Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Multi-Agent Deep Reinforcement Learning Framework for Distributed Resource Allocation in 5G NR-V2X Mode 2 Under Dynamic Traffic Conditions

Ali Nihad Al-Najjar
Universiti Putra Malaysia
Mohd Fadlee A. Rasid
fadlee@upm.edu.my
Universiti Putra Malaysia
Fazirulhisyam Hashim
Universiti Putra Malaysia
Faisul Arif Ahmad
Universiti Putra Malaysia

Article

Keywords: NR-V2X, distributed resource allocation (DRA), sensing-based semi-persistent scheduling (SB-SPS), multi-agent deep reinforcement learning (MADRL), cooperative awareness messages (CAM), Markov decision process (MDP)

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11060584/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Multi-Agent Deep Reinforcement Learning Framework for Distributed Resource Allocation in 5G NR-V2X Mode 2 Under Dynamic Traffic Conditions

Ali Nihad Al-Najjar[1,2,3], Mohd Fadlee A. Rasid[1,2], Fazirulhisyam Hashim[1,2], Faisul Arif Ahmad[1,2]

[1]Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, Serdang, 43400, Malaysia.

[2]Wireless and Photonics Networks Research Centre (WiPNET), Faculty of Engineering, Universiti Putra Malaysia, Serdang, 43400, Malaysia.

[3]Studies Planning and Follow-up Directorate, Ministry of Higher Education and Scientific Research, Baghdad, 10065, Iraq.

Abstract

Efficient Distributed Resource Allocation (DRA) is a critical challenge in Fifth Generation New Radio Vehicle-to-Everything (5G NR-V2X) Mode 2 communication, particularly under dynamic vehicular environments characterized by high mobility and unpredictable traffic patterns. In this mode, vehicles autonomously select sidelink transmission resources without centralized coordination, which increases the risk of resource conflicts and packet collisions. This challenge becomes more severe in the presence of Cooperative Awareness Messages (CAMs), whose transmission behavior often includes both periodic and aperiodic components. Event-driven aperiodic CAMs introduce bursty and non-stationary traffic loads that significantly degrade the performance of the standardized Sensing-Based Semi-Persistent Scheduling (SB-SPS) mechanism. To address these limitations, this paper proposes a learning-assisted DRA framework based on Multi-Agent Deep Reinforcement Learning (MADRL). The dynamic vehicular environment is formulated as a Markov Decision Process (MDP), enabling structured modeling of traffic variability, channel conditions, and resource utilization. A Deep Q-Network (DQN)-based multi-agent learning architecture is integrated into the SB-SPS resource selection process to enable adaptive, decentralized decision-making. Unlike conventional rule-based scheduling, the proposed approach allows each vehicle to learn optimal resource allocation policies through interaction with the environment, thereby improving responsiveness to dynamic traffic conditions. A custom Python-based NR-V2X Mode 2 simulation framework is developed to evaluate the proposed approach under a highway scenario with varying vehicle densities and mixed periodic–aperiodic CAM traffic. Performance is assessed using Packet Reception Ratio (PRR), Collision Loss Ratio (CLR), latency, and Channel Busy Ratio (CBR). Simulation results demonstrate that the proposed MADRL-assisted framework significantly enhances communication reliability, achieving up to 8% improvement in PRR for 190-byte CAMs and 20% improvement for 470-byte CAMs under aperiodic traffic conditions, while substantially reducing collision probability and transmission latency compared with conventional SB-SPS. These results highlight the effectiveness of learning-assisted scheduling in overcoming the structural limitations of SB-SPS and demonstrate the potential of MADRL-based DRA for next-generation intelligent vehicular communication systems.

Keywords: NR-V2X, distributed resource allocation (DRA), sensing-based semi-persistent scheduling (SB-SPS), multi-agent deep reinforcement learning (MADRL), cooperative awareness messages (CAM), Markov decision process (MDP).

Corresponding author: Mohd Fadlee A. Rasid (e-mail: fadlee@upm.edu.my)

1. Introduction

The rapid development of Intelligent Transportation Systems (ITS) has significantly increased the demand for reliable, low-latency, and scalable vehicular communication technologies capable of supporting safety-critical and cooperative applications. Vehicle-to-Everything (V2X) communication enables vehicles to exchange real-time information with surrounding entities, including other vehicles, infrastructure, and pedestrians, thereby enhancing situational awareness and improving road safety and traffic efficiency [1]. Cooperative vehicular networking is widely recognized as a key enabler for advanced applications such as autonomous driving, platooning, and collision avoidance systems [2].

Early vehicular communication systems were primarily based on Dedicated Short-Range Communication (DSRC), which provides direct communication among vehicles with low latency. However, DSRC suffers from scalability limitations and inefficient spectrum utilization in dense vehicular environments [2]. To address these challenges, Cellular Vehicle-to-Everything (C-V2X) communication has emerged as a promising alternative, leveraging cellular technologies to improve coverage, reliability, and spectral efficiency [3]. The evolution toward Fifth Generation (5G) New Radio Vehicle-to-Everything (NR-V2X) further enhances vehicular communication by supporting flexible numerologies, higher data rates, and ultra-reliable low-latency communication (URLLC) capabilities [4].

NR-V2X defines two operational modes for sidelink communication: Mode 1 and Mode 2. Mode 1 relies on centralized scheduling by network infrastructure, whereas Mode 2 enables distributed communication, allowing vehicles to autonomously select transmission resources without centralized coordination [4]. This distributed operation is particularly suitable for safety-critical scenarios and environments with limited network coverage. However, the absence of centralized control introduces significant challenges in resource allocation, especially in dynamic vehicular environments characterized by high mobility, rapidly changing topology, and varying traffic density [5].

A fundamental component of V2X communication is the Cooperative Awareness Message (CAM), which enables vehicles to periodically broadcast their state information, including position, speed, and direction [6]. In realistic vehicular environments, CAM transmissions exhibit both periodic and aperiodic behavior. Aperiodic CAM transmissions are triggered by sudden events such as emergency braking or lane changes, resulting in bursty and unpredictable traffic loads [7]. These dynamic transmission patterns significantly complicate Distributed Resource Allocation (DRA), as conventional scheduling mechanisms are not designed to handle non-stationery and event-driven traffic conditions.

To support DRA in NR-V2X Mode 2, the Third Generation Partnership Project (3GPP) has standardized the Sensing-Based Semi-Persistent Scheduling (SB-SPS) mechanism [8]. SB-SPS allows vehicles to sense channel occupancy and reserve transmission resources semi-persistently, thereby reducing signaling overhead and enabling decentralized scheduling. While SB-SPS performs effectively under moderate traffic conditions, its rule-based and semi-static nature limits its adaptability in dynamic environments. In dense vehicular scenarios, vehicles may independently select identical resources due to incomplete sensing information, leading to increased packet collisions and degraded communication reliability [9]. These limitations become more pronounced in the presence of aperiodic CAM traffic, where unpredictable transmission bursts disrupt semi-persistent reservations.

Recent advances in Machine Learning (ML) have introduced new opportunities for adaptive resource allocation in wireless communication systems. Reinforcement Learning (RL) enables agents to learn optimal decision policies through interaction with dynamic environments, making it suitable for complex and uncertain wireless systems [10]. Deep Reinforcement Learning (DRL), which integrates RL with deep neural networks, further enhances this capability by enabling efficient handling of high-dimensional state spaces and complex decision-making problems [11]. In distributed vehicular networks, Multi-Agent Deep Reinforcement Learning (MADRL) has emerged as a promising approach due to its ability to model interactions among multiple autonomous agents sharing a common wireless medium [12].

Several studies have explored DRL-based resource allocation in wireless networks and vehicular communication systems, demonstrating improved adaptability compared with

conventional rule-based approaches [13]. However, most existing works focus on Long-Term (LTE)-V2X systems or assume centralized coordination, limiting their applicability to NR-V2X Mode 2 communication. Furthermore, many studies rely on simplified traffic models that do not accurately capture the stochastic and aperiodic nature of CAM transmissions in realistic vehicular environments. Consequently, there remains a critical research gap in developing adaptive, learning-assisted DRA mechanisms specifically designed for NR-V2X Mode 2 under dynamic traffic conditions.

To address this gap, this paper proposes a MADRL-based DRA framework for NR-V2X Mode 2 communication. The dynamic vehicular environment is modeled as a Markov Decision Process (MDP), enabling structured representation of system states, actions, and rewards. A Deep Q-Network (DQN)-based multi-agent learning architecture is integrated into the SB-SPS resource selection process, allowing vehicles to adaptively select transmission resources based on realtime observations. This approach enhances the adaptability of distributed scheduling while maintaining compatibility with the standardized SB-SPS framework.

The main contributions of this paper are summarized as follows:

- **MDP-based modeling of dynamic vehicular traffic:** An MDP framework is developed to model dynamic vehicular environments, incorporating both periodic and aperiodic CAM transmissions and capturing their stochastic and time-varying characteristics.
- **MADRL-assisted DRA framework:** An MADRL approach is proposed to enhance DRA in NR-V2X Mode 2, enabling decentralized and adaptive decision-making.
- **Integration of DQN with SB-SPS scheduling:** An DQN model is embedded into the SB-SPS resource selection process, allowing vehicles to learn optimal resource allocation strategies while preserving the underlying distributed scheduling mechanism.
- **Custom NR-V2X Mode 2 simulation framework:** A flexible Python-based simulation platform is developed to evaluate DRA under realistic vehicular scenarios, including dynamic traffic conditions and varying vehicle densities.

The remainder of this paper is organized as follows. Section 2 reviews related work on DRA and learning-based approaches in vehicular networks. Section 3 presents the system model and problem formulation. Section 4 describes the MDP-based traffic modeling and the proposed MADRL framework. Section 5 details the simulation setup, while Section 6 presents and discusses the results. Finally, Section 7 concludes the paper and outlines future research directions.

2. Related Work

Efficient resource allocation is a fundamental challenge in V2X communication systems, particularly in dynamic vehicular environments characterized by high mobility, interference, and time-varying traffic patterns. Recent studies emphasize that resource allocation plays a critical role in ensuring communication reliability, latency, and spectral efficiency in both LTEV2X and NR-V2X systems [14]. As vehicular networks evolve toward 5G NR-V2X, the complexity of distributed scheduling mechanisms has increased significantly due to the absence of centralized coordination and the need to support diverse traffic types.

The SB-SPS mechanism is the standardized approach for DRA in NR-V2X Mode 2 communication. In this scheme, vehicles autonomously select transmission resources based on local sensing of channel occupancy. While SB-SPS reduces signaling overhead and enables decentralized operation, it suffers from several inherent limitations in dynamic vehicular environments.

Recent studies have shown that SB-SPS is highly sensitive to network density and mobility. In particular, packet collisions become more frequent due to hidden node problems and incomplete sensing information [15]. Furthermore, the semi-persistent nature of SB-SPS makes it unsuitable for handling bursty and aperiodic traffic patterns, as resource reservations may become outdated when traffic conditions change rapidly. These issues significantly degrade PRR and increase latency in dense vehicular scenarios.

To address these limitations, several extensions and optimization techniques have been proposed. For example, recent work has investigated dynamic reselection mechanisms and

adaptive reservation strategies to mitigate collisions and improve reliability [15]. Other studies have explored cooperative resource allocation schemes where vehicles exchange scheduling information to reduce conflicts, demonstrating improvements over conventional Mode 2 operation [16]. However, these approaches often require additional signaling or coordination, which may not be feasible in fully distributed environments.

Beyond SB-SPS enhancements, a wide range of optimization-based approaches have been proposed for resource allocation in V2X communication systems. These methods typically aim to optimize key parameters such as channel selection, power allocation, and interference management.

Recent research has introduced intelligent resource allocation frameworks that treat vehicular communication links as competing agents sharing limited spectrum resources. For instance, asynchronous resource allocation strategies have been proposed to allow vehicles to dynamically decide transmission opportunities and adjust power levels, improving overall network stability and reducing interference [17]. Similarly, optimization-based methods focusing on interference-aware scheduling and joint V2V–V2N coordination have demonstrated improved reliability and latency performance in complex vehicular scenarios [18].

Despite these advancements, optimization-based approaches typically rely on accurate system modeling and may suffer from scalability issues in large-scale vehicular networks. Additionally, their performance is often limited in highly dynamic environments where traffic conditions and channel states change rapidly.

To overcome the limitations of rule-based and optimization-based approaches, recent studies have explored the use of Machine Learning (ML) techniques for adaptive resource allocation. In particular, DRL has emerged as a promising solution for handling complex and dynamic wireless environments.

DRL-based approaches enable agents to learn optimal resource allocation policies through interaction with the environment, without requiring explicit modeling of system dynamics. Recent works have demonstrated that DRL can significantly improve resource allocation performance in vehicular networks by adapting to changing channel conditions and traffic patterns [19]. For example, Q-learning-based scheduling mechanisms have been proposed to reduce collision probability and enhance communication reliability in NR-V2X sidelink systems.

More advanced DRL techniques, including multi-agent reinforcement learning (MARL), have been applied to distributed vehicular networks. These approaches model each vehicle as an autonomous agent that interacts with other agents in a shared wireless environment. Recent studies have shown that MARLbased resource allocation can improve spectrum utilization and reduce interference by enabling cooperative decision-making among vehicles [17].

Furthermore, recent work has explored state-aware DRL models that incorporate richer environmental information, such as channel conditions, interference levels, and historical observations, to enhance learning performance. These approaches improve agents' ability to perceive environmental dynamics and achieve better coordination in complex V2X scenarios [13].

Despite significant progress, several limitations remain in the existing literature:

- **Limited focus on NR-V2X Mode 2:** Many DRL-based resource allocation studies are developed for LTE-V2X or assume centralized control, limiting their applicability to fully distributed NR-V2X Mode 2 systems.
- **Simplified traffic modeling:** Existing works often rely on periodic traffic assumptions and do not adequately capture the stochastic and aperiodic nature of CAM transmissions in real-world vehicular environments.
- **Lack of multi-agent coordination:** While multi-agent approaches have been explored, many existing frameworks fail to fully exploit cooperation among agents, leading to suboptimal performance in dense networks [13].

- **Insufficient integration with standardized mechanisms:** Most learning-based approaches replace existing scheduling mechanisms rather than integrating with standardized SB-SPS, limiting their practical applicability.

To address these limitations, this paper proposes a MADRL-assisted DRA framework that:

- Models dynamic vehicular traffic using an MDP framework.
- Integrates multi-agent DQN learning with SB-SPS.
- Captures periodic and aperiodic CAM behavior.
- Enables adaptive and decentralized scheduling.

This approach aims to bridge the gap between standard-compliant scheduling (SB-SPS) and intelligent learning-based resource allocation, providing a practical and scalable solution for NR-V2X Mode 2 communication.

3 System Model and Problem Formulation

3.1 Network Model

We consider a highway vehicular communication scenario where vehicles communicate using 5G NR-V2X Mode 2 sidelink transmission. In this mode, vehicles operate in a fully distributed manner and autonomously select transmission resources from a shared sidelink resource pool without relying on centralized network coordination.

The considered scenario consists of a multi-lane bidirectional highway with varying vehicle densities. Each vehicle periodically or aperiodically transmits CAMs to neighboring vehicles within a predefined communication range. Communication operates over the 5.9 GHz band, which is commonly allocated for vehicular communications.

Let $V = \{1, 2, \ldots, N\}$ denote the set of vehicles in the network, where N represents the total number of active vehicles. Each vehicle acts as an independent transmitter and receiver, forming a decentralized Vehicle-to-Vehicle (V2V) communication network.

The sidelink resource pool is composed of time-frequency resource blocks. Let $R = \{r_1, r_2, \ldots, r_M\}$ denote the set of available resources, where M is the total number of resource blocks. Each resource block corresponds to a combination of subchannels and time slots.

Due to the broadcast nature of CAM transmissions, interference occurs when multiple vehicles select the same resource block within overlapping communication ranges. Therefore, efficient resource allocation is critical to minimize packet collisions and ensure reliable communication.

3.2 NR-V2X Mode 2 Resource Allocation

In NR-V2X Mode 2, resource allocation is performed in a distributed manner based on local sensing and decision-making. Vehicles select transmission resources from the available resource pool using predefined scheduling mechanisms. Each vehicle performs resource selection over a selection window, based on observations collected during a sensing window. The objective is to avoid selecting resources that are likely to be occupied by neighboring vehicles. However, due to limited sensing range, hidden node problems, and dynamic traffic conditions, vehicles may still select conflicting resources, leading to packet collisions and degraded performance.

3.3 SB-SPS Scheduling Mechanism

The baseline resource allocation scheme considered in this work is the SB-SPS mechanism defined in 3GPP NR-V2X standards.

SB-SPS operates in three main steps:

- **Channel Sensing:** Each vehicle monitors the channel over a sensing window to identify occupied resources based on received signal strength and historical usage.
- **Candidate Resource Selection:** A set of candidate resources is selected from the available pool by excluding highly occupied resources.

- **Semi-Persistent Reservation:** The selected resource is reserved for a predefined number of consecutive transmissions, determined by a reselection counter.

Let $r_i(t) \in R$ denote the resources selected by vehicle i at time t. The SB-SPS mechanism aims to minimize collisions by selecting resources that appear unused during the sensing phase.

Despite its effectiveness under moderate traffic conditions, SB-SPS suffers from several limitations:

- Inability to adapt to rapidly changing traffic conditions.
- Performance degradation under high vehicle density.
- Inefficiency in handling aperiodic traffic bursts.
- Lack of predictive capability.

These limitations motivate the need for learning-based resource allocation mechanisms.

3.4 Channel and Interference Model

We consider a simplified interference model suitable for system-level simulation. A successful packet reception occurs when the received signal-to-interference-plus-noise ratio (SINR) exceeds a predefined threshold.

The SINR at receiver *j* from transmitter *i* is given by:

$$\boldsymbol{SNIR_{i,j}} = \frac{\boldsymbol{P_i G_{i,j}}}{\sum_{\boldsymbol{k \in I}} \boldsymbol{P_k G_{k,j}} + \boldsymbol{N_0}} \qquad (1)$$

where:

- P_i is the transmit power of vehicle i.
- $G_{i,j}$ is the channel gain between vehicles *i* and *j*.
- I is the set of interfering vehicles.
- N_0 is the noise power A packet is successfully received if:

$$\boldsymbol{SINR_{i,j}} \geq \boldsymbol{\gamma_{th}} \qquad (2)$$

where:

- π is the resource allocation policy.
- $\gamma \in (0, 1]$ is the discount factor.

3.5 Motivation for Learning-Based Approach

The above optimization problem is difficult to solve using conventional methods due to:

- Dynamic and stochastic traffic patterns.
- Partial observability of the environment.
- Distributed decision-making constraints.

To address these challenges, the resource allocation problem is modeled as an MDP and solved using an MADRL framework, where each vehicle learns an optimal resource selection policy through interaction with the environment.

4 MDP-Based Modeling and Proposed MADRL Framework

4.1 MDP Formulation

To address the challenges of DRA in dynamic vehicular environments, the problem is formulated as an MDP. The MDP framework provides a structured approach for modeling sequential decision-making under uncertainty, enabling vehicles to learn optimal resource allocation policies through interaction with the environment.

The MDP is defined as a tuple:

$$\boldsymbol{\mathcal{M}} = (\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{P}}, \boldsymbol{\mathcal{R}}) \qquad (3)$$

where:

- S is the state space.

- A is the action space.
- P is the state transition probability.
- R is the reward function.

Each vehicle is modeled as an agent interacting with the wireless environment and making resource allocation decisions based on local observations.

4.2 State Space Design

The state representation plays a critical role in capturing the dynamic characteristics of the vehicular environment. In this work, the state observed by each vehicle at time t is defined as:

$$\boldsymbol{s_t} = \{\boldsymbol{C_t}, \boldsymbol{U_t}, \boldsymbol{D_t}, \boldsymbol{H_t}\} \qquad (4)$$

where:

- C_t: Channel occupancy information (sensed resource usage).
- U_t: Resource utilization levels within the selection window.
- D_t: Local vehicle density (number of neighboring vehicles).
- H_t: Historical resource selection and collision feedback.

This state formulation allows the agent to capture both instantaneous channel conditions and temporal traffic patterns, which is essential for handling aperiodic CAM transmissions.

4.3 Action Space

The action space corresponds to the selection of transmission resources from the available sidelink resource pool.

$$\boldsymbol{a_t} \in \boldsymbol{\mathcal{A}} = \{\boldsymbol{r_1}, \boldsymbol{r_2}, \ldots, \boldsymbol{r_M}\} \qquad (5)$$

where each action represents the selection of a specific resource block for transmission.

The size of the action space is determined by the number of candidate resources within the selection window.

4.4 Reward Function Design

The reward function is designed to guide the learning process toward improving communication reliability while minimizing collisions and latency.

$$\boldsymbol{R_t} = \boldsymbol{\alpha} \cdot \boldsymbol{PRR_t} - \boldsymbol{\beta} \cdot \boldsymbol{CLR_t} - \boldsymbol{\delta} \cdot \boldsymbol{D_t} \qquad (6)$$

where:

- PRR_t: Packet Reception Ratio (success metric).
- CLR_t: Collision Loss Ratio (penalty term).
- D_t: Transmission delay.
- α, β, δ: weighting coefficients.

This formulation ensures that:

- Successful transmissions are rewarded.
- Collisions are penalized.
- Delay-sensitive performance is considered.

The reward design directly aligns with the system-level performance metrics used in the evaluation.

4.5 Multi-Agent Learning Environment

In the considered NR-V2X Mode 2 scenario, all vehicles operate in a distributed manner and independently select transmission resources. Therefore, the system is modeled as a multi-agent environment, where:

- Each vehicle acts as an autonomous agent.
- Agents interact indirectly through shared wireless resources.
- The environment is partially observable.

Let N denote the number of agents. The joint action space is given by:

$$\mathcal{A}_{joint} = \mathcal{A}_1 \times \mathcal{A}_2 \times \cdots \times \mathcal{A}_N \quad (7)$$

Due to the distributed nature of the system, each agent learns its own policy $\pi i(a \mid s)$ based on local observations without centralized coordination.

4.6 DQN Framework

To solve the MDP, a DQN is employed for each agent. The DQN approximates the optimal action-value function:

$$Q^*(s,a) = \max_{\pi} \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t R_t | s_0 = s, a_0 = a] \quad (8)$$

The Q-function is approximated using a neural network parameterized by θ:

$$Q(s,a;\theta) \quad (9)$$

The network takes the state st as input and outputs Q-values for all possible actions.

4.6.1 Q-Learning Update Rule

The network parameters are updated using the temporal-difference (TD) learning rule:

$$\theta \leftarrow \theta + \eta(y_t - Q(s_t,a_t;\theta))\nabla_\theta Q(s_t,a_t,\theta) \quad (10)$$

where the target value y_t is defined as:

$$y_t = R_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-) \quad (11)$$

and θ represents the parameters of a target network.

4.6.2 Experience Replay

To stabilize training, an experience replay buffer D is used to store transitions:

$$(s_t, a_t, R_t, s_{t+1}) \quad (12)$$

Mini-batches are randomly sampled from D to update the network, reducing correlation among samples and improving convergence.

4.6.3 Exploration Strategy

A ε-greedy policy is adopted for action selection:

$$a_t = \begin{cases} \text{random action}, & \text{with probability } \epsilon \\ \arg\max_a Q(s_t,a), & \text{otherwise} \end{cases} \quad (13)$$

This balances exploration and exploitation during training.

The overall architecture of the proposed MADRL-based resource allocation framework is illustrated in Fig. 1. The framework integrates multiple decentralized learning agents with the NR-V2X Mode 2 communication stack, where each agent observes local network conditions and selects transmission resources using a DQN. The architecture highlights the interaction between the MDP-based environment, experience replay mechanism, and multi-agent decision-making process. In particular, the integration of the learning module with the SB-SPS scheduling procedure enables adaptive resource selection while preserving compatibility with the standardized sidelink protocol.

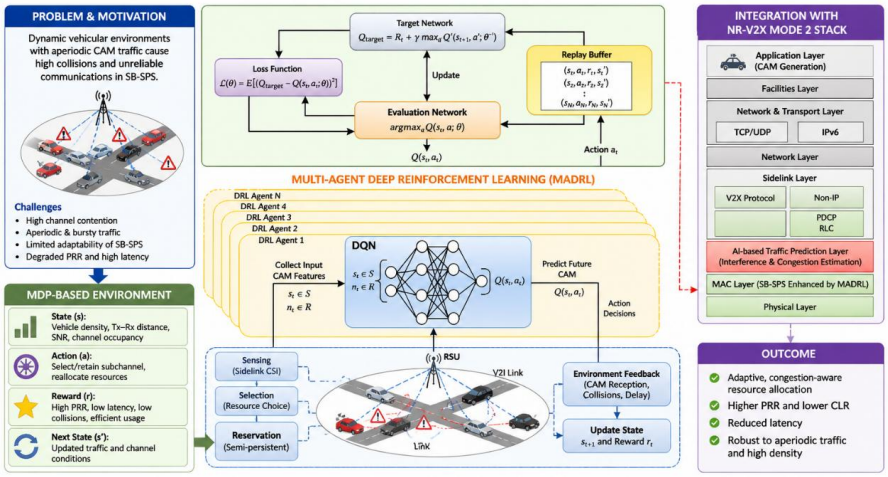


Fig. 1: Proposed MADRL-based DRA architecture for NR-V2X Mode 2 communication, illustrating the interaction between MDP-based environment modeling, multi-agent DQN learning, and integration with the SB-SPS scheduling mechanism.

4.7 Integration with SB-SPS

Instead of replacing the SB-SPS mechanism entirely, the proposed approach enhances the resource selection stage by incorporating MADRL.

Specifically:

- SB-SPS sensing phase is retained.
- Candidate resource pool is generated as in standard SB-SPS.
- MADRL is applied to select the optimal resource within the candidate set.

This hybrid design ensures:

- Compatibility with 3GPP standards.
- Reduced complexity.
- Improved adaptability.

4.8 Algorithm Overview

The overall learning process is summarized as follows:

Algorithm 1: MADRL Training Procedure

1. Initialize replay buffer D
2. Initialize Q-network parameters θ
3. **for** each training episode **do**
4. Observe state s_t
5. Select action a_t using the ε-greedy policy
6. Execute transmission using the selected resource
7. Observe reward R_t and next state s_{t+1}
8. Store transition (s_t, a_t, R_t, s_{t+1}) in D
9. Sample a mini-batch from D
10. Update the Q-network using temporal-difference learning
11. Update the target network periodically
12. **end for**

The proposed MADRL framework enables vehicles to learn adaptive resource allocation strategies that:

- Anticipate channel congestion.
- Avoid resource collisions.
- Respond to dynamic traffic patterns.

By incorporating learning into the SB-SPS framework, the proposed approach improves performance while maintaining standard compliance and distributed operation.

5 Simulation Setup and Evaluation Methodology

5.1 Simulation Framework

To evaluate the performance of the proposed MADRL-assisted resource allocation framework, a custom Python-based NR-V2X Mode 2 simulator is developed. The simulator models distributed sidelink communication at the resource allocation level, focusing on the interaction between vehicles during the scheduling and transmission processes.

Unlike full-stack 5G simulators (e.g., link-level or PHY-accurate engines), the developed simulator abstracts the physical layer and emphasizes MAC-layer resource allocation behavior, which is the primary focus of this study. This abstraction allows efficient modeling of large-scale vehicular scenarios and enables rapid evaluation of learning-based algorithms under dynamic traffic conditions.

The simulator implements:

- NR-V2X Mode 2 DRA,
- SB-SPS scheduling mechanism (baseline).
- MADRL-enhanced resource selection.
- Dynamic vehicular mobility and traffic generation.

This design ensures that the simulation results capture relative performance trends of different resource allocation strategies while maintaining computational efficiency.

5.2 Vehicular Scenario Configuration

A highway scenario is considered to represent a realistic vehicular communication environment with dynamic topology and high mobility.

The key scenario characteristics are summarized as follows:

- **Road type:** Multi-lane bidirectional highway.
- **Number of lanes:** 3 lanes per direction.
- **Road length:** 5 km.
- **Vehicle speed:** 70 km/h (standard deviation: 7 km/h).
- **Vehicle density:** 100, 200, and 300 vehicles/km.

Vehicles are randomly distributed along the road and follow a constant-speed mobility model with slight variations to emulate realistic driving behavior. Each vehicle periodically broadcasts CAMs and may generate additional aperiodic messages based on event-triggered conditions. The communication range is set to 250 m, which is commonly used as the awareness range for vehicular safety applications.

5.3 Communication and Physical Layer Parameters

The simulation parameters are selected based on NR-V2X specifications and commonly used values in literature. The main communication and physical-layer parameters used in the simulation are summarized in Table 1 and 2. These parameters define the sensing, selection, and reselection behavior of the distributed scheduling mechanism and are critical for accurately modeling resource allocation performance under varying traffic and channel conditions.

Table 1: Simulation parameters for NR-V2X MODE 2 DRA communication based on 3GPP TR 38.885.

Parameter	Value
Carrier frequency	5.9 GHz
Bandwidth	10 MHz
Subcarrier spacing (SCS)	15 / 30 / 60 kHz
Subchannel size	12 RB
Modulation scheme	QPSK (MCS 7)
Transmission power	23 dBm
Antenna gain	3 dBi
Noise figure	9 dB

Table 2: SB-SPS configuration parameters for NR-V2X MODE 2 DRA communication based on 3GPP TR 38.885.

Parameter	Value
Sensing window	100 ms
Selection window	1-100 ms
RSRP threshold	−126 dBm
RSSI threshold	−90 dBm
Resource reselection probability P_{keep}	0, 0.4, 0.8

5.4 Traffic Model

The traffic model incorporates both periodic and aperiodic CAMs:

- **Periodic CAMs:** Generated at fixed intervals (e.g., 10 Hz).
- **Aperiodic CAMs:** Triggered by dynamic events such as sudden speed changes or braking.

Two packet sizes are considered:

- 190 bytes (standard CAM size).
- 470 bytes (high-load scenario).

The inclusion of aperiodic traffic is essential to evaluate the robustness of resource allocation strategies under burst traffic conditions, which significantly impact SB-SPS performance.

5.5 MADRL Training Configuration

The training hyperparameters of the proposed MADRL framework are summarized in table 3, each vehicle operates as an independent agent and learns its resource selection policy through interaction with the environment.

Table 3 MADRL training hyperparameters for DRA

Parameter	Value
Discount factor γ	0.99
Learning rate	0.001
Replay buffer size	100,000
Batch size	64
Exploration policy	ε-greedy

ε (initial → final)	1.0 → 0.1
Training episodes	10,000

5.6 Performance Metrics

The performance of the proposed framework is evaluated using the following key metrics:

5.6.1 Packet Reception Ratio (PRR)

$$PRR = \frac{Number\ of\ successfully\ received\ packets}{Total\ transmitted\ packets} \quad (14)$$

PRR measures communication reliability within the awareness range.

5.6.2 Collision Loss Ratio (CLR)

$$CLR = \frac{Number\ of\ collided\ packets}{Total\ transmitted\ packets} \quad (15)$$

CLR quantifies the impact of resource conflicts.

5.6.3 End-to-End Latency

The time delay between packet transmission and successful reception.

5.6.4 Channel Busy Ratio (CBR)

$$CBR = \frac{Occupied\ channel\ time}{Total\ observation\ time} \quad (16)$$

CBR indicates channel utilization and congestion level.

5.7 Simulation Methodology

The evaluation is conducted in two stages:

- **Baseline Evaluation (SB-SPS):** Performance is analyzed under varying numerologies, vehicle densities, and traffic patterns.
- **MADRL Evaluation:** The proposed framework is compared with SB-SPS under identical conditions.

Multiple simulation runs are performed to ensure statistical consistency, and average values are reported for all metrics.

5.8 Validation and Limitations

To ensure the credibility of the simulation results:

- The simulator parameters are aligned with 3GPP specifications.
- Results are compared with established trends in literature.
- A benchmark validation approach (e.g., dataset-based comparison) is used.

However, the following limitations are acknowledged:

- The physical layer is abstracted and does not include full link-level modeling.
- The study focuses on a highway scenario.
- Real-world channel impairments are simplified.

Despite these limitations, the simulator provides a reliable platform for evaluating resource allocation strategies at the MAC level, which is the focus of this work.

6 Results and Discussion

This section presents a comprehensive evaluation of the proposed MADRL-assisted resource allocation framework and compares its performance with the conventional SB-SPS mechanism under dynamic vehicular conditions. The analysis is conducted using the simulation setup described in Section 5, considering varying vehicle densities, transmission distances, and both periodic and aperiodic CAM traffic models. First, the baseline performance of SB-SPS is examined to highlight its limitations in dense and highly dynamic environments. Subsequently, the performance of the proposed MADRL framework is

evaluated in terms of PRR, CLR, latency, and CBR. The results, illustrated in Fig. 2–Fig. 16, demonstrate the effectiveness of the learning-based approach in improving communication reliability, reducing collisions, and enhancing channel utilization. A detailed discussion is provided to explain the observed performance trends and to highlight the advantages of integrating MADRL with the SB-SPS scheduling mechanism.

6.1 Performance of SB-SPS under Dynamic Vehicular Conditions

The performance of the SB-SPS mechanism is first analyzed to establish a reliable baseline for DRA in NR-V2X Mode 2 communication. The evaluation, conducted using the custom Python-based simulator described in Section V, examines the impact of numerology, resource retention probability, traffic patterns, and vehicle density on key performance metrics, including PRR, CLR, latency, and CBR. The results are interpreted as scheduling-level performance trends consistent with 3GPP NR-V2X Mode 2 behavior.

The PRR performance under different numerologies is illustrated in Fig. 2, where a consistent degradation is observed as the transmitter–receiver distance increases. This trend reflects the combined impact of path loss and increased interference due to spatial reuse of sidelink resources. Among the evaluated configurations, higher numerologies (SCS = 60 kHz) achieve better PRR compared to lower numerologies, particularly at medium and long distances. This improvement is mainly attributed to the shorter transmission duration associated with higher SCS, which reduces temporal overlap between transmissions and lowers the probability of collisions. However, the nearly parallel PRR curves indicate that numerology primarily affects the absolute performance level rather than altering the fundamental distance-dependent degradation imposed by SB-SPS.

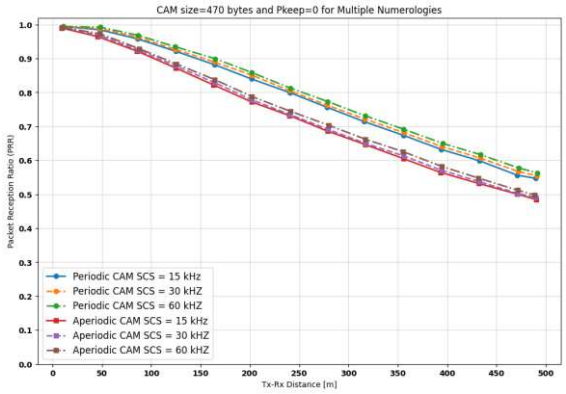
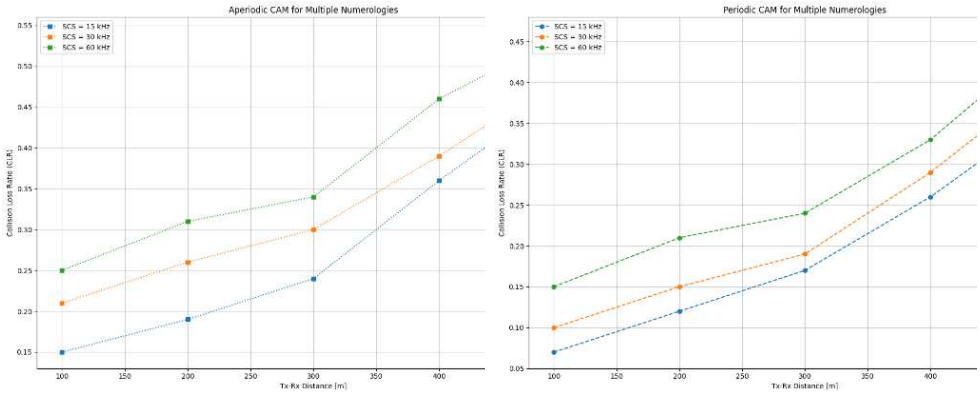


Fig. 2 PRR vs. Tx-Rx distance for periodic and aperiodic CAMs with a packet size of 470 bytes and P_{keep} = 0, evaluated under SCS values of 15, 30, and 60 kHz.

A clear distinction between periodic and aperiodic traffic is also evident in Fig. 2, where aperiodic CAMs consistently exhibit lower PRR, especially at longer distances. This behavior is caused by the irregular transmission patterns of aperiodic traffic, which reduces the effectiveness of sensing-based resource reservation and increases uncertainty in resource selection. In contrast, periodic CAMs benefit from predictable transmission intervals, allowing SB-SPS to exploit historical sensing information more effectively.

The impact of numerology on collision behavior is further illustrated in Fig. 3(a) and Fig. 3(b), which present the CLR for periodic and aperiodic traffic, respectively. In both cases, CLR increases monotonically with distance, indicating a higher probability of packet collisions as the interference region expands. For periodic traffic, higher numerologies result in increased CLR due to more frequent resource usage, which raises the likelihood of overlapping reservations. This trend is more pronounced for aperiodic traffic, where CLR values are consistently higher across all distances. The irregular and event-driven nature of aperiodic CAMs disrupts the sensing process, leading to reduced coordination among vehicles and increased collision probability.



(a) **(b)**

Fig. 3 CLR vs. Tx-Rx distance for periodic and aperiodic CAMs under SB-SPS with SCS values of 15, 30, and 60 kHz.

The divergence between periodic and aperiodic CLR curves beyond medium distances highlights the limited ability of SB-SPS to manage heterogeneous traffic patterns.

The latency performance shown in Fig. 4 reveals an increasing trend with higher numerologies for both traffic types. Although higher SCS reduces transmission duration, it also intensifies scheduling dynamics within SB-SPS, increasing the likelihood of reservation conflicts and deferred transmissions. As a result, packets experience longer waiting times before successful transmission. Aperiodic CAMs consistently exhibit higher latency compared to periodic traffic, as their irregular transmission timing leads to more frequent scheduling conflicts and retransmissions.

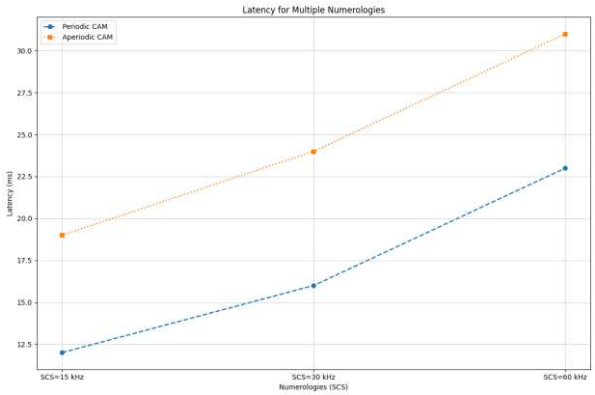


Fig. 4 Average end-to-end latency for periodic and aperiodic CAMs under SB-SPS as a function of SCS (15, 30, and 60 kHz).

The influence of the resource retention probability P_{keep} is analyzed in Fig. 5–Fig. 7. As shown in Fig. 5, increasing P_{keep} improves PRR across all distances, particularly for aperiodic CAMs. This improvement is due to the reduced frequency of resource reselection, which limits exposure to newly occupied resources and stabilizes transmission patterns.

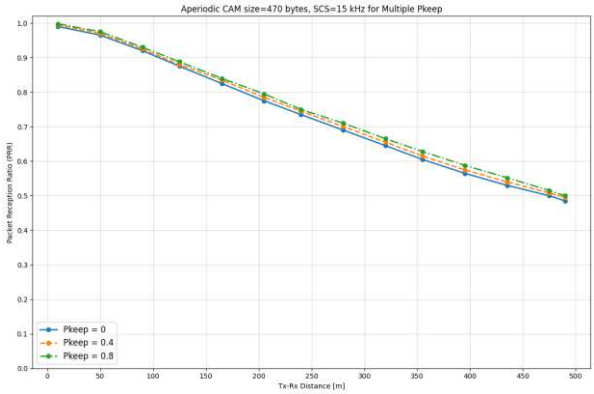
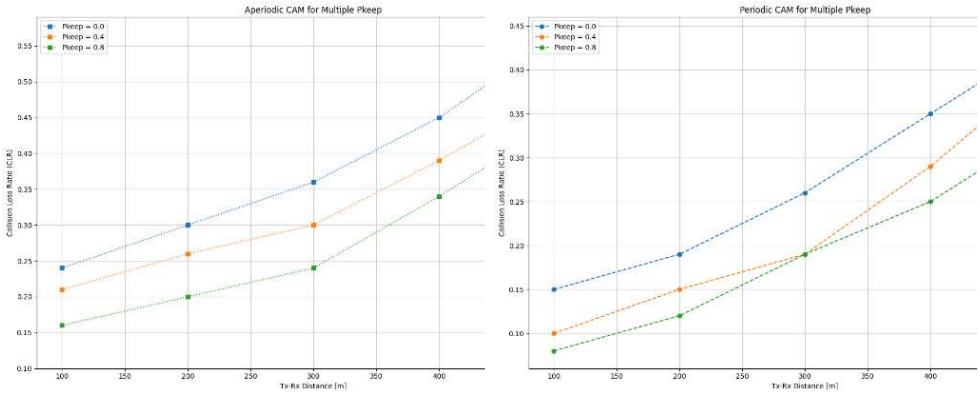


Fig. 5 PRR vs. Tx-Rx distance for aperiodic CAMs under SB-SPS with $P_{keep} \in \{0, 0.4, 0.8\}$, evaluated at SCS = 15 kHz and CAM size = 470 bytes.

Similarly, Fig. 6(a) and Fig. 6(b) demonstrate that higher P_{keep} significantly reduces CLR for both periodic and aperiodic traffic. By retaining previously selected resources, vehicles avoid frequent reselection events that can lead to simultaneous resource conflicts. However, while higher P_{keep} improves short-term stability, it does not eliminate the fundamental increase in collisions with distance and density.



(a) **(b)**

Fig. 6 CLR vs. Tx-Rx distance for periodic and aperiodic CAMs under SB-SPS with $P_{keep} \in \{0, 0.4, 0.8\}$, evaluated at SCS = 15 kHz and CAM size = 470 bytes.

The latency results in Fig. 7 further confirm the stabilizing effect of resource retention. As P_{keep} increases, latency decreases for both traffic types due to reduced reselection overhead and fewer transmission deferrals. Nevertheless, aperiodic CAMs continue to experience higher latency, indicating that resource retention alone cannot fully compensate for traffic irregularity.

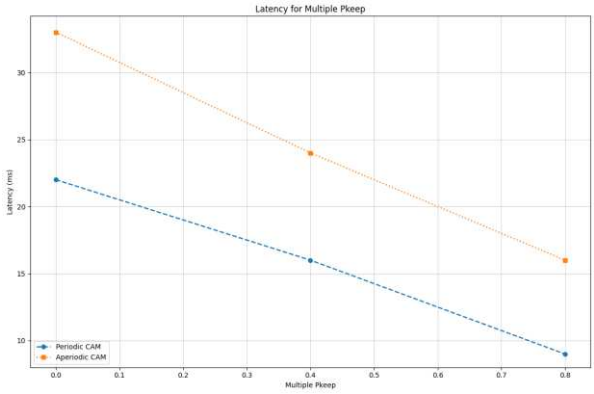


Fig. 7 Average end-to-end latency for periodic and aperiodic CAMs under SB-SPS as a function of P_{keep}, evaluated at SCS = 15 kHz and CAM size = 470 bytes.

Finally, the CBR results shown in Fig. 8 highlight the impact of network density on channel utilization. CBR increases monotonically with vehicle density for both traffic types, with aperiodic CAMs consistently exhibiting higher values. This reflects the more irregular and less coordinated channel access behavior of aperiodic traffic. At high densities, the sharp increase in CBR indicates channel saturation, which directly contributes to the observed degradation in PRR, increase in CLR, and rise in latency.

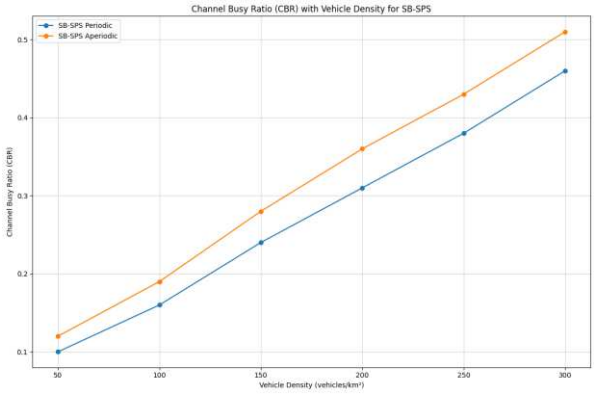


Fig. 8 CBR vs. vehicle density for periodic and aperiodic CAMs under SB-SPS, evaluated at SCS = 15 kHz and CAM size = 470 bytes.

Overall, the results demonstrate that while SB-SPS provides a functional baseline for DRA, its performance is fundamentally limited by its static and sensing-based design. The mechanism struggles to adapt to dynamic traffic conditions, particularly under high density and aperiodic message generation. These limitations motivate the need for adaptive and learning-based approaches, which are explored in the following section.

6.2 MDP-Based Modeling for Dynamic Vehicular Environment

6.2.1 MDP-Based Representation of the Vehicular Environment

Dynamic vehicular environments (DVEs) introduce significant challenges for DRA due to their inherently stochastic and time-varying nature. Variations in vehicle density, mobility patterns, and channel occupancy lead to complex interactions that cannot be effectively captured using static or rule-based models. To address this, the vehicular communication environment is modeled as an MDP, providing a structured framework for representing system dynamics and sequential decision-making under uncertainty.

Within this formulation, the system is described by the tuple (S, A, P, R,), where S denotes the state space, A the action space, P the transition dynamics, and R the reward function. Each state captures the instantaneous network condition using key features including vehicle density, transmitter–receiver distance, and signal-to-noise ratio (SNR), which collectively characterize interference levels and channel quality. The action space represents resource allocation decisions, such as selecting sidelink subchannels and adapting resource usage between periodic and aperiodic CAM transmissions. The reward function evaluates the effectiveness of each action by promoting high communication reliability and low latency while penalizing collisions and inefficient resource usage. State transitions are inherently stochastic and governed by vehicular mobility and dynamic channel conditions, reflecting realistic NR-V2X operation.

6.2.2 Experience Generation for Learning

Based on the MDP formulation, an experience dataset is generated through simulation of the vehicular environment using the custom Python-based NR-V2X Mode 2 framework. Each interaction between the environment and the decision-making entity is represented as a tuple (s_t, a_t, r_t, s_{t+1}), capturing the system state, selected action, obtained reward, and resulting next state. This dataset encapsulates the temporal evolution of resource allocation decisions under varying traffic and channel conditions and serves as the foundation for training the subsequent MADRL framework.

The dataset includes diverse traffic scenarios, ranging from low-density conditions with minimal interference to highly congested environments characterized by frequent collisions. This diversity ensures that the learning agents are exposed to a wide range of operating conditions, enabling them to develop adaptive and generalizable resource allocation strategies.

6.2.3 Dataset Validation

To ensure the reliability of the generated dataset, both structural consistency and statistical properties are analyzed. The average reward distribution as a function of transmitter–receiver distance and channel load is illustrated in Fig. 9. The figure shows a smooth and monotonic decrease in reward as distance and channel load increase, reflecting expected NR-V2X behavior where longer communication ranges and higher congestion lead to increased interference and reduced transmission success. This structured reward surface confirms that the MDP formulation effectively captures the dominant dynamics of the vehicular environment.

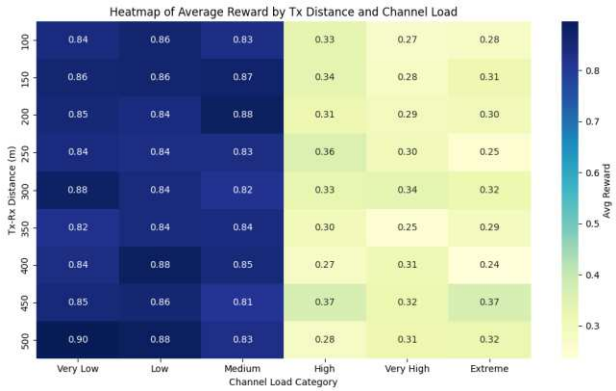


Fig. 9 Heatmap of average reward as a function of Tx-Rx distance and channel-load level, generated by the MDP environment.

Further validation is provided by the reward distribution histogram shown in Fig. 10, which exhibits a well-balanced and unimodal distribution spanning both low and high reward values. This indicates that the dataset is free from reward sparsity and adequately represents diverse system conditions. The presence of a broad reward range ensures that the learning process is not biased toward specific states and supports stable training of reinforcement learning agents.

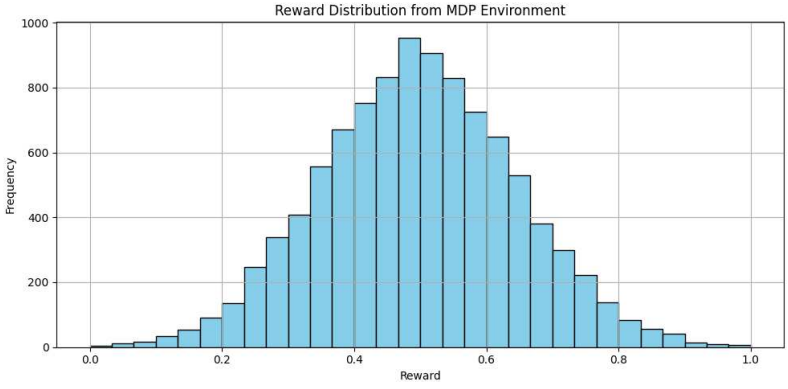


Fig. 10 Histogram of reward values generated by the MDP environment across all sampled experiences, illustrating reward spread and stability for dataset validation.

The MDP-based modeling framework provides a coherent and physically consistent representation of dynamic vehicular communication environments. By capturing the relationships between traffic conditions, channel characteristics, and communication performance, the generated dataset enables effective learning of adaptive resource allocation policies. The validation results confirm that the dataset accurately reflects realistic NR-V2X behavior and is suitable for training the proposed MADRL framework. This modeling

approach forms the foundation for the learning-based optimization of DRA, as presented in the following section.

6.3 Performance of the Proposed MADRL Framework

The performance of the proposed MADRL framework is evaluated using the same simulation environment and conditions as the SB-SPS baseline to ensure a fair and consistent comparison. The results presented in this section are obtained by integrating multiple learning agents into the custom Python-based NR-V2X Mode 2 simulator, where the agents optimize resource selection decisions at the scheduling layer while physical-layer effects remain abstracted. Therefore, the observed performance gains reflect improvements in distributed scheduling behavior rather than enhancements at the physical transmission level.

6.3.1 Packet Reception Ratio (PRR)

The PRR performance of the proposed MADRL framework is illustrated in Fig. 11(a) and Fig. 11(b) as a function of transmitter–receiver distance for aperiodic CAM transmissions with packet sizes of 190 bytes and 470 bytes, respectively. Across all schemes, PRR decreases monotonically with increasing distance due to the combined effects of path loss, expanding interference regions, and increased collision probability. This consistent trend confirms that all evaluated approaches operate under the same underlying system constraints.

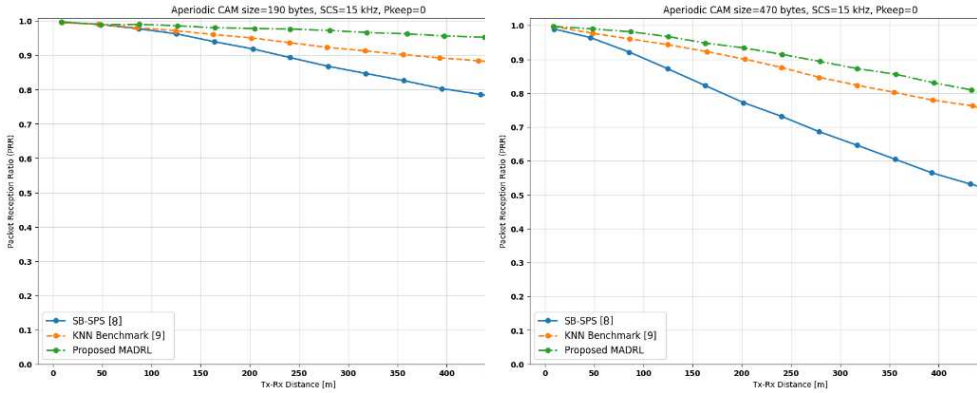
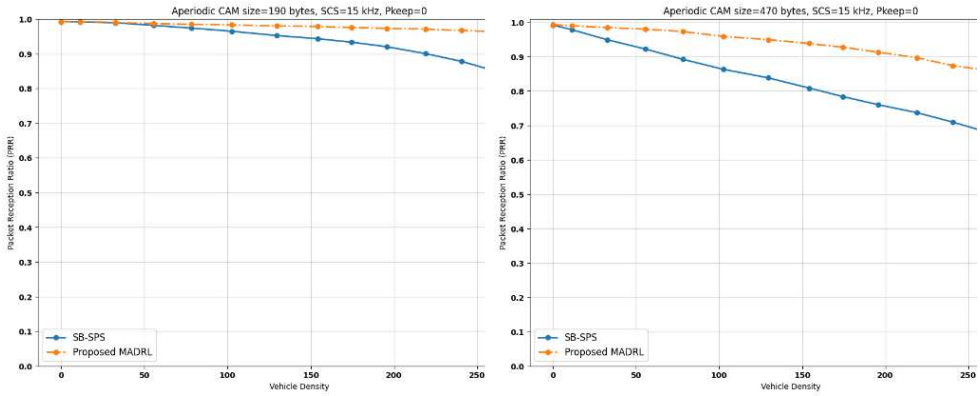


Fig. 11 PRR versus Tx-Rx distance for aperiodic CAMs under SB-SPS [8], the KNN benchmark [9], and the proposed MADRL approach, evaluated at SCS = 15 kHz and P_{keep} = 0.

Despite this common degradation pattern, the proposed MADRL framework consistently outperforms both SB-SPS and the KNN-based benchmark across all distances. The performance gain becomes particularly evident at medium and long packets, MADRL maintains a higher PRR throughout the distance range, while for 470-byte packets, the improvement is even more pronounced, as shown in Fig. 11(b). This behavior indicates that the learning agents effectively capture congestion patterns and adapt resource selection accordingly, enabling more reliable communication under challenging conditions.

The superiority of MADRL becomes further evident when comparing PRR as a function of vehicle density, as shown in Fig. 12(a) and Fig. 12(b). In both cases, PRR decreases with increasing density due to higher channel contention and interference. However, while the SB-SPS baseline exhibits a rapid degradation in PRR at higher densities, the MADRL framework maintains a more gradual decline, demonstrating its ability to mitigate congestion effects. This improvement is particularly significant for 470-byte packets, where larger transmission durations increase sensitivity to collisions. The results confirm that adaptive learning-based scheduling provides substantial benefits in dense vehicular environments.



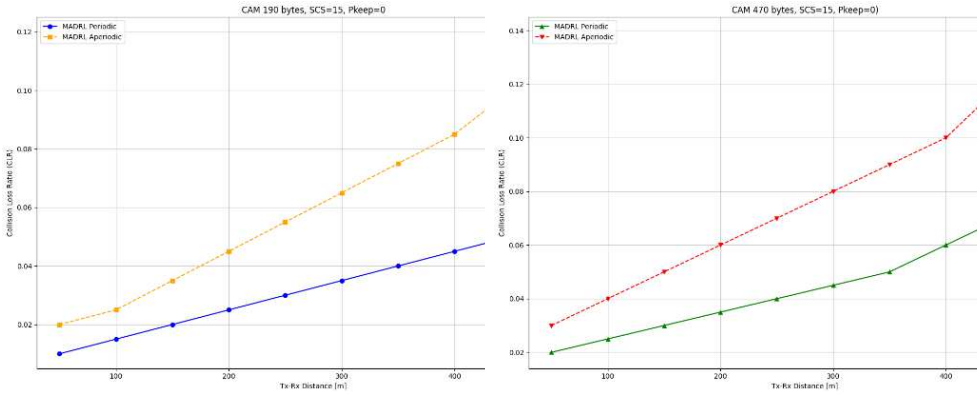
(a) **(b)**

Fig. 12 PRR vs. vehicle density for aperiodic CAMs, comparing SB-SPS and the proposed MADRL approach at SCS = 15 kHz and $P_{keep} = 0$.

6.3.2 Collision Loss Ratio (CLR)

The CLR performance of the MADRL framework is presented in Fig. 13(a) and Fig. 13(b) for packet sizes of 190 bytes and 470 bytes, respectively. In both cases, CLR increases with transmitter– receiver distance due to the expansion of the interference region and the increased likelihood of overlapping transmissions. This trend is consistent for both periodic and aperiodic traffic, confirming that the MADRL framework operates under realistic interference constraints.

Aperiodic CAMs consistently exhibit higher CLR than periodic CAMs, reflecting their irregular transmission patterns and reduced predictability. However, the overall CLR values remain well-controlled, indicating that the MADRL agents successfully learn collision-aware scheduling strategies. Compared to the SB-SPS baseline discussed in Section VI-A, the MADRL framework significantly reduces repeated resource conflicts by enabling agents to anticipate congestion and avoid heavily utilized subchannels.



(a) **(b)**

Fig. 13 CLR vs. Tx-Rx distance for periodic and aperiodic CAMs under the proposed MADRL resource allocation scheme at SCS = 15 kHz and $P_{keep} = 0$.

The impact of packet size is also evident, as larger packets (470 bytes) result in higher CLR due to longer channel occupancy. Nevertheless, the MADRL framework maintains a stable and gradual increase in CLR, avoiding abrupt performance degradation. This indicates that the learned policies are robust and do not introduce aggressive or unstable scheduling behavior.

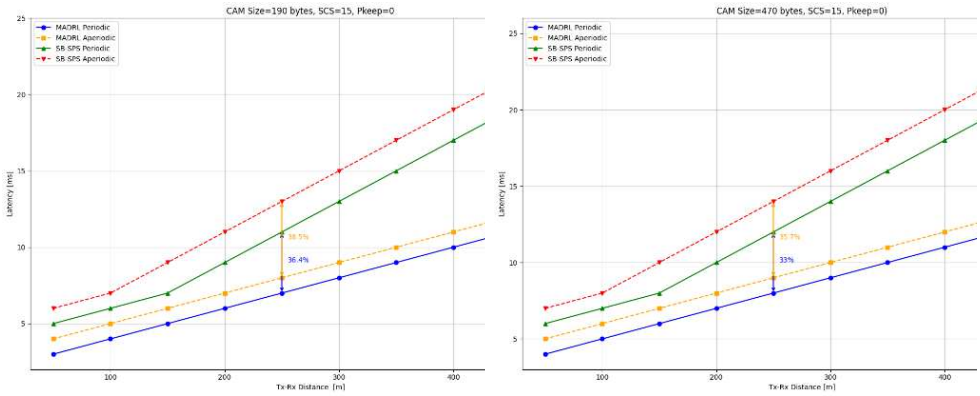
6.3.3 Latency Performance

The latency performance of the proposed framework is shown in Fig. 14(a) and Fig. 14(b), where the average end-to-end delay is plotted as a function of transmitter–receiver distance for both packet sizes. For all configurations, latency increases with distance due to higher

collision probability and increased scheduling delays. This behavior is consistent across both SB-SPS and MADRL, confirming that both operate under the same network constraints.

However, the MADRL framework consistently achieves lower latency compared to SB-SPS across all distances and traffic types. The improvement is particularly noticeable at medium and long distances, where channel congestion is more pronounced. This latency reduction is primarily attributed to the improved resource allocation decisions made by the learning agents, which reduce retransmissions and scheduling deferrals.

Aperiodic CAMs exhibit higher latency than periodic CAMs due to their irregular transmission patterns. Nevertheless, the latency gap between periodic and aperiodic traffic is reduced under MADRL, indicating that the learning-based approach effectively accommodates traffic variability. For representative scenarios, latency reductions of approximately 33–38% are observed, demonstrating the effectiveness of adaptive scheduling in improving time-critical communication performance.



(a) (b)

Fig. 14 Average end-to-end latency versus Tx-Rx distance for periodic and aperiodic CAMs, comparing SB-SPS and the proposed MADRL approach for 190-byte and 470-byte packets.

6.3.4 Channel Busy Ratio (CBR)

The CBR results are presented in Fig. 15, illustrating channel utilization as a function of vehicle density for both packet sizes and traffic types. As expected, CBR increases monotonically with vehicle density due to higher channel occupancy. Aperiodic traffic consistently results in higher CBR than periodic traffic, reflecting its irregular and less predictable transmission behavior.

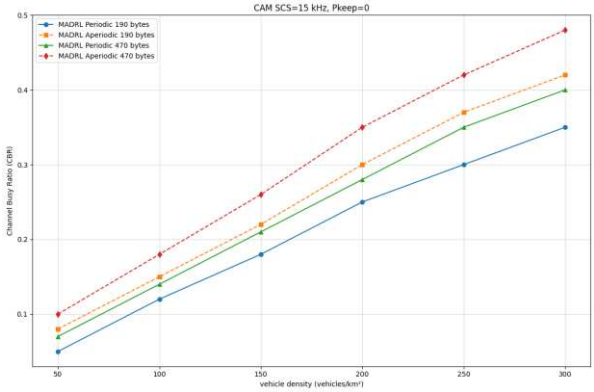


Fig. 15 CBR vs. vehicle density for periodic and aperiodic CAMs with packet sizes of 190 and 470 bytes under the proposed MADRL framework.

Despite the increase in channel load, the MADRL framework maintains a well-controlled and gradual growth in CBR across all densities. The absence of abrupt saturation indicates that the learning agents effectively distribute transmissions across available resources, preventing excessive congestion. When considered alongside the improvements in PRR,

CLR, and latency, these results confirm that the performance gains of MADRL are achieved through more efficient resource utilization rather than reduced channel usage.

6.3.5 Reward Convergence and Learning Stability

The convergence behavior of the MADRL framework is illustrated in reward over 10,000 episodes. During the initial training phase, significant fluctuations are observed due to exploration, as agents interact with the environment to gather diverse experiences. As training progresses, the reward steadily increases and the variance decreases, indicating a transition toward stable policy learning.

Beyond the mid-training phase, the reward converges to a high and stable value, with a maximum reward of approximately 94.55%. This convergence demonstrates that the agents successfully learn effective resource allocation strategies. Occasional reward fluctuations are observed due to the stochastic nature of the vehicular environment; however, these do not affect overall stability, confirming the robustness of the learned policies.

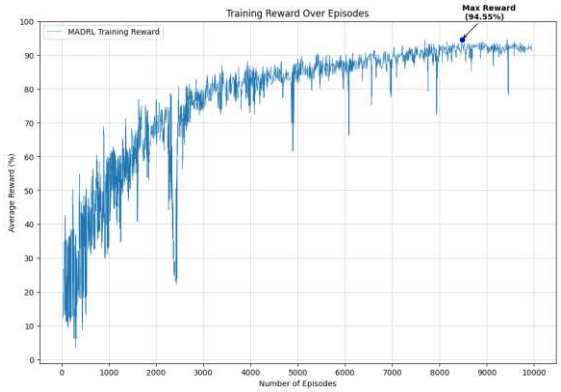


Fig. 16 Average training reward versus the number of training episodes for the proposed MADRL framework, illustrating learning convergence and stability.

The results clearly demonstrate that the proposed MADRL framework significantly enhances DRA performance in NR-V2X Mode 2 communication. By incorporating adaptive and learning-based decision-making, the framework enables vehicles to anticipate channel conditions, avoid resource conflicts, and optimize scheduling decisions.

The improvements observed across PRR, CLR, latency, and CBR confirm that MADRL effectively addresses the limitations of SB-SPS, particularly under high-density and aperiodic traffic conditions. Importantly, these gains are achieved without reducing channel utilization, indicating that the improvement stems from more efficient resource allocation rather than reduced traffic load.

It is important to note that the results are obtained within a system-level simulation framework with abstracted physical-layer modeling. Therefore, the reported performance gains should be interpreted as relative improvements at the scheduling level rather than absolute performance guarantees for real-world deployments.

7 Conclusion

This paper investigated the problem of DRA in 5G NR-V2X Mode 2 communication under dynamic vehicular environments, with a particular focus on the challenges introduced by aperiodic CAM traffic. The limitations of the standardized SB-SPS mechanism were first analyzed, revealing its reduced effectiveness in handling bursty traffic patterns, high vehicle densities, and rapidly changing channel conditions. To address these limitations, an MADRL-based framework was proposed, in which each vehicle operates as an autonomous agent and learns adaptive resource selection policies through interaction with the environment.

Table 4: Performance comparison between SB-SPS and MADRL at the 250 m awareness range.

Metric	SB-SPS	MADRL	Relative Observation at 250 m
PRR (190 bytes)	~0.72	~0.82	MADRL improves PRR through congestion-aware scheduling.
PRR (470 bytes)	~0.60	~0.75	The gain is larger because longer packets are more collision-sensitive.
CLR (190 bytes)	~0.18	~0.10	MADRL reduces collision probability through adaptive resource selection.
CLR (470 bytes)	~0.30	~0.18	MADRL mitigates collision escalation at the awareness range.
Latency (190 bytes)	~18 ms	~12 ms	MADRL reduces scheduling and reselection delays.
Latency (470 bytes)	~26 ms	~16 ms	A stronger latency reduction is obtained for larger packets.
CBR	~0.62	~0.60	Channel occupancy is comparable; MADRL does not artificially reduce CBR.
Stability	Semi-static	Adaptive	MADRL adapts to traffic dynamics at 250 m.
Traffic sensitivity	High (aperiodic)	Lower	MADRL handles bursty traffic more effectively.

The vehicular communication system was modeled as an MDP, enabling structured representation of dynamic network conditions and supporting data-driven learning. A Deep Q-Network (DQN) was integrated into the SB-SPS resource selection stage to enhance decision-making while maintaining compatibility with the distributed scheduling paradigm. The proposed approach was evaluated using a custom Python-based NR-V2X Mode 2 simulator under realistic highway scenarios, considering varying vehicle densities, packet sizes, and both periodic and aperiodic traffic models.

Table 4 demonstrated the results that the proposed MADRL framework significantly improves communication performance compared to the SB-SPS baseline. In particular, the PRR is improved by approximately 8% for 190-byte CAMs and up to 20% for 470-byte CAMs under aperiodic traffic conditions. These gains are accompanied by a substantial reduction in CLR and noticeable improvements in latency, especially in dense and high-load scenarios. Importantly, these performance enhancements are achieved without reducing channel utilization, as indicated by comparable CBR values, confirming that the improvements stem from more efficient and adaptive resource allocation rather than reduced network load.

Despite these promising results, it is important to note that the evaluation is conducted using a system-level simulator with an abstracted physical layer. Consequently, the reported performance gains reflect improvements at the scheduling and resource allocation level rather than absolute performance guarantees for real-world deployments. Additionally, the study focuses on highway scenarios, and further investigation is required to assess performance in more complex urban environments.

Future work will focus on extending the proposed framework to incorporate more detailed physical-layer modeling, including realistic channel propagation effects and advanced interference mechanisms. Moreover, exploring more advanced multi-agent learning techniques, such as cooperative or federated reinforcement learning, may further enhance scalability and coordination among vehicles. Finally, real-world validation using hardware testbeds or standardized simulation platforms represents an important direction toward practical deployment.

Funding Declaration

The authors declare that no external funding was received for this study.

Competing Interests Statement

The authors declare that they have no competing financial or personal interests that could have influenced the results of this research.

References

[1] H. Hartenstein and K. P. Laberteaux, "A tutorial survey on vehicular ad hoc networks," IEEE Communications Magazine, vol. 46, no. 6, pp. 164–171, 2008, doi: 10.1109/MCOM.2008.4539481.

[2] J. B. Kenney, "Dedicated short-range communications (DSRC) standards in the United States," Proceedings of the IEEE, vol. 99, no. 7, pp. 1162–1182, Jul. 2011, doi: 10.1109/JPROC.2011.2132790.

[3] G. Naik, B. Choudhury, and J. M. Park, "IEEE 802.11bd and 5G NR V2X: Evolution of radio access technologies for V2X communications," IEEE Access, vol. 7, pp. 70169–70184, 2019, doi: 10.1109/ACCESS.2019.2919489.

[4] 3GPP, "Study on NR vehicle-to-everything (V2X), Release 16," 3rd Generation Partnership Project, Technical Report 38.885, Version 16.0.0, Mar. 2019.

[5] M. M. Saad, M. A. Tariq, A. Siddiqa, and D. Kim, "Analytical models for distributed resource allocation in NR-V2X Mode 2," IEEE Internet of Things Journal, pp. 1–15, 2023, doi: 10.1109/JIOT.2023.3328885.

[6] ETSI, "Intelligent transport systems (ITS); Vehicular communications; Basic set of applications; Part 2: Specification of cooperative awareness basic service," ETSI Specification, Version 1.4.1, pp. 1–22, 2019.

[7] H. Seon, H. Lee, and H. Kim, "Predicting CAM generation times through machine learning for cellular V2X communication," ICT Express, vol. 9, no. 5, pp. 841–846, 2023, doi: 10.1016/j.icte.2022.08.006.

[8] 3GPP Technical Specification Group Radio Access Network, "NR; Medium Access Control protocol specification, Release 19," 3rd Generation Partnership Project, 2022.

[9] L. Lusvarghi, A. Molina-Galan, B. Coll-Perales, J. Gozalvez, and M. L. Merani, "A comparative analysis of the semi-persistent and dynamic scheduling schemes in NR-V2X Mode 2," Vehicular Communications, vol. 42, 2023, doi: 10.1016/j.vehcom.2023.100628.

[10] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[11] H. Ye, G. Y. Li, and B. H. F. Juang, "Deep reinforcement learning based resource allocation for V2V communications," IEEE Transactions on Vehicular Technology, vol. 68, no. 4, pp. 3163–3173, Apr. 2019, doi: 10.1109/TVT.2019.2897134.

[12] K. Zhang, Z. Yang, and T. Ba¸sar, "Multi-agent reinforcement learning: A selective overview of theories and algorithms," Studies in Systems, Decision and Control, vol. 325, pp. 321–384, 2021, doi: 10.1007/978-3030-60990-0-12.

[13] M. Sun, J. Xu, and J. Wang, "State-aware resource allocation for V2X communications," Sensors, vol. 26, no. 1, pp. 1–25, 2026, doi: 10.3390/s26010344.

[14] A. Nair and S. Tanwar, "Resource allocation in V2X communication: State-of-the-art and research challenges," Physical Communication, vol. 64, Art. no. 102351, 2024, doi: 10.1016/j.phycom.2024.102351.

[15] M. Tabassum and A. Oliveira, "Collision probabilities between user equipment using 5G NR sidelink time-domain-based resource allocation in C-V2X," Electronics, vol. 14, no. 4, 2025, doi: 10.3390/electronics14040751.

[16] S. Nikooroo, J. Estrada-Jimenez, A. Machalek, J. Harri, T. Engel, and I. Turcanu, "Mitigating collisions in sidelink NR V2X: A study on cooperative resource allocation," in Proc. 22nd Mediterranean Communication and Computer Networking Conference (MedComNet), 2024, doi: 10.1109/MedComNet62012.2024.10578242.

[17] J. Zhao et al., “Boosting vehicular connectivity through resource allocation algorithm based on heterogeneous agent proximal policy optimization,” Vehicular Communications, vol. 50, Art. no. 100856, 2024, doi: 10.1016/j.vehcom.2024.100856.

[18] Z. Li, G. Li, Z. Wu, W. Zhang, and A. Bazzi, “Dynamic allocation of C-V2X communication resources based on graph attention network and deep reinforcement learning,” Sensors, vol. 25, no. 16, pp. 1–26, 2025, doi: 10.3390/s25165209.

[19] Z. Li, P. Wang, Y. Shen, and S. Li, “Reinforcement learning-based resource allocation scheme of NR-V2X sidelink for joint communication and sensing,” Sensors, vol. 25, no. 2, 2025, doi: 10.3390/s25020302.