Adaptive Learning-Driven Contention Window Selection for Efficient Channel Access in Vehicular Networks

Lopamudra Hota (Member IEEE), Arun Kumar (Senior Member IEEE, Peter Han Joo Chong (Senior Member IEEE)

***Abstract*—In Vehicular Ad-hoc Networks (VANETs) and major transportation systems, efficient communication protocol is vital for timely data transmission to vehicles. The dense vehicular network poses challenges to efficient channel-sharing. For the proper utilization of the available bandwidth, optimization of channel mechanisms is crucial. The proposed approach enables vehicles to dynamically tune their Contention Windows (CWs) using locally observable MAC-layer information, with the objective of jointly maximizing throughput, minimizing delay, and maintaining fair channel access. Comprehensive simulations and analysis show notable improvement in the overall network efficiency in terms of throughput, collision, and delay. The adaptiveness of the proposed algorithm guarantees flexibility to changing traffic conditions and is well-suited to the evolving Intelligent Transportation Systems (ITS). With an emphasis on high throughput, low latency, and fair channel allocation, the proposed model contributes to the advanced communication protocols for VANETs. The proposed model also highlights the significance of intelligent adaptive techniques in obtaining enhanced network performance.**

***Index Terms*—VANET, MAC, DRL, Multi-Agent, Contention Window, Adaptive, Channel, Actor-Critic.**

I. Introduction

In the last two decades, vehicular communication has emerged as the safer and most sophisticated mode of communication starting from manual vehicles to network-assisted fully automated driving [1]. The new paradigm emphasizes vehicles communicating via wireless networks so that they can perceive their surroundings completely. Wireless vehicular networks support numerous transmitting vehicle stations within the range of one another because DSRC specifies a 1-hop range of up to 1 km Line-of-Sight (LoS). IEEE 802.11p is a member of the IEEE 802.11 family of protocols, which were initially created to be used in Wireless Local Area Networks (WLANs). The Internet of Vehicles (IoVs) is proposing an ever-increasing number of promising applications, novel protocols are required to meet challenging demands not addressed by the conventional standard. The IEEE 802.11p stack manages 50–100 concurrent connected stations within a communication range, hence the DSRC physical (PHY) and Medium Access-Control (MAC) must be scalable. Safety Applications rely on timely message dissemination thus, seek to identify probable collisions on the road on time and warn approaching vehicles. The IEEE 802.11p radio access technology in the 5.9 GHz frequency band, specifies the MAC and PHY layers of Wireless Access in Vehicular Environments (WAVE) [2], used for communication. The Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocols serves as the foundation for the IEEE 802.11p MAC layer, which also contains the Outside of the Context of a Basic Service Set (OCB) operation mode that was recently described in [3]. The IEEE 802.11p PHY layer employs a channel bandwidth of 10 MHz proven to reduce the negative effects of multi-path delay and the Doppler effect [4].

Neighbouring vehicles must constantly exchange status information on a shared control channel (CCH) for providing cooperative vehicular safety. Each vehicle regularly broadcasts beacons to other vehicles to know its presence within the transmission range [5]. In the US and Europe, the beacons are officially known as Basic Safety Messages (BSM) [6] and Cooperative Awareness Messages (CAM) respectively. These messages include details regarding the transmitting vehicle, like position, movement, speed, and acceleration. The receiving vehicle can build a Local Dynamic Map (LDM) using the status data from its vicinity through a beaconing mechanism [7]. Cooperative safety applications use this status information to identify and prevent potential collisions in real-time, like, the crash risk can be evaluated by examining the vehicle's movement status [8]. The successful operation of cooperative safety applications depends on selecting the optimum beacon transmission rate for each vehicle in various scenarios.

When safety messages generated from a vehicular node collide and are dropped, lose connectivity to the wireless vehicular network, which poses challenges for safety applications, as the message will never be sent. Furthermore, ITS applications have very high time-constraint requirements which is also a challenge due to the dynamic nature of the vehicular networks. The CW parameter in IEEE 802.11p networks has a significant impact on performance, especially in networks with high traffic loads. Proper adjustment of the CW parameter provides notable performance improvements for both the network as whole and individual stations. The optimal value for the CW parameter depends on several factors, in-

Department of Computer Science and Engineering,
Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India
Email: lhota@bitmesra.ac.in,
Department of Computer Science and Engineering,
National Institute of Technology, Rourkela, Odisha, India
Email: kumararun@nitrkl.ac.in,
School of Engineering, Computer and Mathematical Sciences,
Auckland University of Technology, New Zealand
Email: peter.chong@aut.ac.nz

0000–0000/00$00.00 © 2021 IEEE

cluding network density (number of active transmitters), traffic volume, required throughput, latency tolerance of applications, and prioritization of specific data exchanges. Choosing the appropriate CW size involves a trade-off [9]. It should be large enough to accommodate network traffic without collisions, minimizing the likelihood of transmission failures due to collisions. However, it should not be unnecessarily large, as this would increase packet transmission latency and cause stations to forfeit transmission opportunities by waiting too long for the channel to become available. In some scenarios, it may be beneficial to set a lower CW for a specific data class to allocate a larger portion of the available bandwidth for prioritization purposes. This allows certain exchanges to have higher priority and better throughput [10].

The use of ML techniques including supervised learning, unsupervised learning, and Reinforcement Learning (RL) has surged to optimise networking protocols to enhance network performances. RL is widely recognized and implemented in areas with multidimensional objective optimization problems [11]. The RL interacts with the real-time environment based on past experience. There are many variables, such as the agents (vehicles) that act (optimising parameters, CW), at the state (channel status) to obtain the reward (throughput) as the network performance in an environment (VANET) [12].

The optimal selection of the CW parameter involves careful consideration of network conditions (traffic density), traffic characteristics, throughput requirements, latency tolerance, and prioritization needs to strike the right balance between maximizing network capacity and minimizing delays. This paper aims to investigate a self-learning channel access mechanism for data dissemination over DSRC networks for safety applications. Multi-Agent Reinforcement Learning (MARL) is implemented for CW adaptation, with the functionality to enable direct wireless interconnection of a large number of mobile and stationary units over IEEE 802.11p interfaces. By using this self-learning-based method, the network performance is enhanced in terms of latency and throughput.

A. Motivation

MAC layer plays a vital role for message dissemination by governing for contention of vehicular nodes to access the shared wireless medium. The IEEE 802.11p DSRC mode of communication is based on CSMA/CA, which is a decentralized approach for channel access, that sense the channel before transmission to avoid collision. Broadcast in VANET is a primary challenge in dense traffic scenario, and for safety message transmission due to the time-constraint. The IEEE 802.11p standard has specified the value of CW randomly based on the channel access and the Listen Before Talk (LBT) mechanism to handle collisions. This mechanism fails under rapidly varying vehicular densities and traffic conditions. The data transmission is via the V2V or V2I link; thus, an efficient MAC layer protocol that adapts to the vehicular density is necessary. The core objective of this research is therefore to design an intelligent CW adaptation mechanism to improve channel utilization, reducing access delay, and providing efficient safety message dissemination in highly dynamic vehicular environments.

B. Contribution

This research takes into account completely cooperative and partially observable multi-agent domains in order to improve network performance. The work is an improvement to our previous work based on an adaptive CW mechanism using the Actor-Critic Network [9]. All vehicular stations aim to boost the system throughput by combining rewards through fully cooperative learning. The set-up framing issue optimization is needed to depict the vehicle's decision-making process in a partially observable environment. Decentralized POMDP (Dec-POMDP) is used, and the Deep R-Learning based Multi-Agent (MA-DRL) algorithm is suggested, to examine the packet transmission efficiency by CW optimization in a multi-agent scenario. The primary contributions of this paper include:

1) The contention window adaptation is formulated as a decentralized multi-agent reinforcement learning problem, enabling each vehicle to autonomously adjust its CW based solely on local MAC-layer observations.
2) Based on decentralized POMDP, fully collaborative vehicular stations are taken into consideration and the efficiency of transmission of packets using the optimized CW designed for the multi-agent scenario.
3) The network performance is analysed for realistic VANET environment with the proposed MA-DRL algorithm.

II. Related Work

Researchers have extensively studied VANET's MAC protocol in recent years. Researchers are primarily focused on improving protocols, evaluating performance, design of channel estimation and modelling schemes [13].

Klapez et al. [14] researched the modelling and optimisation of CSMA/CA in vehicular networks, focusing on the 802.11p standard concerning contention-based safety messages in V2X environment. Theoretically, a closed-form analytic model with a maximum capacity upper bound was derived. The effects of vehicle traffic on available capacity were assessed.

According to Bianchi's theory [15], the CW of a backoff timer in the CSMA scheme is used to calculate the transmission probability of nodes. It is possible to get the theoretical formula for the average throughput based on the two-dimensional (2D) Markov model. It turns out that a high likelihood of collisions will cause the average throughput in dynamic vehicular networks to decrease as traffic density rises. Several modified CSMA algorithms with adaptive CW have been proposed. As an illustration, Ref. [16] suggested a self-adaptive CW update factor to accommodate various traffic densities. By adopting a threshold for traffic density, DASC [17] makes CW size adaptive to whether packet transmission success or collision occurs. Additionally, CSA [18] suggested a bidirectional backoff adaptation using an expanded 2D Markov model. The transmit probability closed-form formulas were developed in the CSA with a higher average throughput.

Additionally, RL techniques have the ability to create intelligent MAC. In V2V communication networks, the MAC protocol has added RL for combined mode selection and power adaption [19]. Underwater sound sensor networks use

a low-complexity RL-based MAC for high channel counts [20]. To reduce delay sensitivity with dense traffic, a self-learning contention technique to optimise CW in vehicular networks was presented in QLMAC [21]. The many link conditions between vehicle nodes in the dynamic environment were thus disregarded because QLMAC employs continual positive rewards for all successful actions undertaken in Q-learning. Theoretically, agents uses Q-learning to evaluate the score on how effective an optimal policy will be in the current environment because it is a model-free policy-based value function [22]. Without prior knowledge of the environment, the agent can decide how to update itself from the current state to a new state through interactive learning from the environment at a maximum Q value. Although Q-learning's mathematical foundation is straightforward, the Q function does in fact depend on the policy's reward model.

Taherkhani et al. [23] proposed a centralized and localized data congestion control strategy to control data congestion using RSUs at intersections. This strategy consists of three units for detecting congestion, clustering messages, and controlling data congestion. The messages are clustered using a machine learning algorithm in each RSU independently. Transmission rate, CW size, and AIFS are the necessary parameters for congestion occurrence. Instead of assigning an optimal value for all the messages, the authors have assigned an optimal value for each cluster, which will improve the processes.

Sepulcre et al. [24] proposed an analytical model for IEEE 802.11P to analyse the performance of V2V communication based on propagation, hidden terminals, interference, and distance between transmitter and receiver. The author's work is one of the first kind to estimate the Channel Busy Ratio (CBR) even in dense traffic scenarios. Wang et al. [25] proposed a Multi-Agent Reinforcement Learning (MARL) based on dynamic channel assignment that focuses on efficient channel access mechanism and adaptation of back-off for real-time scenarios. Here, each vehicular node adapts to the decision for channel selection and back-off for dynamic applications and channel conditions. The result presents a minimised delay and maximized packet delivery ratio.

Khizra et al. [26] proposed an algorithm for setting an optimal contention window called Deep Reinforcement Learning (DRL)-based Contention Window Optimization under different network conditions. With DRL, a Deep Neural Network (DNN) estimates the state action-value function. The NN of Deep Q-Networks is trained to reduce the prediction error caused by the loss function. To further optimize the DQN for the WiFi network, the authors use DDQN. Initially, the WiFi is controlled by the 802.11 standards. The collision probability is observed and evaluated. In the next step, DDQN is trained by maximizing the reward. In the last step, DRL is deployed in the network. Now CW is updated by the optimized procedure.

Andreas et al. [27] proposed an RL mechanism to design a contention-based MAC protocol that selects the most suitable CW parameter based on gained experience from its interactions with the environment. The MAC protocol uses a Q-Learning that adjusts the contention window size based on binary feedback from probabilistic rebroadcasts to avoid packet collisions and increase the network throughput. The authors achieved an increased throughput than the standard IEEE 802.11p.

The literature study highlights the importance to design an adaptive CW scheme by integrating learning algorithms to the MAC for reliable broadcast of safety messages in VANETs. There is still scope for improved throughput, packet delivery handling the time-constraints and fairness. This work presents a multi-agent deep reinforcement learning (MA-DRL)–based MAC-layer solution that dynamically adapts the CW parameter in an IEEE 802.11p–compatible manner. By enabling each vehicle to autonomously learn the optimal channel access behavior from local observation, the proposal target to enhance the network performance.

III. The IEEE 802.11p Mechanism

IEEE 802.11p is a standard that defines the Medium Access Control (MAC) layer and physical layer specifications for wireless access in vehicular environments (WAVE). The main purpose of IEEE 802.11p is to enable communication between vehicles (V2V) and between vehicles and infrastructure (V2I), providing the foundation for various ITS applications and services [28]. The standard emerged as an amendment to the original IEEE 802.11 standard to support low-latency and high-reliability communication between vehicles, allowing them to exchange safety-related information and enable cooperative driving applications.

The IEEE 802.11p standard introduces several MAC layer enhancements to cater to the requirements of vehicular communication. One crucial feature is the addition of the channel access method known as "Enhanced Distributed Channel Access" (EDCA). EDCA prioritizes access categories, allowing for different levels of Quality of Service (QoS) for different types of data traffic (e.g., safety messages, non-safety messages) [29]. The Physical Layer (PHY) specifications of IEEE 802.11p use the 5.9 GHz frequency band, specifically allocated for ITS [30]. This frequency band is chosen because it provides good propagation characteristics and is less susceptible to interference compared to other Wi-Fi frequency bands.

Binary Exponential Backoff (BEB) is a key mechanism used in the IEEE 802.11p Medium Access Control (MAC) layer to manage contention and access the wireless medium in vehicular communication environments. IEEE 802.11p uses a contention-based MAC protocol, where multiple nodes (vehicles) compete for access to the shared wireless medium. In a contention-based MAC, nodes do not have a predefined time slot for transmission but instead listen to the medium and transmit whenever the channel is sensed to be idle. If two or more nodes attempt to transmit simultaneously, a collision occurs, leading to data loss and reduced network efficiency. When a node has data to transmit and senses the channel is idle (i.e., no other transmission is taking place), it initiates the channel access process [13]. This involves selecting a random backoff time before the actual transmission to avoid collisions with other nodes that might also be ready to transmit. The BEB algorithm is used to determine the random backoff time for a node after it detects an idle channel. When a node's initial transmission attempt fails (e.g., due to a collision with another node's transmission), it invokes the BEB algorithm. After a

transmission failure (collision), the node sets a retransmission counter to a predefined value (usually 0). The node then selects a random backoff time from a CW. The CW size is typically initialized to a small value. If the channel remains idle for the entire backoff time, the node starts its transmission. If, however, the channel becomes busy again during the backoff time, the node stops the timer and freezes the countdown. When the channel becomes idle again, the node resumes the backoff countdown from where it stopped. If the node experiences another collision (channel becomes busy while in backoff), it doubles the CW size and selects a new random backoff time from the increased CW range. The process of doubling the CW size and selecting a new backoff time continues with each subsequent collision until the maximum CW size is reached [31].

While BEB is a widely used mechanism for contention resolution in IEEE 802.11p MAC, it does have some disadvantages, especially in highly congested and dynamic vehicular communication environments [32]. Recent advancements have been made to address some of these limitations and improve the overall efficiency of channel access in vehicular networks. In highly dense vehicular networks, collisions still occurs frequently even with BEB's random backoff mechanism. When multiple nodes experience collisions and double their contention window, the contention becomes more prolonged, leading to inefficient channel access and increased latency. BEB leads to the phenomenon known as "backoff collisions." When nodes choose their backoff timers independently and randomly, there is a possibility that two or more nodes may select the same backoff time, causing them to collide again when they attempt to transmit simultaneously after the backoff period. BEB treats all traffic types equally, without considering the priority of data packets. In vehicular communication, some applications, such as safety-critical messages, require higher priority and lower latency than non-safety messages. BEB does not provide mechanisms to differentiate QoS for different types of traffic [33].

Contrary to these disadvantages, EDCA is an extension of the IEEE 802.11e standard that introduces a priority-based channel access method for wireless networks. It uses different contention window values for different access categories, allowing for QoS differentiation. EDCA enhances the fairness and efficiency of channel access in vehicular networks by providing better QoS support. Some recent approaches combine the benefits of different MAC protocols, such as combining contention-based mechanisms like BEB with reservation-based or time-division multiplexing approaches. These hybrid MAC schemes seek to balance efficiency, scalability, and QoS support, leveraging each approach's strengths to mitigate their weaknesses [34]. Researchers have proposed intelligent backoff algorithms that adapt the contention window size based on network conditions and traffic load. These algorithms dynamically adjust the CW size to improve channel access efficiency and reduce collision probability [35]. Although not specific to IEEE 802.11p, the Wi-Fi 6 standard (IEEE 802.11ax) introduces various improvements to MAC and physical layers that indirectly benefit vehicular communication. The Wi-Fi 6 features such as OFDMA (Orthogonal Frequency Division Multiple Access) and MU-MIMO (Multi-User Multiple Input Multiple Output) enhance spectral efficiency and support multiple users simultaneously, which is advantageous in dense vehicular scenarios [36].

IV. Proposed Work

This section presents the proposed MADDPG framework for CW tuning. Table I presents the notation table for symbols used in the proposed mechanism.

red

A. Problem Formulation

The traditional Markovian Decision Process (MDP) consists of only a single agent, in contrast here, in a multi-agent scenario the actions taken by any agent will impact the environment state and rewards of other agents.

The vehicular stations are made capable of learning the environment from observation, by the proposed MA-DRL algorithm that is based on POMDP used for the multi-agent environments. The mathematical formulation stated as $\mathcal{P} = (N, \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \Omega, \mathcal{O})$

$i \epsilon N$ is the finite set of agents (Vehicular station) at time duration t. $s_t \in \mathcal{S}$ represents the current state of vehicular stations i.e. agents connected to the wireless network. $a_t \in \mathcal{A}$ is the cooperative action taken by the agents to set the *CW* value from the action space 0 to 7 taken in our case. A CW_{th} is set whose value varies from $2^7 \times (1 + a)$, where $a \in 0...7$. The state transition probability for the state of the environment changes from s_t to s_{t+1} is denoted by $p\left(s_{t+1} \mid s_t, a_t\right) \in \mathcal{T}$. The finite set of collaborative observations here depends on the collision probability $P_{col} \in \Omega$. The set of current and historical observations is denoted by $P_{col} \in \mathcal{O}$ known as observation probability. $r \in R$ states the reward function obtained after each action to optimize the value of *CW*. The main goal here is to maximize the network throughput, the normalized throughput is considered as the reward that aids in finding a deterministic policy π. The vehicular station learns this policy to maximize the cumulative reward in the future by a discount factor $\gamma \in [1, 0]$. Choosing an optimal action value for a state action pair based on the policy basically is the problem, which is solved by the proposed Deep Deterministic Policy Gradient model for multi-agent (MA-DRL algorithm).

Here *N* represents the number of vehicular nodes, S represents the VANET environment state space, $\mathcal{A}_n$ represents the action space set 0...7, $f : \mathcal{S} \times \mathcal{A}_1 \times \cdots \times \mathcal{A}_N \times \mathcal{S} \rightarrow [0, 1]$ represents teh state transition probability from state s_t to s_{t+1} after an action $\mathrm{A}_n, \mathrm{R}_t$ represents the reward obtained after a vehicular node takes action A_n at state S.

B. Proposed Approach

The proposed model addresses the adaptive contention window problem in VANETs as a multi-agent reinforcement learning task and adopt the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) framework under a centralized training and decentralized execution (CTDE) paradigm. Each vehicle acts as an autonomous agent that observes local features such as queue length, channel busy ratio, collision history, and current contention window size. Based on these observations, the agent selects its contention window within bounded limits, aiming to balance throughput, delay, and collision avoidance.

This article has been accepted for publication in IEEE Internet of Things Journal. 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/JIOT.2026.3677251

TABLE I: Summary of List of Symbols

Notation	Description	Notation	Description
$\mathcal{N} = \{1, \ldots, N\}$	Set of vehicular agents	N	Total number of vehicles
t	Discrete time index	M	Number of training episodes
T	Steps per episode	B	Mini-batch size
s_t	Global environment state at time t	$\mathcal{S}$	Global state space
o_i^t	Local observation of agent i	$\mathcal{O}_i$	Observation space of agent i
$\mathbf{o}_t$	Joint observation vector	a_i^t	Action of agent i
$\mathbf{a}_t$	Joint action vector	$\mathcal{A}_i$	Action space of agent i
$\mathcal{A}$	Joint action space	CW_i^t	CW selected by agent i
$CW_{\min}, CW_{\max}$	CW bounds	$CW_{\min}(N_t)$	Density-aware minimum CW
$CW_{\max}(N_t)$	Density-aware maximum CW	N_t	Active vehicle count
$\hat{N}_i^t$	Neighbor estimate of agent i	q_i^t	Queue length
ρ_i^t	Channel Busy Ratio (CBR)	c_i^t	Collision history
τ_i^t	Average transmission delay	b_i^t	Backoff counter
$p(t)$	Traffic load at time t	o_i^t	Local MAC observation vector
$P(s_{t+1}\|s_t, \mathbf{a}_t)$	State transition probability	r_i^t	Reward of agent i
$\mathcal{R}_i$	Reward function	α, β, δ	Reward weighting factors
γ	Discount factor	J_i	Expected cumulative return of agent i
$J(\theta)$	Global cooperative objective	μ_{θ_i}	Actor policy of agent i
θ_i	Actor network parameters	Q_{ϕ_i}	Centralized critic
ϕ_i	Critic parameters	$Q_{\phi_i}(s, \mathbf{a})$	Joint action–value function
y_i^t	Temporal-difference target	$L(\phi_i)$	Critic loss function
$\nabla_{\theta_i} J$	Policy gradient	τ	Soft target update rate
D	Replay buffer	$\mu'_{\theta_i}, Q'_{\phi_i}$	Target networks

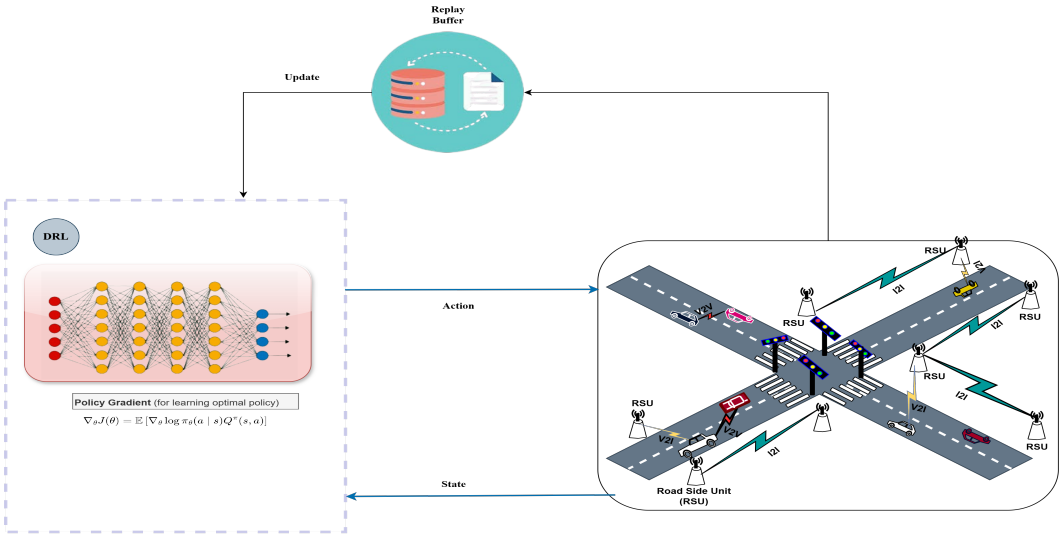

Fig. 1: Architectural Model of the Proposed Framework.

The environment is modeled as a Dec-POMDP where agents jointly interact through the shared wireless channel. To capture cooperative behavior, we design a reward function that encourages successful transmissions while penalizing collisions and excessive delay. The actor networks of individual agents learn deterministic policies, while centralized critics leverage the global state and joint actions to provide accurate value estimates during training.

The problem formulation is demonstrated as,

$$\langle \mathcal{N}, \mathcal{S}, \{\mathcal{O}_i\}, \{\mathcal{A}_i\}, P, R, \gamma \rangle, \tag{1}$$

where $\mathcal{N}$ are the vehicles, $\mathcal{S}$ the global network state, $\mathcal{O}_i$ the local observation space, $\mathcal{A}_i$ its action space, P the state transition function The Global state representation, includes a centralized critic,

$$s_t = \big[\{\ell_i(t)\}_{i \in \mathcal{N}},\ \mathrm{CBR}(t),\ C(t),\ \rho(t)\big], \tag{2}$$

where $\ell_i(t)$ denotes the queue length of vehicle i, $\mathrm{CBR}(t)$ is the channel busy ratio, $C(t)$ represents collision at time t, and $\rho(t)$ presents the traffic load. Each vehicle i operates based solely on locally observable MAC-layer information, defined as,

$$o_i(t) = \big[\ell_i(t),\ \mathrm{CBR}_i(t),\ CW_i(t),\ b_i(t),\ c_i(t)\big], \tag{3}$$

$CW_i(t)$ is the CW value at time t, $b_i(t)$ is the back-off counter, and $c_i(t)$ is the current collision history. The Effects such as hidden terminals, contention, and channel states are captured

via decentralized observations. Each agent select a CW value based on the policy based on local observation,

$$a_i(t) = \mu_{\theta_i}\big(o_i(t)\big), \tag{4}$$

where μ_{θ_i} denotes the deterministic actor policy parameterized by θ_i.

The critic is updated using temporal difference learning with centralized information, whereas each actor is updated through policy gradients derived from its local observation. A soft target update mechanism is employed to stabilize learning. After training, the learned policies are executed in a decentralized manner, allowing each vehicle to independently adapt its contention window in real time without requiring global coordination.

The Figure 1 depicts the architectural model for the proposed algorithm.

C. System Model

We consider a set of vehicles $\mathcal{N} = \{1, \ldots, N\}$ sharing a DSRC channel under CSMA/CA. The global system state at time t is

$$s_t = \big(q_t^1, \ldots, q_t^N, \rho_t^1, \ldots, \rho_t^N, c_t^1, \ldots, c_t^N, \tau_t^1, \ldots, \tau_t^N, CW_t^1, \ldots, CW_t^N\big), \tag{5}$$

where q_t^i is the queue length, ρ_t^i the channel busy ratio, c_t^i the collision count, τ_t^i the average delay, and CW_t^i the current contention window of agent i.

Each agent i observes only local information:

$$o_t^i = \big[q_t^i, \rho_t^i, c_t^i, \tau_t^i, CW_t^i\big]. \tag{6}$$

The action of agent i is the selection of its contention window:

$$a_t^i = \mu_{\theta^i}(o_t^i) \in [CW_{\min}, CW_{\max}], \tag{7}$$

where μ_{θ^i} is the actor network for agent i. The next state evolves as

$$s_{t+1} \sim P\big(s_{t+1} \mid s_t, \mathbf{a}_t\big), \quad \mathbf{a}_t = (a_t^1, \ldots, a_t^N). \tag{8}$$

The reward of agent i at time t is defined as

$$r_t^i = \alpha \cdot \text{SuccTx}_t^i - \beta \cdot \text{Collisions}_t^i - \eta \cdot \tau_t^i, \tag{9}$$

where $\alpha, \beta, \eta > 0$ are weighting parameters. The coefficient α weights the successful transmissions and serves as a utility scaling factor. β and η, are considered as Lagrange multipliers enforcing collision and delay constraints. In our implementation, α is selected based on a scale-normalization principle to ensure balanced reward magnitudes across components:

$$\alpha : \beta : \eta \approx \frac{1}{\mathbb{E}[\text{SuccTx}]} : \frac{1}{\mathbb{E}[\text{Collisions}]} : \frac{1}{\mathbb{E}[\tau]}.$$

The expectations are estimated from various runs with MAC parameters. Sensitivity analysis over $\alpha \in [0.2, 2.0]$ indicates that moderate values yield a Pareto-efficient trade-off between throughput maximization with delay control [37], [38]. The formulation of β and η is presented in the Appendix VII. The objective of each agent is to maximize the discounted cumulative reward:

$$J(\theta^i) = \mathbb{E}_\pi\left[\sum_{t=0}^{\infty} \gamma^t r_t^i\right]. \tag{10}$$

D. Adaptive Contention Window

The VANET contention window adaptation is modelled as a multi-agent Markov game $\mathcal{G} = \langle \mathcal{N}, \mathcal{S}, \{\mathcal{A}_i\}, P, \{r_i\}, \gamma \rangle$:

- $\mathcal{N} = \{1, 2, \ldots, N\}$ is the set of vehicles (agents).
- $\mathcal{S}$ is the global state space representing network conditions (e.g., channel occupancy, number of vehicles, packet collisions).
- $\mathcal{A}_i$ is the action space of agent i, where the action corresponds to the choice of contention window (CW).
- $P : \mathcal{S} \times \mathcal{A}_1 \times \cdots \times \mathcal{A}_N \to \mathcal{S}$ is the state transition probability.
- $r_i : \mathcal{S} \times \mathcal{A}_1 \times \cdots \times \mathcal{A}_N \to \mathbb{R}$ is the reward function for agent i.
- $\gamma \in [0, 1)$ is the discount factor.

Each agent i selects a contention window size CW_i^t at time t. We enforce bounded adaptation based on the number of vehicles N_t:

$$CW_i^t \in \big[CW_{\min}(N_t),\, CW_{\max}(N_t)\big], \tag{11}$$

where

$$CW_{\min}(N_t) = CW_{\min} \cdot f(N_t), \quad CW_{\max}(N_t) = CW_{\max} \cdot f(N_t),$$

and $f(N_t)$ is an increasing function of N_t (e.g., linear or logarithmic scaling), ensuring scalability with traffic density [39].

Each agent maintains a deterministic policy $\mu_{\theta_i} : \mathcal{S} \to \mathcal{A}_i$ parameterized by θ_i. The centralized critic for agent i is defined as:

$$Q_i^\mu(s, a_1, \ldots, a_N) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_i^t \,\middle|\, s^0 = s, a_j^t = \mu_{\theta_j}(o_j^t)\, \forall j \in \mathcal{N}\right]. \tag{12}$$

The actor update is performed by the policy gradient:

$$\nabla_{\theta_i} J(\theta_i) = \mathbb{E}_{s \sim \mathcal{D}}\left[\nabla_{\theta_i} \mu_{\theta_i}(o_i) \nabla_{a_i} Q_i^\mu(s, a_1, \ldots, a_N)\Big|_{a_i = \mu_{\theta_i}(o_i)}\right]. \tag{13}$$

Reward Function

The reward is designed to balance throughput, fairness, and delay:

$$r_i^t = \alpha \cdot \text{Throughput}_i^t - \beta \cdot \text{Collision}_i^t - \delta \cdot \text{Delay}_i^t,$$

where α, β, δ are weighting factors.

E. MADDPG Framework

Each agent has an actor μ_{θ^i} and a centralized critic $Q_{\phi^i}(s, a^1, \ldots, a^N)$.

a) Critic update

The target for agent i is

$$y_t^i = r_t^i + \gamma\, Q_{\phi^{i-}}\Big(s_{t+1}, \mu_{\theta^{1-}}(o_{t+1}^1), \ldots, \mu_{\theta^{N-}}(o_{t+1}^N)\Big), \tag{14}$$

and the critic loss is

$$\mathcal{L}(\phi^i) = \mathbb{E}_{(s_t, \mathbf{a}_t, r_t, s_{t+1})}\Big[\big(Q_{\phi^i}(s_t, \mathbf{a}_t) - y_t^i\big)^2\Big]. \tag{15}$$

b) Actor update.

The deterministic policy gradient for agent i is,

$$\nabla_{\theta^i} J(\theta^i) = \mathbb{E}_{s_t \sim \mathcal{D}} \Big[\nabla_{\theta^i} \mu_{\theta^i}(o_t^i)\, \nabla_{a^i} Q_{\phi^i}\big(s_t, a_t^1, \ldots, a_t^N\big) \Big|_{a^i = \mu_{\theta^i}(o_t^i)} \Big] \tag{16}$$

F. Proposed Algorithm

Algorithm 1: MADDPG for Adaptive CW in VANETs

Require: Number of episodes M, steps per episode T, replay capacity, batch size B, discount γ, soft update τ

Initialize actor networks μ_{θ^i} and critic networks Q_{ϕ^i} for all agents i

Initialize target networks $\theta^{i-} \leftarrow \theta^i$, $\phi^{i-} \leftarrow \phi^i$

Initialize replay buffer $\mathcal{D}$

for episode $= 1$ to M **do**

  Reset environment; obtain initial global state s_0 and local observations o_0^i

  for t = 0 to $T-1$ **do**

    for each agent $i = 1, \ldots, N$ **do**

      Estimate number of active vehicles:

$$\widehat{N}_t^i = 1 + \#\{\text{unique beacons heard in } [t-T, t]\}$$

      Compute analytic target contention window:

$$CW^\star(\widehat{N}_t^i) = \mathrm{clip}\big(2\widehat{N}_t^i - 1,\ CW_{\min}^{\text{proto}},\ CW_{\max}^{\text{proto}}\big)$$

      Define adaptive bounds:

$$CW_{\min}(\widehat{N}_t^i) = \max\{CW_{\min}^{\text{proto}}, \beta_1\, CW^\star(\widehat{N}_t^i)\}$$

$$CW_{\max}(\widehat{N}_t^i) = \min\{CW_{\max}^{\text{proto}}, \beta_2\, CW^\star(\widehat{N}_t^i)\}$$

      Actor generates continuous action:

$$\tilde{a}_t^i = \mu_{\theta^i}(o_t^i) + \mathcal{N}_t^i$$

      Map action to actual contention window:

$$CW_t^i = \left\lfloor \frac{\tilde{a}_t^i + 1}{2} \big(CW_{\max}(\widehat{N}_t^i) - CW_{\min}(\widehat{N}_t^i)\big) + CW_{\min}(\widehat{N}_t^i) \right\rfloor$$

    end for

    Environment executes transmissions under $\mathrm{CW}_t^{1\cdots N}$, returns next global state s_{t+1}, next observations o_{t+1}^i, and rewards r_t^i

    Store transition $(s_t, \mathbf{o}_t, \mathbf{a}_t, \mathbf{r}_t, s_{t+1}, \mathbf{o}_{t+1})$ into $\mathcal{D}$

    if $|\mathcal{D}| \geq B$ **then**

      Sample minibatch of size B from $\mathcal{D}$

      for each agent i **do**

        Compute target actions for next state using target actors:

$$a_{t+1}^{j-} = \mu_{\theta^{j-}}(o_{t+1}^j) \quad \forall j \in \mathcal{N}$$

        Compute TD target:

$$y^i = r_t^i + \gamma\, Q_{\phi^{i-}}\big(s_{t+1}, a_{t+1}^{1-}, \ldots, a_{t+1}^{N-}\big)$$

        Update critic by minimizing:

$$\mathcal{L}(\phi^i) = \frac{1}{B} \sum_{\text{batch}} \big(Q_{\phi^i}(s_t, \mathbf{a}_t) - y^i\big)^2$$

        Update actor using deterministic policy gradient:

$$\nabla_{\theta^i} J \approx \frac{1}{B} \sum_{\text{batch}} \nabla_{\theta^i} \mu_{\theta^i}(o_t^i) \nabla_{a^i} Q_{\phi^i}(s_t, a_t^1, \ldots, a_t^N)\big|_{a^i = \mu_{\theta^i}(o_t^i)}$$

      end for

      for each agent i **do**

        {soft-update targets}

        $\phi^{i-} \leftarrow \tau\phi^i + (1-\tau)\phi^{i-}$

        $\theta^{i-} \leftarrow \tau\theta^i + (1-\tau)\theta^{i-}$

      end for

    end if

    $s_t \leftarrow s_{t+1}$, $o_t^i \leftarrow o_{t+1}^i$

  end for

end for

Ensure: Trained decentralized actors $\{\mu_{\theta^i}\}_{i=1}^N$ =0

The proposed Multi-Agent Deep Deterministic Policy Gradient (MADDPG)-based Adaptive Contention Window (CW) framework formulates the VANET channel access problem as a multi-agent Markov game, where each vehicle acts as an agent interacting with a shared wireless medium. Each agent observes a local state vector comprising queue length, collision count, channel busy ratio, experienced delay, and an estimate of the number of active neighboring vehicles. The actor network outputs a continuous action $a_t^i \in [-1, 1]$ representing the candidate CW, which is then mapped into an adaptive bounded range $[CW_{\min}^i(t), CW_{\max}^i(t)]$ using a vehicle-density-dependent scaling function $f(\widehat{N}_t^i)$ to ensure scalability and compliance with IEEE 802.11p constraints.

The centralized critic approximates the joint action-value function $Q_i(s_t, a_t^1, \ldots, a_t^N)$, capturing both cooperative and competitive interactions among agents, and is trained using temporal-difference updates with soft target network stabilization. Actor networks are updated via deterministic policy gradients to maximize expected cumulative discounted reward, defined as a weighted combination of successful transmission throughput, collision penalization, and latency minimization. The framework employs the centralized training and decentralized execution (CTDE) paradigm, allowing actors to operate autonomously during runtime while exploiting global information during training.

By integrating adaptive CW scaling [40], MADDPG effectively enables vehicles to dynamically modulate their backoff windows, reducing collision probability under high-density scenarios and minimizing delay under sparse traffic, thereby improving overall network throughput, fairness, and stability in dynamic VANET environments.

G. Constraints and Cooperative Objective

The contention window is bounded, as if CW too small, it leads to high collision probability and degraded throughput. And if, the CW i too large it leads to long idle waiting with increased delay, channel under-utilization.

$$CW_{t+1}^i = \min\big(\max(a_t^i, CW_{\min}), CW_{\max}\big). \tag{17}$$

Since multiple vehicles share the same channel, they must cooperate to maximize network performance. The MADDPG uses a CTDE paradigm to handle this. A cooperative global objective is considered:

$$J(\theta) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \frac{1}{N} \sum_{i=1}^{N} r_t^i\right]. \tag{18}$$

V. Simulation and Results

This section describes the simulation model and the results obtained by varying CW with simulation time and number of vehicles for the proposed MADDPG model.

A. Simulation Model

The simulation parameters are set per IEEE 802.11p standard stated in Table I. The simulation model comprises network components for communication and a mobility model for traffic generation. NS-3 is an open-source discrete-event network simulator designed to model and simulate complex communication networks. It is widely used by researchers and developers to study and evaluate various networking protocols, technologies, and scenarios. NS-3 is written in C++ and provides a flexible and modular architecture that allows users to create custom network scenarios and analyze network behavior in a controlled environment. NS-3 offers support for

various communication protocols used in VANETs, such as IEEE 802.11p (for Dedicated Short-Range Communications - DSRC), IEEE 1609.4 (for Multi-Channel Operations), and IEEE 1609.11 (for WAVE Short Message Protocol - WSMP). Researchers evaluate the performance of these protocols under diverse traffic conditions and network topologies. NS3-gym module facilitates this communication by integrating the NS-3 simulator with OpenAI Gym for applying RL algorithms in VANET simulations [41].

Simulation of Urban MObility(SUMO) is an open-source traffic simulation tool that is widely used for modeling and simulating vehicular traffic. It is a popular choice for studying urban traffic scenarios, road networks, and various ITS applications to create realistic mobility scenarios and simulate vehicle movements. The default mobility model in SUMO is Krauss mobility mode, it considers the car following model based on safe speed and safe distance between vehicles.

The simulation environment on which the MAC protocol is implemented uses SUMO for traffic generation and is integrated into NS3-gym for network communication. The network interface includes IEEE 802.11p PHY/MAC and the basic safety message exchange in the application layer. The simulation model is depicted in the Figure 2

B. Simulation Parameters

The vehicles have been configured to broadcast packets to emulate the MAC communication standard 802.11p. The operating system used is Ubuntu 20.04 with 16GB of RAM. The simulation was run in NS3-gym. Two scenarios were considered to validate the model with varied vehicular density. The urban scenario with vehicle density 30-100, average speed 20-50km/hr, and the highway scenario 50-200 vehicles, avearge speed 60-120km/hr. More exploration occurs in the training mode, while the DNN parameters are not updated in the evaluation mode. 32 samples are taken for each batch, learning rate (λ) is taken to be 0.0003, discount factor (γ) as 0.99, and 1e-6 decrement for ϵ-greedy. The network structure of the proposed MA-DRL algorithm consists of three DNN hidden layers with 128 fully connected networks; the rectified linear unit (ReLU) is utilised as a layer activation function. The approach makes use of the Adam optimizer algorithm.

TABLE II: Simulation Settings.

Parameter	**Value**
Simulation time	300s
MAC Protocol	802.11p
Channel Frequency	5.89GHz
Channel Bandwidth	10MHz
Data transmission rate	54Mbits/s
Packet Size	256, 512, 1024 bytes
Number of vehicles/RSU	20-100
Backoff slot time	13μs
Mobility Model	Random Way-Point Mobility Model
Learning rate	0.0003
Discount factor	0.99
Epsilon Greedy ϵ	0.1
Exploration ratio	0.01

C. Results and Discussion

The results presented in this section show the intrinsic characteristics of network performance in terms of throughput, end-to-end delay, and fairness for IEEE 802.11p and proposed protocols. Equation 19 is the formula for the computation of network throughput.

$$\text{Throughput} = \frac{\text{Total Packet Transferred}}{\text{Total Simulation Time}}(Mbps) \tag{19}$$

The channel access delay refers to the time interval from when a data frame enters the transmission queue until it receives a successful acknowledgment, indicating that the frame was received. If a frame exceeds the maximum retry limit, it is discarded, and its time delay is not considered when calculating the channel access delay. The channel access delay is measured in milliseconds by Equation 20 shows the formula for the computation of end-to-end delay.

$$EED = \frac{Trans_{\text{D}} + Prop_{\text{D}} + Queu_{\text{D}} + Proc_{\text{D}}}{\text{Total Packets received}}(ms) \tag{20}$$

In a network with more vehicles in close proximity, 802.11p encounters fairness issues due to some vehicles consistently operating with larger CW sizes. In contrast, the proposed AC addresses this fairness concern by enabling each vehicle to autonomously and intelligently adapt to its environment. Therefore, the Fairness Index is better in the case of AC, EAC and DRL compared to traditional IEEE 802.11p with BEB. Equation 21 represents the fairness index formula. The x_i is the throughput of the vehicle *i* and *N* is the total number of vehicles.

$$\text{Fairness Index} = \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N \cdot \sum_{i=1}^{N}(x_i^2)} \tag{21}$$

The vehicular node is capable to learn the CW value, and the current and future state action pair is stored in RSU. The number of transmitting node and receiving nodes is assumed to be known to the RSU. The number of vehicular nodes is taken to be 10 to 100. The convergence is done after 20 rounds of iterations in the learning phase. The results are compared with the standard IEEE 802.11p and our previous work AC and EAC [9].

The first experiment compared the IEEE 802.11p standard and QL approach for the same environment by evaluating the CW value. The CW value is doubled in IEEE 802.11p when the number of stations increases. The CW value turned into the maximum of CW i.e. 1024 in the high-density vehicular scenario. The larger the value of CW the more is the waiting time of the vehicular nodes to send out the packets. The QL approach for adaptive CW has guaranteed the expected result by optimizing the CW value. After training many times, CW value is determined by the exploration during the learning phase and succeeded in quickly adapting when the light or heavy network loads at more or less competing stations.

For further enhancement, the MA-DRL approach has been implemented and the results are compared with AC and EAC models. Figure 3 shows an improved throughput of the MA-DRL approach compared to other approaches and the throughput remains constant as the CW optimal value is considered for the increase in vehicular density. The increase in CW value means an increase in back-off interval accommodates the high number of transmissions from more vehicles, mitigating

This article has been accepted for publication in IEEE Internet of Things Journal. 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/JIOT.2026.3677251

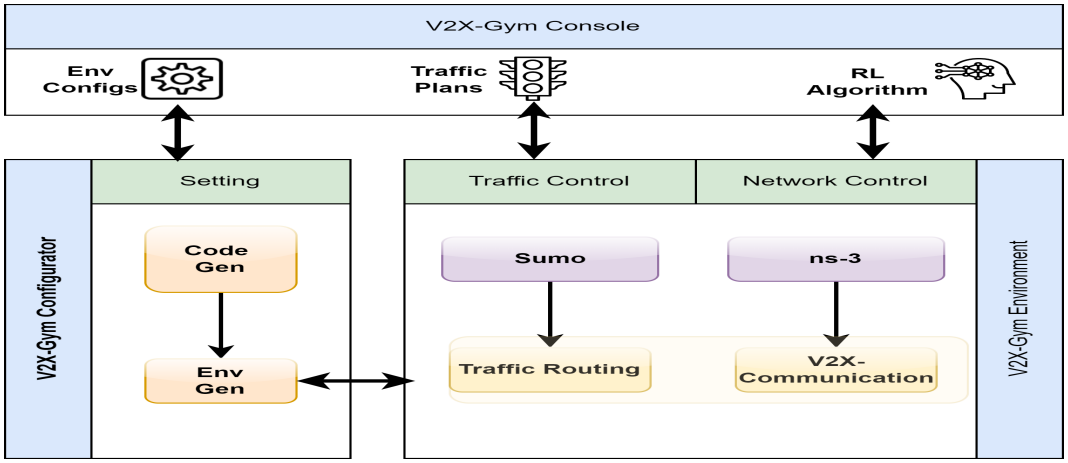

Fig. 2: Schematic Representation of Simulation Model.

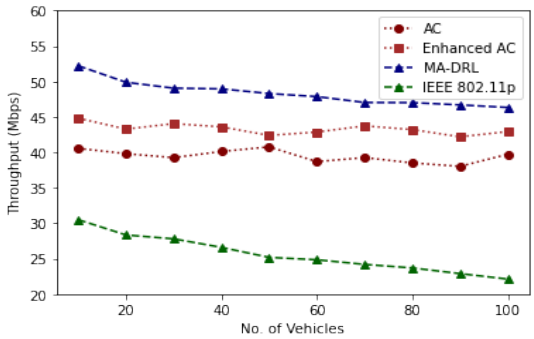

Fig. 3: Throughput Comparison.

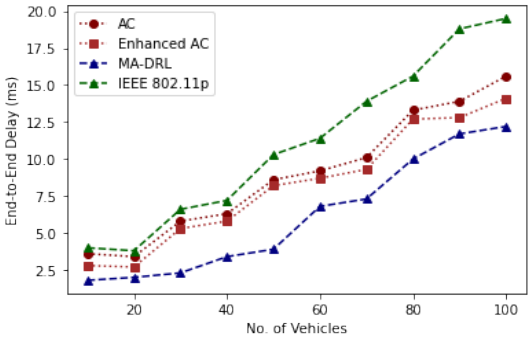

Fig. 4: End-to-End Delay Comparison.

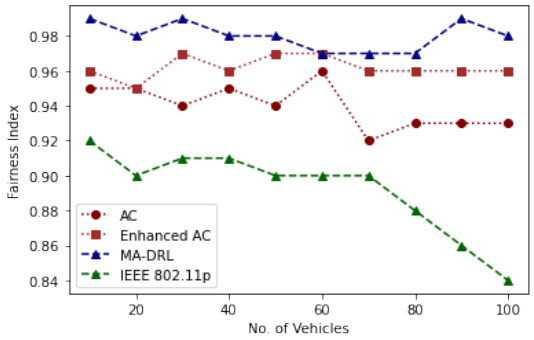

Fig. 5: Fairness Index Comparison.

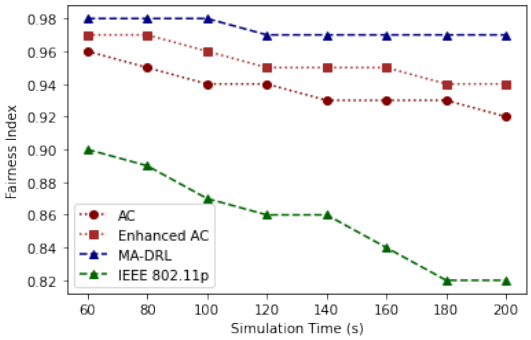

Fig. 6: Fairness Index Comparison varying Simulation Time (For 50 vehicles).

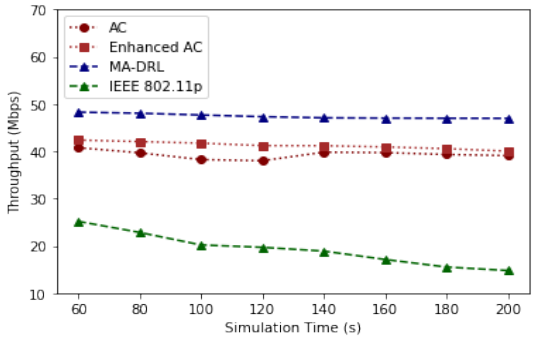

Fig. 7: Throughput Comparison varying Simulation Time (For 50 vehicles).

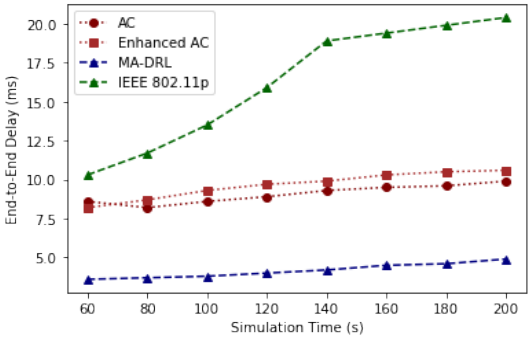

Fig. 8: End-to-End Delay Comparison varying Simulation Time (For 50 vehicles).

the number of packet collisions. As the DRL learns the environment more efficiently and choose discrete CW values

This article has been accepted for publication in IEEE Internet of Things Journal. 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/JIOT.2026.3677251

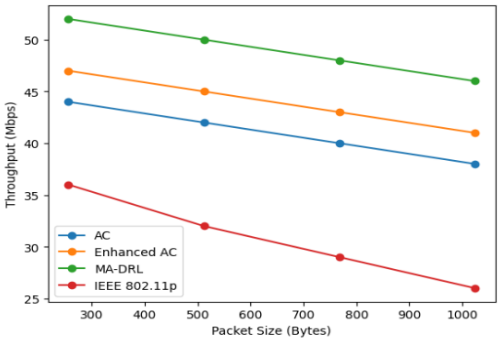

Fig. 9: Throughput Comparison varying Packet Size.

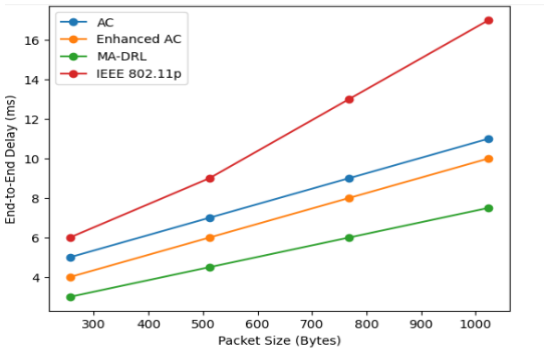

Fig. 10: End-to-End Delay Comparison varying Packet Size.

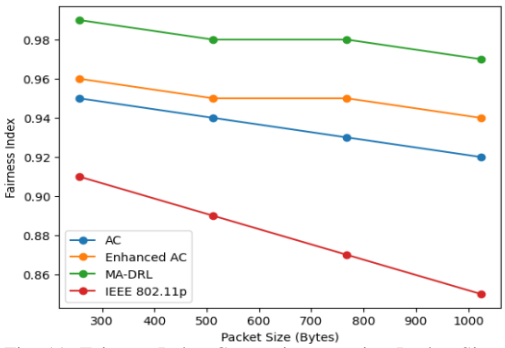

Fig. 11: Fairness Index Comparison varying Packet Size.

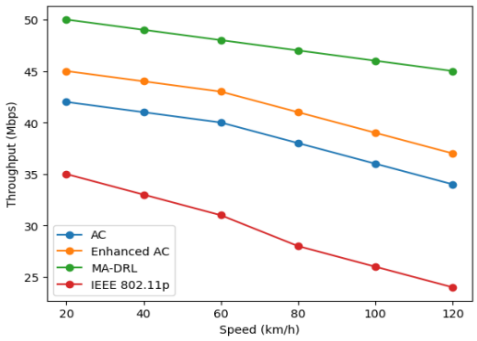

Fig. 12: Throughput Comparison varying Speed

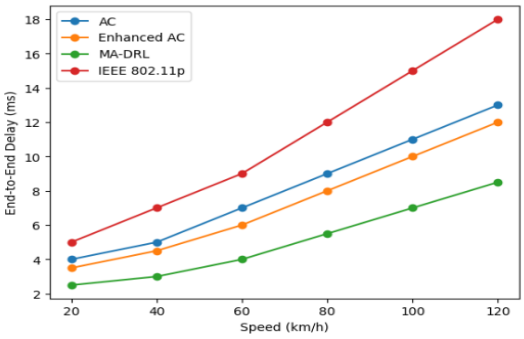

Fig. 13: End-to-End Delay Comparison varying Speed.

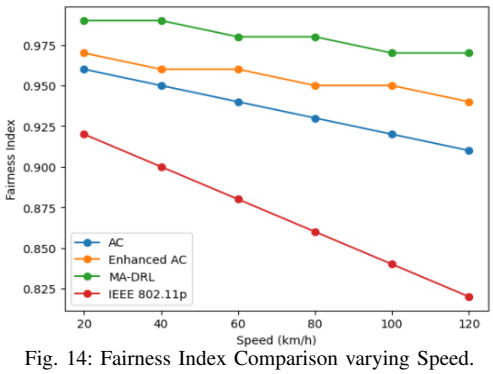

Fig. 14: Fairness Index Comparison varying Speed.

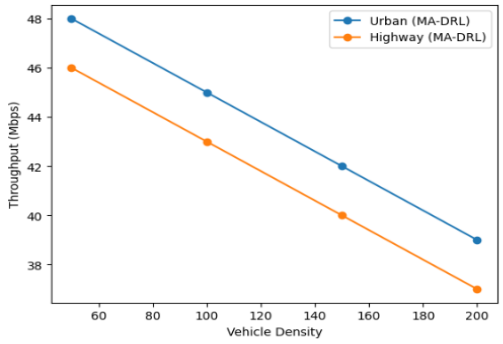

Fig. 15: Throughput Comparison varying Scenarios.

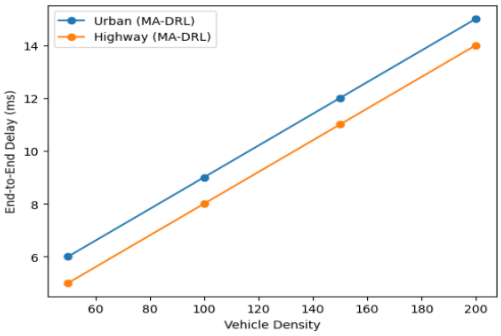

Fig. 16: End-to-End Delay Comparison varying Scenarios.

based on vehicular density, the throughput is better compared to other protocols. Similarly, as depicted in Figure 4, the MA-DRL approach provides less delay compared to AC, EAC, and IEEE 802.11p, as the number of collisions is minimized by adapting the CW value of the network. And, finally, in terms of the fairness index, the MA-DRL outperforms the standard protocol IEEE 802.11p, AC, and EAC as shown in Figure 5.

The simulation is performed for 50 vehicles varying the simulation time, and results are depicted in Figures 6, 7, and 8 . As the CW is adapted based on the number of vehicles, i.e. the number of transmissions, as the simulation time increases the throughput remains nearly consistent and MA-DRL shows

This article has been accepted for publication in IEEE Internet of Things Journal. 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/JIOT.2026.3677251

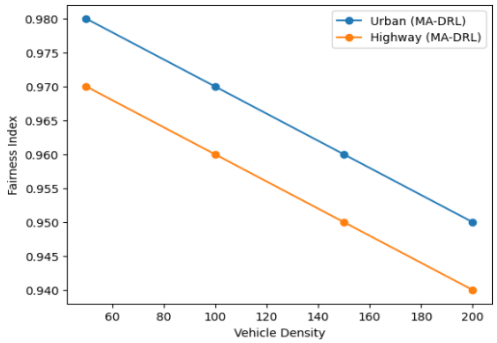

Fig. 17: Fairness Index Comparison varying Scenarios.

improved throughput compared to IEEE 802.11p, AC, and EAC. Similarly, in terms of delay and fairness index also the MA-DRL outperforms IEEE 802.11p, AC, and EAC with an increase in simulation time keeping vehicles constant.

For analysis of the proposed algorithm in rapid topological changes the performance matrices are evaluated with various packet size. The Figure 9, indicates that MA-DRL dynamically adjusts CW to avoid congestion, handling the large packets that tends to increase the channel occupancy. In the traditional, IEEE 802.11p and conventional AC and EAC-based schemes, this results in congestion due to multiple nodes repeatedly attempting for transmissions. In contrast, MA-DRL dynamically adjusts the contention window based on locally observed channel state providing collision analysis, reducing concurrent channel access. This adaptive behavior limits excessive contention and stabilizes channel utilization, thereby balancing the throughput even for increase in packet sizes. The Figure 10, provide the end-to-end delay analysis on increase in packet size. In IEEE 802.11p, increasing packet size amplifies the delay due to retransmission and back-off due to collisions, causing end-to-end delay to increase exponentially. MA-DRL mitigates this effect by learning contention policies that tune the channel access with collision avoidance. By preventing collisions and back-off, MA-DRL ensures that delay increases gradually with packet size, resulting in near-linear scaling rather than exponential growth. The Figure 11, demonstrates the fairness of channel access with respect to packet size. Basically, larger packets bias the channel access leading to unfair allocation in conventional schemes like the IEEE. The MA-DRL policy explicitly incorporates collision history and queue state, avoiding the bias channel access when packet sizes are large. Therefore, the channel access opportunities are evenly distributed across nodes regardless of packet size. This demonstrates the robustness of MA-DRL in supporting mixed VANET applications, including safety broadcasts, emergency messages and infotainment data, without explicit traffic differentiation.

With the increase in packet size, the transmission time per packet increases proportionally, leading to contention and higher channel occupancy dense VANET environments. The conventional IEEE 802.11p keeps the contention window static, causing the nodes to contend aggressively for the channel access. This results in repeated collisions and back-off that degrades the performance of the network. The Figure 12 presents the variation of throughput with increase in packet size. The proposed MA-DRL framework continuously tunes the contention window based on locally observed channel conditions. When packet sizes increase and the channel becomes occupied, the learning agent's policy responds by increasing the CW, thus reducing re-transmissions. This dynamic adjustment limits the collisions and maintain the network throughput. The Figure 13, depicts the trend of the end-to-end Delay with respect to packet size. The MA-DRL prevents excessive backoff by learning contention strategies that handles collision. Also, the retransmissions and prolonged waiting times are minimized balancing the delay with packet size. Whereas, the traditional IEEE 802.11p experiences retransmission delays and back-offs, resulting in exponential increase in delay. The Figure 14, presents the fairness trend with varying packet size. In the proposed MA-DRL the policy learned by the agent checks for queue occupancy and contention state, the algorithm handles channel access opportunities for agents that are vehicles, even under heterogeneous traffic loads.

Further, to discuss on the scenarios, the network performance of the proposed MA-DRL is evaluated under dense urban intersection and highway environments by varying the vehicle density from 50 to 200. As the number of vehicles increases the channel contention increases, reflecting on the throughput and delay of the network. As depicted in Figure 15, the throughput in both the scenarios shows throughput reduction with increasing density leading from higher contention and collision probability. The urban environment achieves slightly higher throughput than the highway scenario due to availability of next hop nodes and mobility pattern. The MA-DRL policy aligns with CW adaptation with the actual channel load, thereby providing higher effective channel utilization. In contrast, frequent topology changes in highway scenarios due to sparse network introduces contention uncertainty, leading to reduced throughput. Despite the high throughput, the urban scenario exhibits higher delay depicted in Figure 16. This is due to vehicle density at intersections, traffic signals, and uneven pattern, which leads to re-transmissions due to channel busy period and large back-off waiting times. Whereas, in highway scenario has spatial dispersion of vehicles and less node contending for channel access at a particular period of time within carrier-sensing range, leading to less delay. The MA-DRL framework tunes the CW in both cases based on the network status learning and policy, handling delay as density increases. The fairness results depicted in Figure 17, indicate that the proposed MA-DRL provides high fairness in both scenarios, of around 0.95 in urban, and 0.94 in highway scenarios at high vehicular densities. This is due the policy and learning factors in MA-DRL that avoids channel over and under utilization, providing proper channel access opportunities to each vehicle at even high density and mobility pattern.

VI. Conclusion

In this paper, an adaptive CW mechanism for VANETs using multi-agent RL is introduced and analysed. In the dynamic VANET environment, this contribution aims to address the crucial challenge of obtaining high throughput, less delay,

and maintaining a balanced fairness index. The extensive simulation and findings show that the proposed MA-DRL using DDPG approach finds the optimal CW for the specified vehicular density. Due to the incorporation of RL principles, there is an enhancement in the network's overall performance. The reduced contention and collision due to the optimal setting of CW leads to enhanced throughput. Less delay is achieved due to the optimized channel access strategy, which enables reliable and timely data transmission to vehicles in the network. The proper channel allocation provides a balanced fairness index, preventing a single node from dominating to transmit in the channel access. The proposed MA-DRL model feature demonstrates the possibility of cooperative behaviour in vehicular communication by enabling vehicles to collaboratively learn and adapt. The MA-DRL-based CW optimization adoption has the potential to build more resilient and effective networks, which will eventually contribute to the development of safer and more intelligent transportation systems.

VII. Future Scope

In this section, few future directions for the proposed framework in next-generation intelligent transportation systems are discussed. The first one includes integration with vehicular fog architectures [42] to support safety-critical vehicular applications. In fog-based VANET environments, computational and storage resources are distributed among RSUs, edge servers, and even parked vehicles, enabling data processing closer to the vehicles. By incorporating fog nodes into the proposed model, resource allocation and channel access mechanisms will be adapted to support real-time vehicular services such as multimedia streaming, traffic monitoring, and cooperative driving. Furthermore, intelligent optimization techniques, such as reinforcement learning or heuristic-based resource management, could be integrated to dynamically allocate fog resources based on vehicle mobility, traffic density, and service requirements. The second one includes deployment of the proposed framework within a Software-Defined Internet of Vehicles (SDN-IoV) architecture [43]. SDN-IoV introduces centralized network control through an SDN controller that separates the control plane from the data plane including vehicles, RSUs and other communicating devices for dynamic network communication. Integrating the proposed approach with an SDN-based framework will enable more efficient coordination of vehicular communication resources, improved traffic routing, and adaptive channel access mechanism. The SDN controller collects real-time network information and dynamically adjust policies to optimize network performance. Incorporating software-defined mechanisms will allow flexible deployment of network functions and improved large-scale vehicular environments supporting demand for high bandwidth applications such as video streaming and other infotainment services.

Acknowledgment

The authors gratefully acknowledge the support of the National Institute of Technology Rourkela, India, for providing the necessary facilities during the course of this research work. The authors extend their sincere thanks to NIT Rourkela, for its continuous encouragement and support throughout the patent drafting and development process. The invention titled "Adaptive Contention Window Optimization in VANETs Using Multi-Agent Deep Reinforcement Learning for Enhanced Performance Model" has been granted an Indian patent (Patent No. 570529, Application No. 202531006631) granted in 2025.

References

[1] Sassi Maaloul, Hasnaa Aniss, Leo Mendiboure, and Marion Berbineau. Performance analysis of existing its technologies: Evaluation and coexistence. *Sensors*, 22(24):9570, 2022.

[2] Neha Septa. The performance analysis of 802.11 p with cooperative communication and dynamic contention window. *Wireless Personal Communications*, pages 1–24, 2023.

[3] Biraja Prasad Nayak, Lopamudra Hota, Arun Kumar, Ashok Kumar Turuk, and Peter HJ Chong. Autonomous vehicles: Resource allocation, security, and data privacy. *IEEE Transactions on Green Communications and Networking*, 6(1):117–131, 2021.

[4] Arshee Ahmed, Haroon Rasheed, Ali Kashif Bashir, and Marwan Omar. Millimeter-wave channel modeling in a vanets using coding techniques. *PeerJ Computer Science*, 9:e1374, 2023.

[5] Lopamudra Hota, Bibhudatta Sahoo, and Arun Kumar. Fair channel allocation in ieee 802.11 p for high throughput and low-latency. *Physical Communication*, 71:102683, 2025.

[6] Wei Liu, Xinxin He, Zhitong Huang, and Yuefeng Ji. Transmission capacity characterization in vanets with enhanced distributed channel access. *Electronics*, 8(3):340, 2019.

[7] Lopamudra Hota, Biraja Prasad Nayak, Bibhudatta Sahoo, Peter HJ Chong, and Arun Kumar. An adaptive traffic-flow management system with a cooperative transitional maneuver for vehicular platoons. *Sensors*, 23(5):2481, 2023.

[8] Wenfeng Li, Wuli Song, Qiang Lu, and Chao Yue. Reliable congestion control mechanism for safety applications in urban vanets. *Ad Hoc Networks*, 98:102033, 2020.

[9] Praveen Kumar, Lopamudra Hota, Biraja Prasad Nayak, and Arun Kumar. An adaptive contention window using actor-critic reinforcement learning algorithm for vehicular ad-hoc networks. *Procedia Computer Science*, 235:3045–3054, 2024.

[10] Yi-Hao Tu, En-Cheng Lin, Chih-Heng Ke, and Yi-Wei Ma. Enhanced-setl: A multi-variable deep reinforcement learning approach for contention window optimization in dense wi-fi networks. *Computer Networks*, 253:110690, 2024.

[11] Shujuan Tian, Xinjie Zhu, Bochao Feng, Zhirun Zheng, Haolin Liu, and Zhetao Li. Partial offloading strategy based on deep reinforcement learning in the internet of vehicles. *IEEE Transactions on Mobile Computing*, 24(7):6517–6531, 2025.

[12] Jeffrey Redondo, Nauman Aslam, Juan Zhang, and Zhenhui Yuan. Multi-agent assessment with qos enhancement for hd map updates in a vehicular network and multi-service environment. *IEEE Transactions on Network Science and Engineering*, 2024.

[13] Lopamudra Hota, Biraja Prasad Nayak, Arun Kumar, GG Md Nawaz Ali, and Peter Han Joo Chong. An analysis on contemporary mac layer protocols in vehicular networks: state-of-the-art and future directions. *Future Internet*, 13(11):287, 2021.

[14] Martin Klapez, Carlo Augusto Grazia, and Maurizio Casoni. Application-level performance of ieee 802.11 p in safety-related v2x field trials. *IEEE Internet of Things Journal*, 7(5):3850–3860, 2020.

[15] Muhammet Ali Karabulut, AFM Shahen Shah, and Haci Ilhan. A novel mimo-ofdm based mac protocol for vanets. *IEEE Transactions on Intelligent Transportation Systems*, 23(11):20255–20267, 2022.

[16] Changsen Zhang, Pengpeng Chen, Jianji Ren, Xiaofei Wang, and Athanasios V Vasilakos. A backoff algorithm based on self-adaptive contention window update factor for ieee 802.11 dcf. *Wireless networks*, 23:749–758, 2017.

[17] Guilu Wu and Pingping Xu. Improving performance by a dynamic adaptive success-collision backoff algorithm for contention-based vehicular network. *IEEE Access*, 6:2496–2505, 2017.

[18] WANG Shuai, LU Yan, ZHU Jie, and WANG Ping. A novel collision supervision and avoidance algorithm for scalable mac of vehicular networks. *Chinese Journal of Electronics*, 30(1):164–170, 2021.

[19] Di Zhao, Hao Qin, Bin Song, Yanli Zhang, Xiaojiang Du, and Mohsen Guizani. A reinforcement learning method for joint mode selection and power adaptation in the v2v communication network in 5g. *IEEE Transactions on Cognitive Communications and Networking*, 6(2):452–463, 2020.

[20] Sung Hyun Park, Paul Daniel Mitchell, and David Grace. Reinforcement learning based mac protocol (uw-aloha-q) for underwater acoustic sensor networks. *IEEE access*, 7:165531–165542, 2019.

[21] Celimuge Wu, Satoshi Ohzahata, Yusheng Ji, and Toshihiko Kato. A mac protocol for delay-sensitive vanet applications with self-learning contention scheme. In *2014 IEEE 11th Consumer Communications and Networking Conference (CCNC)*, pages 438–443. IEEE, 2014.

[22] Guangjie Han, Aini Gong, Hao Wang, Miguel Martínez-García, and Yan Peng. Multi-auv collaborative data collection algorithm based on q-learning in underwater acoustic sensor networks. *IEEE Transactions on Vehicular Technology*, 70(9):9294–9305, 2021.

[23] Nasrin Taherkhani and Samuel Pierre. Centralized and localized data congestion control strategy for vehicular ad hoc networks using a machine learning clustering algorithm. *IEEE Transactions on Intelligent Transportation Systems*, 17(11):3275–3285, 2016.

[24] Miguel Sepulcre, Manuel Gonzalez-Martin, Javier Gozalvez, Rafael Molina-Masegosa, and Baldomero Coll-Perales. Analytical models of the performance of ieee 802.11 p vehicle to vehicle communications. *IEEE Transactions on Vehicular Technology*, 71(1):713–724, 2021.

[25] Yun-peng Wang, Kun-xian Zheng, Da-xin Tian, Xu-ting Duan, and Jian-shan Zhou. Cooperative channel assignment for vanets based on multiagent reinforcement learning. *Frontiers of Information Technology & Electronic Engineering*, 21(7):1047–1058, 2020.

[26] Khizra Asaf, Bilal Khan, and Ga-Young Kim. Wireless lan performance enhancement using double deep q-networks. *Applied Sciences*, 12(9):4145, 2022.

[27] Andreas Pressas, Zhengguo Sheng, Falah Ali, and Daxin Tian. A q-learning approach with collective contention estimation for bandwidth-efficient and fair access control in ieee 802.11 p vehicular networks. *IEEE Transactions on Vehicular Technology*, 68(9):9136–9150, 2019.

[28] Lusheng Miao, Karim Djouani, Barend J van Wyk, and Yskandar Hamam. Evaluation and enhancement of ieee 802.11 p standard: a survey. *Mobile Computing*, 1(1):15–30, 2012.

[29] Chien-Min Wu, Cheng-Tai Tsai, Cheng-Chun Hou, Jun-Jie Yang, Gong-De Lin, and Mi-Yu Kuang. Emergency message broadcast mechanism in vehicular ad-hoc networks based on reinforcement learning with contention estimation. *IEEE Transactions on Intelligent Vehicles*, 2024.

[30] Fernando A Teixeira, Vinicius F e Silva, Jesse L Leoni, Daniel F Macedo, and José MS Nogueira. Vehicular networks using the ieee 802.11 p standard: An experimental analysis. *Vehicular Communications*, 1(2):91–96, 2014.

[31] Zijun Zhao, Xiang Cheng, Miaowen Wen, Bingli Jiao, and Cheng-Xiang Wang. Channel estimation schemes for ieee 802.11 p standard. *IEEE intelligent transportation systems magazine*, 5(4):38–49, 2013.

[32] Imran Ali Qureshi and Sohail Asghar. A systematic review of the ieee-802.11 standard's enhancements and limitations. *Wireless Personal Communications*, 131(4):2539–2572, 2023.

[33] Ahmed Thair Shakir, Barbara M. Masini, Nemer Radhwan Khudhair, Rosdiadee Nordin, and Angela Amphawan. Priority-aware multi-agent deep reinforcement learning for resource scheduling in c-v2x mode 4 communication. *IEEE Access*, 13:129024–129039, 2025.

[34] Ifa Fatihah Mohamed Zain, Azlan Awang, and Anis Laouiti. Hybrid mac protocols in vanet: A survey. In *Vehicular Ad-Hoc Networks for Smart Cities: Second International Workshop, 2016*, pages 3–14. Springer, 2017.

[35] VanDung Nguyen, Chuan Pham, Thant Zin Oo, Nguyen H Tran, Eui-Nam Huh, and Choong Seon Hong. Mac protocols with dynamic interval schemes for vanets. *Vehicular Communications*, 15:40–62, 2019.

[36] Abdulrahman Saad Alqahtani, Azath Mubarakali, M Saravanan, Suresh Babu Changalasetty, Lalitha Saroja Thota, P Parthasarathy, and B Sivakumar. Enhanced machine learning approach with orthogonal frequency division multiplexing to avoid congestion in wireless communication system. *Optical and Quantum Electronics*, 55(10):913, 2023.

[37] Luke Snow and Vikram Krishnamurthy. Multi-agent inverse reinforcement learning for identifying pareto-efficient coordination–a distributionally robust approach. *arXiv preprint arXiv:2509.08956*, 2025.

[38] Yifei Song, Shuai Wang, Xuanhe Yang, Xiaqing Miao, Benedetta Picano, Chee Yen Leow, and Gaofeng Pan. A learning automaton mac protocol for directional fanets with throughput enhancement and fairness control. *IEEE Transactions on Network Science and Engineering*, 13:4472–4489, 2025.

[39] Dingbang Liu, Fenghui Ren, Jun Yan, Guoxin Su, Wen Gu, and Shohei Kato. Scaling up multi-agent reinforcement learning: An extensive survey on scalability issues. *IEEE Access*, 12:94610–94631, 2024.

[40] Jinfang Jiang, Yiling Dong, Guangjie Han, and Gang Su. Underwater acoustic mac protocol for multi-objective optimization based on multi-agent reinforcement learning. *Drones*, 9(2):123, 2025.

[41] Anatolij Zubow et al. ns3-gym: Extending openai gym for networking research. *arXiv preprint arXiv:1810.03943*, 2018.

[42] Yunli Cheng, A Vijayaraj, Kiran Sree Pokkuluri, Taybeh Salehnia, Ahmadreza Montazerolghaem, and Roqia Rateb. Vehicular fog resource allocation approach for vanets based on deep adaptive reinforcement learning combined with heuristic information. *IEEE Access*, 12:139056–139075, 2024.

[43] Ahmadreza Montazerolghaem. Efficient resource allocation for multimedia streaming in software-defined internet of vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 24(12):14718–14731, 2023.

[44] Bingjie Liang, Wenqi Lu, and Bin Ran. Deploying roadside unit efficiently in vanets: A multi-objective delay-based optimization strategy using lagrangian relaxation. *IEEE Transactions on Intelligent Transportation Systems*, 25(2):1646–1660, 2023.

APPENDIX

A. Lagrangian-Based Formulation of MA-DRL for CW Optimization

This appendix provides a theoretical interpretation of the proposed reward design from the perspective of constrained network utility maximization using Lagrangian relaxation [44].

1) Constrained Network Optimization Formulation

The contention window (CW) adaptation problem in dense VANETs can be formulated as a constrained stochastic optimization problem, where the objective is to maximize network throughput while satisfying reliability and latency constraints:

$$\max_{\pi} \quad \mathbb{E}_{\pi}\big[\text{SuccTx}\big], \tag{22}$$

$$\text{s.t.} \quad \mathbb{E}_{\pi}\big[\text{Collisions}\big] \le C_{\max}, \tag{23}$$

$$\mathbb{E}_{\pi}\big[\tau\big] \le \tau_{\max}, \tag{24}$$

where π denotes the joint policy of all agents, SuccTx is the number of successful transmissions, Collisions denotes packet collision events, and τ is the end-to-end transmission delay. The bounds $C_{\max}$ and $\tau_{\max}$ reflect application-level QoS requirements for vehicular safety and reliability.

2) Lagrangian Relaxation

The constrained problem in (22)–(24) can be transformed into an unconstrained optimization via Lagrangian relaxation:

$$\mathcal{L}(\pi) = \mathbb{E}_{\pi}\big[\text{SuccTx}\big] - \beta\left(\mathbb{E}_{\pi}\big[\text{Collisions}\big] - C_{\max}\right) - \eta\left(\mathbb{E}_{\pi}\big[\tau\big] - \tau_{\max}\right), \tag{25}$$

where $\beta \ge 0$ and $\eta \ge 0$ are Lagrange multipliers associated with collision and delay constraints, respectively. Ignoring constant terms $C_{\max}$ and $\tau_{\max}$, the per-time-step stochastic objective yields the instantaneous reward:

$$r_t^i = \alpha \cdot \text{SuccTx}_t^i - \beta \cdot \text{Collisions}_t^i - \eta \cdot \tau_t^i, \tag{26}$$

which matches the reward formulation adopted in Eq. (9) of the paper.

3) Constrained Network Utility Maximization

The contention window (CW) adaptation problem in dense VANETs can be formulated as a constrained optimization problem over the joint policy π:

$$\max_{\pi} \quad \mathbb{E}_{\pi}\big[\text{SuccTx}\big], \tag{27}$$

$$\text{s.t.} \quad \mathbb{E}_{\pi}\big[\text{Collisions}\big] \le C_{\max}, \tag{28}$$

$$\mathbb{E}_{\pi}\big[\tau\big] \le \tau_{\max}, \tag{29}$$

where SuccTx denotes the number of successful transmissions (throughput), Collisions represents packet collision events, and

τ is the end-to-end delay. The constants $C_{\max}$ and $\tau_{\max}$ correspond to application-level reliability and latency constraints, respectively.

4) Lagrangian Relaxation

The constrained problem in (27)–(29) can be transformed into an unconstrained optimization using Lagrangian relaxation:

$$\mathcal{L}(\pi;\beta,\eta) = \mathbb{E}_\pi\big[\text{SuccTx}\big] - \beta\Big(\mathbb{E}_\pi\big[\text{Collisions}\big] - C_{\max}\Big) - \eta\Big(\mathbb{E}_\pi\big[\tau\big] - \tau_{\max}\Big), \tag{30}$$

where $\beta \geq 0$ and $\eta \geq 0$ are Lagrange multipliers associated with the collision and delay constraints, respectively. Ignoring constant terms $C_{\max}$ and $\tau_{\max}$, the instantaneous reward can be written as:

$$r_t^i = \alpha \cdot \text{SuccTx}_t^i - \beta \cdot \text{Collisions}_t^i - \eta \cdot \tau_t^i, \tag{31}$$

which matches the reward formulation used in Eq. (9) of the paper.

5) Interpretation of Weighting Factors

From the Karush–Kuhn–Tucker (KKT) optimality conditions, the multipliers β and η satisfy:

$$\beta \geq 0, \quad \beta\big(\mathbb{E}_\pi[\text{Collisions}] - C_{\max}\big) = 0, \tag{32}$$

$$\eta \geq 0, \quad \eta\big(\mathbb{E}_\pi[\tau] - \tau_{\max}\big) = 0, \tag{33}$$

indicating that the relative magnitude of β and η governs how strongly the learned policy enforces reliability and latency constraints, respectively. Specifically:

- α controls the relative priority of throughput maximization and reflects the aggressiveness of channel access.
- β penalizes collisions and acts as a reliability constraint multiplier.
- η penalizes delay and regulates latency sensitivity for safety-critical ITS applications.

Dr. Lopamudra Hota (lhota@bitmesra.ac.in) is currently an Assistant Professor in the Department of Computer Science and Engineering at the Birla Institute of Technology-Mesra, Ranchi, India. She completed her Ph.D. from the Department of Computer Science and Engineering, National Institute of Technology, Rourkela, India, in 2024. Her research interest includes Vehicular Adhoc Networks, IoT, and Machine Learning

Dr. Arun Kumar (kumararun@nitrkl.ac.in) is an Assistant Professor in the Department of CSE at NIT, Rourkela, India. He has been a Research Fellow at EMDL Lab in the Department of ECE of NUS, Singapore from 2015 to 2018. He has worked as a Post-Doctoral Research Fellow at the Institute of Information Science, Academia Sinica, Taipei, Taiwan from 2014 to 2015. He received his PhD degree from the School of Computer Engineering, NTU, Singapore in 2014. His research interest includes Internet of Things, wireless sensor networks, adhoc and mobile networks, Data analytics, Sentiment Analysis, and Vehicular AdHoc Networks.

Prof. Peter Han Joo Chong (peter.chong@aut.ac.nz) is an Associate Head of School (Research), School of Engineering, Computer and Mathematical Sciences, at Auckland University of Technology, New Zealand. He received the Ph.D. degree from the University of British Columbia, Canada, in 2000. He was previously an Associate Professor (tenured) at Nanyang Technological University, Singapore. His research interests are in the areas of mobile communications systems including V2X, Internet of Things/Vehicles, artificial intelligence for wireless networks, and 5G networks.