Profiling Vehicles for Improved Small Cell Beam-Vehicle Pairing Using Multi-Armed Bandit

Abdulkadir Kose*, Chuan Heng Foh*, Haeyoung Lee*, Klaus Moessner*†

*5GIC & 6GIC, Institute for Communication Systems (ICS), University of Surrey, Guildford, UK
†Faculty of Electronics and information Technology, University of Technology Chemnitz, Germany
Email: {a.kose, c.foh, Haeyoung.Lee}@surrey.ac.uk, klaus.moessner@etit.tu-chemnitz.de

Abstract—The 5G technology has tapped into millimeter wave (mmWave) spectrum to create additional bandwidth for improved network capacity. The use of mmWave for specific applications including vehicular networks has widely discussed. However, applying mmWave to vehicular networks faces challenges of high mobility nodes and narrow coverage along the mmWave beams. In this paper, we focus on a mmWave small cell base station deployed in a city area to support vehicular network application. We propose profiling vehicle mobility for a machine learning agent to learn the performance of serving vehicles with different mobility profiles and utilize the past experiences to select appropriate mmWave beam to service a vehicle. Our machine learning agent is based on multi-armed bandit learning model, where classical multi-armed bandit and contextual multi-armed bandit are used. Particularly for the contextual multi-armed bandit, the contexts are vehicle mobility information. We show that the local street layout has naturally constrained vehicle movement creating distinct mobility information for vehicles, and the vehicle mobility information is highly related to communication performance. By using vehicle mobility information, the machine learning agent is able to identify vehicles that can remain within a beam for longer time period to avoid frequent handovers.

Index Terms—Beam Handover, mmWave Networks, 5G.

I. INTRODUCTION

Networks nowadays are confronted with higher traffic demands, which requires various advancement to current cellular networks [1]. Increasing capacity demand is considered to be met by various enhancements, such as network densification, massive MIMO and beamforming techniques, utilizing higher frequency bands and carrier aggregation [1], [2].

Network densification through the dense deployment of small cells is one of the approaches to meet the increasing capacity demand [3]. Small cells are initially intended to be utilized in hotspot areas to support high data rates. Due to intensive traffic demand, small cells are also deployed under the macro layer creating Heterogeneous Network (HetNet) to offload macro cells.

Limited by spectrum availability and high data rate support in the current sub-6GHz microwave bands, millimetre wave (mmWave) small cells have become one of the promising candidates for 5G systems. In contrast, users often suffer from high propagation loss at higher frequencies ($\gg$ 6GHz) [4]. As a remedy, directional transmission, i.e. mmWave beam, is adopted in mmWave but it introduces new challenges to support some important 5G applications such as vehicle-to-everything (V2X) communication in vehicular network applications. The narrow coverage of mmWave beam along the beam and fast moving vehicles can lead to a short period of sojourn time for a vehicle within a beam which in turns causes frequent handover and high overhead [5]. Traditional approach of choosing a beam that can achieve the strongest SNR may not always result in selecting the vehicle with the longest sojourn time, due to the local street layout, a vehicle with the the strongest SNR may move out of the beam quickly. Thus a different strategy is needed for the beam selection in vehicular network applications.

To harness the benefit of artificial intelligence, multi-armed bandit (MAB) machine learning technique has been considered in beam selection for mmWave systems [6], [7]. The goal of MAB learner is to learn the environment and apply past experience to make decision. There is also an associate question on how much effort of learning is needed to achieve efficient exploitation the system and minimize the regret. The study of balance between exploration and exploitation is particularly important when the resources in terms of time or number of actions for exploration and exploitation are limited. In [6], the adaptive beam selection is modelled as a contextual multi-armed bandit (C-MAB) problem. The traffic pattern, precisely the travelling direction of a vehicle, is included in the contextual information in their proposed contextual online learning algorithm. A mmWave BS, a learning agent, autonomously learns the relationship of beam selection and the data rate performance for given context information (traffic pattern and permanent/temporary blockage). Since this algorithm learns the expected beam performance in different contexts over time, it actually does not require a training phase and is highlighted as a fast learning algorithm.

In [7], the broadcasting clustering possibility of neighbouring vehicles is additionally considered and a two-layer MAB algorithm is proposed. This study assumes that mmWave broadcasting would be useful when multiple vehicles want to download a popular same contents (e.g. movies) of high data rate. First, similar to [6], a mmWave BS learns the correlation between the beam allocation and the achievable throughput and then adapts to the dynamic blockages and traffic patterns without any prior knowledge of its surrounding blockage. Then, the mmWave BS learns the appropriate beams to cover multiple vehicles requiring to download the same contents throughput broadcasting in the cells and the best broadcast angle along these beams. The work emphasizes use of social

preference information of vehicles to enhance learning and decision making for broadcasting.

While we see several past attempts to introduce C-MAB in mmWave V2X application for beam selection decision making, the use of contexts is limited. We notice that the contexts describing vehicle mobility features also implicitly encapsulates the information of local street layout. This is because the local street layout constrains the vehicle mobility, vehicle mobility information can be used to predict a vehicle movement and indicate communication performance. In this paper, we demonstrate the application of C-MAB to explicitly use sufficient vehicle mobility context in order to indirectly achieve learning of vehicle mobility behaviour on the surrounding local street and improve the V2X communication.

Different from other works, we focus on using C-MAB to select the appropriate vehicle context in a realistic simulation setup. We first apply classical MAB to show the benefit of learning from past experiences. We then extend the model by profiling the mobility of vehicles and use the profiles as contexts in C-MAB. We show that C-MAB further improves the performance of MAB in terms of beam sojourn time by 50%. The remainder of the paper is organized as follows. Section II describes the considered scenario and formulation of beam allocation problem. In Section III, the proposed beam selection algorithm based on context-aware multi-arm bandit is elaborated. The performance validation are explained in Section IV to show the effectiveness of our proposed algorithm. Finally, we draw important conclusions in Section V.

II. Scenario Setup and Problem Formulation

We consider a mmWave small cell deployed to offload the V2X data traffic from an existing macro base station. For cost effective consideration, the small cell implements most basic functions, some features such as beam steering will not be available. Similar to [7], we assume that the small cell base station has a number of antennas pointing at different fixed directions, and it has a lower number of RF chains installed than the number of antennas. Each antenna may consist of a set of antenna elements creating a beam at a direction. As defined in [7], the coverage of a beam is called a beam section, and ideally, they do not overlap with each other. Due to the limited number of RF chains, only subset of antennas can be activated at one time. We focus on using the mmWave small cell for downlink transmissions.

Fig. 1 illustrates our scenario setting. In this scenario, the mmWave small cell is deployed around Guildford town center in UK. It has six antennas, each faces a different direction to cover the entire surrounding. With two RF chains, at one time it can only activate two beams to serve two vehicles. The shaded areas in the figure demonstrates two active beams serving two vehicles. The shaded areas are only the illustrative coverage, the actual coverage of service area depends on the channel model and antenna settings.

We consider 3GPP Band n257 operating at 28 GHz with 50 Mbps of bandwidth setting. We use pathloss model for

Parameter	Value
Carrier frequency, f_c	28 GHz
Channel bandwidth	50 MHz
Transmit power	30 dBm for BS, 20 dBm for vehicle
Transmitter antenna gain	14 dB
Thermal noise	-97.2 dBm (for 50 MHz bandwidth)
Noise figure	4 dB at BS, 7 dB at vehicle
SNR threshold	-5 dB

urban micro (UMi) given by [8]

$$PL(d) = 32.4 + 20\log(f_c) + 20\log(d) + X_g \quad (1)$$

where f_c is the carrier frequency in GHz, d is the distance between the transmitter and receiver antennas in meters, and X_g describes the channel fading. The channel fading setting is given by Table I in [8], however we do not include channel fading. We assume antenna height difference of 5m between the base station and vehicle, thus with a distance $\hat{d}$ between two nodes, $d = \sqrt{\hat{d}^2 + 5^2}$.

We follow the mmWave beamforming model used in [7]. In the study, the beamforming gain G_{BF} of the antenna is calculated based on [9] by

$$G_{BF}(\Delta\theta) = \frac{2\pi}{B_{3dB}} 10^{-0.1\eta\left(\frac{\Delta\theta}{B_{3dB}}\right)^2} \quad (2)$$

where B_{3dB} is the beamwidth of 3dB of the antenna, $\Delta\theta$ is the off-center angle which measures the angle between the beam center direction and its pointing direction to the serving vehicle within its beam sector, and η is a constant carrying a value of 12. For 6-sector setting, B_{3dB} is set to $35°$ [10] or 0.61 radian.

While the base station has antennas with fixed pointing direction, we assume that vehicles have a steerable beam antenna that can track and steer towards the base station during the communication with the base station. With the settings described above, the SNR of a serving vehicle can be computed by

$$SNR = p_0 - PL(d) + G_{tx} + G_{BF}(\Delta\theta) + G_{rx} - N \quad (3)$$

where G_tx and G_rx are the transmitter and receiver antenna gains respectively, N is the noise including thermal noise and the receiver noise figure. Since vehicles tracks and steers their beams to the base station while receiving a transmission, the receiver antenna gain also include beamforming gain, and here we set to the beamforming gain. The parameters used in the simulation is given in Table I.

Using mmWave small cell for V2X application faces a unique challenge that the small cell must use a narrow beam to serve fast moving vehicles, and often the sojourn time of a moving vehicle within a narrow beam is short, leading to frequent handover and high overhead. Due to the user mobility, channel state changes rapidly, and thus a conservative approach of using the most robust modulation scheme rather than an adaptive modulation scheme may be more suitable.
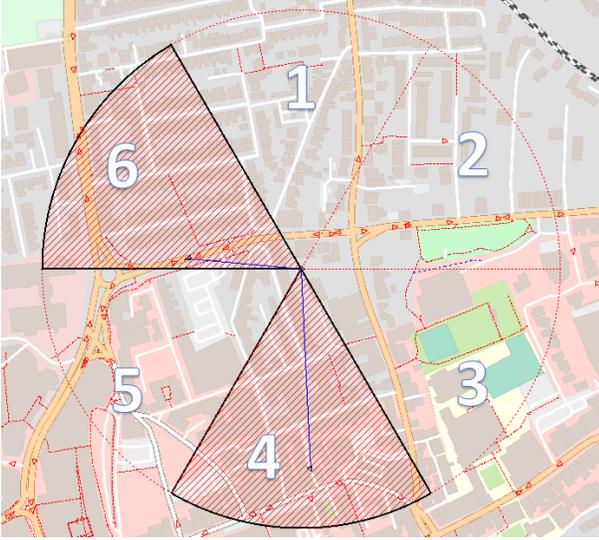
Fig. 1. A scenario of beam selection (among 6 beams) for Guildford town center, UK.

As a result, to maximize the data transmission between the small cell and its serving vehicles, the radio resource allocation strategy for this setup shall aim at serving vehicles that have the longest possible sojourn time within a beam.

For reliability consideration, we assume that adaptive modulation is not used. Instead, the base station uses the most robust modulation scheme for communication regardless of the reporting SNR. Let R be the corresponding fixed data rate for all beams when serving a vehicle resides within its beam sector. Vehicles moving out of the beam service area shall achieve zero rate. Let $\mathcal{B}$ be a set of beams of the mmWave small cell base station, n be the maximum number of active beams can be used by the base station at one time, and $\beta_i(t)$ be the transmission rate of beam i at time t. Thus we have $\beta_i(t) = R$ if beam i is serving a vehicle at time t, otherwise $\beta_i(t) = 0$ if either it is performing handover procedure and waiting the data for the vehicle to be rerouted from the macro base station, or it is inactive, or there is no vehicle within its beam sector to serve. The radio resource allocation strategy is thus to maximize the overall data transmission from all beams given a certain time period T, which is

$$\max \left(\sum_{i \in \mathcal{B}} \int_0^T \beta_i(t)dt \right). \tag{4}$$

As each handover event incurs blackout time resulting in $\beta_i(t) = 0$ during the event, maximizing the above can also be achieved by minimizing the number of handover events since the data rate remains constant during a service. In other words, the vehicle sojourn time within a beam should be kept as long as possible to minimize the handover frequency.

The small cell base station attempts to maximize its service by utilizing all active beams whenever possible. When a serving vehicle has left the beam sector, the beam serving the vehicle turns inactive, and the base station can activate a beam from all available beams to serve another vehicle. The base station may unbiasedly pick an available beam to serve a vehicle in its beam sector. It may also select the vehicle with the highest SNR to pair with the available beam for service. As we shall show later the traditional approach of selecting the highest SNR may not work well for mmWave vehicular networks due to the narrow beamwidth.

Since the mobility of vehicles is constrained to the local street layout, it is beneficial to utilize the information of local street layout as part of the consideration for selecting a vehicle to serve. However, acquiring and processing local street layout for this purpose can be tedious, we propose using mobility information of vehicles which indirectly captures the local street layout instead. To achieve that, we profile vehicles based on their mobility information and use the profile as a context to pair a vehicle with a beam. Mobility information to be used for the context may include the speed, orientation, location and distance away from the small cell base station derived from the location or measured through timing advance.

III. PROPOSED MULTI-ARMED BANDIT LEARNING DESIGN

We apply both classical MAB and C-MAB models to achieve learning of the best outcome to service vehicles. Given that our objective is to select vehicles with the longest sojourn time within a beam, the reward is the connection duration experienced by a serviced vehicle. For C-MAB, mobility context of the vehicle is also used.

While MAB is an online learner, it requires initial learning phase for effective exploitation. In our application, the setting of learning phase is not particularly critical since the layout of the surrounding street will remain unchanged for a long time. Once learned, the knowledge remains valid for a long time. Thus as a new MAB learner is introduced to a new environment, the MAB learner should concentrate on learning the environment. With sufficient learning, the learner can then proceed to exploit the learned knowledge while revising its knowledge based on the new findings during exploitation. Thus we choose Epsilon-First learning strategy [11] for both the proposed MAB models.

A. Classical MAB Learning Algorithm

In our classical MAB learning, the arms in MAB are $\mathcal{B}$, which is a set containing all mmWave beams in the base station. During the learning phase, the MAB learner performs full exploration of the beam selection and learn the outcome via the rewards [12]. Whenever a beam has ended its service, the base station randomly pick an available beam, say b and a vehicle within the beam sector to serve the vehicle. The beam continues to serve the vehicle until the vehicle has moved out of the beam sector, then the connection time is recorded and provided to MAB learner as the reward of selecting beam b, r_b. The reward for each beam is recorded in a set $\mathcal{R} = \{r_b | b \in \mathcal{B}\}$.

With Epsilon-First strategy, the learning phase has a pre-determined period [11]. After the learning phase has ended, the base station switches to exploitation phase. In this phase,

whenever a beam is available for service, the base station greedily chooses the beam that has the best reward on average so far. If there is no vehicle in its beam sector, then the beam with the next best reward is sought, and the process repeats until a beam is chosen. In the case that no vehicle is found in all available beams, no additional beam will be activated until a vehicle appears in an available beam sector. Algorithm 1 illustrates the algorithm of the MAB model.

Algorithm 1 MAB for Beam and Vehicle Selection

    Input (for exploration and exploitation):
        A set of available beams, $\mathbf{B}$.
    Output (for exploration and exploitation):
        Selected (beam,vehicle) pair, or None.

1: **procedure** EXPLORATION
2:     **while** $\mathbf{B} \neq \emptyset$ **do**
3:         Randomly select a beam b from $\mathbf{B}$
4:         **if** No vehicle is available in beam b **then**
5:             $\mathbf{B} \leftarrow \mathbf{B} \setminus \{b\}$
6:         **else**
7:             Randomly select a vehicle v in b
8:             return (b, v)
9:         **end if**
10:     **end while**
11:     return None
12: **end procedure**
13:
14: **procedure** EXPLOITATION
15:     **while** $\mathbf{B} \neq \emptyset$ **do**
16:         $b \leftarrow \arg\max\{r_b | b \in \mathbf{B}\}$
17:         **if** No vehicle is available in beam b **then**
18:             $\mathbf{B} \leftarrow \mathbf{B} \setminus \{b\}$
19:         **else**
20:             Randomly select a vehicle v in b
21:             return (b, v)
22:         **end if**
23:     **end while**
24:     return None
25: **end procedure**
26:
27: **procedure** UPDATEREWARD(beam b, reward r)
28:     $r_b \leftarrow \frac{r_b \cdot k_b + r}{k_b + 1}$ ▷ r_b is set to 0 initially
29:     $k_b \leftarrow k_b + 1$ ▷ k_b is set to 0 initially
30: **end procedure**
31:

B. Contextual MAB Learning Algorithm

Contextual MAB extends classical MAB by including contexts when making decision. In our design, we use mobility information as the contexts. We profile each vehicle based on its mobility information. While the arms of the C-MAB are $\mathcal{B}$, contexts are inspected to decide which beam is used to service a vehicle with a specific mobility profile. Let $\mathcal{C}$ be a set containing all contexts, and $\langle b, c \rangle$ be an ordered pair of beam b and context c. The reward is recorded for all combinations of beams and contexts. We denote $\mathcal{R}_{\mathcal{C}} = \{r_{\langle b,c \rangle} | b \in \mathcal{B}, c \in \mathcal{C}\}$ to be the set containing all rewards.

When a beam becomes available, the base station checks the context c of each vehicle in each available beam b, and collect the corresponding past average reward $r_{\langle b,c \rangle}$. The collected corresponding rewards are ranked, and the vehicle associated with the profile carrying the highest reward is selected. In case that multiple vehicles are associated with the selected profile, one of those vehicles is selected randomly for service. Once the decision is made, the vehicle is immediately scheduled for service. The service continues until the vehicle has moved out of the beam sector. Then the beam reports the connection time as the reward for its selection. The algorithm for C-MAB is presented in Algorithm 2. We omit the exploration procedure as it is the same as that of MAB in Algorithm 1.

The mobility information to be used for contexts may include features such as speed, orientation and location information. The choice of contexts is flexible and dependent on local street layout. Including more features allows more precise separation among vehicles with different performance characteristics, but it introduces a longer learning period to capture the performance characteristics. Thus, selection of features should be prioritized to those features that can significantly influence the performances.

Algorithm 2 Contextual MAB for Beam and Vehicle Selection

    Input (for exploitation):
        A set of available beams, $\mathbf{B}$.
    Output (for exploitation):
        Selected (beam,vehicle) pair, or None.

1: **procedure** EXPLORATION
2:     //same with MAB in Algorithm 1
3: **end procedure**
4: **procedure** EXPLOITATION
5:     $\mathbf{R_C} \leftarrow \{r_{\langle b,c \rangle} | b \in \mathbf{B}, c \in \mathcal{C}\}$
6:     **while** $\mathbf{R_C} \neq \emptyset$ **do**
7:         $\langle b, c \rangle \leftarrow \arg\max \mathbf{R_C}$
8:         **if** No vehicle is associated to c in beam b **then**
9:             $\mathbf{R_C} \leftarrow \mathbf{R_C} \setminus \{r_{\langle b,c \rangle}\}$
10:         **else**
11:             Randomly select a vehicle v in b with c
12:             return (b, v)
13:         **end if**
14:     **end while**
15:     return None
16: **end procedure**
17:
18: **procedure** UPDATEREWARD(beam b, context c, reward r)
19:     $r_{\langle b,c \rangle} \leftarrow \frac{r_{\langle b,c \rangle} \cdot k_{\langle b,c \rangle} + r}{k_{\langle b,c \rangle} + 1}$ ▷ $r_{\langle b,c \rangle}$ is set to 0 initially
20:     $k_{\langle b,c \rangle} \leftarrow k_{\langle b,c \rangle} + 1$ ▷ $k_{\langle b,c \rangle}$ is set to 0 initially
21: **end procedure**
22:

IV. RESULT DISCUSSION

In the section, we present our findings on applying MAB and C-MAB machine learning models for small cell beam selection. The scenario for our experiment is presented in Fig. 1. In our scenario, we focus on a single mmWave small cell base station with 6 beams. At one time, the base station can only activate 2 beams for service. Vehicles are created and absorbed at certain locations on the map. Those locations are the main entrance point to the city and the exit point from the city. Besides, we also included several main parking spots in the city as vehicle creation and absorption locations. A pair of vehicle creation and absorption locations is used to create a route simulating a vehicle either passing through the city, entering to or exiting from the city. We use A-STAR path finding algorithm [13] to establish the route for vehicles. The speed of each vehicle is set randomly between 30 to 50 km/h (considering speed limit around city area). We assume that each of these vehicles requires downlink data service when entering the small cell.

We use our own developed Python Mobility Simulation Platform (PyMoSim[1]) for the simulation. In our simulation, there are over a hundred of vehicles continuously moving on the map. We simulate 10 hours of operation where the base station begins with full exploration for learning, and then it switches to full exploitation after 3 hours which is the first 30% of the simulation time. As Epsilon-First strategy stops triggering exploration after the learning phase, this enables us to focus on the study of learning effectiveness acquired during the learning phase.

We compare our MAB and C-MAB algorithms with traditional best SNR and also random beam selection schemes. In best SNR, the base station greedily selects a vehicle that reports the highest SNR. For our C-MAB, we use orientation and distance from the small cell to form a context. For the orientation, we profile a vehicle into one of the four directions of movement (north-east, north-west, south-east, south-west), and for the distance, we propose the base station to use timing advance to profile a vehicle into one of the three ranges (near, middle, far) with approximately same length for each range. We do not use speed as a feature since we found its impact on the performance is low as the speed range of vehicles is narrow due to the low speed limits around the city area.

For the performance measure, we focus on the beam sojourn time of connections. Fig. 2 plots the mean beam sojourn time of connections of the random selection scheme, best SNR scheme, and our proposed MAB and C-MAB algorithms over the course of the simulation. At the beginning, both MAB and C-MAB perform similarly to the random selection scheme since both learners are operating in full exploration in the first 30% of the simulation duration. Upon switching to exploitation, the mean beam sojourn time of both MAB and C-MAB algorithms jumps as both learners begin applying the learned knowledge to make selection decision. Since C-
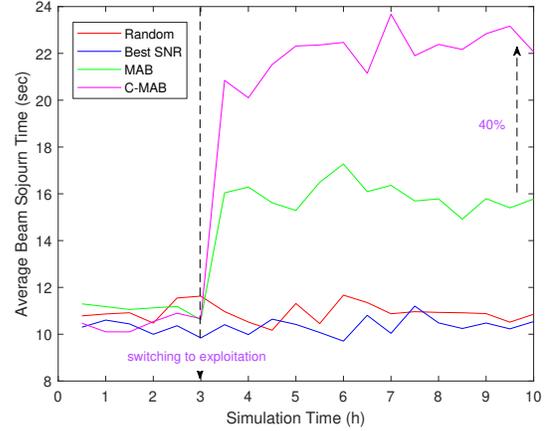
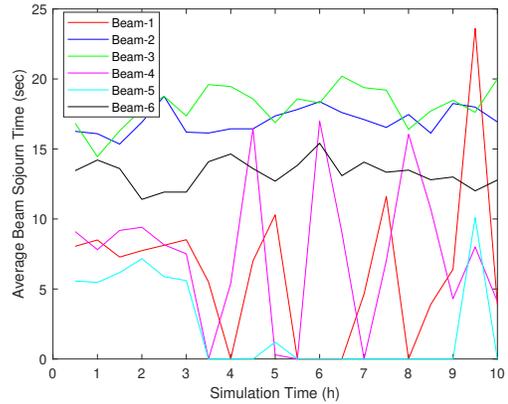Fig. 2. Mean beam sojourn time from different beam selection schemes

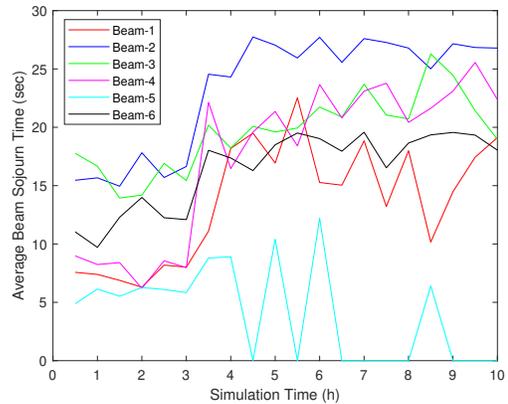Fig. 3. Mean beam sojourn time of six beams with MAB

Fig. 4. Mean beam sojourn time of six beams with C-MAB

MAB algorithm is able to select a vehicle based on contexts, it doubles the duration of mean beam sojourn time when comparing to that of the random selection scheme. It also outperforms classical MAB by about 40%.

Interestingly, we see that the performance of best SNR records the lowest among all. This is mainly due to the unfavorable street layout for best SNR scheme. As best SNR scheme designs to select vehicles closer to the base station for stronger signals. Based on the street layout in Fig. 1, vehicles in Beam-1 and Beam-6 have a high chance to be selected. However, most of these vehicles move across the beam quickly and thus the beam sojourn time is short. This also shows the importance of knowing the local street layout for base station to operate efficiently in V2X application.

To understand the benefit of profiling vehicles, we show the breakdown of the average beam sojourn time for each beam in Figs. 3-4. From Fig. 3 which presents results using MAB, focusing on Beam-1, we see that it performs poorly during the learning phase. Thus in the exploitation phase, the base station avoids utilizing it where the results showing zero indicate no utilization. The poor performance from Beam-1 can be easily understood by inspecting the local street layout presented in Fig. 1. We see that most vehicles moving on the main road in the east-west movement only remain briefly in Beam-1, and thus the sojourn time is very short. Some vehicles moving on the other main road in Beam-1 in north-south movement may have longer sojourn time, but since MAB does not differentiate the two, the resultant average sojourn time remains low.

However for C-MAB, since contexts are explicitly used, C-MAB can differentiate between east-west movement and north-south movement, and hence Beam-1 is efficiently utilized in C-MAB as seen in Fig. 4. Unlike MAB that eventually focuses on specific beams, C-MAB is able to select vehicles experiencing long sojourn time in the past in all beams, and thus C-MAB utilizes a wider range of beams than MAB.

V. Conclusions

In this paper, we developed machine learning algorithms based on MAB models to improve the performance of mmWave small cell communication for vehicular network applications. Due to the narrow coverage along the mmWave beams and fast moving vehicles, connections experienced frequent handover as the sojourn time of a fast moving vehicle in a narrow mmWave beam is often short. Since the movement of vehicles were constrained on the roads, depending on the local street layout, some vehicles remained in the same beam for much longer time than others.

Utilizing the vehicle mobility information, we proposed applying MAB learning models for the machine learning agent to learn from the past experiences when making decision. We applied classical MAB and C-MAB. For the C-MAB, we explicitly use mobility information as the contexts. With our scenario of a small cell deployed at a city center, we found that vehicle speeds had insignificant impact to the performance. However, the vehicle moving direction and its distance from the small cell are sufficient to avoid selecting short beam sojourn time. We also found that the traditional approach of selecting best SNR may not be practical since vehicles near the base station may move out of the coverage quickly. Aiming for longer beam sojourn time, our results confirmed using mobility information as contexts in C-MAB model outperformed classical MAB, best SNR and random beam selection.

References

[1] W. Saad, M. Bennis, and M. Chen, "A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020.

[2] I. Ahmed, H. Khammari, A. Shahid, A. Musa, K. S. Kim, E. De Poorter, and I. Moerman, "A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3060–3097, 2018.

[3] H. Tullberg, P. Popovski, Z. Li, M. A. Uusitalo, A. Hoglund, O. Bulakci, M. Fallgren, and J. F. Monserrat, "The METIS 5G System Concept: Meeting the 5G Requirements," *IEEE Communications Magazine*, vol. 54, no. 12, pp. 132–139, 2016.

[4] A. N. Uwaechia and N. M. Mahyuddin, "A Comprehensive Survey on Millimeter Wave Communications for Fifth-Generation Wireless Networks: Feasibility and Challenges," *IEEE Access*, vol. 8, pp. 62 367–62 414, 2020.

[5] A. Kose, C. H. Foh, H. Lee, and M. Dianati, "Beam-centric Handover Decision in Dense 5G-mmWave Networks," in *IEEE Int'l Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2020)*, Sept. 2020.

[6] G. H. Sim, S. Klos, A. Asadi, A. Klein, and M. Hollick, "An Online Context-Aware Machine Learning Algorithm for 5G mmWave Vehicular Communications," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2487–2500, 2018.

[7] D. Li, S. Wang, H. Zhao, and X. Wang, "Context-and-Social-Aware Online Beam Selection for mmWave Vehicular Communications," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 8603–8615, 2021.

[8] S. Shu Sun, T. S. Rappaport, T. A. Thomas, A. Ghosh, H. C. Nguyen, I. Z. Kovács, I. Rodriguez, O. Koymen, and A. Partyka, "Investigation of Prediction Accuracy, Sensitivity, and Parameter Stability of Large-Scale Propagation Path Loss Models for 5G Wireless Communications," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 2843–2860, 2016.

[9] L. Yan, X. Fang, and Y. Fang, "Stable Beamforming With Low Overhead for C/U-Plane Decoupled HSR Wireless Networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 7, pp. 6075–6086, 2018.

[10] 3GPP TR 25.996, "3rd Generation Partnership Project; Universal Mobile Telecommunications System (UMTS); Spatial channel model for Multiple Input Multiple Output (MIMO) simulations (Release 11)," Tech. Rep., v.11.0.0, 2003.

[11] D. Issa Mattos, J. Bosch, and H. H. Olsson, "Multi-armed bandits in the wild: Pitfalls and strategies in online experiments," *Information and Software Technology*, vol. 113, pp. 68–81, 2019.

[12] F. Li, D. Yu, H. Yang, J. Yu, H. Karl, and X. Cheng, "Multi-Armed-Bandit-Based Spectrum Scheduling Algorithms in Wireless Networks: A Survey," *IEEE Wireless Communications*, vol. 27, no. 1, pp. 24–30, 2020.

[13] F. Duchoň, A. Babinec, M. Kajan, P. Beňo, M. Florek, T. Fico, and L. Jurišica, "Path Planning with Modified a Star Algorithm for a Mobile Robot," *Procedia Engineering*, vol. 96, pp. 59–69, 2014.