Proceedings of Engineering and Technology Innovation, vol. 29, 2025, pp. 47-60

Risk-Aware Multi-Agent Advantage Actor-Critic Based Resource Allocation for C-V2X Communication in Cellular Networks

Irshad Khan*, Manjula Sunkadakatte Haladappa

Department of Computer Science and Engineering, University of Visvesvaraya College of Engineering, Bengaluru, India

Received 15 August 2024; received in revised form 03 December 2024; accepted 05 December 2024

DOI: https://doi.org/10.46604/peti.2024.14136

Abstract

Intelligent transportation systems have emerged promisingly for industries to enable automated and safe driving. However, to satisfy reliability, environmental sustainability, and overall performance, deep reinforcement learning requires massive energy consumption with its computational demands. In this research, the risk-aware multi-agent advantage actor-critic (RA-MA-A2C)-based resource allocation (RA) is proposed for the cellular-vehicle-to-everything (C-V2X) network. The RA-MA-A2C considers collision risk when allocating resources such as frequency and power. By integrating risk assessment into the decision-making process, the RA-MA-A2C adjusts RA to mitigate collision risks and thereby increases the system's effectiveness. The RA-MA-A2C's performance is evaluated in terms of the success rate, completion time, vehicle-to-infrastructure link sum rate, and vehicle-to-vehicle links probability. The RA-MA-A2C demands 108 ms completion time with a 98.81% success rate, surpassing the performance of the existing offloading resource allocation-based deep reinforcement learning (ORAD) method.

Keywords: deep reinforcement learning, intelligent transportation systems, multi-agent, resource allocation, risk-aware

1. Introduction

Vehicular network is an enabling technology for smart vehicles and autonomous driving, offering different on-board data services [1]. This is a key technique that improves transportation by facilitating cooperation between vehicles apropos of proximity and ensures suitable quality of service (QoS) [2]. Dedicated short-range communications (DSRC) and intelligent transportation systems (ITS) rely on the IEEE 802.11 standard [3]. This standard evaluates inter-vehicle collaboration channels to obtain channel state information (CSI) in vehicle-to-everything (V2X) networks, enabling secure data transmission [4-5]. However, associated vehicles pose challenges in mobility management, scalability, and consistent QoS due to their inherent design challenges [6], primarily arising mainly from the design of link structures and physical layers which provide less flexible transport capabilities [7]. This issue is addressed through a standard 3rd generation partnership project (3GPP) which assists in different QoS needs of V2X networks. It also employs device-to-device (D2D) communication in long-term evolution (LTE) and cellular 5G networks [8-9].

Hence, this study focuses on resource allocation (RA) in vehicular networks according to the 3GPP standard, which includes the shared frequency spectrum for vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) links, and the utilization of proximity communication 5 (PC5) and user equipment to universal terrestrial radio access network (Uu) radio boundaries [10]. There are several inherent drawbacks in wireless communication which include hostile channels, an increasingly crowded spectrum, and a linear increase in the number of vehicular devices. On the other hand, excessive

* Corresponding author. E-mail address: research.irshad@gmail.com

English language proofreader: Chih-Wei Chang

flexibility of these systems enables numerous configurations under various conditions, leading to inconsistencies [11-12]. Deep reinforcement learning (DRL) has been implemented in V2V networks to effectively handle challenges and difficulties in large-scale data RA and decision-making under ambiguity [13-15]. Certainly, numerous V2V services rely on DRL methods to analyze and process the collected vehicle data from the cloud to leverage resources for better performance [16-17]. However, the associated vehicles pose a challenge in performance, network management, and effectiveness.

Vehicular networks represent robust dynamics in traffic patterns, propagation channels, and network topologies. Furthermore, the rising popularity of mobile applications like in-vehicle infotainment has brought to the table unprecedented demands in the infrastructure of wireless networks. Therefore, mobile network operators (MNOs) gather a considerable amount of heterogeneous data which manage the network performance and offer better service [18]. The present network management approaches have to maintain the operational state while obligating to different service level agreements (SLAs), which has become an increasingly intricate task. A solution to this problem is establishing intelligence in the network [19]. The time and power constraints of DRL affect the overall performance and the environment. Researchers hence aim to address these limitations for ITS.

Hou et al. [20] introduced joint computation offloading resource allocation-based deep reinforcement learning (ORAD) in cellular-vehicle-to-everything (C-V2X) networks, which monitored dynamic changes and diversity of computation offloading patterns in vehicular environments. RA and dynamic computation offloading were the sequential decision issues that arose during the Markovian decision-making process. The introduced ORAD enabled automated decision-making with a deterministic policy gradient approach which increased the offloading success rate effectively. However, extracting nonlinear feature vectors requires substantial memory and processing time due to complex high-dimensional processing data.

Mafuta et al. [21] developed a multi-agent double deep Q-network (MA-DDQN) to maintain system integrity and increase the capacity of V2I links. The MA-DDQN effectively handled the nonstationary environment by determining various joint actions that offered improved outcomes. However, the critical parameter of adaptability was not considered in vehicular networks due to the dynamic and rapidly changing nature of the environments. Lee and Kim [22] presented a decentralized multi-agent deep Q-network (MADQN)-based RA for heterogeneous traffic in V2X communications. To minimize system complexity, roadside units (RSUs) were utilized as a group of virtual agents in a minimized action space, rather than a single agent which selected numerous sources simultaneously. The flexibility to change resource availability and demands rendered the MADQN approach effective for the V2X environment. The presented approach involved an RSUs with multifarious virtual agents selected using a single resource which reduced the size of the action space with increased learning effectiveness.

Nevertheless, managing decision-making under uncertainties posed significant challenges due to the unpredictable nature of various agents which resulted in non-stationary environments. Waqas et al. [23] implemented a duplex deep reinforcement learning (DDRL)-based RA for V2X communications. Specifically, DDRL incorporated numerous network parameters and effectively utilized resources during the learning stage to address RA issues. The main focus of this approach was to handle RA by considering the performance measures: system throughput, latency, packet received rate, and network utilization to ensure effective RA and optimal performance. The DDRL utilized a deep Q-learning approach which handled the band resource for the next-generation V2X network. However, DDRL did not consider completion time and success rate, which diminished its performance.

Alrubaye and Ghahfarokhi [24] developed a geo-based RA for joint clustered V2V and V2I communications. The V2I vehicles were integrated through density-based spatial clustering of applications with noise (DBSCAN) associated with cluster heads (CHs). In hypothetical scenarios, CHs allocated resource blocks towards the base station (BS). The RA scheme considered the data on V2I routing clusters which retrieved relevant V2V resources for the deployment of V2I communications. Nevertheless, it was non-linear and non-convex due to changes in decision variables for RA, which affected the model's performance.

Wan et al. [25] implemented an improved biogeography-based optimization (IBBO) to minimize the execution time of vehicular ad hoc networks (VANETs) based on RA and joint offloading tasks. The multi-task and multi-vehicle networks were constructed without utilizing sensors in the RSUs edge layer. The other vehicles' computational resources were applied in the device layer based on RA and computation offloading allocation approach which provided a better performance. However, the overall completion time was increased as the model did not assess the success rate.

The existing methods have drawbacks such as not considering adaptability, which is a substantial parameter in vehicular networks. Additionally, the completion time and success rate are not considered, in turn diminishing the performance, and hindering the ability to handle decision-making under uncertainty. These methods need large memory and execution time for extracting nonlinear feature vectors.

The main contributions of this article are as follows:

(1) The research proposes a risk-aware multi-agent advantage actor-critic (RA-MA-A2C), which focuses on RA for C-V2X networks. This allocation addresses challenges associated with DRL, especially in minimizing the completion time.
(2) The RA-MA-A2C completes the tasks at a lesser completion time as the vehicles near the RSUs require less communication time.
(3) RA mitigates collision by integrating risk assessment into the decision-making process.

The remainder of this paper is organized as follows: Section 2 details the proposed RA-MA-A2C approach employed for C-V2X communication, Section 3 represents the empirical results simulated in a Python environment, and Section 4 presents the key findings and highlights the methodological significance of this study.

2. Methodology

The proposed RA-MA-A2C model is employed for RA for C-V2X communication in cellular networks. The approach addresses DRL issues by improving the completion time and success rate by incorporating risk awareness. The system uses power domain non-orthogonal multiple access (PD-NOMA), where every subcarrier is allocated to more than one user. Technically, decoding the strongest signal improves the clarity and reliability of data transmission as robust signals are arranged for processing. Eliminating weaker signals minimizes interference and transmission errors, thereby enhancing the overall system performance. This selective decoding method improves signal-to-noise ratio (SNR) which results in the efficient usage of available resources and enhanced service quality in C-V2X. The RA process of the proposed method, which includes three vehicles, V1, V2, and V3, with three D2D pairs, is illustrated in Fig. 1.

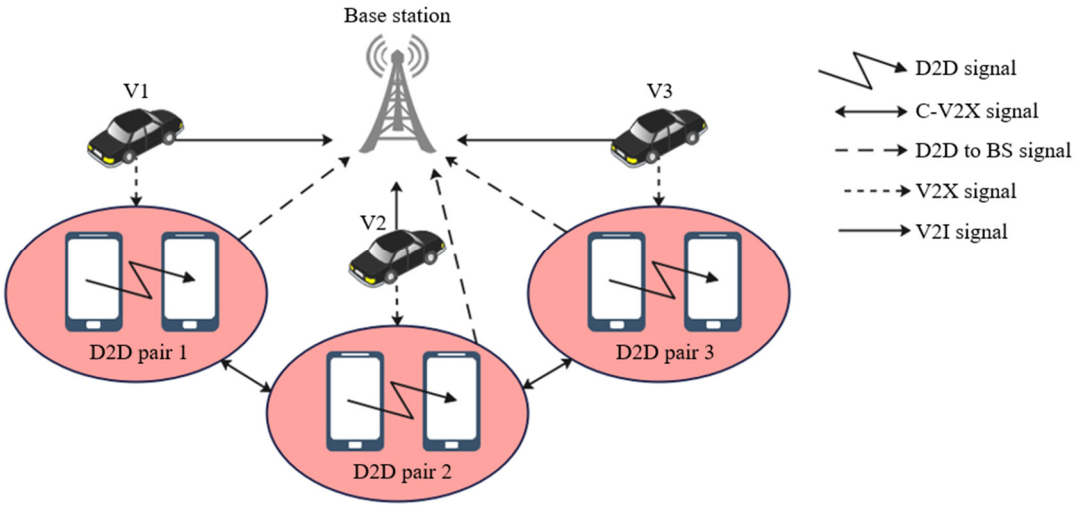

Fig. 1 C-V2X Scenario in cellular networks

2.1. System model

As presented in Fig. 1, it is assumed that the PD-NOMA single-cell system contains vehicles and D2D pairs. PD-NOMA is essential as it enables users to share similar subcarriers depending on power levels which increase the spectral effectiveness and enhance concurrent connections. This approach is especially beneficial for high-density networks as it increases resource utilization without necessitating extra bandwidth. Moreover, in Fig. 1, it is also observable that a single BS handles communication among the vehicles and D2D pairs, alongside handling V2I, V2V, and interference among the D2D signals.

The complexity of handling interference and RA within the single BS coverage is also addressed in the methodology. With multiple BS, the complexity is increased with the coverage of dispersed regions, their associated vehicles, and D2D pairs. The deployment of an omnidirectional antenna is crucial to transmit and receive signals in all directions with 360-degree coverage. The multi-BS is simplified for single BS through aggregating the traffic and coverage of multi-BS into a single model which reduces the computational complexity, rendering it suitable for optimization.

There are L vehicles in set $V = \{v_1, v_2, \cdots, v_L\}$, where L is the number of vehicles that allocate uplink sources through D2D pairs. The set of devices is represented as $D = \{d_1, d_2, \cdots, d_M\}$, where M is the number of D2D pairs. Concerning C-V2X frequency, the binary variable $x_{d_i,u}$ indicates if the subcarrier u is allocated to the device d_i. If $x_{d_i,u} = 1$, subcarrier u is allocated to the device d_i, else $x_{d_i,u} = 0$. Similarly, $y_{v_i,u}$ is a binary variable that defines subcarrier allocation for vehicles. The set of each subcarrier is presented as U, and B is the available bandwidth divided into $|U|$ subcarriers, each with bandwidth $b = B/|U|$. In this process, the signal with higher strength is interpreted and subtracted from the integrated signal, while the weaker signal is eliminated. Successive interference cancellation (SIC) is effectively accomplished for the l^{th} vehicle, with m^{th} D2D pair signals for decoding, as presented:

$$\begin{cases} \forall u \in U \\ \forall l, m \in V \bigcup D,\ l \neq m \\ \left|h_{l,u}\right|^2 > \left|h_{m,u}\right|^2 \end{cases} \tag{1}$$

where U represents the set of all subcarriers. This paper focuses on intra-cell interference produced by users who allocate similar frequency bands. The system experiences interference due to similar spectrum allocation, with the types of interferences described below:

(1) Similarly assigning the same spectrum sources causes interferences between vehicles and their respective D2D pairs.
(2) When the system transmitter allocates similar spectrum resources in the C-V2X network, it incurs interference between signals received at the BS from the vehicle v_i with interference created by D2D.
(3) Allocating similar spectrum resources causes interference with the D2D pair signal present in the D2D pair d_i, with vehicle v_i, and with other C-V2X links. The received interference power at the vehicle v_i on subcarrier u is determined by:

$$I_{v_i,u}^{pow} = \sum_{\substack{d_i \in D, \\ \left|h_{v_i,b,u}\right|^2 < \left|h_{d_i,b,u}\right|^2}} x_{d_i,u} \left|h_{d_i,b,u}\right|^2 G_{v_i}, G_b L_{d_i,b,u}^{-\beta} P_{d_i,u} + \sum_{\substack{v_i, v_j \in V, v_i \neq v_j, \\ \left|h_{v_i,b,u}\right|^2 < \left|h_{v_j,b,u}\right|^2}} y_{v_j,u} \left|h_{v_j,b,u}\right|^2 G_{v_j}, G_b L_{v_j,b,u}^{-\beta} P_{v_j,u} \tag{2}$$

where G is the antenna gain, L is the path loss, P is transmit power, and $h_{v_i,b,u}$ is the channel coefficient for the Gaussian random variable between the D2D pair d_i, the BS on subcarrier u with zero mean, and unit variance. Furthermore, $h_{d_i,b,u}$ represents channel gain between the device d_i and the BS b on communication channel u. The variables $x_{d_i,u}$ and $y_{v_j,u}$ correspond to the subcarrier assignments of the D2D pair and vehicle, respectively. $L_{d_i,b,u}^{-\beta}$ represents the distance between the

transmitter of the D2D pair d_i and BS b on the subcarrier u, with a path loss exponent β. $P_{v_j,u}$ and $P_{d_i,u}$ are the available transmit power values for vehicle v_j and D2D pair d_i on subcarrier u, respectively. The terms G_{v_i} and G_b symbolize the transmit and receive antenna gains for vehicle v_i with i^{th} index vehicle, and the BS b, respectively.

The signal-to-interference-noise ratio (SINR) for vehicle v_i on subcarrier u is detailed below the formula. SINR is a measure of signal quality that compares the desired signal power to the combined interference power from the background signal. It quantifies the clarity of the received signal, enabling efficient data clarity.

$$\varphi_{v_i,u} = \frac{\left|h_{v_i,b,u}\right|^2 G_{v_i} G_b L_{v_i,b,u}^{-\beta} P_{v_i,u}}{I_{v_i,u}^{pow} + N_{v_i,u^b}} \tag{3}$$

where N_{v_i,u^b} signifies noise power that impacts the signal on subcarrier u for vehicle v_i from BS b. The received interference power at the D2D pair d_i on subcarrier u is determined:

$$\begin{aligned} I_{d_i,u}^{pow} = & \sum\nolimits_{\left|h_{d_i,d_i,u}\right|^2 < \left|h_{d_i,v_i,u}\right|^2}^{v_i \in V} y_{v_i,u} \left|h_{d_i,v_i,u}\right|^2 G_{d_i}^2 L_{d_i,v_i,u}^{-\beta} P_{v_i,u} \\ & + \sum\nolimits_{\left|h_{d_i,d_i,u}\right|^2 < \left|h_{d_i,d_j,u}\right|^2}^{d_i,d_j \in D, i \neq j} x_{d_i,u} \left|h_{d_i,d_j,u}\right|^2 G_{d_i} G_{d_j} L_{d_i,d_j,u}^{-\beta} P_{d_j,u} \end{aligned} \tag{4}$$

$h_{d_i,d_j,u}$ is the channel coefficient for a Gaussian random variable between a D2D pair d_i and vehicle v_i with zero mean and unit variance. The variables $x_{d_i,u}$ and $y_{v_i,u}$ indicate the subcarrier assignments for the D2D pair d_i and vehicle v_i, respectively. $L_{d_i,v_i,u}^{-\beta}$ denotes the distance between the transmitter of the D2D pair d_i and vehicle v_i on subcarrier u, with a path loss exponent β. $P_{v_i,u}$ and $P_{d_j,u}$ are the available transmit powers for vehicle v_i and the D2D pair d_j on subcarrier u, respectively. The terms G_{d_i} and G_{b_j} determine the transmit and receive antenna gains for the D2D pairs d_i and d_j.

The SINR of D2D pair d_i on subcarrier u is shown in:

$$\varphi_{d_i,u} = \frac{\left|L_{d_i,d_j,u}\right|^2 G_{d_i}^2 L_{d_i,d_j,u}^{-\beta} P_{d_i,u}}{I_{d_i,u}^{pow} + N_{d_i,u^b}} \tag{5}$$

where, N_{d_i,u^b} characterizes the Gaussian noise with spectral density. In the C-V2X environment, the receiver of the D2D pair d_i is affected by interference from the vehicle v_i due to the allocation of similar spectrum resources. Therefore, the RA for vehicle users in the C-V2X environment aims to enhance the system's energy efficiency, as discussed in the section below.

2.2. Optimization framework

This part allocates resources efficiently to each user while ensuring QoS requirements for both vehicles and D2D pairs in C-V2X. The optimization problem of outage-based energy efficiency is presented in:

$$E = \frac{b\left(\sum_{d_i \in D} \log(1+\varphi_{d_i,u}) + \sum_{v_j \in V} \log(1+\varphi_{v_j,u})\right)}{\sum_{d_i \in D} P_{d_i,u} + \sum_{v_j \in V} P_{v_j,u}} \tag{6}$$

where E is energy efficiency, b signifies the subcarrier bandwidth, $\varphi_{d_i,u}$ and $\varphi_{v_j,u}$ are the SINRs of the D2D pair d_i and vehicle v_j on subcarrier u, respectively. $P_{v_j,u}$ and $P_{d_i,u}$ are the transmitted powers for vehicle v_j and D2D pair d_i on subcarrier u, respectively. The system's constraints are defined in below Section 2.3. Fig. 2 shows the architecture of the RA-MA-A2C model.

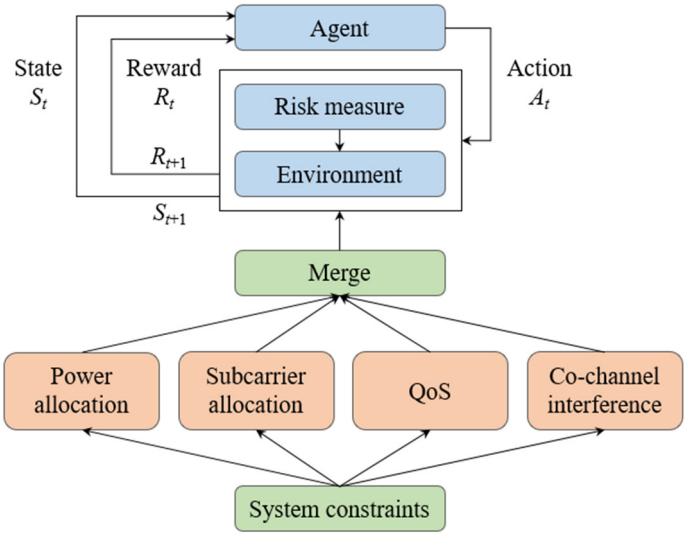

Fig. 2 Architecture of RA-MA-A2C model

2.3. *Reinforcement learning*

The three states: agent, risk measure, and environment are explained in the following sub-sections. The agent is a decision-making entity in RL that selects the actions A_t based on the current state S_t to maximize cumulative rewards. The external system interacts with the agent by receiving actions and generating new states S_{t+1} and rewards R_{t+1} as feedback.

2.3.1. *System constraints*

The RA-MA-A2C model presents a new change in resource allocation by integrating a risk-aware mechanism for C-V2X networks. When compared to existing DRL approaches, the RA-MA-A2C model adapts to the fluctuating demands of vehicular networks dynamically by addressing critical challenges like resource contention and latency. For example, the incorporation of risk metrics such as time headway (THW) and time to threshold crossing (TTC) permits the proposed RA-MA-A2C model to solve and mitigate the potential collisions, which ensures higher reliability and safety. Moreover, the proposed RA-MA-A2C model utilizes a PD-NOMA which allows multiple users to share similar subcarriers to improve spectral efficiency. This PD-NOMA is beneficial for especially high-density networks for significant challenges possessed by the limitations of bandwidth. Multiple system parameters are used to design the system constraints, ensuring the secure operation of the power system. The parameters include power allocation, subcarrier allocation, QoS, and co-channel interference constraints.

2.3.2. *Power allocation constraints*

Power allocation determines the optimal power for each subcarrier, even when the subcarriers operate autonomously. The parameters $P_{d_i,u}$ and $P_{v_i,u}$ are required to satisfy the constraints, as presented in:

$$P_{d_i,u} \geq 0,\ \forall d_i \in D,\ \forall u \in U \tag{7}$$

$$\sum_{u \in U} x_{d_i,u} P_{d_i,u} \leq P_{d_i,u}^{\max},\ \forall d_i \in D,\ \forall u \in U,\ x_{d_i,u} \in \{0,1\} \tag{8}$$

$$P_{v_i,u} \geq 0,\ \forall v_i \in V,\ \forall u \in U \tag{9}$$

$$\sum_{u \in U} y_{v_i,u} P_{v_i,u} \leq P_{v_i,u}^{\max},\ \forall v_i \in V,\ \forall u \in U,\ y_{v_i,u} \in \{0,1\} \tag{10}$$

where $P_{d_i,u}$ and $P_{v_i,u}$ are the available transmit powers for v_i and d_i on subcarrier u, respectively. Eqs. (8) and (10) correspondingly denote the maximum threshold transmit powers $P_{d_i,u}^{max}$ and $P_{v_i,u}^{max}$ for each D2D pair and vehicle user.

2.3.3. Subcarrier allocation constraints

Subcarrier allocation is used to determine which subcarriers are assigned to users and to identify the primary user, considering factors of user importance, data rate requirements, and channel conditions. This process is described following:

$$x_{d_i,u}, y_{v_i,u} \in \{0,1\},\ \forall d_i \in D,\ \forall v_i \in V,\ \forall u \in U \tag{11}$$

$$\sum_{u \in U} x_{d_i,u} \leq 1,\ \forall d_i \in D,\ \forall u \in U \tag{12}$$

Here, $x_{d_i,u}$ and $y_{v_i,u}$ represent the subcarrier assignments for D2D pairs and vehicles, respectively. Eq. (11) defines the binary variables for D2D pair and vehicle subcarrier assignments, with the constraint in Eq. (12) ensuring that each D2D pair is allocated to a single subcarrier. SIC enables the recycling of each subcarrier, as numerically formulated.

$$\sum_{d_i \in D} x_{d_i,u} + \sum_{v_i \in V} y_{v_i,u} \leq U^{\max},\ \forall u \in U,\ x_{d_i,u}, y_{v_i,u} \in \{0,1\} \tag{13}$$

$U^{\max}$ symbolizes the maximum number of users per subcarrier, $x_{d_i,u}$ and $y_{v_i,u}$ denote the subcarrier assignments respectively for D2D pairs and vehicles. The system complexity increases with $U^{\max}$ which is based on the signal processing delay in SIC and the receiver complexity.

2.3.4. QoS constraints

The QoS constraints ensure that the system supports multiple synchronized applications, each operating at different quality levels based on the available system resources. The QoS for all users is determined by the minimum SINR required for D2D pairs and users.

$$\varphi_{v_i,u} \geq \rho_v,\ \forall v_i \in V,\ \forall u \in U \tag{14}$$

$$\varphi_{d_i,u} \geq \rho_d,\ \forall d_i \in D,\ \forall u \in U \tag{15}$$

Here, ρ_v and ρ_d denote the minimum SINR required respectively for vehicle v_i and D2D pair d_i on subcarrier u.

2.3.5. Co-channel interference constraints

Interference constraints are incorporated into the model to restrict co-channel interference between different links (V2V and V2I) that share the same subcarriers in the C-V2X network. The formulated constraints are:

(1) Co-channel interference constraints for V2I links: For each vehicle user v on the subcarrier u, the interferences from D2D pairs sharing the same subcarrier, maintain a predefined threshold as shown in the following:

$$\sum_{d=1}^{M} x_{d_i,u} \times g_{d_i,v_j,u} \times \frac{1}{l_{d_i,v_j}} \times P_{d_i,u} \leq I_{V2I}^{thre} \tag{16}$$

where $x_{d_i,u}$ is a binary subcarrier assignment variable for the D2D pair d_i, $g_{d_i,v_i,u}$ is the channel gain between the D2D transmitter d_i and vehicle v_i on the subcarrier u, l_{d_i,v_j} is the distance between the D2D transmitter d_i and vehicle v_j, $P_{d_i,u}$ is the transmit power of the D2D pair d_i on subcarrier u, and I_{V2I}^{thre} indicates a threshold value for the interference or in a V2I communication.

(2) Co-channel Interference Constraints for D2D Links: Regarding each D2D pair d_i on subcarrier$u$, the interference from vehicles and other D2D pairs sharing the same subcarrier should be below the threshold I_{D2D}^{thre}, as mathematically expressed:

$$\sum_{v=1}^{V} y_{v_i,u} \times g_{d_i,v_j,u} \times \frac{1}{l_{d_i,v_j}} \times P_{v_i,u} + \sum_{d' \neq d} x_{d',u} \times g_{d',d,u} \times \frac{1}{l_{d',d}} \times P_{d',u} \leq I_{D2D}^{thre} \tag{17}$$

where $x_{d',u}$ is a binary subcarrier assignment variable for another D2D pair d', $g_{d',d,u}$ is the channel gain between the D2D transmitters d and d' on subcarrier u, $P_{d',u}$ is the transmit power of d' on u, and I_{D2D}^{thre} signifies threshold value for interference in D2D communication. These interference constraints ensure that the co-channel interference experienced by V2I and D2D links, remains below specified thresholds, thereby maintaining the required signal quality for reliable communications.

2.3.6. Optimization problem

The network is dynamically designed based on transmission power vectors P_v and P_d, and subcarrier allocation vectors y_v and x_d for cellular and D2D pairs, respectively.

$$\max_{P_v, P_d, y_v, x_d} E \tag{18}$$

where E represents the system's energy efficiency, with the vectors $y_v = [y_{11}^1, \cdots, y_{1N}^N, y_{21}^1, \cdots, y_{2N}^N, y_{K1}^1, \cdots, y_{LN}^N]$, $x_d = [x_{11}^1, \cdots, x_{1N}^N, x_{21}^1, \cdots, x_{2N}^N, x_{M1}^1, \cdots, x_{MN}^N]$, $P_v = [P_1^n, P_2^n, \cdots, P_L^n]$, and $P_d = [P_1^n, P_2^n, \cdots, P_M^n]$. In Eq. (18), the optimization problem involves non-convex objective functions, integers, and continuous variables. However, addressing the non-deterministic polynomial time (NP)-hard issue also resolves the convex optimization problem. Non-convex optimization issues are tackled using heuristic methods, owing to their ability to explore high and non-linear search spaces. Moreover, the formulated problem, in its current form, is scarcely solvable in a distributed way.

2.3.7. Risk-Aware deep reinforcement learning

RA-MA-A2C is utilized to address the requirement of a longer completion time with a lower success rate. The proposed model achieves the desired results in a shorter duration due to the proximity of vehicles to the RSUs, thereby minimizing communication delays. The RA-MA-A2C dynamically adjusts to undefined and fluctuating network demands. Additionally, the RA-MA-A2C prioritizes RA to minimize potential bottlenecks and enhance reliability through risk awareness. The TTC is often used as a metric to predict the time until a potential collision, while also considering the vehicles' speed. THW is another key metric that analyzes the time gap before a collision with the vehicle ahead, which is particularly relevant when considering the possibility of lane changes.

However, both TTC and THW fall short of their capabilities in dealing with two vehicles positioned in different lanes, especially when lane changes occur, and there is no vehicle in the original lane. The ego vehicle (EV) refers to the autonomous or self-driving vehicle, with a focus on perception and decision-making within the automated driving system. The EV is divided into four areas (P1, P2, P3, and P4) to facilitate effective monitoring and processing of environmental information in its surroundings. The relevant vehicles in these areas are denoted as: following vehicle (FV), left leading ego vehicle (LLEV), vehicle following vehicle (V_{FV}), vehicle ego vehicle (V_{EV}), vehicle left leading ego vehicle (V_{LLEV}), and direct left leading ego vehicle (d_{LLEV}), all of which describe interactions with these vehicles. Hence, TTC is described as the ratio of relative distance to scalar speed. The mathematical expression for TTC is given:

$$Risk\ measure = \begin{cases} TTC_f = \dfrac{d_f}{v_{EV} - v_f},\ if\ v_{EV} > v_f,\ f = LLEV, RLEV, PV \\ TTC_r = \dfrac{d_r}{v_f - v_{EV}},\ if\ v_f > v_{EV},\ r = LLAV, RLAV, FV \end{cases} \tag{19}$$

where TTC_f and TTC_r represent time to threshold crossing when the subject vehicle is at risk of collision with the vehicle ahead and the rear vehicle, respectively. The d_f and v_f respectively denote the distance and velocity of the front vehicle, d_r means its distance from the rear vehicle, $LLAV$ refers to a left leading adjacent vehicle, $RLEV$ symbolizes a right leading EV,

$RLAV$ signifies a right leading adjacent vehicle, and PV is defined as a preceding vehicle. While TTC serves as a valuable indicator for predicting potential collisions, relying solely on a single risk measure is insufficient for evaluating the driving risk in several conditions.

To address this limitation and provide a more comprehensive assessment of the driving risk, another risk measure, THW is introduced. THW is mathematically expressed in:

$$Risk\ measure = \begin{cases} THW_f = \dfrac{d_f}{v_{EV}},\ f = LLEV, RLEV, PV \\ THW_r = \dfrac{d_f}{v_f},\ r = LLAV, RLAV, FV \end{cases} \tag{20}$$

The THW_f and THW_r denote time headways relative to the front and rear vehicles respectively. When the EV is driving in the right lane, it is considered that the risk indicators P1 and P3 are out of lane boundaries. On the other hand, when the vehicle is placed in the left lane, P2 and P4 are considered to be '0'. However, given their significant differences, the sigmoid normalization function is expressed:

$$sigmoid(x) = \frac{1}{1+e^{-x}} \tag{21}$$

Additionally, the equivalent rate of risk denotes manifold risk levels where TTC and THW are standardized, as given below.

$$F_i = 1 - sigmoid(\alpha(T_i - \beta)),\ i = TTC, THW \tag{22}$$

where α is a shape factor that controls the charge rate, increasing the sharpness of the curve as it rises, and β indicates the equivalent rate of individual variables. T_i (where i represents TTC or THW) denotes the output values before normalization. Table 1 shows the hyperparameters and their ranges in the actor-critic network.

Table 1 Hyperparameter and its values in actor-critic network

Parameters	Values
Actor learning rate	0.001
Critic learning rate	0.0001
Minibatch size	64
Target update factor	0.001
Target update rate	100

The risk for EVs is assessed by integrating various risk metrics and employing the softmax function.

$$\text{Softmax}(x_1, x_2) = \left[\frac{e^{x_1}}{\sum_{i=1}^{2} e^{x_i}}, \frac{e^{x_2}}{\sum_{i=1}^{2} e^{x_i}} \right] \tag{23}$$

Consequently, the comprehensive risk index (CRI) between two vehicles is calculated in:

$$\text{CRI} = \text{Softmax}(F_{TTC}, F_{THW}) \left[\frac{F_{TTC}}{F_{THW}} \right] \tag{24}$$

where x_1 and x_2 are input values, and e^{x_1} and e^{x_2} are the exponential functions of x_1 and x_2, respectively. F_{TTC} and F_{THW} are the scaling factors based on TTC and THW, respectively. The RA-MA-A2C model improves the exploration strategies in RA by guiding agents to explore regions with high uncertainties. Fig. 3 depicts the flowchart for the RA-MA-A2C approach.

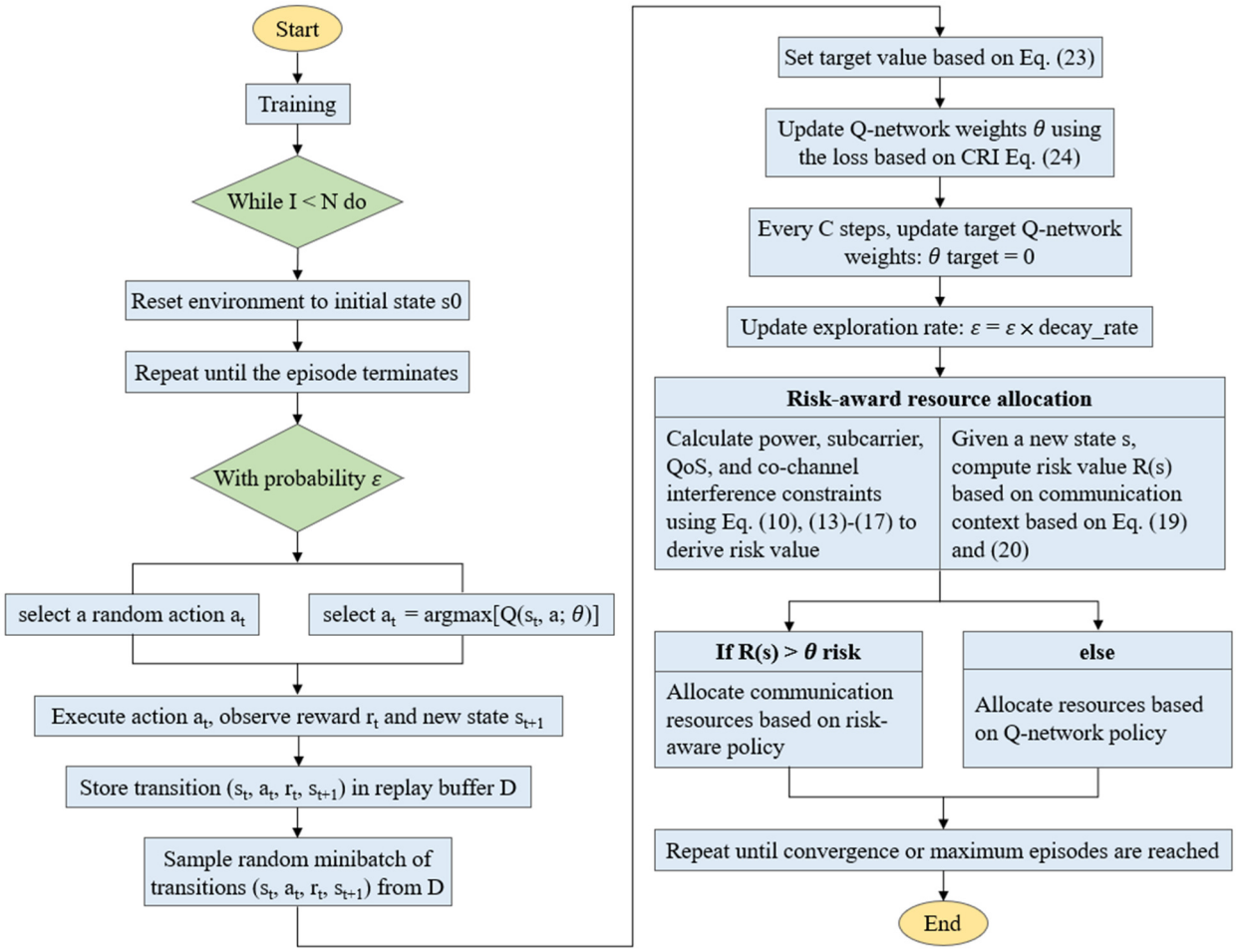

Fig. 3 Flowchart for RA-MA-A2C

3. Results and Analysis

The results obtained from the proposed RA-MA-A2C applied to the RA process in V2X networks are analyzed. RA-MA-A2C exhibits a better success rate and completion time than the existing methods. V2X communication scenarios are simulated on a Python environment with 16GB RAM, Intel Core i7 processor, and Windows 10 operating system. The simulation parameters listed in Table 2 are selected to reflect realistic conditions analyzed in V2X network environments. The number of parameters and vehicle speed are chosen depending on typical urban traffic scenarios. The limit of delay time aligns with the requirements of latency for safety-critical V2X applications. The carrier frequency and bandwidth are set following typical specifications for V2X communications bands. Eventually, the number of sub-bands is set to 10 to provide a balance between RA and the computational complexity. These parameters are reliable with scenarios in C-V2X communication.

Table 2 Simulation parameter

Parameters	Values
No. of vehicles	20, 40, 60, 80, and 100
Vehicle speed	10-25 km/h
Delay time limit	100 ms
Bandwidth and carrier frequency	10 MHz and 2 GHz
No. of sub-bands	10

3.1. Performance analysis

This section presents the performance of the proposed RA-MA-A2C concerning success rate, completion time, V2I link sum rate, and V2V links probability. DRL faces challenges such as vanishing gradients, hyperparameter sensitivity, and non-convex optimization, which hamper effective training, reduce reliability and success rate, and increase completion time. Risk-

aware reinforcement learning (RARL) involves complex computations such as uncertainty evaluations, which further increases the model's computational complexity. Fig. 4 and Fig. 5 illustrate the success rate and completion time of RA-MA-A2C. The outcomes show that RA-MA-A2C attains a higher success rate of 98.81% and a lower completion time of 108ms, compared to existing techniques. RA-MA-A2C consumes less time as vehicles are proximate to RSUs and achieves a high success rate thereby minimizing the communication time.

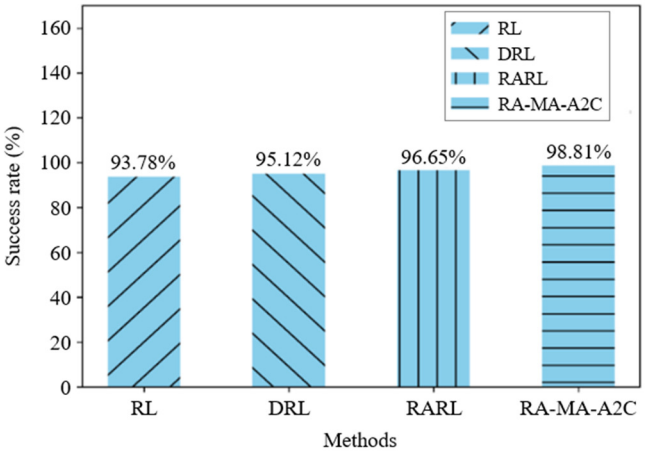

Fig. 4 Performance of success rate

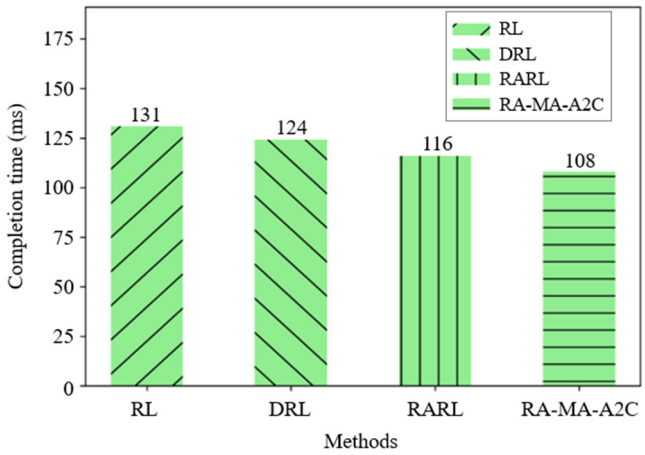

Fig. 5 Performance of completion time

Fig. 6 depicts the V2V link probability for RA-MA-A2C and a comparison with RL, DRL, and RARL. The RA-MA-A2C achieves higher V2V link probabilities of approximately 0.993, 0.989, 0.975, 0.967, and 0.961, for 20, 40, 60, 80, and 100 vehicles. As the number of vehicles increases, V2V link probability decreases across all methods. However, higher probabilities are achieved with an increasing vehicle count which provides better resource management.

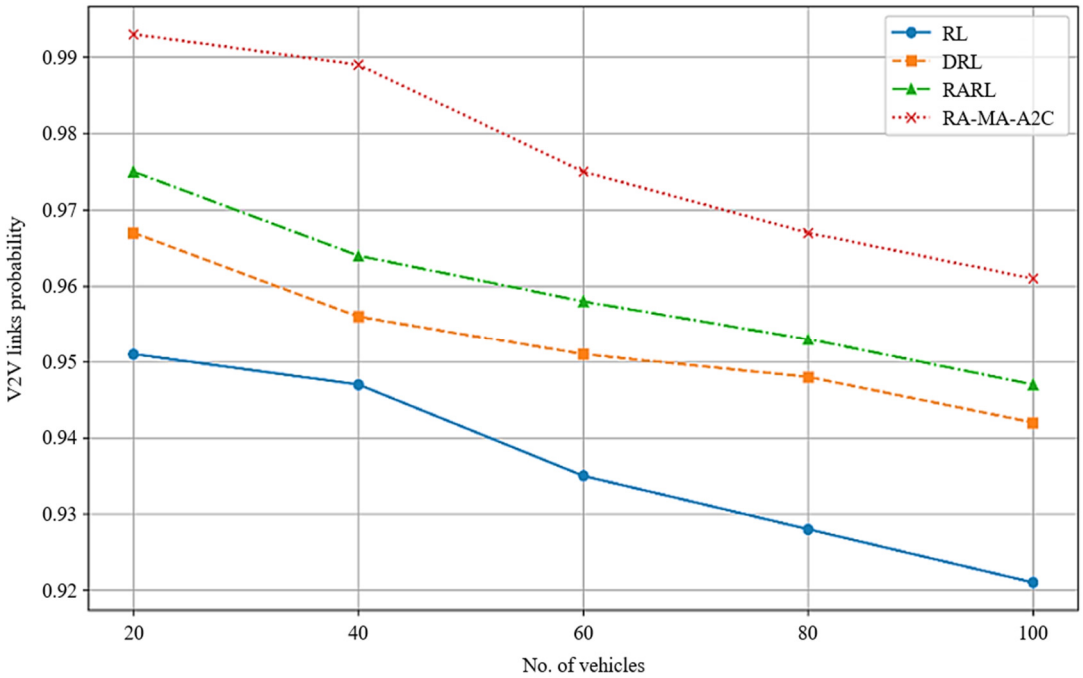

Fig. 6 Performance of V2V links probability

Fig. 7 presents the V2I link sum rate for RA-MA-A2C, and a comparison with RL, DRL, and RARL. RA-MA-A2C achieves higher V2I link sum rates of 197 Mb/s, 185 Mb/s, 179 Mb/s, 171 Mb/s, and 164 Mb/s for 20, 40, 60, 80, and 100 vehicles. As the number of vehicles increases, the sum rate of V2I links diminishes significantly for higher vehicle density, inducing increased interference for available resources and minimizing the overall data rate of V2I and communication links. The proposed RA-MA-A2C hence displays a superior output than other algorithms namely, RL, DRL, and RARL.

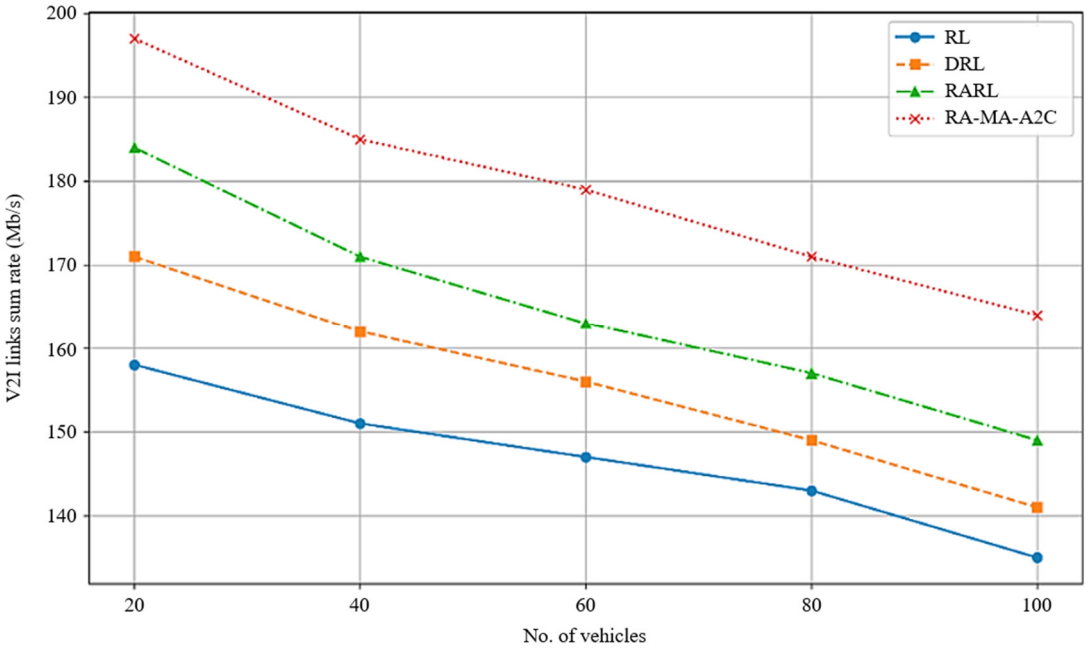

Fig. 7 Performance of V2I links sum rate

3.2. *Comparative analysis*

In this section, the performance of RA-MA-A2C is compared with existing ORAD [20] and MA-DDQN [21] methods, as presented in Tables 3 and 4. The comparison is based on the numbers of vehicles being varied between 20 and 40. In Table 3, the comparison is performed apropos of success rate and completion time. In Table 4, the comparison is performed using V2V link probability and V2I link sum rate. The observations in these resultant graphs prove that the RA-MA-A2C offers commendable results.

The results in Table 3 show the RA-MA-A2C achieves a higher success rate at a lower duration. The success rate of RA-MA-A2C is 98.81%, whereas the rate reported by Hou et al. [20] is 97.68%. The completion time for RA-MA-A2C is 108 ms, which is less than that reported by Hou et al. [20]. In Table 4, the comparison is performed using V2V link probability. RA-MA-A2C attains higher V2V link probability values of 0.993 and 0.989 for 20 and 40 vehicles compared to MA-DDQN [21]. In Table 4, results for the V2I link sum rate exhibit RA-MA-A2C's higher values of 197 Mb/s and 185 Mb/s compared to those reported by Mafuta et al. [21]. It is also observed that as the number of vehicles increases, the V2I link sum rate and V2V link probability diminish for every algorithm.

Table 3 Comparative analysis of success rate and completion time

Methods	Success rate (%)	Completion time (ms)
ORAD [20]	97.68	120
Proposed RA-MA-A2C	98.81	108

Table 4 Comparative analysis of V2V links probability and V2I links sum rate

Methods	No. of vehicles	V2V links probability	V2I links sum rate
MA-DDQN [21]	20	0.986	192
	40	0.981	168
Proposed RA-MA-A2C	20	0.993	197
	40	0.989	185

3.3. *Discussion*

The RA-MA-A2C approach's advantages, alongside the existing method's limitations, are examined in this section. Specifically, ORAD [20] requires significant memory and execution time, in MA-DDQN [21] adaptability was not considered in vehicular networks. RA-MA-A2C overcomes the aforementioned limitations by optimizing memory usage and increasing

adaptability. Besides, it is tailored to adapt to the dynamic environments by integrating risk-aware mechanisms for RA. On these bases, RA-MA-A2C is deployed to address the increased time consumption of DRL. Finally, RA-MA-A2C also enhances exploration strategies in RA by guiding agents to explore RA in regions where uncertainties are high.

Its risk-aware features present agents to ensure precise RA decisions by rapidly changing network conditions in a dynamic vehicular context. By employing multi-agent coordination, RA-MA-A2C permits vehicles to handle resources which reduces network congestion and increases stability. Moreover, RA-MA-A2C increases exploration by directing agents to focus on high-uncertainty areas which makes more balanced and effective RA over the network. This risk-awareness and stability yield better performance than the DRL approach in C-V2X environments.

3.4. Limitation

In an RA-MA-A2C, the model depends substantially on multifarious training data for learning optimal policies over a network condition and traffic range. When the environment diverges significantly from training scenarios like sudden shifts in signal interference, vehicle density, or unexcepted congestion of the network, the proposed approach struggled to generalize efficiently. This reliance on certain patterns of data results in inaccurate or suboptimal RA.

4. Conclusion

In this research, the RA-MA-A2C model was developed for resource allocation in C-V2X networks to address the limitations of traditional DRL methods. The model demonstrates enhanced adaptability and reliability in dynamic vehicular environments. Moreover, the PD-NOMA technique was utilized to improve spectral efficiency, making the model suitable for high-density networks. This RA-MA-A2C achieved a success rate of 98.81% and reduced the completion time to 108 ms, outperforming existing methods such as ORAD and MA-DDQN.

The conclusions of this research are as follows:

(1) The proposed RA-MA-A2C model for resource allocation in C-V2X networks increased the success rate as well as reduced the completion time when compared to the state-of-the-art approaches.
(2) The proposed RA-MA-A2C model with an integrated risk-aware mechanism effectively handled network uncertainties, resulting in reliable decisions in dynamic scenarios.
(3) Multi-agent coordination is incorporated with the proposed RA-MA-A2C model to optimize resource allocation in C-V2X networks, reduce network congestion, and enhancing network stability.
(4) The proposed RA-MA-A2C model enhanced the resource allocation process and attained a superior performance in V2I link sum rates and V2V link probabilities, even with increased vehicle densities.

Future work will focus on addressing the model's limitations in generalization by exploring distributed machine learning algorithms. These algorithms aim to adopt the sudden changes in vehicle density, network congestion, and signal interference, enhancing the model's scalability and broader applicability.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] 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.
[2] T. I. Bayu, Y. F. Huang, and J. K. Chen, "Performance of Fuzzy Inference System for Adaptive Resource Allocation in C-V2X Networks," Electronics, vol. 11, no. 23, article no. 4063, 2022.

[3] S. M. Waqas, Y. Tang, L. Yu, and F. Abbas, "A Joint Cluster-Based RRM and Low-Latency Framework Using the Full-Duplex Mechanism for NR-V2X Networks," Computer Communications, vol. 209, pp. 513-525, 2023.
[4] S. Sabeeh and K. Wesołowski, "Congestion Control in Autonomous Resource Selection of Cellular-V2X," IEEE Access, vol. 11, pp. 7450-7460, 2023.
[5] N. Banitalebi, P. Azmi, N. Mokari, A. H. Arani, and H. Yanikomeroglu, "Distributed Learning-Based Resource Allocation for Self-Organizing C-V2X Communication in Cellular Networks," IEEE Open Journal of the Communications Society, vol. 3, pp. 1719-1736, 2022.
[6] Y. Ding, Y. Huang, L. Tang, X. Qin, and Z. Jia, "Resource Allocation in V2X Communications Based on Multi-Agent Reinforcement Learning with Attention Mechanism," Mathematics, vol. 10, no. 19, article no. 3415, 2022.
[7] Z. Li, K. Wang, T. Yu, and K. Sakaguchi, "Het-SDVN: SDN-Based Radio Resource Management of Heterogeneous V2X for Cooperative Perception," IEEE Access, vol. 11, pp. 76255-76268, 2023.
[8] G. Chai, W. Wu, Q. Yang, M. Qin, Y. Wu, and F. R. Yu, "Platoon Partition and Resource Allocation for Ultra-Reliable V2X Networks," IEEE Transactions on Vehicular Technology, vol. 73, no. 1, pp. 147-161, 2024.
[9] S. U. Nyati, S. R. Suralkar, and U. S. Bhadade, "V2X Spectrum Allocation for Emergency Communication Using Cognitive Radio Transmission," SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology, vol. 14, no. 01, pp. 79-85, 2022.
[10] D. Han and J. So, "Energy-Efficient Resource Allocation Based on Deep Q-Network in V2V Communications," Sensors, vol. 23, no. 3, article no. 1295, 2023.
[11] S. Feki, M. Hamdi, A. Belghith, F. Zarai, and A. D. Algarni, "Multi-Layer Neural Network Algorithm for Vehicle-to-Everything Communication in 5G Networks," International Journal of Communication Systems, vol. 36, no. 7, article no. e5260, 2023.
[12] R. H. Hwang, F. Marzuk, M. Sikora, P. Chołda, and Y. D. Lin, "Resource Management in LADNs Supporting 5G V2X Communications," IEEE Access, vol. 11, pp. 63958-63971, 2023.
[13] A. M. A. Ibrahim, Z. Chen, Y. Wang, H. A. Eljailany, and A. Ipaye, "Optimizing V2X Communication: Spectrum Resource Allocation and Power Control Strategies for Next-Generation Wireless Technologies," Applied Sciences, vol. 14, no. 2, article no. 531, 2024.
[14] P. Ma, P. Zhao, Z. Bai, X. Dong, X. Yang, and K. Kwak, "Coalitional Game Based Resource Allocation in D2D-Enabled V2V Communication," Journal of Systems Engineering and Electronics, vol. 34, no. 6, pp. 1508-1519, 2023.
[15] L. Shan, M. M. Wang, F. Zhang, S. Chen, and J. Zhang, "Resource Allocation for Cellular Device-to-Device-Aided Vehicle-to-Everything Networks with Partial Channel State Information," Transactions on Emerging Telecommunications Technologies, vol. 33, no. 7, article no. e4501, 2022.
[16] N. Zhao, J. Wang, B. Jin, R. Wang, M. Wu, Y. Liu, et al., "Multi-Agent Reinforcement Learning-Based Resource Management for V2X Communication," International Journal of Mobile Computing and Multimedia Communications, vol. 14, no. 1, pp. 1-17, 2023.
[17] Q. Liu, R. Luo, H. Liang, and Q. Liu, "Energy-Efficient Joint Computation Offloading and Resource Allocation Strategy for ISAC-Aided 6G V2X Networks," IEEE Transactions on Green Communications and Networking, vol. 7, no. 1, pp. 413-423, 2023.
[18] M. Christopoulou, S. Barmpounakis, H. Koumaras, and A. Kaloxylos, "Artificial Intelligence and Machine Learning as Key Enablers for V2X Communications: A Comprehensive Survey," Vehicular Communications, vol. 39, article no. 100569, 2023.
[19] M. Parvini, P. Schulz, and G. Fettweis, "Resource Allocation in V2X Networks: From Classical Optimization to Machine Learning-Based Solutions," IEEE Open Journal of the Communications Society, vol. 5, pp. 1958-1974, 2024.
[20] P. Hou, X. Jiang, Z. Lu, B. Li, and Z. Wang, "Joint Computation Offloading and Resource Allocation Based on Deep Reinforcement Learning in C-V2X Edge Computing," Applied Intelligence, vol. 53, no. 19, pp. 22446-22466, 2023.
[21] A. D. Mafuta, B. T. J. Maharaj, and A. S. Alfa, "Decentralized Resource Allocation-Based Multiagent Deep Learning in Vehicular Network," IEEE Systems Journal, vol. 17, no. 1, pp. 87-98, 2023.
[22] I. Lee and D. K. Kim, "Decentralized Multi-Agent DQN-Based Resource Allocation for Heterogeneous Traffic in V2X Communications," IEEE Access, vol. 12, pp. 3070-3084, 2024.
[23] S. M. Waqas, Y. Tang, F. Abbas, H. Chen, and M. Hussain, "A Novel Duplex Deep Reinforcement Learning Based RRM Framework for Next-Generation V2X Communication Networks," Expert Systems with Applications, vol. 233, article no. 121004, 2023.
[24] J. S. Alrubaye and B. S. Ghahfarokhi, "Geo-Based Resource Allocation for Joint Clustered V2I and V2V Communications in Cellular Networks," IEEE Access, vol. 11, pp. 82601-82612, 2023.
[25] N. Wan, Y. Luo, G. Zeng, and X. Zhou, "Minimization of VANET Execution Time Based on Joint Task Offloading and Resource Allocation," Peer-to-Peer Networking and Applications, vol. 16, no. 1, pp. 71-86, 2023.