Multiple Channel Access using Deep Reinforcement Learning for Congested Vehicular Networks

Chungjae Choe, Junsung Choi, Jangyong Ahn, Dongryul Park, and Seungyoung Ahn
The CCS Graduate school of Green Transportation
KAIST
Daejeon, Korea
cjchoe12@kaist.ac.kr

***Abstract*—Vehicular Ad-hoc Network (VANET) is a standard protocol for wireless vehicular communication that enables Vehicle to Vehicle (V2V) and Vehicle to Infrastructure (V2I) communications. VANET safety applications aim to prevent traffic accidents and require a high Packet Delivery Ratio (PDR) and low latency of safety packet broadcast. When a large number of vehicles simultaneously access a limited channel resource for the safety broadcast, the safety requirements impose more challenges; the communication performance will significantly degrade due to network congestion. Especially, infrastructure-less VANETs, which only allow V2V communication, vehicles are supposed to overcome the congestion problem using a self-adaptation scheme without the aid of infrastructures. In this paper, we propose a self-adaptive MAC layer algorithm employing Deep Q Network (DQN) with a novel contention information-based state representation to improve the performance of the V2V safety packet broadcast. The proposed algorithm operates a fully distributed manner, and it is evaluated by simulations considering various levels of traffic congestion.**

***Index Terms*—Vehicular ad-hoc network, channel access, contention window, reinforcement learning**

I. Introduction

Vehicle-to-Everything (V2X) communication is the core function of intelligent transportation system which contributes to prevent traffic accidents and congestions by exchanging local surrounding information between vehicles and infrastructures. Vehicular Ad-hoc Network (VANET) is the standard network configuration of V2X that offers Vehicle to Vehicle (V2V) and Vehicle to Infrastructure (V2I) communications. Unlike the general mobile adhoc network, VANET has dynamically changing environment because of the rapid movement of vehicles. Due to this characteristic, VANET safety applications such as Cooperative Adaptive Cruise Control (CACC) and Forward Collision Warning (FCW) require high Packet Delivery Ratio (PDR) and low latency [1]. When a large number of vehicles access a limited frequency band, the safety requirements impose more challenges. Thus, the development of an efficient algorithm for Medium Access Control (MAC) layer is a key task to enhance the communication stability and performance for congested VANETs and eventually guarantee the safety of fully autonomous vehicles.

Dedicated Short Range Communications (DSRC) [2], which is the widely tested standard for V2X with $75\,MHz$, seven separate channels, in $5.9\,GHz$ band. If multiple vehicles transmit packets simultaneously using the same channel, packet collisions occur, and vehicles may not communicate successfully. As a naive solution, DSRC standard employs Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) for MAC layer protocol [3]. CSMA/CA is a contention oriented random channel access protocol; all vehicles in VANETs have the same priority and access a channel with a carrier sensing method. When a vehicle recognizes that the channel is idle, no-one is using the channel, it waits a random amount of time before transmitting data. This scheme is called random backoff and the waiting duration corresponds to Contention Window (CW). All vehicle select the waiting duration uniform randomly over the range of [0, CW]. After transmitting, senders receive acknowledgment packet (ACK) to identify a successful transmission, whereas its CW doubles for each packet collision (no ACK) within a maximum CW (CW_{max}).

The aforementioned protocol is suitable for sparse VANETs with a small number of vehicles. However, the communication performance degrades in congested VANETs with a large number of vehicles ~~due to the~~ high possibility of overlapped CW values among vehicles [4]. If infrastructures exist, channel access of vehicles can be instantaneously controlled by nearby infrastructures [5]–[7]. On the other hand, network congestion grows to a severe problem when only V2V communication is available. In infrastructure-less case, vehicles suffer from high collision rate of V2V packets since it is unable to obtain status information of VANETs [8]. Furthermore, the V2V-based safety packet broadcast encounters a worse situation because the DSRC standard does not allow ACK transmission for a broadcast packet to prevent the ACK storm phenomenon [9]. Since vehicles cannot recognize whether broadcasting of packets were successful, the adjustment of CW according to a network condition is impossible. As a result, the broadcast performance is significantly deteriorated due to the high packet collision rate. Therefore, an intelligent MAC layer algorithm that performs self-adaptation of CW with a proper ACK scheme is needed for infrastructure-less congested VANET.

In this paper, we propose a self-experience-based CW adaptation algorithm employing Reinforcement Learning (RL) for MAC layer of VANETs. In RL, an agent learns the optimal action policy to receive a maximum reward through numerous trials and errors even if there is no information on the envi-

978-1-7281-5207-3/20/$31.00 ©2020 IEEE

ronment model. According to this scheme, vehicles (*agents*) broadcast (*trials*) the safety packet using V2V communication and receive transmission results (*errors*) from a VANET (*environment*) without the aid of infrastructures. Consequently, vehicles learn to adjust the optimum CW to receive high rewards by self experiences solely. Previous studies [10]–[12] have presented effective CW adaptation algorithms using Q-learning [13], where each vehicle only exploits its own CW for state representation. In the proposed algorithm, we improve the performance of the CW adaptation algorithm employing Deep Q Network (DQN) [14] empowered by a contention information-based state representation that includes CW values received from neighboring vehicles, corresponding success rates and frequency values. From the observed state, vehicles can estimate the congestion level of VANETs [15]. The objective of the algorithm is to achieve high PDR and low end-to-end delay for V2V broadcast in infrastructure-less congested VANETs, and key features of the algorithm are as follows. First, the proposed algorithm follows the multi-channel operation of DSRC standard [16]. Second, the algorithm only adapts CW following the operation protocol of CSMA/CA. Third, vehicles exploit the proposed contention information-based state to learn the adaptive CW policy. We present a software simulation to evaluate the algorithm with various congestion levels such as the number of vehicles 50 to 150.

This paper is organized as follows. In Section II, we provide a literature review of related studies and improvements in this paper. A brief introduction of the multi-channel operation in DSRC standard is presented in Section III. Section IV introduces the proposed algorithm in detail, and Section V evaluates the functionality of the algorithm. Lastly, the conclusion with a discussion of future works is drawn in Section VI.

II. Related Works

The importance of an adaptive MAC protocol for congested VANETs has been highly emphasized to improve the performance of the safety packet broadcast. For this purpose, roadside infrastructures can be employed to control the multiple channel access of vehicles since infrastructures can behave as a channel coordinator. A TDMA-based algorithm using a disjoint set of time slots is proposed to reduce packet collision probability [5]. Also, studies based on CSMA/CA have been conducted in [6] and [7]. An optimization algorithm for multi-channel intervals of the DSRC standard is presented to improve the stability of the broadcast [6]. The minimization of idle service interval increase the performance of the broadcast using the reservation time mechanism [7]. However, they may not be suitable for infrastructure-less congested VANETs.

As a solution of V2V communications in infrastructure-less congested VANETs, RL-based MAC protocols [10]–[12] have been proposed. A Q-learning-based MAC algorithm is proposed that defines each vehicle as a single agent and improves the performance of data transmission, but it only consider V2V unicast case [10]. In order to enhance the V2V broadcast performance, a Q-learning-based MAC protocol employing a novel reward function with a collective contention estimation is proposed, and the authors demonstrate the performance improvement of V2V broadcast from various experiments [11], [12]. However, the aforementioned algorithms assume single-channel operation of the control channel (CCH), not consider multi-channel operation of DSRC standard. Moreover, there is a potential to further improve the V2V broadcast performance in terms of packet delivery and latency.

broadcast

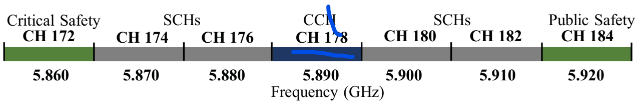


Fig. 1. Description of multi-channel in DSRC

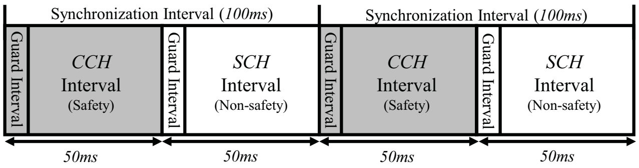


Fig. 2. Alternating CCH and SCH intervals

In this paper, we propose the DQN-based MAC algorithm for V2V safety packet broadcast in infrastructure-less congested VANETs. The distinctive contribution of this paper is that we employ the contention information-based state representation, which includes the communication status of neighboring vehicles. Moreover, the algorithm complies with the DSRC multi-channel protocol. We evaluate the algorithm with two criterions: PDR and end-to-end delay.

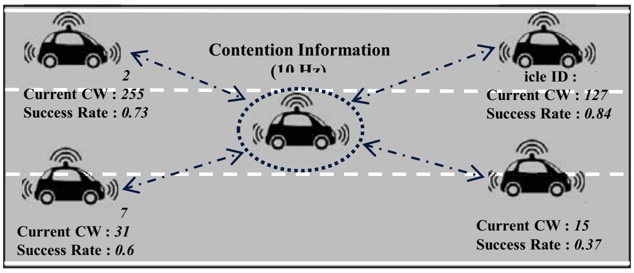
III. Multi-Channel operation in DSRC

As shown in Fig. 1, the spectrum of DSRC consist of seven 10MHz channels [16]. There is channel 178 corresponds to CCH, which is allocated for broadcasting the safety packet. There are six service channels (SCH) used for non-safety packet transmission. Fig. 2 shows a concept of channel intervals for the multi-channel operation in DSRC standard. An operation scheme of DSRC is divided into alternating CCH and SCH intervals; each channel interval is 50ms. In principle, DSRC standard specifies the rule that every vehicle in VANETs broadcasts the safety packet for every 100 ms (10Hz) during the CCH interval (CCHI). This is mainly exchanged by V2V links for safety purposes.

IV. The proposed DQN-based MAC algorithm

The objective of the proposed DQN-based MAC algorithm is to guarantee high PDR and low end-to-end delay for V2V safety packet broadcast in infrastructure-less congested VANETs. Every vehicle acts as a single agent and performs adaptive backoff that adjusts CW through the interaction with a VANET. Besides, vehicles broadcast the safe packet and unicast ACK to a designated vehicle (a closest node ID) for providing feedback during the CCHI and the SCH interval (SCHI), respectively. Here, we describe the proposed Markov

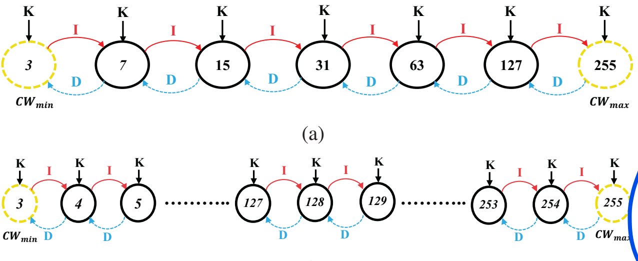
Ack receive



Observed state with the contention information

Vehicle ID	Current CW	Frequency	Success Rate
Itself	*15*	*5*	*0.32*
2	*255*	*10*	*0.73*
17	*31*	*3*	*0.6*
26	*127*	*68*	*0.84*
51	*15*	*17*	*0.37*

Fig. 3. The proposed state representation



(a)

(b)

Fig. 4. (a) Discrete CW space. (b) Continuous CW space.

Decision Process (MDP) model including state, action, and reward function with the DQN structure.

A. State Representation

Previously, the Q-learning-based MAC algorithms [10]–[12] have demonstrated their feasibility in infrastructure-less congested VANET. The simple state-action pair with the corresponding Q-table employed in the previous studies [10]–[12] is expressed as

Q-learn

$$\mathbf{Q}[7][3] = \begin{array}{c} (CW\textbf{-1})/2 \quad CW \quad CW\textbf{*2+1} \\ \begin{bmatrix} \infty & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \infty \end{bmatrix} \end{array} \begin{array}{l} \textbf{state } (CW) \\ \textbf{3} \\ \textbf{7} \\ \textbf{15} \\ \textbf{31} \\ \textbf{63} \\ \textbf{127} \\ \textbf{255} \end{array} \qquad (1)$$

, where Q-function $Q[7][3]$ has the size of the state and action space 7 and 3, respectively. They simplify the problem and represent the simple state representation that a vehicles only exploits a currently selected CW information of itself. However, this method is limited in that vehicles cannot use

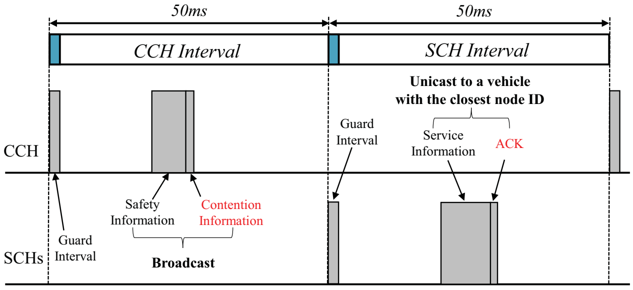


Fig. 5. Packet configurations of CCH and SCH intervals

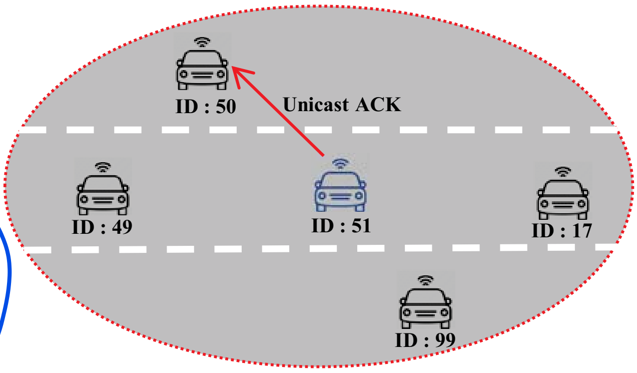


Fig. 6. Example of the vehicle selection method for ACK unicast

the communication information of neighboring vehicles for state representation. Therefore, we propose a fully informative state representation with the contention information. Fig. 3 shows the proposed state representation. During CCHI, vehicles broadcast safety packet appending their node ID, selected CW and the corresponding success rate calculated by broadcast results of the selected CW. According to this, vehicles establish the contention information-based state with collected CW values, corresponding frequency values F and success rates R of other vehicles. Based on the accumulated contention information, vehicles can indirectly estimate the congestion level of the network [15]. Also, the state is function approximated by DQN, which enables vehicles to recognize traffic status information of VANETs and to discover optimal action policy.

B. Action Definition

The action definition consists of three components: Keep (K), Increase (I), and Decrease (D). CW is adjusted according to the selected action. In the proposed algorithm, CW_{min} and CW_{max} are 3 and 255, respectively. In addition, we present the transition rule of CW follows two different spaces: discrete and continuous changes. First, the discrete CW change is based on the binary exponential backoff algorithm [17] as shown in Fig. 4(a); it includes a total of seven state spaces same as the previous studies [11], [12]. Second, in the continuous CW change shown in Fig. 4(b), CW increases or decreases by one-step from CW_{min} to CW_{max}, and it has a

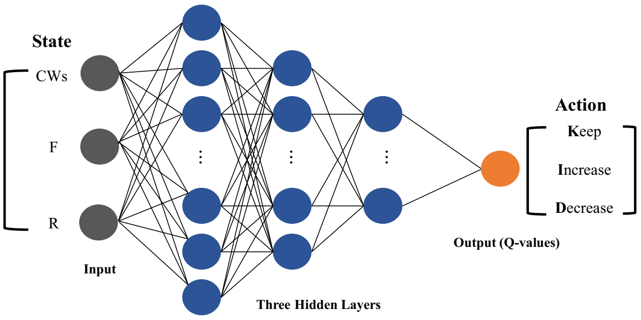


Fig. 7. The proposed DQN structure

total of 253 state spaces. The CW change is applied to both of CCHI and SCHI. In section V, we compare the performance of two different CW changes.

C. ACK Receiving Scheme and Reward Function

All vehicles must receive ACK to identify broadcast results for CW adaptation according to congestion levels. Thus, it is necessary to suggest a proper ACK scheme to define a reward function. Flooding-based ACK scheme such as a probabilistic rebroadcast has proposed in previous studies [11], [12] should be avoided because they can cause serious congestion in VANETs with the flooding storm [9].

We propose a unicast-based ACK scheme during SCHI. Fig. 5 shows the configuration of packets transmitted in each channel interval. In SCHI, every vehicle selects a target vehicle which has the closest *node ID* from itself among received safety packets and transmits ACK to the selected vehicle. If there are two vehicles with closest *node ID*s, ACK is sent to a vehicle that has a higher ID. As shown in Fig. 6, the vehicle, *node ID* 51, selects the destination, *node ID* 52, for ACK unicast among successfully recognized vehicles. According to this scheme, vehicles can identify whether its broadcast was successful during SCHI. We utilize the binary reward function as

$$r_t = \begin{cases} 1, & \text{if the broadcast was successful} \\ -1, & \text{if the broadcast was failed} \end{cases} \tag{2}$$

, where r_t denotes a reward at current time-step t. Vehicles will receive the positive reward 1 or the negative reward –1 from successful and failed broadcast, respectively.

D. The proposed Deep Q Network

As shown in Fig. 7, the network structure of the proposed algorithm consists of a Deep Neural Network (DNN) including three successive hidden layers with 256, 128, and 64 output dimensions, and Leaky Rectied Linear Unit (Leaky-Relu) is used for the activation function. The state S = $<CW, F, S>$ is the input, and the output is the vector of estimated Q values for all actions (Keep, Decrease, Increase). As a precaution against a local optimum problem, we utilize E-greedy algorithm [18] to balance exploration and exploitation in action selection. In order to handle the problem of non-stationary target and correlation between selected training samples, we define a target network and use soft-update scheme [14]. Algorithm 1 shows the overall learning process of the proposed algorithm.

Algorithm 1 Deep Q Network-based MAC algorithm

```
Input transmit packet $P_{tx}$, received packet $P_{rx}$, success rate of contention window $R_{CW}$, replay memory $M$, replay memory size $m$ mini-batch size $B$, number of episodes $E$, epsilon greedy $\epsilon$
Notation
  $CW$: the current contention window.
  $t$: time-step.

for episode=1 to $E$ do
    start CCH interval
    procedure STATE-OBSERVATION
        if $P_{rx}$.IsBroadcast then
            update each element in state $s_t$
        end if
    end procedure
    The agent selects an action $a_{t+1}$ according to $\epsilon$-greedy
    The agent update $CW_{t+1}$ acccording to $a_{t+1}$
    procedure BROADCAST($P_{tx}$,$CW_{t+1}$)
        $P_{tx}$.AddContentionInformation($CW_{t+1}$,$R_{CWt+1}$)
        Transmit($CW_{t+1}$,$P_{tx}$)
    end procedure
    start SCH interval
    procedure RECEIVED-FEEDBACK($P_{rx}$)
        if $P_{rx}$.IsUnicast then
            if $P_{rx}$.CheckACK && $P_{rx}$.ValidLatency then
                $r_t \leftarrow 1$
            else
                $r_t \leftarrow -1$
            end if
        end if
        update $R_{CWt+1}$ acccording to $CW_{t+1}$ and $r_t$
    end procedure
    procedure LEARNING
        store experience $< s_t, a_{t+1}, r_t, s_{t+1} >$ into $M$
        $t \leftarrow t+1$
        if $M > m$ then
            ExcuteTraining($M$, $B$)
        end if
    end procedure
end for
```

V. Evaluation

This section describes the simulation aims to evaluate the performance of the proposed algorithm. We use an open-source simulator, called Simulation of Urban Mobility (SUMO), to build realistic traffic conditions. In addition, the network simulator ns3 is coupled with mobilities of SUMO for the VANET simulation. We also exploit ns3-gym library to build a RL environment [19].

A. Simulation Condition

Fig. 8 shows a simulation environment the four lanes highway with a length of 3 km, and vehicles communicate only using V2V without the aid of infrastructures. The simulation is performed following three conditions: (1) All vehicles of the VANET exist in the one-hop communication range ($1km$). (2) non-safety (service) packets are generated uniform randomly in SCHI. (3) A retransmission of ACK is not executed. From the first condition, packet collisions can be accurately

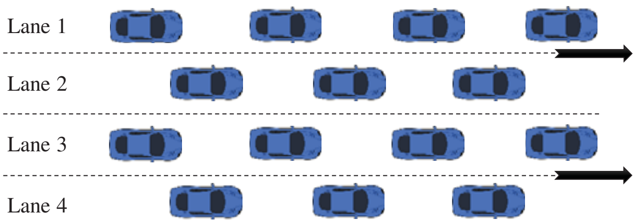


Fig. 8. The traffic environment for the simulation

measured as the number of vehicles increases by eliminating the hidden terminal problem. Through the second and third conditions, all vehicles are able to unicast ACK to target vehicles with a high success rate.

Table I summarizes the key parameters of the VANET considered in the evaluation. Simulations are conducted according to five levels of congestion in 25 steps, ranging from 50 to 150 based on the number of vehicles exist in the VANET; congestion levels are expressed such as 50-VANET, 100-VANET, and 150-VANET. Vehicles follow the Krauss-car following model [20] with default parameters (σ = 0.5, τ = 1) and the maximum velocity 16 ms. The size of the safety packet is 266 bytes including 10 bytes of the contention information, and the non-safety packet consists of 394 bytes including 10 bytes of ACK. Vehicles broadcast the safety packet at 100 ms intervals (10 Hz). The transmission of the non-safety packet is determined following the probability Sn in each SCHI; only the 10 bytes ACK packet may be transmitted in SCHI. We also employ a propagation model, the nakagami propagation model, to reflect a realistic wireless channel [21].

TABLE I
VANET PARAMETERS FOR THE SIMULATION

Parameter	Value
The number of vehicles	$50, 75, 100, 125, 150$
Car-following model	Krauss model
Maximum velocity of vehicles	$16\,m/s$
CCH and SCH frequency	5.890 Ghz and 5.870 Ghz
Transmission power	$100\,mW$
Data rate	$6\,Mbps$
Safety packet size	$256 + 10$ bytes
Non-safety packet size	$384 + 10$ bytes
Non-safety transmission probability S_n	0.2
Propagation model	Nakagami model

B. Simulation Results

We compare four algorithms from simulations, and a description of each is as follows. First, the DSRC CW-31 configures CW_{min} and CW_{max} as 31. It prohibits the change of CW value. Second, the Q-learning-based algorithm, Q-MAC, utilizes the simple state representation that only contains self-CW information. Lastly, the proposed DQN-based algorithm employing the contention information-based state representation is evaluated with two CW space conditions: Discrete CW space-based DQN (DC-DQN) MAC and Continuous CW space-based DQN (CC-DQN) MAC. Table II shows the learning parameters of the proposed algorithm. In the learning process, one episode consists of 10 seconds and $1,000$ learning episodes are consumed for each congestion level.

TABLE II
DQN PARAMETERS FOR TRAINING

Parameter	Value
Time-step t	$0.1\,s$
Replay memory size M	$10,000$
Mini-batch size	10
Epsilon decay rate	0.9995
Starting ϵ	1
Minimum ϵ	0.1
discount factor γ	0.99
Learning rate β	0.0001
Target network update rate α	0.001

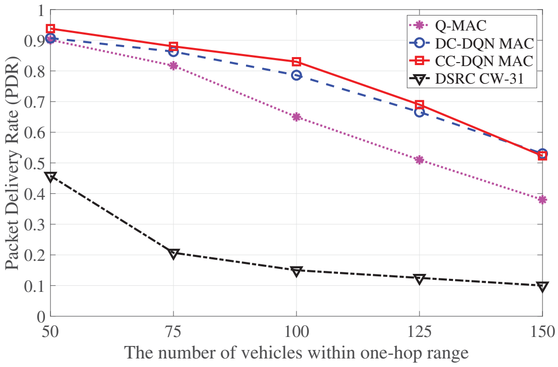


Fig. 9. Average packet delivery ratio

1) Average packet delivery ratio: Fig. 9 shows the average PDRs of the safety packet broadcast regarding the number of vehicles within the one-hop communication range. The DSRC CW-31 is not able to control multiple channel access in congested VANETs, since performance rapidly degrades as the number of vehicles increases. On the contrary, the other three algorithms show robust access control performance with over 90% PDR in the 50-VANET. As the number of vehicles increases, two types of the proposed algorithm, DC-DQN and CC-DQN MAC, outperforms the Q-MAC. Based on the 100-VANET, DC-DQN MAC and CC-DQN MAC have 17% and 21% performance improvement over the Q-MAC, respectively; In particular, the CC-DQN MAC shows the highest PDR of around 83%. In the 150-VANET, the largest congestion level, the Q-MAC has a PDR of 38 %, and that of the DC-DQN MAC and the CC-DQN MAC is around 51%.

2) Average end-to-end delay: As shown in Fig. 10, the end-to-end delay becomes higher as increasing density level.

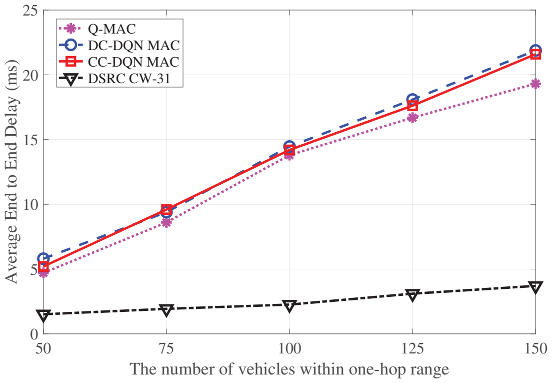


Fig. 10. Average end-to-end delay

The DSRC CW-31 has the lowest delay since the values are measured by a few successful broadcasts due to frequent packet collisions, which indicates that it is not able to control multiple channel access. From a similar phenomenon, the Q-MAC has a relatively low delay compared to the DC-DQN and the CC-DQN MAC. The end-to-end delay increases since the proposed algorithm tends to select high CW values in congested networks for the successful broadcast. It is confirmed that there is a clear trade-off between PDR and latency. Although the degradation, the proposed algorithm satisfies the latency requirement of VANET safety applications [1], $20\,ms$, when the number of vehicles is lower than 125. However, the DC-DQN and CC-DQN show an increasing delay higher than $20ms$ in the 150-VANET. At this extreme congestion case, the proposed algorithm require more improvement to guarantee the acceptable latency.

VI. Conclusion

In this paper, we propose the DQN-based MAC algorithm to improve the performance of V2V safety packet broadcast in infrastructure-less congested VANETs. In the proposed algorithm, fully informative state representation is employed with the contention information collected from neighboring vehicles. In the 100-VANET, the simulation result shows that the proposed algorithm has 21% improvement of PDR performance compared to the simple Q-learning MAC, which uses only self-CW information for the state representation. Furthermore, the proposed algorithm has proved that it is satisfied with the low latency requirement in the VANETs with less than 125 vehicles. However, the performance of the end-to-end delay degrades in the highest congestion level, 150-VANET. Thus, it is required to study of an adaptive MAC algorithm that can further improve PDR and latency performance for severe traffic congestion conditions. As a future work, we expect that combining V2V clustering algorithms with the proposed algorithm can induce further performance improvement.

References

[1] K. A. Hafeez, L. Zhao, B. Ma, and J. W. Mark, “Performance analysis and enhancement of the dsrc for vanet’s safety applications,” *IEEE Transactions on Vehicular Technology*, vol. 62, no. 7, pp. 3069–3083, Sep. 2013.

[2] J. B. Kenney, “Dedicated short-range communications (dsrc) standards in the united states,” *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1162–1182, July 2011.

[3] J. H. Kim and J. K. Lee, “Performance of carrier sense multiple access with collision avoidance protocols in wireless lans,” *Wireless Personal Communications*, vol. 11, no. 2, pp. 161–183, Nov 1999.

[4] X. Ma, X. Chen, and H. H. Refai, “Performance and reliability of dsrc vehicular safety communication: A formal analysis,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, no. 1, p. 969164, Jan 2009.

[5] H. A. Omar, W. Zhuang, and L. Li, “Vemac: A tdma-based mac protocol for reliable broadcast in vanets,” *IEEE Transactions on Mobile Computing*, vol. 12, no. 9, pp. 1724–1736, Sep. 2013.

[6] V. Nguyen, T. T. Khanh, T. Z. Oo, N. H. Tran, E. Huh, and C. S. Hong, “A cooperative and reliable rsu-assisted ieee 802.11p-based multi-channel mac protocol for vanets,” *IEEE Access*, vol. 7, pp. 107 576–107 590, 2019.

[7] Y. Ma, L. Yang, P. Fan, S. Fang, and Y. Hu, “An improved coordinated multichannel mac scheme by efficient use of idle service channels for vanets,” in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, June 2018, pp. 1–5.

[8] K. Ramachandran, M. Gruteser, R. Onishi, and T. Hikita, “Experimental analysis of broadcast reliability in dense vehicular networks,” *IEEE Vehicular Technology Magazine*, vol. 2, no. 4, pp. 26–32, Dec 2007.

[9] Y.-C. Tseng, S.-Y. Ni, Y.-S. Chen, and J.-P. Sheu, “The broadcast storm problem in a mobile ad hoc network,” *Wireless Networks*, vol. 8, no. 2, pp. 153–167, Mar 2002.

[10] C. Wu, S. Ohzahata, Y. Ji, and T. Kato, “A mac protocol for delay-sensitive vanet applications with self-learning contention scheme,” in *2014 IEEE 11th Consumer Communications and Networking Conference (CCNC)*, Jan 2014, pp. 438–443.

[11] A. Pressas, Z. Sheng, F. Ali, D. Tian, and M. Nekovee, “Contention-based learning mac protocol for broadcast vehicle-to-vehicle communication,” in *2017 IEEE Vehicular Networking Conference (VNC)*, Nov 2017, pp. 263–270.

[12] A. Pressas, Z. Sheng, F. Ali, and D. Tian, “A q-learning approach with collective contention estimation for bandwidth-efficient and fair access control in ieee 802.11p vehicular networks,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 9, pp. 9136–9150, Sep. 2019.

[13] C. J. C. H. Watkins and P. Dayan, “Q-learning,” *Machine Learning*, vol. 8, no. 3, pp. 279–292, May 1992.

[14] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, no. 7540, p. 529, 2015.

[15] S. Eichler, “Performance evaluation of the ieee 802.11p wave communication standard,” in *2007 IEEE 66th Vehicular Technology Conference*, Sep. 2007, pp. 2199–2203.

[16] Q. Chen, D. Jiang, and L. Delgrossi, “Ieee 1609.4 dsrc multi-channel operations and its implications on vehicle safety communications,” in *2009 IEEE Vehicular Networking Conference (VNC)*, Oct 2009, pp. 1–8.

[17] J. Goodman, A. G. Greenberg, N. Madras, and P. March, “Stability of binary exponential backoff,” *J. ACM*, vol. 35, no. 3, pp. 579–602, Jun. 1988.

[18] R. S. Sutton and A. G. Barto, *Introduction to Reinforcement Learning*, 1st ed. Cambridge, MA, USA: MIT Press, 1998.

[19] P. Gawłowicz and A. Zubow, “ns-3 meets OpenAI Gym: The Playground for Machine Learning in Networking Research,” in *ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*, November 2019.

[20] S. Krauß, P. Wagner, and C. Gawron, “Metastable states in a microscopic model of traffic flow,” *Physical Review E*, vol. 55, no. 5, p. 5597, 1997.

[21] J. K. Taimoor Abbas, Katrin Sjöberg and F. Tufvesson, “A measurement based shadow fading model for vehicle-to-vehicle network simulations,” *International Journal of Antennas and Propagation*, vol. 2015, p. 12, 2015.