An Adaptive Framework for Resource Allocation Management in 5G Vehicular Networks

Rajilal Manathala Vijayan[1], Fabrizio Granelli[2], A. Umamakeswari[3]*

[1] Rajilal Manathala Vijayan, SASTRA Deemed University, (India)
[2] Fabrizio Granelli, University of Trento, (Italy)
[3] A. Umamakeswari, SASTRA Deemed University, (India)

* Corresponding author: aumaug67@gmail.com

Received 14 August 2023 | Accepted 23 October 2024 | Early Access 11 April 2025

Abstract

Vehicle-to-everything (V2X) communication is crucial in vehicular networks, for enhancing traffic safety by ensuring dependable and low latency services. However, interference has a significant impact on V2X communication when channel states are changed in a high mobility environment. Integration of next generation cellular networks such as 5G in V2X communication can solve this issue. Also, successful resource allocation among users achieves a better interference control in high mobility scenarios. This work proposes a novel resource allocation strategy for 5G cellular V2X communication based on clustering technique and Deep Reinforcement Learning (DRL) with the aim of maximizing systems energy efficiency and MVNO's profit. DRL is used to distribute communication resources for the best interference control in high mobility scenarios. To reduce signalling overhead in DRL deployments, the proposed method adopted RRH grouping and vehicle clustering technique. The overall architecture is implemented in two phases. The first phase addresses the RRH grouping and vehicle clustering technique with the objective of maximising the energy efficiency of the system and the second phase addresses the technique of employing DRL in conjunction with bidding to optimise MVNO's profit. Second phase addresses the resource allocation which is implemented in two level stage. First level addresses the bidding of resources to BS using bidding and DRL techniques and the second level addresses the resource allocation to users using Dueling DQN technique. Through simulations, the proposed algorithm's performance is compared with the existing algorithms and the results depicts the improved performance of the proposed system.

Keywords

5G, Deep Learning, Energy Efficiency, Machine Learning, Resource Allocation.

DOI: 10.9781/ijimai.2025.04.002

I. Introduction

OVER the past decade, vehicular communications have gained popularity and utility due to their potential to drastically decrease traffic fatalities, facilitate safe movement, and enable intelligent vehicular systems [1]. Cellular network topologies have been studied and identified as a possible remedy to achieve this goal. Conventional cellular network systems such as 4G provide voice and data services through the deployment of a single Base Station (BS) in a single cell. Single BS - single cell networks are suited to low-traffic environments. For high-volume traffic environments, multiple BS - multiple cell networks are suited [2].

5G cellular network addresses the limitations of Single BS - single cell networks and proposes the multiple BS - multiple cell network architecture. Numerous topologies, such as heterogeneous networks (HetNet), Cloud Radio Access Networks (C-RAN), and Heterogeneous Cloud Radio Access Networks (H-CRAN), have been investigated to increase the capacity of 5G cellular networks to boast the multiple BS - multiple cell network architecture. To meet the continuously expanding data traffic requirements, HetNet uses the method of overlapping a macro cell and multiple small cells. Small cells have a lesser coverage area than macro cells and are placed closer to consumers in order to deliver a higher data rate. Small cells, on the other hand, need more system power at BSs than standard network topologies. In C-RAN, the reduction in system energy usage is achieved by the division of BS into Remote Radio Head (RRH) and Baseband Unit (BBU) [2]. Signal-processing BBUs are centralised in a BBU pool, whereas radio-frequency transceivers RRHs are spread across the cells.

In the context of vehicular networks, there is need of multiple BS - multiple cell architecture due to high data volume scenarios. In vehicular networks, RRHs are essentially small base stations that are

Please cite this article as:
R. M. Vijayan, F. Granelli, A. Umamakeswari. An Adaptive Framework for Resource Allocation Management in 5G Vehicular Networks, International Journal of Interactive Multimedia and Artificial Intelligence, (2025), http://dx.doi.org/10.9781/ijimai.2025.04.002

positioned at key points, such as roadside spots, traffic lights, and lamp posts. It facilitates the V2X communication, real-time data exchange and high-bandwidth scenarios. Therefore, RRHs are essential for improving vehicle performance on cellular networks (VNets). The following points explain why RRHs are significant [3]:

Increased Network Capacity and Coverage: In regions with a high vehicle density, such as highways or urban centres, traditional base stations may find it difficult to deliver a strong enough signal and adequate capacity. Because they are positioned closer to moving cars, RRHs can greatly enhance reception quality and signal coverage. This corresponds to enhanced user capacity and decreased signal dropouts.

Improved User Experience: Stronger and more dependable signals from RRHs benefit vehicle users in terms of faster data speeds, decreased latency and enhanced call quality.

Scalability and Flexibility: RRHs provide more network deployment options. They provide focused coverage in places where traditional base stations might not be practicable by being simply installed on lamp posts, traffic signals, or existing infrastructure. More RRHs may be easily deployed to scale network capacity efficiently as Vehicular Network applications and user demands increase, satisfying the changing needs of an increasing number of connected vehicles.

Encouraging V2X connectivity: Advanced driver assistance systems (ADAS) and driverless cars rely heavily on V2X connectivity. By offering a low-latency platform for real-time data transmission between cars and infrastructure, and by supporting technologies for increased road safety and traffic efficiency, RRHs can help ensure dependable V2X connection.

In cellular networks, optimising the probable benefits of RRH deployments require efficient resource allocation [4]. Increased energy efficiency, better user experience, and increased network capacity are all advantages associated with efficient RRH resource allocation. Vehicle users will experience additional benefits from resource allocation in terms of Quality of Service (QoS).

In this research, a novel resource allocation strategy for cellular V2X communication based on DRL and a clustering method is proposed. To reduce signalling overhead in DRL deployments which in turn maximises the system energy efficiency, the proposed method in the first step groups the nearby RRHs. Vehicles are grouped to gather recurring messages and seek resource allocation to save system energy usage. A perfect location strategy is learned by the DRL-based resource allocation to optimise the total amount of feasible data rates. The technique of employing DRL in conjunction with bidding is used to address the resource allocation problem which in turn optimises MVNO's profit. Recent studies have suggested novel solutions to these optimization challenges, including game theory techniques [5], linear programming approaches [6], etc. In recent years, a large number of techniques have been suggested in the literature for this resource allocation optimization problem. Most of the recent works heavily concentrated on using meta-heuristic algorithms to solve the resource allocation optimization task [7]. These algorithms have poor convergence rates and a tendency to get stuck in local minima instead of global minima leading to mode complexity, also the parameter settings used in simulation have a significant influence on the outcomes of experiments. The optimization problem in this work has a sizeable state space due to the large number of undisclosed parameters (channel status and customer information). The proposed DRL is a promising method to address this challenging control issue. Therefore, the proposed optimization problem is attempted to be solved in this work using DRL and bidding techniques. Actions having outrageous levels of similarity may be established in the environment to map DRL with states and rewards, which would then undergo training to create stronger tactics. A viable approach to address this challenging control challenge is the proposed DRL technique. As a result, our suggested algorithm's major purpose is to optimize system energy efficiency while maintaining reliability and maximising the profit of MVNO's.

The paper makes the following contributions:

- To improve the system's energy efficiency, RRH grouping and vehicle clustering are proposed.
- To maximise the profit of MVNOs, a novel resource allocation method is proposed using joint bidding and DRL techniques.

The rest of this paper is laid out as follows. The section II describes the recent works concerning resource allocation techniques and clustering techniques. The section III narrates the methodologies and problem formulation. The proposed novel method is elaborated in section IV and the discussion on the results in section V and conclusion of the work in the section VI.

II. Related Works

It is challenging to confirm that the many available algorithms that use standard optimization approaches meet the Quality of Service (QoS) requirements. Machine learning (ML) has shown promise in solving challenges requiring unpredictable network communications [8]. The features of ML help in finding optimal or near-optimal solutions in unpredictable and intrinsically varying vehicular scenarios.

A DeepCog based on 3D CNN was used in [9] for resource management in 5G mobile networks. The infrastructure of the 5G technology network is segmented. Every slice is intended to have its own set of resources allocated by the DeepCog. When tested in a real-world setting, DeepCog is determined to be efficient. Temporal CNN was proposed by authors in [10] for mmWave outdoor location in 5G mobile wireless networks. With an average inaccuracy of 1.78 meters for non-line-of-sight mmWave outdoor sites, the temporal CNN maintained a single actor, moderate bandwidth, and a sample of binary data.

Deep learning was introduced by [11] for the distribution of cooperative resources in 5G mobile wireless networks according to channel circumstances. Using channel data and resource allocation meant for optimization, the study proposed CNN. Especially in a dynamic channel environment, the generated CNN can help make the full-scale channel information optimal use instead of the traditional resource optimal use. It is discovered that the technique works well in reducing optimization complexity, speeding up computing, and delivering good results. A study [12] looked at how 5G wireless cloud network random access networks allocate resources for TV multimedia services. A deep learning paradigm for resource allocation is proposed in this article. The Neural network is connected to user bandwidth and power resource distribution. To reserve resources for Ultra-Reliable Low Latency Communication (URLLC) in a 5G network, authors [13] proposed RL in their work. It is discovered that in terms of resource usage and packet drop probability, the RL outperforms the baseline approach. In order to enhance QoS in a smart grid provided by a 5G network, authors [14] suggested applying reinforcement learning (RL) to the dynamic scheme of the network slice resources. The algorithm has the ability to quickly adjust the network's demand for processing resource allocation.

To allocate radio resources in 5G that satisfied service requirements regardless of the number of slices, the work [15] suggested DRL. In order to optimize resource allocation, computation offload, and caching placement, the authors in their work [16] devised a DRL time scale that consists of rapid and slow timelines for learning processes. Federated learning is used to train the DRL in a distributed manner while maintaining the privacy of edge device data. [17] developed an RL method to meet the varying needs of many entities with severe

TABLE I. Existing Resource Allocation Techniques

Authors	Contributions	Advantages	Disadvantages
Bega et.al.[9]	Ensuring resources for each slice using CNN technique	Each slice has its own resources	Signalling overhead is not addressed
Gante et.al. [10]	Channel information is used for resource allocation using CNN technique	mmWave outdoor sites are considered	Signalling overhead is not addressed
Zhao et.al. [14]	RL technique	Dynamic resource allocation using RL technique	Signalling overhead and maximum utilisation of resources are not addressed
Li et.al. [18]	RL technique	Two-tier routing mechanism to address the dynamic resource allocation	Signalling overhead and maximum utilisation of resources are not addressed
Zhang et.al. [20]	DRL technique	Decentralised method of resource allocation using DRL technique	Signalling overhead and maximum utilisation of resources are not addressed
Sanguanpuak et.al. [23]	Used Q-learning	Maximize power resource allocation	Signalling overhead and maximum utilisation of resources are not addressed
Liu et.al. [29]	Resource allocation based on active users using DQN technique	Addressed the video, VoLTE, and uRLLC slice performance requirements in 5G-RAN	Signalling overhead and maximum utilisation of resources are not addressed

QoS criteria. To accommodate the challenging scenarios of dynamic changes in the network, [18] proposed a user association technique implemented using online RL to balance the network. A two-tier routing mechanism, based on two-tier cluster heads, for efficient data distribution, was established in [19]. The level-1 cluster heads were chosen using fuzzy logic and criteria such as relative velocity, k-connectivity, and connection dependability. Q-learning chose the level-2 cluster heads to reduce communication overhead. The optimization of sum capacity while keeping in mind the latency and reliability limits, a decentralised mode selection and resource allocation method based on DRL was developed in [20]. The Markov decision process was used in selecting mode and issues related to resource allocation, in order to determine the best transmission mode, sub-band, and transmission power level. Graph theory is used to overcome the constraints of local DRL models. Refs. [21] and [22] proposed a resource allocation strategy using Q-learning for eMBB and uRLLC services on 5G-RAN, intending to maximize overall resource usage while considering the need to enhance the performance of the system and the dynamic nature of traffic. The system looked at a space of states in the uplinks and downlinks that was modelled as the number of resource blocks of cell bandwidth. The actions were represented in terms of eMBB and V2X slice allocation ratios. In an edge-computing and multi-tenant setting, Ref. [23] developed a solution using Monte Carlo [24] and Q-learning to maximize power resource allocation in each network slice by ensuring guaranteed QoS. The method used a binary space of states, state one indicating interference in the resource block allotted to a tenant's small cell BS and state zero indicating no interference. The power level to allocate was defined as the action space. Ref. [25] described a Q-learning-based, dynamic, and autonomous compute resource allocation strategy for 5G Fog-RAN that aimed to reduce latency, energy consumption, and cost. The space of states was described as a vector that included allocated computing resources, CPU reservation and average CPU utilization. At the Fog-RAN node level, the action space was specified in terms of resource allocation. To enhance QoS satisfaction and resource usage, references [26] and [27] presented a (Deep Q network) DQN-based and dynamic architecture that reserves and allocates unused bandwidth resources to virtualized RAN. A FNN with four neurons in the input layer, two hidden layers, and twenty neurons in the output layer was utilized in the framework. The proportion of allotted virtual resources and the average resource use of each slice were used to determine the space of states. The proportion for lowering or increasing given resources was used to describe the space of actions. Ref. [28] described a DQN-based resource management solution for MVNO's to reserve and assign cache resources at the edge network to maximize QoS satisfaction and network usage. A FNN with 4 and 11 neurons in the input and output layers, as well as two hidden layers, was employed by the DQN agent. The method envisioned a set of states for resource consumption, QoS fulfilment, reserved resources per slice, and cache resource allocation. The action space was developed to raise or reduce the amount of resources allocated to cache slices. Ref. [29] proposed a resource allocation architecture based on a limited DQN built by a DNN (made of two fully connected layers with 64 and 32 nodes) trained with various RL algorithms and designed to fulfil video, VoLTE, and uRLLC slice performance requirements in 5G-RAN. The number of active users per service was used to represent the space of states in the framework. The bandwidth to assign for each service was determined by the space of actions. According to the dynamics of service demands on 5G-RAN, Ref. [30] developed a resource allocation strategy based on powered DDQN to achieve SLAs as well as maximize resource usage and provider income. Table I summarises the existing techniques.

The following works did the study based on the quality of QoS to the users in terms of security, and proper data migration. In the work [31], the authors developed a novel technique for anticipating server downtime and other delay factors such as delay in processing the user's request for the service etc., in the data communication scenario between the service provider and users. Researchers in their work [32], [33] proposed novel techniques for countering the attacks on the QoS to the users.

III. Problem Formulation

This work describes a resource allocation technique based on DRL for cellular V2X communication with RRH grouping and vehicle clustering while ensuring reliability criteria. By an effective resource allocation technique, the work tries to increase the overall MVNO profits and to maximise the system's energy efficiency. The proposed method is divided into two sections: the first section deals with maximizing the system's energy efficiency and the second one addresses the aim of increasing the profit of MVNOs. Clustering the vehicles and RRH grouping is used to achieve the first aim and a novel DRL-based resource allocation approach is used to achieve the second aim. In vehicular networks, the resource allocation problem is intractable by conventional optimization algorithms because they lack knowledge of channel gain, interference, and other past channel states. Fast determination of the best BS transmission energy allotment method to increase the throughput of the system and to reduce the energy usage is challenging for conventional resource allocation techniques, due to a huge number of the customer's discrete nature of the information in the V2X system. The problem of data

explosion issue brought on by an excessively large action-state space and optimal energy efficiency challenge in vehicular networks, this is addressed using the DRL method.

A. Maximizing System Energy Efficiency

The proposed work considers the V2I and V2V depend on Mode3 of the 3GPP cellular V2X architecture, which encloses multiple RRH and multiple Vehicle Equipment (VEs). The VEs are mobile equipment placed on the roadways, while the RRHs are placed in the middle of junctions and communicated to the BBU pool through the front-haul link. The resources are distributed to VEs in the V2I mode by the RRH and this will be used by VEs in V2V mode. Only internal communication happens between the VEs in the V2V mode. However, the VEs in the V2I mode can communicate with VEs in the V2V mode. $\mathbb{S}$ represents the total group of VEs which comprises the V2I VEs $\mathbb{X}$ and V2V VEs $\mathbb{Y}$. The resource for allocation to different VEs is the bandwidth and it is represented as BW^{total}. The number of RRHs is represented as R. Each V2I VE is given one of the K Resource Blocks (RBs) from the overall bandwidth BW^{total}.

When transmitting safety-critical messages through V2V, it is crucial to meet strict latency and reliability criteria. The latency and reliability criteria must be formulated as constraints of the optimization problem; however, this is challenging. The delay requirement is therefore turned into data queue length by Little's law, while the reliability criteria can be formulated as outage probability [34]. The outage probability defines the probability that the SINR of VEs is less than the established SINR benchmark level [35]. Reduced dependability is caused by an increase in outage likelihood, which causes packet loss and reveals retransmission occurrence. Communication between V2I and V2V suffers from mutual interference when cellular resources are reused. The significance of SINR generation is to address the interference-related issues in the data transmission between V2I and V2V. Equation (1) represents the x^{th} V2I VE's SINR at the k^{th} RB.

$$\gamma_k^x = \frac{p_{k\cdot}^s \cdot g_k^{rx}}{N_p + \sum_{v\in\mathbb{V}} v_k^y \cdot g_k^{xy} + \sum_{r'\in\mathbb{R},\Omega'\neq s} p_k^s \cdot g_k^{r'x_x}} \tag{1}$$

where g_k^{rx} represents the channel gain between x^{th} V2I VE and k^{th} RB. p_k^s and p_k^y represents the transmission power of RRH and y at the k^{th} RB respectively. N_p represents noise power spectral density. Equation (2) represents the y^{th} V2V VE's SINR at the k^{th} RB.

$$\gamma_k^y = \frac{p_k^y . g_k^{yy'}}{N_p + p_k^x . g_k^{xy} + \sum_{j'\in\mathbb{Y}, j'\neq y} p_k^j . g_k^{jy}} \tag{2}$$

where p_k^y denotes the V2V VE transmission power at the k^{th} RB. Equation (3) represents the total system capacity calculation.

$$S_c = BW^{\text{total}} . \sum_{x\in\mathbb{X}} \sum_{k\in\mathbb{K}} \{\log_2(1+\gamma_k^x) + \sum_{y\in\mathbb{Y}} \log_2(1+\gamma_k^y)\} \tag{3}$$

where the system's overall bandwidth is BWtotal. Energy usage of the RRHs and related front-hauls are considered in the calculation. Since the macro-BS only provides voice services, its energy usage is ignored. Equation (4) depicts the RRH energy usage calculation.

$$P_{\mathbb{R}} = \sum_{r=1}^{\mathbb{R}} (\pi_{\mathbb{R}} + \Delta\text{slope} \sum_{x=1}^{\mathbb{X}} a_x^r p_k^s) \tag{4}$$

where $\pi_{\mathbb{R}}$ is the RRH's circuit power, Δslope is the slope of the RRH's load-dependent power consumption, as stated in [36], and a_x^r is the V2I VE x's association indicator, with values of 1 indicating association and 0 indicating non-association. Equation (5) represents the Front-haul links power consumption model.

$$P_{\text{front-link}} = \sum_{r=1}^{\mathbb{R}} (\pi_{\text{front-link}} + \psi . t_r) \tag{5}$$

where $\pi_{\text{front-link}}$ is the combined circuit power transmitted by the front-haul transceiver and switch, ψ is the power usage per bit/s and t_r represents traffic linked with RRH r. Equation (6) represents the entire system's energy usage model formulation.

$$S_e = P_{\mathbb{R}} + P_{\text{front-link}} \tag{6}$$

Equation (7) depicts the energy efficiency of the system.

$$S_{ee} = \frac{S_c}{S_e} \tag{7}$$

The key objective is to increase the system's energy efficiency. Equation (8) and (9) defines the objective.

$$\max_{\{\mathbb{X}\mathbb{Y}\mathbb{K}\}} S_{ee} \tag{8}$$

$$\begin{aligned} &\text{s.t. X1: } \gamma^s \geq \gamma^o, \forall x, y, k \\ &\text{X2: } \theta \leq \theta_{\max} \\ &\text{X3: } 0 \leq p_k^y \leq TP_s, \forall y, k \\ &\text{X4: } \sum_{k\in\mathbb{K}} f_{x,k} \leq 1, f_{x,k} \in \{0,1\} \end{aligned} \tag{9}$$

$f_{x,k}$ represents allocation criteria, with values of 1 and 0 representing allocation and non-allocation to V2I VE x at k^{th} RB respectively. γ^s is the SINR of VE, γ^o is the SINR threshold, TP_s is the maximum transmission power of VE, θ is the system outage probability, and $\theta_{\max}$ is the maximum outage probability limit.

The constraint X1 makes sure that VEs meet the SINR constraint, X2 makes sure that the system's outage probability is within the threshold limit. Due to the serious interference issue, constraint X3 makes sure that V2V transmitters do not surpass the peak threshold level of the transmission power, and X4 represents that each V2I VE can be assigned to an RB. This ensures the exclusive allocation of only one RB to each of the V2I VE.

B. Maximising the Profit of MVNOs

This section discusses the framework concerning to BS and the multiple users. The framework considers a single BS and a group of MVNOs, which is represented as M = {M1, M2, ..., Mm}. There are k connected subscribers, sm= {sk1, sk2, …, skm} to each MVNO and offers unique mobile services to those subscribers. The resources Shared aggregated bandwidth is the resource for this BS. According to the needs of the connected users, each MVNO must bid to the BS and then distribute the resources it receives from the BS to its connected customers. The QoE of the consumers is represented in this study by the SLA satisfaction rate (SSR). Here the challenge is how to coordinate schedules amongst MVNOs, meet connected users' needs, and increase MVNO profits overall. Additionally, this BS's resources are virtualized and cut into several sizes to accommodate user needs. To maximize the profit, the resource allocation is considered in two levels.

1. Higher Level Model

The MVNOs need to determine the sufficient bandwidth and structure the bid for submission to InP according to the number of connected users as well as the guaranteed QoS level. InP allocates the resources based on the algorithm which decides the bidding levels [37]. The parameters r_m and r_k^m represents the resource allocation to m^{th} MVNO by InP and MVNO to users respectively. Equation (11) depicts the formulation of r_m. The MVNOs determine the sufficient bandwidth based on the rate demands from its associated users. The rate demands are represented as $v_{k,0}^m$ and $v_{k,1}^m$. $v_{k,0}^m$ represents the lowest rate of demand and $v_{k,1}^m$ represents the highest rate of demand. Equation (10) depicts the range of b_m in terms of $v_{k,0}^m$ and $v_{k,1}^m$.

$$\sum_{k=0}^{K} v_{k,0}^m < b_m < \sum_{k=0}^{K} v_{k,1}^m \tag{10}$$

$$r_m = \frac{b_m}{\sum_{m=1}^{M} b_m} R, \forall m \in M \tag{11}$$

$$y_m(r_m(b)) = v_m(r_m(b)) - p_m b_m \tag{12}$$

Resources $r_m(b)$ are distributed by the BS to MVNOs in accordance with their bids. An evaluation function p_m is established as a punishment function that will lower MVNOs' profits if they overly increase their bids. The prevention of overpricing the bid value is done using the function $y_m(r_m(b))$.. Equation (12) represents this. Equations (13) and (14) represents the formulation of p_m and β respectively. v'_m represents the bid of the MVNO.

$$p_m = \frac{1}{\beta} v'_m \left(1 - \frac{r_m(b)}{R}\right) \tag{13}$$

$$\beta = \frac{\sum_{m-1}^{M} b_m}{R} \tag{14}$$

The aim of the higher-level model is the maximization of a weighted sum of the utility and benefits of all MVNOs. Equation (15) and (16) represents the aim of the higher model.

$$optF = \sum_{m \in M} g_m + \omega * \sum_{m \in M} y_m(r_m(b)) \tag{15}$$

$$S.t.\, r_m(b) \cap r_n(b) = 0$$
$$\sum_{m \in M} r_m \leq R$$
$$\sum_{k=1}^{K} r_k^m \leq r_m \tag{16}$$

The maximisation function optFmakes sure that the resources allotted to various MVNOs are separated. Given that the BS has a limited amount of bandwidth, constraint $S.t.r_m(b) \cap r_n(b) = 0$ ensures the allocation should be within the BS's available limit and the constraint $\sum_{m \in M} r_m \leq R$ states that the bandwidth allocation to the users by the MVNOs should be within the allocated limit from the BS. DQN can also address the issue of each MVNO competing for resources.

2. Lower-Level Model

The second level aims to develop a resource allocation algorithm to maximize the utility function, g_m, of each MVNO based on the weighted sum of s_m and $SSR_{u_k^m}$. This section explains how g_m is computed.

Equation (17) and (18) depicts the calculation of the downlink rate between the sender BS and receiver, k^{th} user u_k^m, which is connected to the mth MVNO. The calculations are based on Shannon's equation.

$$v_{u_k^m} = r_k^m \log(1 + SNR_{u_k^m}) \tag{17}$$

$$v_m = \sum_{k=1}^{K} v_k^m \tag{18}$$

signal-to-noise ratio associated with BS is represented as $SNR_{u_k^m}$. Equation (19) depicts this.

$$SNR_{u_k^m} = \frac{g_{u_k^m} P_t}{N_{sd} r_k^m} \tag{19}$$

$g_{u_k^m}$ denotes the channel's fading gain between sender and receiver, P_t stand for the disseminated power, and N_{sd} represents the noise spectral density. Equation (20) depicts the weighted sum of S_m.

$$S_m = \frac{\sum_{u_k^m \in U_m} \sum_{k \in K} v_{u_k^m}}{r_m} \tag{20}$$

$SSR_{u_k^m}$ represents the k^{th} user SSR, linked to m^{th} MVNO. Equation (21) depicts this.

$$SSR_{u_k^m} = \frac{\sum_{p_k^m \in P_k^m} \beta_{p_k^m}}{\sum P_k^m} \tag{21}$$

In the above equation numerator represents the value of valid packets that were successfully received by the user and the denominator represents the total number of packets provided by the MVNO. Binary $\beta_{p_k^m}$ indicates the genuineness of the successfully received p_k^m packet, and $\beta_{p_k^m} = 1$ when $v_{u_k^m} > \overline{v_{u_k^m}}$, otherwise $\beta_{p_k^m} = 0$. $\overline{v_{u_k^m}}$ represents the priorly set downlink transmission rate for the user u_k^m depending on the SLA.

$$maxg_m = max(\delta S_m + \sum_{k \in K} \psi_k SSR_{u_k^m}) \tag{22}$$

Maximizing the utility function g_m of each MVNO is the aim of the second-level model and g_m is represented as the weighted sum of $S_m$and $SSR_{u_k^m}$. Equation (22) represents this. The significant weights of S and SSR are shown by the symbols δ and $\psi = \{\psi_1, \psi_2, ..., \psi_s\}$, respectively. The maximization task can be achieved using the DRL technique. The overall architecture is depicted in Fig. 1.

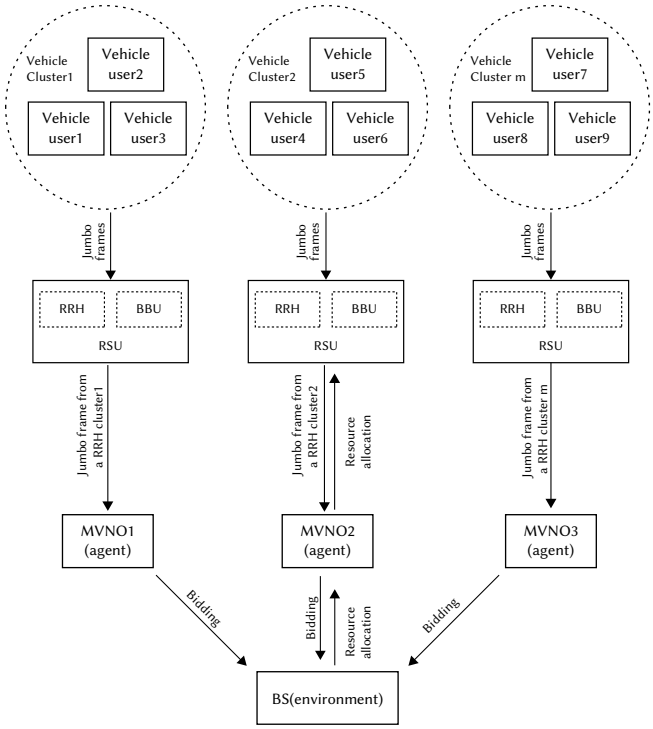


Fig. 1. Resource allocation architectural framework.

IV. Proposed Methodologies

This section describes the proposed novel algorithm for maximising the system's energy efficiency and increasing MVNO's profit. The proposed algorithm is split into two parts, the first part deals with the aim of maximising system energy efficiency and the second part aims at increasing MVNOs' profits. For the first goal, the clustering of the VEs and RRH grouping is adopted, and for the second goal, a novel DRL-based resource allocation approach is used.

A. Maximising System Energy Efficiency

Maximising the system's energy efficiency can be achieved by minimising the signalling overhead and communication complexity and also by maximising system capacity. To minimise the signalling overhead and communication complexity, the RRH grouping and vehicle clustering method is adopted and for maximising system capacity, a novel resource allocation method using DRL is adopted.

1. RRH Grouping

Frequent state updates occur when the RRH perform learning as deep Q-learning agents, leading to significant communication overhead and maintenance costs. Similar RRHs are clustered together

to decrease the overhead. The term related RRHs indicates RRHs with similar volumes of servicing data and automobiles, as well as interference situations. Additionally, because the data flow of vehicle networks exhibits spatial regularities, the position of RRHs is taken into account while evaluating the similarity of RRHs. Each RRH group is considered as an agent in the DRL approach to maximise system energy efficiency. The proposed method considered four parameters when grouping similar RRHs: volume of the traffic, position of the RRH, VEs interference and range of the data services. The number of VEs passing through each RRH in a given time unit is referred to as traffic volume. The resources that the RRH serves VEs in every hour are referred to as the range of data services. VE interference refers to the average SINR values of VEs that are served by the RRH. For every RRH, the likeness with nearby RRHs is calculated and saved in an array by taking into account the above-mentioned variables. Using the minimax technique the cluster centre is formed and certain RRHs are placed in the centre.

2. Vehicle Cluster Formation

The SidelikeUEInformation message in V2X communication, which requests resource allocation, and the MeasurementReport message, which reports location information, share the same communication channel [38]. Because of the frequent notifications, the RRHs are under a heavy traffic strain in situations with many VEs. To address this issue, several VEs join together to form a cluster.

The VEs send out frames to the Cluster Head (CH) which combines the frames from each member and forms a main frame. The inter-communication between each member is through a single frame which consists of fields such as header and data. These frames are used to get information such as location, time, velocity, ID etc. Multiple single frames are combined to form a main frame with a single header and a single substantial payload by CH. The advantage of constructing a single main frame is to reduce the communication overhead caused by headers in multiple frames. Equation (23) represents the traffic calculation associated with a single frame at time t.

$$\sum_{t\in\mathbb{T}} (\mathrm{len}_{\mathrm{header}} + \mathrm{len}^{t}_{\mathrm{payload}}).\, \mathrm{N}^{t}_{v} \tag{23}$$

where $\mathrm{len}_{\mathrm{header}}$ is the header length; $\mathrm{len}^{t}_{\mathrm{payload}}$ is the payload length and N^{t}_{v} represents the total number of users at time t.

Equation (24) depicts the calculation of the traffic associated with a mainframe constructed by CH.

$$\sum_{t\in\mathbb{T}} (\mathrm{len}_{\mathrm{header}} + \mathrm{len}^{t}_{\mathrm{payload}}.\, \mathrm{N}^{t}_{\mathrm{CM}}) \tag{24}$$

where $\mathrm{N}^{t}_{\mathrm{CM}}$ is the number of Cluster Members (CM), $\mathrm{len}_{\mathrm{header}}$ is the header length; $\mathrm{len}^{t}_{\mathrm{payload}}$ is the payload length.

Following are the steps in the cluster creation process.

The very first step is the initialisation step and this is to be done when none of the clusters are configured. In this step, all the VEs are associated with the RRH which provides the highest SINR and is assigned the V2I mode. The CH formation is based on the Prose-Per-Packet-Priority (PPPP^{v}) value. The V2I mode VE with the highest value of $\mathrm{F}^{v}_{\mathrm{CH}}$ is chosen as the first CH in the first phase. Equation (25) depicts the determination of $\mathrm{F}^{v}_{\mathrm{CH}}$.

$$\mathrm{F}^{v}_{\mathrm{CH}} = \gamma^{x}_{k}.\, \mathrm{PPPP}^{v} \tag{25}$$

where γ^{x}_{k} is the SINR of the ith V2I VE at the kth RB and PPPP^{v} is the Prose-Per-Packet-Priority value.

PPPP^{v} is important during data transmission in a V2X environment [39]. The main frame is filled with the payload of each CH and is said to wait until the transmission from each VE. The frames are arranged with the PPPP^{v} priority value to ensure that the CH with the highest $\mathrm{F}^{v}_{\mathrm{CH}}$ is placed as the CH in the first phase. Then the CM with the next PPPP^{v} priority value placed as succeeding CH.

The next step is cluster formation and deformation, which takes place in the following circumstances:(i) when the VE switches to a different cell; (ii) when the jumbo frame's payload exceeds the maximum permissible payload; (iii) when the permissible maximum limit of communication distance in the V2V scenario is crossed by the CH and the next CH. The nearby vehicles compare the receiving SINR between the RRH in action and nearby CH to determine whether or not to associate with the cluster during the cluster formation and deformation step. During this procedure, the broadcasted frame from each user, which includes location, time, velocity and ID at each instance, is used to compare the SINR of the serving RRH with those of the nearby CHs for each vehicle.

The last stage of the step is reached when VEs don't find a suitable cluster or when the SINR value of nearby CHs is lower than RRH. The VEs that are not a part of the cluster are then switched to V2I mode.

The same vehicle will, however, carry a larger load if it repeatedly assumes the CH position, such as while building jumbo frames. In other words, it would be expensive to keep the same car in CH status; as a result, it would need to be turned over to different vehicles periodically. The CH identifies the VE as the next CH in the jumbo frame, after the current CH. According to Fig. 2, the main frame is arranged in chronological sequence of the CM frames that were received. Therefore, CH does not experience an increase in load. A jumbo frame, however, has a less-than-ideal relationship with the cluster's ideal CH.

The load on RRH and the size of the cluster is indirectly proportional to each other. A Larger cluster area defuses the VEs in V2I mode and thus reduces the interference and also will reduce the load on RRH. Whereas in the small area cluster, the number of VEs in V2I mode will increase and thus also increase the load on RRH. Consequently, a dynamic cluster size is necessary. The formulated main frame from each cluster will compete to get resources from the respective MVNOs. The framework for resource allocation by MVNOs to the competing mainframes is described below.

B. Increasing MVNOs Profit

1. Two-level Resource Allocation Method by Integrating DQN and Joint Bidding Techniques

The resource allocation problem is considered in two levels which are termed as higher and lower levels. The jumbo frames submit resource allocation requests to MVNO. MVNO places the bidding on the BS. The BS allocates the resources to MVNO. MVNO allocates the resources to each vehicle cluster. On the higher level, each MVNO is considered as an agent and is given access to get resources from the BS by using a combination of bidding and the DQN method. On the lower level, MVNO distributes the resources it receives to the users connected to it using the Dueling DQN technique.

2. Deep Reinforcement Learning

We outline the core ideas of DRL and Q-Learning in this section. RL is a crucial machine learning technique that combines agents, rewards, and environment [40]. The agents respond differently to the environment and create new states. The agent applies the action based on the new state and gets a reward (positive or negative feedback) [41],[42]. The agent constructs decisions in accordance with the reward signal and learns through trial and error the best state-action pair, or strategy, to maximize the accumulated reward.

Some of the different types of RL that can be used to solve various problems [43] are, Model-based RL (environment modelling, where understanding of the environment is done by an actor), model-free

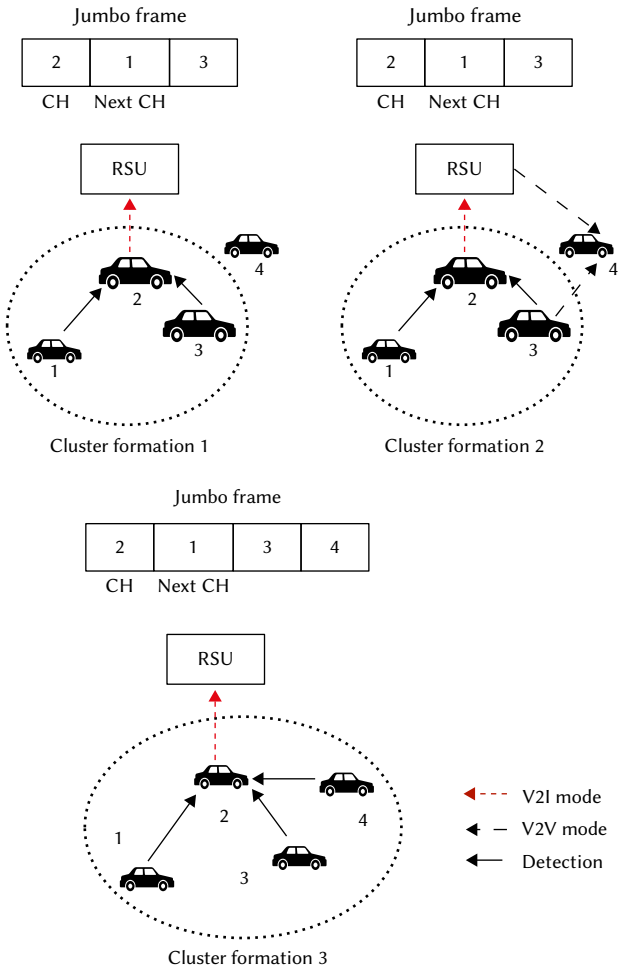


Fig. 2. Cluster formation method.

RL, on-policy RL (same learning and interacting actors), off-policy RL (different learning and interacting actors), Monte-Carlo RL (update once the completion of the learning process is done), temporal difference RL.

Model-free RL algorithms (also known as on-policy algorithms) train an optimal (stochastic or deterministic) policy or optimal Q-value function [39],[44]. On-policy RL algorithms include actor-critic [45], state–action–reward–state–action (SARSA) [46], and proximal policy optimization (PPO) [47]. The most prevalent off-policy RL algorithm is Q-learning [48]. Many interactions are realised by RL algorithms to achieve an optimal policy that meets design constraints. Because of the huge amount of data and the high computational cost, a high number of repetitions results in an expensive procedure in RL algorithms. To address this issue, ML suggests DRL, which combines RL and DL to solve problems with large dimensions and infinite states [49]. DRL use RL to train DNNs (e.g., FNNs and RNNs) to learn optimum policies on time [41],[50].

A common DRL method, DQN is useful for handling complex computation-intensive problems as well as decision-making issues. The learning process will be managed by an agent in DRL. The intelligent agent continuously interacts with the environment to generate a large amount of fresh data. It then applies a group of strategies centred on this information which enables DRL in maximising the accumulated reward when figuring out the apt course of action in each specific circumstance. The interaction of agent with its surroundings can be modelled using Markov decision process (S, A, R, P, γ), where the state space, S, contains the present state s and the next state s′; the action space, A, which contains present action a, and the next action a′; policy π(.|s) that calculates the mapping of s to a; reward function R, gained by carrying out the action a while in the state s in accordance with the policy π(.|s); probability of the transfer, P(.|s,a); γ denotes the discount factor.

Equation (26) represents the derivation of the state value function $U^{\pi}(S)$ according to the policy π(.|s) under the state s.

$$U^{\pi}(s) = F_{\pi,p}[\sum_{t=0}^{\infty}\gamma^{t}R_{t}|S_{o} = s]] \tag{26}$$

where $F_{\pi,p}$ is the expectation under policy π, R_t is the reward received in state s, γ denotes the discount factor.

Similar to this, the action value function $Q^{\pi}(s,a)$ can be formulated by carrying out the action *a* under the state *s* in accordance with the policy π(.|s). Equation (27) represents this.

$$Q^{\pi}(s,a) = F_{\pi,p}[\sum_{t=0}^{\infty}\gamma^{t}R_{t}|S_{o} = s], A_{o} = a] \tag{27}$$

where $F_{\pi,p}$ is the expectation under policy π, R_t is the reward received in state s, γ denotes the discount factor and A denotes the action space.

The agent feeds state *s* to the neural network to obtain all $Q^{\pi}(s,a)$. The state represents the observation from the environment. For the action selection and decision-making from $Q^{\pi}(s,a)$, agent uses the ϵ-greedy strategy. The agent is rewarded by the environment based on the action. Finally, $Q^{\pi}(s,a)$ is used to update the agent in accordance with the reward provided by the environment. Equation (28) represents this.

$$Q^{*}(s,a) = Q(s,a) + \beta(R + \gamma max_{a'}Q(s',a') - Q(s,a)) \tag{28}$$

where β represents the exclusive parameter of the Q network.

DQN is the combination of DL and neural networks with θ as the parameter for selecting the action and updating the parameter. The updation of the target Q-value function network and the Q-value function is done after a specific number of iterations and in real-time respectively. The value function is denoted as $Q\pi(s,a;\theta)$. The minimisation of the squared TD error using the equation (28), leads to calculating the optimum parameter θ, resulting in $Q\pi(s,a;\theta) = Q^{*}(s,a)$. Equation (29) represents the squared TD error.

$$\lambda^{2} = [r + \gamma max_{a'\in A}Q(s',a;\theta) - Q_{\theta}(s,a;\theta)]^{2} \tag{29}$$

Equation (30) represents the target Q-value calculation.

$$TargetvalQ = r + \gamma max'_{a}Q(s',a;\theta) \tag{30}$$

Equation (31) represents the loss function specified in DQN.

$$Loss(\theta) = F[TargetvalQ - Q(s,a;\theta)]^{2} \tag{31}$$

where *F* represents the Mean Square Error.

Dueling DQN guarantees the DQN's network structure, and also splits the Q value for using the in-state value function and advantage function. Equation (32) represents this.

$$Q^{\pi}_{DQNDuelling}(s,a) = U^{\pi}(s) + A^{\pi}(s,a) \tag{32}$$

$A^{\pi}(s,a)$ parameter depends on the action *a*, while parameter $U^{\pi}(s,a)$ is not connected with *a* and returns only one status value. Equation (33) depict the more specifically formulation of $Q^{\pi}_{DQNDuelling}(s,a)$.

$$Q_{Duelling}(s,a;\theta,\beta,\delta) = U^{\pi}(s;\theta,\beta) + A^{\pi}(s,a;\theta,\delta) \tag{33}$$

The parameters β and δ are solely used by two function networks whereas the parameter θ is commonly used between the functions. The dominating function is typically centralised to enhance the recognition of the two functions. Equation (34) represent this.

$$Q_{Duelling}(s,a;\theta,\beta,\delta) = U^{\pi}(s;\theta,\beta) + A^{\pi}(s,a;\theta,\delta) - \frac{1}{A}\sum_{a'\in A}A(s,a',\theta,\delta) \tag{34}$$

This formulation has the advantage of ensuring the stable relative ranking of dominant functions of each action in a given state, removing excess degrees of freedom, reducing the range of Q value, and improving algorithm stability. Dueling DQN simplifies training and accelerates convergence by breaking down the Q value into the value function $U^{\pi}(s,a)$ and the advantage function $A^{\pi}(s,a)$. This is in contrast to the conventional DQN network structure. This benefit intensifies further as the number of actions rises. It is simpler to train the state value function because it only depends on the state and is not influenced by behaviour. In the same state, different behaviours can share the same value $U^{\pi}(s,a)$.

Duelling DQN's benefits are as follows [51]:

(1) Without altering the basic RL algorithm, it can generalise the learning process to all possible environmental actions.

(2) It can swiftly determine the optimal course of action since it does not need to know the effects of every action on every state because it can learn the most critical state for the agent.

(3) From the standpoint of network training, less data is needed, which results in a more straightforward and user-friendly network training process.

(4) It is simpler to maintain the order between actions when the state and advantage functions are trained independently. The value function can be broken down into individual result parts, each of which has practical importance and the combination of results is determined uniquely, which improves the precision and robustness of network learning.

Hence, Dueling DQN, an enhanced algorithm for RL, exhibits superior performance and efficiency, making it suitable for tackling resource allocation problems.

3. Two Level Resource Allocation Method Using Dueling DQN and Bidding

The motivation for adding MVNO between the users and BS guarantees the privacy of user requirements and efficient utilization of physical resources available at the BS for users. Each MVNO gathers the user requirements and the status of the channel in jumbo frames to do a successful bidding to avail the resources available at the BS. Upon winning the bid, MVNO allocates the received resources to the linked users. The above-mentioned scenarios can be realized using the Markov Decision Process, using the two-level resource allocation method as described in the problem formulation. The proposed method is depicted in the Fig. 3.

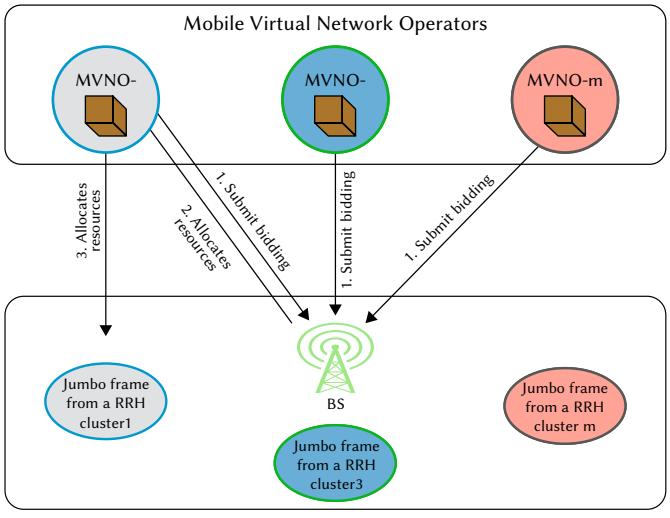


Fig. 3. Resource allocation using Dueling DQN and Bidding techniques.

The optimization task in the first level can be realised using the joint bidding and DQN algorithm which is described as Algorithm 1. Initially the bidding pool B_p and the DQN parameters are initialized. Each MVNO gathers the lower and upper limits of B_p . After that, the lower and higher cap of the user requirements from the jumbo frame is estimated based on the connected user's minimum and maximum total required rate. This is necessary to be done before creating B_p . The B_p is created after converting the needed rate to the lower and upper value of the bid in a particular ratio. The higher-level algorithm uses the B_p as action space and fills Table T with the lower-level resource allocation action in accordance with the higher-level action.

Algorithm 1: Joint Bidding and DQN Algorithm for Allocation of bandwidth in the first Level

1. **Initialize** the MVNO's low-level action selection table T and bidding pool $\boldsymbol{B_p}$;
2. **initilalize** the parameters such as action value function Q, target action value function Q, memory L, capacity N;
3. Each MVNO $\mathbf{m} \in \mathbf{M}$ calculates the entire upper and lower required rates of connected customers before establishing the Bidding pool $\boldsymbol{B_p}$;
4. **While** $\boldsymbol{b_m}$ in $\boldsymbol{B_p}$ **do**
   Find the appropriate allocation strategy for the lower level and record it in Table T;
5. end **while**
6. An action $\boldsymbol{a_t}$ is randomly selected, based on the bidding value $\boldsymbol{b_m} \in \boldsymbol{B_p}$ and BS issues $\boldsymbol{r_m}$ to each MVNO in accordance with (11);
7. Repeat
   While n=1 to N, **do**
   finding the proportionality between the allotted bandwidth and the needed lower rate and assigning this as the previous iteration's state i.e., C = c ;
   While m = 1 to M, **do**
   According to Table T, each MVNO m distributes to its consumers the optimal bandwidth $\boldsymbol{r_k^m}$;
   Calculation of $\boldsymbol{v_m}$ by each MVNO m using (17);
   Calculation of the penalty $\boldsymbol{p_m}$ by each MVNO m using (13);
   Each MVNO m, formulates its profit $\boldsymbol{y_m}$ by (12) to determine its reward $\boldsymbol{r_m}$;
   End **While**
   Determine the overall system utility $\boldsymbol{F_u}$ in accordance with (14);
   Determine the overall reward r;
   Select a course of action $\boldsymbol{a_m}$, say, bidding value $\boldsymbol{b_m} \in \boldsymbol{b}$ in accordance with DQN's policy;
   Each MVNO receives $\boldsymbol{r_m}$ from BS in accordance with (11);
   Obtain the state $C = c'$ following this iteration's selection operation;
   Each MVNO acts as the agent by providing (s, a, s′, and r) to DQN;
   Transition (s, a, s′, and r) are stored by agent in L;
   The agent chooses a subset of transitions (s, a, s′, and r) from Lin random;
   If iteration stops at *step_+1*
   Set the value for $y\_$ as $r\_$;
   Else
   Set the value for $y\_$ as $\boldsymbol{r\_ + \gamma max_{a^*}\, \hat{Q}\,(s'\_, a^*; \theta^-)}$;
   Regarding the network parameters $\boldsymbol{\theta}$, the agent carries out a gradient descent step on $(\boldsymbol{y\_ - Q\,(s\_, a\_; \theta)})^2$;
   $\boldsymbol{\hat{Q} = Q}$ is reset after each step;
   End **While**
8. Until the predetermined maximum number of iterations has been reached.

A higher-level action must be chosen at random to create the initial state before the iteration may begin. Obtaining the present state c, choosing the action in accordance with the rule π(.|s) in present state c, creating the state c_, computation of utility function G, and computation of reward r, constitutes in iteration. The present state is accessible at the start of each iteration. Combining the DQN algorithm and ϵ-greedy DQN policy, the selection of a better action in each iteration is done using $a_t = \text{argmax}_a Q(\psi(s_t), a; \theta$. This action comprises of bids from each MVNO, $a = b_1, b_2, ..., b_m$. The BS allocates the bandwidth resources to each MVNO in proportion, $r = r_1, r_2, ..., r_m$, based on the bid values,b_m, using equation (11). The allotment of bandwidth r_m to the linked users and rate, v_m, adding up is performed by each MVNO. Then the calculation of the proportionality between the allotted bandwidth and minimum required rate is performed by each MVNO and uses the output as the base for the calculation of state, which is s_. When assigning bandwidth to users, the MVNO also creates an action space and table L can be used to determine the best lower-level action a_l for each upper-level activity. MVNO calculates a discount function based on Equation (13). The MVNO then uses Equation (12) to find the profit value y_m in the present iteration based on the addition of v_m and p_m. Once all MVNOs in the present iteration have completed the aforementioned tasks, the system's total reward, r, and entire utility function, F_u are calculated.

The training of DQN is done by taking input as s, a, s_ , and r produced in the current iteration. The memory pool L in DQN is recorded with the transition parameters (s, a, s′, and r) in each iteration and selects random transition parameters (s, a, s′, and r) to train the Q-value net parameters. After this loss function Loss(θ) is used for updation of the target Q-value net parameters.

The Dueling DQN algorithm is used in Algorithm 2 to address the lower-level model's resource allocation task which in turn optimises the MVNO's profit. The Dueling DQN neural network's parameters (Q, θ, Q̂, and N) are initialised initially, as in Algorithm 1, and upon collecting the resources r_m from the BS, MVNOs generate the lower-level action space A_l. The random selection and execution of an action, $a \in A_l$, from the generated action space is performed by each MVNO, prior to the iteration.

The following tasks are performed by the generated action a, block-splitting of resources, allotment of resources among linked customers, calculation of user's successful reception of packets, p_k^m , and representation of this as state c. After this the iteration begins and the agent after obtaining present state c, selects the action in accordance with the Dueling DQN policy, i.e., ϵ-greedy policy. The selection of a better action is done using $a_t = \text{argmax}_a Q(\psi(s_t), a; \theta)$. Following the allocation procedure, MVNO calculates the state c′, utility function f_u and reward r, and then feed parameters (s, a, s, r) to Dueling DQN for training the neural network for the predefined rounds of iterations.

Algorithm 2: Bandwidth allocation using Dueling DQN technique

1. **Initialize** parameters: action-value function Q, target action-value function $\hat{\boldsymbol{Q}}$, memory pool L, capacity N;
2. The BS provides bandwidth $\boldsymbol{r_m}$ to each MVNO;
3. An action space $\boldsymbol{A_l}$ is created by each MVNO;
4. **While** $\boldsymbol{1 \le m \le M}$ **do**

   The MVNO decides a random action $\boldsymbol{a \in A_l}$ and executes it;

   Users connected to the MVNO are given access to the bandwidth r_k^m;

   Determine the p_k^m using state s;

   While $\boldsymbol{1 \le t \le T}$ **do**

   The present state s is obtained by the agent;

   Choose an action $\boldsymbol{a \in A_l}$ in accordance with Dueling DQN's rules;

   Determine the total system utility $\boldsymbol{f_u}$ based on (22);

   Determine the overall reward;

   The agent allocates the available bandwidth among the users and determines the iteration's post-selection state as s';

   Each MVNO acts as the agent by providing (s, a, s', and r) values to Dueling DQN;

   The agent stores transition values (s, a, s', r) in the *memory pool*;

   The agent chooses a subset of transition values ($s\_$, $a\_$, $s'\_$, and $r\_$) from memory pool in random;

   If iteration stops at step_+1

   Set the value for $y\_$ *as* $r\_$;

   Else

   Set the value for y_ as $\boldsymbol{r\_ + \gamma max_{a*}\, \hat{Q}(s'\_, a*; \theta^-, \beta, \delta)}$;

   Regarding the network parameters $\boldsymbol{\theta}$, $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, the agent performs a gradient descent step on $\boldsymbol{(y\_ - Q(s\_, a\_; \theta, \beta, \delta))^2}$;

   $\boldsymbol{\hat{Q} = Q}$ is reset after each step;

   End **While**

5. End **While**

V. Performance Analysis

In this section, the comparison of the proposed algorithm with the existing ones is performed. The evaluation of the proposed technique is carried in two parts. The first part is to evaluate the proposed algorithm with respect to system energy efficiency and the second part is to evaluate the proposed algorithm in terms of MVNO profit.

A. Experimental Environment

Tensor-flow 2.0 and the Simulation of Urban Mobility (SUMO) simulator were used to run the simulations [52]. The multi-cell H-CRAN environment was taken into account in the simulations. Table II summarizes the parameters.

TABLE II. Energy Maximisation Parameters Used in the Simulation

Parameter	Notation	Value
Noise power spectral density	N_p	-174 dBm/Hz
Total bandwidth	BW^{total}	100 MHz
SINR threshold	γ^o	0.5 dBm
Maximum outage probability constraint	θ_{max}	0.05
Circuit power of RRH	$\pi_{\mathbb{R}}$	4.3 W
Slope of RRH	Δslope	4.0
Circuit power of front-haul transceiver and switch	$P_{front\text{-}link}$	13 W
Power consumption per bit/s	ψ	0.83 W
Transmission power of V2I mode	p_k^x	23 dBm
Transmission power of RRH	p_k^s	24 dBm
Cluster size of RRH grouping	N_r	5

The parameters and system specifications have been established in line with the 3GPP specifications Releases 15 and 16 [36], [53]. Additionally, the path-loss model (140.7 + 36.7 log(distance)) and Rayleigh fading were taken into account [54]. A fully connected neural network with an input layer, a hidden layer, and an output layer, is the Dueling DQN used for the simulation in the resource allocation part. In the hidden layer, there were 256 neurons, and ReLu was used as an activation function. For the Dueling DQN, the following parameters were chosen. Learning rate, $\alpha = 0.01$, discount factor $\beta = 0.9$ and γ value of 0.95. The network update frequency is 2, the replay memory capacity is 3000, and the desired network update frequency is 30. Every unit of an episode a network upgrade takes place. 10 ms of delay and a 3 dB outage threshold were required for dependability.

The Python platform which runs on a Mac M1 system with 16GB RAM is used for modelling the proposed two-tier Dueling DQN algorithm along with the existing algorithms. Graphs are plotted and compared once the data from the four algorithms has been obtained. The suggested approach in this research is found to be feasible and provides several advantages over the other three techniques.

B. Results and Discussions

The evaluation of the proposed algorithm is described here. 43 RRHs were placed in the middle of junctions in the multi-cell environment, and the vehicles were then dispersed throughout the routes. The mobility and intersection simulations were performed using the Luxembourg SUMO traffic Scenario [55] dataset to build a scenario that would satisfy standard criteria. The mentioned dataset helps in simulating real-time traffic necessities and mobility patterns.

The steps involved in using SUMO for urban mobility are network creation, defining vehicle types, numbers and routes and defining the simulation-specific parameters [56]. SUMO provides two essential commands, "NetGenerate" and "NetConvert," which are commonly used to create road networks by importing digital road maps. These commands allow the generation of three distinct types of alternative road networks: arbitrary networks, circular "spiderNet" connections, and grid networks like "Manhattan." Each generation algorithm offers a range of settings that allow users to customize the characteristics of the generated networks. Two different types of environments were created using the dataset: less traffic and high traffic. We employed the system energy efficiency metric to assess the efficacy of the suggested method. The possible data rate per unit of energy consumption is the system's energy efficiency. The proposed method is compared with two other existing algorithms, such as resource allocation using clustering and DQN named as "existing1", and resource allocation algorithm using DQN alone named "existing2".

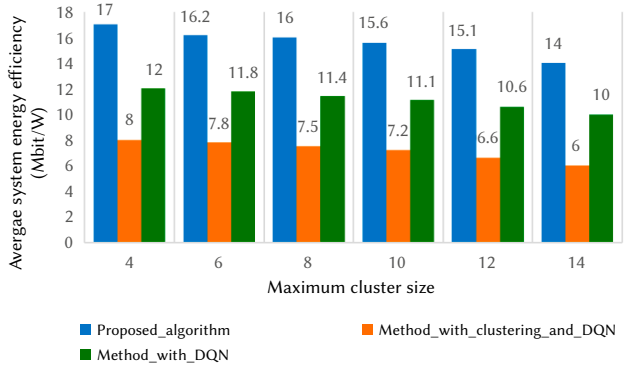


Fig. 4. Energy efficiency comparison in maximum scenario.

In Fig. 4 the comparison of the energy efficiency of the system for different traffic scenarios is performed. The figure depicts that energy efficiency is proportional to high-traffic situations. This is due to the fact that in high-traffic situations, the number of clustered VEs and VEs communicating using RRH are inversely proportional. As a result, the energy efficiency of the system rises as well. Additionally, it has been found that in identical density situations, the suggested algorithm has a greater system energy efficiency than existing1, even though existing1's technique also depends on the user clustering technique. In the proposed algorithm, there are fewer V2I mode VEs than in the existing1. This leads to minimized system energy usage and interference, increasing the possible data throughput.

For evaluating the efficiency of the proposed resource allocation algorithm, three existing algorithms such as the Double DQN algorithm, DQN algorithm, and Q-Learning algorithm are used for comparison. The comparative analysis is shown in Figures 5-9. The maximisation objective's importance weight, which was determined by formulas (6) and (15), is set to $\rho = 0.01$, $\psi = [1, 1, 1]$, and $\omega = 0.1$.

For depicting the enhanced QoE of the proposed algorithm with the existing ones the URLLC service scenario is adopted since this service has more impact on vehicular communication. Fig. 5 shows the QoE curves for the URLLC service type. Compared to the other three algorithms, the Dueling DQN algorithm's QoE curves for the three services are more consistent and less variable. The advantage of DRL over QL is depicted in the SE (Fig. 6) and system utility (Fig. 7) graphs. The Dueling DQN algorithm produces curves for SE and system utility that are greater in value than the curves produced by the other methods. These curves converge and are stable at the greatest values (SE > 300, utility > 6). The real simulation data reveals that after 2200 iterations, the Dueling DQN method's SE is around 1 per cent more than the DQN algorithm, approximately 2.6 per cent more than the Double DQN algorithm, and approximately 79 per cent more than the QL algorithm. There is a marginal enhancement in the utility also.

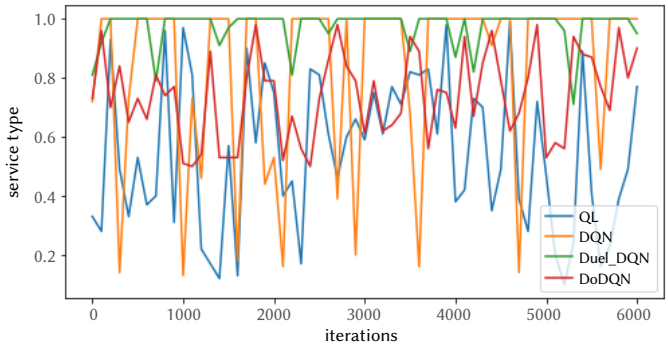


Fig. 5. Comparison of QoE curves.

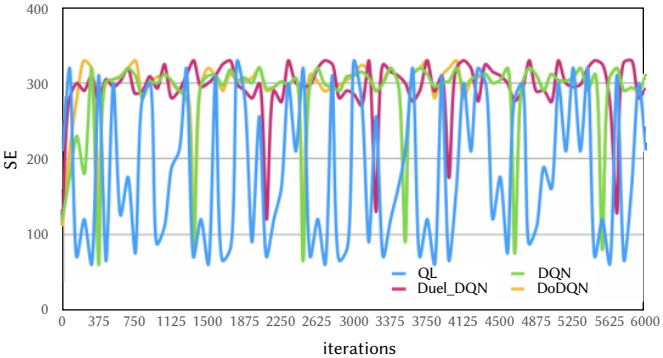


Fig. 6. Comparison of SE curves.

The SE and utility graphs exhibit the advantages of the proposed technique. Through comparison, it is determined that the Dueling

DQN algorithm's graph shows a higher stability than the compared algorithms. The curve produced by the competing DQN algorithms seldom changes significantly, even after 200 iterations, and the average value of the competing algorithms converges to a comparatively large value, demonstrating the effectiveness of the Dueling DQN algorithm in achieving the effective resource allocation which in turn maximises the MVNOs profit.

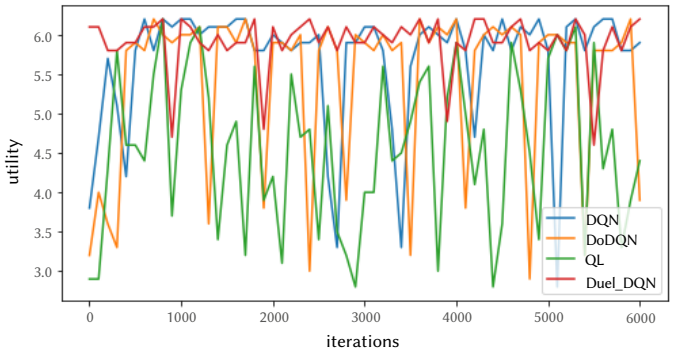


Fig. 7. Comparison of utility curves.

Figures 8 and 9 compare higher model outputs, i.e., MVNO's profit and system's utility, with other existing methods. The line graph clearly shows that the higher model's optimization result utilising the DQN technique (blue curve) is the best. From the graphs, it is clear that after 3500 iterations, profit value and utility value progressively converge to 205 and 8, respectively. The QL algorithm graph's values are noticeably larger than the other two but are more volatile after 3600 iterations than the DQN algorithm. The performance of the curve produced by the Double DQN algorithms is lower.

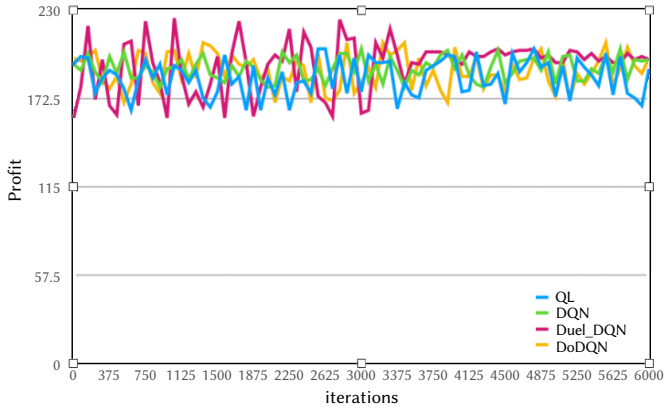


Fig. 8. Comparison of profit curves.

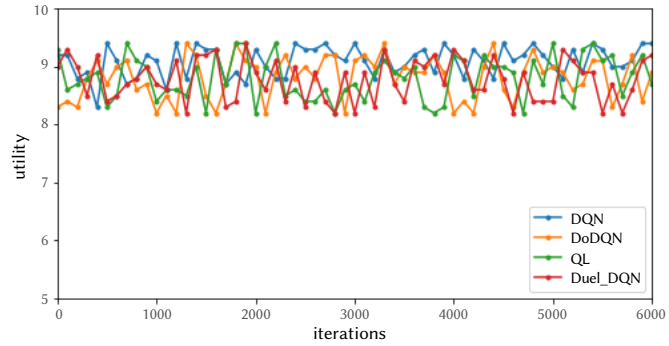


Fig. 9. Comparison of utility curves.

VI. Conclusion and Future Works

In this work, a resource allocation management system for 5G cellular V2X communication is proposed based on the clustering technique and DRL with the aim of maximising system energy efficiency and MVNO's profit. DRL is used to distribute communication resources for the best interference control in high-mobility scenarios. The creation of RRH grouping and vehicle clustering techniques is to reduce communication complexity and signalling overhead in DRL deployments. The overall architecture is implemented in two phases. The first phase addresses the RRH grouping and vehicle clustering technique with the objective of maximising the energy efficiency of the system and the second phase addresses the technique of employing DRL in conjunction with bidding to optimise MVNO's profit. The second phase addresses resource allocation which is implemented in two levels of the stage. At the first stage, the higher level of the proposed work integrates the Dueling DQN and bidding techniques with the aim of maximising the usage of BS resources. After this, the system uses a comprehensive listing to obtain the ideal lower-level actions that correspond to the higher-level actions and then uses a penalty function to limit the upper bidding range of MVNOs. In the lower layer of the scheme, the Dueling DQN is used to distribute resources among the linked users in each MVNO. Additionally, this work performed the merging of bidding with the Q-learning algorithm in the model's higher layer, to draw the conclusion that the Dueling DQN algorithm demonstrates superior performance.

The future work includes the integration of federated learning for accurate dynamic resource allocation with respect to real-world scenarios by collecting channel information from each vehicle in the network and dynamic updates of user location and required services.

References

[1] X. You, CX. Wang, J. Huang, et al. "Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts," *Science China Information Sciences* 64, 110301 (2021). https://doi.org/10.1007/s11432-020-2955-6.

[2] L. Feng, W. Li, Y. Lin, L. Zhu, S. Guo and Z. Zhen, "Joint computation offloading and URLLC resource allocation for collaborative MEC assisted cellular-V2X networks," *IEEE Access*, vol. 8, pp. 24914-24926, 2020.

[3] J. -W. Ke, R. -H. Hwang, C. -Y. Wang, J. -J. Kuo and W. -Y. Chen, "Efficient RRH activation management for 5G V2X," in *IEEE Transactions on Mobile Computing*, vol. 23, no. 2, pp. 1215-1229, Feb. 2024, doi: 10.1109/TMC.2022.3232547.

[4] M.A. Thanedar and S.K. Panda, "A dynamic resource management algorithm for maximizing service capability in fog-empowered vehicular ad-hoc networks," *Peer-to-Peer Networking and Applications* 16, 932–946 (2023). https://doi.org/10.1007/s12083-023-01451-7

[5] B. Fu, Z. Wei, X. Yan, K. Zhang, Z. Feng and Q. Zhang, "A game-theoretic approach for bandwidth allocation and pricing in heterogeneous wireless networks," 2015 *IEEE Wireless Communications and Networking Conference (WCNC)*, New Orleans, LA, USA, 2015, pp. 1684-1689, doi: 10.1109/WCNC.2015.7127721.

[6] A. R. Elsherif, W. -P. Chen, A. Ito and Z. Ding, "Resource allocation and inter-cell interference management for dual-access small cells," in *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 6, pp. 1082-1096, June 2015, doi: 10.1109/JSAC.2015.2416990.

[7] S. Tang, Z. Pan, G. Hu, Y. Wu and Y. Li, "Deep reinforcement learning-based resource allocation for satellite internet of things with diverse QoS guarantee," *Sensors* 22, (2022).

[8] N. C. Luong, D. T. Hoang, S. Gong, et al., "Applications of deep reinforcement learning in communications and networking: A survey," in *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3133-3174, Fourthquarter 2019, doi: 10.1109/COMST.2019.2916583.

[9] D. Bega, M. Gramaglia, M. Fiore, A. Banchs, and X. Costa-Perez, "DeepCog: cognitive network management in sliced 5G networks with deep learning," in *Proceedings of the IEEE INFOCOM 2019-IEEE Conference*

on Computer Communications, pp. 280–288, IEEE, Paris, France, July 2019.

[10] J. Gante, G. Falcão, and L. Sousa, "Deep learning architectures for accurate millimeter wave positioning in 5G," *Neural Processing Letters*, vol. 51, no. 1, pp. 487–514, 2020.

[11] D. Huang, Y. Gao, Y. Li, et al., "Deep learning based cooperative resource allocation in 5G wireless networks," *Mobile Networks and Applications*, pp. 1–8, 2018.

[12] P. Yu, F. Zhou, X. Zhang, X. Qiu, M. Kadoch, and M. Cheriet, "Deep learning-based resource allocation for 5G broadband TV service," *IEEE Transactions on Broadcasting*, vol. 66, no. 4, pp. 800–813, 2020.

[13] A. Pradhan and S. Das, "Reinforcement learning-based resource allocation for adaptive transmission and retransmission scheme for URLLC in 5G," in *Advances in Machine Learning and Computational Intelligence*, Springer, Singapore, 2020.

[14] G. Zhao, M. Wen, J. Hao, and T. Hai, "Application of dynamic management of 5G network slice resource based on reinforcement Learning in Smart Grid," in *International Conference on Computer Engineering and Networks*, pp. 1485–1492, Springer, Singapore, 2020.

[15] Y. Abiko, D. Mochizuki, T. Saito, D. Ikeda, T. Mizuno, and H. Mineno, "Proposal of allocating radio resources to multiple slices in 5G using deep reinforcement learning," in *Proceedings of the 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)*, pp. 1-2, IEEE, Osaka, Japan, October 2019.

[16] P. Yu, J. Guo, Y. Huo, X. Shi, J. Wu, and Y. Ding, "Three-dimensional aerial base station location for sudden traffic with deep reinforcement learning in 5G mmWave networks," *International Journal of Distributed Sensor Networks*, vol. 16, no. 5, Article ID 1550147720926374, 2020.

[17] M. A. Salahuddin, A. Al-Fuqaha and M. Guizani, "Reinforcement learning for resource provisioning in the vehicular cloud," *IEEE Wireless Communications* 23, 128–135 (2016).

[18] Z. Li, C. Wang and C. J. Jiang, "User association for load balancing in vehicular networks: an online reinforcement learning approach," *IEEE Transactions on Intelligent Transportation Systems* 18, 2217–2228 (2017).

[19] Z. Khan, P. Fan, F. Abbas, H. Chen and S. Fang, "Two-level cluster based routing scheme for 5G V2X communication," *IEEE Access 7*, 16194–16205 (2019).

[20] 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 Journal* 7, 6380–6391 (2020).

[21] H. D. R. Albonda and J. Pérez-Romero, "Reinforcement learning-based radio access network slicing for a 5G System with support for cellular V2X," *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 291*, 262–276 (2019).

[22] H. D. R. Albonda and J. Pérez-Romero, "An efficient RAN slicing strategy for a heterogeneous network with eMBB and V2X services," *IEEE Access* 7, 44771–44782 (2019).

[23] T. Sanguanpuak, N. Rajatheva, D. Niyato and M. Latva-Aho, "Network slicing with mobile edge computing for micro-operator networks in beyond 5G," *International Symposium on Wireless Personal Multimedia Communications*, WPMC 2018-November, 352–357 (2018).

[24] B. Kartal, P. Hernandez-Leal and M.E. Taylor, "Using monte carlo tree search as a demonstrator within asynchronous deep RL", (2018). doi:10.48550/arxiv.1812.00045.

[25] N. Khumalo, O. Oyerinde and L. Mfupe, "Reinforcement learning-based computation resource allocation scheme for 5G fog-radio access network," *2020 Fifth International Conference on Fog and Mobile Edge Computing (FMEC)*, Paris, France, 2020, pp. 353-355, doi: 10.1109/FMEC49853.2020.9144787.

[26] G. Sun, Z. T. Gebrekidan, G. O. Boateng, D. Ayepah-Mensah and W. Jiang, "Dynamic reservation and deep reinforcement learning based autonomous resource slicing for virtualized radio access networks," *IEEE Access 7*, 45758–45772 (2019).

[27] G. Sun, G. T. Zemuy and K. Xiong, "Dynamic reservation and deep reinforcement learning based autonomous resource management for wireless virtual networks," *2018 IEEE 37th International Performance Computing and Communications Conference (IPCCC)*, Orlando, FL, USA, 2018, pp. 1-4, doi: 10.1109/PCCC.2018.8710960.

[28] G. Sun, H. Al-Ward, G. O. Boateng and G. Liu, "Autonomous cache resource slicing and content placement at virtualized mobile edge network," *IEEE Access 7*, 84727–84743 (2019).

[29] Y. Liu, J. Ding, Z. L. Zhang and X. Liu, "CLARA: A constrained reinforcement learning based resource allocation framework for network slicing," *Proc. - 2021 IEEE International Conference on Big Data, Big Data 2021* 1427–1437 (2021).

[30] Y. Hua, R. Li, Z. Zhao, X. Chen and H. Zhang, "GAN-powered deep distributional reinforcement learning for resource management in network slicing," *IEEE Journal on Selected Areas in Communications,* 38, 334–349 (2019).

[31] A. Gupta and S. Namasudra, "A Novel Technique for Accelerating Live Migration in Cloud Computing," *Automated Software Engineering* 29, 34 (2022). https://doi.org/10.1007/s10515-022-00332-2.

[32] D. Choudhary and R. Pahuja, "Improvement in quality of service against doppelganger attacks for connected network", *International Journal of Interactive Multimedia and Artificial Intelligence,* 7. 51. 10.9781/ijimai.2022.08.003.

[33] M. B. Ahmad, M. A. Shehu and D. E. Sylvanus, "Enhancing phishing awareness strategy through embedded learning tools: a simulation approach," v1, *OpenAlex*, Dec. 2023, doi:10.60692/Q9Y25-7W438.

[34] J. D. C. Little, "A proof for the queuing formula: L = λW," https://doi.org/10.1287/opre.9.3.383 9, 383–387 (1961).

[35] A. Ghosh, L. Cottatellucci and E. Altman, "Nash Equilibrium for Femto-Cell Power Allocation in HetNets with Channel Uncertainty," *2015 IEEE Global Communications Conference (GLOBECOM)*, San Diego, CA, USA, 2015, pp. 1-7, doi: 10.1109/GLOCOM.2015.7417510.

[36] G. Auer, V. Giannini, C. Desset, et al., „How much energy is needed to run a wireless network?," in *IEEE Wireless Communications*, vol. 18, no. 5, pp. 40-49, October 2011, doi: 10.1109/MWC.2011.6056691

[37] S. K. Sharma and X. Wang, "Toward massive machine type communications in ultra-dense cellular IoT networks: current issues and machine learning-assisted solutions," *IEEE Communications Surveys & Tutorials 22*, 426–471 (2020).

[38] D. Sempere-García, M. Sepulcre and J. Gozalvez, "LTE-V2X mode 3 scheduling based on adaptive spatial reuse of radio resources," *Ad Hoc Networks 113*, 102351 (2021).

[39] Specification # 23.303. Available at: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=840.

[40] T. T. Nguyen, N. D. Nguyen and S. Nahavandi, "Deep reinforcement learning for multiagent systems: a review of challenges, solutions, and applications," *IEEE Transactions on Cyber*netics 50, 3826–3839 (2020).

[41] K. Arulkumaran, M. P. Deisenroth, M. Brundage and A. Bharath, "A. deep reinforcement learning: A brief survey," *IEEE Signal Processing Magzine* 34, 26–38 (2017).

[42] W. Qiang and Z. Zhongli, "Reinforcement learning model, algorithms and its application," *2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC)*, Jilin, China, 2011, pp. 1143-1146, doi: 10.1109/MEC.2011.6025669.

[43] I. H. Sarker, "Machine learning: algorithms, real-world applications and research directions," *SN Computer Science* 2, 1–21 (2021).

[44] R. Li, Z. Zhao, Q. Sun, *et al.,* "Deep reinforcement learning for resource management in network slicing," in *IEEE Access*, vol. 6, pp. 74429-74441, 2018, doi: 10.1109/ACCESS.2018.2881964.

[45] R. S. Sutton and A. G. Barto, "Reinforcement learning: an introduction," in *IEEE Transactions on Neural Networks*, vol. 9, no. 5, pp. 1054-1054, Sept. 1998, doi: 10.1109/TNN.1998.712192.

[46] H. Jiang, R. Gui, Z. Chen, L. Wu, J. Dang and J. Zhou, "An improved Sarsa(λ) reinforcement learning algorithm for wireless communication systems," in *IEEE Access*, vol. 7, pp. 115418-115427, 2019, doi: 10.1109/ACCESS.2019.2935255.

[47] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," *ArXiv*, abs/1707.06347, 2017.

[48] C. J. C. H. Watkins and P. Dayan, "Q-learning. mach. Learn," 1992 83 8, 279–292 (1992).

[49] A. Haydari and Y. Yilmaz, "Deep reinforcement learning for intelligent transportation systems: a survey," *IEEE Transactions on Intelligent Transportation Systems* 23, 11–32 (2022).

[50] Q. Mao, F. Hu and Q. Hao, "Deep learning for intelligent wireless networks: a comprehensive survey," *IEEE Communications Surveys & Tutorials 20*, 2595–2621 (2018).

[51] T. -W. Ban, "An autonomous transmission scheme using dueling DQN for d2d communication networks," in *IEEE Transactions on Vehicular*

Technology, vol. 69, no. 12, pp. 16348-16352, Dec. 2020, doi: 10.1109/TVT.2020.3041458.

[52] P. A. Lopez, M. Behrisch, L. Bieker-Walz, *et al.*, "Microscopic Traffic Simulation using SUMO," *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, HI, USA, 2018, pp. 2575-2582, doi: 10.1109/ITSC.2018.8569938.

[53] Specification # 38.885. Available at: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3497.

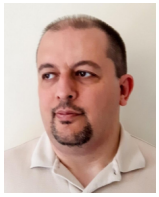
[54] Y. S. Song and H. K. Choi, "Analysis of V2V broadcast performance limit for WAVE communication systems using two-ray path loss model," *ETRI J.* 39, 213–221 (2017).

[55] L. Codeca, R. Frank, S. Faye and T. Engel, "Luxembourg SUMO Traffic (LuST) Scenario: Traffic Demand Evaluation," in *IEEE Intelligent Transportation Systems Magazine*, vol. 9, no. 2, pp. 52-63, Summer 2017, doi: 10.1109/MITS.2017.2666585.

[56] R. Monga and D. Mehta, "Sumo (Simulation of Urban Mobility) and OSM (Open Street Map) implementation," 2022 11th *International Conference on System Modeling & Advancement in Research Trends (SMART)*, Moradabad, India, 2022, pp. 534-538, doi: 10.1109/SMART55829.2022.10046720.

Rajilal Manathala Vijayan

Mr. Rajilal is serving as a Research Scholar in School of Computing at SASTRA University, Thanjavur, Tamil Nadu, India. He received the M.Tech. degree in embedded systems and the B.tech. in electronics and communications engineering from SASTRA Deemed University, Thanjavur, India and ASIET, Kalady, India, in 2006 and 2019, respectively. His research has centered on Machine Learning, 5G cellular network. He is possessing 11 years of rich Industrial experience in the field of Automotive & Embedded systems.



Fabrizio Granelli

Dr. Fabrizio Granelli is Associate Professor at the Dept. of Information Engineering and Computer Science (DISI) – University of Trento (Italy), IEEE ComSoc Director for Educational Services (2018-19) and Chair of Joint IEEE VTS/ComSoc Italian Chapter. He is Research Associate Professor at the University of New Mexico, NM, USA. He received the M.Sc. and Ph.D. degrees from University of Genoa, Italy, in 1997 and 2001. His research interest includes like Cloud computing, Communication networks , Mobile networks, Smart grids, Wireless networks.

A. Umamakeswari

Dr. Umamakeswari Arumugam received her B.E., from A.C.C.E.T, Karaikudi, M.E., from NIT, Trichy and Ph.D., degree from SASTRA University, Thanjavur, India. Currently she is working as Dean in School of Computing, SASTRA University. Her research area includes IoT, Wireless Sensor Network, Cloud computing, Embedded system and Blockchain. She has visited Hungary, Japan and Singapore. She has published 168 papers in SCOPUS / SCI - SCIE indexed journals and conference proceedings.