, such that the Doppler shift is $f_b = 2v_b/\lambda$ where λ is the wavelength. After down-conversion and digitization, the received radar signal is

$$\hat{\hat{\mathbf{x}}}_q[m] = \sum_{b=1}^{B} \sigma_b \mathbf{w}_{BS} \mathbf{u}_\theta \mathbf{u}_\theta^T \mathbf{X}_{q_{uc}}\left[m - m_b\right] e^{-j2\pi f_b q T_P} + \rho, \tag{5}$$

where, m_b is the sample index corresponding to τ_b.

Radar signal processing: The radar received signal at BS-RX gathered over Q packets is first converted to a radar rectangle of dimension $[M \times Q]$ which then goes to the radar signal processing block to obtain the range and Doppler of the corresponding target. The range estimation output, $\chi_\mathbf{q}$, is obtained through the matched filtering for each qth packet, $\chi_\mathbf{q} = \hat{\hat{\mathbf{x}}}_q \otimes \mathbf{x}_q$. The output is processed through the ordered-statistics constant false alarm (OS-CFAR) to estimate each ath peak with amplitude, $\hat{\sigma}_a$, at the range $\hat{r}_a$. This $\hat{\sigma}_a$ information is used subsequently for the *amplitude gated radar-enhanced MAB algorithm* discussed in the next section. Next, Doppler estimation is carried out through one-dimensional multiple signal classification (MUSIC) for each ath peak across the Q packets to estimate the corresponding $\hat{f}_a$. This $\hat{f}_a$ information is used for the *Doppler gated radar-enhanced MAB algorithm* discussed in the next section.

Communication received signal and processing: The one-way propagated communication signal, $\hat{\mathbf{x}}(t)$, is received at the P_{MU} element ULA at the MU receiver (MU-RX) as shown in

$$\hat{\mathbf{x}}(t) = \mathbf{w}_{MU_\phi}\mathbf{u}_\phi\mathbf{H_c}\mathbf{u}_\theta^T\left[\mathbf{X}_{q_{uc}}(t - \tau_b)\right] + \delta(t). \tag{6}$$

Here, $\mathbf{w}_{MU_\phi}$ represents the weights applied at the MU-RX and $\mathbf{u}_\phi = \begin{bmatrix} 1 & e^{jk_c d_{MU}\sin\phi} & \cdots \end{bmatrix}$ is the steering vector for the BS at ϕ for d_{MU} antenna element spacing. H_c is the one-way propagation channel matrix $P_{BS} \times P_{MU}$ model [18] and δ is the additive circular-symmetric white Gaussian noise at MU-RX. The signal $\hat{\mathbf{x}}(t)$ is received by MU and gets processed and the corresponding signal-to-noise ratio (SNR) is sent back to the BS as uplink feedback. Note that the processing time for the MU results in a greater delay for the uplink signal to return to the BS compared to the nearly instantaneous radar-scattered signal. Second, the uplink $\hat{\mathbf{x}}$ is distinguished from $\hat{\hat{\mathbf{x}}}$ at the BS-RX through cross-correlation with $\mathbf{x}$. Due to the nature of the Golay sequence, the peak-to-sidelobe ratio after cross-correlation for $\hat{\mathbf{x}}$ is very high compared to the $\hat{\hat{\mathbf{x}}}$.

III. PROPOSED MAB FRAMEWORK FOR JRC

In this section, we set up the beam-selection problem between BS and MU as MAB and develop the algorithms that speed up the beam selection. The standard stochastic MAB consists of a set of $\mathcal{K}$ arms (predetermined beams) and a single player (the BS-TX/RX) as shown in Fig.1a. In each time slot, the BS-TX/RX, selects a single k^{th} beam and receives the reward - the SNR of the communication link, S_k, obtained from uplink feedback along the beam. For each arm, the reward is assumed to be drawn independently across time from distributions that are stationary and independent across arms. The performance metric is equal to the difference between the SNR of the optimal beam and the SNR over the selected beam. We define this as regret which is given as

$$R = TS_{k'} - \mathbb{E}\left[\sum_{k\in\mathcal{K}} S_k N_k\right] \tag{7}$$

where T is the total number of time slots, N_k is the number of times the beam k is selected by BS and $k' = \underset{k\in\mathcal{K}}{\arg\max}\, S_k$. The expectation here is with respect to the random number of pulls of the arms (N_k). Thus, the regret can be minimized by selecting the optimal beam, k', i.e., the beam with the highest SNR as many times as possible in a given horizon of size T. In this paper, we limit our discussion to the upper confidence bound-based (UCB) MAB algorithm [19] and provide regret bounds. The proposed idea can be easily extended to other MAB algorithms such as UCB variants and Thompson Sampling.

A. SNR Based Beam Selection using UCB

We first present the conventional approach for beam selection using the UCB algorithm given in Algorithm 1: $\mathbf{UCB}_{SNR}$. The communication is assumed to be time-slotted. At each tth time slot, the BS-TX transmits over the beam selected by the UCB algorithm. The algorithm selects each of the $\mathcal{K}$ beams once at the beginning (lines 4-6), and thereafter, the beam selection is based on the UCB index (lines 7-8). Here, T_{RSP} denotes the time slots required for RSP and it is set to 0 for $\mathbf{UCB}_{SNR}$. We denote the index of the selected beam and corresponding reward, i.e., instantaneous normalized SNR, as I_t and W_t, respectively. At the end of each time slot, the parameters are updated (line 12). The UCB index is calculated for each beam separately and it is given as

$$UCB_k(t) = \frac{\hat{S}_k}{N_k} + \sqrt{\frac{2\log(t)}{N_k}}, \tag{8}$$
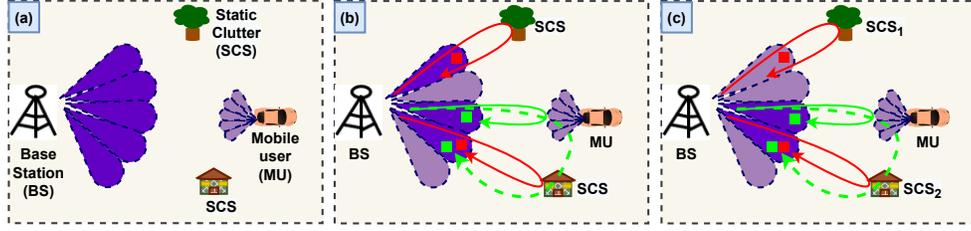
Fig. 1: System model showing (a) Standard non-radar based MAB beam selection, (b) Amplitude Gated Radar-Enhanced MAB, (c) Doppler Gated Radar-Enhanced MAB. Beams with red and green boxes indicate zero and non-zero Doppler targets respectively.

where $\hat{S}_k$ denotes the empirical mean of kth arm using samples obtained till time t. The expected regret of $\mathbf{UCB}_{SNR}$ scales as $\mathcal{O}(\sum_{k \in \mathcal{K} \setminus k'} \frac{\log T}{\Delta_k})$ [20] where $\Delta_k = S_{k'} - S_k$ for all $k \neq k'$. Furthermore, it suffers from high exploration time especially when $\mathcal{K}$ is large. Both these drawbacks limit the usefulness of $\mathbf{UCB}_{SNR}$ for mmW communication with a large number of narrow directional beams.

Algorithm 1 $\mathbf{UCB}_{SNR}$: SNR Based Beam Selection

1: **Input:** $\mathcal{K}, T, T_{RSP}$
2: **Initialize:** $N_k \leftarrow 0$ and $S_k \leftarrow 0$ for all k
3: **for** $t = T_{RSP} + 1, 2 \ldots T$, **do**
4:    **if** $t \leq \mathcal{K}$ **then**
5:       Select beam, $I_t = t$.
6:    **else**
7:       $\forall k \in [\mathcal{K}]$: compute $UCB_k(t)$ as given in Eq. (8)
8:       Select beam, $I_t = \arg\max_{k \in [K]} UCB_k(t)$
9:    **end if**
10:   BS-TX transmits a data frame over I_t
11:   MU observes instantaneous normalized SNR, W_t and communicate to BS-RX over the uplink.
12:   $N_{I_t} \leftarrow N_{I_t} + 1$ and $S_{I_t} \leftarrow S_{I_t} + W_t$.
13: **end for**

B. $\mathbf{UCB}_{SNR\_AG}$: Amplitude Gated Radar-Enhanced MAB

In this section, we augment the $\mathbf{UCB}_{SNR}$ with the proposed radar-based target detection as shown in Fig. 1b and described in Algorithm 2: $\mathbf{UCB}_{SNR\_AG}$. Here, a radar target is detected in a beam when the *amplitude/strength* of the scattered signal in any one or more of the range bins within the beam is above a pre-set threshold determined by CFAR. The number of beams where potential targets are detected is $\tilde{\mathcal{K}}$ (dark-colored beams in the figure) where $\mathcal{K} \geq \tilde{\mathcal{K}}$. Compared to Algorithm 1, the number of available beams is updated based on radar target detection during the first time slot (line 3 of Algorithm.2). Compared to $\mathbf{UCB}_{SNR}$, $\mathbf{UCB}_{SNR\_AG}$ potentially offers lower regret due to the following reasons: 1) Faster target detection: The identification of the presence of targets using radar is significantly faster since returns of the scattered signals from the short-range targets are nearly instantaneous with a short round-trip delay of the order of a few ns. For 5G, one slot is at least 4 ms assuming downlink sub-frame (1 ms), uplink sub-frame for reward feedback (1 ms), downlink (1 ms) and uplink data processing (1 ms). On average, RSP time, T_{RSP} for 10 radar packets is 36 ms, i.e., 9 slots are sufficient to find $\tilde{\mathcal{K}}$ [21]; 2) The proposed algorithm focuses on a subset of the total beams in which a mobile target may be present which in turn reduces the exploration time. To quantify this gain, let us fix a bandit instance. The set of beams is detected by the radar, $\tilde{\mathcal{K}}$, is a random variable depending on the distribution of scatterers. We can assume $\tilde{\mathcal{K}}$ includes the optimal arm in each realization as it has the maximum

---

Algorithm 2(or)3 $\mathbf{UCB}_{SNR\_AG}$ (or) $\mathbf{UCB}_{SNR\_DG}$

1: **Input:** $\mathcal{K}, T, T_{RSP}$
2: **Initialize:** $N_k \leftarrow 0$ and $S_k \leftarrow 0$ for all k
3: Find $\tilde{\mathcal{K}}$ using target detection (**Subroutine 1**)
                        $\triangleright$ $\mathbf{UCB}_{SNR\_AG}$
   Find $\tilde{\tilde{\mathcal{K}}}$ using Doppler detection (**Subroutine 2**)
                        $\triangleright$ $\mathbf{UCB}_{SNR\_DG}$
4: Run $\mathbf{UCB}_{SNR}$ with $\tilde{\mathcal{K}}$ arms for rest of the time horizon.
                        $\triangleright$ $\mathbf{UCB}_{SNR\_AG}$
   Run $\mathbf{UCB}_{SNR}$ with $\tilde{\tilde{\mathcal{K}}}$ arms for rest of the time horizon.
                        $\triangleright$ $\mathbf{UCB}_{SNR\_DG}$

Subroutine 1 AG : $\tilde{\mathcal{K}}$ Beams Selection Using Radar Target Detection

---

1: **Input:** $\mathcal{K}$
2: **Output:** $\tilde{\mathcal{K}}$
3: **for** $\theta = 1, 2 \ldots \mathcal{K}$, **do**
4:    χ_θ: Matched filtering across fast time samples
5:    CFAR detection:
6:    **if** $\chi_\theta \geq \gamma$ **then**
7:       Include beam, θ in subset $\tilde{\mathcal{K}}$
8:    **end if**
9: **end for**

---

signal strength and radar is unlikely to miss the MU. Hence the optimal arm is the same in any realized set $\tilde{\mathcal{K}}$. Expected regret over the set $\tilde{\mathcal{K}}$ is $\mathcal{O}(\sum_{k \in \tilde{\mathcal{K}} \setminus k'} \frac{\log T}{\Delta_k})$. Clearly this bound is smaller than $\mathcal{O}(\sum_{k \in \mathcal{K} \setminus k'} \frac{\log T}{\Delta_k})$ obtained for the previous case. Taking expectation over the random realizations $\tilde{\mathcal{K}}$, we get expect regret of $\mathbf{UCB}_{SNR\_AG}$ as $\mathbb{E}\left[\mathcal{O}(\sum_{k \in \tilde{\mathcal{K}} \setminus k'} \frac{\log T}{\Delta_k})\right] \leq \mathcal{O}(\sum_{k \in \mathcal{K} \setminus k'} \frac{\log T}{\Delta_k})$. Thus $\mathbf{UCB}_{SNR\_AG}$ is better than that of $\mathbf{UCB}_{SNR}$ resulting in an improvement in the performance.

C. $\mathbf{UCB}_{SNR\_DG}$: Doppler Gated Radar-Enhanced MAB

In the $\mathbf{UCB}_{SNR\_AG}$ algorithm, all the beams where radar targets are present are selected. However, some of the beams correspond to SCS as shown in Fig.1c. The SCS is distinguished to be of two types: some give rise to direct scattering at the radar (termed SCS_1) with zero Doppler while others give rise to Doppler-shifted returns at the radar through multipath with respect to the MU (termed SCS_2). The proposed Doppler-enhanced MAB algorithm is described in Algorithm 3: $\mathbf{UCB}_{SNR\_DG}$. Since direct path returns from SCS do not give any type of information regarding the MU, they can be excluded from the list of candidate beams based on the Doppler shift estimated from the radar signal processing described earlier. Hence the total number of beams where potential MU are detected is $\tilde{\tilde{\mathcal{K}}}$ and $\mathcal{K} \geq \tilde{\mathcal{K}} \geq \tilde{\tilde{\mathcal{K}}}$. Thus, fewer candidate beams (dark-colored beams in the figure) result in lower exploration time. The theoretical explanation for the reduction in beams follows the same logic provided for the previous algorithm and hence is not repeated here. Note

that static communication targets are not likely to have first triggered the necessity for the selection of a new beam by the BS and hence, static targets can be interpreted safely as SCS.

Subroutine 2 DG: $\tilde{\mathcal{K}}$ Beams Selection Based on Doppler Estimation

---

1: **Input:** $\mathcal{K}$
2: **Output:** $\tilde{K}$
3: **for** $\theta = 1, 2 \ldots \mathcal{K}$, **do**
4:     1D-MUSIC for $\hat{r}_a$ across Q packets
5:     **if** $\hat{f}_a \neq 0$ **then**
6:         Include beam, θ, in subset $\tilde{\mathcal{K}}$
7:     **end if**
8: **end for**

IV. PERFORMANCE ANALYSIS

We consider a three-dimensional (3D) Cartesian coordinate space with the ground plane defined by the x and y axes and the height axis along z. The BS is located at $[0, 0, 0]$ m with a y-aligned uniform linear array (ULA) of 32 antennas with an antenna spacing of $\lambda/2$ where λ is the wavelength corresponding to the center frequency f_c of 60 GHz. We adopt a 16QAM modulation and coding scheme with 512 OFDM subcarriers with a signal bandwidth of 1.76 GHz. We assume that the channel consists of a single MU and multiple SCS as shown in Fig. 1. The radar scattered returns of each of these are confined to a single beam. We assume that the MU is also with a 32-element ULA and is initially located at $[50, 20, 0]$m and subsequently moves with a constant velocity of v m/s along the x axis. Both MU and SCS are modeled as isotropic point scatterers and the SCS are distributed randomly across the 3D Cartesian space.

The throughput, Υ, is calculated as $\left(1 - \frac{\sum_{i=1}^{N_t} BER_i}{N_t}\right) \frac{D}{T_d}$ where, BER_i corresponds to the bit error rate of ith time slot, D and T_d correspond to the total number of bits and the time duration for each slot respectively. We benchmark the Υ performance of the proposed algorithms, $\mathbf{UCB}_{SNR\_AG}$ and $\mathbf{UCB}_{SNR\_DG}$ with conventional $\mathbf{UCB}_{SNR}$, lower upper confidence bound (LUCB) described in [22], digital beamforming and trivial random beam selection approach. We present the effects of the number of targets, number of beams, the Doppler velocity resolution, and radar receiver SNR on Υ. Each result presented in this section is obtained after averaging over 15 independent experiments and each experiment's duration/horizon is 2000 time slots.
Effect of Number of Radar Scatterers: In Fig. 2(a), we compare Υ of all the algorithms at different instants of the time horizon. Here, we assume one MU and vary the number of SCS. The Doppler velocity of the MU is fixed to 3 m/s and the angular resolution is 4° resulting in a total of 41 candidate beams spanning from -80° to 80° and the Doppler velocity resolution is 1 m/s. It can be observed in the figure that the proposed algorithms offer higher Υ than the benchmarked approaches - except for DBF - due to faster identification of the optimal beam. DBF provides the best-case results since all the beams are tested simultaneously. However, this approach is not pursued since the implementation of multiple synchronized receiver chains is costly and complex. In Fig. 2(b), we compare Υ of all algorithms at the end of the horizon as the number of SCS increases. We observe that the performance of all MAB-based approaches is significantly better than the random selection approach validating the need for a learning algorithm.
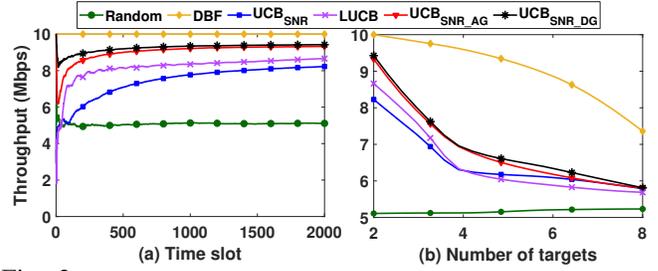

Fig. 2: (a) Throughput at different time slots for two targets, and (b) Throughput at the end of the horizon for different numbers of targets.

As expected, the difference between $\mathbf{UCB}_{SNR}$ (non-radar based beam selection) and $\mathbf{UCB}_{SNR\_AG}$ or $\mathbf{UCB}_{SNR\_DG}$ (radar-based beam selection) reduces as the number of targets increases. This is because when there are a large number of SCS, there is a proportionate increase in the number of candidate beams. Further, we can observe that there is a slight improvement in Υ for $\mathbf{UCB}_{SNR\_DG}$ than $\mathbf{UCB}_{SNR\_AG}$ as beams with zero Doppler get eliminated.
Effect of Angular Resolution Next, we fix the number of targets to three, which includes one MU and two SCS, and compare the performance for different angular resolutions or the number of candidate beams. The rest of the parameters are the same as Fig. 2. In Fig. 3(a), we compare the Υ
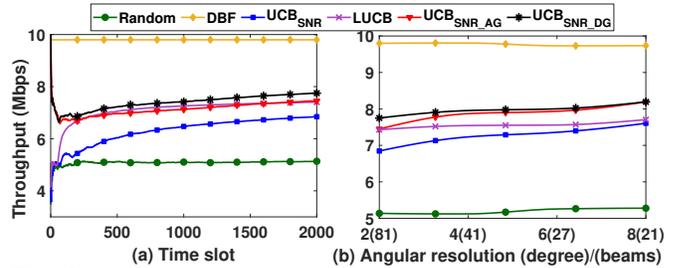

Fig. 3: (a) Throughput at different time slots for angular resolution of 2°, and (b) Throughput at the end of the horizon for different angular resolutions/number of candidate beams.

for angular resolution of 2° and it can be observed that the proposed approaches offer higher Υ than $\mathbf{UCB}_{SNR}$ and random approaches at all time slots. In Fig. 3(b), we compare the Υ at the end of the horizon for different numbers of candidate beams. As expected, Υ increases as the number of beams reduces due to lower exploration time. However, proposed $\mathbf{UCB}_{SNR\_AG}$ or $\mathbf{UCB}_{SNR\_DG}$ algorithms significantly outperform $\mathbf{UCB}_{SNR}$ and random approaches in all cases. The $\mathbf{UCB}_{SNR\_DG}$ offers higher Υ than $\mathbf{UCB}_{SNR\_AG}$ for smaller angular resolution (i.e., large $\mathcal{K}$) due to more opportunities of beam elimination. However, the impact of Doppler processing reduces with the increase in angular resolution due to the occurrence of fewer beams.
Effect of Velocity Resolution: Next, we study the effect of Doppler velocity resolution on Υ. The simulation is carried out for Doppler velocity resolution of 1 m/s, 2 m/s, 3 m/s, 4 m/s and 5 m/s. We assume the MU with a Doppler velocity of 3 m/s and three targets (one MU and two SCS). In Fig. 4(a), we consider the velocity resolution of 5 m/s, which is not sufficient to detect the MU and hence, the performance of $\mathbf{UCB}_{SNR\_DG}$ is poorer than that of $\mathbf{UCB}_{SNR\_AG}$. As shown in Fig. 4(b), the velocity resolution of radar signal processing should be carefully chosen. Thus, $\mathbf{UCB}_{SNR\_DG}$ may not offer better performance than $\mathbf{UCB}_{SNR\_AG}$ even though it incurs higher computational cost due to Doppler estimation. As expected, Υ of $\mathbf{UCB}_{SNR}$ and random approaches are independent of the velocity resolution since they do not use this information for beam selection.
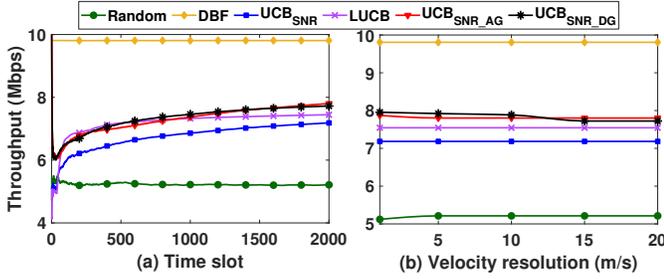
Fig. 4: (a) Throughput at different time slots for velocity resolution of 15m/s, and (b) Throughput at the end of the horizon for different velocity resolutions.

Effect of Radar Receiver SNR: Further, we analyze all the algorithms for four different SNRs at the radar receiver ranging from -10 dB to 10 dB. Here, also we simulate three targets (one MU and two SCS), and the velocity resolution is set as 1 m/s. Fig. 5(a) represents the viewgraph of Υ with respect to
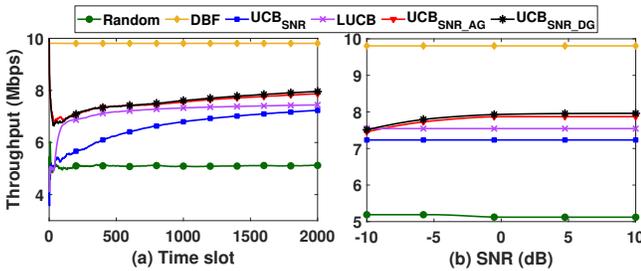

Fig. 5: (a) Throughput at different time slots for 10dB SNR, and (b) Throughput at the end of the horizon for different SNRs.

time slots for a high SNR (10 dB). We observe in Fig. 5(b) that the regret improves for the $\mathbf{UCB}_{SNR\_AG}$ and $\mathbf{UCB}_{SNR\_DG}$ with increase in SNR. At lower SNR, the poor prediction of the target presence amidst noise results in higher regret. Note that since we consider the SNR with respect to the radar receiver, change in SNR does not have any impact on the random beam selection as well as $\mathbf{UCB}_{SNR}$ algorithms.

Impact of Doppler Processing: Further, we discuss the scenario where the number of SCS_1 is greater than SCS_2 such that the candidate beams selected by the $\mathbf{UCB}_{SNR\_DG}$ are significantly lesser than those selected by the $\mathbf{UCB}_{SNR\_AG}$. In Fig. 6(a), we consider one MU, one SCS_2 and two SCS_1. The angular resolution is 4° and the velocity resolution is 1 m/s. Since SCS_2 gives rise to
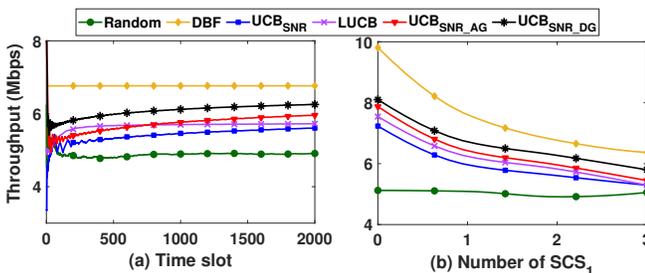

Fig. 6: (a) Throughput at different time slots for four targets (including two SCS_1, one SCS_2 and one MU, and (b) Throughput at the end of the horizon for different number of SCS_1.

Doppler shift at the BS through multipath from the MU, the corresponding beam is retained for further scanning while the beams corresponding to the two SCS_1 that give rise to zero-Doppler shift are excluded in $\mathbf{UCB}_{SNR\_DG}$. This results in better performance of the algorithm compared to $\mathbf{UCB}_{SNR\_AG}$. Further, the improvement is greater as the

number of SCS_1 increases for a fixed number of SCS_2 as seen in Fig. 6(b).

V. Conclusion

In this work, we demonstrate how radar-enhanced MAB within a JRC BS can substantially reduce the exploration time by selecting only those candidate beams that detect the presence of radar targets of which the MU may be one. Further reduction in the exploration time is realized by distinguishing SCS from MU through radar-based Doppler estimation. Simulation results demonstrate an overall improvement in the communication link metrics with the reduction in the exploration time through radar-enhanced MAB when compared with conventional MAB algorithms.

References

[1] "Federal Communications Commission," https://www.fcc.gov/general/unlicensed-operations-57-64-ghz-band, accessed: 2023-05-02.
[2] "Innovation, Science and Economic Development Canada (ISED)," https://www.ic.gc.ca/eic/site/smt-gst.nsf/eng/sf11298.html, accessed: 2023-05-02.
[3] "European Conference of Postal and Telecommunications Administrations (CEPT)," https://www.cept.org/ecc/topics/radio-spectrum-matters, accessed: 2023-05-02.
[4] "Ministry of Internal Affairs and Communications (MIC)," https://www.tele.soumu.go.jp/e/data/alloc/freq/index.html, accessed: 2023-05-02.
[5] "Ministry of Science and ICT (MSIT) ," https://www.msit.go.kr/web/msipContents/contents.do?mId=MTY5, accessed: 2023-05-02.
[6] B. Yang, Z. Yu *et al.*, "Digital beamforming-based massive mimo transceiver for 5g millimeter-wave communications," *IEEE Trans. on Microwave Theory and Techniques*, vol. 66, no. 7, pp. 3403–3418, 2018.
[7] J. Zhang, Y. Huang, Y. Zhou, and X. You, "Beam alignment and tracking for millimeter wave communications via bandit learning," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5519–5533, 2020.
[8] M. B. Booth *et al.*, "Multi-armed bandit beam alignment and tracking for mobile millimeter wave communications," *IEEE Communications Letters*, vol. 23, no. 7, pp. 1244–1248, 2019.
[9] I. Aykin, B. Akgun *et al.*, "Mamba: A multi-armed bandit framework for beam tracking in millimeter-wave systems," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 2020, pp. 1469–1478.
[10] I. Chafaa, E. V. Belmega, and M. Debbah, "Adversarial multi-armed bandit for mmwave beam alignment with one-bit feedback," in *Proceedings of the 12th EAI International Conference on performance evaluation methodologies and tools*, 2019, pp. 23–30.
[11] E. Grossi, M. Lops, L. Venturino, and A. Zappone, "Opportunistic radar in IEEE 802.11 ad networks," *IEEE Transactions on Signal Processing*, vol. 66, no. 9, pp. 2441–2454, 2018.
[12] P. Kumari, J. Choi *et al.*, "IEEE 802.11ad-based radar: An approach to joint vehicular communication-radar system," *IEEE Transactions on Vehicular Technology*, vol. 67, pp. 3012–3027, Apr. 2018.
[13] F. Liu, C. Masouros *et al.*, "Joint radar and communication design: Applications, state-of-the-art, and the road ahead," *IEEE Transactions on Communications*, vol. 68, no. 6, pp. 3834–3862, 2020.
[14] H. Deng and B. Himed, "Interference mitigation processing for spectrum-sharing between radar and wireless communications systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 49, no. 3, pp. 1911–1919, 2013.
[15] Y. Zeng *et al.*, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 36–42, 2016.
[16] A. Hassanien, M. G. Amin *et al.*, "Dual-function radar communication systems: A solution to the spectrum congestion problem," *IEEE Signal Processing Magazine*, vol. 36, no. 5, pp. 115–126, 2019.
[17] G. Duggal, S. Vishwakarma, K. V. Mishra, and S. S. Ram, "Doppler-resilient 802.11ad-based ultra-short range automotive joint radar-communications system," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 56, pp. 4035–4048, 2020.
[18] R. W. Heath, N. Gonzalez-Prelcic *et al.*, "An overview of signal processing techniques for millimeter wave mimo systems," *IEEE journal of selected topics in signal processing*, vol. 10, no. 3, pp. 436–453, 2016.
[19] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, no. 2, pp. 235–256, 2002.
[20] S. Bubeck, N. Cesa-Bianchi *et al.*, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Foundations and Trends® in Machine Learning*, vol. 5, no. 1, pp. 1–122, 2012.
[21] Y. Choi, D. Jeong *et al.*, "Fpga implementation of the range-doppler algorithm for real-time synthetic aperture radar imaging," *Electronics*, vol. 10, no. 17, 2021.
[22] S. Kalyanakrishnan, A. Tewari, P. Auer, and P. Stone, "Pac subset selection in stochastic multi-armed bandits." in *ICML*, vol. 12, 2012, pp. 655–662.