Contextual Multi-Armed Bandit based Beam Allocation in mmWave V2X Communication under Blockage

Arturo Medina Cassillas*, Abdulkadir Kose†, Haeyoung Lee‡, Chuan Heng Foh§, Chee Yen Leow¶

*Department of Engineering, King's College London, London, UK
†Department of Computer Engineering, Abdullah Gul University, Kayseri, Turkey
‡School of Physics, Engineering and Computer Science, University of Hertfordshire, Hatfield, UK
§5GIC & 6GIC, Institute for Communication Systems (ICS), University of Surrey, Guildford, UK
¶Wireless Communication Centre, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Johor, Malaysia
arturo.medina_cassillas@kcl.ac.uk, abdulkadir.kose@agu.edu.tr, h.lee@herts.ac.uk, c.foh@surrey.ac.uk, bruceleow@fke.utm.my

***Abstract*—Due to its low latency and high data rates support, mmWave communication has been an important player for vehicular communication. However, this carries some disadvantages such as lower transmission distances and inability to transmit through obstacles. This work presents a Contextual Multi-Armed Bandit Algorithm based beam selection to improve connection stability in next generation communications for vehicular networks. The algorithm, through machine learning (ML), learns about the mobility contexts of the vehicles (location and route) and helps the base station make decisions on which of its beam sectors will provide connection to a vehicle. In addition, the proposed algorithm also smartly extends, via relay vehicles, beam coverage to outage vehicles which are either in NLOS condition due to blockages or not served any available beam. Through a set of experiments on the city map, the effectiveness of the algorithm is demonstrated, and the best possible solution is presented.**

***Index Terms*—beam allocation, mmWave networks, V2X.**

I. Introduction

Since the 1970s, vehicular communication networks, which enable vehicles to communicate with roadside units or other vehicles, have been the subject of extensive research [1]. These technologies began with basic implementations that guided drivers through routes using simple messages indicating turns and maneuvers. Over time, they evolved into automotive navigation systems in the 1990s that included maps of the surroundings of the vehicle as well as visible route calculations [2]. With the development of faster transmission technologies, autonomous driving has become more feasible, and many vehicles now include features such as assisted parking, braking, and steering. The development of 5G communications has made fully autonomous traffic a realistic possibility in the future.

5G communications have become an integral part of the communications industry, with millimeter wave (mmWave) becoming the standard due to its high carrier frequencies compared to the sub-6Hz band, which allows for high bandwidth and high transmission quality [3]. In the automotive industry, higher connection speeds could help design safer and more effective autonomous driving vehicles with increased control and visibility [4]. However, mmWave communications also have inherent disadvantages, such as severe signal attenuation, easily blocked signals, and limited coverage due to their short wavelengths [3]. As stated in [5], numerous efforts have been made by academia and industry to overcome the challenges of applying mmWave bands in 5G communications, especially in vehicular scenarios. Addressing these issues is crucial to fully realize the advantages of mmWave communication for connected autonomous vehicles.

In order to mitigate the drawbacks of mmWave-based vehicular communication, the benefits of vehicle-to-vehicle (V2V) relaying assisted vehicle-to-infrastructure (V2I) communication has manifolds. The use of V2V relaying can lead to improved connectivity and extended coverage for mmWave V2I communication by forwarding data to the destination, ensuring that vehicular networks function properly even in the presence of frequent blockages [5]–[8]. In [6], the authors propose the use of relays to mitigate the impact of blockages and demonstrate the positive effects of these relays by analyzing blockage probability, average blockage duration, and signal-to-interference-noise ratio (SINR) distribution. They also discuss the possibility of using a multi-hop relaying scheme, which allows relays to forward traffic to another relay to improve the range of mmWave signals and reduce the need for a high number of roadside units. However, the effectiveness of relays located far from a vehicle is limited due to higher blockage probability compared to close relays. Additionally, the reliability and latency performance of mmWave vehicular communication can be improved with provided multiple paths for data transmission via cooperative relaying, which is important for safety-critical applications in vehicular networks. Note that, the trade-off between latency and reliability shall be carefully optimized as multi-hop V2V communications can improve the reliability of mmWave signal transmission due to

reduced propagation loss and high probability of LOS at the possible cost of increased transmission latency due to the increased number of communication hops [9]. Furthermore, due to coverage blindness in directional beams in mmWave, V2V relaying can also help with the reduction of data interruptions during handovers. In [5], a V2V communication-assisted soft V2I handover scheme is proposed. This scheme utilizes V2V communication to forward handover preparation information over an actively served vehicle by the next predicted mmWave BS which is determined using machine learning.

One the other hand, the dynamic mobility of vehicular networks leads to frequent relay selection and beam alignment, which can cause high signaling overhead and significant data interruptions. To address this, relaying vehicle and beam selection shall be jointly selected to support end-to-end stable connection [7], [8]. Machine learning (ML) is increasingly gaining attention in the design and optimization of wireless communication systems, especially in vehicular networks with high mobility, heterogeneous connectivity, and diverse QoS requirements [10]. Utilizing data available in V2X networks such as vehicle kinetics, road conditions, and traffic flows, ML can be applied to various operational aspects of V2X for making the investigation of ML's applicability for relay selection and beam management in mmWave V2X networks [7], [8], [11]. In this regard, [7], [8] proposed deep reinforcement learning (DRL) based joint relay selection and beam management. The work proposed by [7] efficiently finds unblocked relays and maintains high spectral efficiency under fast-varying channels. The work also considers beam management overhead to account for beam alignment due to relay (re)selection. Similarly, via joint selection, the proposed scheme in [8] improves overall communication capacity where an imposed capacity threshold constraint ensures each user receives a satisfactory level of service.

Note that, the proposed schemes in [7], [8] have not been tested for realistic vehicular environments. In addition, [7] has only considered relaying for V2V communication, and blockage is only considered between mmWave BS and relaying vehicle in [8]. In this paper, we have Context-aware Multi-Arm Bandit (C-MAB) based relay-aided communication by incorporating vehicle mobility context under a realistic vehicular mobility environment. The aim is to establish long-lasting connections, unlike signal strength-based approaches for end-to-end V2I and V2V communication. The remainder of the paper is organized as follows: Section II describes the considered scenario and formulation of the beam allocation problem. In Section III, the proposed beam selection algorithm based on C-MAB is elaborated. The experimental results 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

In this work, we assume a mmWave small cell which is deployed to enhance the data transfer rates and boost the network capacity. The considered small cell base station comprises an array of antennas pointing towards predetermined directions, with fewer radio frequency (RF) chains installed than antennas. Each antenna is potentially composed of a set of antenna elements that produce a beam directed towards a particular direction. As defined in [12], the coverage area of a beam is referred to as a beam sector, where beams are ideally spatially separated. Owing to the limited number of RF chains, only a subset of antennas may be activated simultaneously for downlink transmission.

The depicted scenario setting is shown in Fig. 1. The mmWave base station (mmBS) is positioned in an area of Guildford town center located in the UK, consisting of a total of 12 antennas that are oriented towards distinct directions in order to provide comprehensive coverage of the surrounding environment. Due to the limited number of radio frequency (RF) chains, the mmBS is only capable of activating four beams simultaneously, enabling it to serve up to four vehicles at any given time. It is worth noting that the sector areas represented in the illustration are merely indicative of the coverage areas, and the actual coverage of the service region may vary depending on the particular channel model and antenna settings employed.

We consider the pathloss model for the 28 GHz mmWave channel as given by [13]. The model is expressed as

$$PL(d) = PL(d_0) + 10n\log_{10}(d/d_0) + X_g, \qquad (1)$$

where d is the distance between the transmitter and receiver antennas in meters, X_g describes the channel fading (which we do not include in our analysis), and $PL(d_0)$ is the free space path loss (FSPL) in dB. The FSPL is a function of the carrier frequency f_c, given by $10\log_{10}((4\pi d_0 f_c/c)^2)$, with $d_0 = 1$ m. We assume an antenna height difference of 5 m between the base station and the vehicle, and thus the distance $\hat{d}$ between the two nodes is related to d as $d = \sqrt{\hat{d}^2 + 5^2}$.

We assume that the vehicles involved in the communication have access to steerable beam antennas that can track and adjust their orientation towards the base station during the communication. With the aforementioned setup, the signal-to-noise ratio (SNR) of a given vehicle being served can be calculated as follows:

$$SNR = p_0 - PL(d) + G_{tx} + G_{rx} - N, \qquad (2)$$

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.

The utilization of mmWave small cells for vehicle-to-everything (V2X) communication presents a distinctive challenge whereby a narrow beam must be utilized to serve fast-moving vehicles. The short duration of a moving vehicle's presence within a narrow beam results in frequent handovers and signaling overheads. Moreover, the channel state is highly volatile owing to user mobility, thus a conservative approach involving the use of the most robust modulation scheme instead of an adaptive modulation scheme may be more appropriate. Consequently, to optimize the transmission of data

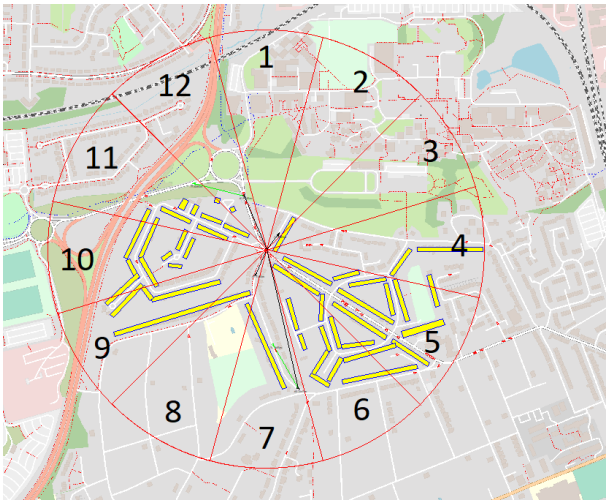

Fig. 1. A scenario of beam selection (max 4 beams out of 12 beams) for area of Guildford, UK.

between the small cell and the vehicles it serves, the radio resource allocation strategy for this particular arrangement should endeavor to serve vehicles that remain within a beam for the longest possible duration.

In the context of reliability, it is assumed that an adaptive modulation scheme is not employed. Instead, the base station utilizes the most resilient modulation scheme for communication irrespective of the reporting SNR. Let R denote the corresponding fixed data rate assigned to all beams when serving a vehicle situated within its beam sector. However, vehicles that traverse out of the beam service area experience a zero rate. Consider a set of beams $\mathcal{B}$ belonging to the mmWave small cell base station, wherein n is the maximum number of beams that can be actively deployed by the base station at any given time. Furthermore, let $\beta_i(t)$ represent the transmission rate of beam i at time t. Specifically, $\beta_i(t)$ is equal to R if beam i is serving a vehicle at time t, and $\beta_i(t) = 0$ if it is either undergoing a handover procedure, awaiting data to be resumed from the macro base station, or is inactive, or no vehicle is present within its beam sector for service. The proposed radio resource allocation strategy aims to maximize the overall data transmission from all beams over a specified period T, which is given by:

$$\max \left(\sum_{i \in \mathcal{B}} \int_0^T \beta_i(t) dt \right). \tag{3}$$

Each handover event incurs a data interruption time in $\beta_i(t) = 0$ during the event, the maximization of the above expression can also be achieved by minimizing the number of handover events since the data rate remains constant during a vehicle's service. Thus, the duration for which a vehicle remains within a beam sector, also referred to as the vehicle sojourn time, should be maximized to minimize the frequency of handovers.

In order to maximize its total transmission time, the mmBS employs all active beams for transmission. Upon a serving vehicle's departure from a beam sector, the beam previously serving the vehicle becomes inactive, and the base station can activate a beam from all available beams to serve another vehicle. The base station may opt to serve a vehicle in its beam sector in an unbiased manner or select the vehicle with the highest SNR to pair with the available beam for service. However, as we will demonstrate later, the traditional approach of selecting the highest SNR may not be suitable for mmWave vehicular networks owing to the narrow beamwidth. Due to the constrained mobility of vehicles within the local street layout, we propose utilizing mobility information of vehicles which can serve as an indirect measure of the local street layout. In this regard, we establish vehicle profiles based on their mobility information and utilize such profiles as contextual inputs for pairing vehicles with beams. The mobility information used as context may include vehicle orientation, location, and distance from the mm, which can be derived from the location or measured through timing advance.

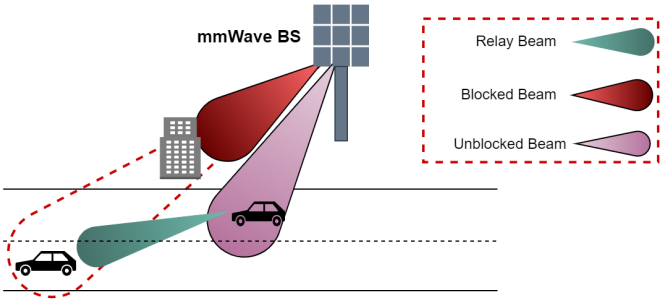

Fig. 2. Relay-assisted beam communication under blockages

III. Proposed Contextual Multi-Armed Bandit Learning Design

In our previous work [14], the effectiveness of the contextual multi-armed bandit algorithm for beam-level communication in mmWave vehicular networks has been shown. In this work, a more complex road layout scenario with roundabouts and permanent blockages due to buildings is considered. Due to different road layout and connection disruptions by blockages, context information should be carefully selected in order to capture different features about vehicular environment. In this case, adding the relevant context to the decision-making will improve the performance [14], [15]. In this work, we use the mobility information of the vehicle as a feature, specifically the travelling routes of the vehicles which start their trips from one of the twelve parking spots with ending at another parking spot. We also include the distance between the vehicle and the BS as a feature. It is classified into three levels, namely Near, Middle and Far. In addition, to mitigate the impact of the blockages, connected vehicles are used as relay nodes to extend the coverage to those deprived vehicles using V2V communications as shown in Fig. 2. The proposed C-MAB algorithm is designed also to consider the availability of second-hop vehicles around the relay vehicle in order to improve the coverage while making beam allocation.

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. Once a beam b_i is initiated for transmission at time t, an instant context $c_i(t)$ will be observed. After serving the vehicle for a duration of Δt, the reward $r_i(t + \Delta t)$ associated with the current context is measured and used to update the average reward value of the context $c_i(t)$ using the following formula:

$$\bar{r}_{c_i}(t + \Delta t) = \frac{\bar{r}_{c_i}(t) \cdot k_{c_i}(t) + r_i(t + \Delta t)}{k_{c_i}(t) + 1}, \tag{4}$$

where $k_{c_i}(t)$ represents the number of times the current context has been observed in the past.

The Alg.1 provides the steps for the contextual learning algorithm. Similar to [15], the contexts summarize both vehicle profiles (as user features) and beam sector (as arms). The objective of the ML agent is to establish the relationship between contexts and rewards, and develop a policy to derive the best arm given a context. Unlike [15], we make no assumption on the relationship between contexts and rewards. For the learning strategy, we apply exploration-first strategy. During the exploration time (30% of the simulation) the arms that can serve a vehicle are selected randomly from a set of available arms. The vehicle, say v, stays connected until it loses connection due to a blockage or leaves the sector. While being connected, another vehicle say z that is outside of the BS coverage but near vehicle v can be served by the BS sector via transmission relaying through vehicle v. We shall focus on the connection time as the main key performance indicator (KPI), and thus the reward derives directly from this KPI. The reward is measured by the overall service time of both vehicles v and z when they are connected to the mmBS sector directly or indirectly. Once vehicle v lost its connection to the mmBS sector, the reward is computed and associated with the context that encapsulates the arm and the feature vectors of vehicles v and z. After the exploration time has finished, the algorithm changes to exploitation for the rest of the simulation. During this period, the context of the vehicles within each available beam is checked. The corresponding rewards of each context are ranked and the vehicle associated with the highest reward profile will immediately receive service from the beam after the decision is made. The beam service continues until the connection is lost due to blockage or the vehicle leaving the sector. Once the connection is lost, rewards are updated for each context and beam pair. This guarantees that the best context-beam pair is always selected, improving the connection times in the long run.

IV. Result Discussion

In the section, we present experimental results for the proposed C-MAB machine learning based beam selection. The scenario for considered experiment is presented in Fig. 1. In this scenario, we consider a standalone single mmBS with 12 beams. At one time, the base station can only activate 4 beams for service. Vehicles are created and absorbed at certain

Algorithm 1 Contextual MAB for Beam and Relay Vehicle Selection

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

procedure EXPLORATION
  while $\mathbf{B} \neq \emptyset$ **do**
    Randomly select a beam b from $\mathbf{B}$
    if No vehicle is available in beam b **then**
      $\mathbf{B} \leftarrow \mathbf{B} \setminus \{b\}$
    else
      Randomly select a vehicle v in b
      if No available vehicle nearby v **then**
        return (b, v)
      else
        Randomly select a nearby vehicle z by using vehicle v as relay
        return (b, v, z)
      end if
    end if
  end while
  return None
end procedure

procedure EXPLOITATION
  $\mathbf{R_C} \leftarrow \{r_{\langle b,c\rangle} | b \in \mathbf{B}, c \in \mathcal{C}\}$
  while $\mathbf{R_C} \neq \emptyset$ **do**
    $\langle b, c\rangle \leftarrow \arg\max \mathbf{R_C}$
    if No vehicle is associated to c in beam b **then**
      $\mathbf{R_C} \leftarrow \mathbf{R_C} \setminus \{r_{\langle b,c\rangle}\}$
    else
      Randomly select a vehicle v in b with c
      Randomly select a nearby vehicle z by using vehicle v as relay
      return (b, v, z)
    end if
  end while
  return None
end procedure

procedure UPDATEREWARD(beam b, context c, reward r)
  $r_{\langle b,c\rangle} \leftarrow \frac{r_{\langle b,c\rangle} \cdot k_{\langle b,c\rangle} + r}{k_{\langle b,c\rangle} + 1}$ ▷ $r_{\langle b,c\rangle}$ is set to 0 initially
  $k_{\langle b,c\rangle} \leftarrow k_{\langle b,c\rangle} + 1$ ▷ $k_{\langle b,c\rangle}$ is set to 0 initially
end procedure

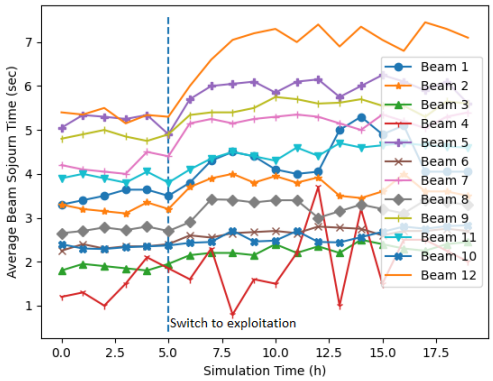

Fig. 3. Average beam sojourn time of twelve beams with C-MAB

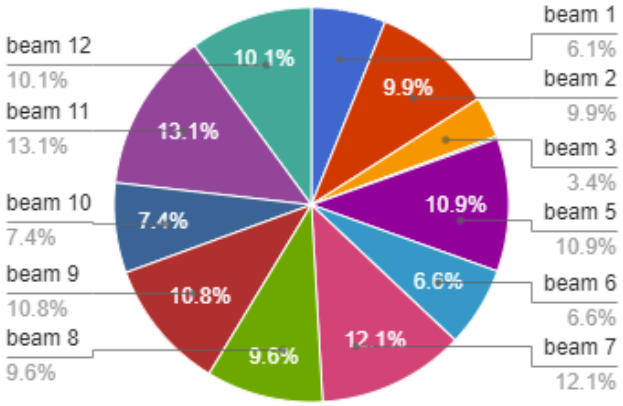

Fig. 4. Total beam utilisation ratio

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 [16] to establish the route for vehicles. We assume that each of these vehicles requires downlink data service when entering the small cell. Experiments are conducted via our own developed Python Mobility Simulation Platform (PyMoSim[1]).

In our simulation, there are 50 vehicles which continuously travel on the map. We simulate 20 hours of operation where the base station begins with full exploration for learning, and then it switches to full exploitation after the first 30% of the simulation time. As Explore-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.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Number of beams per mmBS (N_b)	12
Carrier Frequency (f_c)	28 GHz
Transmit power	30 dBm for BS/ 23 dBm for vehicles
Path loss exponent (LOS/NLOS)	1.9 / 3.8
Noise power	-90dBm
Vehicle Speed	20 m/s
Simulation time	20 hrs
Sampling time slot	0.1sec.
Simulation area	800 m x 450 m

In Fig. 3, the beam sojourn time performance of the proposed C-MAB is shown for all beams. After the exploration time, there is an increase in the mean beam sojourn time of most of the beams once the algorithm switches to exploration phase. This can be explained as beams are utilised more for specific context due to longer beam sojourn time observed in the learning phase. For example, among others Beam-7, Beam-5 and Beam-12, Beam-9, Beam-11 achieved higher mean beam sojourn time. It can be observed in the scenario depicted by Fig. 1, the shaded area for these beams have mostly the longest road coverages and higher LOS case due to less blockages. A similar pattern is depicted in Fig. 4 which shows total utilisation ratio of each beam during the whole simulation. Therefore we can state that the proposed C-MAB efficiently learns the environmental and mobility dynamics to pair a given context with an available beam with highest reward.

[1]We plan to release the full source code of PyMoSim and our scenario setup code in the near future.

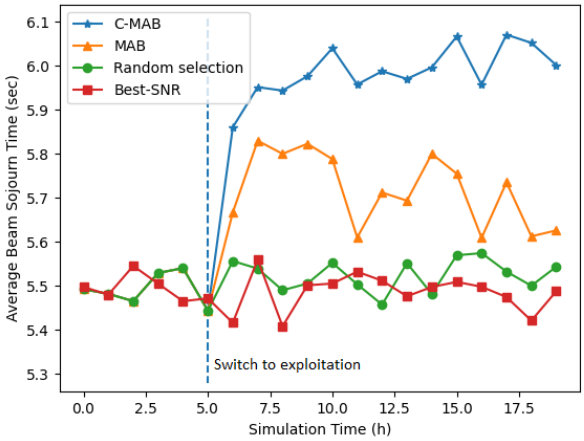

Fig. 5. Average beam beam sojourn time performance of different beam selection schemes

We also compare C-MAB algorithm with classical MAB, Best-SNR and random beam selection schemes in terms of average beam sojourn time. In best SNR, the base station greedily selects a vehicle that reports the strongest received signal strength. For our C-MAB, we use vehicle travelling route and distance from the small cell to form a context. For the travelling route, we profile a vehicle into start and

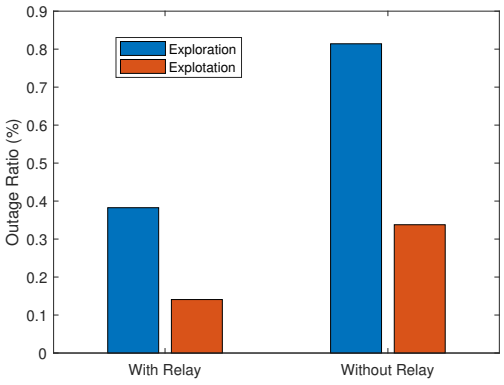

Fig. 6. Outage time percentage considering maximum available beams for service

end points of its travelling route, 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. Fig. 5 is provided for comparison of the beam sojourn time performance of the considered schemes. After the exploration time, there is a steep increase in the mean beam sojourn time for C-MAB and MAB once the algorithm switches to exploration phase. However, as C-MAB utilizes context information to make decision, it outperforms MAB by about 7%. This improvement can be high as 10% when it is compared to Best-SNR and random selection schemes.

To also evaluate the performance of C-MAB based relay-aided communication, Fig. 6 is given. For relaying, we assume that once a vehicle connected to a beam it can relay this to one nearby vehicle which has no service by any beam. Thus, outage time is nearly halved via relaying as more vehicles gain access even they are out of coverage serving beam or have a very weak signal to establish a communication link with a transmission point. The proposed C-MAB have also reduced the outage time in w/o and with relay cases by 53% and 63% respectively. The reason for the additional gain in with relay case obtained by proposed C-MAB is that it also takes the second-hop vehicle route context to achieve better decision making. In other words, if a second-hop vehicle have a opposite direction with the relay vehicle it is likely that the V2V link will be interrupted shortly. Thus, C-MAB can also learn long-lasting V2V link contact times between relay and second-hop vehicles to avoid frequent V2V link changes.

V. Conclusion

In this paper, we develop a C-MAB based beam selection and relay-aided communication to improve the connectivity robustness by enabling long-lasting V2I beam communication and V2V connections. The street layouts in real-world scenarios make mobility prediction and communication robustness a challenging task due to varying conditions, such as roundabouts, curved paths and signal blockages at the corners which need to be addressed. In such a complex scenario, due to its ability to learn various dynamics of a considered vehicular environment with appropriate context and relay selection, the proposed contextual MAB outperformed the strongest received signal-based selection and classical MAB in terms of longer beam residence time.

Acknowledgment

This work was partly sponsored by Horizon 2020 Marie Skłodowska-Curie Actions under the project SwiftV2X (grant agreement ID 101008085) and DEDICAT 6G (grant no. 101016499). We would also like to recognise the contributions from 5GIC/6GIC members to this study.

References

[1] M. Noor-A-Rahim, Z. Liu, H. Lee, G. G. M. N. Ali, D. Pesch, and P. Xiao, "A Survey on Resource Allocation in Vehicular Networks," *IEEE Trans. Intell. Transp. Sys.*, vol. 23, no. 2, pp. 701–721, 2022.

[2] N. R. Velaga, M. A. Quddus, A. L. Bristow, and Y. Zheng, "Map-aided integrity monitoring of a land vehicle navigation system," *IEEE Trans. Intell. Transp. Sys.*, vol. 13, no. 2, pp. 848–858, 2012.

[3] X. Wang, L. Kong, F. Kong, F. Qiu, M. Xia, S. Arnon, and G. Chen, "Millimeter wave communication: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1616–1653, 2018.

[4] 3GPP TR 36.872, "5G Automotive Vision," Tech. Rep., [online] Available at: https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-White-Paper-on-Automotive-Vertical-Sectors.pdf, 2015.

[5] L. Yan, H. Ding, L. Zhang, J. Liu, X. Fang, Y. Fang, M. Xiao, and X. Huang, "Machine learning-based handovers for sub-6 GHz and mmWave integrated vehicular networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4873–4885, 2019.

[6] C. Tunc and S. S. Panwar, "Mitigating the impact of blockages in millimeter-wave vehicular networks through vehicular relays," *IEEE Open J. Intell. Transp. Syst.*, vol. 2, pp. 225–239, 2021.

[7] D. Kim, M. R. Castellanos, and R. W. Heath Jr, "Joint relay selection and beam management based on deep reinforcement learning for millimeter wave vehicular communication," *arXiv preprint arXiv:2212.09862*, 2022.

[8] Y. Ju, H. Wang, Y. Chen, L. Liu, T.-X. Zheng, Q. Pei, and M. Xiao, "Drl-based beam allocation in relay-aided multi-user mmwave vehicular networks," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2022.

[9] L. Xiang, D. W. K. Ng, T. Islam, R. Schober, V. W. Wong, and J. Wang, "Cross-layer optimization of fast video delivery in cache-and buffer-enabled relaying networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 11 366–11 382, 2017.

[10] M. Noor-A-Rahim, Z. Liu, H. Lee, M. O. Khyam, J. He, D. Pesch, K. Moessner, W. Saad, and H. V. Poor, "6G for Vehicle-to-Everything (V2X) Communications: Enabling Technologies, Challenges, and Opportunities," *Proc. IEEE*, vol. 110, no. 6, pp. 712–734, 2022.

[11] A. Kose, H. Lee, C. H. Foh, and M. Dianati, "Beam-based mobility management in 5G millimetre wave V2X communications: A survey and outlook," *IEEE Open J. Intell. Transp. Syst.*, vol. 2, pp. 347–363, 2021.

[12] 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.

[13] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, 2015.

[14] A. Kose, C. H. Foh, H. Lee, and K. Moessner, "Profiling Vehicles for Improved Small Cell Beam-Vehicle Pairing Using Multi-Armed Bandit," in *2021 IEEE International Conference on Information and Communication Technology Convergence (ICTC)*, 2021, pp. 221–226.

[15] L. Li, W. Chu, J. Langford, and R. E. Schapire, "A contextual-bandit approach to personalized news article recommendation," in *Proceedings of the 19th int'l conference on World wide web*, 2010, pp. 661–670.

[16] 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.