Resource Allocation in NR-V2X Mode 2 Using Multi Agent DQN

Insung Lee, Duk Kyung Kim†
Department of Electrical and Computer Engineering, Inha University, Incheon, Republic of Korea
† *Corresponding Author*
insung237@inha.edu, kdk@inha.ac.kr

***Abstract*—V2X communication has been studied to increase road safety and traffic efficiency, and the most recently standardized technology is NR-V2X developed by third generation partnership project (3GPP). NR-V2X supports two modes of communication mode 1 and mode 2. In mode 2, vehicles reserve the resources based on their local observations using semi-persistent scheduling (SPS). In this method, if two or more vehicles select the same resource, a continuous resource collision occurs, and it makes communication performance greatly degraded. To overcome this, we propose a resource allocation method using multi-agent reinforcement learning (MARL). As agents, vehicles that transmit periodic cooperative awareness messages (CAM) are modeled. The state is composed of the received signals strength indicator (RSSI) that the agents received from each resource, and we set the total sum rate of all agents as the shared reward. The proposed method is compared with the random resource allocation in a highway scenario. The results show that the proposed method outperforms in terms of throughput performance.**

***Index Terms*—New Radio vehicle-to-everything (NR-V2X), Resource Allocation, Deep reinforcement Learning**

I. Introduction

Vehicular communication technology has steadily emerged for the purpose of increasing road safety and traffic efficiency. Accordingly, IEEE introduced Dedicated Short Range Communication (DSRC), the first vehicular communication standard in 2010[1]. However, DSRC couldn't show good performance in terms of reliability and Quality-of-Service (QoS) due to the dynamic characteristics of the vehicular environment[2]. Therefore, 3GPP introduced a new vehicular communication standard NR-V2X in release 16, that can achieve higher performance and also be able to cover every V2X communication such as Vehicles-to-Pedestrian (V2P), Vehicles-to-Vehicles (V2V), and Vehicles-to-Infrastructure (V2I). In V2X communication, the resource allocation method greatly affects the V2X communication performance, and 3GPP proposes two resource allocation methods, Mode 1 and Mode 2[3].

- Mode 1: A technique in which vehicles within range of a Base Station (BS) are allocated resources with the support of the BS.
- Mode 2: SPS technique in which each vehicle detects resources by itself without BS support, and selects an appropriate resource to utilize repeatedly for a certain period of time.

However, since wireless resources are always in a non-stationary environment, there are always performance limitations in mode 2[4]. In order to overcome the above limitations, a lot of research has been conducted on resource allocation methods using reinforcement learning (RL) in the V2X environment, and these are some examples of which are as follows.

H. Ye et. al. proposed a MARL DQN structure that can be applied to both V2V unicast and broadcast communication where a V2I link exists [5]. Each V2V link was modeled as a DQN agent, and each V2V transmitter vehicle was modeled as a DQN agent. Their research was a way to improve not only the communication performance of V2V but also the V2I. However, in the learning process of DQN, the throughput and sum rate of V2V communication were not considered, only latency constraints were considered as a reward. And also, since they didn't show throughput-related performance in a broadcast environment, it is difficult to measure the throughput performance in a V2V broadcast environment. M. M. Saad et. al. proposed a discrete power control technique using RL in LTE-based C-V2X communication [6]. They modeled each transmitting vehicle as an agent of DQN and utilized a multi-agent reinforcement learning (MARL) DQN structure that receives shared rewards as feedback. However, since communication performance degradation due to interference is inevitable in adjusting the transmission power only, it is more inefficient than selecting different sub-channels within wireless resources. Y. Yuan et. al. proposed a method of performing both resource selection and continuous power control for V2V by utilizing DQN and DDPG together in a V2V unicast environment where a V2I link exists [7]. They regarded the transmitting and receiving vehicles as a pair of V2V links, and modeled them as DQN agents. They also utilized the MARL DQN structure that receives shared rewards as feedback. However, since they considered unicast V2V link only, didn't mention the broadcast environment.

In this paper, we have proposed a resource allocation method using MARL DQN in an NR-V2X broadcast environment. We aimed to improve the overall throughput performance of the NR-V2X broadcast scenario by modeling each transmitting vehicle as an agent and setting the average throughput of all vehicles as a reward. In the remainder of this paper is organized as follows. Section II presents the system model. Section III presents deep reinforcement learning-based mechanism we propose. Section IV presents the performance evaluation by a comparison of throughput performance with the random resource selection scheme according to whether or not NLOSv is considered. Finally, the conclusion is drawn in Section V.

II. System Model

We assumed a 6-lane highway environment in which N_{Tx} transmitting UEs among N UEs share M resources in the same resource pool. The lane width is 4m on a 1km long highway, and vehicles move to the right on the upper 3 lanes and on the left on the lower 3 lanes. We considered two scenarios where the speed of vehicles is 30 km/h and 70 km/h, and the location of the vehicles is updated every 50ms. If the vehicle crosses the road boundary, it is set to appear again on the opposite side of the same lane so we keep the number of vehicles on the road the same.

There are N_{Tx} transmitting UEs which periodically broadcast CAM messages using M resources, and these transmitting UEs are not changed during DQN training session. Fig 1 below shows the road environment when $N_{Tx} = 4$.

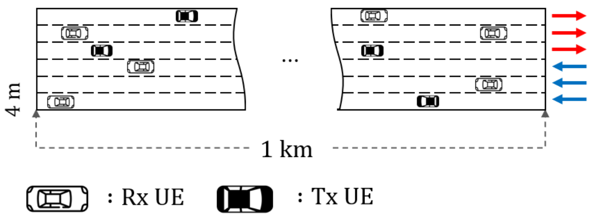

Fig. 1. Assumed road environment

We assumed UE_{Tx}^{i} as i^{th} transmitting UE and UE_{Rx}^{n} as n^{th} receiving UE. In this case, when we set the throughput between UE_{Tx}^{i} and UE_{Rx}^{n} as $THP_{i,n}$, it can be expressed as below.

$$THP_{i,n} = W \cdot \log_2\left(1 + SINR_{i,n}\right) \tag{1}$$

W means the bandwidth of the utilized wireless resource, and $SINR_{i,n}$ means the SINR between transmitter UE_{Tx}^{i} and receiver UE_{Rx}^{n} . Interference used in calculating SINR considers only CAM messages of other transmitting UEs using the same resource. When we set the transmission power of UEs as P_{Tx} and channel gain between UE_{Tx}^{i} and UE_{Rx}^{n} as $h_{i,n}$, $SINR_{i,n}$ can be expressed as below. In this case, UE_{Tx}^{i} broadcasts its CAM messages using m^{th} resource.

$$SINR_{i,n} = \frac{P_{Tx} \cdot h_{i,n}}{\sigma^2 + \sum_{\substack{k=1 \\ k \neq i}}^{N_{Tx}} \rho_k[m] \cdot P_{Tx} \cdot h_{k,n}} \tag{2}$$

σ^2 means the noise power, and $\rho_k[m]$ is an indicator which means whether UE_{Tx}^{k} has used the $m \in [1,2,\ldots,M]^{th}$ resource or not. If UE_{Tx}^{k} has used the m^{th} resource, $\rho_k[m]$ represents 1, otherwise 0. And also, $RSSI_i[m]$, which means the RSSI received by UE_{Tx}^{i} from m^{th} resource can be expressed as below. At this time, the received power for the message transmitted by itself is excluded from RSSI measurement.

$$RSSI_i[m] = \sum_{\substack{k=1 \\ k \neq i}}^{N_{Tx}} \rho_k[m] \cdot P_{Tx} \cdot h_{k,i} \tag{3}$$

III. Deep Reinforcement Learning

We modeled N_{Tx} number of transmitting UEs as agents, and each agent performs an action in the direction of maximizing the Q-value through its own state. The agent can notice the least utilized resource through its own state information, and appropriate action can be performed within the action space $\{1,2,\ldots,M\}$. When we set the state of agent i as $state_i$, it can be expressed as below.

$$\text{State}_i = \{RSSI_i[1], \ldots, RSSI_i[M]\} \tag{4}$$

After all the agents take action, the shared reward is calculated as the mean of the average throughput of all agents for the receiving vehicles. For the throughput calculation, only vehicles within the communication range of each transmitting UE are considered. When we set the number of receiving UEs located within the communication range of UE_{Tx}^{i} as N_i, the average throughput $THP_{i,mean}$ of UE_{Tx}^{i} can be expressed as below.

$$THP_{i,mean} = \frac{1}{N_i}\left(\sum_{r=1}^{N_i} THP_{i,r}\right) \tag{5}$$

And shared reward R_{sh} can be expressed as below.

$$R_{sh} = \frac{1}{N_{Tx}}\left(\sum_{i=1}^{N_{Tx}} THP_{i,mean}\right) \tag{6}$$

IV. Performance Evaluation

In the highway scenario shown in Fig. 1, we compared the communication performance between the case of resource allocation using the proposed DQN method and the case of random resource allocation. At this time, the performance was shown by changing the number of vehicles, the communication range of the transmission vehicle, and whether or not NLOSv was considered. NLOSv was considered by referring to the ETSI TR document [8], assuming that each vehicle is a circle with a radius of 1.5m, and attenuating the channel gain by 6.9dB for each number of vehicles existing between the transmitting and receiving UE.

As a channel model, the WINNER + B1 channel was used for path loss calculation, and time-correlated fast fading was considered. DQN was trained for 10,000 episodes, and each episode consisted of 50-time steps. Detailed parameters are shown in table I below.

TABLE I

Simulation Parameters

Center frequency	5.9 GHz
No of Resources M	4
Noise power σ^2	-114 dBm
Channel	WINNER + B1
No of Vehicles N	80, 160
No of Tx Vehicles N_{Tx}	4
Communication range	{50, 100}m
Vehicle speed	{30, 70}km/h

Fig 2 shows the CDF distribution of the throughput when 80 vehicles, 50 m of communication range, and 30 km/h of vehicle's speed are set. The red line indicates resource allocation using the proposed DQN, the black line indicates random resource allocation, solid line, and the dashed line indicates whether NLOSv is considered or not. Before and after NLOSv's consideration, DQN showed much better performance than the random resource allocation. The average throughput of DQN and random resource allocation in the case of not considering NLOSv is 178, 109, and the case of considering NLOSv is 160, 100 Mbps. In both cases, DQN showed about 60% better throughput performance than the random resource allocation.

When comparing the results of the 5%-tile throughput, the performance difference between DQN and random allocation is being more clear. In the case of DQN, it shows about 116 and 39 Mbps before and after the NLOSv consideration. However, the random scheme showed about 0.3 Mbps regardless of NLOSv consideration, which is a significantly low value compared to DQN method. Thus, it is confirmed that the use of DQN method is much better than random resource allocation.

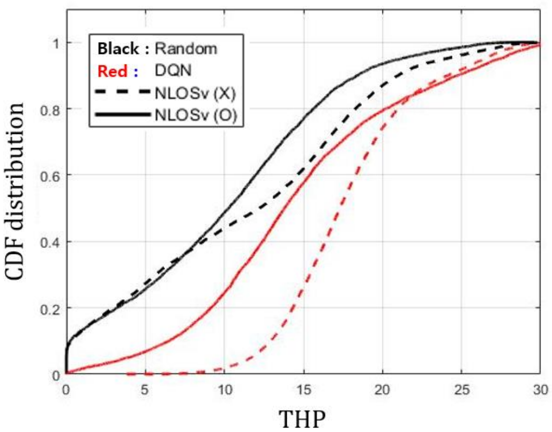

Fig. 2. CDF of **THP**

In addition, we analyzed the change in throughput performance when several parameters related to the environment were changed. Such as the number of vehicles, communication range, and speed of vehicles. Table II shows the average and 5%-tile throughput values for the changes in the parameters mentioned above. In this case, NLOSv was not considered.

Table II shows that the performance degradation was the most severe when the communication range was changed. This is because vehicles far from transmission UE are also used to calculate throughput. In contrast, the parameter that showed the least variation in throughput performance when changed was the vehicle's speed.

TABLE II

THP PERFORMANCE FOR PARAMETERS [Mbps]

	Baseline	Veh num 80 → 160	Comm range (m) 50 → 100	Veh speed (km/h) 30 → 70
Mean	178	170	115	181
5%-tile	116	104	109	119

V. CONCLUSION

In this paper, a resource allocation method using RL in NR-V2X broadcast communication environment is presented. MARL DQN was used as the RL structure, and vehicles that transmit periodic CAM messages were modeled as agents. By comparing the throughput performance of DQN based scheme and random resource allocation in highway scenario, it is confirmed that the DQN based scheme showed much better performance.

ACKNOWLEDGEMENT

This work was supported by the framework of the international cooperation program managed by the National Research Foundation of Korea (NRF-2021K1A3A1A6100 3229), and by IITP grant funded by the Korea government (MSIT) (2022-0-00521, Development of 5G-NR-V2X communication technology to support over Lv.4 autonomous driving service).

REFERENCES

[1] J. B. Kenney, “Dedicated short-range communications (DSRC) standards in the United States,” *Proceedings of the IEEE*, vol. 99, No. 7, pp. 1162-1182, 2011.

[2] W. Sun, E. G. Strom, F. Brannstrom, Y. Sui, and K. C. Sou, “D2D-based V2V communications with latency and reliability constraints,” in *2014 IEEE Globecom Workshops (GC Wkshps)*, pp. 1414–1419, IEEE, 2014.

[3] Z. Ali, S. Lagén, L. Giupponi, and R . Rouil, “3GPP NR V2X mode 2: overview, models and system-level evaluation,” *IEEE Access*, vol. 9, pp. 89554-89579, 2021.

[4] X. He, J. Lv, J. Zhao, X. Hou, and T. Luo, “Design and analysis of a short-term sensing-based resource selection scheme for C-V2X networks,” *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 11209–11222, 2020.

[5] H. Ye, G. Y. Li, and B.-H. F. Juang, “Deep reinforcement learning based resource allocation for V2V communications,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3163–3173, 2019.

[6] M. M. Saad, M. M. Islam, M. A. Tariq, M. T. R. Khan, and D. Kim, “Collaborative Multi-Agent Resource Allocation in C-V2X Mode 4,” In *2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN),* pp. 7-10, IEEE, 2021.

[7] Y. Yuan, G. Zheng, K. K. Wong, and K. B. Letaief, “Meta-reinforcement learning based resource allocation for dynamic V2X communications,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, pp. 8964-8977, 2021.

[8] S.Antipolis, “Intelligent Transportation Systems(ITS); Acess Layer; Part 1: channel Models for the 5,9 GHz frequency band,” ETSI Tech. Rep. document TR103 257-1, V1.1.1, May. 2019.