A Beam Tracking Scheme Based on Deep Reinforcement Learning for Multiple Vehicles

Binyao Cheng(✉), Long Zhao, Zibo He, and Ping Zhang

The Key Lab of Universal Wireless Communication, Ministry of Education, Beijing University of Posts and Telecommunications (BUPT), Beijing, China
chengbinyao98@163.com

Abstract. In Internet of Vehicles (IoV), beam tracking for multiple vehicles is a challenging topic due to the nonlinear mobility and inter-vehicle interference (IVI). This paper considers the scenario that multiple vehicles with high mobility are periodically served by the radiated beams of millimeter wave (mmWave) massive multiple-input multiple-output (MIMO) systems. The main objective is to maximize the probability of successful information transmission in each beam tracking period, where successful transmission is defined by signal-to-interference-plus-noise ratio (SINR) exceeding a threshold. Based on deep reinforcement learning, we propose a position prediction and joint selection (PPJS) scheme for beam selection of multiple vehicles in consideration of both the coverage and IVI. On one hand, long short-term memory (LSTM) network is employed to predict the future trajectory in upcoming beam tracking period for providing better beam coverage. On the other hand, multi-layer perception (MLP) network is designed to select the served beams by taking into account the IVI, where the vehicles are divided into clusters and the objective of beam tracking in each cluster is decomposed to reduce the scheme complexity. Simulation results demonstrate that the proposed PPJS scheme performs better than both the traditional position-based algorithm and deep Q-learning (DQN) algorithm.

Keywords: Internet of Vehicles · Beam tracking · Deep reinforcement learning

1 Introduction

The Internet of Vehicles (IoV) is one solution to the Ultra-Reliable and Low Latency Communications (URLLC) scenarios in the next generation mobile communication [1,2]. The main objectives of IoV are to improve the transportation efficiency, traffic safety and quality of information services on vehicles. In order to improve the communication quality of high-speed multi-vehicles, millimeter wave (mmWave) [3] and massive multiple-input multiple-output (MIMO) are adopted for IoV [4]. On one hand, mmWave frequency has a large bandwidth,

H. Gao et al. (Eds.): ChinaCom 2021, LNICST 433, pp. 291–305, 2022.
https://doi.org/10.1007/978-3-030-99200-2_23

which can provide high transmission rate according to Shannon formula; on the other hand, the extremely narrow beams generated by massive MIMO base station (BS) can provide high transmission gain. However, due to high-speed mobility of vehicles, it is difficult to guarantee the accurate coverage in time; meanwhile, dense vehicle distribution leads to the high inter vehicle interference (IVI) [5]. In order to balance the coverage and IVI for multiple vehicles, beam selection technology is significant for mmWave massive MIMO in IoV [6,7].

In the existing literature, beam selection for mmWave systems has been studied in the context of traditional communication scenario. A single beam selection scheme has been studied by tracking the users' path with analog beamforming architecture, which is realized by extended Kalman filter (EKF) [8]. But conventional filter scheme is mainly fit for stationary channels, which can not be directly applied for the terminals, such as vehicles or users with nonlinear trajectory. Therefore, the classical approaches, e.g., EKF and particle filter (PF), are modified for non-stationary scenarios, as well as reinforcement learning (RL)-based approaches are introduced for typical intersection scenario [9]. Taking into account mobility, a two-phase RL framework is studied to perform adaptive beam selection [10]. In this two-phase scheme, the inner agent adjusts the beam direction based on instantaneous signal-to-interference-plus-noise-ratio (SINR) reward, and the outer agent selects the number of utilized antennas based on long-term SINR reward. However, the IVI among vehicles has not been considered in both [9] and [10].

In this paper, the mmWave and massive MIMO are employed to serve multiple vehicles on the considered road segment. In order to improve the transmission quality by considering both the beam coverage and IVI, we propose a position prediction and joint selection (PPJS) scheme. In the proposed PPJS scheme, long short-term memory (LSTM) network is employed to predict the future vehicle trajectory, based on which we can improve the beam coverage; meanwhile the multi-layer perception (MLP) network is adopted to jointly select beams for minimizing the IVI among vehicles with low complexity, where multiple vehicles are divided into clusters and the objective in each cluster is decomposed. The simulation results indicate that the PPJS algorithm achieves better transmission quality than other compared schemes. In addition, the influence of beam width and the size of discrete beam direction set are discussed under different setups of moving condition, which provides a reference for practical systems.

The rest of the paper is organized as follows. Section 2 illustrates the IoV system model served by mmWave massive MIMO; Sect. 3 formulates the problem for beam selection based on reinforcement learning; The PPJS scheme is proposed in Sect. 4; Sect. 5 presents our simulations and corresponding analysis before summarizing our conclusion in Sect. 6.

2 System Model

As shown in Fig. 1(a), we consider a IoV scenario where the BS employing mmWave and N_{BS} uniform linear array (ULA) antennas simultaneously serves

K single-antenna vehicles. The BS is located on the top of a rectangle road with the coordinate of $[\frac{h_\mathrm{r}}{2}, h_\mathrm{d}]$, where $[0, 0]$ and $[h_\mathrm{r}, 0]$ are the range of single lane road, and h_d is the distance between the BS and lane. The coordinates of the K moving vehicles are $[x_k, 0]$ $(k \in \{1, 2, \cdots, K\})$. We assume that the BS periodically changes the beams to provide coverage and reduce IVI for vehicles. As shown in Fig. 1(b), a beam selection period consists of I uplink intervals and J downlink intervals, and a time interval lasts for T_i seconds. The vehicles could periodically report their locations to the BS by uplink and the BS intelligently selects the beams for the downlink data transmission.

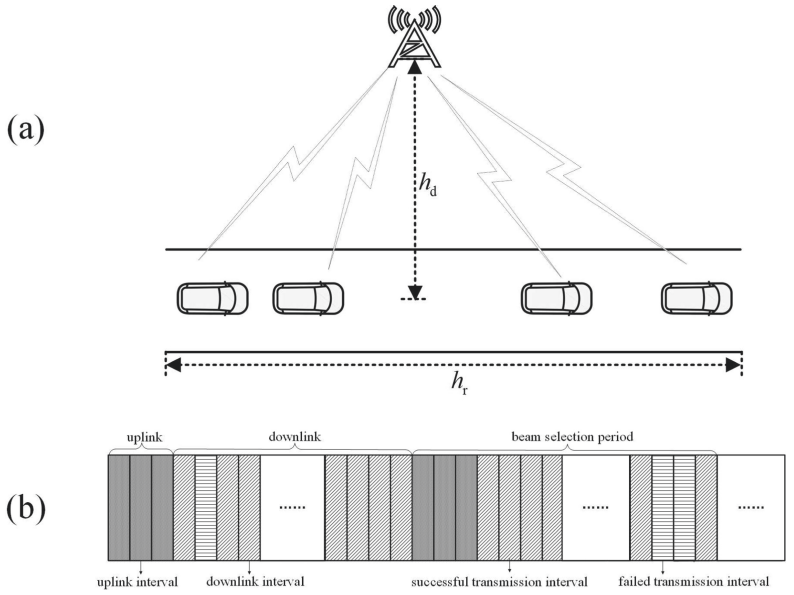

Fig. 1. Beam selection scenario and frame structure.

2.1 Channel Model

We assume that the channels from the BS to the vehicles satisfy the standard Line of Sight (LoS) path-loss model. Therefore, the complex channel vector between the BS and the kth vehicle at the jth downlink interval of the tth beam selection period can be written as

$$\mathbf{h}_{t,j,k} = \alpha_{t,j,k} \left[1, \mathrm{e}^{-j\frac{2\pi}{\lambda} d \cos\delta_{t,j,k}}, \cdots, \mathrm{e}^{-j\frac{2\pi}{\lambda} d(N_\mathrm{BS}-1)\cos\delta_{t,j,k}}\right]^\mathrm{T}, \tag{1}$$

where $\delta_{t,j,k}$ denotes the angle of departure (AoD) of the signal from the BS to the kth vehicle at the jth interval of the tth selection period; $\alpha_{t,j,k}$ is the channel fading coefficient; $d = \frac{\lambda}{2}$ denotes the antenna spacing of ULA antennas with the carrier wavelength λ [11].

2.2 Signal Model

In order to reduce the system complexity, the discrete beams are generally adopted in practical systems, i.e.,

$$\mathbf{v}(\phi) = \frac{1}{\sqrt{N_{\mathrm{BS}}}} \left[1, \mathrm{e}^{-j\frac{2\pi}{\lambda}d\cos\phi}, \cdots, \mathrm{e}^{-j\frac{2\pi}{\lambda}d(N_{\mathrm{BS}}-1)\cos\phi}\right]^{\mathrm{T}}, \tag{2}$$

where ϕ is the beam direction, which belongs to

$$\mathbf{\Phi} = \{\phi_1, \phi_2, \cdots, \phi_M\}, \tag{3}$$

and M is the total number of beam directions.

Assume that $\phi_{t,k} \in \mathbf{\Phi}$ is the beam direction of the kth vehicle at the jth interval of the tth selection period; $s_{t,j,k} \in \mathbb{C}$ with $|s_{t,j,k}| = 1$ denotes the transmitted modulation symbol; equal power p is allocated for each vehicle at the BS in every interval. The received signal of the kth vehicle at the jth interval of the tth selection period is given by

$$y_{t,j,k} = \sqrt{p}\mathbf{h}_{t,j,k}^{\mathrm{H}}\mathbf{v}_{t,k}\left(\phi_{t,k}\right)s_{t,j,k} + \sum_{i=1,i\neq k}^{K}\sqrt{p}\mathbf{h}_{t,j,k}^{\mathrm{H}}\mathbf{v}_{t,i}\left(\phi_{t,i}\right)s_{t,i} + w_{t,j,k}, \tag{4}$$

where $w_{t,j,k}$ is the complex additive white Gaussian noise (AWGN) and follows $\mathcal{CN}\left(0,\sigma^2\right)$,

Based on (4), we can obtain the SINR

$$\zeta_{t,j,k}(\phi_{t,k}) = \frac{p\left|\mathbf{h}_{t,j,k}^{\mathrm{H}}\mathbf{v}_{t,k}\left(\phi_{t,k}\right)s_{t,j,k}\right|^2}{p\sum_{i=1,i\neq k}^{K}\left|\mathbf{h}_{t,j,k}^{\mathrm{H}}\mathbf{v}_{t,i}\left(\phi_{t,i}\right)s_{t,j,i}\right|^2 + \sigma^2}. \tag{5}$$

From (5), we know that accurate beam coverage can increase the SINR, while IVI can decrease it.

Because the downlink interval is very short, the SINR can be regarded as a constant. If the SINR exceeds a given threshold ζ_{th}, the received signal can be considered to be decoded successfully, which is indicated by

$$a_{t,j,k}(\phi_{t,k}) = \begin{cases} 1, & \zeta_{t,j,k}(\phi_{t,k}) \geq \zeta_{\mathrm{th}}, \\ 0, & \text{else}. \end{cases} \tag{6}$$

By improving the transmission quality for multiple vehicles, the objective of this paper is to maximize the successful transmission probability, i.e.,

$$\max_{\phi_{t,k}\in\Phi}\left\{\frac{1}{KJ}\sum_{k=1}^{K}\sum_{j=1}^{J}a_{t,j,k}(\phi_{t,k})\right\}. \tag{7}$$

3 Problem Formulation

3.1 Deep Beam Selection Model

We use RL to formulate the beam selection problem. The basic elements of the beam selection problem are given as follows.

- State set: In order to efficiently cover the vehicles, the BS should predict the future vehicle positions according to their history trajectories. Therefore, we assume the state set to be

$$S = \{\mathbf{s}_t \mid \mathbf{s}_t = [\mathbf{x}_{t,1}, \mathbf{x}_{t,2}, \cdots, \mathbf{x}_{t,K}]\}, \tag{8}$$

  where $\mathbf{x}_{t,k} = [x_{t-t_\mathrm{h},k}, x_{t-t_\mathrm{h}+1,k}, \cdots, x_{t,k}]^\mathrm{T}$, $x_{t_i,k}$ is the kth vehicle's location at the tth beam selection period, and t_h represents the number of history selection periods.
- Action set: The actions are defined to be a set of beam directions for K vehicles, i.e.,

$$A = \left\{\mathbf{a}_t \mid \mathbf{a}_t = [\phi_{t,1}, \phi_{t,2}, \cdots, \phi_{t,K}]^\mathrm{T}\right\}, \tag{9}$$

  where $\phi_{t,k}$ denotes the beam direction of the kth vehicle at selection period t.
- Reward: The instant reward of K vehicles is defined as the average successful transmission probability in the tth beam selection period, i.e.,

$$r_t = \frac{1}{KJ}\sum_{k=1}^{K}\sum_{j=1}^{J} a_{t,j,k}(\phi_{t,k}). \tag{10}$$

3.2 Beam Selection Problem

Assume the beam selection strategy is $\pi : S \to A$, and based on the above model, the cumulative successful transmission probability in T beam selection periods can be defined as

$$R^\pi(\tau) = \sum_{t=0}^{T} \gamma^t r_t, \tag{11}$$

where $\tau = [\mathbf{s}_0, \mathbf{a}_0, r_0, \mathbf{s}_1, \mathbf{a}_1, r_1, \cdots, \mathbf{s}_T, \mathbf{a}_T, r_T]$ is a trail under beam selection strategy π, and γ represents the discount factor.

The objective of this paper is to maximize the average cumulative successful transmission probability $J(\pi)$ for any trail τ, i.e.,

$$\begin{aligned} J(\pi) &= \mathbb{E}_{\tau\sim p(\tau)}\left[R^\pi(\tau)\right] \\ &= \mathbb{E}_{\mathbf{s}\sim p(\mathbf{s}_0)}\left[\mathbb{E}_{\tau\sim p(\tau)}\left[\sum_{t=0}^{T}\gamma^t r_t \mid \tau_{\mathbf{s}_0} = \mathbf{s}\right]\right] \\ &\triangleq \mathbb{E}_{\mathbf{s}\sim p(\mathbf{s}_0)}\left[V^\pi(\mathbf{s})\right], \end{aligned} \tag{12}$$

where $p(\tau)$ is the probability of beam selection trail. Therefore the optimal beam selection strategy is

$$\pi^* = \arg\max_{\pi} J(\pi) = \arg\max_{\pi} V^\pi(\mathbf{s}). \tag{13}$$

4 PPJS Scheme Based on Deep Reinforcement Learning

As shown in Fig. 2(a) to solve the formulated problem in Sect. 3, we propose a PPJS scheme by considering both the coverage and IVI, consists of three modules. The state (8) is first fed into trajectory prediction module to obtain the future positions; Based on the predicted and current positions, the K vehicles can be divided into U clusters in vehicle division module; In beam selection module, the joint objective of each cluster is decomposed to simplify the selection complexity. The three modules will be discussed in details as follows.

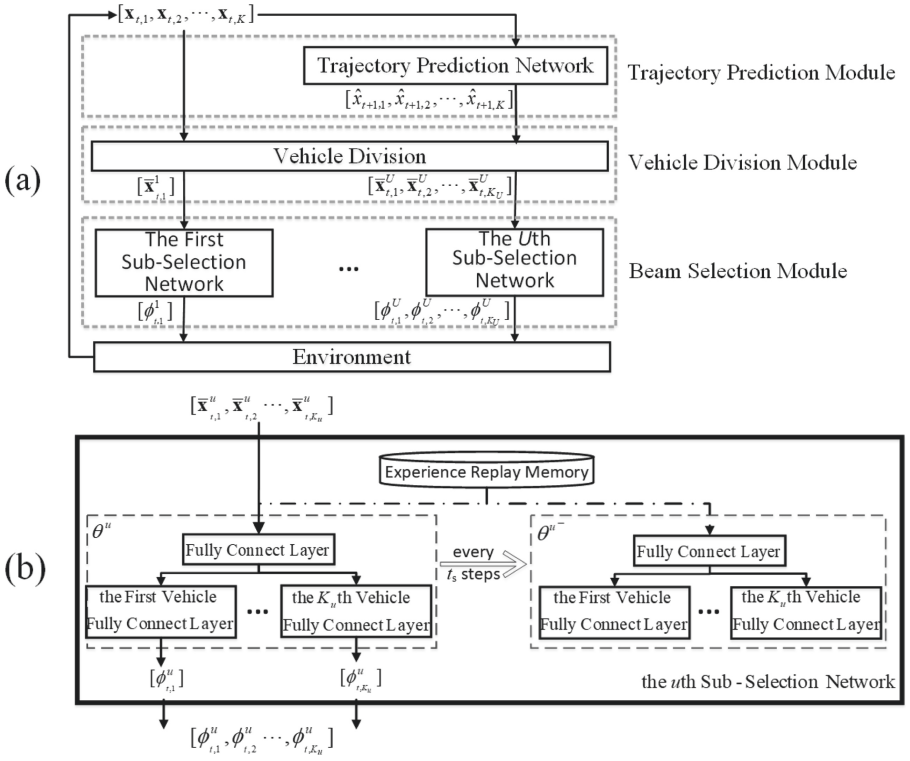

Fig. 2. The proposed PPJS scheme.

4.1 Trajectory Prediction Module

In order to ensure the accurate beam coverage, the positions of vehicles at the beginning of current and next beam selection periods should be known. So we use the LSTM network to predict the future positions according to the historical trajectory in this module.

The input of this module is a series of historical positions, which is given in (8), and the output is the position at the beginning of the next beam selection period, i.e.,

$$[\hat{x}_{t+1,1}, \hat{x}_{t+1,2}, \cdots, \hat{x}_{t+1,K}]. \tag{14}$$

The distance loss function for LSTM training is adopted in our scheme, i.e.,

$$\text{Loss} = \frac{1}{K}\sum_{k=1}^{K}(x_{t+1,k} - \hat{x}_{t+1,k})^2, \tag{15}$$

where $x_{t+1,k}$ is the future position in practical.

In order to predict an accurate future position for optimal beams, the trajectory prediction module should be pretrained before being embedded into the PPJS scheme based on practical data set. Here we should note that the beams are not directly selected to point to the current position or the next position. Because the vehicles are moving during each selection period, each beam generally points to the position that between the beginning and ending positions. However, the accurate position should be selected by considering both the coverage gain and IVI, which is achieved by beam selection module.

4.2 Vehicle Division Module

According to (10), all the IVIs should been taken into account so the scheme needs to jointly select beams for all the served vehicles, which leads to overwhelming computational complexity. To simplify the PPJS scheme, the K vehicles are decomposed into U clusters according to vehicle spacing, where the IVI between any vehicles in different clusters is small and can be ignored. Specifically, the IVI is only considered in each cluster.

Therefore, vehicle division module first combines the kth vehicle's current position $x_{t,k}$ in state (8) and predicted position $\hat{x}_{t+1,k}$ in (14) into $\overline{\mathbf{x}}_{t,k} = [x_{t,k}, \hat{x}_{t+1,k}]^{\mathrm{T}}$, and then decompose the K vehicles into U clusters. The uth cluster's state can be written as

$$\mathbf{s}_t^u = \left[\overline{\mathbf{x}}_{t,1}^u, \overline{\mathbf{x}}_{t,2}^u, \cdots, \overline{\mathbf{x}}_{t,K^u}^u\right], \tag{16}$$

where K^u is the number of vehicles in the uth cluster. Hence, each cluster can respectively select their own joint beams

$$\mathbf{a}_t^u = \left[\phi_{t,1}^u, \phi_{t,2}^u, \cdots, \phi_{t,K^u}^u\right]^{\mathrm{T}}. \tag{17}$$

Moreover, the instance reward (10) can also be decomposed into each cluster's average successful transmission probability r_t^u in the tth selection period, i.e.,

$$\begin{aligned} r_t &= \frac{1}{KJ}\sum_{k=1}^{K}\sum_{j=1}^{J} a_{t,j,k}(\phi_{t,k}) \\ &= \frac{1}{KJ}\sum_{u=1}^{U}\sum_{k^u=1}^{K^u}\sum_{j=1}^{J} a^u_{t,j,k^u}(\phi^u_{t,k^u}) \\ &= \sum_{u=1}^{U}\frac{K^u}{K}\frac{1}{K^uJ}\sum_{k^u=1}^{K^u}\sum_{j=1}^{J} a^u_{t,j,k^u}(\phi^u_{t,k^u}) \\ &= \sum_{u=1}^{U} p^u r_t^u, \end{aligned} \tag{18}$$

where $p^u = K^u/K$ is the proportion the number of vehicles in the uth cluster to the total number of vehicles K.

Then the original objective in (12) can also be decomposed into multiple sub-objective $V^{\pi^u}(\mathbf{s}^u)$ for multiple clusters, i.e.,

$$\begin{aligned} V^{\pi}(\mathbf{s}) &= \mathbb{E}_{\tau\sim p(\tau)}\left[\sum_{t=0}^{T}\gamma^t r_t \mid \tau_{\mathbf{s}_0} = \mathbf{s}\right] \\ &= \mathbb{E}_{\tau\sim\prod_{u=1}^{U} p(\tau^u)}\left[\sum_{t=0}^{T}\gamma^t\sum_{u=1}^{U} p^u r_t^u \mid \tau_{\mathbf{s}_0} = \mathbf{s}\right] \\ &= \sum_{u=1}^{U} p^u \mathbb{E}_{\tau^u\sim p(\tau^u)}\left[\sum_{t=0}^{T}\gamma^t r_t^u \mid \tau^u_{\mathbf{s}_0^u} = \mathbf{s}^u\right] \\ &= \sum_{u=1}^{U} p^u V^{\pi^u}(\mathbf{s}^u), \end{aligned} \tag{19}$$

where $\tau^u = [\mathbf{s}_0^u, \mathbf{a}_0^u, r_0^u, \mathbf{s}_1^u, \mathbf{a}_1^u, r_1^u \cdots, \mathbf{s}_T^u, \mathbf{a}_T^u, r_T^u]$ is the uth cluster's trail under beam selection strategy π^u.

Therefore, the optimal beam vectors for K vehicles can be alternatively obtained by selecting the optimal beams for the vehicles in each cluster, i.e.,

$$\begin{aligned} \pi^* &= [\pi^{1^*}, \pi^{2^*}, \cdots, \pi^{U^*}] \\ &= [\arg\max_{\pi^1} V^{\pi^1}(\mathbf{s}^1), \arg\max_{\pi^2} V^{\pi^2}(\mathbf{s}^2), \cdots, \arg\max_{\pi^U} V^{\pi^U}(\mathbf{s}^U)]. \end{aligned} \tag{20}$$

After vehicle division module, the optimal beam selection for the vehicles only to be studied in each cluster.

4.3 Beam Selection Module

According to (17), the increasing action set of each cluster causes gradient explosion. Based on Bellman equation $V^{\pi}(\mathbf{s}) = \mathbb{E}_{\mathbf{a}\sim\pi(\mathbf{a}|\mathbf{s})}Q^{\pi}(\mathbf{s},\mathbf{a})$

and value-decomposition networks (VDN) decomposing equation $Q^{\pi}(\mathbf{s},\mathbf{a}) \approx \sum_{n=1}^{N} Q_n^{\pi}(\mathbf{s},a_n)$ under the condition that $r(\mathbf{s},\mathbf{a}) = \sum_{n=1}^{N} r(\mathbf{s},a_n)$, we can reduce the complexity of the PPJS scheme by decomposing the objective of each cluster into multiple objectives of single vehicle, i.e.,

$$\begin{aligned}
&V^{\pi^u}(\mathbf{s}^u)\\
&= \mathbb{E}_{\tau^u_{0:T}\sim p(\tau^u)}\left[\sum_{t=0}^{T}\gamma^t r_t^u \mid \tau^u_{\mathbf{s}_0^u} = \mathbf{s}^u\right]\\
&= \mathbb{E}_{\tau^u_{0:T}\sim p(\tau^u)}\left[r(\mathbf{s}^u,\mathbf{a}^u,\mathbf{s}^{u'}) + \gamma\sum_{t=1}^{T}\gamma^{t-1} r_t^u \mid \tau^u_{\mathbf{s}_1^u} = \mathbf{s}^{u'}\right]\\
&= \mathbb{E}_{\mathbf{a}^u\sim\pi^u(\mathbf{a}^u|\mathbf{s}^u)}\mathbb{E}_{\mathbf{s}^{u'}\sim\pi^u(\mathbf{s}^{u'}|\mathbf{s}^u,\mathbf{a}^u)}\\
&\quad\left[\frac{1}{K^u J}\sum_{k^u=1}^{K^u}\sum_{j=1}^{J} a^u_{t,j,k^u}(\phi^u_{t,k^u}) + \gamma\mathbb{E}_{\tau^u_{1:T}\sim p(\tau^u)}\sum_{t=1}^{T}\gamma^{t-1} r_t^u \mid \tau^u_{\mathbf{s}_0^u} = \mathbf{s}^{u'}\right]\\
&= \frac{1}{K^u}\sum_{k^u=1}^{K^u}\mathbb{E}_{\mathbf{a}^u\sim\pi^u(\mathbf{a}^u|\mathbf{s}^u)}\\
&\quad\mathbb{E}_{\mathbf{s}^{u'}\sim\pi^u(\mathbf{s}^{u'}|\mathbf{s}^u,\mathbf{a}^u)}\left[r^u_{k_u} + \gamma\mathbb{E}_{\tau^u_{1:T}\sim p(\tau^u)}\sum_{t=1}^{T}\gamma^{t-1} r_t^u \mid \tau^u_{\mathbf{s}_0^u} = \mathbf{s}^{u'}\right]\\
&\triangleq \frac{1}{K^u}\sum_{k^u=1}^{K^u} Q^{\pi^u}_{k_u}(\mathbf{s}^u,\mathbf{a}^u)\\
&\approx \frac{1}{K^u}\sum_{k^u=1}^{K^u} Q^{\pi^u}_{k_u}(\mathbf{s}^u,\phi^u_{k_u}),
\end{aligned} \tag{21}$$

where $r^u_{k_u}$ is the k_uth reward in the uth cluster, $\phi^u_{k_u}$ is the k_uth beam direction in the uth cluster. According to (21), we can know that the optimal beam selection strategy for each cluster is equivalent to choose the optimal beam for each vehicle, respectively, i.e.,

$$\begin{aligned}
\pi^{u*} &= \arg\max_{\pi^u} V^{\pi^u}(\mathbf{s}^u)\\
&= \left[\arg\max_{\phi^u_1} Q_1^{\pi^u}(\mathbf{s}^u,\phi^u_1), \arg\max_{\phi^u_2} Q_2^{\pi^u}(\mathbf{s}^u,\phi^u_2), \cdots, \arg\max_{\phi^u_{K_u}} Q_{K_u}^{\pi^u}(\mathbf{s}^u,\phi^u_{K_u})\right]\\
&= [\phi_1^{u*}, \phi_2^{u*}, \cdots, \phi_{K_u}^{u*}].
\end{aligned} \tag{22}$$

Based on the above analysis, the uth cluster network is shown in Fig. 2(b). To realize the beam selection state of the proposed PPJS scheme, deep Q-learning network structure is employed in this paper. The beam selection network has two specialized structures for the purpose of stabilization. One is experience replay memory unit, which stores the set of beam selection transition $[\mathbf{s}_t^u, \mathbf{a}_t^u, r_t^u, \mathbf{s}_{t+1}^u]$ for the random-batch-training of clusters. The other is the quasi-static beam

selection network, which contains two sub-networks with the same network structures, which is designed based on MLP by us, i.e., policy network and target network with different parameters θ and θ^-, respectively. The policy network is used to select beams in real time according to V^{π_θ}; the target network is used to provide a target y for the policy network's parameters update. And then we can obtain the loss function $(y - V^{\pi_\theta})^2$. Based on the loss function, we can use the gradient descent method to get the parameters θ, and the target network parameter θ^- is updated by θ every t_c steps. From the output layer of Fig. 2, we can see that beam selection is significantly simplified according to (19) and (21).

Algorithm 1: PPJS scheme

Input: $\mathbf{s}_t^u$ for the uth cluster ($u = \{1, 2, \cdots, U\}$).

for *cluster u=1 to U* **do**
  Initialize buffer with capacity $C^u = C$, the policy parameter θ^u, the target network parameter θ^{u-}.
end
for *episode=1 to M* **do**
  for *cluster u=1 to U* **do**
    Initialize $\mathbf{s}_t^u = \left[\overline{\mathbf{x}}_{t,1}^u, \overline{\mathbf{x}}_{t,2}^u, \cdots, \overline{\mathbf{x}}_{t,K^u}^u\right]$.
    for k_u *=1 to* K_u **do**
      Generate a random number ξ in [0,1],

$$\phi_{t,k_u}^u = \begin{cases} \text{random beam direction } \phi_{t,k_u}^u \in \Phi & \xi \geq \epsilon, \\ \text{argmax } Q_{a_{t,k_u}^u}^u\left(\mathbf{s}_t^u, \phi_{t,k_u}^u; \theta_t^u\right) & \text{otherwise .} \end{cases}$$

    end
    Form the uth cluster action $\mathbf{a}_t^u = \left[\phi_{t,1}^u, \phi_{t,2}^u, \cdots, \phi_{t,K^u}^u\right]^{\mathrm{T}}$.
  end
  Execute action $\mathbf{a}_t = \left[\mathbf{a}_t^1, \mathbf{a}_t^2, \cdots, \mathbf{a}_t^U\right]$ to get the next selection period state $\mathbf{s}_{t+1}$ of K vehicles, and observe U clusters reward $\left[r_t^1, r_t^2, \cdots, r_t^U\right]$.
  Feed $\mathbf{s}_{t+1}$ into the trained trajectory prediction module and obtain each cluster's own state $\mathbf{s}_{t+1}^u = \left[\overline{\mathbf{x}}_{t+1,1}^u, \overline{\mathbf{x}}_{t+1,2}^u, \cdots, \overline{\mathbf{x}}_{t,K^u}^u\right]$.
  for *cluster u=1 to U* **do**
    Store selection transition $\left(\mathbf{s}_t^u, \mathbf{a}_t^u, r_t^u, \mathbf{s}_{t+1}^u\right)$ in M_u.
    Sample random minibatch of transitions $\left(\mathbf{s}_i^u, \mathbf{a}_i^u, r_i^u, \mathbf{s}_{i+1}^u\right)$ from M_u.
    Set $y_i^u = \begin{cases} r_i^u & \text{if } i+1 = T, \\ r_i^u + \gamma \max V^{\pi^{\theta^{u-}}}\left(\mathbf{s}_{i+1}^u\right) & \text{otherwise} \end{cases}$
    Perform a gradient descent on $(y_i^u - V^{\pi^{\theta^u}}(\mathbf{s}_i^u))^2$ with respect to the network parameters θ^u.
    Reset $\theta^{u-} = \theta^u$ every t_{s} steps .
  end
end

5 Simulation Results and Performance Analysis

In this section, we first give the simulation parameters and then evaluate the performance of the proposed PPJS scheme under different scenarios.

Table 1. Environment parameters and network parameters.

Trajectory prediction network parameter	Value
Input size N_{i}	2
Output size N_{o}	2
Cluster size N_{c}	10
Learning rate l_{t}	0.06
Batch size L	500

Sub-selection network parameter	Value
Learning rate l_{s}	0.01
Discount factor γ	0.9
Threshold ϵ	0.9
Update time t_{s}	300
Batch size N_{b}	32
Memory size C	500

Environment parameter	Value
Number of antennas N_{BS}	32
Number of uplink intervals I	30
Number of downlink intervals J	70
Time interval T_{i}	10 ms
Road length h_{r}	200 m
Distance between BS and road h_{d}	100 m

5.1 Parameter Setup

The default parameters are listed in Table 1. Considering a practical traffic scenario [12,13], both the vehicle speed and distance distributions follow the truncated normal distribution. The relationship between them can be characterized by traffic flow theory [14], that is,

$$V = V_m \ln\left(\frac{K_m}{K}\right). \tag{23}$$

Moreover, the vehicles could speed up and down in order to avoid crash between vehicles.

5.2 Results and Analysis

Figure 3 depicts the successful transmission probability versus the number of iterations. We can see that the successful transmission probability gradually converges with the increasing number of iterations, which validates the convergence of the proposed PPJS.

The successful transmission probabilities versus different traffic parameters are shown in Fig. 4 under three beam tracking schemes. Apart from the proposed PPIA scheme, the DQN and position-based beam selection schemes are adopted for comparison, i.e., the beams for vehicles are directly selected based on DQN network and the current vehicle positions, respectively. Fig. 4 indicates that the proposed PPIA scheme performs better than the other two schemes,

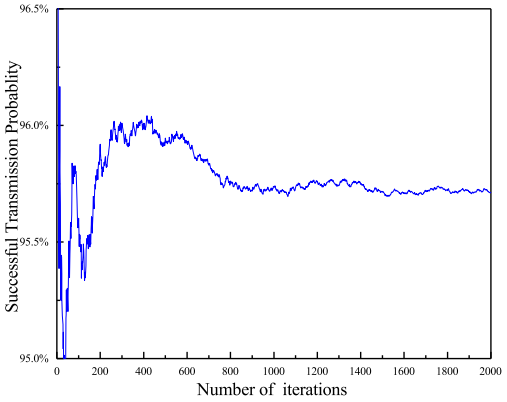

Fig. 3. Successful transmission probability versus iteration number

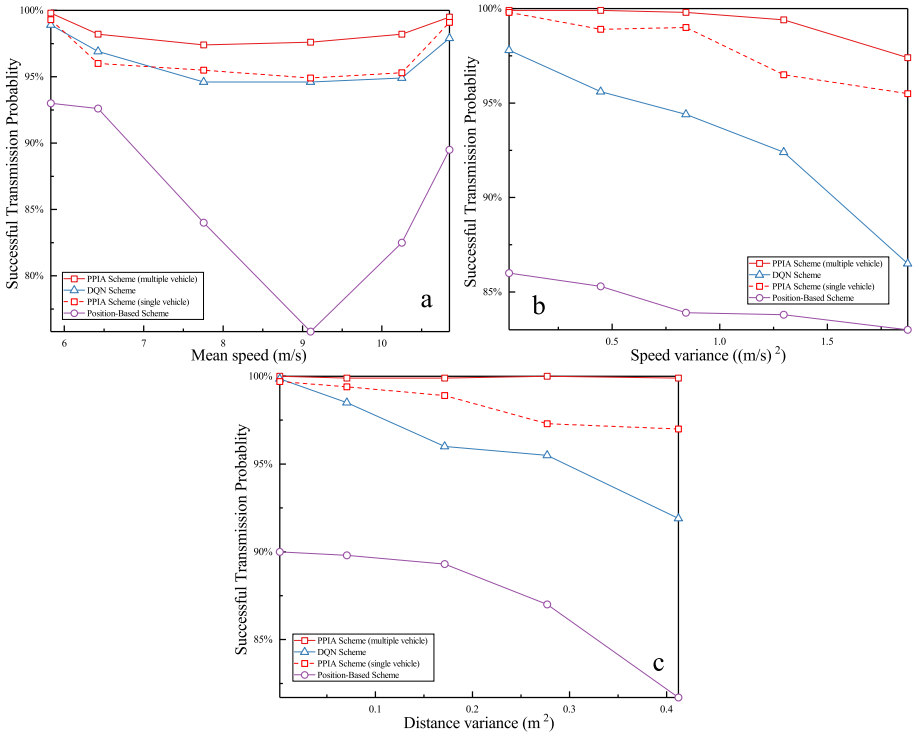

Fig. 4. The performance under different traffic parameters.

because both the beam coverage and IVI are taken into account. In Fig. 4(a), the successful transmission probability first increases and then decreases with the increasing vehicle mean speed. Because for a given number of antennas at the

BS, the beam width is fixed, then the average coverage gain will drop with the increasing speed mean of vehicles; meanwhile the mean distance between vehicles will increase based on traffic theory and therefore the IVI could be mitigated. In the low speed mean, the gain degradation dominates the performance, while the IVI mitigation dominates the performance in the high speed mean. Figure 4(b) indicates that the successful transmission probability becomes worse with the increasing speed variance. This is because some inter-vehicle distances decrease with the increasing speed variance, which results in severer IVI among vehicles. Although other inter-vehicle distances become large, the contribution of the decreasing IVI is tiny. Figure 4(c) also shows a decreasing tendency in successful transmission probability with the increasing vehicle distance variance, which could be explained by the same reason for Fig. 4(b).

Figure 5 illustrates the successful transmission probability versus the number of antennas at the BS under different moving speed setups. Based on the antenna theory, the beam width decreases with the increasing number of antennas. The wide beam could achieve better coverage with high interference, while the narrow beam has low IVI with limited coverage region. During a beam selection period, the coverage and beam gain should match the moving speed and inter vehicle distance. Therefore, the performance first increases and then decreases with the increasing number of antennas in Fig. 5.

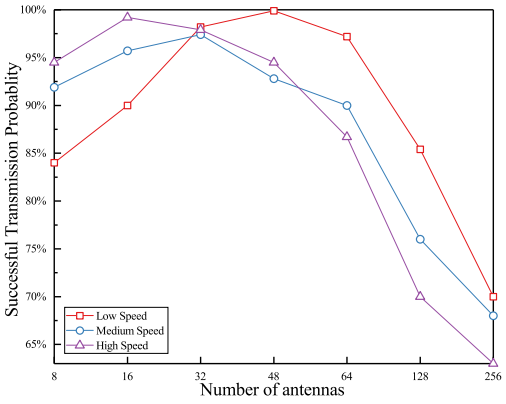

Fig. 5. Successful transmission probability versus the number of antennas.

Figure 6 indicates the relationship between the successful transmission probability and the size M of the beam direction set under different moving speed scenarios. With the increasing number of candidate beams in the beam direction set, the transmission performance gradually becomes better. Because the beam direction set with increasing number of discrete beams could gradually approximate to the continuous beams, then the optimal beam selection could be realized at cost of slight complexity increase.

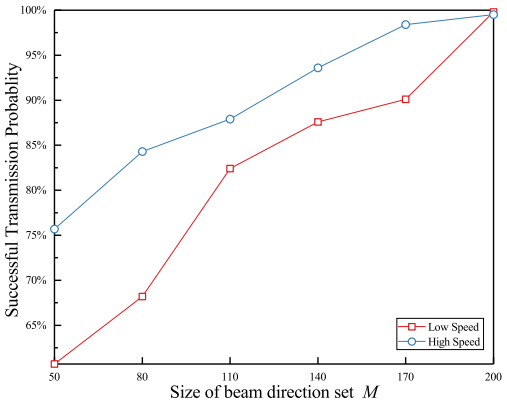

Fig. 6. Successful transmission probability versus the number of discrete beams.

6 Conclusion

This paper considers a scenario that an IoV system with mmWave and massive MIMO serves multiple mobile vehicles by periodical beam selection. In order to improve the transmission performance influenced by both beam coverage and IVI on our considered scenario, an intelligent PPJS scheme is proposed. The PPJS scheme first predicts the future trajectories of vehicles; and then selects beams with a low complexity by dividing the vehicles into decoupled clusters and decomposing the cluster's joint objective. Simulation results indicate that the proposed PPJS scheme performs better than other schemes through balancing the beam coverage and interference. Moreover, the effects of traffic parameters, beam width, and the size of discrete beam direction set on the performance are studied in details, which could provide a reference for practical systems.

Acknowledgment. This work was supported in part by the China Nature Science Funding under Grant 61731004.

References

1. Bujari, A., Gottardo, J., Palazzi, C.E., Ronzani, D.: Message dissemination in urban IoV. In: 2019 IEEE/ACM 23rd International Symposium on Distributed Simulation and Real Time Applications (DS-RT), pp. 1–4 (2019). https://doi.org/10.1109/DS-RT47707.2019.8958708
2. Hou, Z., She, C., Li, Y., Zhuo, L., Vucetic, B.: Prediction and communication co-design for ultra-reliable and low-latency communications. IEEE Trans. Wireless Commun. **19**(2), 1196–1209 (2020). https://doi.org/10.1109/TWC.2019.2951660
3. Bencivenni, C., Gustafsson, M., Haddadi, A., Zaman, A.U., Emanuelsson, T.: 5G mmWave beam steering antenna development and testing. In: 2019 13th European Conference on Antennas and Propagation (EuCAP), pp. 1–4 (2019)

4. Ali, M.Y., Hossain, T., Mowla, M.M.: A trade-off between energy and spectral efficiency in massive MIMO 5G system. In: 2019 3rd International Conference on Electrical, Computer Telecommunication Engineering (ICECTE), pp. 209–212 (2019). https://doi.org/10.1109/ICECTE48615.2019.9303551
5. Raza, A., Junaid Nawaz, S., Wyne, S., Ahmed, A., Javed, M.A., Patwary, M.N.: Spatial modeling of interference in inter-vehicular communications for 3-D volumetric wireless networks. IEEE Access **8**, 108281–108299 (2020). https://doi.org/10.1109/ACCESS.2020.3001052
6. Zhao, L., Zhao, H., Zheng, K., Xiang, W.: Massive MIMO in 5G Networks: Selected Applications. Springer, Heidelberg (2018). https://doi.org/10.1007/978-3-319-68409-3
7. Zheng, K., Zhao, L., Mei, J., Shao, B., Xiang, W., Hanzo, L.: Survey of large-scale MIMO systems. IEEE Commun. Surv. Tutor. **17**(3), 1738–1760 (2015). https://doi.org/10.1109/COMST.2015.2425294
8. Va, V., Vikalo, H., Heath, R.W.: Beam tracking for mobile millimeter wave communication systems. In: 2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP), pp. 743–747 (2016). https://doi.org/10.1109/GlobalSIP.2016.7905941
9. Liu, Y., Jiang, Z., Zhang, S., Xu, S.: Deep reinforcement learning-based beam tracking for low-latency services in vehicular networks. In: ICC 2020–2020 IEEE International Conference on Communications (ICC), pp. 1–7 (2020). https://doi.org/10.1109/ICC40277.2020.9148759
10. Jeong, J., Lim, S.H., Song, Y., Jeon, S.W.: Online learning for joint beam tracking and pattern optimization in massive MIMO systems. In: IEEE INFOCOM 2020 - IEEE Conference on Computer Communications, pp. 764–773 (2020). https://doi.org/10.1109/INFOCOM41043.2020.9155475
11. Tran, X.V., Pham, V.H.: An analytical method for calculating the limitation of beam scanning in uniform linear array (ULA). In: 2009 International Conference on Advanced Technologies for Communications, pp. 257–260 (2009). https://doi.org/10.1109/ATC.2009.5349450
12. Cvetek, D., Muštra, M., Jelušić, N., Abramović, B.: Traffic flow forecasting at micro-locations in urban network using bluetooth detector. In: 2020 International Symposium ELMAR, pp. 57–60 (2020). https://doi.org/10.1109/ELMAR49956.2020.9219023
13. Dai, G., Ma, C., Xu, X.: Short-term traffic flow prediction method for urban road sections based on space-time analysis and GRU. IEEE Access **7**, 143025–143035 (2019). https://doi.org/10.1109/ACCESS.2019.2941280
14. Abadi, A., Rajabioun, T., Ioannou, P.A.: Traffic flow prediction for road transportation networks with limited traffic data. IEEE Trans. Intell. Transp. Syst. **16**(2), 653–662 (2015). https://doi.org/10.1109/TITS.2014.2337238