A Recent Reinforcement Learning Trend for Vehicular Ad Hoc Networks Routing

Woongsoo Kim, Junhong Min, Yongseok Son, and Jeongyeup Paek
Department of Computer Science & Engineering, Chung-Ang University, Seoul, Republic of Korea
{woongsu0614, dmc93, sysganda, jpaek}@cau.ac.kr

***Abstract*—Vehicular ad hoc networks (VANETs) are one of the most essential parts of intelligent transportation system (ITS). VANETs fulfill a crucial role in continuous traffic monitoring, emergency message transmission, and in-vehicle infotainment services. Given the short communication distance, constraints of unpredictable traffic environments, and the dynamic topology caused by high mobility, effective VANET routing is vital for network performance. However, prior ad hoc routing schemes in the literature are unsuitable for VANETs dynamic environments. For this reason, recent works focusing on VANETs have proposed *reinforcement learning* (RL)-based approaches. In this paper, we survey the literature that tackles the VANET routing problem using RL, summarizing which RL algorithms are used and their optimization goals. In addition, we analyze and discuss the limitations of RL-based approaches to propose guidelines for promising VANET routing solutions for constructing the future of ITS.**

***Index Terms*—Vehicular Ad Hoc Networks (VANETs), Intelligent Transportation System (ITS), Routing, Reinforcement Learning**

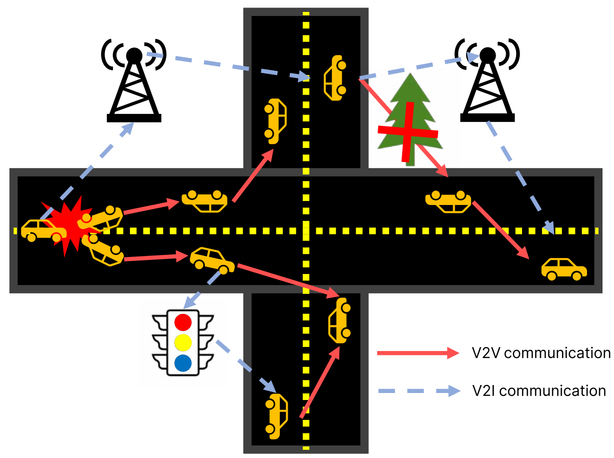

Fig. 1: ITS scenario with VANET routing

I. Introduction

Vehicular ad hoc networks (VANETs), an essential part of intelligent transportation system (ITS), offer a suitable technology to enhance road safety and traffic efficiency. VANETs support applications like emergency alerts, road safety, and collision prevention through vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), and vehicle-to-everything (V2X) communication, helping to implement key functionalities of ITS such as traffic monitoring [1], emergency message communication [2], and in-vehicle infotainment services [3].

However, there are several challenges to realizing VANETs, including the high mobility of vehicles, unpredictable traffic environments, and unstable wireless links. Without suitable approaches to overcome such difficulties, VANETs can undergo frequent communication link disconnections due to their dynamic network topologies. For VANET routing, many prior works based on heuristics have been proposed. Generally, they can be classified into broadcast-based, geocast-based, and unicast-based methods. Broadcast-based routing [4] facilitates multi-hop communication through flooding, while network congestion can occur due to excessive packet transmission. Geocast routing [5], on the other hand, sends data traffic in a specific direction to decrease congestion than broadcasting, but it requires knowledge of vehicle locations beforehand. Unicast routing [6] is efficient in using bandwidth and minimizing traffic; however, it is vulnerable to frequent changes in network topology and communication disconnection. For instance, Fig. 1 illustrates an emergency message communication scenario using unicast-based VANET routing in ITS when an accident occurs.

Recently, reinforcement learning (RL) has been considered a promising solution for VANET routing, which can overcome traditional drawbacks. RL is one of the machine learning approaches which enables agents to learn optimal decision-making through interactions with the environment. Using the RL model for VANET routing can be more effective than traditional routing protocols in terms of scalability and robustness because they can interact with dynamic environments with feedback to learn. However, there are still some limitations. For example, RL-based approaches cannot guarantee successful routing with observation of a single agent determined without observations of other agents [7]. For this reason, efforts are emerging to model and solve the VANET routing problem by utilizing *multi-agent reinforcement learning (MARL)*.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1A4A5034130 & No. 2021R1A2C1008840), and also by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2023-RS-2022-00156353) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation)

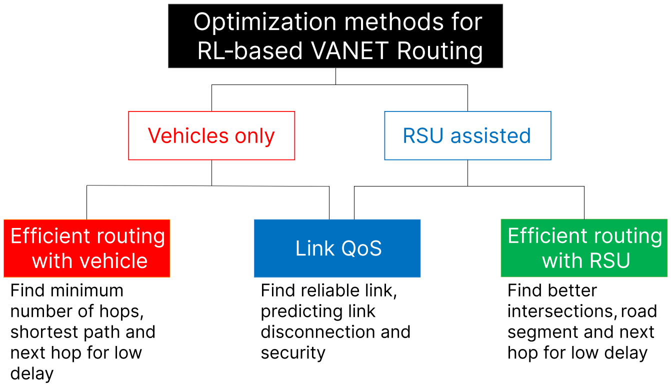

Fig. 2: Optimization objectives for RL-based VANET routing algorithms

This paper aims to provide future direction for RL-based VANET routing. Therefore, we explore prior works for RL-based VANET routing. Specifically, we compare several metrics and algorithms for their different objectives. Then, we discuss the limitations of prior works. Finally, we would like to find out which parts of RL-based VANET routing are currently on-problem and present guidelines for improvement in future research.

II. Overview of RL-Based VANET Routing

Unicast is considered the least resource-consuming approach in VANET routing. However, it requires designing the optimal routing from the source to the destination. In this section, we present an overview of the RL-based VANET routing topics, and what is needed for efficient VANET routing.

According to our survey, the most studied RL algorithm in RL-based VANET routing is a model-free algorithm, and q-learning is the most used among them. Most studies use the model-free algorithms for more realistic assumptions [8]. The road side unit (RSU) is an important device for collecting and disseminating traffic information to vehicles. However, deploying RSU everywhere around the road is difficult. Furthermore, network congestion may lead to potential RSU unavailability. To address this, we consider classifying VANET routing into two scenarios as shown in Fig. 2; RSU-assisted VANET routing and vehicle-only communication based VANET routing. The primary objective of VANET routing is to discover the optimal communication path for routing and optimize link quality to maintain stable, reliable routing. Efficient routing algorithms aim to achieve several goals, such as minimizing the number of hops, reducing end-to-end delay, and finding the shortest path. These algorithms utilize information provided by the surrounding vehicle environment to find the optimal path for efficient routing. In addition, improving link quality focuses on enhancing bandwidth efficiency, ensuring uninterrupted and secure link connections, and mitigating potential threats from malicious nodes. With this overview, we introduce RL-based VANET routing studies according to the perspective with and without RSU, and present consideration for improving RL algorithms which are used in subsequent sections.

III. Survey of VANET Routing with RL

In this section, we take a detailed look at the studies by algorithm from the point of view related to the goal of VANET routing. In the following subsections, we provide explanations for the references cited in Table. I, which are organized based on the objectives of efficient routing. We categorize the literature of routing without RSU focusing on vehicle-only solutions based on routing relaying strategies such as greedy, clustering, and various link Quality of Service (QoS) metrics. On the other hand, RSU-assisted solutions focus on selecting efficient routing by selecting the intersection, road, and next hop using one huge RSU or all RSU placed in multiple intersections. We present the result of surveyes in the latest works of RL-based VANET routing protocols, both in without RSU and RSU-integrated scenarios.

A. Q-learning

Q-learning is the most popular and widely used RL-based routing algorithm recently among researchers. It employs a q-table to store and update values representing the expected rewards for different state-action or action values.

Li *et al.* [9] propose a greedy algorithm based on q-learning, where the geographical area is divided into grid regions. This algorithm continuously selects the next grid to head towards. By offline learning with road data, it is possible to achieve high-performance routing focusing on areas with high traffic density to ensure connectivity. Wu *et al.* [10] propose a protocol that utilizes hello packets to periodically update the information in the one hop table and q-table, enabling adaptive learning of the efficient route. And also a novel hello packet structure is designed to avoid routing loops. Ji *et al.* [11] propose an RL-based real-time path exploration hybrid routing that detects blind paths with potential link disconnection risks, even without time expiration. The routing table is updated to mitigate intermittent link disconnections in dynamic VANET environments. Additionally, the q-table size is limited to prevent high computational costs. Several studies [12], [13] evaluate and enhance the performance of cluster-based VANET routing algorithms using q-learning approaches. These algorithms define link states through techniques such as fuzzy logic or the Gaussian mixture model to enhance the stability of links between clusters. Wang *et al.* [14] applying q-learning a VANET routing algorithm to utilize software-defined network (SDN)-enabled RSU to detect the traffic situation and dynamically select a more suitable routing algorithm between GPSR and AODV. Several other works [15]–[17] also aim to enhance the performance of efficient routing finding using RSU. A common characteristic among these works is the selection of the optimal next hop within the road segment at an intersection. Notably, these works construct scenarios assuming the existence of an RSU at each intersection.

B. Deep Reinforcement Learning

Deep reinforcement learning (DRL) is a solution that combines RL and deep learning, a technique in which an agent

TABLE I: Comparison objectives among RL algorithms

Algorithm	Objective	Routing strategy	Link QoS	Metrics	Improvement	Reference
Q-learning	Vehicle-only efficient grid-based routing	Grid-based greedy routing with q-learning	Reliable link	Vehicle position	Link reliability, transmission delay, throughput	[9]
	Vehicle-only efficient routing	Greedy routing with q-learning	Stable link	Vehicle position	End-to-end latency, packet delivery ratio, number of hop	[10]
	Vehicle-only efficient routing	Link life time with distance between one hop	Fresh link	Link life time, one hop distance	Packet delivery ratio, rount trip time, overhead	[11]
	Vehicle-only efficient cluster-based routing	Cluster-based routing with q-learning	Fuzzy logic or Gaussian mixture model based stable link	Cluster distance, link reliability	Link reliability, transmission delay, throughput	[12], [13]
	RSU-assisted efficient routing algorithm selection	Routing algorithm selection with q-learning	Vehicle speed and vehicle density based stable link	Vehicle speed, vehicle density, latency	Packet delivery ratio	[14]
	RSU-assisted efficient intersection, road, next hop selection	Greedy (Dijkstra) routing with q-learning	Vehicle position-based reliabile link	Vehicle position information, link reliability	End-to-end latency, packet drop rate, packet delivery speed, throughput	[15]–[17]
DRL	Vehicle-only efficient routing	Evolved Dijkstra routing with DQN	Stable link	Vehicle position, one hop delay	End-to-end latency	[18]
	Vehicle-only secure message routing	Secure intelligent message routing strategy with DRL	Reliabile link in urgency situation	Vehicle distance, message type, security attribute	End-to-end latency, high transmission security	[19]
	RSU-assisted efficient trust management routing	Trust value-based routing with DRL / DQL	Malicious node detection for reliable link	Vehicle location, forwarding ratio, trust value, ETX	End-to-end latency, throughput	[20]–[22]
MARL	Vehicle-only efficient routing	Routing with MARL	Fuzzy logic based stable link	Vehicle speed, vehicle density, link quality	End-to-end latency, packet delivery ratio, overhead	[23]
	RSU-assisted efficient routing	Routing algorithm selection with MARL	Stable link	Vehicle distance	End-to-end latency, packet failed ratio	[24]
	RSU-assisted efficient trust management routing	Grid-based secure routing with MARL	Malicious node detection for reliable link	Vehicle location, time-to-live, angle between mobility and relay node	Packet delivery ratio, number of hop	[25]

interacts with its environment and learns to make decisions that maximize rewards. DRL can simplify the solution by expressing it in the form of a function through a neural network, enabling efficient solutions and large-scale data processing.

Zhou *et al.* [18] propose a evolved Dijkstra algorithm with a deep q-network (DQN). In such cases, DRL-based routing method with evolved Dijkstra algorithm can achieve near minimum end-to-end delay in more feasible time to compare with a graph-based method. Liu *et al.* [19] propose a DRL-based intelligent message routing strategy for secure message transmission considering multiple message types and vehicle security in VANETs. The strategy adaptively selects routing based on node distances and security attributes, enabling fast and secure message transmission in VANETs. Zhang *et al.* [20]–[22] propose several algorithms to choose the next hop to enhance security and avoid malicious vehicles. SDN agent learns based on DRL, deep q-learning (DQL) based framework based on vehicle trust model and evaluates communication link and expected transmission count (ETX) delay.

C. Multi-Agent Reinforcement Learning

MARL is a technique that performs RL by considering the interaction among multiple agents and the environment. MARL-based VANET routing studies are used to explore efficient routing through cooperative decision-making by multiple agents and sharing experiences.

Jafarzadeh *et al.* [23] propose a MARL algorithm using a model-based RL algorithm and fuzzy logic, aiming to select links that can last as long as possible for routing. Each node transmits its q-value and received packets are evaluated for link connectivity based on fuzzy logic, then a neighboring node is selected using softmax. Lu *et al.* [24] solve routing selection optimization protocol using MARL. It uses the epsilon greedy algorithm to select actions and stores them in a replay memory to distribute the training of agents and reduce network delay through optimized router selection. They use a deep double q-learning network and dueling DQN to improve convergence speed and stability perspective. Zhang *et al.* [25] propose a grid-based approach, an online and adaptive MARL algorithm considering security, to utilize route mutation, commonly used in fixed topology, in VANET. They use grid-based states to avoid routing through malicious nodes and achieve routing based on the minimum angle relay nodes for possible movement directions through joint action learning using the estimated q-value.

IV. Discussion

In this section, we explore the primary challenges currently under investigation and outline the future research guidelines for RL-based VANET routing, particularly focused on the efficient unicast method. Some studies [9], [16], [17], [19], [23], [24] clearly describe state, action, and reward, yet the agent's definition remains unclear. This ambiguity could potentially confuse researchers during model creation. To the best of our knowledge, there has been no study dealing with the latency of reward in a vehicle-only environment. In the real world, reward feedback via VANETs communication might not be instantaneous, and the state change needs to reflect a result. In addition, studies using RSU assume that there is an RSU at every intersection, or attempt to address the matter via centralized structures such as SDN. However, these assumptions are challenging to accept as realistic. Our main investigation revolves around RL algorithms, categorized into q-learning, DRL, and MARL, revealing prevalent challenges within each. Difficulties arise when using q-learning because the environment becomes too simple or the q-table becomes too large in large-scale problems. Overcoming limitations of q-learning can be achieved through DRL, which employs functions instead of q-tables for solutions. Nevertheless, the intricate tasks of data labeling and state correlation complicate the learning process, and training with centralized aspects of the DRL model for VANET agents demands difficult formalization. Addressing these complexities, MARL, which can learn in various scenarios by cooperating with multiple agents, seems to be a realistic solution in the VANET environment where topology dynamically changes. However, challenges such as potential overfitting problems, high computational cost, and different agent targets still exist. We recommend MARL as a promising solution for RL-based VANET routing and simultaneously emphasize the need for continuous efforts and research to solve the remaining problems.

V. CONCLUSION

VANETs are an essential component in ITS applications. In this paper, we have surveyed the literature that tackles the VANET routing problem using RL, and investigated how to improve performance of VANETs by applying RL algorithms. Additionally, we have briefly summarized methods and limitations dealt with by the latest literature and presented research directions to achieve realistic VANET routing.

REFERENCES

[1] M. Won, "Intelligent Traffic Monitoring Systems for Vehicle Classification: A Survey," *IEEE Access*, vol. 8, pp. 73 340–73 358, 2020.

[2] X. Yang, L. Liu, N. H. Vaidya, and F. Zhao, "A Vehicle-to-Vehicle Communication Protocol for Cooperative Collision Warning," in *The First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, MOBIQUITOUS.* IEEE, 2004, pp. 114–123.

[3] H. T. Cheng, H. Shan, and W. Zhuang, "Infotainment and road safety service support in vehicular networking: From a communication perspective," *Mechanical systems and signal processing*, vol. 25, no. 6, pp. 2020–2038, 2011.

[4] G. Korkmaz, E. Ekici, F. Özgüner, and Ü. Özgüner, "Urban Multi-Hop Broadcast Protocol for Inter-Vehicle Communication Systems," in *Proceedings of the 1st ACM international workshop on Vehicular ad hoc networks*, 2004, pp. 76–85.

[5] S. Allal and S. Boudjit, "Geocast Routing Protocols for VANETs: Survey and Guidelines," in *International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*. IEEE, 2012, pp. 323–328.

[6] B. Karp and H.-T. Kung, "GPSR: Greedy Perimeter Stateless Routing for Wireless Networks," in *Proceedings of the 6th annual international conference on Mobile computing and networking*, 2000, pp. 243–254.

[7] L. Buşoniu, R. Babuška, and B. De Schutter, "Multi-agent Reinforcement Learning: An Overview," *Innovations in multi-agent systems and applications-1*, pp. 183–221, 2010.

[8] J. Achiam, "Spinning Up in Deep Reinforcement Learning," 2018, [Accessed: Aug 2023]. [Online]. Available: https://spinningup.openai.com/en/latest/

[9] F. Li, X. Song, H. Chen, X. Li, and Y. Wang, "Hierarchical Routing for Vehicular Ad Hoc Networks via Reinforcement Learning," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1852–1865, 2018.

[10] J. Wu, M. Fang, and X. Li, "Reinforcement Learning Based Mobility Adaptive Routing for Vehicular Ad-Hoc Networks," *Wireless Personal Communications*, vol. 101, pp. 2143–2171, 2018.

[11] X. Ji, W. Xu, C. Zhang, T. Yun, G. Zhang, X. Wang, Y. Wang, and B. Liu, "Keep Forwarding Path Freshest in VANET via Applying Reinforcement Learning," in *IEEE First International Workshop on Network Meets Intelligent Computations (NMIC)*, 2019, pp. 13–18.

[12] C. Wu, T. Yoshinaga, Y. Ji, T. Murase, and Y. Zhang, "A Reinforcement Learning-Based Data Storage Scheme for Vehicular Ad Hoc Networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 6336–6348, 2016.

[13] A. Nahar and D. Das, "Adaptive Reinforcement Routing in Software Defined Vehicular Networks," in *International Wireless Communications and Mobile Computing (IWCMC)*, 2020, pp. 2118–2123.

[14] C. Wang, L. Zhang, Z. Li, and C. Jiang, "SDCoR: Software Defined Cognitive Routing for Internet of Vehicles," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3513–3520, 2018.

[15] L. Rui, Z. Yan, Z. Tan, Z. Gao, Y. Yang, X. Chen, and H. Liu, "An Intersection-Based QoS Routing for Vehicular Ad Hoc Networks With Reinforcement Learning," *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[16] A. Lolai, X. Wang, A. Hawbani, F. A. Dharejo, T. Qureshi, M. U. Farooq, M. Mujahid, and A. H. Babar, "Reinforcement learning based on routing with infrastructure nodes for data dissemination in vehicular networks (RRIN)," *Wireless Networks*, vol. 28, no. 5, pp. 2169–2184, 2022.

[17] L. Luo, L. Sheng, H. Yu, and G. Sun, "Intersection-Based V2X Routing via Reinforcement Learning in Vehicular Ad Hoc Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 5446–5459, 2021.

[18] M. Zhou, L. Liu, Y. Sun, K. Wang, M. Dong, M. Atiquzzaman, and S. Dustdar, "On Vehicular Ad-Hoc Networks With Full-Duplex Radios: An End-to-End Delay Perspective," *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[19] B. Liu, G. Xu, G. Xu, C. Wang, and P. Zuo, "Deep Reinforcement Learning-Based Intelligent Security Forwarding Strategy for VANET," *Sensors*, vol. 23, no. 3, p. 1204, 2023.

[20] D. Zhang, F. R. Yu, R. Yang, and H. Tang, "A Deep Reinforcement Learning-based Trust Management Scheme for Software-defined Vehicular Networks," in *Proceedings of the 8th ACM symposium on design and analysis of intelligent vehicular networks and applications*, 2018, pp. 1–7.

[21] D. Zhang, F. R. Yu, and R. Yang, "A Machine Learning Approach for Software-defined Vehicular Ad Hoc Networks with Trust Management," in *IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–6.

[22] D. Zhang, F. R. Yu, R. Yang, and L. Zhu, "Software-Defined Vehicular Networks With Trust Management: A Deep Reinforcement Learning Approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 1400–1414, 2020.

[23] O. Jafarzadeh, M. Dehghan, H. Sargolzaey, and M. M. Esnaashari, "A Model-Based Reinforcement Learning Protocol for Routing in Vehicular Ad hoc Network," *Wireless Personal Communications*, vol. 123, no. 1, pp. 975–1001, 2022.

[24] C. Lu, Z. Wang, W. Ding, G. Li, S. Liu, and L. Cheng, "MARVEL: Multi-Agent Reinforcement Learning for VANET Delay Minimization," *China Communications*, vol. 18, no. 6, pp. 1–11, 2021.

[25] 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.