A Survey on Multi-Agent Reinforcement Learning Applications in the Internet of Vehicles

Elham Mohammadzadeh Mianji*, Mohammad Fardad*, Gabriel-Miro Muntean†, Irina Tal*
*School of Computing, †School of Electronic Engineering
Dublin City University
Dublin, Ireland
{elham.mohammadzadehmianji3, mohammad.fardad2}@mail.dcu.ie
{gabriel.muntean,irina.tal}@dcu.ie

2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring) | 979-8-3503-8741-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/VTC2024-SPRING62846.2024.10683524

***Abstract*—The development of the Internet of Vehicles (IoV) and autonomous vehicles plays a significant role in intelligent transportation systems (ITS) that are empowered by vehicular networks. However, the dynamic nature of these networks presents challenges that need to be addressed. Reinforcement learning (RL) has emerged as an effective technique for strengthening vehicular networks. The use of standard single-agent RL and deep reinforcement learning (DRL) has recently been demonstrated to enable each network entity as a decision-making agent to adapt to unknown environments by learning an optimal decision-making policy. However, in the complex and dynamic environments of vehicular networks, the limitations of single-agent approaches become apparent. Multi-agent reinforcement learning (MARL) offers a compelling alternative, enabling network entities to learn their optimal policies by observing the environment as well as the policies of other network entities. Due to this, MARL has recently been used to solve various problems in IoV by improving its learning efficiency. In this paper, we review the applications of MARL in IoV networks. Following the review, four main application areas for MARL in IoV were identified, namely: resource management, task offloading, trust management, and privacy preservation. Furthermore, the MARL-based approaches in IoV were classified into three main categories: fully centralized, fully decentralized, and centralized training with decentralized execution (CTDE) depending on the MARL architecture employed. Finally, we discuss the challenges, open issues, and future directions related to the applications of MARL in the IoV.**

***Index Terms*—Multi-agent reinforcement learning, Internet of vehicles, Resource allocation, Task offloading, trust management, Privacy-preserving**

I. Introduction

Vehicular networks, also known as vehicular ad-hoc networks (VANETs), are self-organizing networks of vehicles that communicate with each other and with fixed infrastructure networks to exchange information. The broader concept of the Internet of Vehicles (IoV) encompasses not only vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications, but also all types of communication between vehicles and other devices, including V2V, V2I, vehicle-to-pedestrian (V2P), and vehicle-to-network (V2N), as illustrated in Fig. 1 [1]–[3]. This collective communication between vehicles and various entities is referred to as vehicle-to-everything (V2X) communication. IoV is a rapidly developing technology with the potential to revolutionize transportation by enhancing road safety, reducing traffic congestion, and optimizing efficiency. This will redefine the automotive industry, paving the way for safer and more efficient autonomous vehicles.

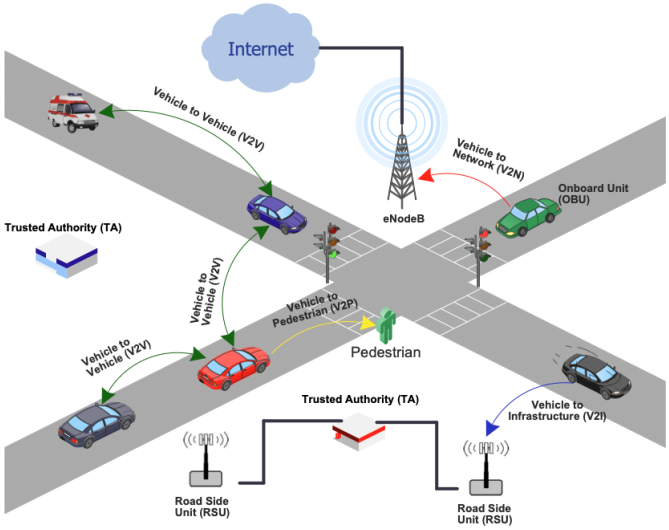

Fig. 1. IoV Network

Machine learning (ML) is increasingly important in dealing with multifaceted challenges, from those in the medical field [4], [5] to issues related to the IoV. Among the diverse ML paradigms, reinforcement learning (RL) has emerged as a particularly potent approach, offering a unique ability to address sequential decision-making problems within intricate networks [6]. RL empowers devices to continually interact with their environments, tuning their decision-making strategies through iterative learning. These strategies are often formalized using models such as Markov decision processes (MDP) in specific scenarios [7]. RL proves to be a versatile tool, demonstrating efficacy in tackling an array of challenges, from resource allocation and offloading to caching and, significantly, security and privacy concerns. The dynamic nature of vehicular environments necessitates intelligent and adaptive decision-making, making RL a suitable candidate for optimizing resource allocation strategies. Additionally, the use of multi-agent RL (MARL) due to the collaborative nature of vehicular networks is particularly well-suited to addressing challenges that extend

979-8-3503-8741-4/24/$31.00 ©2024 IEEE

beyond individual vehicles. MARL enables more efficient resource allocation, improved task offloading strategies, and enhanced privacy-preserving and trust management through coordinated decision-making. In this manner, IoV challenges can be addressed comprehensively, promoting cooperative interactions between vehicles and infrastructure.

This survey provides a comprehensive categorization of MARL solutions into trust management, privacy-preserving, task offloading, and resource allocation domains and provides a systematic framework for understanding how MARL techniques contribute to addressing key challenges and objectives in vehicular networks. To the best of our knowledge, this is the first survey that focuses primarily on providing a detailed discussion of various MARL features and classifying available MARL-based proposed approaches in IoV.

The remaining parts of the paper are organized as follows: Section II presents related works and our scope while Section III highlights the designing of RL and MARL architectures. Analyzing and comparing proposed approaches based on MARL in IoV has been discussed in Section IV. Open challenges and future directions are presented in Section V, while Section VI concludes the paper.

II. Related Surveys and Our Scope

There has been increasing interest in utilizing ML to solve various IoV challenges in the past few years, with a focus on vehicular network challenges specifically. A variety of surveys have been conducted on vehicular networks that address different issues and discuss various solutions. In [8], the concept of trust is critically analyzed for existing machine learning-based approaches in mobile ad-hoc networks. As well, it compares the different distributed trust computing mechanisms that can be divided into three categories, direct trust, indirect trust, and hybrid trust. [9] focused on ML techniques, but also highlights other methods for addressing the various challenges in VANETs. Additionally, [10] includes a discussion of proposed artificial intelligence (AI) techniques such as ML, deep learning, and swarm intelligence in security, resource allocation, and mobility management areas. [11] described the state-of-the-art methods used to allocate and utilize the available wireless network resources in an ultra-efficient manner. It presented a comprehensive survey on resource allocation schemes for the two dominant vehicular network technologies, e.g. dedicated short range communications (DSRC) and cellular based vehicular networks. Also, it discussed the challenges and opportunities for resource allocation in modern vehicular networks based on network slicing, ML, and context awareness. In another work, [12] focused on vehicular networks and presented a comprehensive review of research that integrates RL and deep reinforcement learning (DRL) algorithms for vehicular networks management, with an emphasis on vehicular telecommunications. Two types of schemes are categorized in this survey, i.e. vehicular resource management and vehicular infrastructure management. Security in vehicular networks is not covered in this study. In [13], a comprehensive examination of AI and ML solutions is presented, alongside cryptography-based techniques, specifically within the domain of V2X communication. This study prioritizes data-driven solutions tailored to mitigate security, privacy, and trust concerns, paying special attention to the emerging threat vectors that these innovations introduce. Similarly, another survey, [14], provided an extensive overview of VANETs and explored trust management core concepts within this context. It highlighted the use of AI-based approaches to trust management in VANETs, including clustering, RL, fuzzy logic, and game theory techniques.

While numerous surveys have examined the various ML and AI algorithms that can be applied to the IoV, this survey takes a different approach by focusing specifically on MARL methods. Unlike traditional single-agent RL approaches, which train a single agent to interact with its environment to maximize a reward signal, MARL is better suited to address the complex challenges that arise in dynamic networks such as IoV. This is because MARL enables multiple agents to collaborate and coordinate their actions to achieve a common goal, leading to improved performance and efficiency in complex environments. By examining various MARL-based solutions designed to address resource allocation, task offloading, trust management, and privacy preservation challenges in IoV environments, this survey provides a comprehensive review of the state-of-the-art in this emerging field. The survey highlights the potential advantages and limitations of MARL-based approaches and identifies future research directions, contributing to the advancement of IoV research and development.

III. RL and MARL

A. Designing RL Systems

RL is a type of ML that allows agents to learn from their experiences and improve their performance over time by receiving rewards from the environment. Training agents to make the best decisions based on feedback is the focus of RL. Each action by an agent is rewarded or penalized based on its impact on the environment. An agent learns to take actions that maximize rewards and minimize penalties [6]. The problem of RL can be expressed as an MDP. MDPs provide a mathematical structure for capturing the complexities of decision-making problems where agents interact with their surroundings over time. The decision-making process of the agent is referred to as the policy. Upon being presented with a state, a policy generates either a specific action or a probability distribution covering multiple actions. By observing the environment, a policy provides guidance on the action the agent should take (or multiple probabilities for each action). An MDP is defined by the tuple $(\mathcal{S}, \mathcal{A}, P, r, \gamma)$, where:

- $\mathcal{S}$ represents the state space, encompassing discrete or continuous states $s \in \mathcal{S}$.
- $\mathcal{A}$ denotes the action space, comprising discrete or continuous actions $a \in \mathcal{A}$ available to the agent.
- P signifies the transition probability matrix. At time step t, it governs the probability of transitioning from state s_t to state s_{t+1}. Mathematically, $P_{ij} = p(s_{t+1} = j | s_t = i)$, where i and j are state indices.

- $r : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ is the reward function $r(s_t, a_t)$, providing the agent with a reward when transitioning from state s_t to state s_{t+1} through action a_t.
- γ is the discount factor, constrained within $[0, 1]$, influencing the agent's consideration of future rewards. $\gamma = 0$ implies sole emphasis on immediate rewards, while $\gamma = 1$ indicates equal importance assigned to all future rewards.

The discounted return, R_t, is the cumulative sum of all future rewards, discounted by γ. Essentially, it represents the expected value of rewards the agent anticipates after executing a specific action in a given state. the agent's objective is to determine an optimal policy, maximizing the expected discounted return, i.e., $\max \mathbb{E}[R_t]$. According to Fig. 2, three primary types of RL methods can be used to achieve this optimal policy and thus resolve the RL problem.

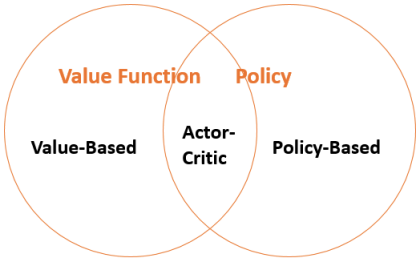

Fig. 2. Value-Based vs Policy-Based vs Actor-Critic

1) Policy-based methods: These methods involve directly training the policy to discern the appropriate action corresponding to a given state.

2) Value-based methods: In contrast, value-based methods focus on training a value function, enabling the agent to learn the relative value of different states. Subsequently, this value function is utilized to determine the action that leads to the most valuable state.

3) Actor-critic methods: Actor-critic algorithms use two neural networks, an actor and a critic, to learn both the policy and the value function.

B. Designing MARL Systems

MARL is essential for modeling scalable and self-organizing systems consisting of connected agents. It addresses the limitations of single-agent RL by enabling agents to learn effective policies while considering the actions and observations of other agents. MARL has found significant success in various applications, such as autonomous driving vehicles [15]. To design a MARL system, there are three main architectures: fully decentralized learning, fully centralized and centralized training with decentralized execution (CTDE).

1) Fully Decentralized Architecture: In this approach, each agent is trained independently from the others as shown in Fig. 3, hence, the agent can only use information gathered from its local observations to make action decisions. This method has the benefit of allowing agents to be designed and trained like single agents, considering other agents as part of the environment dynamics. Agents maintain their own critics (value functions) and policies, which are updated based on their own experiences. However, this technique can make the environment non-stationary, as the underlying MDP changes over time due to the interactions of other agents. This can be problematic for many RL algorithms that cannot reach a global optimum in non-stationary environments [16].

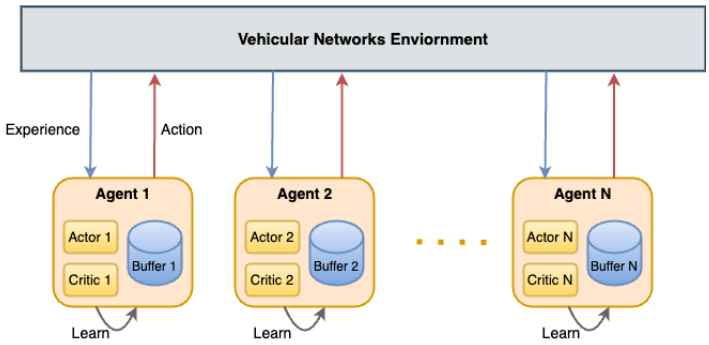

Fig. 3. Fully Decentralized MARL Architecture

2) Fully Centralized Architecture: In this approach, a high-level process collects agents' experiences, which are used to learn a common policy. In centralized training which is shown in Fig. 4, agents share the collected experiences and learn from them together. This method allows learning from all agents, taking into account the present state of the environment and the policy outputs joint actions. The reward is global, as it considers the actions and states of all agents. [17]–[19].

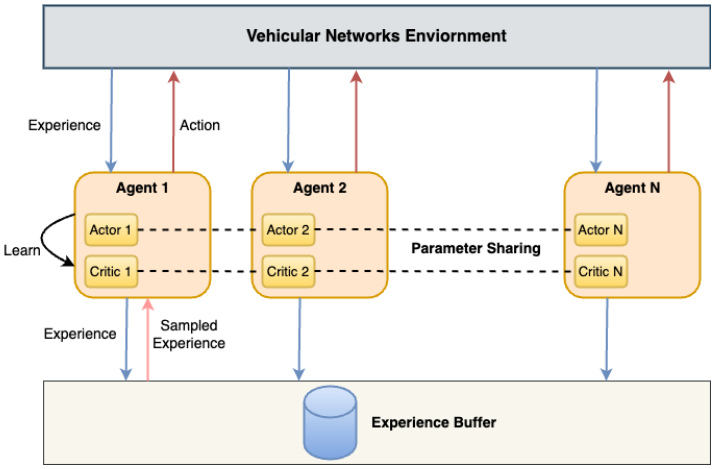

Fig. 4. Fully Centralized MARL Architecture

3) Centralized training with decentralized execution: CTDE refers to training agents in a centralized manner, where they share a common critic and policy, but during execution or deployment, agents act independently without communication or coordination. CTDE enables the agent to learn a decentralized policy through global optimization, as it trains a unique policy for each agent through learning the experiences of all the agents. Several MARL algorithms use the CTDE framework, including multi-agent deep deterministic policy gradient (MADDPG) where each agent has a decentralized

actor and a centralized critic, and counterfactual multi-agent policy gradients (COMA). However, unlike the MADDPG, the COMA estimates the action-value function through a single centralized critic [15].

IV. MARL-based Approaches in IoV

This section systematically reviews the existing literature on MARL-based solutions in IoV, categorizing them based on the problems they addressed: privacy preservation, trust management, task offloading, and resource allocation. Each category is subsequently evaluated, emphasizing the unique characteristics of individual approaches. This classification attempts to offer a well-organized overview of MARL-based solutions in IoV. A comparison of different approaches and reviewed works are summarized in Table I. As seen from the table, most of the works adopt the CTDE framework.

A. MARL-based Resource Allocation

Effective resource management is essential for ensuring the efficient operation of IoV networks. MARL can be used to develop resource management algorithms that allocate network resources, such as bandwidth and power, among vehicles in a fair and equitable manner. By dynamically adjusting resource allocation based on real-time demands, MARL-based solutions can optimize network performance and prevent resource contention. Resource allocation using MARL has been widely employed in various applications of the IoV; For instance, Kumar *et al.* [20] proposed a mobility-aware channel allocation scheme for V2X networks that maximizes spectrum efficiency by adapting to the dynamic nature of vehicular environments. They used MARL to predict the mobility patterns of vehicles in order to allocate resources in a proactive manner based on anticipated movements and channel conditions. Fu *et al.* [21] proposed a digital twin (DT)-assisted resource allocation framework for vehicle platooning applications that jointly optimizes sub-band and power allocation to meet quality-of-service (QoS) requirements. To address the challenges of cooperation aging in their proposed framework, they utilized a tactic-interactive MARL method. Ji *et al.* [22] proposed a MARL based approach for resource allocation in V2X communication to maximize the sum rate of V2I link occupancy while ensuring the maximum transmission rate of cellular links and meeting the reliability and delay requirements of V2V communication. The approach allocates resources among vehicles and infrastructure nodes to prevent eavesdropping and ensure secure communication. Dai *et al.* [23] proposed a MARL solution to optimize the allocation of available resources for live video streaming (LVS) services in the IoV. Their approach involves grouping vehicles and assigning the required video quality, and then using MARL techniques to allocate the required resources. Cui *et al.* proposed a dynamic radio access network (RAN) slicing framework in [24] to minimize the long-term system cost in a multi-access edge computing (MEC) assisted heterogeneous vehicular network. They transformed the resource allocation problem into a partially observable Markov decision process and then presented a MARL solution to cooperatively allocate spectrum and computing resources by considering base stations (BSs) as cooperative agents.

B. MARL-based Task Offloading

Traditionally, cloud computing has been utilized to handle computationally intensive tasks in mobile networks. Nevertheless, cloud computing yields high response times, which are unsuitable for dynamic and latency-critical environments like vehicular networks [25]. To address these issues, MEC has emerged as a promising solution [26]–[28], particularly in the context of vehicular networks, where it is referred to as vehicular edge computing (VEC) [29], [30]. VEC effectively deploys several computing and storage resources in close proximity to vehicles, capitalizing on the presence of road side units (RSUs) as a viable solution. VEC servers are deployed along roads in vehicular networks, and vehicles can offload computation tasks to them using wireless communication. By offloading computations, the vehicles are able to reduce the task delay and improve the quality of service. Single-agent deep RL (DRL) struggles to handle the intricate dynamics of multi-agent vehicular scenarios, particularly when addressing reliability and latency concerns. Recent research has shifted towards MARL approaches to enhance communication efficiency and reduce latency. In this regard, a multi-agent DRL based computation offloading scheme for vehicular edge networks has been proposed in [31]. This scheme determines the most optimal offloading decisions to nearby MEC servers to minimize the total processing delay for multiple vehicles. In this approach, each vehicle acts as an agent, and the system state includes vehicle location and task information. The action involves selecting an MEC server for offloading and the reward function is based on task delay. In a similar work [32] examined the problem of computation offloading in an IoV environment to guarantee latency, energy consumption, and cost requirements for offloading. A multi-agent DRL based Hungarian algorithm is proposed to solve the dynamic task offloading problem formulated as a stochastic game. This enables cooperative decentralized offloading decisions. In terms of task offloading, [33] combined DT technology which brings data-driven representations of the physical world to mirrored virtual space and MARL into the design of the automotive edge computing network and propose a coordination graph-driven vehicle task offloading scheme to minimize offloading costs. Hazarika *et al.* [34] Introduced a multi-agent (MA-DRL) based task offloading strategy for an IoV network aided by multiple re-configurable intelligent surfaces. Tasks are classified based on priority (high, common, low), size (small, medium, large) and network criticality. Then the MA-DRL algorithm modeled as a Markov game aims to maximize the utility of the network by optimizing the task offloading decisions strategy.

C. MARL-based Trust Management

Establishing and maintaining trust among vehicles is essential for ensuring reliable and secure communication in IoV. MARL

can be used to develop trust management systems that evaluate the trustworthiness of vehicles based on their behavior and interactions. By dynamically assessing trust levels, MARL-based solutions can prevent malicious vehicles from gaining access to critical network resources or disrupting communication. The absence of centralized infrastructure in IoV necessitates decentralized trust management schemes for trusted routing in vehicular networks. One such approach, proposed by [35], employs fuzzy logic to assess the trustworthiness of one-hop neighbors based on factors like cooperativeness, honesty, and responsibility. Additionally, for estimating trust between non-neighbors, a Q-learning method is introduced by synthesizing reports from multiple nodes. By avoiding malicious nodes in routing decisions, this trust management scheme enhances network performance, increasing throughput and packet delivery ratio. Trusted routing in vehicular networks can also be achieved by avoiding malicious vehicles. In this regard, [36] developed a secure routing approach to defend against packet drop attacks in VANETs. This paper introduces a novel grid-based extended joint action learning approach, Grid-eJAL, that employs online and adaptive MARL to implement route mutation for attack mitigation in VANETs. Grid-eJAL divides the area of interest into grids, and during packet transmission, the optimal next hop is determined by the learning policy's minimum angle of mobility. Moreover, in our previous work [37], a method for detecting and preventing malicious vehicles in VANETs was proposed using a trusted routing technique based on Q-learning (QL-TRT). The trust level and reliability of links between pairs of vehicles are evaluated by considering factors such as packet forwarding ratios, energy consumption, and expected transmission times. To select the most trustworthy route from the source to the destination, Q-learning is used to learn the trust value of links. Authors in [38] proposed a distributed trust sharing mechanism for IoV. This approach utilizes Echo State Networks, which simplify the training of recurrent neural networks, and DQNs to achieve trust information sharing and predict vehicle routes at intersections based on observed traffic patterns. Current trust assessments and traffic conditions are incorporated into the state space to improve safety. Furthermore, [39] presented a trust model to prevent bogus information from reaching decision-making entities. In this model, a vehicle requests trust values from the trust evaluation model regarding a driving decision-making event. A RL algorithm is used to determine the optimal strategy for trust schemes based on feedback from historical evaluation results.

D. MARL-based Privacy Preservation

Protecting user privacy in IoV is crucial, as sensitive data such as location information and identity of vehicles or other vehicle's data can be easily compromised. MARL can be leveraged to develop privacy-preserving mechanisms that allow vehicles to share information while maintaining anonymity. Regarding privacy protection, in [40], [41], a methodology is delineated for the concurrent enhancement of throughput from RSUs to vehicles, while concurrently preserving vehicle privacy through the provision of obscured location data to the server. The authors introduced a strategy involving furnishing the server with obfuscated location details. To address this issue, they leveraged the capabilities of Q-learning, where a state-action matrix was trained. This matrix played a crucial role in guiding the decision-making process concerning the formulation of location obfuscation strategies. These strategies were intricately tied to both the present and past reported locations of the vehicles. Wei *et al.* in [47] proposed a privacy-aware MADRL approach for task offloading over VANET. A vulnerability in MADRL's policy learning process was exploited to cause vehicles to offload tasks to malicious RSUs through an offloading preference inference attack. The authors proposed a joint optimization of offloading action and transmission power to minimize system costs, including local and edge costs, while protecting privacy preferences. A related work in edge content cache (ECC) [46] proposed a joint optimization problem involving real-time cache replacement strategies and cache prefetching strategies, aimed to enhance ECC efficiency, reduce overall communication costs, and safeguard user privacy.

Other papers have utilized multi-agent federated RL (MAFRL) to allocate resources in vehicular networks while maintaining privacy. For instance, [42] developed a privacy-preserving, distributed MAFRL approach for cooperative content caching in vehicular edge networks. The DQN-based agents, RSUs, learn to make caching decisions that improve cache efficiency and meet content delay requirements. However, [43] introduced a MAFRL framework for joint edge association and power allocation in vehicular networks, which achieves a good connectivity-cost trade-off while preserving privacy. Using this approach, the agents are trained centrally at a macro BS by sharing encrypted Q-values, but then execute policies in a distributed manner based on local observations. Consequently, it is a centralized training, decentralized execution approach. This problem is similarly formulated in [44] by introducing an actor-critic algorithm for device selection, unmanned aerial vehicle (UAV) placement, and resource management to enhance federated convergence performance. Similarly, [45] proposed actor-critic RL algorithms that incorporate federated learning (FL) for trajectory optimization and resource allocation in networks with multiple UAVs serving as aerial BSs. Specifically, the authors develop FL actor-critic and distillation FL actor-critic methods where each UAV agent learns decentralized policies but coordinates with other UAVs periodically to reach an overall optimal solution. The key innovation is to use FL for privacy-preserving collaboration between UAVs, while actor-critic methods handle the continuous action spaces.

V. Open challenges and Future Directions

A. Real-World Deployment and Evaluation:

- Challenge: Despite significant progress in MARL research for IoV, the real-world deployment and evaluation of MARL-based solutions remain limited. Validating the effectiveness and scalability of MARL algorithms in real-world settings poses practical challenges.

TABLE I
COMPARISON OF DIFFERENT APPROACHES IN MARL FOR IOV

MARL Framework	Classification	Applications of MARL in IoV			
		Resource Allocation	Task Offloading	Trust Management	Privacy Preserving
Centralized MARL	Value-Based	-	-	-	-
	Policy-Based	-	-	-	-
	Actor-Critic	[20]	-	-	-
Decentralized MARL	Value-Based	[22]	-	[35]–[39]	[40], [41]
	Policy-Based	-	-	-	-
	Actor-Critic	-	[33]	-	-
CTDE	Value-Based	[23]	[32]	-	[42]–[45]
	Policy-Based	-	-	-	-
	Actor-Critic	[21], [24]	[31], [34]	[46]	[44], [45], [47]

- Future Directions: Future research should focus on conducting large-scale field trials and simulations to evaluate the performance of MARL-based approaches in real-world vehicular environments. This includes collaborating with industry partners and regulatory bodies to address practical deployment challenges and ensure the safety and reliability of MARL-enabled vehicular networks.

B. Multi-Objective Optimization:

- Challenge: An effectively crafted reward function holds the potential to greatly enhance the performance of an agent and expedite the learning process. However, in future networks comprising various network entities such as BSs and users, each with distinct objectives such as maximizing throughput or energy efficiency, devising an optimal reward design remains an unsolved challenge.
- Future Directions: Future research should focus on developing multi-objective optimization techniques tailored to MARL in IoV. This includes exploring algorithms such as evolutionary algorithms, and multi-objective RL to efficiently trade off between competing objectives. Additionally, the reviewed works typically set a global reward (i.e., a common reward) for all the agents or a local reward for an individual agent. Furthermore, the conflicting objectives challenge the reward design, which is common in network transmission objectives such as maximizing the network throughput while minimizing the interference to each other. Hence, it is critical to design an appropriate reward to optimize the trade-off between different objectives.

VI. CONCLUSION

The purpose of this paper is to provide an in-depth examination of the applications of MARL in the IoV. In the first section, we have introduced the concept of IoV and vehicular networks, a subset of IoV, followed by a description of the background of RL and MARL. Next, we have reviewed, analyzed, and compared recent research contributions that utilize MARL to address emerging issues in IoV in general and vehicular networks in particular. These emerging issues involve resource allocation, task offloading, task offloading, and privacy protection. Finally, we have outlined the key open challenges and future directions that can assist readers in further exploring these topics.

ACKNOWLEDGMENT

This publication has emanated from research supported in part by a grant from Science Foundation Ireland under grant number 18/CRT/6183. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any author accepted manuscript version arising from this submission. G.-M. Muntean acknowledges the Science Foundation Ireland's support via grant 12/RC/2289_P2 (INSIGHT).

REFERENCES

[1] M. Fardad, E. M. Mianji, G.-M. Muntean, and I. Tal, "A fast and effective graph-based resource allocation and power control scheme in vehicular network slicing," in *2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2022, pp. 1–6.

[2] M. Fardad, G.-M. Muntean, and I. Tal, "Latency-aware v2x operation mode coordination in vehicular network slicing," in *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*, 2023, pp. 1–6.

[3] A. Bazzi, A. O. Berthet, C. Campolo, B. M. Masini, A. Molinaro, and A. Zanella, "On the design of sidelink for cellular V2X: A literature review and outlook for future," *IEEE Access*, vol. 9, pp. 97 953–97 980, 2021.

[4] R. Ranjbarzadeh, P. Zarbakhsh, A. Caputo, E. B. Tirkolaee, and M. Bendechache, "Brain tumor segmentation based on optimized convolutional neural network and improved chimp optimization algorithm," *Computers in Biology and Medicine*, vol. 168, p. 107723, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482523011885

[5] R. Ranjbarzadeh, N. Tataei Sarshar, S. Jafarzadeh Ghoushchi, M. Saleh Esfahani, M. Parhizkar, Y. Pourasad, S. Anari, and M. Bendechache, "MRFE-CNN: multi-route feature extraction model for breast tumor segmentation in mammograms using a convolutional neural network," *Annals of Operations Research*, vol. 328, no. 1, pp. 1021–1042, 2023.

[6] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[7] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[8] G. Jinarajadasa and S. Liyange, "A survey on applying machine learning to enhance trust in mobile adhoc networks," in *2020 International Research Conference on Smart Computing and Systems Engineering (SCSE)*. IEEE, 2020, pp. 195–201.

[9] N. H. Hussein, C. T. Yaw, S. P. Koh, S. K. Tiong, and K. H. Chong, "A comprehensive survey on vehicular networking: Communications, applications, challenges, and upcoming research directions," *IEEE Access*, vol. 10, pp. 86 127–86 180, 2022.

[10] A. Mchergui, T. Moulahi, and S. Zeadally, "Survey on artificial intelligence (ai) techniques for vehicular ad-hoc networks (vanets)," *Vehicular Communications*, vol. 34, p. 100403, 2022.

[11] M. Noor-A-Rahim, Z. Liu, H. Lee, G. M. N. Ali, D. Pesch, and P. Xiao, "A survey on resource allocation in vehicular networks," *IEEE transactions on intelligent transportation systems*, 2020.

[12] A. Mekrache, A. Bradai, E. Moulay, and S. Dawaliby, "Deep reinforcement learning techniques for vehicular networks: recent advances and future trends towards 6g," *Vehicular Communications*, p. 100398, 2021.

[13] R. Sedar, C. Kalalas, F. Vázquez-Gallego, L. Alonso, and J. Alonso-Zarate, "A comprehensive survey of v2x cybersecurity mechanisms and future research paths," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 325–391, 2023.

[14] H. Amari, Z. A. E. Houda, L. Khoukhi, and L. H. Belguith, "Trust management in vehicular ad-hoc networks: Extensive survey," *IEEE Access*, vol. 11, pp. 47 659–47 680, 2023.

[15] T. Li, K. Zhu, N. C. Luong, D. Niyato, Q. Wu, Y. Zhang, and B. Chen, "Applications of multi-agent reinforcement learning in future internet: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1240–1279, 2022.

[16] K. Zhang, Z. Yang, H. Liu, T. Zhang, and T. Başar, "Fully decentralized multi-agent reinforcement learning with networked agents," 2018.

[17] Y. Zhou, S. Liu, Y. Qing, K. Chen, T. Zheng, Y. Huang, J. Song, and M. Song, "Is centralized training with decentralized execution framework centralized enough for marl?" 2023.

[18] A. Gao, S. Zhang, Y. Hu, W. Liang, and S. X. Ng, "Game-combined multi-agent drl for tasks offloading in wireless powered mec networks," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 7, pp. 9131–9144, 2023.

[19] O. Urmonov, H. Aliev, and H. Kim, "Multi-agent deep reinforcement learning for enhancement of distributed resource allocation in vehicular network," *IEEE Systems Journal*, vol. 17, no. 1, pp. 491–502, 2023.

[20] A. S. Kumar, L. Zhao, and X. Fernando, "Multi-agent deep reinforcement learning-empowered channel allocation in vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 1726–1736, 2022.

[21] X. Fu, Q. Yuan, Z. Zhuang, Y. Li, J. Liao, and D. Zhao, "TacNet: A tactic-interactive resource allocation method for vehicular networks," *IEEE Internet of Things Journal*, pp. 1–1, 2023.

[22] B. Ji, B. Dong, D. Li, Y. Wang, L. Yang, C. Tsimenidis, and V. G. Menon, "Optimization of resource allocation for V2X security communication based on multi-agent reinforcement learning," *IEEE Transactions on Vehicular Technology*, pp. 1–12, 2023.

[23] P. Dai, M. Wu, K. Li, X. Wu, and Y. Ding, "Joint optimization for quality selection and resource allocation of live video streaming in internet of vehicles," *IEEE Transactions on Services Computing*, pp. 1–14, 2023.

[24] Y. Cui, H. Shi, R. Wang, P. He, D. Wu, and X. Huang, "Multi-agent reinforcement learning for slicing resource allocation in vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–12, 2023.

[25] S. S. Shinde, A. Bozorgchenani, D. Tarchi, and Q. Ni, "On the design of federated learning in latency and energy constrained computation offloading operations in vehicular edge computing systems," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 2, pp. 2041–2057, Feb 2022.

[26] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1657–1681, 2017.

[27] W. Yang *et al.*, "Edgekeeper: A trusted edge computing framework for ubiquitous power internet of things," *Frontiers of Information Technology & Electronic Engineering*, vol. 22, no. 3, pp. 374–399, 2021.

[28] P.-Q. Huang, Y. Wang, and K.-Z. Wang, "Energy-efficient trajectory planning for a multi-UAV-assisted mobile edge computing system," *Frontiers of Information Technology & Electronic Engineering*, vol. 21, no. 12, pp. 1713–1725, 2020.

[29] Y. Gong, Y. Wei, Z. Feng, F. R. Yu, and Y. Zhang, "Resource allocation for integrated sensing and communication in digital twin enabled internet of vehicles," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 4, pp. 4510–4524, 2023.

[30] Y. Dai, D. Xu, K. Zhang, S. Maharjan, and Y. Zhang, "Deep reinforcement learning and permissioned blockchain for content caching in vehicular edge computing and networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 4312–4324, 2020.

[31] X. Zhu, Y. Luo, A. Liu, M. Z. A. Bhuiyan, and S. Zhang, "Multiagent deep reinforcement learning for vehicular computation offloading in iot," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 9763–9773, 2021.

[32] M. Z. Alam and A. Jamalipour, "Multi-agent drl-based hungarian algorithm (madrlha) for task offloading in multi-access edge computing internet of vehicles (iovs)," *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 7641–7652, 2022.

[33] K. Zhang, J. Cao, and Y. Zhang, "Adaptive digital twin and multiagent deep reinforcement learning for vehicular edge computing and networks," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 2, pp. 1405–1413, 2022.

[34] B. Hazarika, K. Singh, S. Biswas, S. Mumtaz, and C.-P. Li, "Multi-agent drl-based task offloading in multiple ris-aided iov networks," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 1, pp. 1175–1190, 2024.

[35] S. Guleng, C. Wu, X. Chen, X. Wang, T. Yoshinaga, and Y. Ji, "Decentralized trust evaluation in vehicular internet of things," *IEEE Access*, vol. 7, pp. 15 980–15 988, 2019.

[36] T. Zhang, C. Xu, B. Zhang, J. Shen, X. Kuang, and L. A. Grieco, "Toward attack-resistant route mutation for vanets: An online and adaptive multiagent reinforcement learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 23 254–23 267, 2022.

[37] E. M. Mianji, G.-M. Muntean, and I. Tal, "Trustworthy routing in vanet: A q-learning approach to protect against black hole and gray hole attacks," in *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*, 2023, pp. 1–6.

[38] T. Jing, Y. Liu, X. Wang, and Q. Gao, "A reservoir computing-based distributed trust sharing provisioning for internet of vehicle," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, 2022, pp. 3038–3043.

[39] J. Guo, X. Li, Z. Liu, J. Ma, C. Yang, J. Zhang, and D. Wu, "Trove: A context-awareness trust model for vanets using reinforcement learning," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6647–6662, 2020.

[40] S. Berri, J. Zhang, B. Bensaou, and H. Labiod, "Privacy-preserving data-prefetching in vehicular networks via reinforcement learning," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[41] S. Berri, J. Zhang, B. Bensaou, and H.Labiod, "Preserving location-privacy in vehicular networks via reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 18 535–18 545, 2022.

[42] Y. Liu and B. Mao, "On a novel content edge caching approach based on multi-agent federated reinforcement learning in internet of vehicles," in *2023 32nd Wireless and Optical Communications Conference (WOCC)*, 2023, pp. 1–5.

[43] Y. Lin, J. Bao, Y. Zhang, J. Li, F. Shu, and L. Hanzo, "Privacy-preserving joint edge association and power optimization for the internet of vehicles via federated multi-agent reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 6, pp. 8256–8261, 2023.

[44] H. Yang, J. Zhao, Z. Xiong, K.-Y. Lam, S. Sun, and L. Xiao, "Privacy-preserving federated learning for uav-enabled networks: Learning-based joint scheduling and resource management," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 10, pp. 3144–3159, 2021.

[45] M. Nasr-Azadani, J. Abouei, and K. N. Plataniotis, "Distillation and ordinary federated learning actor-critic algorithms in heterogeneous uav-aided networks," *IEEE Access*, vol. 11, pp. 44 205–44 220, 2023.

[46] S. Wang, Q. Zhu, H. Huang, Y. Lei, W. Zhan, and H. Duan, "Deep reinforcement learning based real-time proactive edge caching in intelligent transportation system," in *2023 8th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, 2023, pp. 162–166.

[47] D. Wei, J. Zhang, M. Shojafar, S. Kumari, N. Xi, and J. Ma, "Privacy-aware multiagent deep reinforcement learning for task offloading in vanet," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–15, 2022.