Collaborative Multi-Agent Resource Allocation in C-V2X Mode 4

Malik Muhammad Saad, Md. Mahmudul Islam, Muhammad Ashar Tariq
Muhammad Toaha Raza Khan, Dongkyun Kim†
School of Computer Science and Engineering, Kyungpook National University, Daegu, Republic of Korea
† *Corresponding Author*
{maliksaad,mislam,tariqashar,toaha,dongkyun}@knu.ac.kr

***Abstract*—Intelligent Transport System (ITS) provides an efficient solution to road safety traffic. To support safety applications, cellular vehicle-to-everything (C-V2X) is developed by third generation partnership project (3GPP). C-V2X support two modes of communication as mode 3 and mode 4. In mode 4, vehicles reserve the resources based on their local observations using semi-persistent scheduling (SPS). If two vehicles, simultaneously select the same resources, it will lead to resource contention. This arises the consensus problem. To overcome this, in this paper we proposed the multi agent collaborative deep reinforcement learning based scheme. A single deep Q network (DQN) is trained for each zone. Each zone is preconfigured with resources which constitute a resource pool. A reward function is shared between the vehicles that belong to the same pool. This approach makes the vehicles to collaborate rather than compete in selecting the resources for their transmission. The proposed scheme is compared with the random resource allocation in C-V2X. The results show that the proposed scheme outperforms even in dense vehicular environment.**

***Index Terms*—Cellular vehicle-to-everything (C-V2X), Semi-Persistent Scheduling (SPS), Distributed Resource Allocation, Deep reinforcement Learning**

I. Introduction

C-V2X is an emerging vehicular technology that support critical safety applications and provide efficient transportation on the road. Dedicated short range communication also known as IEEE802.11p was the first standard introduced in 2010. IEEE802.11p is developed our IEEE802.11 WiFi based technology for vehicular communication. To meet high reliability and dynamic characteristics of vehicular environment, 802.11p based solutions are not satisfactory [1]. Usually, the remote evolutions offer system to vehicular correspondences, which can accept a basic activity in capable resource the board similarly as transportation for the provisioning of quality-of-service (QoS) and quality-of-experience (QoE). 3GPP is among the joining rules for strong correspondences among Vehicles-to-Pedestrian (V2P), Vehicle-to-Vehicle (V2V), Vehicle-to-Infrastructure (V2I), and cutting-edge Vehicle-to-Everything (V2X). The improvement in V2X technologies require multiplexing resources across vehicular frameworks. Regardless, the communication of remote cell systems is somewhat costly regarding inertness for time basic situations in a vehicular system. 3GPP is widely known standard for LTE-V2V communication standard that provides connectivity by exchanging messages with the LTE based infrastructure.

In 2017, C-V2X was first introduced by 3GPP, in its release Rel.14. C-V2X is developed over the LTE Rel.12 device-to-device (D2D) module [2]. Two transmission modes are defined in C-V2X as mode 3 and mode 4. Mode 3 operates in coverage region i.e., vehicles are in the coverage range of enode B (eNB), whereas mode 4 operates both in coverage and out of coverage region. In mode 3 vehicles communicate over the Uu interface. The eNB schedule the resources for vehicles to communicate. In mode 4 vehicles communicate over the PC-5 interface [2]. The resources are schedule by the vehicles themselves for their transmission using SPS. [3].

SPS is prone to hidden node problem, as vehicles reserve the resources based on their local observations. If two vehicles select the same resources simultaneously, will lead to resource collision which results in packet dropped due to interference [4]. Therefore, the problem of resource contention should be address. To improve the SPS algorithm, Yongseok et. al. [3] proposed a mechanism based on three parts as explicit, early, and repeated to enhance SPS. Vehicle will explicitly inform other vehicles in its proximity upon resource selection to avoid resource contention. The selection of resources is made earlier before transmission to let other vehicles, know in advance. This selection is also reserved for next transmissions for multiple times depends upon the vehicle application. Xinxin et. al. [4] proposed a short-term sensing (STS) based resource allocation. However, if vehicles could not find the resources in the short duration of time (i.e., having short sensing window) will increase the latency to find the available resources. Alessandro et. al. [5] optimized the parameters required for resource allocation to support aperiodic messages. Fakhar et. al. [6] proposed a hybrid scheme based on both IEEE 802.11p and C-V2X technology for improving the end-to-end delay and the reliability of the vehicular networks.

Reinforcement learning is embedded in various applications for real time problem solving. Hao et. al. [7], [8] proposed a decentralized resource allocation scheme based on deep reinforcement learning to solve the latency problem of the vehicular communication link. Here the vehicle will act as an agent and based on the current channel condition the agent will select the transmission power and frequency at

each time slot. Oshri et. al. [9] proposed a deep-Q learning-based dynamic resource allocation algorithm to improve the network performance in multichannel wireless networks. Since each vehicle is model as an agent, their proposed scheme is computationally inefficient and not scalable.

Real systems are non-stationary, have safety constraints and have negative effects of poor actions, unlike the simulated environment. In this paper we have proposed the collaborative deep reinforcement learning based mechanism for resource selection in C-V2X to complement SPS mode 4. In our scheme, the reward is shared with the vehicles belonging to the same pool of resources to make the environment collaborative rather than competitive. This will also mitigate the non-stationary problem and since centralized training is performed, the proposed scheme is robust and scalable. In the remainder of this paper section II presents the system model. In section III the deep reinforcement learning based mechanism is presented. In section IV proposed scheme is evaluated and compared with the random resource allocation mechanism in C-V2X. Finally, the conclusion is drawn in section V.

II. System Model

We consider $M = \{1, 2, 3, \ldots, M\}$ vehicles, which share the common pool of resources as shown in equation (1). Resources are distributed in a pool over two dimensional frequency and time space. Frequency domain is divided into subchannels and time domain into subframes. In C-V2X mode 4, vehicles communicate with each other over the PC-5 interface. For the exchange of cooperative awareness messages (CAMs), vehicles reserve the resources for its transmission from the pool. Data packets are transmitted over the physical sidelink shared channel (PSSCH), whereas sidelink control information (SCI) is transmitted across the each associated data packet over physical sidelink control channel (PSCCH). SCI carries information such as modulation coding symbols (MCS) which assist in the decoding of data packet. Along with MCS, SCI also carries additional information which assist in SPS. The additional information include retransmission counter (RC), received signal strength information (RSSI) and resource reservation interval (RRI). RC indicates the remaining number of times, the vehicle will perform next transmissions on the periodically reserved slots after each RRI.

$$N = SF * SC \tag{1}$$

According to new standards set by the european telecommunication standard institute (ETSI) [10], [11], implies that generation of CAMs is no more periodic. The generation of CAM depends upon the vehicle speed, heading and direction. From this it can be concluded that the already reserved slots can increase the probability of collision for the aperiodic CAMs transmission. This arises the consensus problem, where vehicles based on their local observations select the resources for their transmission. If two vehicles select the same resources simultaneously, lead to resource contention. We model these constraints into deep reinforcement learning in order to improve the distributed SPS.

We assume that vehicles belonging to a common pool of subchannels, approximate their neighbours position via neighbor discovery as proposed by geo-based scheduling scheme in [12]. Vehicles belonging to common pool of subchannels, share their actions in order to avoid resource contention. Moreover a reward function is shared between vehicles which share the common pool of resources in order to increase the capacity of each V2V link. Equation (2) shows the calculated capacity, where B is the bandwidth of the channel and $SINR$ is the signal-to-interference-plus-noise ratio.

$$C = Blog_2(1 + SINR) \tag{2}$$

The SINR is given in equation (3), for a m^{th} vehicle that wants to transmit a packet. P_m is the m^{th} vehicle transmitting power, h_m is the channel gain of V2V link, σ is the channel noise and P_d is the interfering power from other vehicles in a same pool.

$$SINR = \frac{P_m * h_m}{\sigma^2 + \sum_{i=1}^{M} P_i * h_i} \; i \neq m \tag{3}$$

The reward is modelled in order to satisfy the latency constraint of the V2V link and to minimize the interference between vehicles. The penalty is given to the vehicle if any of the constraint is violated.

III. Deep Reinforcement Learning

Each vehicle is modelled as an agent. The state of the environment observed by the vehicle consist of channel state information C_t, channel gain G_t, $RSSI$, channel indices occupied by the vehicles in the previous time slot from a pool N_{t-1}, RC, number of bits to be transmitted by the vehicle L and the latency associated with the generated packet U. Vehicle chooses an action from the action space set $\{23dBm, 10dBm, 5dBm\}$ based on the current state. Equation (4) shows the state representation.

$$S_t = \{G_t, C_t, RSSI, RC, N_{t-1}, L, U\} \tag{4}$$

Equation (5), shows the reward function. It shows the sum of capacities of V2V link and the penalty is given if latency constraint is violated. In equation (5) λ is the weighting factor, T_0 is the maximum tolerable latency.

$$R = \sum_{i=1}^{M} C_i - \lambda(T_0 – U) \tag{5}$$

Equation (6), shows the discounted cumulative reward. Long term rewards are given more preference rather than relying on a greedy approach. β is in between [0,1], the more it is close to 1, means future rewards are given more preference.

$$R_t = \sum_{i=0}^{\infty} \beta^i r_t \tag{6}$$

The state transitions are modelled using markov decision process (MDP). The state transitions are generated by the

simulation environment. The transition probability from one state into other state depends upon the MDP.

The agent choose the action based on the Q-values. However, with large Q table i.e, having large action state pair, the Q-values converge slowly to the optimum value. To alleviate this, deep Q learning is considered. DQN maps the state to an action instead of storing Q-values in look-up table. A single DQN is trained for each pool. Vehicles that belong to the same pool, share the reward function and access the actions taken by the other vehicles. The reward function is shared between vehicles, so that the vehicles collaborate to increase the capacity of each V2V link rather than compete. This parameter sharing approach in centralized training assist favorable computation and is robust to large scale scenarios.

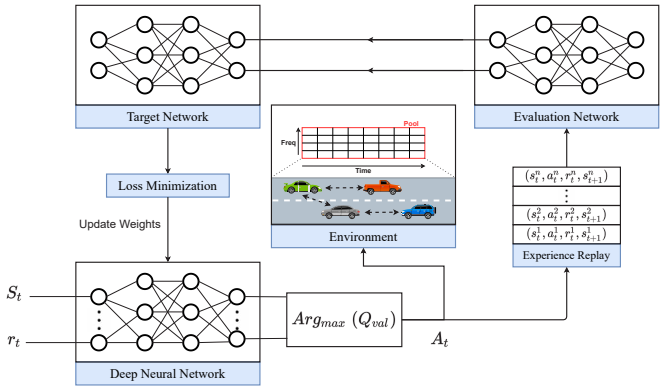

Fig. 1. Deep Reinforcement Learning based Resource Selection in C-V2X

Deep Q networks (DQN) consist of neural network along with the Q learning as shown in Fig 1. The evaluation and target network is used for stabilize learning. The weights of the Q network are updated after each iteration to reduce the loss as given in equation (7).

$$loss = \left\{(r_t + R_t * max'_a Q'(s_t', a_t') - Q(s_t, a_t))\right\}^2 \quad (7)$$

The intuition behind calculating the loss is to feedback through backward propagation process and update the weights of the neurons using gradient descent process. This can reduce the loss and optimize the Q-values.

IV. Performance Evaluation

We considered a Manhattan grid scenario of $750X250m^2$. It consist of 3 lanes urban scenario with both line of sight (LOS) and non-line of sight (NLOS) channel model. The WINNER + B1 channel is considered for the evaluation. A deep neural network is built with three hidden layers. Table I, shows the hyper parameters and the detailed simulation parameters. DQN is trained for thousands of episodes. The centralized training is performed i.e., agents that belong to same pool of subchannels shared the reward. Once the training is performed decentralize distributed execution takes place.

Fig 2 shows the packet delivery ratio (PDR) with respect to varying payload size. The number of vehicles are kept 20 and the message packet size is varied between 90-1500 bytes. The number of resources are configured to the 25% less than the number of vehicles. The collaborative multi agent reinforcement learning (C-MARL) scheme overall performs better than the random resource allocation. It is observed that with the small packet size, the PDR acheived is more than 96% using C-MARL based scheme. The PDR decreases with the increase of payload size as it results in large transmission time which in turn increases the penalty. However with the increase of payload size the PDR decreased but still it is quite high as compared to random resource allocation scheme.

TABLE I
Simulation Parameters

Center frequency	5.9 GHz
No of hidden layers	3
α	0.001
Noise power	-114dBm
Transmission delay budget	100ms
Channel	WINNER+ B1
No of vehicles	20-200
Transmission power	$\{23, 15, 5\}$dBm
message packet Size	$\{90, 300, 450, 600, 900, 1200, 1500\}$ Bytes

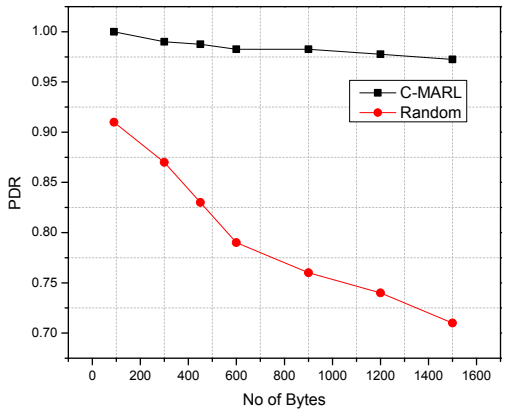

Fig. 2. Impact of Payload Size on PDR

According to ETSI standards the generation of CAMs is aperiodic in nature. Fig 3 shows the PDR with the increase in number of vehicles. The resources are configured in a pool to 35% less than the number of vehicles. The generation of message data bytes are considered aperiodic in nature. It is observed that as the number of vehicles are increased, the PDR decreases in both schemes. However, the over all impact of number of vehicles is less using C-MARL based scheme as compare to random resource allocation scheme.

From Fig 4, it is observed that vehicles choose the power level from the action space set $\{23dBm, 10dBm, 5dBm\}$ for their transmission. The agents learned and the probability of selecting 23dBm is more in less congested scenarios. However, in the congested scenarios, vehicles learn to choose low power

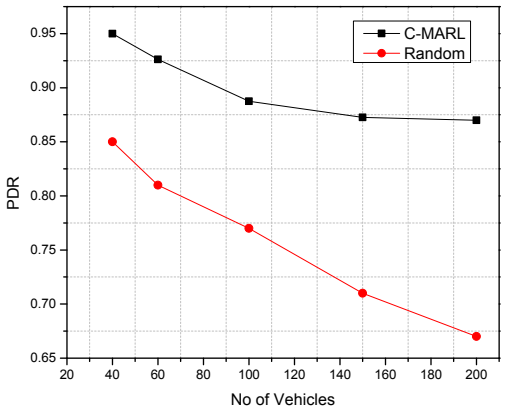

Fig. 3. Impact of number of Vehicles on PDR

level for their transmission in order to avoid interference with neighbour vehicles. Far vehicles learn to reuse the resources for their transmission at low transmit power.

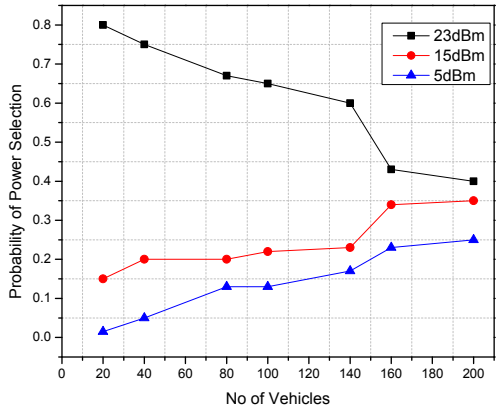

Fig. 4. Probability of Power Selection with increase in number of vehicles

V. Conclusions

In this paper resource allocation problem in C-V2X is modelled using deep reinforcement leaning. Based on their geo-locations, vehicles which belong to same pool share the reward and actions, to collaborate with neighbouring vehicles in terms of resource selection. C-MARL based scheme is compared with the random resource allocation in C-V2X. C-MARL based scheme shows the performance satisfactory in terms of PDR as compare to random resource allocation.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1D1A3B01015510). In addition this research was supported by 2021 Kyungpook National University BK21 FOUR Graduate Innovation Project (International Joint Research Project for Graduate Students).

References

[1] W. Sun, E. G. Ström, F. Brännström, 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.

[2] G. Naik, B. Choudhury, and J.-M. Park, "IEEE 802.11 bd & 5G NR V2X: Evolution of radio access technologies for V2X communications," *IEEE Access*, vol. 7, pp. 70169–70184, 2019.

[3] Y. Jeon and H. Kim, "An explicit reservation-augmented resource allocation scheme for c-v2x sidelink mode 4," *IEEE Access*, vol. 8, pp. 147241–147255, 2020.

[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] A. Bazzi, A. Zanella, and B. M. Masini, "Optimizing the resource allocation of periodic messages with different sizes in LTE-V2V," *IEEE Access*, vol. 7, pp. 43820–43830, 2019.

[6] F. Abbas, P. Fan, and Z. Khan, "A novel low-latency V2V resource allocation scheme based on cellular V2X communications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2185–2197, 2018.

[7] 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.

[8] L. Liang, H. Ye, and G. Y. Li, "Spectrum sharing in vehicular networks based on multi-agent reinforcement learning," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 10, pp. 2282–2292, 2019.

[9] O. Naparstek and K. Cohen, "Deep multi-user reinforcement learning for distributed dynamic spectrum access," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 310–323, 2018.

[10] S. Bartoletti, B. M. Masini, V. Martinez, I. Sarris, and A. Bazzi, "Impact of the generation interval on the performance of sidelink c-v2x autonomous mode," *IEEE Access*, vol. 9, pp. 35121–35135, 2021.

[11] C. . C. C. Consortium, "Survey on ITS-G5 CAM Statistics," Tech. Rep. document TR2052, V1.0.1, Dec. 2018.

[12] R. Molina-Masegosa, M. Sepulcre, and J. Gozalvez, "Geo-based scheduling for c-v2x networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 9, pp. 8397–8407, 2019.