Distributed Joint Congestion Control for V2X Using Multi-agent Reinforcement Learning

Hojeong Lee
Korea University
Seoul, Korea
hojeong0507@korea.ac.kr

Chanwoo Kim
Seoul National University
Seoul, Korea
chanwoo.kim@snu.ac.kr

Eugene Yang
Sungkyunkwan University
Suwon, Korea
cneyang@skku.edu

Hyogon Kim
Korea University
Seoul, Korea
hyogon@korea.ac.kr

***Abstract*—As research on autonomous driving advances, wireless communication between autonomous vehicles to exchange information about their kinematics becomes increasingly important. Specifically, each vehicle periodically sends Basic Safety Messages (BSM) or Cooperative Awareness Messages (CAM) containing the vehicle's position, speed, and acceleration to prevent a car crash. However, the Intelligent Transportation System (ITS) band allocated for vehicle-to-everything (V2X) communication is typically narrow. As a result, when vehicle traffic density increases, the wireless channel can become congested. Although there are standard congestion control algorithms, they are heuristically designed and have been shown to perform poorly due to the difficulties involved in the joint control. In this paper, we show that Multi-Agent Reinforcement Learning (MARL) can be harnessed to produce an effective joint rate and power control for cellular V2X environment. Our results show that the learned policy can achieve significant Packet Inter-Reception time (PIR) improvement over the standard algorithms while maintaining an appropriate Channel Busy Ratio (CBR). Moreover, the congestion control function can be extracted from the trained network, and is simple enough to be specified in the standard documents.**

***Index Terms*—V2X, congestion control, reinforcement learning (RL), SAE J3161/1, joint control**

I. Introduction

Cellular Vehicle-to-Everything (C-V2X) communication, based on Long Term Evolution (LTE), enables vehicles to transmit Basic Safety Messages (BSM) or Cooperative Awareness Messages (CAM) periodically, containing safety-critical information such as position, speed, and acceleration [1]. The bandwidth allocation for C-V2X varies across countries, but typically does not exceed 30 MHz and can suffer congestion as the density of vehicles increases. The channel congestion can increase packet collisions, negatively impacting safety due to the lack of updates from BSM or CAM packets. Although congestion control algorithm for C-V2X environment has been recently specified in the Society of Automotive Engineers (SAE) J3161/1 [2] standard, several studies [3]–[5] have shown that the design of the algorithm and its predecessor algorithm in J2945/1 [6] have a problem of large time gaps between information updates and there is room for improvement.

As can be seen in the cases of J3161/1 and J2945/1, a major difficulty in designing congestion control is the coordination of multiple control modes. It may be behind the decision to abandon the transmit (Tx) power control in the newer J3161/1 whereas J2945/1 had both the power control and message rate control. However, congestion control using only rate transmission to mitigate channel congestion can be a safety issue. In other words, reducing the BSM transmission rate is not a desirable solution as it can compromise safety by reducing the frequency of information updates. We believe that this conundrum is difficult to solve with the traditional heuristic design methodology. To solve the problem of designing effective joint congestion control that uses both the transmission rate and power, therefore, we propose to use Deep Reinforcement Learning (DRL) approach.

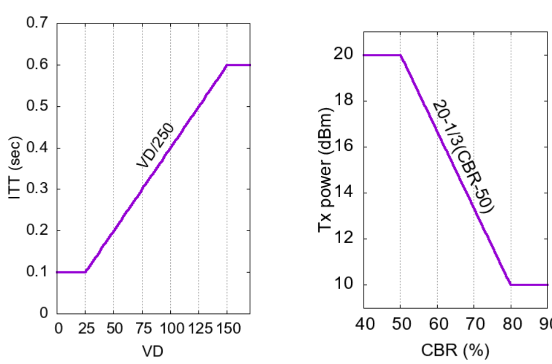
Although the trained network can be used to prescribe the joint control, it cannot be used in the standard specification. Therefore, we queried and extracted the policy in the neural network so that the policy can be specified in future specifications. It turns out that the RL-generated policy is simple enough to serve the purpose, although it also has some aspects that do not align well with our heuristics. Whether in the neural network or in the extracted function form, the RL-generated congestion control policy that jointly uses the two control modes significantly improves the packet delivery performance over the standard algorithms. It roughly halves the Packet Inter-Reception (PIR) time that reflects how frequently vehicles are informed as to the movements of their neighbor vehicles. It also controls the Channel Busy Ratio (CBR) more in line with the original specification of the standards.

II. Related work

There is a rich literature on congestion control in V2X communications. In particular, various attempts have been made to solve congestion control problems in C-V2X communication environments. In Choi et al. [5], authors suggested that adopting Quality of Service (QoS)-aware power or rate control can improve the QoS aspect for various applications using V2X communication. However, the lack of mathematical optimization led to only a small improvement in CBR and Packet Reception Ratio (PRR). Egea-Lopez et al. [7] explored mathematical optimization in joint power and rate adaptation. Unfortunately, it is not clear if the optimized control could be used in real-time due to high computation requirements. DRL can be an alternative to address the computation time requirement, so DRL-based rate control was proposed in Choi et al. [8]. However, using only the BSM transmission rate has a problem of increased inter-packet time gap during congestion, which adversely affects the tracking error of neighboring vehicles' kinematics due to less frequent updates. Moreover, because each vehicle was learned through independent networks, a single final policy could not be derived in the previous work. In order to supplement the rate control and suppress the increase of inter-packet gap is to use Tx power control in combination with the rate control. Nguyen et al. [9] proposed a DRL algorithm that accounts for the safety distance in congestion control. However, calculating the safety distance for each neighboring vehicle may be time-consuming compared to other algorithms. In Kang et al. [10], authors demonstrated that joint control can be achieved by utilizing a stochastic approach that incorporates information from neighboring vehicles. Finally, what is noticeable in the research thread using DRL is the lack of multi-agent Reinforcement Learning (MARL)-based algorithm that jointly controls the transmission rate and power. Because vehicle-to-vehicle (V2V) communication is a highly distributed environment where vehicles react to other vehicles' control actions, MARL is deemed a better approach if only its complexity can be overcome in the training phase.

III. SAE J2945/1 and J3161/1's Congestion Control

SAE J2945/1 established an algorithm for congestion control through two parameters of Inter-Transmit Time (ITT) and Tx power in Dedicated Short Range Communication (DSRC) environment. ITT at each On-Board Unit (OBU) is regulated by Vehicle Density (VD), and Tx power is regulated by CBR. Note that ITT is inversely proportional to the transmission rate. VD is the number of vehicles within a radius of 100m (`vPERRange`) and CBR is the percentage of time the DSRC channel is sensed busy. The J2945/1 congestion control is illustrated in Fig. 1.



(a) Rate control (b) Power control

Fig. 1: J2945/1 congestion control

Unfortunately, the J2945/1 control has been shown to be hardly synergistic. Specifically, the Tx power control stays inactive in both DSRC [3] and C-V2X environments [4]. The fact that rate control single-handedly copes with congestion beats the whole purpose of joint control. It also implies that the inter-packet time gap, a.k.a. PIR time, becomes high and threatens safety due to less frequent updates on neighbor vehicles' kinematics. This is probably the reason that the more recent SAE J3161/1 standard for the C-V2X environment has entirely abandoned the Tx power control, retaining only the rate control from J2945/1. However, the importance and need of the joint control as originally intended in J2945/1 should not be discounted from the perspective of safety achievable through more frequent periodic message exchanges. Although it has been proven to be prohibitively difficult to design an effective joint control heuristically, we demonstrate in this paper that DRL makes it achievable. Moreover, the DRL-generated control policy is simple enough to be specified in the standard. The joint control functions do not have to remain intangible in the neural network, but can be queried and extracted for explicit standard specification.

IV. Solution Approach

A. MARL framework for joint congestion control

The main purpose of distributed congestion control (DCC) is to regulate the channel congestion level from excessive increase. This is a common objective for all vehicles. Therefore, we take *centralized-training-decentralized-execution* approach for our MARL model. The overall architecture of the proposed MARL-based DCC algorithm is illustrated in Fig. 2. Below,

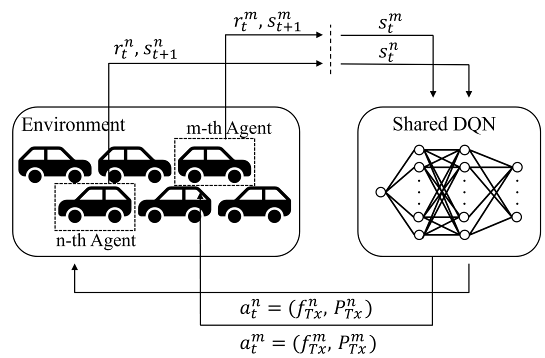


Fig. 2: Overall framework of RL-based congestion control policy learning in C-V2X environment

we define agent, state, action and reward to train a DCC policy through MARL. After training, we aim to extract the learned policy from the trained neural network.

1) Agent: Each vehicle is an agent. The vehicles move along the road and broadcast BSMs. They take joint rate and power control actions based on their own observation of CBR.

2) State: As each vehicle measures local CBR that provides an estimation on the current channel congestion level, we define CBR value to be the state of a vehicle. CBR is determined by the proportion of subchannels in the resource pool with Received Signal Strength Indicator (RSSI) exceeding a preconfigured threshold, observed over the last 100 subframes.

3) Action: We define both the message rate space $\mathcal{F}$ and Tx power space $\mathcal{P}$ as a discrete space. Specifically, the set of actions is defined as $\mathcal{A} = \mathcal{F} \times \mathcal{P}$. We set $\mathcal{F} = \{1.66, 2, 2.5, 3.33, 5, 10\}$ in Hz and $\mathcal{P} = \{10, 12, 14, 16, 18, 20\}$ in dBm, hence $|\mathcal{A}| = 36$.

4) Reward: All vehicles receive a single shared reward at time step t based on the average CBR and average PIR, as

$$r_t = (1 - \mathrm{CBR}_t) + \alpha \times (0.1 - \mathrm{PIR}_t) \tag{1}$$

where α is a weight factor. To obtain the average values of PIR and CBR for reward computation, we exploit that individual PIR and CBR measurements are available in the simulator. The reward is based on the fact that suppressing CBR is the primary goal of DCC. Also, a conflicting goal of minimizing PIR is accommodated. Therefore, at each vehicle m our MARL model is tasked to solve an optimization problem:

$$\begin{aligned} \underset{f_{Tx}^m, P_{Tx}^m}{\operatorname{argmax}} \quad & \mathbb{E}[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots | s_t] \\ \text{where} \quad & r_t = (1 - \mathrm{CBR}_t) + \alpha \times (0.1 - \mathrm{PIR}_t) \\ \text{subject to} \quad & 0 \le \mathrm{CBR}_t \le 1, \\ & 0 \le \mathrm{PIR}_t \le 1. \end{aligned} \tag{2}$$

where f_{Tx}^m and P_{Tx}^m denote the message transmission rate and the transmission power of vehicle m. That is, our reward system is designed to find joint rate and power control that properly control channel congestion while minimizing PIR. Here, PIR value is defined based on vehicles within 150 m and calculated every time step, i.e., 1 s. This is why PIR is constrained to 1 s at the maximum.

As to the use of PIR in the reward, PRR is an important performance metric when rate control is not used, but otherwise PIR becomes the primary metric as in our scheme and the two baseline schemes J2945/1 and J3161/1 [11], [12]. It is because aiming at high PRR may lead to the trivial solution of transmitting at a very low transmission rate and thereby minimizing packet collisions. This will increase the time interval between two successful packet receptions, beating the very purpose of periodic safety beaconing.

B. Learning

Once the policy is updated, the updated policy is distributed to individual vehicles. The vehicles then take actions based on the newly distributed policy and the resulting PIR and CBR averages are used to update the policy in the next step. The iteration repeats until the policy update is stopped. The learning process is described in Algorithm 1 and the RL parameters are presented in Table I.

TABLE I: RL Parameters Configuration

Parameter	Value
Multi-step	3
Batch size	64
Target DQN update period (C)	200 [steps]
Optimizer	Adam
DQN hidden layer	(100, 100)
Learning rate	1e-4
Discount factor γ	0.95
α	2.43

The vehicles learn a policy through Deep Q-Network (DQN) [13], parameterized by Multi-Layer Perceptron (MLP) consisting of two hidden layers. All vehicles share one main DQN and one target DQN. Note that the shared networks are an artefact of the centralized training and distributed execution approach. At each time step, the network's parameters θ are updated to minimize the loss function $L(\theta) = \mathbb{E}[(y - Q(s, a; \theta))^2]$. The target value y is calculated using the target DQN Q^*. Here, we employ the target network which is updated every 200 steps to overcome the non-stationary target problem [13]. The discount factor γ is used to weight future rewards, and multi-step RL is for stabilizing effect [14]. In Algorithm 1, D indicates the number of time steps for multi-step deep reinforcement learning. The transitions from the vehicles are stored in the replay buffer. Note that the transition (s_t, a_t, r_t, s_{t+1}) is not directly stored in the replay memory $\mathcal{E}_{replay}$ because we use multi-step RL. For balancing exploration and exploitation, we adopted the ϵ-greedy method. Each vehicle chooses a random action with probability ϵ rather than the action that maximizes the Q-value. This enables the vehicle to explore different actions and values. As the number of encountered transitions increases, the policy improves and the need for exploration decreases. Therefore, we linearly decrease ϵ from 0.8 in each step. In order to avoid training separate networks for different vehicle traffic densities, we assume ρ = 500 veh./km during training. Nevertheless, we will see in Section V that the trained policy for this density works effectively for other traffic densities. The training takes 3,000 steps to find out the policy

Algorithm 1 MARL-Based Learning Algorithm for DCC Policy Generation

Initialize replay memory $\mathcal{E}_{replay}$ to capacity N
Initialize transition memory $\mathcal{E}_{trans}$ to capacity M
Initialize Q-function with random weights θ
Initialize target Q^*-function with $\theta^- = \theta$

for episode = 1, E **do**
  for time step =1, T **do**
    For all vehicles,
    if $B = 1$ for $B \sim$ Bernoulli(ϵ) **then**
      Choose a random a_t
    else
      Choose $a_t = \operatorname{argmax}_a Q(s_t, a; \theta)$
    end if
    Vehicles transmit the packet at the chosen transmission rate f_{Tx} and power P_{Tx} (a_t)
    Observe reward r_t and calculate s_{t+1}
    Store the transition (s_t, a_t, r_t, s_{t+1}) in $\mathcal{E}_{trans}$
    Store the experience $(s_{t-D+1}, a_{t-D+1}, \mathbf{r}_{t-D+1}, s_{t+1})$ in $\mathcal{E}_{replay}$ where $\mathbf{r}_{t-D+1} = (r_{t-D+1}, r_{t-D+2}, \cdots, r_t)$
    Sample mini-batch of experiences from $\mathcal{E}_{replay}$
    Set $y_j = \mathbf{d} \cdot \mathbf{r}_j + \gamma^D \max_{a'} Q^*(s_{j+D}, a'; \theta^-)$ where $\mathbf{d} = (1, \gamma, \gamma^2, \cdots, \gamma^{D-1})$
    Perform optimization via $L(\theta)$ and update θ
    Update Q^*-function with $\theta^- = \theta$ for every C steps
  end for
end for

of joint congestion control. As training should start after the replay buffer is filled to some extent, so the network is first updated after step 300, after which the update is performed every 200 steps.

C. Interfacing with V2X simulator

As we need an environment for vehicles in V2X communication, we use the open-source V2X network simulator LTEV2Vsim written in MATLAB [15]. The simulator can track the status of packets sent during communication based on the V2X wireless channel configuration and provides CBR and PIR measurements. To integrate the V2X simulator with the Python-based MARL framework, we establish a socket communication between them. The V2X simulator includes the physical parameters and Medium Access Control (MAC) algorithms of C-V2X communications. It also implements in-band emission interference. Since sufficient time is needed for the actions to be reflected in the environment, each time step is set to 1 second. At every second, the MATLAB environment measures the CBR state s_t and reward r_t for each vehicle and then sends this information to the MARL process. The LTEV2Vsim process waits for the MARL to calculate the corresponding actions, which consist of the message rate f_{Tx} and power P_{Tx} for each vehicle. The LTEV2Vsim is resumed after it receives actions f_{Tx}^{m} and P_{Tx}^{m} then reflects it to each vehicle. Then the vehicle can measure a new CBR in the next state s_{t+1}^{m}. Through this process, the MARL algorithm calculates the expected cumulative reward, allowing the vehicles to learn how to maximize their reward over time.

D. Extracted policy

After training, an optimized policy that maximizes the reward was extracted from the neural network by querying the trained network with CBR values, and is shown in Fig. 3. Comparing with Fig. 1, we observe a few interesting aspects from the RL-generated policy.

- The joint control policy is not complex. It implies that it can be readily specified in the standard.
- It is intriguing that it selects the opposite action for Tx power as congestion aggravates: it increases the power whereas J2945/1 decreases it.
- The joint control has discontinuity. It makes abrupt transitions at the CBR values of 56% and 78%.
- Both control modes remain active and cooperative. It is in contrast to J2945/1 whose power control is quickly rendered ineffective [3], [4].

Although this RL-generated policy has some aspects that do not align well with our heuristics (i.e., discontinuity and power control direction), we show below that it outperforms the standard algorithms under all vehicle traffic densities.

Finally, note that our control policy uses CBR as the only input. It contrasts to the baselines, because J2945/1 uses VD and CBR and J3161/1 uses VD. The main reason to change the input is to align with 3GPP standard that defines C-V2X communication. Although 3GPP does not specify a congestion control algorithm, it stipulates that the control input is CBR

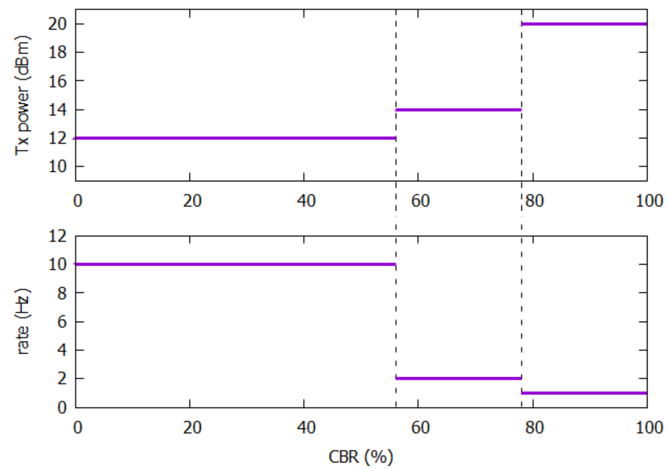


Fig. 3: Learned joint control policy

[16]. Moreover, VD and CBR are not independent. Therefore, we believe that CBR summarizes the channel load more directly than VD and use it as the sole input to the joint congestion control policy.

V. Performance Evaluation

To evaluate the performance of the proposed method, we conduct experiments on various levels of vehicle density and consequent wireless channel congestion: ρ =300, 400, 500, and 600 km/h. Note that the simulation parameters during training and evaluation are all the same except for the policy. Since there is no standardized joint congestion control in the C-V2X environment, a C-V2X adaptation of J2945/1 that jointly controls congestion in DSRC environment is used as a baseline. As another baseline, J3161/1 that has only rate control in the C-V2X environment is used. We show that an effective joint congestion control policy can be designed through DRL via a comparison between the DRL-designed method and the above two baselines.

A. Simulation Parameters

Table II shows the parameters used for the experiments. In V2X simulator, the Modulation and Coding Scheme (MCS) index represents the modulation and coding level for the message, while the resource keep probability and Resource Reselection Counter (RRC) are used for sidelink resource selection in the V2X Semi-Persistent Scheduling (SPS) algorithm [16]. Note that the values of f_{Tx} are inversely related to the ITT or resource reservation interval (RRI) values. The P_{Tx} value was set to between 10 and 20 dBm as in the baseline schemes [2]. All other simulation parameters followed the values presented in the J2945/1 and J3161/1 standards.

B. Results

1) PIR: Fig. 4 shows PIR of the three compared schemes. We observe that the proposed RL-generated control significantly reduces PIR compared to the baselines at all vehicle

TABLE II: Simulation Parameters for V2X Communication

Parameter	Value
Message size	300 [bytes]
MCS index [2]	11
Pathloss model	WINNER+B1
Antenna height	1.5 [m]
Resource keep probability	0.8
RRC	[5:15]
Bandwidth [2]	20 [MHz]
Tx Rate (f_{Tx})	{1.66, 2, 2.5, 3.33, 5, 10} [Hz]
Tx Power (P_{Tx})	{10, 12, 14, 16, 18, 20} [dBm]
Vehicle density (ρ)	{300, 400, 500, 600} [veh./km]
Speed of vehicles	(m, σ) = (50, 3) [km/h]

densities. Note that the baselines are almost indistinguishable as the J2945/1 power control stays inactive (which is a pathology [3], [4]), and J3161/1 has given up power control. In all congestion scenarios, the RL-generated joint control enables each vehicle to more frequently update surrounding vehicles, which is important in terms of driving safety.

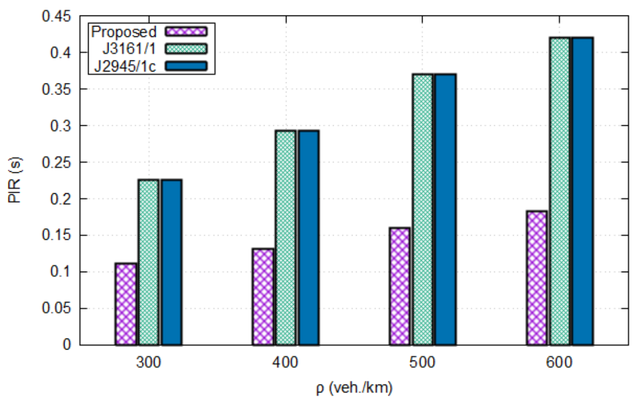
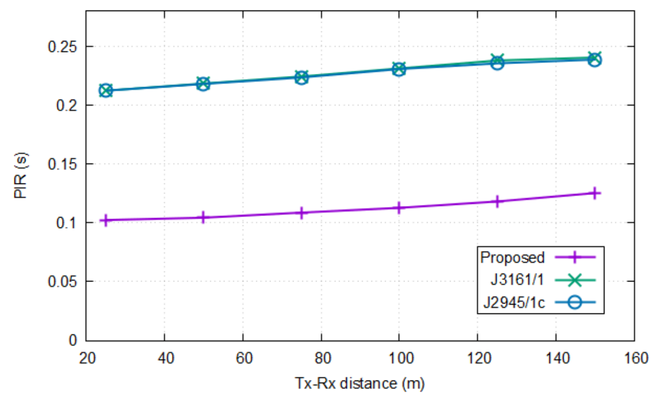


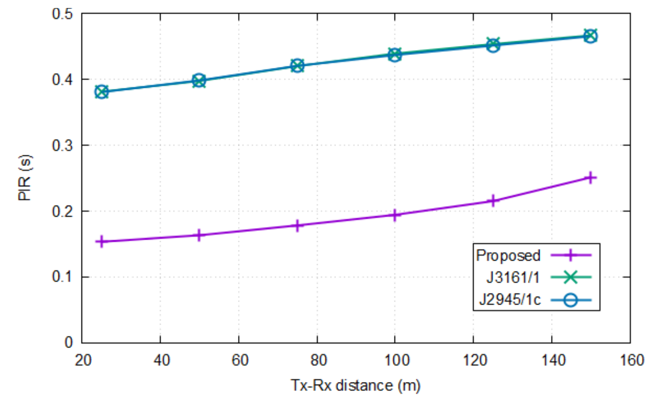
Fig. 4: Average PIR

The RL-generated control not only shows better average PIR, but also much improved performance over all Tx-Rx distances. Fig. 5 shows the cases for the lowest and the highest vehicle traffic densities: $\rho = 300$ veh./km and $\rho = 600$ veh./km. Although we do not present the other vehicle density scenarios in between them for space reason, we obtain the same qualitative result. Note that the PIR is defined and measured within 150 m of the transmitter in this paper, so the PIR values are shown up to 150 m.

2) PRR: Note that in V2X safety communication through periodic beacons such as BSM or CAM, updating neighboring vehicles as to the ego vehicle's kinematics must be done as frequently as the channel allows. Therefore, PIR is the most important performance metric for periodic safety beaconing. In determining the PIR, two factors are involved: how frequently messages are sent (ITT or inversely, the Tx rate) and how much among them survive (PRR). Therefore, the fact that the proposed RL scheme is inferior to the baselines in terms of the PRR metric should be considered in combination with ITT [11], [12]. Fig. 6 shows the PRR value for $\rho = 600$ veh./km. We obtain the same qualitative results for other vehicle traffic densities, which are omitted for space.



(a) $\rho = 300$ veh./km



(b) $\rho = 600$ veh./km

Fig. 5: PIR as functions of Tx-Rx distance

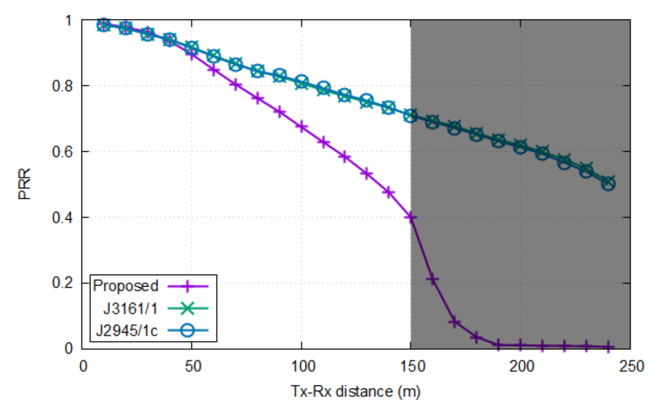


Fig. 6: PRR of individual packets as functions of Tx-Rx distance for ρ = 600 veh./km

There are two reasons why PRR of individual packets in the baselines are higher than the proposed method. First, in the baselines, maximum power is used for transmission in each vehicle. J3161/1 simply fixes the Tx power at the maximum, and J2945/1's power remains high in almost all vehicle traffic densities due to aggressive intervention of rate control [4]. Since the proposed method more effectively controls the power along with ITT, the PRR becomes inevitably lower over distance although PIR is compensated for by smaller ITT.

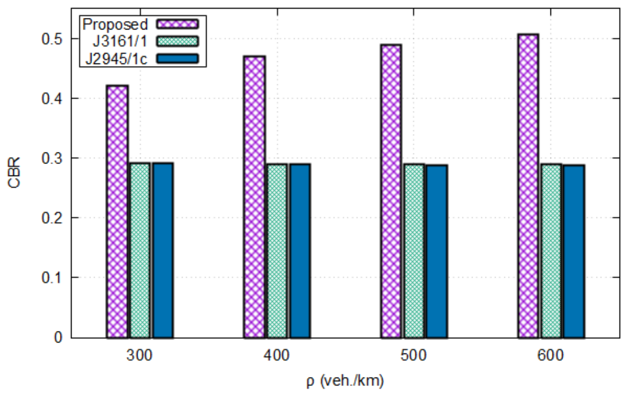


Fig. 7: Average CBR

Second, as mentioned above, the baselines transmit messages at a lower rate (or at a higher ITT) than the proposed method so that packet collision losses are minimized. Nevertheless, the excessively lowered rate directly leads to the deterioration of PIR in the baselines as shown by Fig. 4.

Another observation that we make from the figure is that the PRR performance of the proposed scheme more rapidly drops beyond the Tx-Rx distance of 150 m. This is because the PIR is averaged in the communication range of 150 m for the reward function in (2). It shows that the communication range requirement can be effectively accommodated in the RL-generated control.

3) CBR: Fig. 7 compares the CBR of the three schemes. We observe that the baseline schemes suppress the CBR at approximately 30%. However, it is excessive and unnecessary because J2945/1 targets to contain the CBR between 50% and 80% [6]. Unilaterally reducing CBR is not desirable because it leads to the excessive rate control and consequent inflation of PIR due to large ITT values used. In contrast, the proposed scheme yields CBR in the range of 40% to 50%. It is still stronger than J2945/1 aims at, but much closer to target CBR than the baselines. These results confirm that the MARL-generated joint rate and power congestion control is better designed than the heuristically designed standard algorithms. Moreover, the MARL-generated control is not a black box contained in a neural network, but is simple enough to be explicitly specified in standard documents, as given by Fig. 3.

VI. Conclusion

This paper proposed a DQN-based joint congestion control for V2X communication. It demonstrated the feasibility of RL-generated congestion control policy simple enough to be specified in the standard and more effective than the heuristically designed control algorithms in the current standards. Joint control with multiple control modes poses prohibitive difficulties in predicting the complex interactions between the control modes and is likely to lead to suboptimal performance. However, RL-based joint control policy generation overcomes such difficulties and can produce a better performing control. This paper demonstrated that the automatically generated control achieves much better PIR performance while containing CBR in a range closer to the design goal of the current standard. For future work, we plan to extend the 2-dimensional discrete action space to a multi-dimensional one that additionally considers other control modes and inputs, such as MCS index and QoS requirements.

Acknowledgment

This research was supported by the Ministry of Science, ICT(MIST), Korea, under the Project-based AI Talent Fostering Program (RS-2022-00143911) supervised by the Institute for Information & Communications Technology Planning & Evaluation (IITP).

References

[1] Society of Automotive Engineers (SAE), *V2X Communications Message Set Dictionary*, jul 2020. [Online]. Available: https://doi.org/10.4271/J2735_202007

[2] ——, *On-Board System Requirements for LTE-V2X V2V Safety Communications*, mar 2022. [Online]. Available: https://doi.org/10.4271/J3161/1_202203

[3] S. Lim and H. Kim, "Improving information age in sae j2945 congestion-controlled beaconing," *IEEE Communications Letters*, vol. 23, no. 2, pp. 358–361, 2018.

[4] Y. Yoon and H. Kim, "Balancing power and rate control for improved congestion control in cellular v2x communication environments," *IEEE Access*, vol. 8, pp. 105 071–105 081, 2020.

[5] J. Choi and H. Kim, "A qos-aware congestion control scheme for c-v2x safety communications," in *2020 IEEE Vehicular Networking Conference (VNC)*. IEEE, 2020, pp. 1–4.

[6] Society of Automotive Engineers (SAE), *On-Board System Requirements for V2V Safety Communications*, apr 2020. [Online]. Available: https://doi.org/10.4271/J2945/1_202004

[7] E. Egea-Lopez, P. Pavon-Mariño, and J. Santa, "Optimal joint power and rate adaptation for awareness and congestion control in vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 25 033–25 046, 2022.

[8] J.-Y. Choi, H.-S. Jo, C. Mun, and J.-G. Yook, "Deep reinforcement learning-based distributed congestion control in cellular v2x networks," *IEEE Wireless Communications Letters*, vol. 10, no. 11, pp. 2582–2586, 2021.

[9] L.-H. Nguyen, V.-L. Nguyen, and J.-J. Kuo, "Efficient reinforcement learning-based transmission control for mitigating channel congestion in 5g v2x sidelink," *IEEE Access*, vol. 10, pp. 62 268–62 281, 2022.

[10] B. Kang, J. Yang, J. Paek, and S. Bahk, "Atomic: Adaptive transmission power and message interval control for c-v2x mode 4," *IEEE Access*, vol. 9, pp. 12 309–12 321, 2021.

[11] B. Toghi, M. Saifuddin, Y. P. Fallah, and M. O. Mughal, "Analysis of distributed congestion control in cellular vehicle-to-everything networks," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*. IEEE, 2019, pp. 1–7.

[12] T. Shimizu, B. Cheng, H. Lu, and J. Kenney, "Comparative analysis of dsrc and lte-v2x pc5 mode 4 with sae congestion control," in *2020 IEEE Vehicular Networking Conference (VNC)*. IEEE, 2020, pp. 1–8.

[13] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[14] J. F. Hernandez-Garcia and R. S. Sutton, "Understanding multi-step deep reinforcement learning: A systematic study of the dqn target," *arXiv preprint arXiv:1901.07510*, 2019.

[15] G. Cecchini, A. Bazzi, B. M. Masini, and A. Zanella, "Ltev2vsim: An lte-v2v simulator for the investigation of resource allocation for cooperative awareness," in *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*. IEEE, 2017, pp. 80–85.

[16] 3GPP, "NR; Physical layer procedures for data," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.214, 9 2022, version 17.3.0. [Online]. Available: https://portal.3gpp.org/#/55936-specifications.