IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 December 2023, accepted 24 December 2023, date of publication 2 January 2024,
date of current version 9 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3349350

RESEARCH ARTICLE

Decentralized Multi-Agent DQN-Based Resource Allocation for Heterogeneous Traffic in V2X Communications

INSUNG LEE AND DUK KYUNG KIM, (Member, IEEE)
Department of Information and Communication Engineering, Inha University, Incheon 22212, South Korea

Corresponding author: Duk Kyung Kim (kdk@inha.ac.kr)

This work was supported by IITP Grant funded by the Korea Government (MSIT) (Development of 5G-NR-V2X communication technology to support over Lv.4 autonomous driving service) under Grant 2022-0-00521.

ABSTRACT Vehicle-to-everything (V2X) communication is a pivotal technology for advanced driving, encompassing autonomous driving and Intelligent Transportation Systems (ITS). Beyond direct vehicle-to-vehicle (V2V) communication, vehicle-to-infrastructure (V2I) communication via Road Side Unit (RSU) can play an important role for efficient traffic management and enhancement of advanced driving, providing surrounding vehicles with proper road information. To accommodate diverse V2X scenarios, heterogeneous traffic with varied objectives, formats, and sizes needs to be supported for V2X communication. We tackle the challenge of resource allocation for heterogeneous traffic in the RSU-deployed V2X communications, proposing a decentralized Multi-Agent Reinforcement Learning (MARL) based resource allocation scheme with limited shared resources. To reduce the model complexity, RSU is modeled as a collection of virtual agents with a small action space instead of a single agent selecting multiple resources at the same time. A weighted global reward is introduced to incorporate traffic heterogeneity efficiently. The performance is evaluated and compared with random, 5G NR mode 2, and optimal allocation schemes in terms of Packet Reception Ratio (PRR) and communication range. The proposed scheme nearly matches the performance of the optimal scheme and significantly outperforms the random allocation scheme in both underload and overload situations.

INDEX TERMS Vehicle-to-everything (V2X), resource allocation, heterogeneous traffic, decentralized multi-agent DQN.

I. INTRODUCTION

Autonomous driving is recognized as a solution for enhancing road safety, traffic efficiency, and environmental sustainability [1]. Modern autonomous vehicles are equipped with advanced sensors like cameras, radar/LiDAR, and Advanced Driver Assistance Systems (ADAS). These sensors are pivotal in enabling autonomous driving functions and ensuring on-road safety. Nonetheless, they can be influenced by adverse weather conditions, poor visibility, and obstructed sightlines [2]. Thus, to foster information exchange and coordination among road users and traffic managers, the European Commission (EU) introduced the Cooperative Intelligent Transport Systems (C-ITS) [3] which has played a key role in deploying diverse use cases and introducing advanced mechanisms to elevate transportation systems. Among these advancements, Cooperative Vehicle-to-Everything (C-V2X) technology emerges as cutting-edge, encompassing vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), vehicle-to-pedestrian (V2P), and vehicle-to-network (V2N). In particular, 5G New Radio (NR)-based V2X, introduced in the 3rd Generation Partnership Project (3GPP) Rel. 16, offers wider coverage and enhanced Quality of Service (QoS) compared to its predecessors.

V2X communications encompass diverse communication casting types: unicast, groupcast, and broadcast. These types

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Quan.

cater to specific communication goals and settings. Unicast and groupcast are utilized locally to enhance road safety in specific scenarios. In contrast, broadcast serves broader regions for comprehensive road safety. Our focus is on enhancing road safety through V2X broadcast communication. Catering to diverse scenarios entails sending messages with specific formats and sizes that match their objectives. Incorporating a heterogeneous traffic model becomes crucial for achieving authentic V2X communications. The prevalent channel model employed in V2X broadcast communication is outlined in the 3GPP document [4]. Furthermore, investigations into channel models for 6G and tera-Hz bands have been conducted in [5] and [6], with the incorporation of MIMO discussed in [7].

3GPP introduces two resource allocation modes for 5G NR: mode 1 and mode 2 [8]. In mode 1, the base station (BS) directly assigns wireless resources to user equipment (UE) for sidelink (SL) via the Uu interface. A UE sends a scheduling request (SR) to the BS using Physical Uplink Control Channel (PUCCH), and the BS specifies the resource through the Physical Downlink Control Channel (PDCCH). Conversely, in 5G NR mode 2, UEs autonomously select SL resources using sensed Reference Signals Received Power (RSRPs) and the information from the 1st Sidelink Channel Information (SCI) of other UEs. UEs sort unoccupied and unreserved resources, randomly select the required amount of resources, and reuse them repeatedly for a defined number of cycles in a semi-persistent fashion. Despite their merits, both modes have their own limitations. Mode 1 is geographically constrained by the BS's coverage, limiting its scope. In 5G NR mode 2, the performance suffers from random resource selection, uncertainty, and potential interference. Thus, an appropriate resource allocation scheme is imperative for V2X environments.

The Roadside Unit (RSU) serves as a crucial infrastructure, amplifying communication performance through extended coverage, blind-spot detection, and multi-hop relay capabilities. By directly broadcasting road information to nearby vehicles, RSUs notably enhance road safety. RSUs also function as multi-hop relays, as studied in [9], [10], and [11], ushering in opportunities for improved communication, expanded coverage, and heightened performance in V2X systems. The concept of Smart RSUs, equipped with diverse sensors like LiDAR and cameras, is actively explored to gather comprehensive road data [12]. Additionally, Smart RSUs can offload tasks and leverage Multi-Access Edge Computing (MEC) technology, easing user computational loads and enhancing their overall experience [13], [14].

Reinforcement learning (RL) offers a key advantage in addressing intricate tasks and requirements that conventional mathematical models struggle with. An adeptly crafted RL model can account for diverse factors and optimize actions accordingly, proving flexible in dynamic and uncertain settings. This adaptability suits problems with multiple objectives, crucial in rapidly changing V2X environments where traditional models falter. Particularly in decentralized multi-agent (MA) RL, each agent (e.g., vehicles or RSUs) independently acts based on local observations, mirroring real-world resource allocation behavior in V2X settings.

This paper introduces an RL-driven resource allocation strategy for a mixed RSU-vehicle scenario. In this approach, both RSU and vehicles transmit packets of varying sizes using a shared resource pool. The framework employs a decentralized Multi-Agent Deep Q-Network (DQN) model. Each agent autonomously makes decisions based solely on locally observed environment states, without sharing RSRP values, thus minimizing resource consumption and exchange delays. A shared global reward, informed by priority weights, guides MARL model training. This encourages agents to act cooperatively, optimizing not just individual performance but also supporting diverse traffic efficiently.

The main contributions of this paper are as follows:

- In addressing diverse V2X traffic with varying packet sizes, we initially compute individual rewards for V2X entities. Subsequently, priority weights are applied to generate a weighted global reward. Through modulation of these weights for messages of differing sizes, we enhance the Packet Reception Ratio (PRR) performance for larger messages, despite their resource-intensive nature in comparison to smaller messages.
- Our approach accommodates both overload and underload situations within the shared resource pool. We showcase the efficacy of our resource allocation scheme across diverse channel congestion levels. This ensures efficient resource utilization for adequate communication range. The adaptability to fluctuating resource demand and availability renders our approach well-suited for real-world V2X settings marked by dynamic resource demands.
- Our proposal involves deploying an RSU with multiple virtual agents, each selecting a single resource. This contrasts with a single agent choosing multiple resources simultaneously in the MA-DQN model. This implementation minimizes the action space size and enhances learning efficiency compared to traditional approaches.
- In order to thoroughly examine the performance of our proposed scheme, we carry out an extensive simulation campaign and conduct a comparative analysis against random, 5G NR mode 2, and optimal resource allocation schemes. Our MARL-based scheme demonstrates remarkable superiority over the random and 5G NR mode 2 schemes, and closely approaches the performance of the optimal scheme.
- The evaluation of performance involves the consideration of two distinct sizes for Decentralized Environmental Notification Message (DENM), aiming to underscore the influence of larger messages on PRR performance. Additionally, an in-depth analysis of the performance decline associated with larger messages is conducted, delving into collision probability and SINR distribution.

TABLE 1. List of symbols.

Symbol	Definition
N	Number of all the vehicles
K	Number of transmitters
D	Number of TBs for DENM message transmission
N_K	Number of receiving vehicles for the k -th transmitter
α	Large-scale fading component
$\tilde{g}$	Small-scale fading component
$h_{k,j}$	Channel gain between the k-th transmitter and the j-th receiver
$\gamma_{k,j}$	SINR at the j -th receiver from the k -th transmitter
$q_k[m]$	Received signal strength on the m-th TB at the k-th transmitter
$\boldsymbol{s_k}$	State of agent k
a_k^s	Action of agent k at state s
$C_{k,j}^{(f)}$	Channel capacity of the j-th receiver from the k-th transmitter when the reception is failed
$N_k^{(f)}$	Total number of receiving vehicles that experience reception failures for the k -th transmitter
D_k	Penalty component
R_k	Individual reward of agent k
R_G	Global reward

- The computational complexities of both the optimal and proposed MA-DQN schemes are assessed in terms of the number of multiplications. Notably, the proposed scheme demonstrates markedly lower execution complexity, albeit with an associated increase in complexity during the training phase.

The rest of the paper is organized as follows. Section II presents the related works. Section III introduces the system model and problem formulation. Section IV describes the proposed MARL model and Section V provides the simulation results. Finally, Section VI draws conclusions.

II. RELATED WORKS

A. 5G NR MODE 2 RESOURCE ALLOCATION SCHEME

The resource structure within 5G NR is comprised of two primary elements: frequency and time. In the realm of sidelink communication, the smallest unit in the time domain is referred to as a ‘slot,’ encompassing 14 OFDM symbols. The duration of a slot varies based on the Subcarrier Spacing (SCS). For example, with an SCS of 15k Hz, a slot lasts for 1 ms. As the SCS increases to 30 kHz and 60 kHz, the slot's duration shortens to 0.25 ms and 0.125 ms, respectively. Typically, the default SCS is set at 15 kHz, with higher values reserved for applications necessitating low latency. In the frequency domain, 12 sub-carriers with identical numerology constitute a Resource Block (RB). The bandwidth of this RB is dictated by the SCS. N_{RB} RBs can be consecutively aggregated to form a single subchannel, where $N_{RB} \in \{10,12,15,20,25,50,75,100\}$. Ultimately, a Transport Block (TB) can be composed of 1 slot × 1 subchannel. This comprehensive structure ensures the efficient allocation and utilization of resources in 5G NR communications.

In Release 16, 3GPP introduced an autonomous resource allocation scheme for 5G NR sidelink, referred to as 5G NR mode 2. Under this scheme, each UE is equipped with its own sensing window, selection window, and Reselection Counter (RC). The sensing window has a predefined size ranging from 100 to 1100 ms. During this period, a UE actively scans wireless channels, monitoring the resource usage by neighboring UEs. The selection window, on the other hand, spans from T_{min} to the Packet Delay Budget (PDB). Here, T_{min}, a minimum duration determined by packet priority and the numerology coefficient μ, takes values from the set $\{1, 5, 10, 20\} \cdot 2^{\mu}$. The size of the selection window is constrained by the latency deadline, PDB. Within this window, specific resources are allocated based on information gathered during the earlier sensing window. The RC is a randomly chosen integer ranging from 5 to 15. Notably, the UE consecutively utilizes the same allocated resource for a number of RC times, providing a degree of resource stability.

The operational process unfolds in the following manner: A UE employs the allocated resource for packet transmission, decrementing its RC with each transmission. When the RC reaches 1, the UE faces a decision point. It assesses whether to persist with the current resource or allocate a new one, guided by a predetermined probability, P_{keep}. If the decision is to maintain continuity with the current resource, the RC value is duly updated. Alternatively, if the UE opts for a resource change, it embarks on a selection process. Available resources are scrutinized, filtering out those not in use by other UEs and possessing a RSRP lower than a predetermined threshold. Furthermore, the number of available resources should meet at least $X\%$ of the total resources within the selection window, where X takes values from the set $\{20, 35, 50\}$. If this $X\%$ criterion is unmet, the UE iterates the resource sorting process by incrementing the RSRP threshold by 3 dB. This iterative refinement continues until the $X\%$ requirement is satisfied. Eventually, when the available resources are identified, the UE randomly selects one from the pool while updating its RC, completing the resource allocation process. This comprehensive framework in 5G NR mode 2 enhances autonomous resource management, optimizing communication efficiency for UEs.

B. RESOURCE ALLOCATION USING RL IN V2X COMMUNICATIONS

Recently, numerous studies have proposed various RL-based resource allocation schemes for V2X communications [15],

[16], [17], [18], [19], [20], [21], [22]. These studies have harnessed distinct RL models to pursue varied objectives with different information availability in diverse V2X settings. In [15], the authors utilized an MA-DQN model for optimal resource allocation and discrete power level assignment for V2V links in an intersection environment. In this context, V2V and V2I links share the spectrum resource, aiming to maximize the V2I link channel capacity while satisfying V2V link latency and reliability constraints. However, this approach necessitated an extensive information set including channel information for all the V2I and V2V links, allocation details of neighboring vehicles, interference power on resources, and remaining packet transmission time constraints for V2V links. In [16] and [17], agent decision encompassed mode selection alongside resource and power allocation in underlay situations. For instance, a decentralized MA actor-critic (AC) model deftly handled continuous-valued channel information and transmit power [16], while a federated Deep Reinforcement Learning (DRL) model facilitated distributed learning among agents [17]. The latter model allowed agents to tap into collective network knowledge while considering individual observations and goals. Another contribution [18] combined DQN and Deep Deterministic Policy Gradient (DDPG) to craft an RL model. The DQN element determined optimal resource allocation among vehicles, while DDPG enabled agents to select power levels with continuous variability. In [19], a model known as Double Dueling Deep Recurrent Q-Network (D3RQN) was introduced. This model aimed to address resource allocation and the assignment of discrete power levels for V2V links within an underlay scenario. In [20], the authors employed a multi-actor-attention-critic (MAAC) algorithm. The primary goal was to mitigate packet collisions that arose from the random resource selection process in V2X Mode 4. Within this approach, each vehicle takes into account factors like RSRP for each resource, QoS requirements, latency considerations, and reliability constraints when selecting a resource. In [21], a couple of resource allocation strategies were proposed within a highway environment. One strategy involved incorporating Long Short-Term Memory (LSTM) into a Deep Q-Network (DQN) model, while the other strategy incorporated LSTM into an Actor-Critic (AC) model. These strategies were designed to leverage the predictive capabilities of RSUs regarding vehicle mobility. This predictive information aided in allocating resources based on the priorities of the vehicles. In [22] the authors introduced the utilization of a centralized Reinforcement Learning (RL) model for the BS. This model serves the purpose of centralized resource allocation for V2I links, encompassing a spectrum of varying QoS requirements and numerological considerations.

Most previous studies on V2X communications share two prevalent characteristics. Firstly, they adopted a decentralized MARL model. It is noteworthy that in 5G NR mode 2, each vehicle autonomously selects the resources. This makes a decentralized multi-agent framework, where agents make decisions based on their own observations and learning, more fitting for capturing the V2X context than other RL models. Secondly, they incorporated global rewards for RL learning. The decentralized nature of MARL introduces non-stationarity due to actions by multiple agents, which can impair RL training performance. Nonetheless, supplying all agents with an identical global reward feedback mitigates this non-stationarity. Furthermore, global rewards shift the focus from competitive agent-based resource allocation to a fully cooperative endeavor, directed at improving overall network performance. This approach stands to enhance the V2X network's overall efficacy. In a prior study [23], we introduced a decentralized MA-DQN strategy featuring a global reward for resource allocation, aimed at maximizing V2X communication performance, particularly throughput. However, this study omitted RSUs and considered a homogenous traffic environment, centered solely on Cooperative Awareness Message (CAM) messages.

In this study, our focus lies on dedicated frequency bands for Intelligent Transportation Systems (ITS), ensuring no resource sharing with V2I links. Consequently, the need for mode selection diminishes, and the role of power control becomes less pronounced. This shifts the core emphasis of resource allocation toward efficient distribution among road entities. Additionally, we account for two traffic load situations. In cases of underload, where available resources exceed required allocation, the focus of resource allocation is to avert collisions. In contrast, during overload scenarios where resources are scarce, the objective shifts to minimizing interference stemming from collisions.

C. RSU

Numerous studies have underscored the pivotal role of Roadside Units (RSUs) in V2X communications [24], [25], [26], [27]. In one study [24], the authors underscored RSU deployment as a crucial solution for addressing urban traffic issues, particularly in densely populated and congested areas. RSUs equipped with sensors like LiDAR and cameras can collect road data and disseminate it to vehicles and pedestrians, thus ameliorating traffic congestion, enhancing safety, and expediting emergency responses. In another work [25], an overview of autonomous driving research directions and challenges reiterated the importance of RSU deployment to maximize road efficiency and elevate advanced driving performance. Another study [26] introduced the use cases and structure of Sensor Data Sharing Messages (SDSM), emphasizing RSUs' significant role in road sensor data sharing. Moreover, research highlighted in [27] reveals plans by China and South Korea to practically deploy RSUs in V2X environments, showcasing active utilization of RSUs in both academic and industrial contexts.

D. HETEROGENEOUS TRAFFIC

Several Standard Development Organizations (SDOs) have been actively involved in shaping Intelligent Transportation

TABLE 2. V2X messages and their characteristics.

Standard Organization	Messages	Continuity†	Casting††	Message size††† [bytes]	Link type	Abbreviation / Objective
SAE	BSM	C	Broadcast	320~350	V2V, V2I, V2P	**Basic Safety Message** Vehicles inform their speed, direction, and position to their surroundings.
	MAP	C	Broadcast	-	I2V	**Map Data** Provide geographical information of intersections to vehicles
	SPAT	C	Broadcast	-	I2V	**Signal Phase and Timing Message** Provide status of intersection and signal information such as priority in signal processing
	TIM	C	Broadcast	-	I2V, I2P	**Traveler Information Message** Transmit various information to V2X devices.
	PVD	E	All	-	I2V, V2I	**Probe Vehicle Data** Exchange of vehicle information with surrounding V2X devices (mainly RSUs).
ETSI	CAM	C	Broadcast	Around 350 (200 ~ 800)	V2V, V2I	**Cooperative Awareness Message** Vehicles inform their speed, direction, and position to their surroundings
	DENM	E	All	Up to1200	V2V, I2V, V2P, I2P	**Decentralized Environmental Notification Message** Notify the information of accidents, personal issues (engine malfunctions, brake failures, etc.) or hazardous situations on the road to their surroundings
	CPM	E	All	121 for header, 35 per included information	V2V, V2I, V2P, I2V, I2P	**Collective Perception Message** Alert the surroundings about uncooperative road users and obstacles on the road
	MAPEM	C	Broadcast	-	I2V	**Map Message** Provide geographical information of intersections to vehicles
	SPATEM	C	Broadcast	-	I2V	**Signal Phase and Timing Message** Provide status of intersection and signal information such as priority in signal processing
	IVIM	C	Broadcast	-	I2V	**In-Vehicle Information Message** Provide road information (speed limits, road conditions, etc.) to vehicles

† C: Continuous. E: Event-triggered.
†† All: Unicast, groupcast, and broadcast.
††† Hyphen: Not specified clearly.

Systems (ITS) implementation. Notably, the Society of Automotive Engineers (SAE) initially established V2X communication standards, harmonizing with Wireless Access in Vehicular Environments (WAVE) communication based on IEEE 1609.3 and IEEE 802.11p. These standards were later adapted to accommodate C-V2X communication, a product of 3GPP. The European Telecommunications Standards Institute (ETSI) contributed with standards for V2X communication, utilizing ETSI-ITS-G5 based on WiFi technology. Subsequently, ETSI embraced C-V2X from 3GPP and amended its standards to encompass C-V2X as well. The 5G Automotive Association (5GAA), an international consortium uniting communication and vehicle industry entities, also plays a pivotal role in this landscape.

In diverse V2X environments, messages with distinct objectives, formats, and sizes are defined [4], [28]. A comprehensive overview of key V2X messages established by SAE and ETSI is presented in Table 2. We provide insight into message characteristics encompassing continuity, casting type, message size, V2X link type, and objective. As per 5GAA classification [29], V2X messages fall into two categories: continuous and event-triggered. Continuous messages, like BSM and CAM, facilitate continuous information sharing among road users and infrastructure entities through periodic broadcasts. These messages share details such as heading, speed, acceleration, and position. Conversely, event-triggered messages serve to alert surroundings about specific incidents, such as accidents, road obstacles, or vehicle-related issues like engine trouble or low fuel. These messages also serve as information requests or exchanges. Unlike continuous messages, event-triggered ones can adopt unicast, groupcast, or broadcast modes, contingent on the use case. Unicast might be chosen for Emergency Stops, alerting a specific road user; groupcast may enable Cooperative Adaptive Cruise Control (CACC) for platooning; and broadcast proves fitting for Emergency Vehicle Warning (EVW), notifying hazardous road situations.

Message sizes fluctuate depending on the volume of information they convey. A case in point is the MAP message, which broadcasts intersection-based geographic data. Its size varies based on factors like intersection count and detail level. For extensively studied messages like CAM, the average size approximates 350 bytes. However, it spans a range of 200 to 800 bytes due to optional data, such as vehicle category and security content [30]. Many messages are exchanged between infrastructure and entities via I2V, I2P, and V2I links. This owes to the infrastructure's broader field of view and array of sensors, enabling the transmission of information that might pose challenges for standard road users to obtain.

The array of aforementioned messages can collaboratively contribute to a singular use case. Consider the scenario of Automated Intersection Crossing, where an autonomous vehicle negotiates an intersection. Here, the vehicle draws upon diverse messages to shape its driving plan. It acquires SPAT and MAP messages from infrastructure, obtaining insights into traffic light phase, timing, and geographical specifics. Simultaneously, it relies on CAM or BSM messages from fellow vehicles to decipher their driving intentions. This amalgamation of messages empowers the autonomous vehicle to confidently traverse the intersection and accomplish its crossing maneuver securely.

Beyond the V2X messages detailed in Table 2, several supplementary messages are also established [31], [32]. SAE remains actively engaged in formulating fresh messages to suit diverse use cases. For instance, the most recent version of the SAE J2735 document, released in November 2022, introduces novel messages. These include the Road Weather Message (RWM), facilitating the exchange of weather data, the Pedestrian Safety Message (PSM) designed to bolster the safety of vulnerable road users, and the Sensor Data Sharing Message (SDSM) intended for the sharing of road hazard information stemmed from vehicles and infrastructure, utilizing their individual sensors for detection.

The existing resource allocation schemes have primarily focused on managing singular traffic types, showing limited interest in accommodating heterogeneous traffic. Therefore, it becomes essential to incorporate the diverse nature of V2X messages, particularly considering varying sizes, to offer a more comprehensive solution that aligns with real-world scenarios. This study delves into the resource allocation challenge posed by heterogeneous traffic with RSU-deployed V2X communications. To address this, a novel multi-agent reinforcement learning-based resource allocation scheme is introduced. The scheme's effectiveness is assessed based on Packet Reception Ratio (PRR) and communication range. In order to gauge its performance, the proposed scheme is systematically compared against random allocation, 5G NR mode 2, and optimal allocation schemes.

III. SYSTEM MODEL AND PROBLEM FORMULATION

Figure 1 illustrates a system model in a V2X network deployed on a highway with an RSU. The RSU is located alongside the road's left side, proximate to the upper lane. The highway accommodates N vehicles. In this scenario, a RSU transmits a message with a large size, whereas $K - 1$ vehicles are assumed to dispatch messages with a smaller size. For simplicity, we designate DENM and CAM for messages with large and small sizes, respectively, given that DENM typically surpasses CAM in size [33], [34]. Both messages sustain a default transmission interval of 100 msec.

In the NR frame structure, the smallest resource block is a Physical Resource Block (PRB), encompassing 7 OFDM symbols and 12 subcarriers. In this paper, we presume that CAM necessitates a single TB, constructed from multiple PRBs. Moreover, we assume DENM requires D TBs. The required number of PRBs for a TB transmission hinges on message size, Modulation and Coding Scheme (MCS), and numerology. In our scenario, all entities on the road, including the RSU, access a shared resource pool containing M TBs. Among these, seven vehicles in red denote the receiving vehicles, while the three blue vehicles signify the transmitting vehicles requiring resource allocation.

On the left portion of Figure 1, both UE 1 and UE 2 share the same TB (red), causing interference between them. Similarly, UE 3 and the RSU employ the same yellow-colored TB, resulting in interference as well. Consequently, the overall performance of the V2X network degrades. Each road entity employs its individual DQN for autonomous decision-making. The process heavily relies on RSRP measurements across all TBs. For instance, analyzing UE 1's RSRP indicates

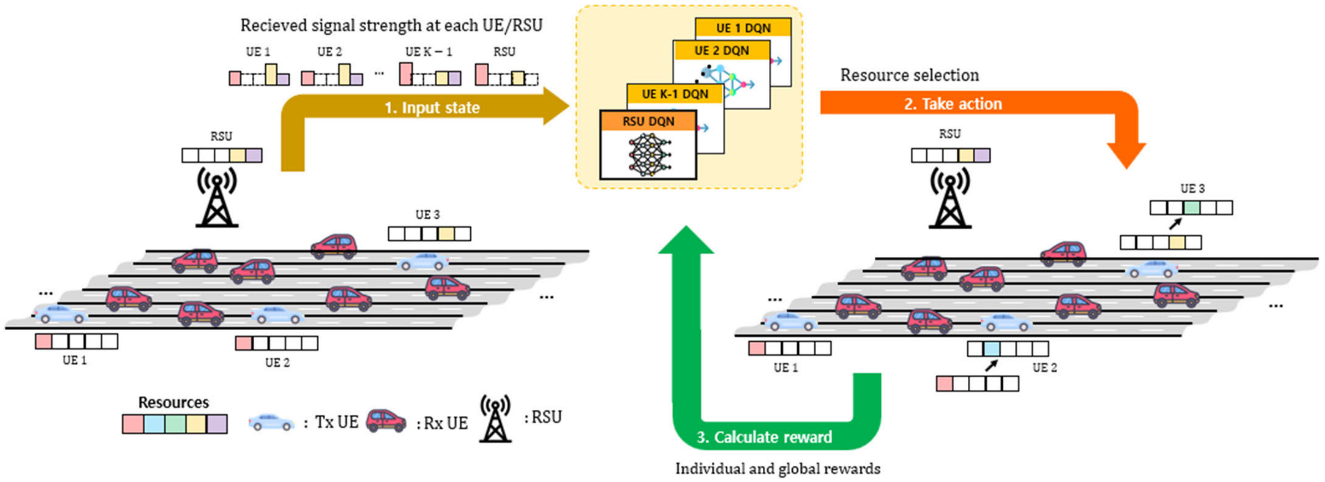


FIGURE 1. System model with heterogeneous V2X traffic in an underload situation (*N* = 10, *K* = 3, *M* = 5, and *D* = 2).

stronger reception on the red TB from nearby UE 2 compared to the yellow TB, which receives signals from the RSU and UE 3. Moreover, the violet TB, allocated solely to the distant RSU, exhibits weak RSRP. Armed with this RSRP data, each road entity reallocates resources. Following reallocation, a global reward is computed, encompassing average PRR performance and channel capacity for vehicles experiencing reception failures. This global reward is then furnished as feedback for DQN training. On the right side of Figure 1, the scenario unfolds after resource reallocation using trained MA-DQNs. UE 2 and UE 3 are reassigned to untapped blue and green resources. This targeted reallocation eradicates interference across all road entities, significantly boosting overall performance.

We consider a large-scale fading component α, which incorporates path loss and shadowing effects. By considering a small-scale fading component $\tilde{g}$, the composite channel gain h can be expressed as $h = \alpha\tilde{g}$. The parameter $\tilde{g}$ follows a Rayleigh distribution with zero mean and unit variance. Let the channel gain between the k-th transmitter and the j-th receiver is denoted as $h_{k,j}$. Then, the signal-to-interference-plus-noise ratio (SINR) received at the j-th receiver from the k-th transmitter, $\gamma_{k,j}$ can be expressed as

$$\gamma_{k,j} = \frac{P_{Tx} \cdot h_{k,j}}{P_{Tx} \cdot \sum_{m=1}^{M} \rho_k[m] \cdot \sum_{\substack{l=1 \\ l \neq k}}^{K} \rho_l[m] \cdot h_{l,j} + \sigma^2}, \tag{1}$$

where σ^2 is the noise power, P_{Tx} is the transmission power. We introduce binary resource selection indicators, $\rho_k[m] \in \{0, 1\}$ and $\rho_l[m] \in \{0, 1\}$, which are 1 if the k-th and l-th transmitter use the m-th TB and otherwise 0, respectively.

To measure PRR, we utilize the SINR threshold γ', i.e., the message reception is considered as successful if the received SINR is equal to or greater than γ'. We introduce an indicator, $\omega_{k,j} \in \{0, 1\}$, which is 1 when $\gamma_{k,j} \geq \gamma'$ and 0 otherwise. Let N_k be the number of receiving vehicles from the k-th transmitter. The PRR of the k-th transmitter, P_k can be expressed as

$$P_k = \frac{1}{N_k}\left(\sum_{j=1}^{N_k} \omega_{k,j}\right). \tag{2}$$

Therefore, the resource allocation problem can be formulated as

$$\arg\max_{\boldsymbol{\rho}} P_k \tag{3a}$$

$$\text{s.t.} \sum_{m=1}^{M} \rho_k[m] = 1, (1 \leq k < K) \tag{3b}$$

$$\sum_{m=1}^{M} \rho_K[m] = D, \tag{3c}$$

where $\boldsymbol{\rho} = \{\rho_1[1], \ldots, \rho_k[m], \ldots, \rho_K[M]\}$ is the set of the resource selection indicators. (3b) implies that only a single TB is assigned for CAM transmission. (3c) means that the last transmitter is an RSU that requires D TBs to broadcast DENM. The formulated problem (3) is a non-convex combination and NP-hard problem, which is very difficult to be directly solved. However, reinforcement learning may offer an effective strategy to discover the optimal policy for attaining the highest reward.

IV. MULTI AGENT REINFORCEMENT LEARNING MODEL

We first consider an independent Q-learning (IQL)-based MARL model. The IQL is a decentralized policy-based model in multi-agent RL [35], where each agent trains its network with its own observations, considering other agents as part of the environment. However, the IQL has a drawback in that all agents learn in a non-stationary environment since other agents also adjust their behavior during the learning process. This drawback becomes more pronounced when

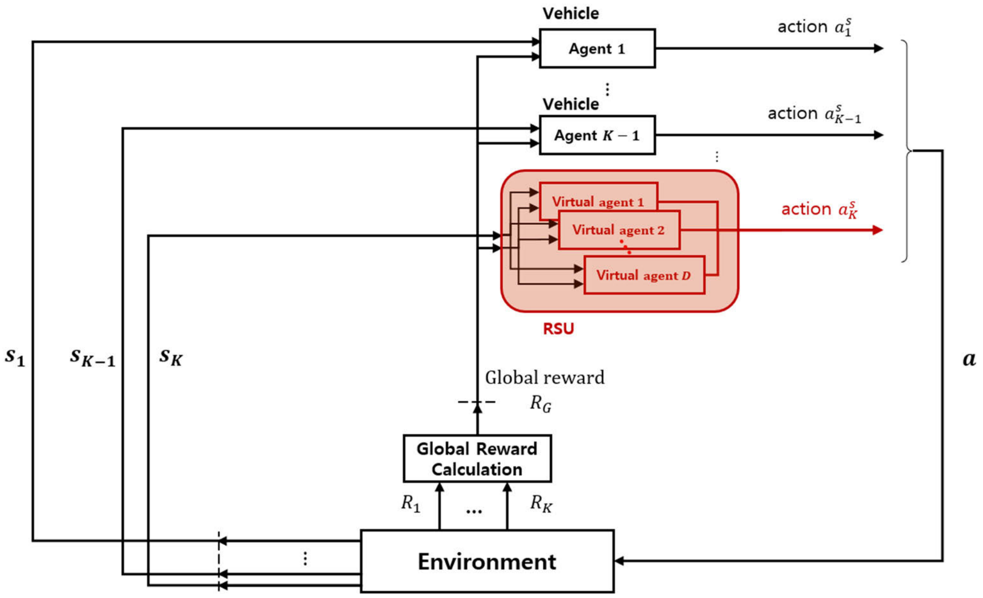


FIGURE 2. Structure of MARL.

using the DQN structure, where the learning process relies on past experiences stored in the replay memory. In a non-stationary environment, the correlation between the current state-action pair and the past state-action pair may be weakened, leading to unstable learning [36]. Figure 2 illustrates the structure of the proposed MA-DQN model with reward sharing, where agent, state, action, and reward are defined as follows.

A. AGENT

Each transmitting V2X entity is an agent, where the k-th agent decides an action, a_k^s based on the observed state, $\boldsymbol{s_k}$. For transmission of a message, vehicle agents require 1 TB while the RSU agent needs to select D TBs. This heterogeneity incurs different action spaces for vehicle agent and RSU agent. In the proposed MA-DQN model, a single RSU agent is formed by a group of D virtual agents, each of which allocates a single TB as a vehicle agent does.

B. STATE

We define the state based on the previous RSRP values on all the TBs. By denoting RSRP as $q_k[m]$, which represents the received signal strength on the m-th TB sensed by the m-th transmitter, can be expressed as

$$q_k[m] = \sum_{\substack{l=1 \\ l \neq k}}^{K} \rho_l[m] \cdot P_{Tx} \cdot h_{l,k}. \tag{4}$$

Then, the state of the k-th agent can be expressed as

$$\boldsymbol{s_k} = \{q_k[1], \ldots, q_k[M]\}. \tag{5}$$

C. ACTION

The action is defined as the selection of TB or TBs by each agent. The size of the action space depends on whether the agent is an RSU or a vehicle. For vehicles, they need to select only one TB at a time. Thus, size of the action space for vehicles is equal to M. On the other hand, the RSU has to select D TBs among M available TBs. Thus, the size of action space for RSU becomes ${}_M C_D$, representing the number of combinations of selecting D TBs from M TBs. In this case, we observe that the action space for RSU becomes larger than the action space for vehicles, which hinders training of the RSU agent. To mitigate this problem, instead of a single DQN model selecting D TBs at the same time, we implement an RSU agent with D virtual agents, each of which only selects an TB as in Figure 2. This approach ensures a more balanced and manageable action space for the RSU agent, enabling effective learning.

D. REWARD

An individual reward of the k-th agent, R_k is calculated based on local observation of the k-th agent after execution of its action. And then, a global reward, R_G is calculated, which is associated with the actions of all agents, and is commonly applied to the learning of all the agents.

We introduce a penalty in reward calculation based on the channel capacities of the vehicles experiencing reception failures. Then, a penalty of the k-th agent, D_k can be expressed as

$$D_k = \sum_{j=1}^{N_k^{(f)}} \left(\frac{1}{\max\left\{C_{min}, C_{k,j}^{(f)}\right\}} \right), \tag{6}$$

where $N_k^{(f)}$ is the total number of receiving vehicles that experience reception failures for the k-th transmitter within its service area. And $C_{k,j}^{(f)}$ represents the channel capacity of the j-th reception failed receiver. We introduce a minimum threshold value for $C_{k,j}^{(f)}$, denoted as C_{min}, which ensures that the penalty does not become excessively large, particularly in the situations with extremely poor channel conditions. $C_{k,j}^{(f)}$ can be obtained by

$$C_{k,j}^{(f)} = W_k \cdot \log_2 \left(1 + \gamma_{k,j}^{(f)}\right), \tag{7}$$

where W_k is the bandwidth, and $\gamma_{k,j}^{(f)}$ is the SINR at the j-th receiver from the k-th transmitter. Then, the individual reward R_k can be expressed as

$$R_k = \lambda_P P_k - \lambda_D D_k, \tag{8}$$

where λ_P and λ_D are constant weights to balance the PRR and penalty, respectively. For the resource allocation of heterogeneous traffic with different message sizes, we introduce priority weights for the calculation of the global reward. Thus, the global reward R_G can be expressed as

$$R_G = \lambda_V \left(\frac{1}{K-1}\sum_{k=1}^{K-1} R_k\right) + \lambda_R R_K. \tag{9}$$

In this case, λ_R and λ_V are priority weights for RSU and vehicle, respectively. By increasing the value of λ_R, while satisfying $\lambda_R + \lambda_V = 1$, the RSU can be controlled to have a higher priority than the vehicle in the global reward.

Each DQN model is composed of three hidden layers with a ReLU activation function and a 1-dimensional batch normalization process. The DQNs are trained to select an optimal policy π^* that maximizes the expected reward G_k. Here, considering the decaying component of the reward as γ, G_k can be expressed as

$$G_k = \sum_{t=0}^{T-1} \gamma^t R_k, \quad (0 \leq \gamma \leq 1) \tag{10}$$

where R_k represents the individual reward of the k-th agent obtained at each time step, and T is the total number of time steps in an episode. The discount factor γ determines the importance of future rewards relative to immediate rewards. By maximizing the expected reward G_k, the agents aim to learn an optimal policy that leads to higher cumulative rewards over time. The action a_k^s can be expressed as

$$a_k^s = \underset{a_k^s \in A_k}{\arg\max} \left[Q\left(\mathbf{s_k}, a_k^s, \theta_k\right)\right], \forall k \tag{11}$$

where $Q\left(\mathbf{s_k}, a_k^s, \theta_k\right)$, θ_k, and A_k denote output Q-value, weights of Q-network, and action space of the k-th agent, respectively. After all agents have taken their actions, the global reward R_G and the next state for each agent are determined. And then, each agent stores the state, action, global reward, and the next state as a tuple in its own replay memory. Subsequently, the k-th agent randomly selects a mini-batch of experiences with a size of $\mathcal{D}$ from its replay memory to compute the loss function $L_k^{\mathcal{D}}(\theta_k)$. $L_k^{\mathcal{D}}(\theta_k)$ is defined based on the difference between the Q-values of the Q-network and the target Q-network. When we denote the target Q-network for the k-th agent, as $\tilde{Q}\left(\mathbf{s_k}, a_k^s, \tilde{\theta}_k\right)$, $L_k^{\mathcal{D}}(\theta_k)$ can be expressed as.

$$\begin{aligned} L_k^{\mathcal{D}}(\theta_k) = \sum_{i=1}^{\mathcal{D}} [&R_G(i) \\ &+ \gamma \max_{a_k^s(i+1)} \tilde{Q}(\mathbf{s_k}(i+1), a_k^s(i+1)\, \tilde{\theta}_k) \\ &- Q\left(\mathbf{s_k}(i+1), a_k^s(i+1), \theta_k\right)]^2 \end{aligned} \tag{12}$$

Afterward, the weights θ_k are updated using stochastic gradient descent to minimize the loss function. Specifically, the weight update is performed in the direction that minimizes the loss function. Similarly, the weights $\tilde{\theta}_k$ of the target Q-network for the k-th agent are periodically updated to match the weights θ_k. This synchronization ensures that the target Q-network closely follows the learned Q-network and stabilizes the learning process. Algorithm 1 summarizes the learning process of the proposed MARL model.

V. PERFORMANCE EVALUATION

We consider a heterogeneous traffic environment in a highway, where a single RSU is deployed. Here, the RSU and vehicles periodically disseminate DENM and CAM messages, respectively, utilizing a shared resource pool containing M TBs. Each TB has a 5.6 MHz bandwidth and a time duration of 1 msec. Our simulation framework aligns with the configuration outlined in 3GPP NR document, for a highway environment, characterized by a six-lane layout with a 4 m lane width and 2 km length [4]. The upper trio of lanes accommodates rightward traffic, while the lower trio caters to leftward traffic. The RSU is assumed to be positioned 800 m from the center of the highway towards the left side, and at a distance of 4 m from the top lane. N vehicles are randomly generated on the highway, following a spatial Poisson distribution. Among them, $K-1$ vehicles are randomly selected as transmitting vehicles. When a vehicle moves beyond the boundary of the highway, we set it to reappear on the opposite side of the same lane. $K-1+D$ TBs are required for the V2X network, where $K-1$ TBs for the transmitting vehicles, and D TBs for the RSU. We consider two situations with different levels of channel congestion; (1) an underload situation where the required amount of resource is equal to or smaller than the allocated amount of resource, i.e., $K-1+D \leq M$ and (2) an overload situation where $K-1+D$ is larger than M, so some resources are allocated overlapped.

For NLOSv channel, we adopt the knife-edge diffraction model [37], [38]. The detailed parameters can be found in Table 3. Our approach adheres to 3GPP guidelines for calculating the SINR threshold γ_{th}, adopting MCS 6 and numerology 1 [39]. With a 900-bytes DENM message, it is determined that its transmission requires 2 TBs, i.e., $D = 2$.

Algorithm 1 Learning Process of the Proposed MARL Model

```
Generate $N$ vehicles randomly and 1 RSU at the left side of the highway;
Initialize Q-networks for all agents
for each position do
    Update vehicle positions and large scale fading component
    for each step $t$ do
        for each agent $k$ do
            Observe $s_k$
            Choose action $a_k^s$ according to $\epsilon$-greedy policy
        end for
        Update small scale fading component
        All agents take actions and receive global reward $R_G$
        for each agent $k$ do
            Observe $\boldsymbol{s_k}$
            Store $\left(\boldsymbol{s_k}, a_k^s, R_G\right)$ in replay memory
        end for
    end for
    for each agent $k$ do
        Randomly sample mini-batches from replay memory
        Optimize error between Q-network and target network using stochastic gradient descent
    end for
end for
```

TABLE 3. Simulation parameters.

Parameters	Value
Carrier frequency	5.9 GHz
Bandwidth per TB, W_k	5.6 MHz
Number of TBs, M	10
Noise power, σ^2	-114 dBm/Hz
Transmission power, P_{Tx}	23 dBm
Pathloss model	WINNER + B1
Vehicle velocity	70 km/h
Vehicle density	5/lane/km
Number of transmitting vehicles	6, 10
Size of CAM	350 bytes
Size of DENM	900 bytes
Number of TBs for DENM message transmission, D	2, 3
SINR threshold, γ_{th}	4.93 dB

The SINR across D TBs is accessed using the equivalent effective SINR mapping (EESM) method [40].

The architecture of each agent's DQN is designed with a fully connected structure, comprising a total of five layers, including three hidden layers. The hidden layers consist of n_1, n_2, and n_3 neurons, respectively, employing ReLU as the activation function. To strike a balance between exploration and exploitation, a ϵ-greedy policy is applied during training, with ϵ linearly reduced. This approach permits each DQN to explore a broad spectrum of action-state pairs through extensive initial exploration. As training advances, the exploitation ratio increases, enabling the agent to increasingly rely on the acquired network knowledge. This strategy accelerates convergence, fostering efficient learning. We train each agent's Q-network for a total of 20,000 episodes. The exploration rate ϵ is reduced from 1 to 0 over the initial 10,000 episodes. Afterward, the exploration rate remains constant. The specific parameters employed within the MARL framework are detailed in Table 4.

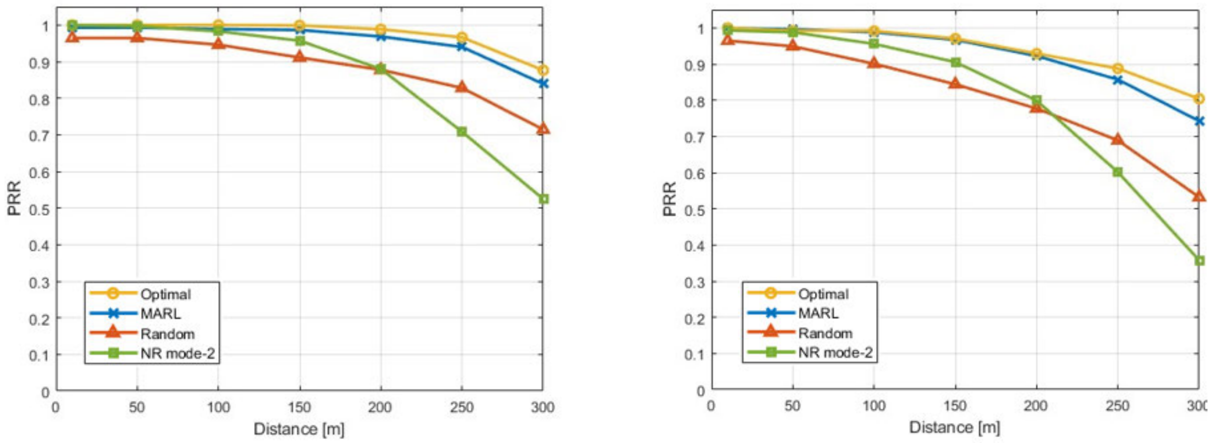
TABLE 4. MARL parameters.

Parameters	Value
Weight for the positive reward, λ_P	10
Weight for the negative reward, λ_D	1
Pair of priority weights, $[\lambda_R, \lambda_V]$	[0.5,0.5], [0.7,0.3]
Exploration rate, ϵ	1.0 → 0
Discount factor, γ	0.9
Number of neurons of hidden layers, $[n_1, n_2, n_3]$	[500, 250, 120]

The proposed scheme is compared with three schemes: random, 5G NR mode 2, and optimal. With a random resource allocation scheme, agents randomly select TBs for each transmission. As an optimal resource allocation scheme, all agents select TBs to maximize the PRR performance across the entire V2X network. Achieving optimal performance involves exploring all feasible resource allocation combinations.

A. TWO CHANNEL CONGESTION LEVELS WITH $D = 2$

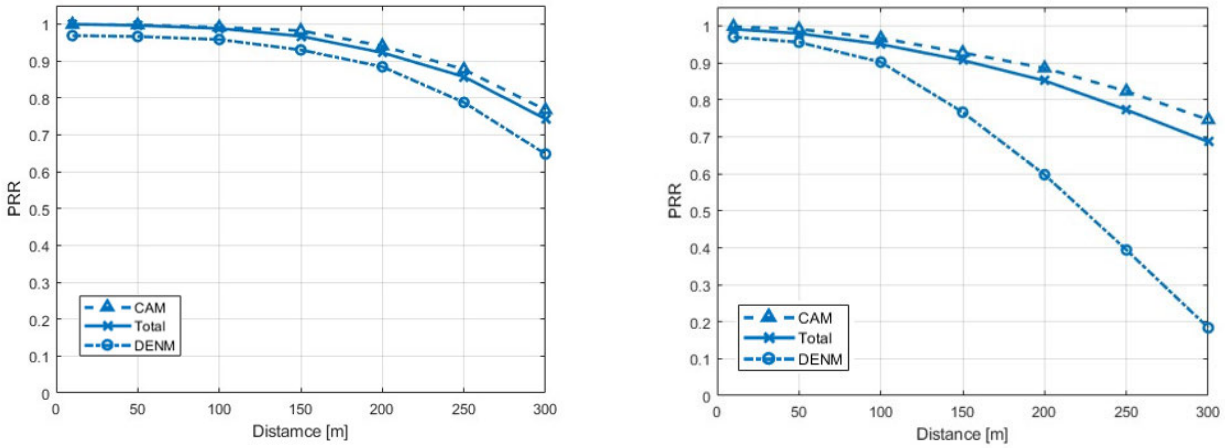
Figure 3 conducts a comparative analysis of four resource allocation strategies: random, 5G NR mode 2, optimal, and the proposed MA-DQN schemes, focusing on PRR performance variation across transmitter-receiver distances. Two congestion situations, involving 8 or 12 required TBs with 10 allocated TBs, are explored in Figures 3 (a) and (b), respectively. At a 200 m distance in the underload situation, the MA-DQN scheme closely approximates optimal PRR performance, while the random scheme achieves a PRR of 0.88. Further, for the communication range satisfying a PRR of 0.9, the proposed MA-DQN scheme attains 94.7% of the optimal result extending up to about 270 m, which is a 35% higher performance compared to the conventional 5G NR mode 2 scheme. Conversely, the random scheme's coverage is notably limited, spanning 155m. It is crucial to emphasize that beyond a distance of 200 m, the 5G NR mode 2 scheme exhibits inferior performance compared to



(a)

(b)

FIGURE 3. PRR performance of random, 5G NR mode 2, optimal, and proposed MA-DQN resource allocation schemes. (a) Underload situation with 6 transmitting vehicles and 1 RSU. (b) Overload situation with 10 transmitting vehicles and 1 RSU.



(a)

(b)

FIGURE 4. PRR performance of CAM and DENM messages of the proposed MA-DQN scheme with $\lambda_R = 0.5$. (a) $D = 2$. (b) $D = 3$.

the random scheme. In the 5G NR mode 2 scheme, a UE filters out resources with low RSRP, minimizing the selection of resources utilized by neighboring UEs in close proximity. This advantageous feature contributes to achieving a commendable PRR performance, especially when the transmitter and receiver are within a short distance, such as less than 200 m. However, as the distance increases, the 5G NR mode 2 scheme becomes more susceptible to resource collisions. In contrast, the random scheme selects resources without considering proximity, resulting in consistent collision probability irrespective of distance. Consequently, the resource collision probability of the 5G NR mode 2 scheme surpasses that of the random scheme when the distance exceeds 200 m. The performance of the 5G NR mode 2 scheme degrades rapidly and becomes inferior to the random scheme with increasing distance.

Similar trends emerge within the overload situation illustrated in Figure 3 (b). However, it's important to note that performance in the overload situation is dampened due to interference among network entities assigned the same TB. For instance, the optimal scheme registers a PRR of 0.98 at 200 m in the underload situation, which drops to 0.92 in the overload situation. In comparison, the MA-DQN, 5G NR mode 2 and random schemes yield PRRs of 0.91, 0.8, and 0.78, respectively. Notably, the communication range dwindles to 100 m under the random scheme, while the proposed MA-DQN scheme dramatically extends it to 219 m. This substantial decline in performance underscores the amplified significance of a suitable resource allocation strategy under overload conditions.

B. PRR PERFORMANCE OF CAM AND DENM

The specific PRR performance of CAM and DENM messages is subject to further examination in Figure 4. Notably, the DENM message with D set to 3 is included in the simulation under overload conditions. Figures 4(a) and 4(b)

TABLE 5. Resource collision probabilities according to *D* and the number of collided TBs.

D	Message	Total collision probability	Number of collided TBs		
			1	2	3
2	CAM	29%	29%	-	-
	DENM	56%	47%	9%	-
3	CAM	40%	40%	-	-
	DENM	99%	63%	33%	3%

illustrate the PRR performance of the proposed scheme for DENM messages with D values of 2 and 3, respectively.

As anticipated, the PRR of DENM message is observed to be inferior to that of CAM message, with the gap widening at longer distances. Notably, as the size of the DENM message (D) increases from two to three, the degradation in PRR becomes more pronounced. Concerning the distance required to achieve a PRR of 0.9, it is 250 m for CAM messages, whereas for DENM messages with $D = 2$, it is only 200 m. In the case of $D = 3$, this distance further diminishes to 150 m for CAM messages and a mere 100 m for DENM messages, as depicted in Figure 4(b). At a distance of 200 m, the PRR of DENM messages stands at 0.6, underscoring a substantial need for improvement to meet the performance demands of V2X applications.

To delve into this phenomenon, we turn our attention to the resource collision probabilities of CAM and DENM messages, as detailed in Table 5. A collision occurs for a transmitting vehicle when another vehicle selects the same resource. Given that a DENM message occupies D Transport Blocks (TBs), the collision probability is assessed even if only one of the D TBs is selected by other transmitting vehicles. The collision probability is further analyzed based on the number of collided TBs, contributing to the overall collision probability. For CAM messages, the collision probability with $D = 3$ is 11% higher compared to that with $D = 2$, resulting in a poorer PRR performance. Remarkably, collisions occur almost consistently for DENM messages with $D = 3$ in overload situations. Since the selection procedure can exclude resources occupied by neighboring transmitting vehicles, the PRR performance doesn't significantly degrade within distances less than 100 m. However, it deteriorates rapidly with increasing distance due to the substantial collision probability.

In Figure 5, the CDFs of SINR for CAM and DENM messages at a distance of 100 m are presented. Evidently, the SINR is higher in the case of $D = 2$ compared to $D = 3$. As expected, the SINR of DENM messages is inferior to that of CAM messages, and this difference becomes more pronounced with $D = 3$ due to the heightened collision probability. For instance, the gap between the median SINRs of CAM and DENM messages is approximately 7 dB for $D = 2$, whereas it widens to about 13 dB for $D = 3$. Moreover, in the case of $D = 2$, the tail probability at γ_{th} (a specified threshold) is nearly identical for both CAM and DENM messages. However, in the case of $D = 3$, the DENM message exhibits a higher probability of SINR $\leq \gamma_{th}$, indicating a more challenging communication environment for DENM messages with larger sizes.

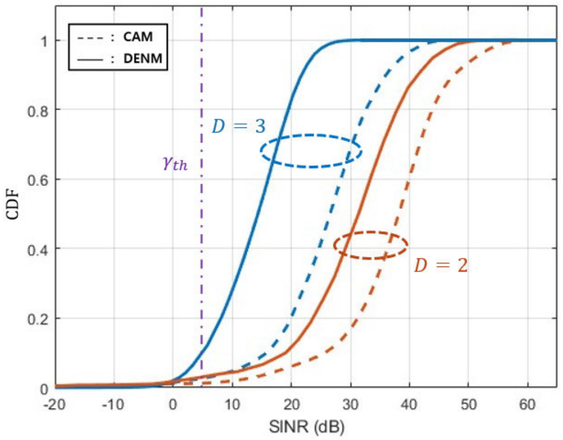


FIGURE 5. CDF of received SINR of CAM and DENM messages at 100 m for *D* = 2 and 3.

C. IMPACT OF MESSAGE PRIORITY

Our analysis now extends to the PRR performance with varying priority weights for small and large messages, λ_R and λ_V, respectively. In Figure 4 (a), λ_R and λ_V both stand at 0.5, indicating equal message priority. Conversely, Figure 6 confers a higher priority to large messages, with λ_R set at 0.7. When messages carry equal priority, the PRR of DENM messages trails CAM messages. This discrepancy is attributed to DENM messages necessitating more TBs for transmission than CAM messages, rendering them more susceptible to resource collisions. By elevating the priority of DENM over CAM messages in Figure 6, we observe an enhancement in the PRR of DENM messages, slightly surpassing that of CAM messages. This outcome underscores the efficacy of our MARL approach in accommodating heterogeneous messages.

Figure 7 illustrates the communication ranges for CAM and DENM messages within an overload scenario, spanning diverse priority constants for λ_R. As the DENM priority constant ascends from 0.5 to 0.7, the communication range of DENM experiences a linear upsurge, albeit less pronounced beyond 0.8. In contrast, increasing the DENM priority constant adversely impacts the communication range of CAM. Consequently, the selection of an apt value for λ_R assumes significance, aiming to strike a harmonious performance equilibrium between CAM and DENM messages.

D. ADVANTAGE OF VIRTUAL AGENTS

Figure 8 illustrates the global reward progression across training episodes in overload situation for the proposed scheme

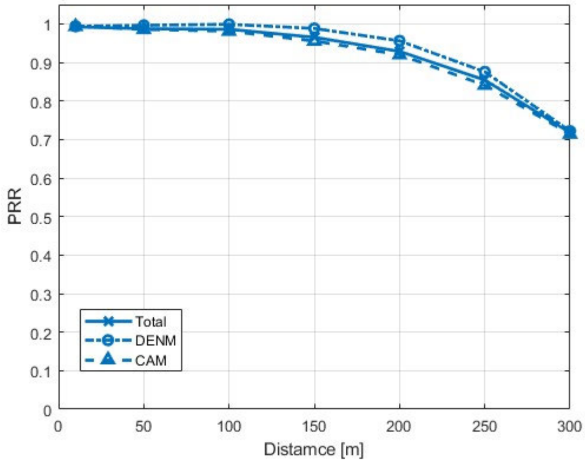


FIGURE 6. PRR performance of CAM and DENM messages with the proposed MA-DQN scheme ($\lambda_R = 0.7$).

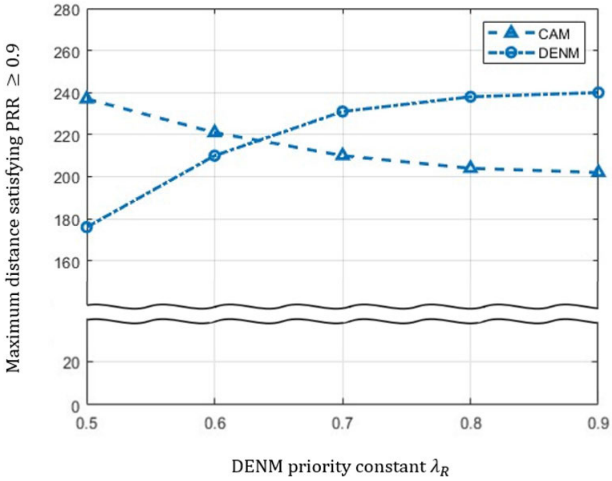


FIGURE 7. Communication range for CAM and DENM messages with different DENM priority constant λ_R.

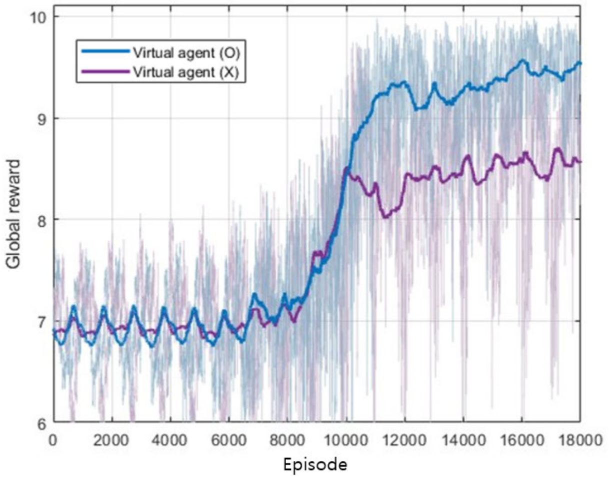


FIGURE 8. Global reward per episode with and without adopting virtual agents.

with and without adopting virtual agents. The lighter-colored graph captures average rewards over every 10 episodes, while the darker-colored graph averages rewards across 300 episodes. In the initial phase, where exploration is emphasized, reward values show limited growth. However, as exploitation takes precedence, rewards undergo a rapid ascent. In the latter stages, marked by exploitation-based actions, the model consistently sustains high rewards. This progression indicates the proficient training of our MA-DQN model. Through the implementation of virtual agents with a reduced action space size, there is a notable enhancement in the global reward, resulting in an approximate 11% improvement. This, in turn, translates to a commendable 9% enhancement in PRR at 200 m. The integration of virtual agents contributes positively to the overall system performance.

E. COMPUTATIONAL COMPLEXITY COMPARISON BETWEEN OPTIMAL AND PROPOSED MA-DQN SCHEMES

In the random scheme, a transmitting vehicle opts for a TB or TBs through a random selection process, incurring no computational complexity. In the 5G NR mode 2 scheme, after measuring RSRPs on each TB, a transmitting vehicle randomly selects a specified number of TBs from those with the $X\%$ lowest RSRP values. As a result, the primary computational complexity arises from the sorting procedure. While both schemes are computationally straightforward due to their reliance on random selection, their performances are deemed inadequate for handling the intricacies of heterogeneous V2X traffic environments.

The computational complexity is assessed in terms of the number of multiplications for both the optimal and proposed schemes. In the optimal scheme, actions for an agent are determined by identifying the best-performing one among all possible combinations. The complexity is calculated by multiplying the total number of combinations by the number of multiplications needed to compute the SINR per combination. In this context, there are K-1 vehicles transmitting CAM messages, and each of them selects one TB out of M TBs. The RSU transmits DENM messages and selects D TBs among M TBs. The total number of possible combinations can be given by $M^{K-1} \cdot {}_{M}C_{D}$. The SINR calculation involves 1 multiplication in the numerator and $2M \cdot (K-1)$ multiplications in the denominator, resulting in a total of $2M \cdot (K-1)+1$ multiplications. Therefore, the complexity for the optimal scheme in the overload situation is approximately 9×10^{13}. The complexity of a DQN model is determined by the sum of the number of multiplications across all layers, where the number of multiplications at each layer is the product of the neurons in that layer and the previous layer [41]. The complexity of the MA-DQN becomes the number of agents times the complexity of a DQN model. Consequently, in the overload situation, the complexity of the proposed MA-DQN is approximately $9.2 \cdot 10^6$. It is evident that the proposed scheme requires significantly lower execution complexity than the optimal scheme, even though it may involve increased computational complexity during the training phase.

VI. CONCLUSION

This paper proposed a resource allocation scheme for V2X broadcast communication, employing a decentralized MA-DQN model with shared global rewards. The scheme effectively managed a heterogeneous traffic environment by incorporating prioritized global weights, expanding the communication range for DENM messages while accommodating their larger resource demand compared to CAM messages. Additionally, the paper proposed a virtual RSU implementation where groups of agents each select a single TB, optimizing learning efficiency. The MA-DQN model's performance was assessed and contrasted against random, 5G NR mode 2, and optimal schemes across varying channel congestion situations and different DENM message sizes. The proposed scheme showcased superior performance over the random and 5G NR mode 2 schemes, and nearly matched the optimal scheme. For instance, in the overload situation with $D = 2$, the communication range increased from 100 m with the random scheme to 219 m with the proposed scheme, representing 92.5% of the optimal scheme's range.

The envisaged MARL-based resource allocation scheme offers potential for expansion to accommodate various casting types, such as unicast and groupcast. This extension can also facilitate energy-efficient provisions for pedestrians and integration with a UAV-assisted network.

REFERENCES

[1] M. H. C. Garcia, A. Molina-Galan, M. Boban, J. Gozalvez, B. Coll-Perales, T. Sahin, and A. Kousaridas, "A tutorial on 5G NR V2X communications," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1972–2026, 3rd Quart., 2021.

[2] A. Alalewi, I. Dayoub, and S. Cherkaoui, "On 5G-V2X use cases and enabling technologies: A comprehensive survey," *IEEE Access*, vol. 9, pp. 107710–107737, 2021.

[3] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, "Connected vehicles: Solutions and challenges," *IEEE Internet Things J.*, vol. 1, no. 4, pp. 289–299, Aug. 2014.

[4] *Study on Evaluation Methodology of New Vehicle-to-Everything (V2X) Use Cases for LTE and NR, (V15.3.0, Release 15)*, 3GPP, Standard TR 37.885, Jun. 2019.

[5] H. Jiang, M. Mukherjee, J. Zhou, and J. Lloret, "Channel modeling and characteristics for 6G wireless communications," *IEEE Netw.*, vol. 35, no. 1, pp. 296–303, Jan. 2021.

[6] K. Guan, B. Peng, D. He, J. M. Eckhardt, S. Rey, B. Ai, Z. Zhong, and T. Kürner, "Measurement, simulation, and characterization of train-to-infrastructure inside-station channel at the terahertz band," *IEEE Trans. THz Sci. Technol.*, vol. 9, no. 3, pp. 291–306, May 2019.

[7] B. Xiong, Z. Zhang, J. Zhang, H. Jiang, J. Dang, and L. Wu, "Novel multi-mobility V2X channel model in the presence of randomly moving clusters," *IEEE Trans. Wireless Commun.*, vol. 20, no. 5, pp. 3180–3195, May 2021.

[8] *LTE; 5G; Overall Description of Radio Access Network (RAN) Aspects for Vehicle-to-Everything (V2X) Based on LTE and NR*, document ETSI TR 137 985, ETSI, Sophia Antipolis, France, 2020.

[9] B. C. Nguyen, X. N. Tran, and L. T. Dung, "On the performance of roadside unit-assisted energy harvesting full-duplex amplify-and-forward vehicle-to-vehicle relay systems," *AEU-Int. J. Electron. Commun.*, vol. 123, Aug. 2020, Art. no. 153289.

[10] S. Wang, D. Wang, C. Li, and W. Xu, "Full duplex AF and DF relaying under channel estimation errors for V2V communications," *IEEE Access*, vol. 6, pp. 65321–65332, 2018.

[11] Y. Ai, F. A. P. de Figueiredo, L. Kong, M. Cheffena, S. Chatzinotas, and B. Ottersten, "Secure vehicular communications through reconfigurable intelligent surfaces," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 7272–7276, Jul. 2021.

[12] Advantech. (Oct. 2020). *Intelligent Roadside Units for Smarter Traffic Management*. [Online]. Available: https://www.advantech.com/en/resources/white-papers/intelligent-roadside-units-for-smarter-traffic-management

[13] F. Busacca, C. Grasso, S. Palazzo, and G. Schembra, "A smart road side unit in a microeolic box to provide edge computing for vehicular applications," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 1, pp. 194–210, Mar. 2023.

[14] C. Tang, C. Zhu, X. Wei, Q. Li, and J. J. P. C. Rodrigues, "Task caching in vehicular edge computing," in *Proc. IEEE Conf. Comput. Commun. Workshops*, May 2021, pp. 1–6.

[15] H. Ye, G. Y. Li, and B. F. Juang, "Deep reinforcement learning based resource allocation for V2V communications," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3163–3173, Apr. 2019.

[16] H. Yang, X. Xie, and M. Kadoch, "Intelligent resource management based on reinforcement learning for ultra-reliable and low-latency IoV communication networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4157–4169, May 2019.

[17] X. Zhang, M. Peng, S. Yan, and Y. Sun, "Deep-reinforcement-learning-based mode selection and resource allocation for cellular V2X communications," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6380–6391, Jul. 2020.

[18] Y. Yuan, G. Zheng, K.-K. Wong, and K. B. Letaief, "Meta-reinforcement learning based resource allocation for dynamic V2X communications," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 8964–8977, Sep. 2021.

[19] P. Xiang, H. Shan, M. Wang, Z. Xiang, and Z. Zhu, "Multi-agent RL enables decentralized spectrum access in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 10750–10762, Oct. 2021.

[20] B. Gu, W. Chen, M. Alazab, X. Tan, and M. Guizani, "Multiagent reinforcement learning-based semi-persistent scheduling scheme in C-V2X mode 4," *IEEE Trans. Veh. Technol.*, vol. 71, no. 11, pp. 12044–12056, Nov. 2022.

[21] A. S. Kumar, L. Zhao, and X. Fernando, "Multi-agent deep reinforcement learning-empowered channel allocation in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 1726–1736, Feb. 2022.

[22] H. Jin, J. Seo, J. Park, and S. C. Kim, "A deep reinforcement learning-based two-dimensional resource allocation technique for V2I communications," *IEEE Access*, vol. 11, pp. 78867–78878, 2023.

[23] I. Lee and D. K. Kim, "Resource allocation in NR-V2X mode 2 using multi agent DQN," in *Proc. 14th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Jul. 2023, pp. 17–19.

[24] A. Guerna, S. Bitam, and C. T. Calafate, "Roadside unit deployment in Internet of Vehicles systems: A survey," *Sensors*, vol. 22, no. 9, p. 3190, Apr. 2022.

[25] L. Miao, S.-F. Chen, Y.-L. Hsu, and K.-L. Hua, "How does C-V2X help autonomous driving to avoid accidents?" *Sensors*, vol. 22, no. 2, p. 686, Jan. 2022.

[26] *V2X Sensor-Sharing for Cooperative and Automated Driving*, SAE Standard J3224, Aug. 2022.

[27] 5GAA. *C-V2X in Action*. Accessed: Jun. 4, 2023. [Online]. Available: https://5gaa.org/c-v2x-in-action

[28] 5GAA. (Jan. 2023). *Conclusions and Recommendations for Communications Service Providers Supporting Road Operator Priorities and Expectations 2*. [Online]. Available: https://5gaa.org/content/uploads/2023/01/5gaa-conclusions-and-recommendations-for-communications-service-providers-supporting-road-operator-priorities-and-expectations-2.pdf

[29] 5GAA. (Oct. 2021). *Study of Spectrum Needs for Safety Related Intelligent Transportation Systems—Day 1 and Advanced Use Cases*. [Online]. Available: https://5gaa.org/content/uploads/2021/10/5GAA_Day1_and_adv_Use_Cases_Spectrum_Needs_Study_V2.0.pdf

[30] C2CCC. *TR 2052—Survey on CAM Statistics*. [Online]. Available: https://www.car-2-car.org/fileadmin/documents/General_Documents/C2CCC_TR_2052Survey_on_CAM_statistics.pdf

[31] *V2X Communications Message Set Dictionary*, SAE Standard J2735, Nov. 2022.

[32] A. Correa, F. Andert, R. Blokpoel, N. Wojke, G. Thandavarayan, B. C. Perales, and C. Böker, "TransAID deliverable 5.2: V2X-based cooperative sensing and driving in transition areas (second iteration)," Universidad Miguel Hernandez (UMH), Elche, Spain, Tech. Rep. 2.0, 2020. [Online]. Available: https://elib.dlr.de/140794/

[33] *Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 3: Specifications of Decentralized Environmental Notification Basic Service*, document ETSI EN 302 637-3, Sophia Antipolis, France, Apr. 2019.

[34] *Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 2: Specification of Cooperative Awareness Basic Service*, document ETSI EN 302 637-2, ETSI, Sophia Antipolis, France, Jan. 2019.

[35] L. Buşoniu, R. Babuška, and B. De Schutter, ''Multi-agent reinforcement learning: An overview,'' in *Innovations in Multi-Agent Systems and Applications*, vol. 310, D. Srinivasan and L. C. Jain, Eds. Berlin, Germany: Springer, 2010, pp. 183–221.

[36] K. Zhang, Z. Yang, H. Liu, T. Zhang, and T. Basar, ''Fully decentralized multi-agent reinforcement learning with networked agents,'' in *Proc. ICML*, 2018, pp. 5872–5881.

[37] *NR; Multiplexing and Channel Coding (V16.3.0, Release 16)*, 3GPP, Standard TS 38.212, Dec. 2020.

[38] *Intelligent Transport Systems (ITS); Access Layer; Part 1: Channel Models for the 5,9 GHz Frequency Band*, document ETSI TR 103 257-1, ETSI, Sophia Antipolis, France, May 2019.

[39] *Radio Frequency (RF) System Scenarios (V13.0.0, Release 13)*, 3GPP, document TR 36.942, Jan. 2016.

[40] *Consideration on the System-Performance Evaluation of HSDPA Using OFDM Modulation*, 3GPP, document TS 25.211 V3.1.0 R1-030999, Ericsson, Oct. 2003.

[41] P. J. Freire, Y. Osadchuk, B. Spinnler, A. Napoli, W. Schairer, N. Costa, J. E. Prilepsky, and S. K. Turitsyn, ''Performance versus complexity study of neural network equalizers in coherent optical systems,'' *J. Lightw. Technol.*, vol. 39, no. 19, pp. 6085–6096, Oct. 15, 2021.

INSUNG LEE received the B.S. degree in information and communication engineering from Inha University, Incheon, South Korea, in 2022, where he is currently pursuing the M.S. degree. His research interests include link/system level performance evaluation and algorithm development in wireless systems and next-generation wireless systems, such as 5G new radio, V2X communication, and underwater acoustic communication.

DUK KYUNG KIM (Member, IEEE) received the B.S. degree in electrical engineering from Yonsei University, Seoul, South Korea, in 1992, and the M.S. and Ph.D. degrees from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 1994 and 1999, respectively. From 1999 to 2000, he was a Postdoctoral Researcher with Wireless Laboratories, NTT DoCoMo, Japan. From 2000 to 2002, he was with the Research and Development Center, SK Telecom, South Korea, where he was involved in the standardization of the third generation partnership project long-term evolution (3GPP-LTE) and in fourth-generation system development. Since 2002, he has been with Inha University, Incheon, South Korea. From 2009 to 2010, he was a Guest Researcher with NIST, Gaithersburg, MD, USA. His research interests include mobility management, radio resource management in 5G and 6G communication, V2X, A2X, IRS, and AI-enabled communications and underwater acoustic communications.