Multi-Agent Context Learning Strategy for Interference-Aware Beam Allocation in mmWave Vehicular Communications

Abdulkadir Kose*, Haeyoung Lee†, Chuan Heng Foh‡, Mohammad Shojafar‡
*Department of Computer Engineering, Abdullah Gul University, Kayseri, TR
†School of Physics, Engineering and Computer Science, University of Hertfordshire, Hatfield, UK
‡5GIC & 6GIC, Institute for Communication Systems (ICS), University of Surrey, Guildford, UK
Email: abdulkadir.kose@agu.edu.tr, h.lee@herts.ac.uk, {c.foh, m.shojafar}@surrey.ac.uk

***Abstract*—Millimeter wave (mmWave) has been recognized as one of key technologies for 5G and beyond networks due to its potential to enhance channel bandwidth and network capacity. The use of mmWave for various applications including vehicular communications has been extensively discussed. However, applying mmWave to vehicular communications faces challenges of high mobility nodes and narrow coverage along the mmWave beams. Due to high mobility in dense networks, overlapping beams can cause strong interference which leads to performance degradation. As a remedy, beam switching capability in mmWave can be utilized. Then, frequent beam switching and cell change become inevitable to manage interference, which increase computational and signalling complexity. In order to deal with the complexity in interference control, we develop a new strategy called Multi-Agent Context Learning (MACOL), which utilizes Contextual Bandit to manage interference while allocating mmWave beams to serve vehicles in the network. Our approach demonstrates that by leveraging knowledge of neighbouring beam status, the machine learning agent can identify and avoid potential interfering transmissions to other ongoing transmissions. Furthermore, we show that even under heavy traffic loads, our proposed MACOL strategy is able to maintain low interference levels at around 10%.**

***Index Terms*—Vehicular networks, mmWave, beam management, machine learning, multi-armed bandit.**

I. Introduction

THE trends towards connected and autonomous vehicles (CAVs) have gained significant momentum in both industry and academia, driven by the numerous benefits they offer, such as reduced traffic congestion, enhanced driving safety, and decreased CO2 emission [1], [2]. Traditional autonomous driving research focused on a vehicle's independent perception using onboard sensors like radar, lidar, and camera [3]. However, these sensors have limitations in field of vision, radar range and object recognition [4], [5]. To overcome these, there is a growing interest in leveraging connected intelligence and communication networks, with Vehicle-to-Everything (V2X) communication, especially 5G NR-based cellular V2X, emerging as a crucial solution, offering high-speed, low-latency, and reliable wireless links [6], [7].

V2X facilitates the proactive exchange of vital information among different entities including networks and vehicles, resulting in an overall improvement in driving quality [8]. This capability significantly enhances a multitude of applications, including platooning, cooperative adaptive cruise control (CACC), lane merging and other intelligent transportation systems (ITS)s, thereby optimizing the traffic flow and operational efficiency [9]. Notably, a study in [10] demonstrated the efficacy of 5G-V2X communications in successfully activating emergency brakes with a response time of under one millisecond, thereby improving safety. Furthermore, V2X aids CAVs in safely passing intersections, reducing congestion [5].

In the realm of V2X applications, encompassing cooperative automated driving and enhanced infotainment services, the need for high data rates is crucial, particularly for the exchange of sensory image data, real-time traffic updates, and efficient path planning. For example, for safe and efficient driving, having up-to-date high definition (HD) 3D map would be useful [11], [12]. However, this map can quickly become outdated due to factors such as road construction. In response, vehicles can leverage V2X communication with high data rates to download the latest HD 3D map data on-demand from the infrastructure. This approach ensures keeping HD 3D map data updated, and allows it to promptly reflect changes on shorter time scales. Moreover, scenarios may arise where vehicles upload onboard sensor data to cloud and/or edge servers. These servers can then utilize the data for constructing crowdsourced HD 3D map information and performing traffic analysis. To fulfill the demanding requirements for high data rates in these diverse applications, millimeter-wave (mmWave) technology has merged as a promising candidate for 5G and beyond networks [13], particularly within the context of vehicular applications [6], [7], [14], [15].

Despite the benefits of mmWave access, transmissions in such high frequency range (above 6GHz) are more susceptible to blockages and often suffer from significant propagation loss which can hinder communication reliability [13]. Dense deployment of mmWave cells and directional transmission are two measures taken to counter non-line of sight (NLOS) cases caused by blockages and high propagation loss. However, these measures introduce new challenges for the cell level and beam level mobility management in vehicular networks due to dynamic topology. The narrow and short coverage of mmWave beams along with fast moving vehicles can result in a short period of sojourn time for a vehicle within a beam which in turns causes frequent beam selection and switching, leading to increased overhead [16]. For the connectivity establishment, there is a need for accurate beam alignment and robust connection management against rapid channel changes [17]. Moreover, while it may seem favorable to provide users with long sojourn time connectivity, the longest sojourn time does not always guarantee the best service throughput performance throughout the transmission period in the presence of interference. When a user is situated in an area where the serving beam overlaps with another beam, this service period will not be interference-free, especially if both beams are simultaneously serving different users. This situation can lead to significant performance degradation. Additionally, the need for dense deployment of mmWave small cells to increase network capacity can exacerbate the inter-beam interference, ultimately diminishing the expected throughput gain from network densification.

To capitalise the densification gain, interference-aware beam selection schemes should be considered to reduce the impact of interference. In multiple antenna-based systems, beamforming has emerged as a key technique for handling inter-cell interference. By maximizing power in desired directions while suppressing it on interfering links, beamforming plays a crucial role in both the transmitter and receiver sides. In general, despite the potential, receiver-based beamforming aims to maximize received power based on the user feedback for the received beamforming vector and does not take into account the interference caused to the other ongoing transmissions. To account for the impact of interference on neighbouring users, beamforming weights can be selected to maximize the signal-to-caused interference ratio (SCIR) [18]. When multi-antenna users are considered, joint optimization of both transmitter and receiver beamforming vectors is required to achieve high SCIR. In multi-cell multi-antenna networks, however, devising effective coordination and cooperation strategies among multiple transmission points for joint decision-making on interference-aware beam transmissions remains an open research challenge [19]–[28].

A. Related Works

In the literature, beamforming systems are commonly considered under two catagories; switched-beam system (SBS) and adaptive array (AAS). In the SBS case, beams patterns are fixed with predetermined pointing directions. In AAS, special beams can be created for each user by series of antanna array processors which apply transmitter and receiver beamforming weight vectors to adjust phase and amplitude of antenna elements. This allows the main beam lobe to be directed towards the intended user, while sidelobes are aligned towards interfering links.

The following studies [19], [20] considered AAS. The authors in [19] proposed an efficient inter-cell interference avoidance scheme to maximize sum-rate in multi-cell antenna systems. In this system, adjacent base stations (BSs) and candidate receivers cooperatively optimize their transmitter/receiver beamforming weights. The goal is to maximize the receiver power at the served user while minimizing interference caused to users in other cells. However, this work assumes that certain cells are grouped to create a virtual cell for transmission coordination for multiple users. It requires strict time synchronization and efficient coordination among cells, which can introduce computational complexity on the BS sites. To offload computational complexity to edge and enable more efficient coordination among transmission points, the authors in [20] proposed a downlink interference alignment scheme for fog radio access networks where each radio remote head (RRH) is connected to a fog access point (F-AP) via a wired link. To reduce inter-cell interference hence to maximize sum-rate of the network, F-AP decides the optimum beams for each transmission pairs upon receiving channel feedback from users in response to randomly generated beamforming matrices at each RRH. The work also emphasizes the existence of the trade-off between sum-rate performance and feedback overhead, which requires further investigations, as the feedback overhead from users to RRH increases with the number random beamforming matrices.

Since practical implementation of fully reconfigurable front-ends in AAS system are expensive [24], some studies [22], [24], [28] adopted SBS due to its low cost and less complexity. Nevertheless, frequent beam switching and cell changes are the main concerns in SBS that need to be mitigated while minimizing interference. In [28], the authors proposed a lightweight beam interference suppression method that involves beam switching and multi-cell cooperative transmission. The lightweight feature is obtained through a two-fold threshold mechanism for user grouping, which helps mitigate the problem of frequent beam switching. For user grouping, interference-to-power ratio of each user is compared with upper and lower interference power thresholds. If interference-to-power ratio of a user is higher than the upper threshold value, the user will be added to the high interference group which requires multi-cell cooperation to suppress the interference. In case interference-to-power ratio is higher than the lower threshold value, intra-cell beam switching within the same BS is performed to avoid interference. To address the complexity of coordination for user and beam pair for each transmission in massive antenna array using beamforming techniques, the work in [24] proposed a two-stage beam coordination approach which groups cells into clusters to reduce intra-cluster and inter-cluster interference. In the intra-cluster stage, based on mutual interference level, several sectors where each sector has width of 120° and 48 beams are grouped into a same cluster. In order to manage interference within a cluster, the coordinated scheduling is used to schedule transmissions for users at the transmission time interval (TTI) level. On the other hand, lack of coordination between clusters leads to interference at cell edge users. To address this issue, in the inter-cluster stage, by using time domain interference coordination approach, cluster-edge users are dynamically allocated among clusters based on interference condition to reduce inter-cluster interference. In [22], the authors proposed a coordinated beamforming among multiple BSs to reduce inter-cell interference and improve cell edge throughput performance. The coordinated BSs are connected via a beamforming control unit. Each mobile terminal measures the signal strength from each beam and feeds the received signal strength back to the serving BS. Then, each BS sends these measurements to the control unit for beamforming coordination. Coordinated beamforming is based on code-book beamforming, where beam directions are predefined, and one beam is selected for each terminal to maximize the minimum throughput of the each terminal and the total throughput.

Despite efforts made in cooperative and coordinated transmission in multi-cell scenarios and optimization of the beam management to reduce interference, the above solutions have taken a passive approach which is inadequate to proactively deal with the complication in the modeling and predicting beam radiation footprint, channel conditions, and dynamics of the vehicular mobility. In recent years, the use of machine learning (ML) [29] for beam management in mmWave systems has received a significant level of attention [23], [30]–[33]. While the conventional beam management studies including [19], [20], [22], [24], [28] rely on mathematical optimization methods, these approaches model-based. However, accurate modelings of interference model, channel estimation, and dynamics of moving vehicle can be challenging [34]. The use of ML can support the adaptive learning of the channel features, enabling reliable beam management without relying on accurate models [30]. Moreover, in high-dimensional features of the propagation scenarios that involve extensive blockage, a data-driven ML approach can be employed for more efficient beam selection by learning and adapting to changing environments.

In recent studies [23], [33], the authors apply ML techniques to address the challenge of interference-aware beam allocation. Ju *et al.* [33] design a centralized solution for user scheduling and precoding in a multi-cell mmWave network. They utilize deep neural network (DNN) at the central node to predict beamforming vector and user groups, enabling efficient assignment of cells to multiple users based on spectral efficiency considerations. The effectiveness of DNN-based prediction is demonstrated, showcasing improvements in spectral efficiency. In another work [23], Elsayed *et al.* consider a multi-cell scenario employing non-orthogonal multiple access (NOMA) technology and design a ML algorithm to manage the inter-beam and inter-cell interference. Recognizing that users located within the intersection region of multiple cells are more susceptible to interference, the authors propose distributing a ML agent in each cell to coordinate and learn the optimal policies for user-cell association and inter-beam power allocation to support these vulnerable users. This work adopts Q-learning [29], a reinforcement learning approach, and its performance is evaluated in a scenario including two base stations and static users, demonstrating an increase in the overall sum rate.

Among various ML techniques, multi-armed bandit (MAB) is considered in [35], [36] for beam selection in mmWave systems. The goal of MAB learning agent is to learn the environment and apply past experience to make its decision. For learning, a balance between exploration and exploitation becomes crucial when there are limitations on resources such as time or number of actions for exploration and exploitation are limited [37]. A MAB training beam selection policy can be used to balance exploration of the set of feasible beam.

In [35], the adaptive beam selection is formulated as a contextual multi-armed bandit (C-MAB) problem. The proposed online learning approach incorporates contextual information, such as the traffic pattern and the vehicle's traveling direction. A mmWave base station (BS), serving as a learning agent, autonomously learns the relationship between beam selection and data rate performance given the contextual information, including the traffic pattern and the presence of permanent or temporary blockages. Furthermore, in [36], the focus extends to broadcasting clustering of neighbouring vehicles that intend to download the same popular contents (e.g. movies), while considering their traveling directions. The mmWave BS learns the appropriate beams to cover multiple vehicles requesting the same contents within the cells, as well as the most suitable broadcast angle along these beams. Since a MAB based algorithm learns the expected performance in different contexts over time, it actually does not require a training phase and is highlighted as a fast learning algorithm.

Despite the effort made for ML-based interference management, none of the work above has considered vehicular communication on highway scenarios. To consider mmWave vehicular communication performance on a highway scenario where BSs are deployed on both sides of the road, [38], [39] have considered stochastic geometry to analyze the performance of beam coverage and connectivity in mmWave vehicular networks. [39] focuses on beam alignment to enhance connectivity and data rate. The authors demonstrated that the performance of vehicle communication performance relies heavily on the periodicity of beam alignment, vehicle speed, beamwidth, and base station density. It is also stated that the stability of connections relies on both beam alignment and sufficient signal quality. [38] focuses on the modeling SINR outage probability and rate coverage probability. It is found that SINR outage probability increases with BS density increases in sparse networks and large beamwidths as interference becomes apparent. However, these works have not proposed an ML-based solution to improve network performance.

B. Our Contribution

Cooperation among multiple cells to mitigate interference requires various information exchange in order to collaboratively decide which beams from which cell should be used to allocate to which users. Considering the fast changing of network topology due to vehicle mobility, keeping up with the topology changes requires low latency message exchanges among BSs for prompt decision making. To deal with the fast-changing environment, we propose a ML algorithm to adaptively respond to the changing environment without significant signalling overheads. Our design uses a central node (CN) to collect operating status of all beams and provide the status information whenever needed by any BS. In a typical mobile network setup, a set of BSs are clustered and managed by a central unit, the central unit can be served as the CN. As the collection only occurs when any BS changes its status between idle and active which does not happen frequently, the signaling overhead is low. The proposed method employs an edge-assisted contextual MAB to reduce inter-cell and inter-beam interference utilising the beam operating status of the neighbouring cells as a context. In this regards, each BS makes decision independently in a distributed manner with the context shared by the CN. In our design, the CN does not control the beam transmissions directly, rather its role is simply pulling and pushing information between BSs without making any decision in the process. This role can be easily fulfilled by existing radio access network (RAN) architecture where the information can be delivered promptly using existing control channels. In this paper, we apply our above mentioned design to a highway scenario with densely deployed multi-cell mmWave BSs along the highway to provide high data rate transmissions to the passing vehicles. The following describes the main contributions of this paper.

- We develop an analytical model which derives the upper-bound for vehicle sojourn time in a beam of a dense mmWave network. Specifically, we formulate directly the vehicle displacement within a beam. The derived upper-bound results are also used to benchmark the performance of a particular setup.
- We also study the impact of the interference on vehicle service period withing a beam. By using geometrical analysis, we derived the probability of a vehicle being interfered by other beams which it also resides.
- We propose a lightweight, fast and intelligent distributed beam allocation ML solution. The proposed ML algorithm called *Multi-Agent Context Learning* (MACOL) algorithm extends MAB by introducing context learning in the process to enable the ML agent to learn potential transmission interference based on the shared contexts among distributed ML agents.

The remainder of the paper is organized as follows. Section II presents the modeling and derivation of beam service period which is used to study the impact of interference in a densely deployed mmWave network. In Section III, the proposed interference-aware beam allocation algorithm based on contextual MAB is introduced. The performance validation are given in Section IV showing the effectiveness of our proposed algorithm with key findings and shares potential directions for future research. Finally, the conclusions are given in Section V.

TABLE I
NOTATION TABLE

Notations	Descriptions
B_N	The total number of beams in the scenario.
b_i	The i-th beam in the scenario, where $i \in \mathcal{B}_N$ and $\mathcal{B}_{\mathcal{N}} = \{1, 2, ..., B_N\}$.
(x_i, y_i, θ_i)	A 3-tuple describing the BS location and pointing direction for beam b_i.
$\mathcal{B}_{\mathcal{L}}$	The layout of the deployment in the scenario. It is a set of (x_i, y_i, θ_i) tuples, $\forall i \in \mathcal{B}_{\mathcal{N}}$.
R_B	The radius of a beam.
Ω_B	The beamwidth of a beam.
$\|D_i\|$	In a highway scenario, it is the distance between the BS operating beam b_i and the edge of the highway.
$r_i\angle\phi_i$	A location relative to beam b_i where r_i is the distance from the beam radiation origin and ϕ_i is the angle from the beam pointing angle.
$r_{k\|i}\angle\phi_{k\|i}$	A location relative to beam b_k translated from $r_i\angle\phi_i$. That is, the location $r_i\angle\phi_i$ is re-written as $r_{k\|i}\angle\phi_{k\|i}$ which is relative to beam b_k.
A_B	The area of a beam, $A_B = \frac{1}{2}R_B^2\Omega_B$
$A_I(k)$	The area of beam b_k overlapped with other beams given the layout $\mathcal{B}_{\mathcal{L}}$.
p_i	The probability that beam b_i is active at an arbitrary observation time.
$\tilde{L}$	The random variable describing the vehicle travelling distance in a beam.
$F_L(l)$	The Cumulative Distribution Function (CDF) of $\tilde{L}$. We further modify $F_L(l)$ to $F_L^*(l)$ for highway scenario, and $\check{F}_L^*(l)$ to include interference consideration for highway scenario.
$J_k(\cdot)$	The probability that a vehicle within beam b_k starting at (r_k, ϕ_k) departs the beam after travelling l distance. We also modify $J_k(\cdot)$ for several considerations: (i) $J_k^*(\cdot)$ for highway scenario, (ii) $\check{J}_k^*(\cdot)$ to also include interference consideration for highway scenario.
$\vec{S}(t)$	The context represents the environment state at time t. ML agent i then derives its own context $\vec{S}_i(t)$ using masking.
$\mathcal{C}_i$	The set containing all contexts observed by ML agent i.
$\rho_i(t)$	The reward computed by ML agent i for the connection established at time t.

II. Analysis of Vehicle Service Period within a Beam

In this section, we perform beam analysis where we derive the *service period* of a vehicle while passing a beam. We define the vehicle service period within a beam to be the time duration that the vehicle is served by a beam from when the vehicle has connected to the beam to when the vehicle has lost the connection. The connection and disconnection are triggered by a handover event. A handover event occurs when a vehicle is about to move out of the coverage of its current beam connection. When this happens, the vehicle seeks for a target beam to handover to. Considering random beam deployment and a target beam is found, the location of the vehicle in its target beam follows a Poisson Point Process (PPP) uniformly distributed within the target beam. Each beam, say b_i in a deployment scenario can be characterized by its location of radiating origin and direction (x_i, y_i, θ_i), its beam sector radius R_B and beamwidth setting Ω_B.

A. Channel Model

We consider 3GPP Band n257 channel for the mmWave transmission. The bandwidth is set to 50 MHz. For the analysis work, we use geometric framework for our modeling approach [40]–[42]. With the directional antenna feature of a mmWave BS, we approximate the beam coverage to have a shape of a sector, and its coverage follows the definition given below.

Definition 1 (Beam Coverage for Geometric Framework)**.** In geometric framework, the coverage of a beam is defined by a particular geometric shape. In this paper, we assume that the shape of the footprint of a beam is a sector. A vehicle can communicate with the base station radiating the beam only if it is located inside the footprint of the beam.

In practice, the range of a beam transmission depends on the channel pathloss and other settings related to antennas

TABLE II
TRANSMISSION PARAMETERS SETTING

Parameter	Value
Central frequency, f_c	28GHz
Bandwidth	50MHz
Transmit power	20dBm
SNR threshold	-5dB
Beamforming gain	9dB
Noise figure	7dB

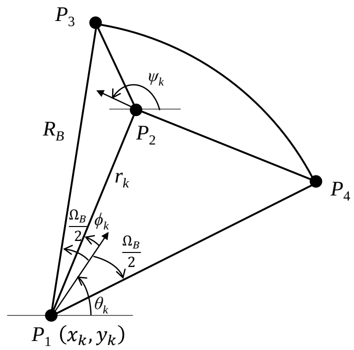


Fig. 1. Illustration of a typical beam sector geometry layout with a BS at P_1 radiating a beam pointing at θ_k, and a vehicle located at P_2 travelling in the direction ψ_k within the beam. From the perspective of the beam, the vehicle is location at $r_k\angle\phi_k$ relative to the beam.

and transceivers. In the simulation, we also consider beam coverage for practical setup to show the effectiveness of our proposed ML algorithm. In this setup, following pathloss for mmWave channel [43]

$$PL(d) = 32.4 + 20\log(f_c) + 10\alpha\log(d) \tag{1}$$

where α is 2.1 when $d < 54$ and 3.4 otherwise. Other settings are given in Table II. In this setup, while the beamwidth setting follows that of geometric framework, the beam radius R_B is determined by the received signal strength. The beam radius is the maximum range where the Signal-to-Noise Ratio (SNR) of the received transmission exceeds a predefined threshold.

Fig. 1 illustrates a typical beam sector b_k radiating from the point (x_k, y_k) towards the direction θ_k with radiation range of R_B and beamwidth of Ω_B. The figure also shows a vehicle located at P_2 within beam b_k travelling in the direction ψ_k. We specify the location of the vehicle as $r_k\angle\phi_k$ which is relative to beam b_k. Precisely, the vehicle is located r_k away from the origin of beam b_k with an angle ϕ_k from the beam pointing angle. The vehicle is said to reside within beam b_k if $r_k \in [0, R_B]$ and $\phi_k \in [-\frac{\Omega_B}{2}, \frac{\Omega_B}{2}]$.

We consider that a vehicle v is about to lose its current connection and has decided to perform a handover to a target beam b_k. At the time of the handover, vehicle v is located at $r_k\angle\phi_k$ relative to beam b_k moving at a particular direction with a constant speed. Assuming that the vehicle has handed over to this beam and it continues to move in the same direction with the same speed, the vehicle service time period depends on its travelling distance within the beam, and the distance is determined by the length between the current location of the vehicle and a point at the edge of the beam where the vehicle departs some time later. During this period, the vehicle receives connection service from the beam, thus the period is the vehicle service period.

Without loss of generality, we consider a north pointing beam in our analysis as shown in Fig. 1. In other words, we apply condition where $0 \leq \theta_k \leq \pi$ in our analysis. For south pointing beams, we can simply apply a 180^{o} rotation transformation about the location of beam b_k and reuse the same derivation approach for the analysis.

Let $\tilde{L}$ be the random variable describing the vehicle travelling distance within the beam. The cumulative distribution function (CDF) of $\tilde{L}$, which is the probability that the vehicle travelling distance in beam b_k is shorter than a particular value l can be determined by

$$F_L(l) = \frac{1}{A_B}\int_0^{R_B}\int_{\Theta_1}^{\Theta_2} r_k J_k(r_k, \phi_k, l)\, d\phi_k\, dr_k \tag{2}$$

where $\Theta_1 = -\frac{\Omega_B}{2}$, $\Theta_2 = \frac{\Omega_B}{2}$, A_B is the area of the sector, and $J_k(r_k, \phi_k, l)$ is the probability that the travelling distance of a vehicle in beam b_k is shorter than l. In other words, it is the probability that a vehicle starting at location $r_k\angle\phi_k$ relative to beam b_k will depart the beam after travelling l distance. The derivation of $J_k(r_k, \phi_k, l)$ is given in Appendix A. Finally, knowing the travelling distance, the vehicle service period can be readily obtained with the vehicle speed.

B. Highway Scenario

In the above, we have presented the formulation of a vehicle travelling distance within its target beam. While this setup is valid for mobility scenario in an urban region in general, the setup for a highway scenario is slightly different. For highway scenario, BSs are usually deployed on the road side with a distance away from the highway. Vehicles mobility on the highway is limited to two directions. The results developed in the previous subsection requires further treatment.

Fig. 2 shows a scenario of dense small cell BSs deployed along a highway. We assume that there are a number of BSs deployed along a section of a straight highway on both sides, which are grouped into a cluster. The highway section can appear in any orientation on a map, for the purpose of explanation, we normalize the orientation to consider that the highway section runs horizontally along the x-axis in the east-west direction. The location of each BS is shown in the figure.

In our scenario, each BS consists of 6 sectors with and each sector radiates an mmWave beam with a beamwidth of 60^o. Out of the 6 beams from each BS, only 3 beams cover the highway. In the scenario, 3 BSs are located on the north and 3 on the south side of the highway. For the BSs located on the south (resp. north) side of the highway, only the beam pointing north, north-east, north-west (resp. south, south-east, south-west) are serving the highway. As a result, the highway is served by 18 beams.

Let (x_k, y_k) and θ_k be the location and pointing direction of beam b_k, respectively. The beam is deployed with a distance of D_k away from the edge of the highway, and the width of the highway is H. Since all vehicles must travel on the highway,

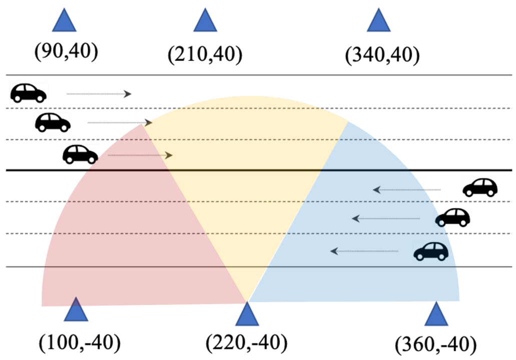


Fig. 2. Illustration of a highway layout with positions of BSs. The illustration shows the center BS situated at the south side of the highway radiating three north-pointing beams each with beamwidth of 60°

(2) is revised as follows to limit the vehicle starting location to be on the highway

$$F_L^*(l) = \frac{\int_{\Theta_1}^{\Theta_2}\int_{R_n}^{R_f} r_k J_k^*(r_k,\phi_k,l,\psi_k)dr_k d\phi_k}{\int_{\Theta_1}^{\Theta_2}\int_{R_n}^{R_f} r_k dr_k d\phi_k} \tag{3}$$

with

$$\begin{aligned} R_n &= \min\left\{\frac{D_k}{\sin(\theta_k+\phi_k)}, R_B\right\} \\ R_f &= \min\left\{\frac{D_k+H}{\sin(\theta_k+\phi_k)}, R_B\right\} \end{aligned} \tag{4}$$

where R_n and R_f are the nearest and farther valid distances to bound locations of vehicles within the highway. The quantity $J_k^*(r_k,\phi_k,l,\psi_k)$ is the probability that the travelling distance of a vehicle starting at location $r_k\angle\phi_k$ relative to beam b_k travelling towards ψ_k direction within beam b_k is shorter than l, and thus the vehicle will depart the beam after moving l distance. The derivation of $J_k^*(r_k,\phi_k,l,\psi_k)$ is given in Appendix B.

C. Impact of Interference

We further study the impact of interference on the vehicle service period. Particularly, we investigate the reduced vehicle service period which is the service period when the connection does not experience interference, i.e. interference-free. Based on geometric framework, the interference condition is defined as follows.

Definition 2 (Interference for Geometric Framework)**.** In geometric framework, a vehicle currently served by a beam is said to experience interference if it is also located within the coverage of another beam that is serving another vehicle at the same time.

The above definition corresponds to a directional transmitter and an onmi-directional receiver. The definition of interference also represents the pessimistic performance. In practice, a transmission can tolerate some level of interference. A transmission being interfered by another transmission will suffer from degraded SINR, but if the received SINR is higher than the SNR threshold, the transmission may survive. In the simulation, we also consider interference for practical setup. We shall show the impact of interference on the performance for this setup in simulation. For the following analysis, we focus on the interference for geometric framework.

Let beam b_k be the beam sector serving a moving vehicle, and another beam $b_i, i \neq k$ be the interfering beams. Let $r_k\angle\phi_k$ be the vehicle location relative and residing in beam b_k. While the location $r_k\angle\phi_k$ is expressed relative to beam b_k, we can translate and rewrite its location as $r_{i|k}\angle\phi_{i|k}$ which is relative to another beam b_i. We first find its absolute Cartesian location $(\hat{x}_k, \hat{y}_k)$ by

$$\begin{aligned} \hat{x}_k &= x_k + r_k\cos(\theta_k+\phi_k) \\ \hat{y}_k &= y_k + r_k\sin(\theta_k+\phi_k) \end{aligned} \tag{5}$$

then we translate the above to the location expression $r_{i|k}\angle\phi_{i|k}$ relative to beam b_i by

$$\begin{aligned} r_{i|k} &= \sqrt{(x_i-\hat{x}_k)^2+(y_i-\hat{y}_k)^2} \\ \phi_{i|k} &= \tan^{-1}\frac{\hat{y}_k-y_k}{\hat{x}_k-x_k} - \theta_i \end{aligned} \tag{6}$$

where $\phi_{i|k}$ is conditioned to ensure that $|\phi_{i|k}-\theta_i| \le 2\pi$ if needed. Given the above results, we can seamlessly translate the location $r_k\angle\phi_k$ relatively to beam b_k to $r_{i|k}\angle\phi_{i|k}$ which is relative to beam b_i.

The vehicle experiences interference from other beams if it is also resided in other beam sectors. We can determine whether a vehicle at $r_k\angle\phi_k$ within beam b_k experiences no interference from another beam b_i by translating its location to $r_{i|k}\angle\phi_{i|k}$ and then determine using the following calculation

$$\mathcal{P}_{\mathcal{I}}(r_{i|k},\phi_{i|k},p_i) = \begin{cases} 1-p_i, & r_{i|k}\le R_B, -\frac{\Omega_B}{2}\le\phi_{i|k}\le\frac{\Omega_B}{2} \\ 1, & \text{otherwise.} \end{cases} \tag{7}$$

In the above, p_i is the probability that beam b_i is active at an arbitrary observation time, and $\mathcal{P}_{\mathcal{I}}(r_{i|k},\phi_{i|k},p_i)$ is the probability that a vehicle at $r_k\angle\phi_k$ experiences no interference from beam b_i.

Given a particular deployment layout $\mathcal{B}_{\mathcal{L}}$ (see also Table I for its definition), assuming that all beams are always active, the area of beam b_k that is interfered by other beams can be computed by

$$A_I(k) = A_B - \int_0^{R_B}\int_{\Theta_1}^{\Theta_2} r_k \prod_{\substack{i=1\\ i\neq k}}^{B_N} \mathcal{P}_I(r_{i|k},\phi_{i|k},1)d\phi_k dr_k \tag{8}$$

where $r_{i|k}\angle\phi_{i|k}$ is the location relative to b_i given that it is at $r_k\angle\phi_k$ relative to beam b_k.

For an interference-free beam serving a vehicle, the provided service time period depends on the vehicle displacement in the beam. However, when a beam is interfered by other beams, its interference-free region is reduced, and the vehicle may experience a reduced service period due to interference. The service period is reduced to the period that the vehicle is not experiencing any interference while travelling across the serving beam.

To derive the reduced service time period, we first determine the trajectory of the vehicle within beam b_k characterized by its starting point $r_k\angle\phi_k$ and moving direction ψ_k. Since the moving direction is fixed, trajectory is a straight line. By adding the starting vector $r_k\angle\phi_k$ and the movement vector

$d\angle\psi_k$, the trajectory of the vehicle mobility can be computed using the following parametric equations

$$\begin{aligned} r_k(d) &= \sqrt{r_k^2 + d^2 + 2r_k d\cos(\gamma_k)} \\ \phi_k(d) &= \phi_k + \text{atan2}(d\sin(\gamma_k), r_k + d\cos(\gamma_k)) \end{aligned} \tag{9}$$

where $\gamma_k = \psi_k - (\theta_k + \phi_k)$, and $r_k(0)\angle\phi_k(0)$ is the starting location relative to beam b_k, $r_k(d)\angle\phi_k(d)$ is the location after travelling d distance. Likewise, at any location $r_k(d)\angle\phi_k(d)$, we denote $r_{i|k}(d)\angle\phi_{i|k}(d)$ to be the translated location relative to beam b_i.

Given a trajectory within beam b_k, say ζ characterized by $(r_k, \phi_k, \psi_k, b_k)$, we can determine the length of the trajectory τ_ζ. The derivation of τ_ζ is already given in Appendix A for various cases. Then the distance that the vehicle travelled experiencing no interference is given by

$$d_\zeta^{(r_k,\phi_k,\psi_k,b_k)} = \int_0^{\tau_\zeta} \prod_{\substack{i=1 \\ i\neq k}}^{B_N} \mathcal{P}_I\left(r_{i|k}(d), \phi_{i|k}(d), p_i\right) dd. \tag{10}$$

Recall that we define $J_k(r_k, \phi_k, l)$ to be the probability that the *service distance* of beam b_k for a vehicle starting at $r_k\angle\phi_k$ is shorter than l. Considering the highway scenario with interference from other beams, we revise $J_k(\cdot)$ and denote it as $\breve{J}_k^*(\cdot)$.

Given $d_\zeta^{(r_k,\phi_k,\psi_k,b_k)}$ which is the distance travelled by a vehicle without experiencing interference, we can derive $\breve{J}_k^*(r_k, \phi_k, l, \psi_k)$ directly by

$$\breve{J}_k^*(r_k, \phi_k, l, \psi_k) = \begin{cases} 1 & d_\zeta^{(r_k,\phi_k,\psi_k,b_k)} \leq l \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Rewriting (3) using $\breve{J}_k^*(\cdot)$, we yield

$$\breve{F}_L^*(l) = \frac{\int_{\Theta_1}^{\Theta_2}\int_{R_n}^{R_f} r_k \breve{J}_k^*(r_k, \phi_k, l, \psi_k) dr_k d\phi_k}{\int_{\Theta_1}^{\Theta_2}\int_{R_n}^{R_f} r_k dr_k d\phi_k} \tag{12}$$

which describes $F_L^*(l)$ for highway scenario capturing the interference from other beams.

III. Multi-Agent Contextual Bandit for Interference Management

We recognize that managing transmission interference in a dense small cell deployment such as one we consider in this paper requires precise knowledge of the interference condition so that beam transmissions can avoid competing and interfering with each other. Designing an algorithm that allows beams to coordinate with each other precisely is challenging since each deployment scenario has a different interference condition, hand crafting an algorithm to suit each deployment scenario is not only tedious but also prone to error due to environment change. In this paper, we present an online reinforcement learning algorithm that can not only self-learn the interference condition but also adapt to the changing environment.

We extend contextual multi-armed bandit (C-MAB) to enable context learning and multi-agent coordination. While the design objective of C-MAB is to learn and discover the option given a specific context, our design has slightly different objective. Instead of discovering the best option, our ML agent attempts to learn and cluster the contexts into two groups in an unsupervised manner, one of which is deemed to cause interference if a transmission is made, and another is deemed to yield an interference free transmission. With the proper classification of contexts, during the exploitation, the ML agent will only initiate a transmission when the context is classified under interference free group, or perform backoff for context classified under interference. Next, we shall describe our proposed Multi-Agent Context Learning algorithm (MACOL).

A. Proposed Algorithm

In a classical MAB algorithm, machine learning agent explores possible options known as *actions* or *arms* to learn the outcome of each option. The outcomes are often quantified into measurable *rewards*, and the objective of MAB is to maximize the rewards which equates to better decision making. As the environment changes, so are the outcomes and rewards. In MAB, the rewards are often assumed to follow some probability distribution function. As a result, learning the statistical characteristics of the reward distribution is important to make optimal decision in order to maximize the overall rewards. MAB also assumes that the environment remains in the same state during the entire operation. Recognizing the underlying environment may also undergo different phases or states, and reward distribution in each state of the environment may exhibit a different statistical characteristic, it is thus necessary to observe the environment state and apply different treatments for different environment states.

Contextual MAB (C-MAB) addresses this issue by introducing the concept of *contexts* into the classical MAB. In C-MAB, different contexts represent different observed environment states. For observed environment states (contexts), C-MAB learns the reward distribution for each environment state during exploration, associates the learned reward distribution to the observed state, and applies corresponding reward distribution for decision making during exploitation.

1) Multiple Agents and Contexts: A necessary step in C-MAB design is to define the environment contexts. In our design, we use the operating status of all beams in the BS cluster within the deployment scenario to form the context. The operating status of a beam can either be *active* or *idle*. When a beam is serving a vehicle, it is said to be in an active status, otherwise it is said to be idle as it is not servicing any vehicle. Let $\vec{S}(t)$ be a context vector collecting the operating status of all beams at the observing time t, and define as follows.

$$\vec{S}(t) = \langle s_1(t), s_2(t), \cdots, s_{B_N}(t)\rangle, \tag{13}$$

where $s_i(t) \in \{0, 1\}$ is the element describing the operating status of beam b_i at time t with 0 indicating idle and 1 indicating active.

Our proposed MACOL uses multiple agents for coordination. Each beam in the BS cluster has an ML agent which independently observes the contexts and interacts with the environment. We call ML agent i for the agent associated with beam b_i. Each ML agent coordinates with others through sharing of its operating status in a form of contexts.

The objective of the learning for each ML agent is to avoid transmission interference by coordinating transmissions among the ML agents, resulting in longer vehicle service period. For each ML agent, it is sufficient to observe the status of those beams that will cause interference to its transmission rather than all beams. In other words, instead of applying the observed context $\vec{S}(t)$ in the learning, each ML agent masks out the status of non-interfering beams from $\vec{S}(t)$ to form its own version of contexts. The purpose of masking out non-interfering beams is to reduce the context space for each ML agent for more focused learning and improved learning convergence. Based on the underlying deployment layout, the non-interfering beams for each beam can be easily identified, and a corresponding mask setting can be applied. The applied mask will remain unchanged throughout the operation unless the underlying deployment layout has changed. The masking vector indicating non-interfering beams for beam b_i can be defined as follows.

$$\vec{M}_i = \langle m_{1|i}, m_{2|i}, \cdots, \ m_{B_N|i} \rangle, \tag{14}$$

where $m_{k|i} \in \{1, 0\}$ indicates whether activation of beam k does not interfere beam i, i.e., $m_{k|i}$ becomes 1 when beam k can interfere beam i and 0 for otherwise. We denote $\vec{S}_i(t)$ to be the observed context of beam b_i after masking all non-interfering beam status to 0. By using (14), $\vec{S}_i(t)$ can be expressed as follows.

$$\vec{S}_i(t) = \vec{S}(t) \circ \vec{M}_i = \langle s_{1|i}, \cdots, s_{k|i}, \cdots, s_{B_N|i} \rangle \tag{15}$$

where $\circ$ is the Hadamard Product operator and $s_{k|i}(t)$ is equal to 1 only if beam k (one of interfering beams to beam i) is activated. For $s_{k|i}(t) = 0$, it can indicate two cases, (i) beam k is not activated (while beam k can interfere to beam i) or (ii) the status of beam k does not affect to beam i. This context vector is an important information enabling the coordination among ML agents to avoid transmission interference.

2) Actions, Rewards and Context Learning: In classical bandit algorithms, the ML agent picks an arm using a particular policy with a model, observes the reward from the environment, and refines the model based on the observed reward. The aim of the ML agent is to learn which arm can produce the best reward. If contexts are used, the arm picking decision will also be based on the observed context. To apply contextual MAB to our problem, we need to design appropriate actions and rewards. We first discuss the actions. When a beam is available for the next transmission service, the ML agent can greedily choose to serve a vehicle. However, this greedy option might cause its transmission to interfere with other ongoing transmissions, thus it might choose not to perform a service and backoff for some period of time. As a result, the ML agent has two arms, one of which is to serve a vehicle and another is to perform a backoff.

We shall now discuss the rewards. The ML agent learns the goodness of its action by observing the reward. In our problem, the reward should indicate whether a transmission has suffered an interference. This can be done by measuring the quality of the transmission during the transmission service. There are several ways we can measure the quality of a transmission, for example by measuring Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), Received Signal Strength Indicator (RSSI), or the goodput rate in bps measuring the average number of bits per second successfully received by the vehicle over the transmission period. In this work, we shall use the goodput rate for the reward. A higher goodput rate indicates good channel quality whereas a lower goodput rate may be due to poor channel quality or transmission interference, or both.

Given the above design of actions and rewards defined for our problem, we see that a classical MAB is unsuitable to apply to the problem. In a classical MAB, a reward is observed after take an action. If no action is taken, no reward can be observed. In our problem, as the ML agent can choose to backoff, no transmission will take place and thus it is impossible to obtain a reward if the backoff action is chosen. It is also not trivial to assign a reward to a backoff action as this requires other ML agents to confirm that a backoff action has indeed avoided an interference and quantify the benefit. To overcome these issues, we introduce a key process into a classical MAB. We add an unsupervised learning model into MAB for the ML agent to learn whether a transmission will cause interference with others. Precisely, based on the contexts received from the environment, the ML agent learns which contexts will cause interference to others and which will not. In other words, the ML agent classifies contexts into two groups, one labeled as "transmission-interference" group and another as "interference-free" group.

The classification of the contexts relies on the observed rewards after each transmission is made. Upon completion of a transmission service, a reward is produced. This reward is associated with the context at the time when the decision of performing a transmission service is made. We further define the value of each context to be the average reward associated with the context. As each context has a value, we can classify each context to be either "transmission-interference" or "interference-free" based on its value. With one-dimensional data, we can simply rank contexts based on their value, set a threshold to separate them into two groups. Those in the higher (resp. lower) value group signify "interference-free" (resp. "transmission-interference"), as a higher reward indicates a non-interfering transmission, and a lower reward indicates a transmission suffering interference or poor channel condition.

With our design, we not only avoid the need to quantify a reward for a backoff action, but also achieve the coordination among all ML agents by using contexts to share status. Unlike classical contextual bandit algorithms that learn directly which action is best given a context, our algorithm learns to classify contexts containing inputs from other ML agents, and take appropriate action based on the classified outcome. Hence we call our bandit algorithm the Multi-Agent Context Learning (MACOL) algorithm.

Let $\mathcal{C}_i(t)$ denote a set containing all contexts already observed by beam b_i at time t, and these contexts carry a valid reward value. When a transmission service is initiated at beam b_i, context, $\vec{S}_i(t)$, is immediately observed. Beam b_i continues to serve the vehicle for Δt period until it has left the coverage of beam b_i, then the reward is measured and the average reward value of context $\vec{S}_i(t)$ is updated. Let $\rho_i(t+\Delta t)$ be the measured reward which is associated with context $\vec{S}_i(t)$, the average reward value of context $\vec{S}_i(t)$ is updated by

$$\bar{\rho}_{\vec{S}_i}(t+\Delta t) = \frac{\bar{\rho}_{\vec{S}_i}(t) \cdot k_{\vec{S}_i}(t) + \rho_i(t+\Delta t)}{k_{\vec{S}_i}(t)+1}, \tag{16}$$

where $k_{\vec{S}_i}(t)$ is the trial number of transmission actions under context $\vec{S}_i(t)$. Besides, the set of observed contexts of beam b_i is also updated accordingly by $\mathcal{C}_i(t+\Delta t) = \mathcal{C}_i(t) \cup \{\vec{S}_i(t)\}$.

After providing a service, beam b_i is ready for the next service. It first observes the context and performs classification of the observed context. It proceeds with a transmission service if the observed context is classified as "interference-free", otherwise, a backoff is performed. As mentioned earlier, the ML agent ranks all already seen contexts based on their reward value, and set a threshold to partition those above the threshold to be "interference-free". In our design, we use the average context value as the threshold. That is,

$$\dot{\rho}(t+\Delta t) = \frac{1}{|\mathcal{C}_i(t+\Delta t)|} \sum_{\vec{S}_i(t+\Delta t) \in \mathcal{C}_i(t+\Delta t)} \bar{\rho}_{\vec{S}_i(t+\Delta t)}, \tag{17}$$

where $\dot{\rho}(t+\Delta t)$ is the average context value over all observed contexts.

In the case that a backoff action is performed, the ML agent remains silence for some period before resuming back to the normal operation. The backoff duration can be as simple as a predefined constant value. Here we use an adaptive duration for the backoff. The backoff duration is set to the average connection duration observed for the current context.

3) Exploration and Exploitation: Algorithm 1 describes our proposed MACOL algorithm for ML agent i. Our algorithm uses the explore-first strategy including two phases, firstly exploration and then exploitation. During the exploration phase of the predefined period, we let the ML agent explore and learn the environment for a specific duration. To do this, ML agent i continuously serves vehicles within its coverage area one at a time. Random vehicle selection is used when there are more than two vehicles to choose for a service. The ML agent only manages the beam of a particular frequency band. The entire frequency band is allocated to the vehicle, and thus the ML agent only manages one vehicle at a time.

The service is said to be completed whenever a connected vehicle has departed the beam. Then the ML agent computes and updates the reward for the context accordingly using (16) (lines 30-34 in Algorithm 1).

As the exploration period expires, ML agent i immediately switches to exploitation mode. During exploitation, for an observed context $\vec{S}_i(t)$ (derived from $\vec{S}(t) \circ \vec{M}_i$) at time t, the ML agent updates the classification threshold and checks the belonging of the context (lines 16-17). If the average reward value of the context is higher than the threshold the context is classified as "interference-free", and the ML agent proceeds to perform the transmission service action (lines 18-21). Otherwise, it performs the backoff action (line 23). Note that the ML agent may occasionally perform exploration with ϵ probability (lines 15, 26).

Algorithm 1 Context Learning MAB for Interference Management for ML agent i

Input:
The context vector $\vec{S}_i$ after masking.
Initialization:
$\mathcal{C}_i \leftarrow \emptyset$.

```
1: procedure EXPLORATION
2:     if No vehicle is available in the beam then
3:         return
4:     end if
5:     Select a vehicle randomly
6:     Serve the vehicle until it moves out of the beam
7:     Compute the reward ρ_i
8:     UpdateReward(S_i, ρ_i)
9: end procedure
10:
11: procedure EXPLOITATION
12:     if No vehicle is available in the beam then
13:         return
14:     end if
15:     if rand() is larger than ϵ then
16:         ρ̇ = 1/|C_i| Σ_{s∈C_i} ρ̄_s
17:         if ρ̄_{S_i} > ρ̇ then
18:             Select a vehicle randomly
19:             Serve the vehicle until it moves out of the beam
20:             Compute the reward ρ_i
21:             UpdateReward(S_i, ρ_i)
22:         else
23:             Wait until the backoff process ends
24:         end if
25:     else
26:         Exploration()
27:     end if
28: end procedure
29:
30: procedure UPDATEREWARD(context s, reward ρ)
31:     ρ̄_s ← (ρ̄_s · k_s + ρ)/(k_s + 1)      ▷ ρ̄_s is set to 0 initially
32:     k_s ← k_s + 1                          ▷ k_s is set to 0 initially
33:     C_i ← C_i ∪ {s}
34: end procedure
```

IV. Simulation and Result Discussion

A. Scenario Setup

In this section, we present the simulation results to show the performance of our proposed MACOL algorithm. We use our own developed Python Mobility Simulation Platform (PyMoSim) for the simulation. Our main purpose is to confirm that the MACOL ML agent can indeed learn whether initiating a transmission by activating its beam will potentially cause an interference to other transmissions by neighbouring beams, and use this learned knowledge to avoid interference.

We focus on a highway scenario illustrated in Fig. 2 including 6 BSs having 3 mmWave beams. In Fig. 3, we show a screen snapshot of our simulation. In our study, we use ideal sector-shaped radiation pattern with the radiation range of 80

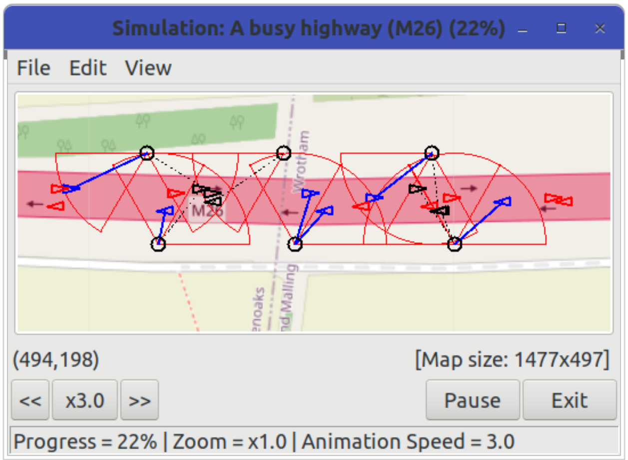


Fig. 3. Screen snapshot of Pymosim simulation running our proposed highway scenario based on geometric framework. Simulation code is given in https://github.com/cfoh/beam-analysis

m and beamwidth of 60°. The ideal sector-shaped radiation pattern is illustrated in Fig. 3. While a BS consists of 3 beams, these 3 beams are independently controlled and operated. They are mainly used for macro BS to offloading infotainment data to vehicles [37]. Whenever the beams are available, they may decide to serve a passing vehicle. When they decide to serve a passing vehicle, we assume that the downlink infotainment data can be quickly channelled to the BS to transmit the data to the served vehicle. During the downlink transmission, the amount of data successfully transmitted will be observed to calculate the goodput rate as the reward. Assuming full downlink buffer rerouted from the macro BS to the local small cell BS, the downlink transmission is performed until the vehicle has departed the beam (i.e., the connection is lost). For transmission to service vehicles, the beam will allocate a specific frequency band. The BS may serve multiple vehicles using multiple beams, each with an orthogonal frequency band, our experiment focuses on a specific frequency band used by all vehicles and BSs. That is, neighbouring beams can interfere each other and only one vehicle can be served by a beam at one time.

We consider different traffic load conditions for the highway scenario, from the light traffic load of 6 vehicles to the heavy traffic load of 30 vehicles. To ensure we maintain the constant number of vehicles on the highway for a specific experimental setup, in our simulation, we create equal amount of vehicles from the two edges of the highway section, and set a random speed for each vehicle to travel across the highway section. Once the vehicle has reached the other end of the highway section, it immediately reappears at its original start location as a new vehicle and start to travel immediately with a new speed randomly set. The speed is set to between 80 km/h and 110 km/h which is the typical travelling speed on highway in most countries.

When a vehicle travels across the highway section, it may receive a service from a beam without any interference, it may be served but experiencing interference from other nearby beams, or it may not be served due to all nearby beams being busy. To evaluate the impact of such interference on transmission to vehicles, we are particularly interested in the ratio of overall service period to the entire period that a vehicle experienced while passing the segment of the highway. That is, if a vehicle takes T_E period of time to pass the entire segment of the highway, and during its journey, it receives transmission service for T_S among of time, then the ratio of service period to the entire period is $\frac{T_S}{T_E}$. Additionally, we also measure the ratio of its interference duration and outage duration to the entire period.

Three beam allocation algorithms are considered in our testing. The first method is the random vehicle selection. Whenever a beam is available to provide service, and vehicles are found within its beam coverage, the beam randomly selects a vehicle to provide service. In the second method, rather than randomly select a vehicle, the BS measures the uplink transmission signal strength and allocates the service to the vehicle with the strongest signal. This is also known as Best SNR which is a widely used strategy. With two reference schemes, our proposed MACOL algorithm is evaluated.

In the following experiments, we first study the impact of interference on service distance. We use geometric framework in this study to show the numerical and simulation results. We next investigate the effectiveness of MACOL for interference management where both geometric framework and practical model are considered. Finally, we study the learning efficiency and computational complexity of our proposed MACOL.

TABLE III
SYSTEM PARAMETERS SETTING

Parameter	Value
Target scenario	DL transmission in a highway covered by mmWave small cell BSs
no. of BSs	6
no. of mmWave beams per BS	3
beam transmission range and beamwidth	80 m & 60^o
max no. of interfering beams	5
no. of vehicles	varying in $[6, 40]$
vehicle traveling speed	random in $[80, 110]$ km/h when entering the highway segment, remains unchanged while passing the highway
overall simulation time	2000 sec
exploration period	one of 120, 180, 240, 300, 600 sec
ϵ (in the exploitation phase)	0.05
no. of observable contexts	32

B. Impact of Interference on Service Distance

Considering the interference for geometric framework, in (3), we derive the probability that a vehicle travelled a distance of l or less within a beam in the highway scenario. By applying (11) to (3), we can focus only on the service distance which is the distance including only when interference-free transmission service is received by the vehicle. Fig. 4 plots the numerical results of service distance CDF for various p_i settings where p_i is the probability that neighbouring beam b_i is active during an observation. In our numerical computation, we set all p_i to carry the same value of p. Note that the derivation of (3) assumes a saturated traffic load condition

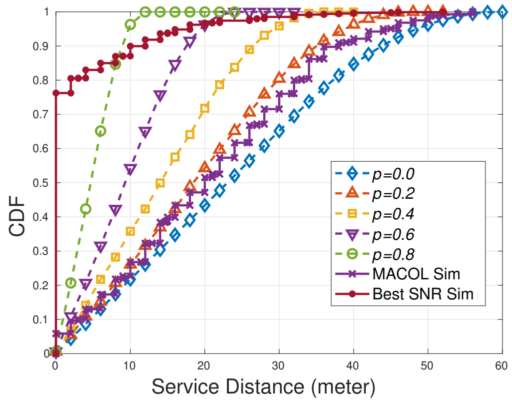


Fig. 4. Illustration of numerical service distance CDF for various p settings and simulated service distance CDF for MACOL and BestSNR techniques with traffic load of 30 vehicles, where p is the probability that its neighbouring beam is active during a transmission service by a beam.

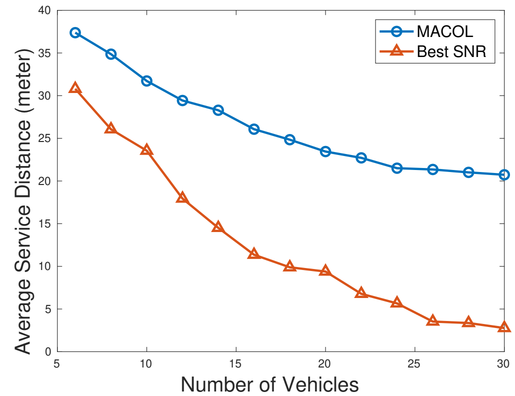


Fig. 5. Mean service distance for MACOL and Best SNR techniques with various traffic load conditions.

where when a service is completed and the beam has decided to initiate a new service, the highway is full of vehicles in all locations within the beam coverage, and the beam will unbiasedly select one vehicle for service which can be at any location. To match this high load condition in our simulation, we use traffic load setting of 30 vehicles. In Fig. 4, we plot the simulated service distance CDF for MACOL and the Best SNR techniques.

We first focus on the case of $p = 0.0$ which is equivalent to the case of no interference. The CDF result represents the best service distance performance which indicates the upper-bound service distance performance. We see that our simulated service distance CDF for MACOL resides between cases of $p = 0.0$ and $p = 0.2$ showing that our proposed MACOL can indeed maintain low interference equivalent to around 0.1 of probability of interference. As p increases, the service distance for the same percentile reduces. We also include the simulated service distance CDF for the Best SNR technique for comparison. We see that over 75% of the connections experience total interference with zero service distance, over 85% fall below that of $p = 0.8$, and over 95% fall below that of $p = 0.6$. The results suggest very intensive interference due to its greedy transmission characteristics.

Next, we compare the mean service distance for our proposed MACOL and Best SNR techniques over a range of numbers of vehicles in Fig. 5. From the figure, we see that with light traffic load of 6 vehicles, both algorithms yields over 30m of mean service distance. As the traffic load increases, we see that the Best SNR algorithm drops, and the gap between the two algorithms continue to widen. On the other hand, for the mean service distance of MACOL, the drop is relatively slow. Interestingly, it stabilizes at above 20m when traffic load becomes intense suggesting effective management of interference even under very high traffic load conditions.

C. Effectiveness of MACOL for Interference Management

We shall now focus on the effective of our proposed MACOL interference management. We first consider the interference for geometric framework. We vary the traffic intensity conditions from low to high and compare the service period ratios of a passing vehicle with various methods in Fig. 6. As can be seen, the service period ratio of a vehicle drops as the traffic condition becomes more intensive. A vehicle receives communication services over 50% of the time while traveling across the highway when there are 6 vehicles on section of the highway. As the number of vehicles increases, the communication services for each vehicle drops. There are two reasons contributing to the drop. Firstly, as the communication services are shared among all vehicles, with more vehicles on the highway, each vehicle receives reduced service duration. Secondly, more vehicles cause increased communication interference leading to lower service duration. While the drop in service period ratio is inevitable due to the sharing of the same frequency band, our proposed MACOL manages to mitigate the interference even with increasing traffic intensity. We see in Fig. 7 that MACOL maintains interference at around 10% even when the traffic intensity increases, whereas growing interference is seen for other methods when traffic intensity increases. The ability to maintain interference when traffic intensity increases shows that our proposed MACOL is effective in managing the interference.

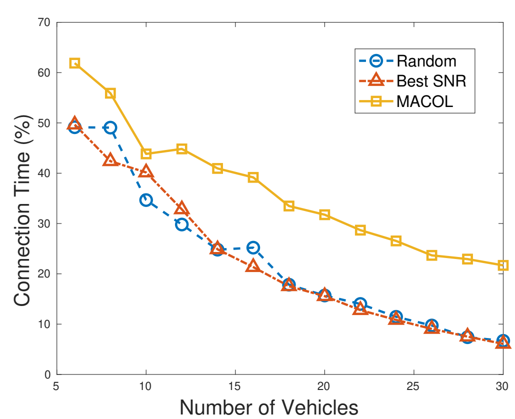


Fig. 6. Percentage of service period experienced by a vehicle passing the highway.

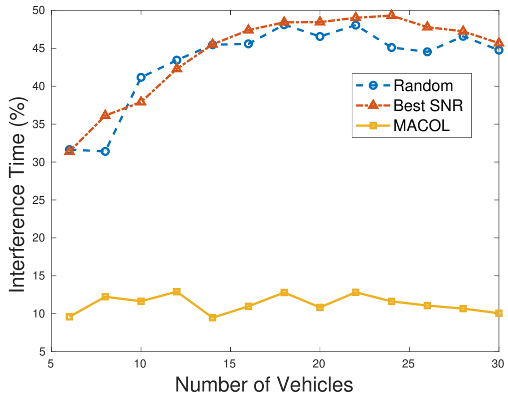


Fig. 7. Percentage of interference time experienced by a vehicle passing the highway.

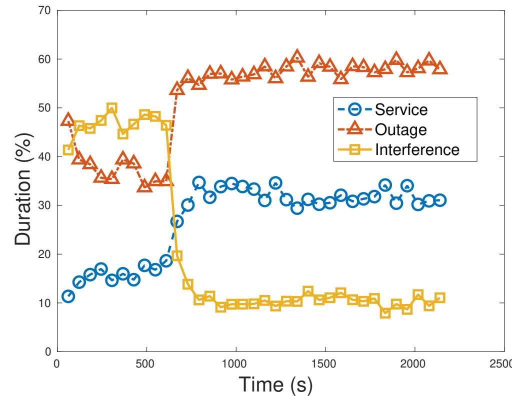


Fig. 8. Service, outage and interference duration over time for 20 vehicle setup.

To understand how MACOL manages the interference, we show the fraction of time that a vehicle receives connection service, experiences interference or outage while passing the highway in Fig. 8. Here we set the number of vehicles to 20. Note that our proposed MACOL algorithm uses explore-first strategy, during the first 10 minutes (or 600 seconds), the ML agent acts greedily to learn the contexts. During this phase, we see that a vehicle typically experiences around 45% of interference, 15% of service period without interference, and the remaining time not being able to receive a service as all nearby beams are occupied by other vehicles. Due to the high number of vehicles, and the aggressiveness of each beam attempting to serve passing vehicles whenever possible, many pairs of vehicles are served within the same region causing high interference experienced by each vehicle. Our proposed MACOL ML agent overcomes the interference by using backoff. When the ML agent judges that a service may cause an interference, it performs backoff instead to avoid the potential interference. As a result, during the exploitation, the ML agent uses the classified contexts to identify and avoid potential interference. We see that the outage ratio increases because ML agent backing off where fewer transmissions are performed. The backing off of services leads to translation of interference to service where interference ratio drops and service ratio increases. This is clearly shown in Fig. 8 that when interference ratio drops by 35% after the exploitation, outage and service ratios each jumps by over 15%.

We now consider the interference model for practical setup. In this setup, the channel model described in (1) is used instead, and SINR is used to determine a decodeability of a transmission. In this case, instead of simply classifying a transmission to either an transmission-interference or interference-free, we measure the transmission quality using SINR. In Fig. 9, we plot the SINR of transmissions from a specific beam over the time. The beam we pick is the the north pointing beam located at the south of the highway. As can be seen, the transmission SINR is below 10dB during exploration. After the learning, the transmission SINR jumps to around above 15dB. This again confirms the effectiveness of the interference management of MACOL.

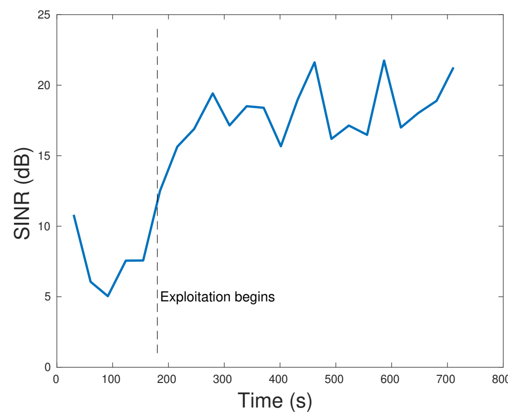


Fig. 9. Transmission quality measured in SINR of a beam service over time for 20 vehicles setup.

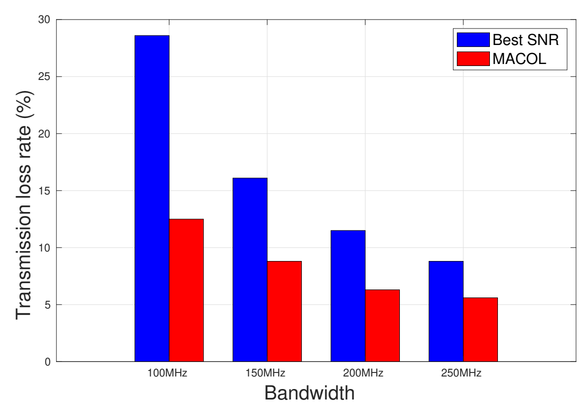


Fig. 10. Impact of bandwidth on transmission loss rate for Best SNR and MACOL for 10 vehicles setup.

D. Transmission Reliability

To support the effective V2X communication, one of the requirement would be the transmission reliability. Depending on the ITS use case and application, the required levels of transmission reliability can be varied. For example, Cooperative Awareness applications like forward collision warnings typically demand reliability levels in the range of 90-95%, whereas Cooperative Sensing applications such as crash mitigation require an even higher reliability threshold of 99% [44]. In this experiment, we study the performance of MACOL for transmission reliability measured by the loss rate due to interference under different bandwidths. As given in Table II, the mmWave BS uses Frequency Range 2 (FR2) 50MHz bandwidth for transmission to serve a vehicle. For a given number of vehicles travelling on the highway, by increasing the bandwidth to $n \times 50$MHz, the BS can serve n vehicles simultaneously on different frequency bands, and this offers a means to reduce interference. As n increases, transmission interference reduces and reliability improved. We have implemented this support in our simulation and measured the transmission loss rate for a given transmission. The results are presented in Fig. 10 for Best SNR and our proposed MACOL solution. As can be seen, Best SNR has over two times higher transmission loss rate than MACOL at 100MHz bandwidth. To support over 90% of transmission reliability, while Best SNR requires 250MHz (or 5 blocks of 50MHz frequency bands) to operate, our proposed MACOL algorithm only requires 150MHz (or 3 blocks of 50MHz frequency bands) to match the performance, representing a bandwidth saving of 100MHz. This outcome underlines the efficacy of MACOL in enhancing reliability, particularly in scenarios where heavy interference and limited resources are consideration. However, it is crucial to acknowledge that as heavy interference setup scenario is considered in this paper, under some relaxed setup and together with interference mitigation techniques, MACOL can achieve very high reliability (99%) to support some mission critical ITS use cases.

E. Learning Efficiency

Here we are interested in how effective the ML agent learns during the exploration. Fig. 11 illustrates the average interference time in percentage experienced by vehicles crossing the highway after the ML agent has switched to exploitation. In this experiment, we set the number of vehicles to 20. The time at 0 on the x-axis signifies the start of the exploitation phase for various exploration period settings. During the exploration phase shown before time $t = 0$, vehicles experience high interference at the level of just below 50% as all ML agents greedily learn the environment leading to high interference. After switching to the exploitation phase, each ML agent begins to coordinate by using shared context information to make decision, and hence we expect a drastic drop in the interference level. From the figure, we see that the ML agent can lower interference level quickly after switching to exploitation. The shortest exploration setting used in our experiment is 120s. Except for the 120s exploration duration setting where it drops slower than others and remains above 20% for 120s during the exploitation period before dropping below 20%, the other settings show similar performance in terms where the interference percentage level quickly drops below 20% after switching to exploitation. The results show that our proposed MACOL does not require a long training period. In a busy traffic condition, several minutes are sufficient for the ML agent to learn how to manage interference which makes the algorithm practical for deployment.

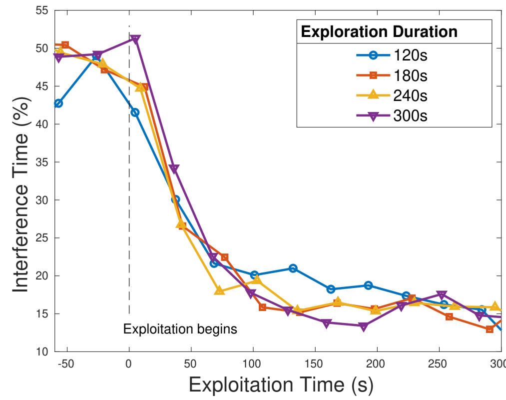


Fig. 11. Comparison of learning effectiveness for different exploration duration periods. All results are aligned based on their exploitation time such that it begins at $t = 0$.

The main knowledge that the ML agent needs to learn is the contexts and their relationship to the decision outcomes. To do so, the ML agent must explore the reward of each context by serving a vehicle during when the context is reported by the environment. The number of contexts that each ML agent can observe depends on the number of neighbouring beams that may cause interference. In our setup, a beam, say at the south site, can potentially be interfered by 3 beams from the opposite side of the highway at the north site, and 2 adjacent beams on the same side of the highway at the south site. The same applies to the beam at the north site. With 5 potential interfering beam each may report a status of either active or idle, this gives $2^5 = 32$ different contexts. Ideally, the ML agent needs to serve a vehicle in each of these 32 contexts to establish the full picture of rewards in order to derive an appropriate threshold to cluster which contexts may cause interference and which not. In our experiment, with 120s of exploration time and 20 vehicles constantly appearing on the highway, we observe from our simulation that each beam has already served around 50 vehicles, which has provided sufficient statistics for the ML agent to classify seen contexts. Since the number of contexts is not high, with sufficient traffic on the highway, the ML agent can quickly learn to differentiate and classify contexts into interfering and non-interfering groups.

F. Computational Complexity and Signaling Overhead

As illustrated in Algorithm 1, the computation of MACOL does not involve in any loop. In other words, the computational complexity of MACOL is linear with respect to the number of agents. During exploration, the ML agent greedily explore the environment independently. The computational complexity is independent of the number of ML agents, traffic load condition and number of interfering beams. During exploitation, the ML agent requires to compute the threshold using (17) which depends on the number of interfering beams. The number of interfering beams depends on the surrounding of the beam rather than the total number of ML agents. In our case, the largest number of interfering beams is 5. This makes our proposed MACOL solution scalable for deployment.

As a multi-agent ML algorithm, ML agents need to coordinate with each other which introduces signaling overhead. In MACOL, agents coordinate among each other by pushing their status and pulling the global context vector. Both status information and global context vector are small in size and can be represented by several bytes of information. For the uplink signaling, each agent updates its status whenever the beam changes its status between idle and active. Since the beam usually remains in active mode serving a vehicle for seconds, the changes are infrequent, and uplink signaling overhead is negligible. Likewise, for the downlink signaling, each agent pull the global context vector when a decision is required to decide whether a service or a backoff is to be performed. This will occur only when an active beam turns idle which is also infrequent. Thus the downlink signaling overhead is also very small. Together with the small size of signaling information, the signaling overhead of MACOL is small to negligible.

V. Conclusion and Future Work

In this paper, we studied the interference of a densely deployed mmWave small cell network for vehicular communications. Focusing on a highway scenario, we derived an analytical model to compute the vehicle sojourn time within a beam under the influence of interference caused by neighbouring beam transmissions, and we showed the impact of interference on the service distance performance of a vehicle while travelling across a beam.

To manage the interference, we proposed context learning bandit algorithm based on C-MAB with extension to enable learning of contexts. Different from classical MAB which aimed to maximise rewards, our proposed MACOL aimed to learn and classify the contexts using the rewards, and made decision based on the observed contexts. While our proposed algorithm used multi-agent design, it only involved in infrequent exchanges of small sized environment contexts. Simulation experiment was used to test our proposed design. Our simulation results showed that our proposed algorithm managed to maintain low interference level of around 10% even under a heavy load condition. The simulation results were compared and the low interference level of MACOL was confirmed by our numerical results. With supports from further simulation results, we illustrated the effectiveness of interference avoidance of MACOL showing reduced interference and increased service duration during the exploitation of MACOL. Using a practical channel model, we presented the SINR performance showing improved SINR when MACOL the exploiting the previously learned information. Considering transmission reliability for ITS safety applications, MACOL can reach nearly 95% reliability at 250MHz bandwidth, whereas Best SNR can slightly exceed 90%. Note that, MACOL only required 150MHz of bandwidth to achieve same level, which is a 40% reduction of radio resource. Finally, we discussed the learning rate and illustrated that MACOL achieved effective operation with very low exploration iteration. Particularly in our scenario, with busy road traffic, MACOL can quickly learn and operate even with a cold start since it only requires a few minutes of exploration to learn the environment.

In the future, we shall extend MACOL to other scenarios such as a dense mmWave small cell deployment scenario to study its effectiveness in interference management and perform further refinement to the algorithm. We shall also extend our analytical framework to include channel models. In the practicability aspect, we shall investigate the technical and regulatory challenges to bring MACOL to the real-world network, as well as the security and privacy of the solution.

Appendix A
Derivation of $J_k(r_k, \phi_k, l)$

Here we derive $J_k(r_k, \phi_k, l)$ which is the probability that a vehicle travelling distance within beam sector b_k is shorter than l. Note that beam b_k is characterized by its Cartesian location (x_k, y_k) and pointing direction θ_k. Since the starting location of the vehicle $r_k\angle\phi_k$ is specified relative to b_k, the beam location (x_k, y_k) is not required in the derivation. Additionally, if $r_k\angle\phi_k$ is not within beam b_k, we simply have $J_k(r_k, \phi_k, l) = 0$.

Fig. 1 in the earlier section illustrates the geometry layout of the beam sector and the vehicle within. The beam sector has three points P_1, P_3 and P_4 where P_1 is the beam origin located at (x_k, y_k). The vehicle within the beam located at $P2$ may travel in any direction and may depart the beam from either the left edge $\overline{P_1P_3}$, the right edge $\overline{P_1P_4}$, or the arc $\widehat{P_3P_4}$. Depending on which side the vehicle departing the beam, the derivation of $J_k(r_k, \phi_k, l)$ requires different treatments. For each case, we shall derive the probability $\mathcal{P}_i(l)$ that the vehicle travelled distance is shorter than l out of all possible cases. Then $J_k(r_k, \phi_k, l)$ is simply the sum of all probabilities, that is

$$J_k(r_k, \phi_k, l) = \sum_{i=1}^{3} \mathcal{P}_i(l). \tag{18}$$

In the following, we shall focus on each individual case and derive the corresponding $\mathcal{P}_i(l)$.

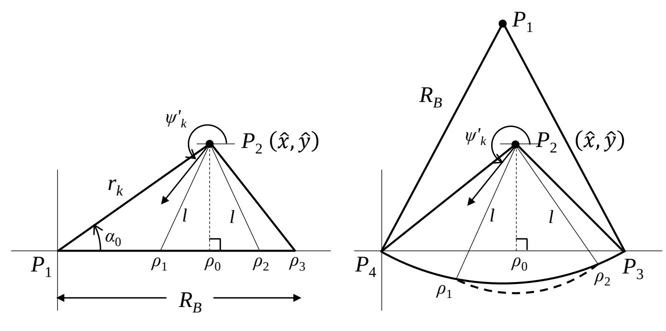


Fig. 12. Illustration of the transformed geometry layout for Cases 1 and 2 (left), and for Case 3 (right). For Case 1, rotation and reflection are performed in the transformation. For Case 2, only rotation is needed. For Case 3, translation and rotation are performed to bring P_4 to the origin and P_3 onto the x-axis.

Case 1: Departing From Left Edge

We first derive Case 1 where the vehicle departs from left edge. We begin by performing geometry transformation including translation, rotation and reflection such that P_1 is at origin, and the edge $\overline{P_1P_3}$ of left triangle $\Delta P_1P_2P_3$ coincide on x-axis as shown in Fig. 12. In the transformed system, point ρ_3 is P_3. We derive the location of P_2 as $(\hat{x}, \hat{y})$ and the transformed travelling angle ψ_k' as

$$\alpha_0 = \frac{\Omega_B}{2} - \phi_k \tag{19a}$$
$$\hat{x} = r_k \cos(\alpha_0) \tag{19b}$$
$$\hat{y} = r_k \sin(\alpha_0) \tag{19c}$$
$$\psi_k' = \frac{\Omega_B}{2} + \theta_k - \psi_k \tag{19d}$$

where ψ_k' is conditioned to fall within $[0, 2\pi]$.

Consider the angles inside the triangle, let α_1 be $\angle P_1P_2\rho_0$, α_2 be $\angle \rho_0P_2\rho_3$, α_3 be $\angle \rho_1P_2\rho_0$, and α_4 be $\angle \rho_0P_2\rho_2$. Given l, we also derive the quantities of the angles as follows

$$\alpha_1 = \tan^{-1}\left(\left|\frac{\hat{x}}{\hat{y}}\right|\right) \tag{20a}$$
$$\alpha_2 = \tan^{-1}\left(\left|\frac{R_B - \hat{x}}{\hat{y}}\right|\right) \tag{20b}$$
$$\alpha_3 = \begin{cases} \min\left(\cos^{-1}\left(\left|\frac{\hat{y}}{l}\right|\right), \alpha_1\right), & l > \hat{y} \\ 0, & \text{otherwise} \end{cases} \tag{20c}$$
$$\alpha_4 = \begin{cases} \min\left(\cos^{-1}\left(\left|\frac{\hat{y}}{l}\right|\right), \alpha_2\right), & l > \hat{y} \\ 0, & \text{otherwise} \end{cases} \tag{20d}$$

and in an extreme case where $\hat{y} = 0$, we simply apply $\alpha_1 = \alpha_2 = \frac{\pi}{2}$ and $\alpha_3 = \alpha_4 = 0$.

The condition that the vehicle will depart from the left edge, i.e. Case 1, is

$$\frac{3\pi}{2} - \alpha_1 \leq \psi_k' \leq \frac{3\pi}{2} + \alpha_2 \tag{21}$$

and the probability that the vehicle travelling distance is shorter than l is

$$\mathcal{P}_1(l) = \frac{\alpha_3 + \alpha_4}{2\pi}. \tag{22}$$

Finally, with the knowledge of the travelling direction, the departure point of the vehicle relative to beam b_k is $\acute{r}_k \angle \acute{\phi}_k$ which can be computed by

$$\acute{r}_k = \hat{x} - \hat{y}\tan\left(\frac{3\pi}{2} - \psi_k'\right) \tag{23a}$$
$$\acute{\phi}_k = \frac{\Omega_B}{2} \tag{23b}$$

and the distance travelled τ_1 is

$$\tau_1 = \sqrt{\hat{y}^2 + (\hat{x} - \acute{r})^2}. \tag{24}$$

Case 2: Departing From Right Edge

For Case 2, we perform linear transformation such that P_1 is at the origin and the edge $\overline{P_1P_4}$ coincides on x-axis. In Fig. 12, ρ_3 is P_4. The quantities α_0, $\hat{x}$ and $\hat{y}$ for Case 2 are thus

$$\alpha_0 = \frac{\Omega_B}{2} + \phi_k \tag{25a}$$
$$\hat{x} = r_k \cos(\alpha_0) \tag{25b}$$
$$\hat{y} = r_k \sin(\alpha_0) \tag{25c}$$
$$\tag{25d}$$

Similar to Case 1, given l, the quantities α_1, α_2, α_3 and α_4 can be computed using (20), and the probability that the vehicle travelling distance is shorter than l can be determined by

$$\mathcal{P}_2(l) = \frac{\alpha_3 + \alpha_4}{2\pi}. \tag{26}$$

The condition for Case 2 is

$$\frac{3\pi}{2} - \alpha_1 \leq \psi_k' \leq \frac{3\pi}{2} + \alpha_2 \tag{27}$$

where in this case

$$\psi_k' = \frac{\Omega_B}{2} - \theta_k + \psi_k \tag{28}$$

and we further condition ψ_k' to ensure that its value falls within $[0, 2\pi]$.

Given the vehicle travelling direction, the departure point follows the computation as shown in Case 1 given in (23). The distance travelled τ_2 also uses the same computation as Case 1 given in (24).

Case 3: Departing From Arc Edge

For Case 3, we perform linear transformation such that P_4 is shifted to the origin and the edge $\overline{P_4P_3}$ coincides on x-axis. The transformation involves translation and rotation of the beam.

As shown in Fig. 12 for Case 3, the location of the transformed P_2 is $(\hat{x}, \hat{y})$ and the transformed direction of movement ψ_k' can be computed by the following

$$\hat{x} = r_k \sin(\phi_k) + R_B \sin\left(\frac{\Omega_B}{2}\right) \tag{29a}$$
$$\hat{y} = -r_k \cos(\phi_k) + R_B \cos\left(\frac{\Omega_B}{2}\right) \tag{29b}$$
$$\psi_k' = \frac{3\pi}{2} - \theta_k + \psi_k. \tag{29c}$$

Consider the angles in the triangle $\Delta P_4P_2P_3$ shown in Fig. 12, let α_1 be $\angle P_4P_2\rho_0$, α_2 be $\angle \rho_0P_2P_3$, α_3 be $\angle \rho_1P_2\rho_0$, and α_4 be $\angle \rho_0P_2\rho_2$. Given l, we can derive the quantities of the angles as follows

$$\alpha_1 = \tan^{-1}\left(\left|\frac{\hat{x}}{\hat{y}}\right|\right) \tag{30a}$$
$$\alpha_2 = \tan^{-1}\left(\left|\frac{R_B - \hat{x}}{\hat{y}}\right|\right) \tag{30b}$$
$$\alpha_3 = \begin{cases} \min\left(\sin^{-1}\left(\frac{\hat{x} - x_{\rho_1}}{l}\right), \alpha_1\right), & \hat{x} > x_{\rho_1} \\ 0, & \text{otherwise} \end{cases} \tag{30c}$$
$$\alpha_4 = \begin{cases} \min\left(\sin^{-1}\left(\frac{x_{\rho_2} - \hat{x}}{l}\right), \alpha_2\right), & x_{\rho_2} > \hat{x} \\ 0, & \text{otherwise} \end{cases} \tag{30d}$$

where x_{ρ_1} and x_{ρ_2} are the x-coordinates of points ρ_1 and ρ_2 respectively. As shown in Fig. 12, the points ρ_1 and ρ_2 are the two intersection points between the circle centered at P_1 with radius R_B and the circle centered at P_2 with radius l. They can be determined by finding the intersection points of the two circles. In the case when l is small and no intersection points can be found, we apply $\alpha_3 = \alpha_4 = 0$.

For completeness, we provide the solution of the intersection points of two circles in the following. Let C_i be a circle centered at $(\breve{x}_i, \breve{y}_i)$ with radius $\breve{r}_i$, and here we define two circles C_1 and C_2. The intersection points of the two circles are

$$\begin{aligned}
d_o &= \sqrt{(\breve{x}_1 - \breve{x}_2)^2 + (\breve{y}_1 - \breve{y}_2)^2} \\
l_o &= \frac{\breve{r}_1^2 - \breve{r}_2^2 + d_o^2}{2d_o} \\
h_o &= \sqrt{\breve{r}_1^2 - l_o^2} \\
\dot{x}_1 &= \frac{l_o}{d_o}(\breve{x}_2 - \breve{x}_1) + \frac{h_o}{d_o}(\breve{y}_2 - \breve{y}_1) + \breve{x}_1 \\
\dot{y}_1 &= \frac{l_o}{d_o}(\breve{y}_2 - \breve{y}_1) - \frac{h_o}{d_o}(\breve{x}_2 - \breve{x}_1) + \breve{y}_1 \\
\dot{x}_2 &= \frac{l_o}{d_o}(\breve{x}_2 - \breve{x}_1) - \frac{h_o}{d_o}(\breve{y}_2 - \breve{y}_1) + \dot{x}_1 \\
\dot{y}_2 &= \frac{l_o}{d_o}(\breve{y}_2 - \breve{y}_1) + \frac{h_o}{d_o}(\breve{x}_2 - \breve{x}_1) + \breve{y}_1
\end{aligned}$$

where $(\dot{x}_1, \dot{y}_1)$ and $(\dot{x}_2, \dot{y}_2)$ are two intersection points.

The probability that the vehicle travelling distance is shorter than l can be determined by

$$\mathcal{P}_3(l) = \frac{\alpha_3 + \alpha_4}{2\pi}. \tag{32}$$

To derive the distance travelled within the beam for Case 3, we first formulate the departure location by using the starting location and the movement vector. The vehicle starting location in a vector form is given by $r_k\angle\phi_k$ and its movement vector is given by $d\angle\psi_k$ where d is the distance travelled. By summing the two vectors, the resultant vector is pointing to the location after the vehicle travelled d distance. The magnitude of the resultant vector is the distance away from the beam origin P_1. Since the vehicle departs the beam at the point when it is R_B distance away from P_1, then we have

$$|r_k\angle\phi_k + d\angle\psi_k| = R_B. \tag{33}$$

The magnitude of the resultant vector is already given in (9). By setting it to R_B and solving for d, we get

$$d = -r_k\cos(\gamma_k) \pm \sqrt{(r_k\cos(\gamma_k))^2 - (r_k^2 - R_B^2)}. \tag{34}$$

The above result produces two solutions, a positive and negative values indicate forward and backward movements respective. The distance travelled τ_3 for Case 3 always takes the positive solution of d.

Appendix B
Derivation of $J_k^*(r_k, \phi_k, l, \psi_k)$

We first note that $J_k^*(r_k, \phi_k, l, \psi_k)$ is a special case of $J_k(r_k, \phi_k, l)$ where the travelling direction is explicitly given. In our highway scenario, the vehicles can only move in one of the two directions. Since we orient our map such that the highway runs horizontally, vehicles can only move in the direction either $\psi_k = 0$ or π depending on which lane they are travelling on which is determined by $r_k\angle\phi_k$.

The calculation of $J_k^*(r_k, \phi_k, l, \psi_k)$ is straightforward. Knowing the setup of beam b_k, the starting location of the vehicle $r_k\angle\phi_k$, and the travelling direction ψ_k, using the working in Appendix A, we can test whether the vehicle is departing beam b_k from its left, right or arc side. We can also calculate the corresponding distance travelled $\tau^*(\psi_k)$ to be either τ_1 for Case 1, τ_2 for Case 2, or τ_3 otherwise. With the above, we have

$$J_k^*(r_k, \phi_k, l, \psi_k) = \begin{cases} 1, & \tau^*(\psi_k) \le l \\ 0, & \text{otherwise.} \end{cases} \tag{35}$$

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). The authors would like to acknowledge the support of 5GIC/6GIC at the University of Surrey.

References

[1] C. Suthaputchakun, Z. Sun, and M. Dianati, "Applications of vehicular communications for reducing fuel consumption and CO2 emission: The state of the art and research challenges," *IEEE Communications Magazine*, vol. 50, no. 12, Dec. 2012.

[2] 5G Americas, "Cellular V2X Communications Towards 5G," *White paper*, Mar. 2018.

[3] D. Parekh, N. Poddar, A. Rajpurkar, M. Chahal, N. Kumar, G. P. Joshi, and W. Cho, "A review on autonomous vehicles: Progress, methods and challenges," *Electronics*, vol. 11, no. 14, p. 2162, 2022.

[4] H. Bagheri, M. Noor-A-Rahim, Z. Liu, H. Lee, D. Pesch, K. Moessner, and P. Xiao, "5G NR-V2X: Toward connected and cooperative autonomous driving," *IEEE Communications Standards Magazine*, vol. 5, no. 1, pp. 48–54, 2021.

[5] S. Fu, W. Zhang, and Z. Jiang, "A network-level connected autonomous driving evaluation platform implementing c-v2x technology," *China Communications*, vol. 18, no. 6, pp. 77–88, 2021.

[6] I. Rasheed, F. Hu, Y.-K. Hong, and B. Balasubramanian, "Intelligent vehicle network routing with adaptive 3D beam alignment for mmWave 5G-based V2X communications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 5, pp. 2706–2718, 2020.

[7] K. Sakaguchi, R. Fukatsu, T. Yu, E. Fukuda, K. Mahler, R. Heath, T. Fujii, K. Takahashi, A. Khoryaev, S. Nagata *et al.*, "Towards mmwave v2x in 5g and beyond to support automated driving," *IEICE Transactions on Communications*, vol. 104, no. 6, pp. 587–603, 2021.

[8] S. Lee, Y. Jung, Y.-H. Park, and S.-W. Kim, "Design of V2X-based vehicular contents centric networks for autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 13 526–13 537, 2021.

[9] F. Bouali, J. Pinola, V. Karyotis, B. Wissingh, M. Mitrou, P. Krishnan, and K. Moessner, "5G for Vehicular Use Cases: Analysis of Technical Requirements, Value Propositions and Outlook," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 2, pp. 73–96, 2021.

[10] H. Cao, S. Gangakhedkar, A. R. Ali, M. Gharba, and J. Eichinger, "A testbed for experimenting 5G-V2X requiring ultra reliability and low-latency," in *WSA 2017; 21th international ITG workshop on smart antennas*. VDE, 2017, pp. 1–4.

[11] K. SAKAGUCHI, R. FUKATSU, T. YU, E. FUKUDA, K. MAHLER, R. HEATH, T. FUJII, K. TAKAHASHI, A. KHORYAEV, S. NAGATA, and T. SHIMIZU, "Towards mmWave V2X in 5G and Beyond to Support Automated Driving," *IEICE Transactions on Communications*, vol. E104.B, no. 6, pp. 587–603, 2021.

[12] 3GPP TR 26.985, "Vehicle-to-everything (V2X); Media handling and interaction," Tech. Rep., 17.0.0, 2022.

[13] 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 Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 701–721, 2022.

[14] H. Bagheri, M. Noor-A-Rahim, Z. Liu, H. Lee, D. Pesch, K. Moessner, and P. Xiao, "5G NR-V2X: Toward Connected and Cooperative Autonomous Driving," *IEEE Communications Standards Magazine*, vol. 5, no. 1, pp. 48–54, Mar. 2021.

[15] L. Zhao, X. Li, B. Gu, Z. Zhou, S. Mumtaz, V. Frascolla, H. Gacanin, M. I. Ashraf, J. Rodriguez, M. Yang *et al.*, "Vehicular communications: standardization and open issues," *IEEE Communications Standards Magazine*, vol. 2, no. 4, pp. 74–80, Dec. 2018.

[16] 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)*, Sept. 2020.

[17] 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 Journal of Intelligent Transportation Systems*, vol. 2, pp. 347–363, 2021.

[18] N. Hassanpour, J. E. Smee, J. Hou, and J. B. Soriaga, "Distributed Beamforming Based on Signal-to Caused-Interference Ratio," in *IEEE 10th Int'l Symposium on Spread Spectrum Techniques and Applications (ISSST)*, Aug. 2008.

[19] H. B. Salameh and F. H. Tha'er, "Efficient beamforming in multi-cell multi-antenna networks: Exploiting network duality," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2016.

[20] J. Yoon, Y. Kim, H. S. Jang, and B. C. Jung, "Downlink interference alignment with multi-user and multi-beam diversity for fog RANs," in *IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, Sep. 2019.

[21] J. Deng, O. Tirkkonen, R. Freij-Hollanti, T. Chen, and N. Nikaein, "Resource allocation and interference management for opportunistic relaying in integrated mmWave/sub-6 GHz 5G networks," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 94–101, June 2017.

[22] K. Hoshino and T. Fujii, "A Study on Optimal Beam Selection Algorithm for Multi-cell Coordinated Beamforming," in *IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, April 2021.

[23] M. Elsayed, K. Shimotakahara, and M. Erol-Kantarci, "Machine learning-based inter-beam inter-cell interference mitigation in mmWave," in *IEEE International Conference on Communications (ICC)*, June 2020.

[24] J. Wang, J. Weitzen, O. Bayat, V. Sevindik, and M. Li, "Interference coordination for millimeter wave communications in 5G networks for performance optimization," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 46, Feb. 2019.

[25] M. S. J. Solaija, H. Salman, A. B. Kihero, M. İ. Sağlam, and H. Arslan, "Generalized coordinated multipoint framework for 5G and beyond," *IEEE Access*, vol. 9, pp. 72 499–72 515, May 2021.

[26] F. Irram, M. Ali, Z. Maqbool, F. Qamar, and J. J. Rodrigues, "Coordinated Multi-Point Transmission in 5G and Beyond Heterogeneous Networks," in *IEEE 23rd International Multitopic Conference (INMIC)*, Nov. 2020.

[27] T. B. de Oliveira, A. L. Bazzan, B. C. da Silva, and R. Grunitzki, "Comparing multi-armed bandit algorithms and Q-learning for multiagent action selection: a case study in route choice," in *International Joint Conference on Neural Networks (IJCNN)*, July 2018.

[28] Z. Zhang and H. Yu, "Beam interference suppression in multi-cell millimeter wave communications," *Digital Communications and Networks*, vol. 5, no. 3, pp. 196–202, Aug. 2019.

[29] M. E. Morocho-Cayamcela, H. Lee, and W. Lim, "Machine Learning for 5G/B5G Mobile and Wireless Communications: Potential, Limitations, and Future Directions," *IEEE Access*, vol. 7, pp. 137 184–137 206, Sep. 2019.

[30] K. Ma, Z. Wang, W. Tian, S. Chen, and L. Hanzo, "Deep Learning for Beam-Management: State-of-the-Art, Opportunities and Challenges," 2021. [Online]. Available: https://arxiv.org/abs/2111.11177

[31] Y. Yang, Z. Gao, Y. Ma, B. Cao, and D. He, "Machine Learning Enabling Analog Beam Selection for Concurrent Transmissions in Millimeter-Wave V2V Communications," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 9185–9189, June 2020.

[32] Y. Long, Z. Chen, J. Fang, and C. Tellambura, "Data-Driven-Based Analog Beam Selection for Hybrid Beamforming Under mm-Wave Channels," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 2, pp. 340–352, May 2018.

[33] S.-L. Ju, N.-i. Kim, and K.-S. Kim, "Machine-Learning-Based User Group and Beam Selection for Coordinated Millimeter-wave Systems," *International journal of advanced smart convergence*, vol. 9, no. 4, pp. 2288–2847, Dec. 2020.

[34] 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," *Proceedings of the IEEE*, pp. 1–23, May 2022.

[35] 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, Dec. 2018.

[36] 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, May 2021.

[37] A. Kose, C. H. Foh, H. Lee, and K. Moessner, "Profiling Vehicles for Improved Small Cell Beam-Vehicle Pairing Using Multi-Armed Bandit," in *IEEE Int'l Conference on Information and Communication Technology Convergence (ICTC)*, Oct. 2021.

[38] A. Tassi, M. Egan, R. J. Piechocki, and A. Nix, "Modeling and design of millimeter-wave networks for highway vehicular communication." [Online]. Available: http://arxiv.org/pdf/1706.00298v5

[39] M. Giordani, M. Rebato, A. Zanella, and M. Zorzi, "Coverage and connectivity analysis of millimeter wave vehicular networks," *Ad Hoc Networks*, vol. 80, pp. 158–171, 2018.

[40] O. Semiari, W. Saad, M. Bennis, and B. Maham, "Caching Meets Millimeter Wave Communications for Enhanced Mobility Management in 5G Networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 779–793, 2018.

[41] L. Li, C. Zheng, and H. Liu, "Handover performance in 5G HetNets with millimeter wave cells," in *2016 16th Int'l Symposium on Communications and Information Technologies (ISCIT)*, 2016, pp. 11–16.

[42] C. Tatino, I. Malanchini, D. Aziz, and D. Yuan, "Beam based stochastic model of the coverage probability in 5G millimeter wave systems," in *2017 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*. IEEE, 2017, pp. 1–6.

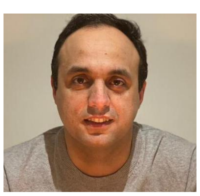
[43] S. Shu Sun, T. S. Rappaport, T. A. Thomas, A. Ghosh, H. C. Nguyen, I. Z. Kovács, I. Rodriguez, O. Koymen, and A. Partyka, "Investigation of Prediction Accuracy, Sensitivity, and Parameter Stability of Large-Scale Propagation Path Loss Models for 5G Wireless Communications," *IEEE Transactions on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

[44] M. Boban, A. Kousaridas, K. Manolakis, J. Eichinger, and W. Xu, "Use cases, requirements, and design considerations for 5G V2X," *arXiv preprint arXiv:1712.01754*, 2017.

Abdulkadir Kose received the M.Sc. degree from Yeditepe University, Turkey, in 2016, and the Ph.D. degree from the Institute for Communication Systems (ICS), the University of Surrey, U.K., in 2021. He is currently an Assistant Professor at Abdullah Gül University, Turkey. His recent research interests include mmWave networks, vehicular communications, mobility management, reconfigurable intelligent surface (RIS) assisted wireless communication.

Haeyoung Lee (Member, IEEE) earned her Ph.D. degree from the Centre for Communication Systems (CCSR) at the University of Surrey, U.K., in 2015. Between 2004 and 2007, she served as a Research Engineer at the Mobile Communication R&D Centre, Samsung Electronics, South Korea, contributing to the development of a CDMA mobile internet platform. Following this, from 2007 to 2015, she held the position of Research Officer at the National Radio Research Agency in South Korea. In 2016, she joined the 5G Innovation Centre at the University of Surrey as a (senior) research fellow. Currently, she holds the position of Senior Lecturer in Mobile Communications at the University of Hertfordshire. Her research interests encompass radio resource management in wireless communication, optimization, machine learning, and vehicular networks. Since 2023, she has served as the Technical Program Committee (TPC) chair for the International Conference on Information Networking (ICOIN). Additionally, she has taken on the role of Editor for the MDPI Topic on Machine Learning in Communication Systems and Networks.

Chuan Heng Foh (Senior Member, IEEE) received the M.Sc. degree from Monash University, Melbourne, VIC, Australia, in 1999 and the Ph.D. degree from the University of Melbourne, Melbourne, VIC, Australia, in 2002. After his Ph.D., he spent six months as a Lecturer with Monash University. In December 2002, he joined Nanyang Technological University, Singapore, as an Assistant Professor until 2012. He is currently a Senior Lecturer with the University of Surrey, Guildford, U.K. He has authored or coauthored more than 180 refereed papers in international journals and conferences. His research interests include protocol design, machine learning application and performance analysis of various computer networks including wireless local area networks, mobile ad hoc and sensor networks, vehicular networks, Internet of Things, 5G/6G networks and Open RAN. He served as a Vice Chair (Europe/Africa) for IEEE TCGCC during 2015 and 2017. He is currently the Vice-Chair of the IEEE VTS Ad Hoc Committee on Mission Critical Communications. He is on the editorial boards of several international journals.



Mohammad Shojafar (M'17-SM'19) is a Senior Lecturer (Associate Professor) in the network security and an Intel Innovator, Professional ACM member and ACM Distinguished Speaker, a Fellow of the Higher Education Academy, and a Marie Curie Alumni, working in the 5G & 6G Innovation Centre (5GIC & 6GIC), Institute for Communication Systems (ICS), at the University of Surrey, UK. Before joining 5GIC & 6GIC, he was a Senior Researcher and a Marie Curie Fellow in the SPRITZ Security and Privacy Research group at the University of Padua, Italy. Dr Mohammad secured around £1.2M as PI in various EU/UK projects, including 5G Mode (funded by DSIT/UK;2023), TRACE-V2X (funded by EU/MSCA-SE;2023), D-Xpert (funded by I-UK;2023), AUTOTRUST (funded by ESA/EU;2021), PRISENODE (funded by EU/MSCA-IF:2019), and SDN-Sec (funded by Italian Government:2018). He was also COI of various UK/EU projects like HiPER-RAN (funded by DSIT/UK;2023), APTd5G project (funded by EPSRC/UKI-FNI:2022), ESKMARALD (funded by UK/NCSC;2022), GAUChO, S2C and SAMMClouds (funded by Italian Government;2016-2018). He received his Ph.D. degree in ICT from Sapienza University of Rome, Rome, Italy, in 2016 with an "Excellent" degree. He is an Associate Editor in *IEEE Transactions on Network and Service Management*, *IEEE Transactions on Intelligent Transportation Systems*, *IEEE Transactions on Circuits and Systems for Video Technology* and Computer Networks. For additional information: http://mshojafar.com