Emergency Message Broadcast Mechanism in Vehicular Ad-Hoc Networks Based on Reinforcement Learning with Contention Estimation

Chien-Min Wu, Cheng-Tai Tsai, Cheng-Chun Hou, Jun-Jie Yang, Gong-De Lin and Mi-Yu Kuang

***Abstract*—In vehicular ad-hoc networks (VANETs), ensuring passenger safety requires fast and reliable emergency message broadcasts. The current communication standard for messaging in VANETs is IEEE 802.11p. As IEEE 802.11p allows carrier-sense multiple access with collision avoidance (CSMA/CA) in the media access control (MAC) layer. A large contention window (CW) value will increase delay, whereas a small CW value will increase the probability of collision. Therefore, adaptive regulation of the CW value is needed to achieve high reliability and low delay in VANETs, in accordance with variations in the environment. However, the traditional MAC protocol cannot achieve the aforementioned requirements. Reinforcement learning (RL) emphasizes the selection of optimal action according to observations of the environment to achieve optimal system performance. In this study, a Q-learning (QL) RL algorithm based on IEEE 802.11p was used to achieve the requirements of adaptive broadcasting. Adaptive broadcasting was achieved based on a reward definition of high reliability and low delay for the QL algorithm. In this approach, the learning state is the CW size, the system sets up a Q-table using RL, and the optimal action is based on the maximum Q-value. The CW size can be provided with adaptive self-regulation by RL, providing high reliability and low delay for the broadcast of emergency messages. We also compared our proposed scheme to other QL-based MAC protocols in VANETs by performing simulations and demonstrated that it can achieve high reliability and low delay for the broadcast of emergency messages.**

***Index Terms*—VANET, IEEE 802.11p, contention window, reinforcement learning, Q-learning.**

I. INTRODUCTION

MOBILE ultra-reliable low-latency communication enables low-delay and high-reliability data exchange and therefore has applications in unmanned driving, intelligent transportation systems (ITSs), industrial automation, smart grids, and vehicle mobile communication networks. A mobile communication network with high reliability and low delay can enhance the safety of vehicle travel. It can reduce loss of life, injuries, and property damage caused by automobile accidents, as well as losses at the social and economic levels [1], [2], [3].

Network architectures comprising IVC (inter-vehicle communications) and RVC (roadside-to-vehicle communications) are based on and derived from the mobile ad-hoc network (MANET) and are thus also known as vehicular ad-hoc networks (VANETs) [4], [5].

To realize the development and implementation of ITSs, wireless access in vehicular environments (WAVE) systems are being studied extensively. The communication protocol of WAVE is IEEE 802.11p, which is a part of IEEE 802.11 that is designed for wireless local area networks. The medium access control (MAC) protocol of DSRC (Dedicated Short-Range Communications) and IEEE 802.11p should be capable of managing large numbers of nodes [6], [7]. However, in a DSRC-based network, the contention window (CW) design cannot adapt dynamically to the traffic. On the other hand, IEEE 802.11p uses a distributed contention-based carrier-sense multiple access (CSMA) algorithm and therefore can manage competition between large numbers of nodes [8].

The safety communication in VANETs takes two forms: period safety messages, which are called beacons, and event driven messages, which are called emergency messages. Emergency messages are usually triggered by events and transmitted by multi-hop broadcasts, also known as decentralized environmental notification messages in the European Telecommunications Standards Institute (ETSI) standard [9].

The transmission of emergency messages in multi-hop scenarios involves relaying messages through neighboring vehicles in a hop-by-hop manner to reach all vehicles in the affected area. The high mobility of vehicles causes dynamic changes in the network topology, resulting in transmission channel competition and collisions during emergency message delivery. This poses significant challenges in the design of a multi-hop emergency message delivery mechanism. Strategically selecting the next hop in VANETs is essential for enhancing the system performance.

In [10], the authors propose a Cooperative Adaptive Cruise Control Lane Change (CACCLC) controller to enhance lane changes for CACC in congested traffic, improving lane change capability, success rate and stability. In [11], the authors suggest using system dynamics modeling for mixed traffic flow to gain insight into traffic dynamics and their effects on traffic conditions. In [12], the authors introduce a Route Segmented Broadcast Protocol using RFID technology (RSBP-RF) and proactive relay vehicle selection to enhance broadcasting efficiency.

In [13], the authors suggest using a probabilistic broadcast protocol that considers the distance, link availability, and packet reception rate as weighted probabilities for potential

C.M. Wu is with the Department of Computer Science and Information Engineering, Nanhua University, Chiayi, Taiwan, 62248.
E-mail: cmwu@nhu.edu.tw
The authors would like to thank the editor and the reviewers for their valuable comments and suggestions. This work is supported by Ministry of Science and Technology, Taiwan under grant MOST 110-2221-E-343-002 - and Research Project of Nanhua University under grant Y112000445.

relay candidates. In [14], the authors propose a reliable multi-hop broadcast mechanism using Bayesian networks and unipolar orthogonal codes to ensure high reliability and low delay in emergency message broadcasts under different channel conditions. In [15], the authors suggest using an adaptive topological area partitioning and broadcasting technique to enhance the reliability of emergency message broadcasts and beacon retransmissions. In [16], a mechanism for multi-hop emergency message transmission was proposed, along with an assessment of the reliability of radio transmitters. In [17], a multi-hop emergency message broadcast algorithm that emphasizes the success rate and distance was introduced.

If the problem of collision associated with the broadcast of emergency information is overcome, then the transmission efficiency will be improved. Thus, vehicles can both immediately receive emergency information to avoid danger and provide safety for smart transportation. Therefore, the selection of the CW value for the IEEE 802.11p backoff mechanism will play a pivotal role for emergency broadcasts in VANETs. However, the traditional MAC protocol cannot fulfill the requirements for high reliability and low delay simultaneously.

Machine learning (ML) is an artificial intelligence method through which a system learns from past data and experiences. QL is an RL method, whereas RL is a branch of ML. ML, RL, and QL are discussed in the existing literature on various aspects of wireless networks [18], [19], [20], [21], [22], [23].

Reinforcement learning leverages the spatial and temporal characteristics of nodes to enhance model accuracy. The system efficiency is consequently enhanced by transmitting regular messages to neighboring nodes to ascertain the next optimal hop node [18]. In [19], the authors proposed dynamically adjusting the reinforcement learning parameters and HELLO message intervals to improve the accuracy of detecting neighboring node positions in flying ad hoc networks (FANETs).

Previous studies have demonstrated that QL can help communication networks learn their optimal transmission strategies. This technology, in conjunction with a proper learning time, can achieve performance comparable to those of real-time statistical communication networks [20]. RL has also been proposed for the channel access problem of wireless networks. Amuru et al. [21] optimized the IEEE 802.11 backoff mechanism in the form of a Markov decision process (MDP) and proposed an RL solution. Liu and Elhanany [22] used RL to find optimal solutions for efficient channel-sharing techniques in wireless sensor networks. Wu et al. [23] explored the use of the QL-based MAC protocol to reduce transmission delay in VANETs. However, in these past efforts, the characteristics of broadcasting, convergence problem, and immediacy of information for the MAC protocol in VANETs have not been considered.

The main focus of this paper is the design of an adaptive regulated contention window for emergency message broadcasts according to traffic for IEEE 802.11p in VANETS. It surpasses the previous related works because it also illustrates the following unique features of IEEE 802.11p in VANETS, which are expected to affect the system performance considerably:

1) multi-hop emergency message broadcasts;
2) an adaptive regulation CW due to environment variation and related phenomena, such as variation of the number of nodes and emergency message broadcasts collision for using same CW for IEEE 802.11p in VANETS;
3) critical requirements for emergency message broadcasts for IEEE 802.11p in VANETS, such as reliability and delay;
4) an optimal CW regulation method for emergency message broadcasts for IEEE 802.11p in VANETS, such as machine learning (ML), reinforcement learning (RL), and Q-learning (QL).

As a result of bullet point 4 mentioned above, the proposed RL method focuses on adaptive regulation of the CW to achieve high reliability and low delay for IEEE 802.11p in VANETS.

RL emphasizes how to act based on the environment to maximize expected benefits. In a system that operates according to an RL algorithm, reward determines the value of an action, and the value of each action determines subsequent steps. Therefore, RL is suitable for VANET environments and is expected to provide high reliability and low delay for emergency information broadcasts. Hence, in this study, IVC was used as the framework to develop a highly reliable and low-delay broadcast control mechanism based on RL for VANETs.The contributions of our study are as follows.

Firstly, we leveraged an optimal CW size based on RL to mitigate the collision probability associated with emergency message broadcasts. If the CW size can be suitably selected for all broadcasting vehicles, then the probability of successful broadcasting will be increased.

Secondly, we defined the reliability reward for emergency message broadcasts for performance optimization, the estimated number of neighbor vehicles and the current CW size are used to calculate the probability of successful emergency broadcasting. Each vehicle sends data, the original vehicular ID, the retransmit vehicular ID, the data sequence number, the number of transmissions, and the CW size for each emergency message broadcast by piggybacking. Therefore, each vehicle can estimate the number of one-hop neighbor vehicles by overhearing. The reliability reward can then be calculated using the collected information and the current CW size.

Thirdly, we achieved high reliability and low delay of emergency message broadcasts for performance optimization, the weighting parameters of the reliability reward and delay reward are determined. If high reliability and low CW size are achieved, then system throughput is maximized and backoff delay is minimized. We determine the optimal weighting parameters of the reliability reward and delay reward while considering reliability and delay simultaneously.

The remainder of this paper is organized as follows. The system model is described in Section II. The proposed QL-based high-reliability and low-delay MAC protocol for VANETs is detailed introduced in Section III. The performance of the proposed QL-based MAC for VANETs is evaluated and some numerical results obtained by simulation are presented in Section IV. Finally, the paper is concluded in Section V.

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2024.3418778

II. System Model

This section introduces our emergency message broadcast model in a multi-hop VANET. In our scenarios, the following assumptions apply [24].

1) Each vehicle in our emergency message broadcast model has a fixed transmission range, which is denoted by R.
2) The vehicles in our emergency message broadcast model are uniformly distributed over a three-lane highway.

The message broadcast model is utilized to transmit non-urgent information, such as general communication or data sharing, on a periodic basis. To save bandwidth and energy, the broadcast protocol is limited to a one-hop range instead of broadcasting to all the nodes. The emergency message broadcast model is similar to the message broadcast model, but is specifically designed to deliver emergency messages such as disaster alerts and first aid instructions. The aim is to reach as many nearby affected vehicles as possible. In this study, the emergency message broadcast transmission range was defined within a 3-hop range, with emergency messages prioritized over general messages. However, this study focuses solely on emergency message broadcasts.

The agent-environment interaction model explains how vehicle agents perceive and adapt to changes in the VANET using reinforcement learning. The reliability model examines potential collisions in emergency message broadcasts in VANET that lead to decreased broadcast reliability. The vehicle agent determines its next action by sensing changes in the environmental state of the map and by using the reliability and delay values from the reliability model.

A. Emergency Message Broadcast Model

In Fig. 1, the three-lane straight highway scenario in two directions for our emergency message broadcast model was set as a one-dimensional model. The emergency message source vehicle Src triggers an emergency event. Src sends an emergency message to its one-hop neighbors within its transmission range r. Then, all the one-hop neighbor vehicles select one CW value using the QL-based MAC protocol and broadcast the first emergency message. The two-hop neighbor vehicles receive the first broadcast emergency message. All two-hop neighbor vehicles then select one CW value using the QL-based MAC protocol and broadcast the second emergency message. The three-hop neighbor vehicles receive the second broadcast emergency message. Then, all three-hop neighbor vehicles select one CW value using the QL-based MAC protocol and broadcast the third emergency message.

In our model, a multi-hop QL-based MAC protocol for emergency message broadcasts in VANETs is employed to use a optimal CW value, and then to maximize the probability of successful broadcasting. In a VANET setting, using optimal CW values allows vehicles to stagger message broadcast times, reducing the risk of collisions. The decreased collision probability improves the likelihood of successful emergency message broadcasts and reduces message delivery delays. An increase in successful broadcasts indicates improved reliability of emergency message broadcasts. Thus, achieving high reliability and low delay maximizes the probability of successful broadcasting.

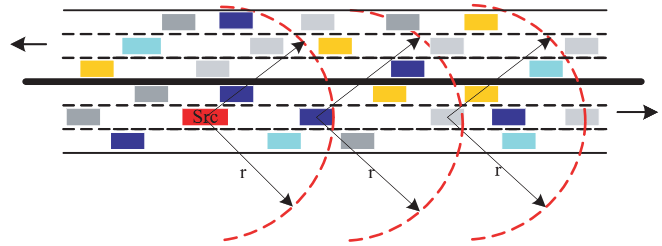


Fig. 1. Demonstration of emergency message broadcast scenario in multi-hop VANET.

As shown in Fig. 2, after the QL mechanism is executed, the agent learns the results and decide which action will be used in the VANET. Each vehicle in the VANET must interact with other vehicles in the VANET. The agent selects the optimal action and creates a new situation. Then, the environment creates a reward after the selected action is executed. The agent selects the optimal action to maximize the reward. The procedure is repeated until a maximized reward is achieved.

Here, S_t represents the contention window size at time t in VANET, indicating the time vehicles must wait before attempting to broadcast again after a collision. This factor significantly impacts channel acquisition efficiency and emergency message broadcasting effectiveness in VANETs. R_t represents the reward for emergency message broadcasting at time t in the VANET environment. In this paper, R_t examines factors such as reliability and delay in emergency message broadcasting and evaluates the effectiveness of broadcasting emergency messages at a specific time. In the VANET, a_t represents the optimal action chosen by the agent at time t under S_t after receiving R_t. This action may involve considerations related to the reliability and timeliness of emergency message broadcasting. The QL algorithm uses the optimal action to maximize overall reward, enhancing emergency message broadcasting performance in VANET.

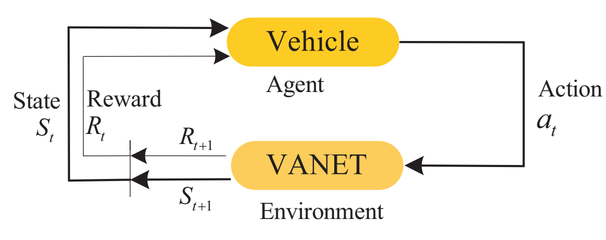


Fig. 2. Agent-environment interaction in an MDP.

Therefore, the reciprocal actions or influences among the vehicles (agents) in the VANETs and the environment at time step t are as follows [25], [26].

1) The vehicle in the VANET observes the environment and obtains the current CW value, S_t, after one beacon interval.

2) The vehicle in the VANET obtains the new Q-table, and the next action a_t is determined by itself.
3) a_t is performed by an emergency message broadcasts application in the MAC protocol. The vehicle receives reward R_{t+1} reaction after one time step in the VANET environment.
4) The vehicle in the VANET moves the state from S_t to S_{t+1}.

B. Reliability of Emergency Message Broadcasting

Reliability is defined as the ability of a system to perform the specified function within a specified period of time. The system performance may be affected by the signal interference when the data transmitted. The interference of emergency message broadcast in VANETs among one-hop neighbors can cause unreliability [27].

The single hop reliability issue in VANETs has been investigated [28]. However, a vehicle broadcasts an emergency message to its neighbors when an emergency condition is detected in the VANET in a multi-hop fashion. The reliability of emergency message broadcasts can be modeled the broadcasting successful probability before the expiry of maximum hop count. The reliability of emergency message broadcasts cab be evaluated by the packet loss probability and hop count as the metrics. The packet loss probability is a function of the channel error rate, message collision probability, and number of vehicles in the VANET. When the hop count for emergency message broadcasts increases, the packet loss probability increases and reliability decreases. The packet loss probability increases when the number of contending vehicles in the VANETs increases. Reliability is defined as the successful probability of emergency message broadcasts [27].

Sattar et al. [29] reported that in IEEE 802.11p, a vehicle will enter the backoff time mode when its channel is busy. They assumed that each emergency message could be broadcasted only once. While the emergency message is being broadcasted by the vehicle, the state of the channel (i.e., whether it is busy) can be determined. For IEEE 802.11p, the range of CW values for emergency message broadcasts in VANETs is [0, CW_{min}]. The no-collision probability of each emergency message broadcast in a VANET is defined as follows [29]:

$$P_c' = (1 - \frac{1}{CW_{min}})^{n-1}, \tag{1}$$

where CW_{min} is the minimum size of the contention window, $n = \rho\pi r^2$ denotes the expected value of the number of neighbor vehicles in the VANET, ρ is the number of vehicles per unit area, and r is the transmission range of the vehicle.

Then the reliability that a vehicle can successfully receive a packet is defined as follows [29]:

$$R = 1 - (1 - P_c')^n. \tag{2}$$

However, in general, the total number of vehicles in a VANET cannot be known, and the vehicles will not be evenly distributed. The number of vehicles per unit area cannot be known. Furthermore, the number of broadcasts for each emergency message in a VANET should not be limited to only one. In this study, the reward generated by the maximum number of broadcasts of emergency messages was included in the definition of the reliability reward. Therefore, the probability of successfully broadcasting an emergency message at each time can be defined as follows:

$$P_{suc}^t = (1 - \frac{1}{CW_t})^{d_{avg}^t - 1}, \tag{3}$$

where CW_t is the size of the contention window at time t. The term d_{avg}^t can be estimated by the average number of neighbor vehicles based on overhearing the vehicle number, data sequence number, and rebroadcasting times, which are carried by rebroadcasts from the neighbors. d_{avg}^t for each vehicle will vary over time due to different relative mobility and corresponds to CW_t at time t. Then, P_{suc}^t will also vary over time due to the differences in d_{avg}^t and CW_t at time t.

Then, the reliability that a vehicle can successfully receive a broadcasting emergency message at time t can be redefined as follows:

$$Rel_{rcv}^t = 1 - (1 - P_{suc}^t)^{d_{avg}^t}. \tag{4}$$

The reliability of emergency message broadcast in a VANET depends on the success probability of message transmission and number of neighboring vehicles. In a VANET, the number of neighboring vehicles can be estimated using a fixed broadcast transmission distance. The probability of node collision can be reduced by adjusting the competition window size. Therefore, adjusting the competitive window size can improve the reliability of emergency message broadcasts by decreasing the collision probability.

III. QL-Based High-Reliability and Low-Delay MAC Protocol for VANETs

For our proposed protocol, an IVC architecture was designed, and an adaptive broadcast control mechanism for emergency messages based on RL for use in VANETs was developed. The neighbor vehicle broadcasts the received emergency message, which contains the original vehicle ID, current vehicle ID, data number, transmission times, and CW value, by piggybacking. All neighbor vehicles can then learn the CW values used by other neighbor vehicles and the numbers of rebroadcasting by overhearing. The mechanism for the selection of the optimized CW value is designed based on a QL reward mechanism to achieve high reliability and low delay in the VANET. Table I lists the symbols used in the performance evaluation of our proposed protocol.

A. Control Channel Descriptions

For our proposed QL-based μRr+νRd-MAC protocols for VANETs, time is divided into several time slots. Each beacon interval has two periods: a backoff period and a data period. As shown in Fig. 3, each vehicle wants to broadcast an emergency message after receiving the original emergency message from the original vehicle. It select a backoff time according to the

TABLE I
SYMBOLS FOR PERFORMANCE EVALUATION.

S_t	State at time t
R_t	Reward at time t
a_t	Action at time t
S_{t+1}	State at time $t+1$
R_{t+1}	Reward at time $t+1$
a_{t+1}	Action at time $t+1$
a	Action
P_c'	No-collision probability of broadcasting
CW_{min}	Minimum size of the contention window
n	Expected number of neighboring vehicles
ρ	Number of vehicles per unit area
r	Transmission range of the vehicle
π	Circumference ratio
R	Reliability for packet reception
P_{suc}^t	Broadcast success probability at time t
CW_t	Contention window size at time t
CW_{t+1}	Contention window size at time $t+1$
d_{avg}^t	Average neighbors at time t
Rel_{rcv}^t	Reliability of message reception at time t
μ	Regulated weight for reliability reward
ν	Regulated weight for delay reward
$Q(s,a)$	Q-value of state s and action a
$\pi(s)$	Policy at state s
$\underset{a}{argmax}\, Q(s,a)$	Action maximizing the Q-value in state s
T_{simu}	System simulation time
T_{run}	Running time
λ	Decay constant for ε-greedy method
ε	Value of ε
P_ε	Selected randomly probability from (0, 1)
$CW^{a_{t+1}}$	Contention window size for action a_{t+1}
N_{max}	Maximum number of emergency broadcasts
R_{rel}	Emergency broadcasts reliability reward
R_{rb}	Reward of rebroadcasting time
R_d	Delay reward
R_{tol}	Total reward
γ	Discount factor
α	Learning rate
$\zeta_{n_{rb}}$	Total throughput from the 1st to n_{rb}th broadcasts
R_{CH}	Data transmission rate for an unlicensed channel
n_{rb}	Broadcasting times
$t_{trans}^{(i,j)}$	Time of jth vehicle for ith successful broadcast
n_{succ}^i	Number of vehicles for ith successful broadcast
$t_d^{(i,j)}$	Delay of jth vehicle for ith successful broadcast
$T_d^{n_{rb}}$	Average delay of n_{rb} successful broadcast
i	ith successful broadcast
j	jth vehicle

QL-based μRr+νRd-MAC protocol and then sends it. The control channel of the QL-based μRr+νRd-MAC protocol for VANETs is shown in Fig. 3. The two phases of the proposed QL-based μRr+νRd-MAC protocol can be described in detail as follows.

- Backoff period: In the backoff period, each vehicle sends the emergency message by selecting one CW value according to the proposed high reliability and low delay QL-based MAC protocol. The neighbor vehicles that receive the emergency message broadcast it again using the optimal CW value for the VANET.
- Data period: After the backoff period, each vehicle in the VANET can transmit an emergency message. The original vehicle ID, retransmit vehicle ID, data sequence, the number of transmissions, and CW value will be included in the emergency message by piggybacking.

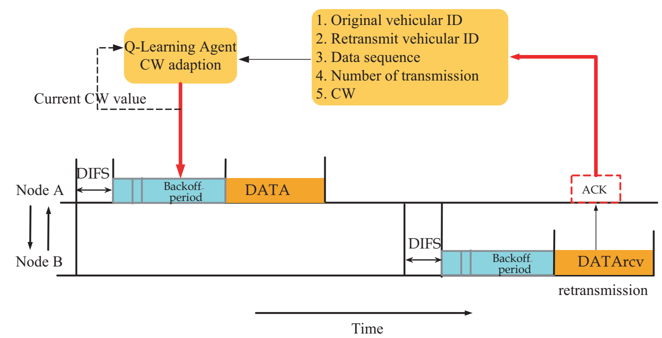


Fig. 3. QL-based high reliability and low delay MAC (QL-based μRr+νRd-MAC) protocol.

B. Action Selection

The main purpose of applying QL in VANETs is to achieve high reliability and low delay system performance. The optimized action is selected using the optimized Q-value of the QL mechanism. The emergency message broadcasts are performed using different CW values (state S). The optimized CW value is determined based on overhearing the rebroadcast results of neighbor vehicles. The experience gained from the CW value achieves the mechanism of self-learning channel access control proposed in this paper.

Watkins and Dayan [20] proved that as long as all actions are performed repeatedly in all states (state S, action a), QL can make the system converge and achieve the goal of a convergence probability of 1. In [20], the generated Q-table is related only to the number of CW used times, to fulfill the requirements for fairness. However, this mechanism cannot guarantee that the system will be able to achieve high reliability and low delay.

In our study, to maintain the original IEEE 802.11p implementation specification, the BEB algorithm was also adopted for the intervals of the CW values. In this approach, the seven CW values are [3, 7, 15, 31, 63, 127, 255]. Therefore, seven states [3, 7, 15, 31, 63, 127, 255] of QL and three kinds of actions exist. Emergency messages will be transmitted through these three actions to learn the optimized Q-value. Here, the states and actions are similar to the defined values in [7], [30]. The selected CW is the common value of all collected CW values for all the one-hop neighbors to achieve fairness in [7], [30]. In this study, the CW value was adaptively regulated based on the reward function for reliability and delay requirements.

In state CW_t, the vehicle agent takes the optimized action by observing the optimal Q-value in the Q-table and enters state CW_{t+1}. The state will then be changed by the action selection, as follows:

$$CW_t \xrightarrow{a\in((CW_t-1)/2,CW_t,2*CW_t+1)} CW_{t+1}. \quad (5)$$

The equation shows that in a VANET, when a vehicle node agent broadcasts emergency messages, it must use Q-learning to select the action at with the highest Q value from the Q table. For $a \in ((CW_t-1)/2, CW_t, 2*CW_t+1)$, the next action a is chosen from the above three options to determine the best action a. The QL action a determines the optimal competition window size for selecting the next emergency message broadcast to prevent collisions.

One of the actions with the largest estimated value is selected by the simplest action selection method. The action is selected randomly when there are many "greedy" actions. The selection mechanism for a greedy action can be presented as [25]:

$$\pi(s) \doteq \underset{a}{argmax}\, Q(s,a), \tag{6}$$

where $\underset{a}{argmax}$ denotes that the expression is maximized under action a.

However, first convergence cannot be achieved by continuous use of the greedy mechanism. Moreover, the greedy mechanism cannot correctly discover all pairs of (State,Action). On the other hand, pairs of (State, Action) will be ongoingly discovered when the random mechanism is used. However, the random mechanism for the controller is suboptimal. A proposed compromise is to utilize the ε-greedy method between random and greedy mechanisms [25]. The ε-greedy method is executed with a probability of 1. Good system performance ensures a balance between exploration and exploitation. The value of ε is defined as follows:

$$\varepsilon = e^{-\lambda \frac{T_{run}}{T_{simu}}}, \tag{7}$$

where T_{run} is the running time and T_{simu} is the system simulation time. In our proposed QL-based MAC for VANETs, convergence to an optimal policy is guaranteed by the decay function.

C. Reward function formulation

Pressas et al. [7] proposed the use of constant rewards, specifically, [1/7, 2/7, 3/7, 4/7, 5/7, 6/7, 1]. These rewards are defined according to the order of the number of used times of different CW values. However, this method cannot reflect the characteristics of high reliability and low delay for emergency message broadcasts in VANETs.

In this study, the reliability reward was redefined according to the requirements for reliability, the transmission frequency reward was defined according to the number of emergency message broadcasts, and the delay reward was defined according to the requirements for low delay. These defined rewards help the proposed QL-based MAC protocol achieve high reliability and low delay in VANETs.

1) Reliability reward: Each vehicle can receive the sequence number and number of rebroadcasting of each emergency message by piggybacking. Each emergency message will be sorted according to the number of rebroadcasts. A higher number of rebroadcasts will yield a lower reward, whereas a lower number of rebroadcasts will yield a greater reward. In this way, the number of rebroadcasting of emergency message for each vehicle will remain fair.

If the numbers of transmissions are $[1, 2, \cdots, N_{max}]$, then the rewards of rebroadcasting time R_{rb} will be $[1, (N_{max} - 1)/N_{max}, \cdots, 1/N_{max}]$, where N_{max} denotes the maximum number of emergency broadcasts. R_{rel} for each vehicle will vary over time due to different P_{suc}^t and CW_t at time t under different R_{rb}. Then, the reward of reliability for emergency message broadcasts can be defined as follows:

$$R_{rel} = [1 - (1 - P_{suc}^t)^{d_{avg}^t}] R_{rb}. \tag{8}$$

Here, P_{suc}^t denotes the likelihood of a vehicle effectively transmitting an emergency message to adjacent vehicles. Hence, it is essential to assess the dependability of the broadcasting procedure. d_{avg}^t has a direct influence on the impact and effectiveness of broadcasting. A higher value of d_{avg}^t indicates that each vehicle has a greater number of neighbors to whom messages can be forwarded, thus enhancing the probability of message broadcasting. R_{rb} introduces the temporal dimension into reward calculations (Eq.(8)) and acknowledges the urgency of emergency message broadcasting.

2) Delay reward: The delay of an emergency message broadcast in a VANET has a close relationship with the CW value: if the CW value is small, then the backoff delay will be small, whereas if the CW value is large, then the backoff delay will be large.

In this study, a large CW value will result in a low delay reward R_d, whereas a small CW value will result in a high R_d. CW values of [3, 7, 15, 31, 63, 127, 255] correspond to R_d values of [1, 6/7, 5/7, 4/7, 3/7, 2/7, 1/7] [7].

3) Total reward: To be consistent with the objective of the QL algorithm, the agent is to obtain the maximum reward for each iteration of learning. If the reward is small, then the action is not good. Therefore, the total reward of high reliability and low delay for emergency message broadcasts in VANETs can be defined as the sum of the high-reliability reward and low-delay reward with two adjusting weighting parameters:

$$R_{tol} = \mu R_{rel} + \nu R_d, \tag{9}$$

where μ and ν are regulated weighting parameters for the reliability and delay reward, respectively. If μ increases, then R_{tol} needs to put more emphasize on the reliability for emergency message broadcast and the system throughput will be increased. If ν increases, then R_{tol} needs to put more emphasize on the delay for emergency message broadcast and the end-to-end delay will be decreased.

This study examines performance indicators, such as reliability and latency, across different scenarios. A tradeoff exists between reliability and latency. The weighting parameters are assessed and adjusted based on the observed performance metrics. This discussion focuses on determining the optimal parameters by considering the number of vehicles and topology structures to achieve high reliability and low latency. We adjusted the parameters μ and ν based on the stability of the state for different numbers of vehicle agents. First, we set μ to one and ν to zero to guarantee positive reliability returns during the learning process. We set μ to 0 and ν to 1 to ensure positive returns on transmission delay during the learning process. During the extreme adjustment processes, we observed a significant difference in system throughput and delivery delays caused by reliability improvements. Subsequently, we adjusted μ and ν while monitoring the changes in R_{tol} during the simulation. The vehicle agent uses QL to maintain learning

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2024.3418778

and achieve stability. Finally, the system is monitored to assess whether the throughput and delay for emergency message broadcasts reach the optimal values to determine the values of μ and ν.

In general, the discount factor γ is set to range from 0.6 to 0.99 [30]. Its value is considered to be part of the problem. Additionally, the learning speed increases in accordance with the increases of learning rate α and vice versa. No new messages are generated when α is 0. On the other hand, only the most recent new message will be considered when α is 1.

In the QL-based MAC protocol for VANETs, the action is selected based on P_ε. The CW value selects one value randomly from $((CW_t - 1)/2, CW_t, 2 * CW_t + 1)$ when $P_\varepsilon < \varepsilon$. Otherwise, the action is determined by the controller (Eq.(6)). P_ε is selected randomly from (0, 1).

Expanding on the prior design, each state has the potential to display a distinct reward gradient, presenting a challenge for the agent in addressing variations in gradients across states. A more efficient reward function not only yields greater information, but also markedly expedites the convergence speed of QL algorithms. It is important to note that rewards customized for specific objectives adapt dynamically in response to the feedback received. This paper introduces the proposed QL-based μRr+νRd-MAC protocol, which is designed to accelerate the convergence of CW in the network, with the goal of achieving high reliability and low delay objectives.

Algorithm 1: QL-based μRr+νRd-MAC Protocol

01: Initialize $Q_0(CW, a)$ at $t = 0$
02: if $T_{run} < T_{simu}$ then
03: $\quad\varepsilon, \alpha \longleftarrow$ decay function
04: else
05: $\quad\varepsilon, \alpha \longleftarrow$ constant
06: end if

07: procedure Action-selection(CW_t)
08: $\quad$randomly select $P_\varepsilon \in$ random (0, 1)
09: $\quad$if $P_\varepsilon < \varepsilon$ then
10: $\quad\quad a_{t+1} \longleftarrow$ random $((CW_t - 1)/2, CW_t, 2 * CW_t + 1)$
11: $\quad$else if $P_\varepsilon \geq 1 - \varepsilon$ then
12: $\quad\quad a_{t+1} \longleftarrow a_\pi$
13: $\quad$end if
14: $\quad CW_{t+1} \longleftarrow CW^{a_{t+1}}$
15: end procedure

16: procedure Feedback(CW_{t+1}, a_{t+1})
17: $\quad$each vehicle sends the Src ID, broadcasting vehicle ID, CW, and rebroadcasting times by piggybacking.
18: $\quad$each vehicle collects the information about its neighbors by listening.
19: $\quad$each vehicle estimates d^t_{avg}.
20: $\quad$each vehicle calculates P^t_{suc}.
21: $\quad$each vehicle calculates R_{rb}.
22: $\quad$each vehicle calculates R_{rel} of Eq.(8).
23: $\quad R_{rel} = [1 - (1 - P^t_{suc})^{d^t_{avg}}]R_{rb}$
24: $\quad CW$ values of [3, 7, 15, 31, 63, 127, 255] corresponds R_d of [1, 6/7, 5/7, 4/7, 3/7, 2/7, 1/7].
25: $\quad$each vehicle calculates R_d.
26: $\quad$each vehicle calculates R_{tol} of Eq.(9).
27: $\quad$set μ and ν.
28: $\quad\mu$ and ν are corresponding parameters for the reliability and delay reward, respectively.
29: $\quad R_{tol} = \mu R_{rel} + \nu R_d$
30: $\quad$update $Q(CW_{t+1}, a_{t+1})$
31: $\quad$Action-selection(CW_{t+1})
32: end procedure

Algorithm 1 outlines the steps for our proposed QL-based μRr+νRd-MAC protocol for the emergency message broadcast mechanism in VANETs based on reinforcement learning with contention estimation. The following instructions outline each step.

1) **Initializing Q-Values in Q-Learning (Step 1): The initialization of Q-values impacts how vehicle agents balance exploration and exploitation, the speed of convergence, and the quality of the learned strategy. Here, $Q_0(CW, a)$ is set to encourage thorough exploration, ensuring that vehicle agents can effectively learn the optimal strategy for emergency message broadcasting through comprehensive exploration.**
2) **Decay function (Steps 2-6): Initially, a high ε value ensures thorough exploration of the state-action space. As ε decreases, vehicle agents prioritize exploiting high-reward behaviors they are familiar with. A constant value for ε is established after a specific simulation time to ensure the adaptability of vehicle agents and prevent premature convergence-related system performance issues.**
3) **Action-selection (Steps 7-15): Vehicle agents balance exploration and exploitation using the QL-based μRr+νRd-MAC protocol with an ε-greedy strategy. During exploration, vehicle agents randomly select contention window sizes to explore new actions. Vehicle agents utilize the Q-learning algorithm to choose actions with the highest Q-values. By balancing exploration and exploitation, the optimal contention window size is selected to improve the efficiency of emergency message broadcasting in VANETs.**
4) **Information collection (Steps 17-19): Vehicle agents share their state and actions with neighbors through piggybacking. They collect data from nearby vehicles by monitoring the network. The collected data is utilized to calculate the average number of neighbors, aiding in understanding network conditions and optimizing broadcast strategies to enhance the reliability of emergency message broadcasting in VANETs.**
5) **Probability calculation (Steps 20-21): Each vehicle agent assess the effectiveness of its broadcasting strategy using success probability and broadcast reward. Transmission success assessment evaluates the probability of successful emergency message broadcasts. Reward calculation quantifies the effectiveness of broadcasting and impacts the learning process.**

This optimizes the broadcast strategy by accurately assessing and improving the reliability and efficiency of emergency message broadcasting in VANETs.

6) **Reliability reward calculation (Steps 22-23): The reliability reward is calculated by considering the probability of successful transmission to neighbors and the effectiveness of rebroadcasting. This enhances the overall performance and robustness of VANETs.**
7) **Delay reward calculation (Steps 24-25): Vehicle agents assign delay rewards based on predefined contention window sizes to quantify the timeliness of emergency message broadcasting. This ensures a balance between reduced delay and improved reliability in emergency message broadcasting in VANETs.**
8) **Total reward calculation (Steps 26-29): The strategy for broadcasting emergency messages by vehicle agents combines reliability and delay rewards. The proposed Q-learning algorithm utilizing total reward helps to determine the optimal balance between reducing delay and enhancing reliability in emergency message broadcasting.**
9) **Convergence (Step 30): The vehicle agents update Q-values based on the total reward. This allows the system to learn from experience and converge towards the optimal solution over time.**
10) **After completing the convergence step, the process returns to Action-selection. Vehicle agents determine the contention window size for the next emergency message broadcast based on the current state and learned knowledge represented by Q-values.**

IV. Performance evaluation of a VANET

We ran our simulation programs on a computer with a single CPU (Intel Core i7-9700K) and a single GPU (NVIDIA Quadro RTX 4000). We used Keras as the learning platform to train neural networks [31]. The simulation results of our proposed QL-based μRr+νRd-MAC protocol as applied in the VANET are presented in this section. The simulation was implemented using the C an Python programming languages.

For QL-based Rcce-MAC [7], the optimal CW value is based on fairness and cannot be regulated according to the broadcast traffic. In addition, selecting the CW value based on the order of CW usage times cannot fulfill the requirements of reducing transmission delay and improving the reliability of emergency message broadcasts. By contrast, for QL-based RcceRd-MAC [7], the reward is the product of fairness and delay reward, thus achieving high fairness and low delay in VANETs.

In this section, the system performance of QL-based μRr+νRd-MAC will be compared with those of other protocols. Here, the vehicles in each VANET exhibit a multi-hop relationship. For our proposed QL-based μRr+νRd-MAC protocol, the CW value is optimized according to the reliability and delay reward. The bandwidth of the unlicensed band channel is 2 $Mbps$.

In the simulations, the number of vehicles was 50 or 100 stations and one vehicle broadcasted an emergency message to its one-hop neighbor vehicles. The one-hop neighbors for 100 stations had a high density in the bounded region. The bounded region was a three-lane highway with a length of 1200 m and a square region with an area of 200 × 200 m^2. The original emergency message broadcasts vehicle generated 32 packets per second. The mobility model was a Krauss model with default parameters ($\sigma = 0.5, \tau = 0.1$) [32]. All vehicles were placed on the three-lane highway with mobility velocities of 22 $m/s \sim 28$ m/s. The number of one-hop neighbors for each vehicle varied over time due to their different relative mobilities. Table II outlines the other parameters of all QL-based MAC protocols tested in the VANET simulations.

With regard to training and learning for QL, the discount factor γ was fixed at 0.9 and the learning rate α was fixed at 0.1. The simulation time was 100,000 s. Five different topologies were created using five seeds, and all simulation results are presented as the average values for these five seeds.

TABLE II
Parameters for our proposed QL-based μRr+νRd-MAC protocol.

Parameter	Value
Simulation time	100,000 s
Number of vehicles	10, 50, 60, 70 ,80, 100, 200
Three-lane highway length	1200 m
Square area	200x200 m^2
Transmission range of vehicle	100 m
Channel data rate	2 $Mbps$
Original message broadcasts rate	32 pkt/sec
Time slot	13 μs
Maximum times of rebroadcast (N_{max})	3
Rebroadcasting times (n_{rb})	1, 2, 3
Discount factor (γ)	0.9
Learning rate (α)	0.1
Mobility model	Krauss model (σ = 0.5, τ = 1)
Vehicle velocity	$22 - 28$ m/s
Vehicle acceleration ability	2.6 m/s^2
Vehicle deceleration ability	4.5 m/s^2

A. *Metrics of Emergency Message Broadcasting*

1) Throughput: When emergency message broadcasts reliability improves, its success probability also increases. As the probability of successful broadcast increases, the message delivery ratio also rises, maximizing system output. Therefore, system throughput is used as a performance indicator to investigate if reliability has been optimized.

Let $t_{trans}^{(i,j)}$ denote the transmission time of the jth vehicle for the ith broadcast of a successful emergency message in the VANET. Then, $\zeta_{n_{rb}}$ is the sum of throughput from the 1st to n_{rb}th broadcasts in the VANET and can be defined as follows [33]:

$$\zeta_{n_{rb}} = \frac{R_{CH} \sum_{i=1}^{n_{rb}} \sum_{j=1}^{n_{succ}^{i}} t_{trans}^{(i,j)}}{T_{simu}}, \tag{10}$$

where R_{CH} is the data transmission rate for an unlicensed channel, T_{simu} is the system simulation time, and n_{succ}^{i} de-

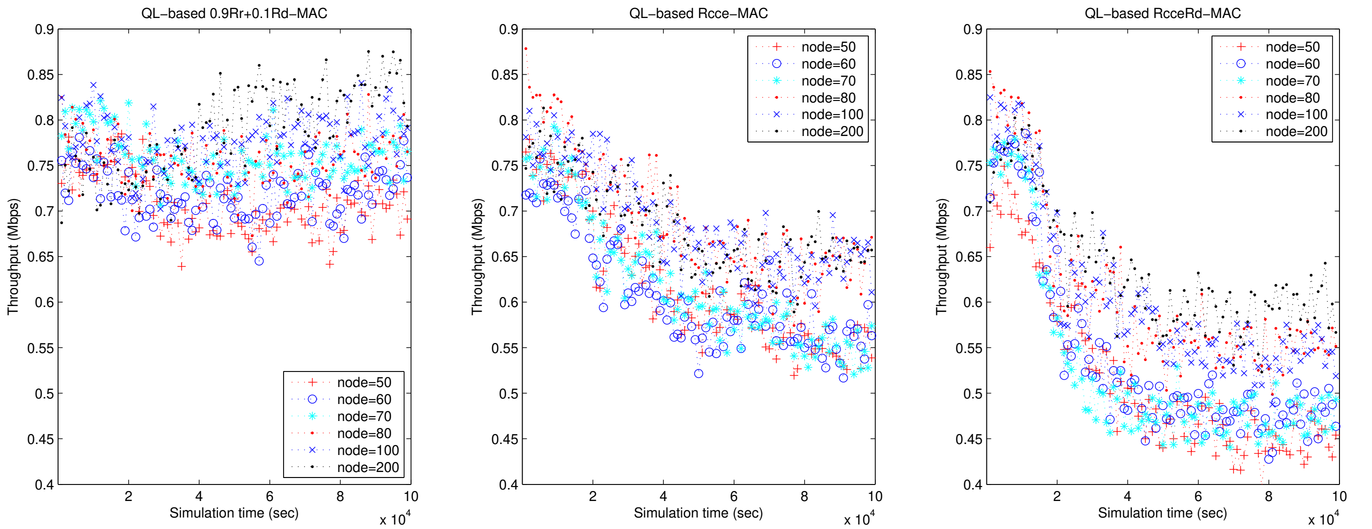


Fig. 4. System throughput of emergency message broadcasts for QL-based μRr+νRd-MAC comparison with other MAC protocols for different number of vehicles in VANET.

notes the number of vehicles for the ith successful emergency message broadcast in the VANET.

2) Delay: The average delay of each successful emergency broadcast with n_{rb} broadcasting times is denoted by $T_d^{n_{rb}}$. Let $t_d^{(i,j)}$ be the delay of the jth vehicle for the ith broadcasting for a successful emergency message in the VANET. The value of $T_d^{n_{rb}}$ can then be computed as follows [34]:

$$T_d^{n_{rb}} = \frac{\sum_{i=1}^{n_{rb}} \sum_{j=1}^{n_{succ}^i} t_d^{(i,j)}}{\sum_{i=1}^{n_{rb}} n_{succ}^i}. \tag{11}$$

B. Sensitivity analysis on the number of vehicles

In a VANET, the probability of emergency message broadcasts among vehicles increases with the growing number of vehicles. This also enhances the likelihood of message rebroadcasting through adjacent nodes, thereby enhancing the overall network throughput and reducing broadcast delays. As the number of vehicles increases and the system nears saturation, an excess of vehicles could lead to the occurrence of emergency message broadcast storms. The occurrence of such storms can lead to traffic congestion, accidents, reduced system capacity, and consequently, delays in emergency message broadcasts.

1) Throughput measurement: In a VANET on a three-lane highway, emergency message broadcast can be accelerated and enhanced with more nearby vehicles. This phenomenon enhances the overall network throughput. Hence, as the number of vehicles in a VANET on a three-lane highway increases, it enhances the potential for emergency message broadcast, introduces multiple broadcast paths, and offers more redundancy opportunities for relay nodes. This enhances system reliability and throughput. Therefore, based on Fig. 4, the proposed QL-based 0.9Rr+0.1Rd-MAC method for a three-lane environment with 50 nodes achieves a throughput range of 0.639 – 0.785 Mbps. When the node=200, the throughput ranges from 0.688 to 0.875 Mbps.

Additionally, as shown in Fig. 4, for a three-lane highway VANET, the system throughput increases with the number of vehicles, regardless of the MAC protocol used. When the number of nodes is between 50 and 200, the system throughput continues to increase as the vehicle density increases, indicating that the system has not reached saturation. This topic is closely related to the topology. In the following discussion, we further explain the saturation point problem by examining how various space distributions affect the system throughput. In the figure, QL-based μRr+νRd-MAC indicates that the system throughput remains stable as the simulation progresses. However, it is anticipated that the system throughput of QL-based Rcce-MAC and QL-based RcceRd-MAC will decrease as the simulation progresses.

TABLE III
SYSTEM THROUGHPUT OF EMERGENCY MESSAGE BROADCASTS FOR QL-BASED μRR+νRD-MAC COMPARISON WITH OTHER MAC PROTOCOLS IN 100-VEHICLE VANET.

100 − vehicle	First	Second	Third
Rr-MAC vs. Rcce-MAC	37.9 %	37.5 %	31.1 %
0.9Rr+0.1Rd-MAC vs. Rcce-MAC	34.0 %	39.7 %	42.2 %
0.5Rr+0.5Rd-MAC vs. Rcce-MAC	26.5 %	16.7 %	19.3 %
100 − vehicle	First	Second	Third
Rr-MAC vs. RcceRd-MAC	54.9 %	93.2 %	93.8 %
0.9Rr+0.1Rd-MAC vs. RcceRd-MAC	60.1 %	107.6 %	105.5 %
0.5Rr+0.5Rd-MAC vs. RcceRd-MAC	48.2 %	59.7 %	59.8 %

As shown in Table III, the greatest improvements in system throughput for the first broadcast of QL-based Rr-MAC, QL-based 0.9Rr+0.1Rd-MAC, and QL-based 0.5Rr+0.5Rd-MAC compared to that of QL-based Rcce-MAC are 37.9%, 34.0%, and 26.5%, respectively, whereas the greatest improvements in system throughput for the first broadcast of QL-based Rr-MAC, QL-based 0.9Rr+0.1Rd-MAC, and QL-based

This article has been accepted for publication in IEEE Transactions on Intelligent Vehicles. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIV.2024.3418778

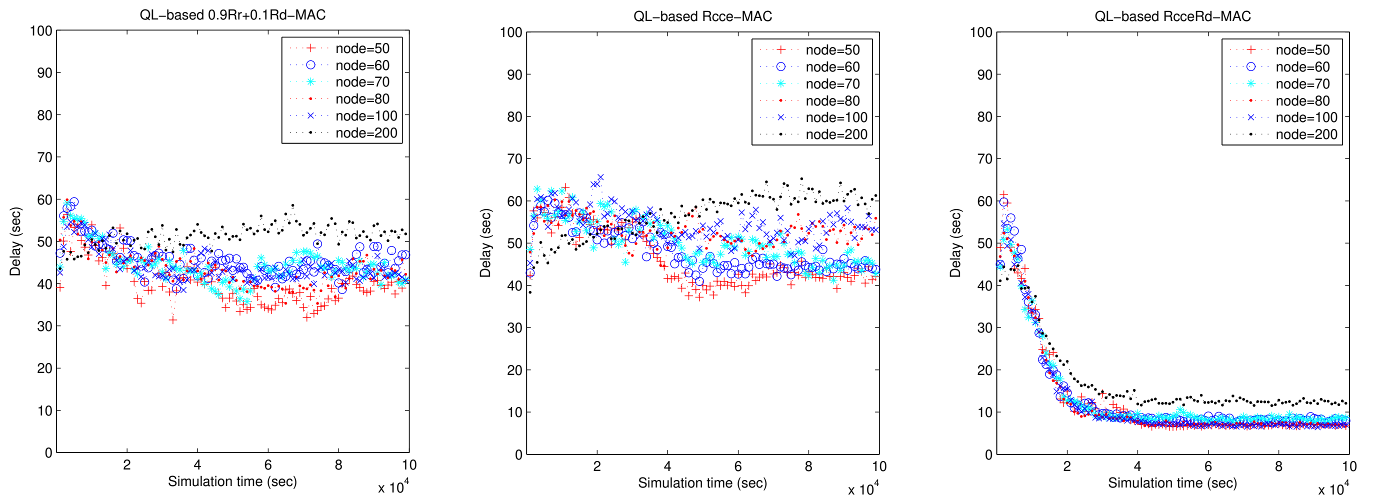


Fig. 5. Delay of emergency message broadcasts for QL-based μRr+νRd-MAC comparison with other MAC protocols for different number of vehicles in VANET.

0.7Rr+0.3Rd-MAC compared to that of QL-based RcceRd-MAC are 54.9%, 60.1%, and 48.2%, respectively.

The greatest improvements in throughput for the second broadcast of QL-based Rr-MAC, QL-based 0.9Rr+0.1Rd-MAC, and QL-based 0.5Rr+0.5Rd-MAC compared to that of QL-based Rcce-MAC are 37.5%, 39.7%, and 16.7%, respectively, whereas the greatest improvements in throughput for the second broadcast of QL-based Rr-MAC, QL-based 0.9Rr+0.1Rd-MAC, and QL-based 0.5Rr+0.5Rd-MAC compared to that of QL-based RcceRd-MAC are 93.2%, 107.6%, and 59.7

The greatest improvements in throughput for the third broadcast of QL-based Rr-MAC, QL-based 0.9Rr+0.1Rd-MAC, and QL-based 0.5Rr+0.5Rd-MAC compared to that of QL-based Rcce-MAC are 31.1%, 42.2%, and 19.3%, respectively, whereas the greatest improvements in throughput for the third broadcast of QL-based Rr-MAC, QL-based 0.9Rr+0.1Rd-MAC, and QL-based 0.5Rr+0.5Rd-MAC compared to that of QL-based RcceRd-MAC are 93.8%, 105.5%, and 59.8%, respectively.

2) Delay measurement: In densely populated VANET environments, an increase in neighboring vehicles generally reduces propagation delay. If vehicle density has already led to congestion, emergency messages may queue in the message queue, waiting for an available channel, causing delivery delays even when vehicles are nearby. Therefore, as illustrated in Fig. 5, the proposed QL-based 0.9Rr+0.1Rd-MAC method in a three-lane environment with 50 nodes yields a delay range of 31.396 – 56.311 seconds. When the node=200, the delay ranges from 43.547 to 58.537 seconds.

Additionally, in Fig. 5, for a three-lane highway VANET, such as a QL-based μRr+νRd-MAC or other MAC protocols, the transmission delay is expected to increase as the number of vehicle nodes increases. The transmission delay of the QL-based μRr+νRd-MAC is expected to be lower than that of the QL-based Rcce-MAC. Nevertheless, QL-based RcceRd-MAC considers the transmission delay caused by communication protocols. Regardless of the variations in the number of nodes, the delivery delay remains consistently minimal.

TABLE IV
AVERAGE DELAY OF SUCCESSFUL EMERGENCY MESSAGE BROADCASTS FOR QL-BASED μRR+νRD-MAC COMPARISON WITH QL-BASED RCCE-MAC IN 100-VEHICLE VANET.

100 − vehicle	First	Second	Third
Rr-MAC vs. Rcce-MAC	21.5 %	52.3 %	53.6 %
0.9Rr+0.1Rd-MAC vs. Rcce-MAC	26.7 %	56.5 %	62.8 %
0.5Rr+0.5Rd-MAC vs. Rcce-MAC	62.3 %	75.9 %	83.7 %

As shown in Table IV, the greatest improvements in the delay for the first broadcast of QL-based Rr-MAC, QL-based 0.9Rr+0.1Rd-MAC, and QL-based 0.5Rr+0.5Rd-MAC compared to that of QL-based Rcce-MAC in a 100-vehicle VANET are 21.5%, 26.7%, and 62.3%, respectively. The greatest improvements in the delay for the second broadcast of QL-based Rr-MAC, QL-based 0.9Rr+0.1Rd-MAC, and QL-based 0.5Rr+0.5Rd-MAC compared to that of QL-based Rcce-MAC in a 100-vehicle VANET are 52.3%, 56.5%, and 75.9%, respectively. Finally, the greatest improvements in the delay for the third broadcast of QL-based Rr-MAC, QL-based 0.9Rr+0.1Rd-MAC, and QL-based 0.5Rr+0.5Rd-MAC compared to that of QL-based Rcce-MAC in a 100-vehicle VANET are 53.6%, 62.8%, and 83.7%, respectively.

As shown in Table IV, the delays for the first, second, and third broadcasts of QL-based μRr+νRd-MAC are higher than those for QL-based RcceRd-MAC. However, QL-based μRr+νRd-MAC will have lower delays than QL-based Rcce-MAC at any value of ν. QL-based μRr+νRd-MAC at higher ν will have low delays similar to those for QL-based RcceRd-MAC.

C. Sensitivity analysis on the space distribution of vehicles

In VANETs, the impact of vehicle density on throughput is influenced by the surrounding environment's topology. In

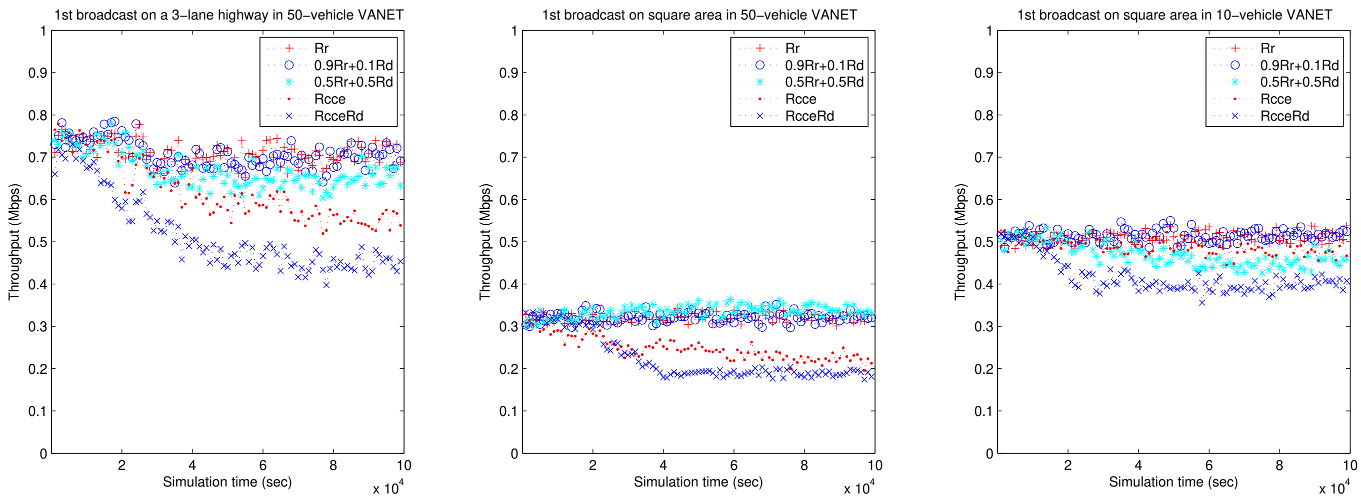


Fig. 6. System throughput of emergency message broadcasts for QL-based μRr+νRd-MAC in a 200x200 (m^2) square area region with 50- and 10-vehicles VANET.

three-lane straight highways, topological changes are generally less pronounced than in square areas. Emergency message broadcasts in three lanes typically follow fixed directions. In a vehicle environment, throughput is expected to be higher when vehicles travel in the same direction compared to a square area. This reduces the likelihood of head-on collisions, enabling more effective broadcast of emergency messages.

In square areas, emergency message broadcasts may exhibit greater diffusion and lower directionality than on highways. Vehicles move in various directions, resulting in intricate traffic patterns. The presence of multidirectional traffic may lead to a decrease in the throughput of emergency message broadcasts compared to highways, as vehicles are required to take detours more frequently. Such a scenario can lead to delays in broadcasting emergency messages and decrease throughput as vehicles are forced to wait before proceeding.

1) Throughput measurement: In the proposed QL-based 0.9Rr+0.1Rd-MAC method, Fig. 6 shows that in a tree-lane environment with 50 nodes, the throughput ranges from 0.639 to 0.785 Mbps. The throughput in a square area with 50 nodes ranges from 0.299 to 0.344 Mbps, while in an area with 10 nodes, it varies from 0.486 to 0.550 Mbps.

Additionally, in Fig. 6, when considering QL-based μRr+νRd-MAC or other MAC protocols, it is observed that the three-lane highway outperforms the square area in a 50-node VANET scenario, resulting in a higher system throughput for emergency message broadcasts. In a square area with 50 nodes moving at high speeds, significant changes occur in the network topology. Competition for channels leads to an increase in collisions during emergency message broadcasts, thereby intensifying and reducing system throughput. In the square area, the system throughput is higher with ten vehicle nodes than with 50 nodes because congestion and collisions are reduced. This observation indicates that when the number of vehicle nodes reaches 50, the capacity of the square area is surpassed, resulting in collisions that significantly diminish the system throughput.

2) Delay measurement: In the proposed QL-based 0.9Rr+0.1Rd-MAC method, depicted in Fig. 7, the delay range in a tree-lane environment with 50 nodes is 31.396 – 56.311 seconds. In a square area with 50 nodes, the delay ranges from 53.026 to 62.646 seconds, and in a square area with 10 nodes, it ranges from 30.384 to 68.052 seconds.

Additionally, in Fig. 7, whether considering QL-based μRr+νRd-MAC or other MAC protocols, it is observed that in a 50-node VANET scenario, the three-lane highway outperforms the square area in terms of emergency message broadcasts, resulting in lower delivery delays. In a square area with 50 nodes moving at high speeds, which results in frequent changes in the network topology, channel contention exacerbates collisions in emergency message broadcasts, thereby amplifying delivery delays. In a square area with 10 vehicle nodes, the transmission delay caused by the reduced congestion and collisions is lower than when there are 50 nodes. Within the square area, 50 nodes surpassed the saturation threshold of the system.

For QL-based μRr+νRd-MAC, the delay reward is considered to be the same as that for QL-based RcceRd-MAC, and the reliability and delay metrics are considered simultaneously. Therefore, the proposed QL-based μRr+νRd-MAC, when applied in VANETs, has a higher system throughput than QL-based Rcce-MAC and QL-based RcceRd-MAC, when the reliability reward is considered with suitable weightings. In addition, improvements in the delay for QL-based μRr+νRd-MAC also imply good performance with a higher ν value.

V. CONCLUSION

In this paper, we proposed a QL-based μRr+νRd-MAC for multi-hop VANETs. To mitigate collision probability in emergency message broadcasts, an optimal CW based on RL is proposed to separate the broadcasting time in the contention region. To achieve high reliability and low delay for emergency message broadcasts for performance optimization, the weighting parameters of reliability reward and delay reward were

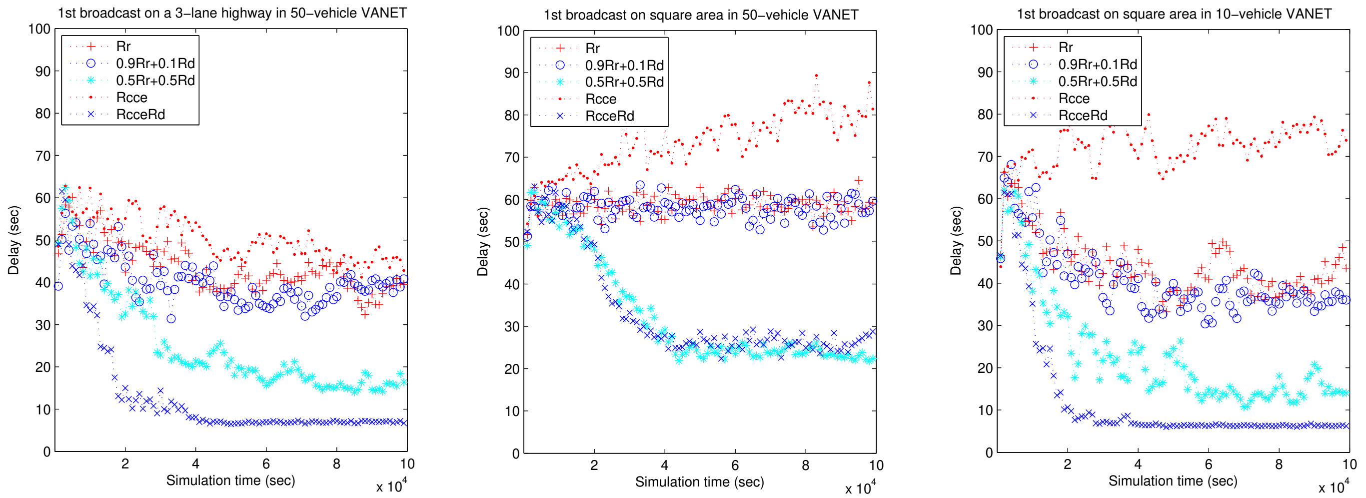


Fig. 7. Delay of emergency message broadcasts for QL-based μRr+νRd-MAC in a 200x200 (m^2) square area region with 50- and 10-vehicles VANET.

developed. The system throughput increases as μ is increased, and the backoff decreases as ν is increased. The simulation results show that the proposed QL-based μRr+νRd-MAC-MAC for VANETs has a higher system throughput and lower delay than other protocols under varying vehicle densities and spatial distributions. Future development of the proposed QL-based μRr+νRd-MAC protocol in VANETs aims to improve vehicle communication efficiency and system performance. In addition, the security of smart transportation information transmission can be enhanced by incorporating security mechanisms to reduce the risk of attacks and interference. In the future, this technology can be applied to various areas, such as traffic optimization, environmental protection, water resource management, smart grids, and medical care in smart cities.

References

[1] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. D. Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5G: a tutorial overview of standards, trials, challenges, deployment and practice", *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201–1220, 2017.

[2] M. M. Elsayeda, K. M. Hosnya, M. M. Foudab, and M. M. Khashabaa, "Vehicles communications handover in 5G: a survey", *ICT Express* , available online https://doi.org/10.1016/j.icte.2022.01.005, 2022.

[3] M. M. Saad, M. T. R. Khan, S. H. A. Shah, and D. Kim, "Advancements in vehicular communication technologies: C-V2X and NR-V2X comparison", *IEEE Commun. Mag.* , vol. 59, no 8, pp. 107–113, 2021.

[4] G. Abbasab, S. Ullahab, M. Waqas, Z. H. Abbas, and M. Bilal, "A position-based reliable emergency message routing scheme for road safety in VANETs", *Comput. Netw.* , vol. 213, pp. 1–14, 2022.

[5] S. Ullah, G. Abbas, M. Waqas, Z. H. Abbas, S. Tu, and I. A. Hameed, "EEMDS: an effective emergency message broadcasts scheme for urban VANETs", *Sensors* , vol. 21, no. 5, pp. 1–19, 2021.

[6] Y. Wang, J. Shi, and L. Chen, "Performance Analysis of IEEE 802.11p MAC with Considering Capture Effect under Nakagami-m Fading Channel in VANETs", *Entropy* , vol. 25, no. 2, pp. 1–13, 2023.

[7] 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 Trans. Veh. Technol.* , vol. 68, no. 9, pp. 9136–9150, 2019.

[8] M. N. Tahir and M. Katz, "Performance evaluation of IEEE 802.11p, LTE and 5G in connected vehicles for cooperative awareness", *Engineering Reports* , vol. 4, no. 4, pp. 1–13, 2021.

[9] X. Lei and S. H. Rhee, "Performance analysis and enhancement of IEEE 802.11p beaconing", *EURASIP J. Wirel. Commun. Netw.* , vol. 61, pp. 1–10, 2019.

[10] H. Wang, J. Lai, X. Zhang, Y. Zhou, S. Li, and J. Hu, "Make space to change lane: a cooperative adaptive cruise control lane change controller", *Transp. Res. Part C Emerg. Technol.* , vol. 13, no. 103847, 2022.

[11] L. An, X. Yang, and J. Hu, "Modeling system dynamics of mixed traffic with partial connected and automated vehicles", *IEEE Trans. Intell. Transp. Syst.* , vol. 23, no. 9, pp. 15755–15764, 2022.

[12] M. Afrashteh and S. Babaie, "A route segmented broadcast protocol based on RFID for emergency message dissemination in vehicular ad-hoc networks", *IEEE Trans. Veh. Technol.* , vol. 69, no. 12, pp. 16017–16026, 2020.

[13] P. Li, Y. Zeng, C. Li, L. Chen, H. Wang, and C. Chen, "A probabilistic broadcasting scheme for emergent message broadcasting in urban internet of vehicles", *IEEE Access*, vol. 9, pp. 13187–13198, 2021.

[14] W. Benrhaiem and A. S. Hafid, "Bayesian networks based reliable broadcast in vehicular networks", *Veh. Commun.*, vol. 21, pp. 1–13, 2020.

[15] M. R. Devi, I. J. S. Jeya, and S. Lokesh, "Adaptive scheduled partitioning technique for reliable emergency message broadcasting in VANET for intelligent transportation systems", *Automatika* , vol. 64, no. 2, pp. 341–354, 2023.

[16] Y. H. Chou, T. H. Chu, S. Y. Kuo, and C. Y. Chen, "An adaptive emergency broadcast strategy for vehicular ad hoc networks", *IEEE Sens. J.* , vol. 8, no. 12, pp. 4814–4821, 2018.

[17] Z. Pei, X. Wang, Z. Lei, H. Zheng, L. Du, and W. Chen, "Joint optimization of multi-hop broadcast protocol and MAC protocol in vehicular ad hoc networks", *Sensors* , vol. 21, no. 18, pp. 1–23, 2021.

[18] K. Sridevi, B. Phijik, K. Helini, G. Rajesh, and A. Rupa, "MANET's node secure mobility predictions using enhanced adaptive learning techniques", in *2023 8th Int. Conf. Commun. Electron. Syst. (ICCES)* , Coimbatore, India, pp. 1–5, 01-03 June 2023.

[19] C. Wei, Y. Wang, X. Wang, and Y. Tang, "QFAGR: A Q-learning-based fast adaptive geographic routing protocol for flying ad hoc networks", in *GLOBECOM 2023 - 2023 IEEE Global Commun. Conf.* , Kuala Lumpur, Malaysia, pp. 1–6, 04-08, December 2023.

[20] J. S. Park and J. H. Park, "Enhanced machine learning algorithms: deep learning, reinforcement learning and Q-learning", *J. Inf. Process. Syst.*, vol. 16, no. 5, pp. 1001–1007, 2020.

[21] S. C. da S. Cruz, M. A. Ouameur, and F. A. P. de Figueiredo, "Reinforcement learning-based wi-fi contention window optimization", *Math. Comput. Sci.*, doi: 10.20944/preprints202211.0011.v1, 2022.

[22] Y. H. Xu, J. W. Xie, Y. G. Zhang, M. Hua, and W. Zhou, "Reinforcement learning (RL)-based energy efficient resource allocation for energy harvesting-powered wireless body area network", *Sensors/*, vol. 20, no. 44, pp. 1–22, 2020.

[23] C. Lu, Z. Wang, W. Ding, G. Li, S. Liu, and L. Cheng, "MARVEL: multi-agent reinforcement learning for VANET delay minimization", *China Commun.* , vol. 18, no. 6, pp. 1–11, 2021.

[24] X. Zeng, M. Yu, and D. Wang, "A new probabilistic multi-hop broadcast

protocol for vehicular networks", *IEEE Trans. Veh. Technol.* , vol. 68, no. 9, pp. 9136–9150, 2019.

[25] C. Paduraru, D. Mankowitz, G. Dulac-Arnold, J. Li, N. Levine, S. Gowal, and T. Hester, "Challenges of real-world reinforcement learning: definitions, benchmarks and analysis", *Machine-mediated learning* , vol. 110, pp. 2419–2468, 2021.

[26] R. Ali, Y. A. Qadri, Y. B. Zikria, T. Umer, B. S. Kim, and S. W. Kim, "Q-learning-enabled channel access in next-generation dense wireless networks for iot-based ehealth systems", *EURASIP J. Wirel. Commun. Netw.* , vol. 178, pp. 1–12, 2019.

[27] S. Dharmaraja, R. Vinayak, and K. S. Trivedi, "Reliability and survivability of vehicular ad hoc networks: an analytical approach", *Reliab. Eng. Syst. Saf.* , vol. 153, pp. 28-38, 2016.

[28] H. Zhang, X. Zhang, and D. K. Sung, "An efficient cooperative transmission based opportunistic broadcast scheme in VANETs", *IEEE Trans. Mob. Comput.*, vol. 22, no. 3, pp. 1327–1342, 2023.

[29] S. Sattar, H. K. Qureshi, M. Saleem, S. Mumtaz, and J. Rodriguez, "Reliability and energy-efficiency analysis of safety message broadcast in VANETs", *Comput. Commun.*, vol. 119, pp. 118-126, 2018.

[30] 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 Veh. Netw. Conf. (VNC)*, Torino, pp. 263–270, 2017.

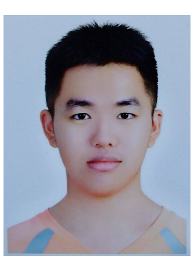
[31] F. Chollet et al., "Keras," https://keras.io, 2015.

[32] Y. Guo, L. Zhu, Y. Li, Y. Liu, J. Zhang, and R. He, "Car-following model based analysis of mixed traffic flow characteristics", in *2021 China Auto. Cong. (CAC)* , Beijing, China, pp. 1-5, 22-24 October 2021.

[33] Z. Sadreddini, Ö. Makul, T. Çavdar, and F. B. Günay, "Performance analysis of licensed shared access based secondary users activity on cognitive radio networks", in *2018 Electr. Electron., Comput. Sci., Biomed. Engineerings' Meeting (EBBT)*, Istanbul, Turkey, pp. 1–4, 2018.

[34] N. T. T. Hang, N. C. Trinh, N. T. Ban, M. Raza, and H. Nguyen, "Delay and reliability analysis of p-persistent carrier sense multiple access for multi-event industrial wireless sensor networks", *IEEE Sens. J.* , vol. 20, no. 20, pp. 12402–12414, 2020.

Chien-Min Wu was born in Yunlin City, Taiwan, in 1966. He received the B.S. degree in automatic control engineering from the Feng Chia University, Taichung, Taiwan, in 1989, the M.S. degree in electrical and information engineering from Yuan Ze University, Chung-Li, Taiwan, in 1994 and the Ph.D. degree with the Department of Electrical Engineering, National Chung Cheng University, Chia-Yi, Taiwan, in 2004. In July 1994, he joined the Technical Development Department, Philips Ltd. Co., where he was a Member of the Technical Staff. He is currently a professor of the Department of the Computer Science and Information Engineering, Nanhua University. His current research focuses on reinforcement learning and deep reinforcement learning applied to *IoT* and *vehicular systems*, *cognitive radio* networks, *ad hoc* networks, and MAC protocol design.



Cheng-Tai Tsai was born in Pingtung City, Taiwan, in 1999. He received the B.S. degree in engineering from Nanhua University, Chiayi, Taiwan. He is currently working toward the M.S. degree in communications engineering at National Chung Cheng University, Chiayi, Taiwan. His research interests include *reinforcement learning*, *cognitive radio* networks, *ad hoc* networks, and MAC protocol design.

Cheng-Chun Hou was born in Taoyuan City, Taiwan, in 1995. He received the B.S. degree in computer science and information engineering from Nanhua University, Chiayi, Taiwan. He is currently working toward the M.S. degree in communications engineering at National Chung Cheng University, Chiayi, Taiwan. His research interests include *reinforcement learning*, *cognitive radio* networks, *ad hoc* networks, and MAC protocol design.

Jun-Jie Yang was born in Changhua City, Taiwan, in 2001. He received the B.S. degree in engineering from Nanhua University, Chiayi, Taiwan. He is currently working toward the M.S. degree in communications engineering at National Chung Cheng University, Chiayi, Taiwan. His research interests include reinforcement learning, *cognitive radio* networks, *ad hoc* networks, and MAC protocol design.

Gong-De Lin was born in Chiayi City, Taiwan, in 2001. He received the B.S. degree in engineering from Nanhua University, Chiayi, Taiwan. He is currently working toward the M.S. degree in communications engineering at National Chung Cheng University, Chiayi, Taiwan. His research interests include reinforcement learning, *cognitive radio* networks, *ad hoc* networks, and MAC protocol design.

Mi-Yu Kuang was born in Taichung City, Taiwan, in 2003. He is currently working toward the B.S. degree in engineering from Nanhua University, Chiayi, Taiwan. His research interests include deep reinforcement learning, reinforcement learning, *cognitive radio* networks, *ad hoc* networks, and MAC protocol design.