Received: October 25, 2024. Revised: November 15, 2024.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Information Volume Evidential Markov Decision Process Based Hierarchical Meta Reinforcement Learning for Resource Allocation in Vehicular Networks

Irshad Khan[1]* **Manjula Sunkadakatte Haladappa[1]**

[1]*Department of Computer Science and Engineering,*
University of Visvesvaraya College of Engineering, Bengaluru, India
* Corresponding author's Email: research.irshad@gmail.com

Abstract: Resource allocation in vehicular networks defines the strategic distribution of accessible communication resources like time slots, power, and bandwidth between infrastructures and vehicles to ensure the effective transmission of data. It handles a dynamic network condition and balances the requirements of different users while reducing interference and increasing network effectiveness. However, resource allocation faces challenges in effectively handling limited communication resources due to high energy consumption and network delays caused by frequent changes in network topology and varying user demands by the high mobility of vehicles. This research proposes the Information volume Evidential Markov Decision Process-based Hierarchical Meta Reinforcement Learning (IEMDP-HMRL) for resource allocation in vehicular networks. In traditional RL, the IEMDP-HM is incorporated to enhance resource allocation by enabling adaptive decision-making in dynamic environments. HMRL enables rapid adaptation to new tasks by leveraging learned policies from prior experiences which increases both efficiency and flexibility in resource management. In scenario 3, the IEMDP-HMRL achieves a lesser average task delay of 0.34 ms for the number of vehicles 10 compared to existing methods like Intelligent Distributed Resource Allocation and Task Scheduling (IRATS).

Keywords: Delay, Hierarchical meta reinforcement learning, Information volume evidential Markov decision process, Resource allocation, Vehicular networks.

1. Introduction

Resource allocation refers to the process of effectively allocating available resources across different tasks to increase performance or outcomes. In vehicular networks, the goal is to determine road safety, maximise traffic effectiveness, and create a new level of onboard entertainment. To achieve this, vehicles are required to communicate with other entities for data exchange, which is known as Vehicle-to-Everything (V2X) [1]. V2X involves Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure/Network (V2I/N) communication, which enables data exchange among vehicles and nearby infrastructure [2]. In vehicular networks, V2V is a significant communication mode that generates a primary basis for vehicular data exchange among vehicles [3]. Furthermore, vehicular network differs significantly from traditional cellular networks for strong dynamics in network topology and channel conditions, which leads to high mobility of vehicles [4, 5]. These features enable the design of effective resource allocation methods, which becomes challenging for V2V systems [6, 7]. Vehicles are unified with Onboard Units (OBU) in vehicular communication for communicating with Road Side Units (RSU) organised with the road [8]. Also, RSUs play a significant role as Base Station (BS) and function as internet access points and routers [9].

Moreover, the computational volume limitation for RSU strengthens the resource competition [10] [11]. Meanwhile, vehicular networks consume an enhancing demand for wide resources associated with networking, caching, and computing to efficiently handle the resources is a significant problem in vehicular networks [12, 13]. The resource allocation process is difficult in Virtual Machine

(VM) migration as it involves determining an appropriate physical server for locating VMs [14]. With the rise of innovative vehicular applications, particularly in the era of autonomous driving, vehicular networks are established to offer high-bandwidth services [15]. V2V offloading significantly increases the system's computing ability and eases the load of RSU [16, 17]. Artificial Intelligence (AI) based methods are increasingly being utilised to optimise resource allocation, which improves the decision-making process and enhances overall system effectiveness in vehicular networks [18]. Nevertheless, it struggles with difficulties in managing limited communication resources due to high energy consumption and network delays that lead to frequent changes in network topology and varying user demands by high mobility of vehicles. To address this problem, the IEMDP-HMRL is proposed for resource allocation in vehicular networks by applying hierarchical structures to adaptively handle limited communication resources effectively.

The main contribution of this research is explained below,

- IEMDP-HMRL manages the dynamic vehicular network by its hierarchical structure, which permits rapid adaptation to changes in network topology and user demands.
- Meta-learning enables the system to rapidly adapt learned policies to new scenarios, which enhances overall performance in dynamic environments.
- By performing this process, resource allocation in vehicular networks becomes more effective and results in less delay and enhanced connectivity for users.

This research paper is structured as follows: Section 2 provides an existing method's literature survey and Section 3 illustrates a brief explanation of the proposed methodology. Section 4 indicates an experimental result, and the conclusion of this research paper is given in Section 5.

2. Literature survey

Zhang [19] introduced Multiagent Reinforcement Learning (MARL) termed as Complete-Game-MARL (CG-MARL) and Mean-Field MARL (MF-MARL), for resource allocation in vehicular networks. The power allocation and joint spectrum management in vehicular communication systems were determined by considering the interactions between environments and vehicles. This was achieved by integrating the cooperative stochastic game theory with MARL, which led to improved stability and faster convergence. However, MARL was challenging in managing dynamic and unpredictable environments because the agents struggled to adapt to rapidly changing network conditions.

Mafuta [20] presented a Multi-Agent Double Deep Q-Network (MA-DDQN) to alleviate the system and increase the Vehicle-to-Infrastructure (V2I) capacity links by satisfying the delay and reliability constraint for V2V links. The selection of transmission mode was applied to avoid interference produced by unstable V2V links in the scheme design. Also, a binarised weight approach was established to increase the learning process of deep neural networks, which assists with computational complexity. Nevertheless, the MA-DDQN struggled in scaling with a large number of agents because of the exponential increase in state-action space.

Gao [21] developed a two-layer optimization approach to address the problem of resource allocation to reduce the task completion of energy consumption and delay in vehicular networks. In the upper layer, tasks offloading and scheduling were applied to improve the CPU frequency allocation, which obtained the scheduling and offloading decisions. However, the two-layer optimization suffer from high energy consumption due to the complexity of addressing both layers simultaneously particularly in dynamic environments with limited resources.

Shu and Li [22] established a Quantum particle swarm optimization for a joint offloading and resource allocation scheme in a vehicular network. To reduce the energy consumption and delay cost, a task execution optimisation was designed to evaluate the task to available service nodes that contain service vehicles and RSU. Then, a vehicle selection approach was used to acquire the best offloading decision sequence for the task offloading process through V2V communication. Nevertheless, the established approach face inefficiency in exploring the large and dynamic solution space because of quantum state collapse which degrades the model performance in vehicular network.

Jamil [23] suggested a Proximal Policy Optimization (PPO) for Intelligent Distributed Resource Allocation and Task Scheduling (IRATS) in vehicular networks. IRATS determined the resource allocation issue using the Markov decision process to reduce the delay of tasks and waiting time. The task scheduler was designed for vehicles to share their idle resources tasks based on priorities by employing multi-level queues. However, the PPO cause high average task delay because of inefficient

exploration in highly dynamic environments which leads to suboptimal performance.

Hang Fu [24] established a Dense Multi-Agent RL (DMARL) aided Multi-Unmanned Aerial Vehicle (UAV) coverage data for vehicular networks. By generating the critical state approach, the MDP was applied which allows the selection of state for a more effective training process. Furthermore, the established DMARL improves the effectiveness in training and distribution within the multi-UAV system. Nevertheless, the DMARL struggled with scalability as the number of UAVs and network agents increases cause potential ineffectiveness and delayed responses in dynamic environments.

Amjad Alam [25] developed a Particle Swarm Optimization (PSO) for resource allocation and joint computational task offloading in vehicular edge networks. The computational efficiency for Connected Autonomous Vehicle (CAV) was determined by enabling an optimized decision on allocating resources. The developed PSO enhance the overall system energy effectiveness by increasing resource allocation in energy constraints. However, the developed approach suffers from premature convergence which leads to suboptimal resource allocation.

In the overall analysis, the existing method had limitations like challenges in scaling with a large number of agents, high energy consumption, inefficiency in exploring the large and dynamic solution space, and high average task delay. To solve this issue, the IEMDP-HMRL is proposed for resource allocation in vehicular networks by using a hierarchical approach that effectively handles the dynamic nature. The proposed approach minimizes energy consumption and adapts policies depending on information volume that makes better energy usage. Moreover, the IEMDP-HMRL reduces average task delay via meta-learning by learning optimal decision-making policies which provides rapid adaptation in network conditions.

3. Proposed methodology

This research proposes IEMDP-MHRL for resource allocation in vehicular networks. The IEMDP processes the data and defines possible states, actions, and rewards for optimal decision-making.

In HRL, the top-level handles broad strategic decisions, whereas the low-level manages more detailed with fine-tuned decisions. The Meta-learning mechanism makes the system adapt rapidly to new scenarios, which enhances decision-making efficiency. Fig. 1 shows a block diagram for the IEMDP-MHRL technique.

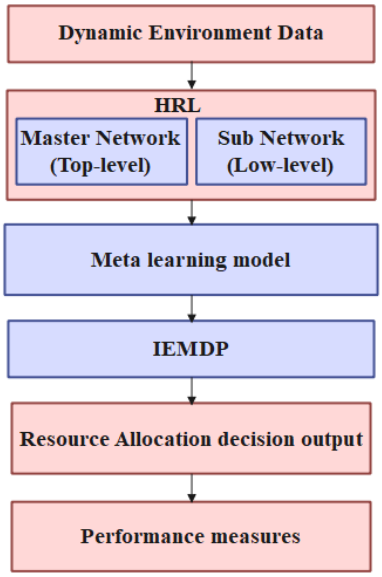

Figure. 1 Block diagram of the proposed IEMDP-MHRLtechnique

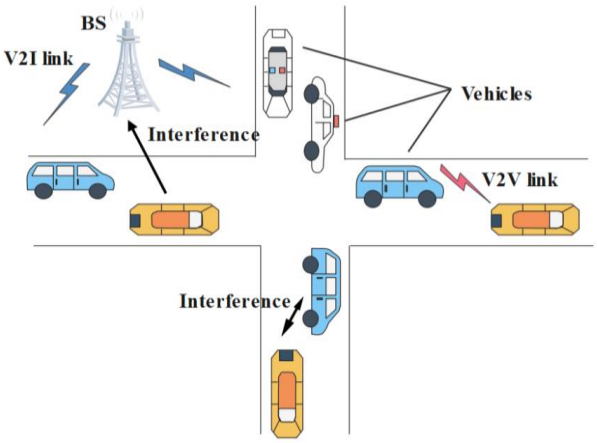

Figure. 2 System model

3.1 System model

Fig. 2 represents a system model and a vehicular network consists of a BS and numerous Vehicle User Equipment (VUE) devices. As shown in Fig 2, the BS is located at a crossroad's center, whereas the VUEs are located on the roads. The BS and VUEs are all evaluated with a single antenna. V2VU and V2IU represent VUE communicating through V2V and V2I links correspondingly. Consider that there are M V2I Unit (V2IU) and K V2V Unit (V2VU) in the network environment. Especially, the V2IU necessitates a V2I connection link via the interface to generate high-ability communication with BS, whereas V2VU requires a V2V link to share the data via the PC5 interface for effective management of traffic safety. Each BS b It associates with a cache to determine if the content is cached or not, depending on a certain probability. The binary variable for state i with BS b

is applied to define it which is represented as $\emptyset_i^b$. It involves two cases: $\emptyset_i^b = 1$, the state i occurs in cache and $\emptyset_i^b = 0$ otherwise. It establishes a two-state Markov chain $H = \{0,1\}$ and employs a $\emptyset_i^b(t)$ to evaluate cache at time t where $t = \{0,1, \dots, T-1\}$. The cache state transmits from one state to another based on the probability of state transition $\Gamma_i(t)$ for state i at time t using Eq. (1).

$$\Gamma_i(t) = [Y_{z1,z2}(t)]_{2\times 2} \quad (1)$$

$$Y_{z1,z2}(t) = \Pr\left(\emptyset_i^b(t+1) = z_2 | \emptyset_i^b(t) = z_1\right) \quad (2)$$

Where $\emptyset_i^b(t)$ determines binary variable at time t for state i with BS b,Pr represents probability, and $Y_{z1,z2}(t)$ denotes probability of transitioning from cache state z_1 to z_2 at time t which is indicated in Eq. (2).

3.2 Computing model

This model contains two components: content size and an appropriate number of CPU cycles to provide the request from vehicles using $Tr = \{o_r, q_r\}$. Where Tr indicates vector at vehicle r, o_r represents content size for vehicle r, and q_r determines CPU cycles to process vehicle r. Based on the probability of state transition, the $M_e^b(t)$ denotes computing ability at time t for an entity e with BS b that is changed from 1 state to another. For instance, changing from pair of states q_1 to q_2 at time t is represented as $\Omega_{q_1,q_2}(t)$. Hence, the $N_e^b(t)$ determines computing power of state transition probability at time t for an entity e with BS b is applied as $n \times n$ the matrix using Eqs. (3) and (4).

$$N_e^b(t) = [\Omega_{q_1,q_2}(t)]_{n\times n} \quad (3)$$

$$\Omega_{q_1,q_2}(t) = \Pr\left(M_e^b(t+1) = q_2 | M_e^b(t) = q_1\right) \quad (4)$$

q_r represents the task's computation execution for vehicle r at a multi-user server is expressed using Eq. (5), and the computation rate is represented in Eq. (6).

$$t_e^b = \frac{q_r}{M_e^b(t)} \quad (5)$$

$$R_{r,e}^{comp}(t) = \frac{o_r}{t_e^b} = \frac{M_e^b(t) o_r}{q_r} \quad (6)$$

Also, the $U_r^b(t)$ is used to evaluate whether the multi-user server for vehicle r with BS b at time t is related with BS b is chosen; if it is $U_r^b(t) = 1$ else, $U_r^b(t) = 0$. $R_{r,e}^{comp}(t)$ represents computing resource for vehicle r and entity e at time t and $\frac{o_r}{t_e^b}$ represents ratio of content size vehicle r to the time required to process the vehicle at entity e with BS b. The Mx_e indicates maximum amount of requests that is established simultaneously multi-user server using Eq. (7).

$$\sum_{r\epsilon R} \sum_{b\epsilon B} U_r^b(t) o_r \le Mx_e \quad (7)$$

3.3 Communication model

The data is transferred over a wireless channel in a communication model, and those wireless channel among vehicles and BS is considered as a realistic time-varying channel. The received Signal-to-Noise Ratio (SNR) is applied as a parameter to determine the channel quality. The random variable k_r^b Is defined to evaluate the received SNR among BS b and vehicle r. Based on transition probability, the obtained SNR$C_r^b(t)$ changes from 1 state to another. For instance, changing from a state d_1 to d_2 at time t is represented as $\Upsilon_{d_1,d_2}(t)$. The $h \times h$ matrix $G_r^b(t)$ is applied to determine state transition probability among BS b and vehicle r using Eqs. (8) and (9).

$$G_r^b(t) = [\Upsilon_{d_1,d_2}(t)]_{h\times h} \quad (8)$$

$$\Upsilon_{d_1,d_2}(t) = Pr(C_r^b(t+1) = d_2 | C_r^b(t) = d_1) \quad (9)$$

The G_r^b Hz is applied to determine the bandwidth which is allocated to vehicle r for available bandwidth of BS b, $C_r^b(t)$ indicates current state at time t, ($C_r^b(t+1)$ represents current state at next time step $t+1$, d_1 and d_2 denotes distance. The backhaul capability of BS b represented as L_b and the overall vehicle rate of BS do not exceed it backhaul ability, and their vehicle communication rate related to BS is determined using Eqs. (10) and (11). Where L_b indicates maximum allowable resource limit for BS b.

$$R_{r,b}^{comm}(t) = U_r^b(t) D_r^b(t) C_r^b \quad (10)$$

$$\sum_{b\epsilon B} \sum_{r\epsilon R} R_{r,b}^{comm} \le L_b, \forall b \epsilon B \quad (11)$$

3.4 Problem formulation

In this section, the resource allocation issue is transformed into the RL process and its has N number of various MDP scenarios. For all scenario, it is considered that there involves a BS, multi-user server, and content cache Ξ. Every BS is related to

the content cache and server while the effectiveness of the wireless channel between BS and vehicle is split into h levels and follows a Markov process. To determine channel effectiveness, the $C_r^b(t)$ is used among BS and vehicles at time t. Likewise, the computing power of multi-user server split into n levels and $M_e^b(t)$ is applied to evaluate the multi-user server computing power with BS at time t. Consider $H_\xi^b(t)$ indicate content saved in cache ξ related with BS and $H_\xi^b(t) = [\Phi_1(t), \Phi_2(t), \dots, \Phi_i(t), \dots, \Phi_n(t)], \xi\epsilon\{1,2, \dots, \Xi\}, \Phi_i(t) \epsilon \{0,1\}$ To effectively allocate resources in a vehicular environment is essential. The vehicular environment changes as vehicles move and during these changes, the proposed approach shares a low-level set of primitives that enhance the generalization of the learning model.

3.5 Information Volume Evidential Markov Decision Process based Hierarchical Meta Reinforcement Learning (IEMDP-MHRL)

The IEMDP-MHRL is used for resource allocation to manage uncertainty by incorporating evidence-based reasoning method in vehicular networks. The hierarchical structure enhances scalability by dividing allocation into low-level and high-level decisions which increase both immediate action and long-term performance. Also, IEMDP assists in adapting to changing network conditions which makes more effective and reliable resource distribution. Conventional RL [26] typically acts infeasible in complex tasks: the action and state spaces are greater; trajectories are higher; the tasks are complex domain and reward signals are sparse; and so on. Irradiated by human societies where a hierarchical organization is applied to address complex tasks, a hierarchical process is used in RL to solve more complicated issues. The purpose of HRL is observed as a divide-and-conquer scheme where a complicated task is hierarchically split into various smaller sub-tasks and then limited solutions are combined into a full and more cost-efficient solution for the issue. The hierarchical process establishes a reduction in computation, space, and time complexity for both learning and overall task execution. A sub-policies set is shared and switched among various tasks by master policy and meta-learning is integrated in learning new unseen tasks more rapidly.

While a controller obtains a vehicle request, it assigns a BS via HMRL that has two parts: master-policy network φ and sub-policy network y as $\{\omega_1, \omega_2, \dots, \omega_y\}$. When a request is received, the controller initially evaluates the master policy to determine the appropriate sub-policy based on present observation S_t in the vehicular network. Then, the chosen sub-policy is applied to generate a particular allocation policy. Based on state S_t, the master policy process specifies a sub-policy network using Eq. (12).

$$sub_k \sim \pi_\varphi(sub_k|S_t), sub_k \epsilon \{sub_1, sub_2, \dots, sub_y\} \quad (12)$$

Where $\pi_\varphi(.|S_t)$ represents master policy at present state S_t and k indicates sub-policies index. The master policy φ is reorganized by gradient method using Eq. (13).

$$\varphi^i = \varphi^{i-1} + \mu \nabla_{\varphi^i} r_H \quad (13)$$

Where μ determines learning rate, r_H indicates gain of master policy from sub-policy sub_k at S_t. Also, it computes a rewarded value at timestep T. Then, the gradient update φ is represented using Eq. (14).

$$\varphi^i = \varphi^{i-1} + \mu \sum_{t=0} \sum_t \beta^t R_H(S_t, Sub_t) \nabla_\varphi log \pi_\varphi(sub_k|S_t) \quad (14)$$

Once, the sub-policy is chosen, then the resource allocation phase is performed in that master-policy and sub-policies are focused based on environment. Then, the 2nd phase is determined using Eq. (15).

$$b_j \sim \pi_\omega sub_k(b_j|S_t), b_j \epsilon \{1,2, \dots, B\} \quad (15)$$

Where $\pi_\omega sub_k(.|S_t)$ represents a sub-policy network that evaluates a BS to request a vehicle in S_t, b_j determined chosen BS. Also, the sub-policy parameter is reorganized by the gradient approach and it computes the reward function at time t. Moreover, the BS information is reorganized; the request data is verified in the BS table, and then computing power, channel, and present cache environment are reorganized based on IEMDP. For modelling the decision-making progress, MDP [27] is utilized. The primary goal of IEMDP is to create an uncertain state in the decision system, various beliefs and actions below certain and uncertain states are established. Then, a mass function of various actions and beliefs is determined by the dynamic evolution of Markov model. Then, instead of Deng's entropy, the Information Volume (IV) of mass function is introduced to distribute uncertain belief states that consider the overall amount of data. At last, the probability distribution is acquired and its difference determines the disjunction effect. A new uncertainty

parameter η is calculated depending on IV using Eq. (16) for a probability distribution.

$$\eta = \frac{H_{IV}^{D} - H_{IV}^{CD}}{H_{IV}^{D} + H_{IV}^{CD}} \quad (16)$$

Where H_{IV}^{D} and H_{IV}^{CD} represents the amount of information at Information Volume (IV) in C-D condition. Hence, the transformation rules in the C-D condition and D-condition for the mass function is calculated using Eqs. (17) to (18).

$$P_{D}^{\prime n}(A) = m^{n}(AU) + \left(\frac{1}{2} + \eta\right) * m^{n}(UU) \quad (17)$$

$$P_{CD}^{\prime n}(A) = m^{n}(AU) + \frac{1}{2} * m^{n}(UG) + m^{n}(AB) + \frac{1}{2} * m^{n}(UB) \quad (18)$$

Where AU indicates interaction between input signal A and distortion or noise term U, UU represents self-interaction of a distortion and noise term U, UG determines interaction among distortion U and variable G, AB denotes interaction between input signal A and Bias B, m^{n}indicates measure m at n^{th}iteration, and UB represents distortion or noise term U and Bias B .The disjunction effects represented as $D_{is}^{\prime}$ is reproduced by the probability difference in C-D condition and D-condition (i.e., the difference among Eqs. (17) and (18).

$$D_{is}^{\prime} = P_{D}^{\prime n}(A) - P_{CD}^{\prime n}(A) \quad (19)$$

Where n represents the number of states in Eq. (19). IEMDP-MHRL improves resource allocation in vehicular networks by enhancing decision-making in dynamic environments. It incorporates information uncertainty management by evidential reasoning, making resource allocation more reliable. Also, hierarchical structure rapidly adapts to changing network conditions whereas meta-learning increases learning over tasks, leading to more effective resource allocation in vehicular communication.

Pseudo code for proposed IEMDP-HMRL method

Step 1: Initialize parameters for Master policy and Sub-policy networks

initialize master policyparameters (φ)

initialize subpolicy parameters($\omega_1, \omega_2, \ldots, \omega_y$)

Step 2: Define Learning Rates, Reward Functions and State Transitions

define learningrate (μ) using Eq. (13)

define rewardfunction (RH, RL)which is represented using Eq. (15)

define statetransitionmatrix (S) by applying Eq. (1)

Step 3: Master Policy Network chooses a Sub-Policy depending on present state

for each request obtained in a dynamic vehicular environment:

Observe present state S_t from vehicular network

Master policy chooses a sub-policy depending on the state

sub_k = selects policy depends on state

Update master policy network utilizing gradient approach

$\varphi = \varphi + \mu * \nabla\varphi *$ reward master policy ($R_H(S_t, sub\_policy\_k)$) using Eq. (13)

Calculate reward for master policy over time steps

Rewardmaster $\varphi^{i} = \varphi^{i-1} + \mu \sum_{t=0} \sum_{t} \beta^{t} R_H(S_t, Sub_t) \nabla_{\varphi} log \pi_{\varphi}(sub_k|S_t)$ by applying Eq. (14)

Update master policy gradient

φ= update master policy (φ, reward master)

Step 4: Sub-policy network chooses Base Station (BS) and allocates resources for subpolicy in selectedsubpolicy$k$:

for subpolicy in selectedsubpolicy$k$:

Sub- policy chooses a BS for resource allocation

Basestation j = ChooseBS ($\pi_{\omega}(sub\ policy\ k|S_t)$)

Update subpolicy parameters depending on reward at timestep t

ω subpolicy k = ω subpolicy $k + \mu *$ $\nabla\omega$subpolicy+ $k *$ rewardsubpolicy($RL(S_t, BS_j)$)

Calculate subpolicy reward

Rewardsub = sum ($\beta_t *$ π_{ω}(subpolicy$k|S_t$)*RL(S_t, BS_j)) using Eq. (12)

Update subpolicy gradient

ω subpolicy k = update subpolicy(ωsubpolicy$k$, rewardsub)

Step 5: Use Evidential Reasoning for uncertainty management for each belief and action:

massfunction = Calculatemassfunction (actions, beliefs) using Eq. (16)

informationvolume = Calculate informationvolume (mass function) by utilizing Eq. (17) and (18)

uncertaintyparameter = Calculate uncertainty (η)which is formulated in Eq. (16)

Use IVbased probability distribution

Pdistribution = updateprobabilitydistribution (informationvolume) using Eq. (19)

Step 6: Resource allocation and request processing in a dynamic environment

AllocateresourcestoBS (BS_j)

Updatestateinformation (BS_j, network conditions, computing power)

Step 7: Repeat steps for new vehicles requests and environment changes until a predetermined maximum number of iterations is reached.

4. Results

The proposed IEMDP-MHRL is simulated based on MATLAB R2020b with an i7 intel processor, 128 GB RAM, and Windows 10 operating system. The performance measures like Probability of satisfied V2V links, cost, energy consumption, task completion rate, Packet loss, Average task wait (ms), and Average task delay (ms) are calculated to identify the model performance for resource allocation in vehicular networks. The mathematical formula for the performance metrics is represented using (20) to (25). Table 1 represents a simulation setting.

$$Delay = \frac{Sum\ of\ transmitting\ and\ receiving\ time\ of\ packets}{Total\ number\ of\ packets} \quad (20)$$

$$Energy\ consumption = E_{transmitted} + E_{received} \quad (21)$$

$$Task\ Completion\ rate = \frac{Number\ of\ tasks\ completed}{Total\ no.of.tasks\ assigned} \quad (22)$$

$$Packet\ loss = \frac{Total\ packets\ sent - Total\ packets\ received}{Total\ packets\ sent} \quad (23)$$

$$Average\ task\ rate = \frac{Total\ number\ of\ tasks\ completed}{Total\ time\ taken} \quad (24)$$

$$Average\ task\ delay = \frac{\sum_{i=1}^{n}(Actual\ completion\ time_i - Expected\ completion\ time_i}{n} \quad (25)$$

Where n represent a total number of tasks and i indicates the number of tasks respectively.

4.1 Performance analysis

Table 2 determines a performance evaluation of energy consumption (J). The existing methods like RL, HRL, and MHRL are used to compare with the proposed IEMDP-MHRL method. When compared to these existing methods, the IEMDP-MHRL achieves less energy consumption of 0.09 J for many vehicles 2 because it leverages evidential reasoning to manage uncertainty which enhances decision-making efficiency. This results in better resource allocation and optimized action selection which minimize unwanted energy use.

Table 3 represents a performance analysis of task completion rate. The proposed IEMDP-MHRL method achieves a high task completion rate of 0.96 for vehicle 2 due to its efficient evidential decision-making process which enhances accuracy in choosing optimal actions under uncertainty. Its hierarchical structure makes exact task coordination over various layers which results in more effective task execution compared to RL, HRL, and MHRL.

Fig. 3 indicates a graphical representation of the average task wait (ms). The existing methods like RL, HRL, and MHRL are compared with IEMDP-MHRL. The proposed IEMDP-MHRL achieve a lower average task wait of 0.95ms for vehicle 10 by using evidential reasoning to enable rapid process and more accurate decisions under uncertainty which minimizes delays in task execution.

Its hierarchical meta-learning structure enables effective task prioritization and resource management over various layers. This leads to rapid task completion and reduced waiting times in vehicular environments. Fig. 4 denotes a graphical representation of the average task delay (ms). The proposed IEMDP-MHRL obtains a less average task delay of 0.35ms for vehicle 10 by including meta

Table 1. Simulation settings

Parameters	Values
Number of vehicles	2 to 10
Carrier frequency	2 GHZ
Speed of vehicles	36 to 54 km/h
Bandwidth	4 MHz
Remaining time for message delivery	100 ms

Table 2. Performance evaluation of energy consumption (J)

Methods	No. of vehicles				
	2	4	6	8	10
RL	1.19	1.21	2.25	2.29	2.46
HRL	0.21	0.24	0.33	0.56	0.60
MHRL	0.16	0.19	0.21	0.24	0.29
IEMDP-MHRL	0.09	0.12	0.14	0.17	0.23

Table 3. Performance analysis of task completion rate

Methods	No. of vehicles				
	2	4	6	8	10
RL	0.86	0.82	0.76	0.74	0.72
HRL	0.91	0.89	0.88	0.86	0.84
MHRL	0.93	0.87	0.86	0.84	0.81
IEMDP-MHRL	0.96	0.94	0.91	0.90	0.86

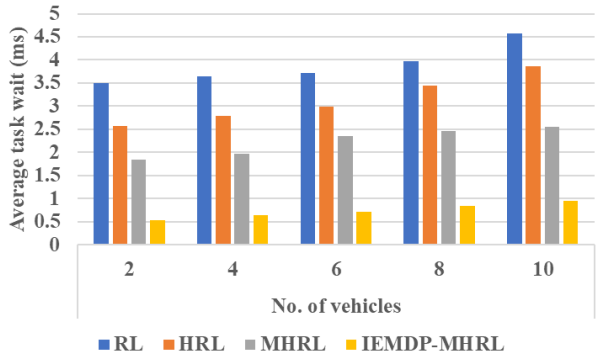

Figure. 3 Graphical representation of average task wait (ms)

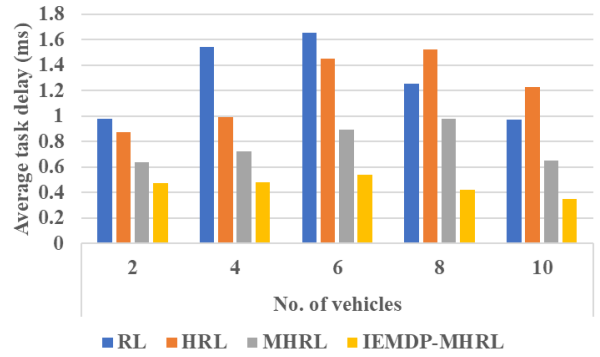

Figure. 4 Graphical representation of average task delay(ms)

learning and evidential methods that enhance the decision-making accuracy under uncertainty and minimize delayed or incorrect actions. Its hierarchical structure enables more effective task decomposition and rapid response in dynamic environments.

4.2 Comparative analysis

Table 4 demonstrates an analysis of simulation parameters. Tables 5, 6, and 7 present a performance analysis for the proposed IEMDP-MHRL with existing methods [20-23]. The scenario 1 is for existing method [20], scenario 2 is for [21], whereas scenario 3 is for [23]. The Vehicle speed, Number. of vehicles/ vehicle density, Bandwidth, and Diameter of RSU coverage/ Transmission range of RSU are considered for simulation parameter and those parameters values are presented clearly in Table 4. Vehicle speed reflects different traffic conditions, Vehicle density modeling varying congestion levels, bandwidth provides different network conditions, and Diameter of RSU coverage/ Transmission range of RSU represents varying infrastructure availability. When compared to these existing methods, the proposed IEMDP-MHRL obtains a better performance. For example, the proposed IEMDP-MHRL obtains a lesser average task delay of 0.34 ms for 10 vehicles due to its evidential decision-making process which minimizes uncertainty in state transition and results in rapid task execution in Table 7 compared to existing methods like IRATS [23]. The HRL enable parallelized task management over different layers which optimize resource distribution. Moreover, meta-learning increases the adaptability to dynamic environments which reduces delay in response time compared to [23].

Table 4. Simulation parameter

Parameters	Scenarios		
	1	2	3
Vehicle speed	36 km/h	22 m/s	12~20 km/h
Number. of vehicles/ vehicle density	20 to 100	16	10~50
Bandwidth	10 MHz	5×10^6 Hz	N/A
Diameter of RSU coverage/ Transmission range of RSU	N/A	500 m	3000 m

Table 5. Performance analysis of IEMDP-MHRL with existing MA-DDQN

Methods	Scenario	Performance measures	Number of vehicles				
			20	40	60	80	100
MA-DDQN [20]	1	Probability of satisfied V2V links	0.990	0.985	0.970	0.975	0.96
ProposedIEMDP-MHRL			0.999	0.992	0.986	0.989	0.98

Table 6. Performance analysis of IEMDP-MHRL with existing JDGO

Methods	Scenario	Performance measures	Number of vehicles						
			4	6	8	10	12	14	16
JDGO [21]	2	Cost	65	100	150	200	230	240	250
		Energy consumption	10	18	23	30	38	42	49
		Task completion rate	0.80	0.82	0.85	0.81	0.84	0.85	0.88
Proposed IEMDP-MHRL		Cost	45	63	72	87	120	155	179
		Energy consumption	5	12	17	23	29	34	38
		Task completion rate	1.00	0.97	0.92	0.90	0.89	0.88	0.87

Table 7. Performance analysis of IEMDP-MHRL with existing IRATS

Methods	Scenario	Performance measures	Number of vehicles				
			10	20	30	40	50
IRATS [23]	3	Percentage of completion task	72	55	59	51	42
		Packet loss	100	380	480	790	1000
		Average task wait (ms)	1.87	2.00	2.00	1.98	1.50
		Average task delay (ms)	1.98	2.10	2.15	2.00	1.50
Proposed IEMDP-MHRL		Percentage of completion task	92	81	87	78	75
		Packet loss	87	231	376	694	872
		Average task wait (ms)	0.76	1.36	1.36	0.82	0.54
		Average task delay (ms)	0.34	1.23	1.25	1.43	1.44

4.3 Discussion

The advantages of the proposed IEMDP-MHRL method and the disadvantages of existing methods are presented in this section. The existing method limitations like MA-DDQN [20] struggled in scaling with a large number of agents because of the exponential increase in state-action space. The two-layer optimization [21] suffer from high energy consumption due to the complexity of addressing both layers simultaneously particularly in dynamic environments with limited resources. PPO [23] cause high average task delay because of inefficient exploration in highly dynamic environments which leads to suboptimal performance. The proposed IEMDP-MHRL overcomes these existing method limitations by incorporating evidential reasoning and meat-learning strategies to handle uncertainty. Its hierarchical structure increases scalability and allows for better coordination between different vehicles and tasks. Also, using a meta-learning approach changes vehicular conditions and maintains stability as well as decision-making progress. This analysis not only minimizes delays but also reduces energy consumption across different scenarios.

5. Conclusion

This research proposed IEMDP-MHRL for resource allocation in vehicular networks. By effectively adding uncertainty via evidential reasoning, the proposed method enhances decision-making in dynamic environments. The HRL enables effective and scalable resource management which adapts to varying demands over various network levels. Moreover, the meta-learning process provides an information transfer across tasks which enhances learning efficiency. It considers both long-term and short-term rewards which results in a more effective and sustainable process. From the overall analysis, the proposed IEMDP-MHRL makes a more robust and enhanced resource allocation mechanism in vehicular networks. When compared to existing methods like IRATS, the proposed IEMDP-MHRL achieves a lesser average task delay of 0.34 ms for vehicle 10 respectively. In the future, the advanced technique like transformer model will be considered in RL for further increasing the model performance in resource allocation.

Notation Description

Symbols	Description
b	Base Station (BS)
$\emptyset_i^b$	Binary variable for state i with BS b
$\Gamma_i(t)$	Probability of state transition for state i at time t
$\emptyset_i^b(t)$	Binary variable at time t for state i with BS b
Pr	Probability
$Y_{z1,z2}(t)$	Probability of transitioning from cache state z_1 to z_2 at time t
Tr	Vector at vehicle r
o_r	Content size for vehicle r
q_r	CPU cycles to process request r
$M_e^b(t)$	Computing ability at time t for an entity e with parameter b
$\Omega_{q_1,q_2}(t)$	Pair of states q_1 to q_2 at time t
$N_e^b(t)$	Computing power of state transition probability at time t for an entity e with parameter b
qr	Task's computation execution for request r
$U_r^b(t)$	Multi-user server for vehicle r with BS b at time t
$R_{r,e}^{comp}(t)$	Computing resource for request r and entity e at time t
$\frac{o_r}{t_e^b}$	Ratio of content size request r to the time required to process the request at entity e with parameter b
Mx_e	Maximum amount of requests that is established simultaneously multi-user server
k_r^b	Random variable among BS b and vehicle r
$\Upsilon_{d_1,d_2}(t).$	Changing from a state d_1 to d_2 at time t
$h \times h$ matrix $G_r^b(t)$	State transition probability among BS b and vehicle r

G_r^b	Bandwidth which is allocated to vehicle r for available bandwidth of BS b
$H_\xi^b(t)$	Content saved in cache ξ related with BS
Ξ	Content cache
$\pi_\varphi(.\|S_t)$	Master policy at present state S_t
k	Sub-policies index
φ	Master policy is reorganized by gradient method
μ	Learning rate
r_H	Gain of master policy from sub-policy sub_k at S_t
$\pi_\omega sub_k(.\|S_t)$	Sub-policy network that evaluates a BS to request a vehicle in S_t
b_j	Chosen BS
H_{IV}^D and H_{IV}^{CD}	Amount of information in Information Volume (IV) in C-D condition
η	New uncertainty par
AU	Interaction between input signal A and distortion or noise term U
UU	Self-interaction of a distortion and noise term U
UG	Interaction among distortion U and variable G
AB	Interaction between input signal A and Bias B
UB	Distortion or noise term U and Bias B
D'_{is}	Disjunction effect is reproduced by the probability difference in C-D condition and D-condition
n	Number of states
m^n	Measure m at n^{th}iteration

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, have been done by 1st author. The supervision and project administration, have been done by 2nd author.

References

[1] K. Tan, D. Bremner, J.L. Kernec, Y. Sambo, L. Zhang, and M.A. Imran, "Intelligent handover algorithm for vehicle-to-network communications with double-deep Q-learning", *IEEE Transactions on Vehicular Technology*, Vol. 71, No. 7, pp. 7848-7862, 2022.

[2] Hemavathi, S.R. Akhila, Y. Alotaibi, O.I. Khalaf, and S. Alghamdi, "Authentication and resource allocation strategies during handoff for 5G IoVs using deep learning", *Energies*, Vol. 15, No. 6, p. 2006, 2022.

[3] X. Li, L. Lu, W. Ni, A. Jamalipour, D. Zhang, and H. Du, "Federated multi-agent deep reinforcement learning for resource allocation of vehicle-to-vehicle communications", *IEEE Transactions on Vehicular Technology*, Vol. 71, No. 8, pp. 8810-8824, 2022.

[4] Y. Zhi, J. Tian, X. Deng, J. Qiao, and D. Lu, "Deep reinforcement learning-based resource allocation for D2D communications in heterogeneous cellular networks", *Digital Communications and Networks*, Vol. 8, No. 5, pp. 834-842, 2022.

[5] Y. Fu, and X. Wang, "Traffic prediction-enabled energy-efficient dynamic computing resource allocation in cran based on deep learning", *IEEE Open Journal of the Communications Society*, Vol. 3, pp. 159-175, 2022.

[6] K. Tan, L. Feng, G. Dán, and M. Törngren, "Decentralized convex optimization for joint task offloading and resource allocation of vehicular edge computing systems", *IEEE Transactions on Vehicular Technology*, Vol. 71, No. 12, pp. 13226-13241, 2022.

[7] G. Sun, L. Sheng, L. Luo, and H. Yu, "Game theoretic approach for multipriority data transmission in 5G vehicular networks", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 23, No. 12, pp.24672-24685, 2022.

[8] B. Hazarika, K. Singh, S. Biswas, and C.P. Li, "DRL-based resource allocation for computation offloading in IoV networks", *IEEE Transactions on Industrial Informatics*, Vol. 18, No. 11, pp. 8027-8038, 2022.

[9] A.K. Singh, R. Pamula, P.K. Jain, and G. Srivastava, "An efficient vehicular-relay selection scheme for vehicular communication", *Soft Computing*, Vol. 27, No. 6, pp. 3443-3459, 2023.

[10] L. Liu, J. Feng, X. Mu, Q. Pei, D. Lan, and M. Xiao, "Asynchronous deep reinforcement learning for collaborative task computing and on-demand resource allocation in vehicular edge computing", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 24, No. 12, pp. 15513-15526, 2023.

[11] W. Feng, S. Lin, N. Zhang, G. Wang, B. Ai, and L. Cai, "Joint C-V2X based offloading and resource allocation in multi-tier vehicular edge computing system", *IEEE Journal on Selected Areas in Communications*, Vol. 41, No. 2, pp. 432-445, 2023.

[12] Y. He, Y. Wang, Q. Lin, and J. Li, "Meta-hierarchical reinforcement learning (MHRL)-based dynamic resource allocation for dynamic vehicular networks", *IEEE Transactions on Vehicular Technology*, Vol. 71, No. 4, pp. 3495-3506, 2022.

[13] Q.L. Dang, W. Xu, and Y.F. Yuan, "A dynamic resource allocation strategy with reinforcement learning for multimodal multi-objective optimization", *Machine Intelligence Research*, Vol. 19, No. 2, pp. 138-152, 2022.

[14] F.N. Al-Wesabi, M. Obayya, M.A. Hamza, J.S. Alzahrani, D. Gupta, and S. Kumar, "Energy aware resource optimization using unified metaheuristic optimization algorithm allocation for cloud computing environment", *Sustainable Computing: Informatics and Systems*, Vol. 35, p. 100686, 2022.

[15] W. Qi, Q. Song, L. Guo, and A. Jamalipour, "Energy-efficient resource allocation for UAV-assisted vehicular networks with spectrum sharing", *IEEE Transactions on Vehicular Technology*, Vol. 71, No. 7, pp. 7691-7702, 2022.

[16] W. Fan, Y. Su, J. Liu, S. Li, W. Huang, F. Wu, and Y.A. Liu, "Joint task offloading and resource allocation for vehicular edge computing based on V2I and V2V modes", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 24, No. 4, pp. 4277-4292, 2023.

[17] Y. Cong, K. Xue, C. Wang, W. Sun, S. Sun, and F. Hu, "Latency-energy joint optimization for task offloading and resource allocation in mec-assisted vehicular networks", *IEEE Transactions on Vehicular Technology*, Vol. 72, No. 12, pp. 16369-16381, 2023.

[18] F.M. Talaat, "Effective deep Q-networks (EDQN) strategy for resource allocation based on optimized reinforcement learning algorithm", *Multimedia Tools and Applications*, Vol. 81, No. 28, pp. 39945-39961, 2022.

[19] H. Zhang, C. Lu, H. Tang, X. Wei, L. Liang, L. Cheng, W. Ding, and Z. Han, "Mean-field-aided multiagent reinforcement learning for resource allocation in vehicular networks", *IEEE Internet of Things Journal*, Vol. 10, No. 3, pp. 2667-2679, 2023.

[20] 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.

[21] J. Gao, Z. Kuang, J. Gao, and L. Zhao, "Joint offloading scheduling and resource allocation in vehicular edge computing: A two layer solution", *IEEE Transactions on Vehicular Technology*, Vol. 72, No. 3, pp. 3999-4009, 2023.

[22] W. Shu, and Y. Li, "Joint offloading strategy based on quantum particle swarm optimization for MEC-enabled vehicular networks", *Digital Communications and Networks*, Vol. 9, No. 1, pp. 56-66, 2023.

[23] B. Jamil, H. Ijaz, M. Shojafar, and K. Munir, "IRATS: A DRL-based intelligent priority and deadline-aware online resource allocation and task scheduling algorithm in a vehicular fog network", *Ad Hoc Networks*, Vol. 141, p. 103090, 2023.

[24] H. Fu, J. Wang, J. Chen, P. Ren, Z. Zhang, and G. Zhao, "Dense Multi-Agent Reinforcement Learning Aided Multi-UAV Information Coverage for Vehicular Networks", *IEEE Internet of Things Journal*, Vol. 11, No. 12, pp. 21274 – 21286, 2024.

[25] A. Alam, P. Shah, R. Trestian, K. Ali, and G. Mapp, "Energy Efficiency Optimisation of Joint Computational Task Offloading and Resource Allocation Using Particle Swarm Optimisation Approach in Vehicular Edge Networks", *Sensors*, Vol. 24, No. 10, p.3001, 2024.

[26] H. Alavizadeh, H. Alavizadeh, and J. Jang-Jaccard, "Deep Q-learning based reinforcement learning approach for network intrusion detection", *Computers*, Vol. 11, No. 3, p. 41, 2022.

[27] A. Asadi, S.N. Pinkley, and M. Mes, "A Markov decision process approach for managing medical drone deliveries", *Expert Systems with Applications*, Vol. 204, p. 117490, 2022.