Intersection-Based V2X Routing via Reinforcement Learning in Vehicular Ad Hoc Networks

Long Luo, *Member, IEEE*, Li Sheng, Hongfang Yu, *Member, IEEE*, and Gang Sun, *Member, IEEE*

***Abstract*— With the rapid development of the Internet of vehicles (IoV), routing in vehicular ad hoc networks (VANETs) has become a popular research topic. Due to the features of the dynamic network structure, constraints of road topology and variable states of vehicle nodes, VANET routing protocols face many challenges, including intermittent connectivity, large delay and high communication overhead. Location-based geographic routing is the most suitable method for VANETs, and such routing performs well on paths with an appropriate vehicle density and network load. We propose an intersection-based V2X routing protocol that includes a learning routing strategy based on historical traffic flows via Q-learning and monitoring real-time network status. The hierarchical routing protocol consists of two parts: a multidimensional Q-table, which is established to select the optimal road segments for packet forwarding at intersections; and an improved greedy strategy, which is implemented to select the optimal relays on paths. The monitoring models can detect network load and adjust routing decisions in a timely manner to prevent network congestion. This method minimizes the communication overhead and latency and ensures reliable transmission of packets. We compare our algorithm with three benchmark algorithms in an extensive simulation. The results show that our algorithm outperforms the existing methods in terms of network performance, including packet delivery ratio, end-to-end delay, and communication overhead.**

***Index Terms*— VANET, intersection-based routing, congestion control, Q-learning.**

I. Introduction

IN RECENT years, intelligent transport systems (ITSs) have become a popular topic. The Internet of vehicles (IoV) is an important part of ITS and has emerged as a research hotspot in wireless network research. In vehicular ad hoc networks (VANETs), vehicles share status information with other vehicles or roadside units (RSUs) to promote safe driving decisions and obtain location-based services to improve road safety and achieve a relaxed driving experience.

To deliver messages in VANETs, data packets must be sent from source nodes to destination nodes through multihop wireless communication via vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications. Nodes in VANETs have more powerful equipment and greater computing capability than normal nodes in mobile ad hoc networks (MANETs); thus, they also are responsible for the routing function of calculating the optimal next hop. The features of VANETs make it difficult to design routing protocols. The high-speed mobility of vehicles causes an uneven distribution of nodes and disconnections of communication links. Furthermore, the network structure is constrained by the road topology. Intersections, buildings and obstacles may decrease signal intensity and have shadowing effects. Therefore, it is important to design an efficient and reliable routing protocol.

With the rapid development of the IoV, researchers have made improvements to traditional MANET routing protocols to adapt to highly dynamic vehicular networks and have applied various emerging technologies to the design of VANET routing protocols. Topology-based routing protocols use routing tables to store information about routing paths. The routes to all destinations must be established before communication [1], and each node periodically broadcasts control packets to share information with other nodes in the network and updates the routing table. Nodes forward data packets to hops on the selected route in the routing table [2]. Due to the storage of a large number of routing entries, the complete routes from the source node to the destination node are sustainably available. Therefore, the topology-based routing protocols can effectively reduce the packet loss rate but also increase the bandwidth overhead. With the application of global positioning system (GPS) technology, vehicular nodes can obtain their location coordinates on a real-time map and then forward data packets according to the location information shared between nodes. Location-based routing can reduce the impact of the high mobility of nodes and avoid the need to maintain the routing table of the entire network. Outstanding performance in terms of a high delivery ratio, short transmission delay and low communication overhead make location-based routing protocols the most suitable, scalable and promising strategies for VANETs [3].

To adapt to the dynamic nature and changeable environment of VANETs, segment metrics and traffic awareness have been introduced in recent location-based geographical routing protocols [4], [5]. Because the vehicular network topology is

Manuscript received August 13, 2020; revised December 2, 2020 and January 19, 2021; accepted January 19, 2021. Date of publication February 2, 2021; date of current version May 31, 2022. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFB1802800 and in part by the PCL Future Greater-Bay Area Network Facilities for Large-Scale Experiments and Applications under Grant PCL2018KP001. The Associate Editor for this article was S.-H. Kong. *(Corresponding author: Gang Sun.)*

Long Luo and Li Sheng are with the Key Laboratory of Optical Fiber Sensing and Communications (Ministry of Education), University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: longluo.uestc@gmail.com; 191249310@qq.com).

Hongfang Yu is with the Key Laboratory of Optical Fiber Sensing and Communications (Ministry of Education), University of Electronic Science and Technology of China, Chengdu 611731, China, and also with the Peng Cheng Laboratory, Shenzhen 518066, China (e-mail: yuhf@uestc.edu.cn).

Gang Sun is with the Key Laboratory of Optical Fiber Sensing and Communications (Ministry of Education), University of Electronic Science and Technology of China, Chengdu 611731, China, and also with the Agile and Intelligent Computing Key Laboratory of Sichuan Province, Chengdu 611731, China (e-mail: gangsun@uestc.edu.cn).

Digital Object Identifier 10.1109/TITS.2021.3053958

restricted by roads and buildings in urban scenes, the best routing path is selected by evaluating the adjacent road segments at each intersection. A group of intersection sequences between source nodes and destination nodes is selected as the packet transmission trajectory [6]. A widely used method in the field of IoV to obtain information about segments and traffic is real-time link assessment by transmitting collector packets (CPs) at intersections. CPs are forwarded from the current intersection to the next intersection to collect the network and traffic status of each segment between intersections; then, a weight is assigned to the segment according to the link quality. When transmitting data packets, a path with the maximum weight is selected to meet the demand according to the packet delivery ratio or delay. Although these protocols are applicable to a variety of dynamic VANET scenarios, the uninterrupted availability of routing metrics during packet communication requires frequent interactions of nodes and constant updating of calculated results [7]. The continuous measurement processes take additional communication bandwidth and time and thus increase the overhead and delay.

Since vehicles in daily life usually exhibit some obvious pattern characteristics from a macro perspective, such as regular visits to certain places and relatively stable vehicle density in different regions, the latest routing technology introduces machine learning methods to train adaptive network models and improve network performance. In these protocols, network characteristics can be learned by machine learning (ML) or reinforcement learning (RL) methods based on historical traffic data without sending CPs, thereby saving bandwidth resources and shortening communication delays. The existing routing protocols implemented via ML and RL predict the movement and trajectory of vehicles to ensure more reliable and effective packet transmission. However, the vehicle behavior is characterized by a large number of states and high uncertainty, and it is difficult to train a network model to make accurate predictions. As an increasing number of vehicles join VANETs, the algorithm will become increasingly complicated and difficult to converge.

In view of the above deficiencies, we propose a routing protocol based on Q-learning that makes full use of macroscopic traffic flow characteristics and microscopic vehicle position behavior for hierarchical packet routing. The distribution of vehicles on different road sections has a certain regularity. Bustling roads close to the city center have large traffic volumes and high vehicle densities, while roads close to the suburbs have less traffic volume and low vehicle density. Vehicle density is a key factor that affects routing performance. For example, frequent disconnections of links will occur in the low vehicle density scenario, whereas packet collision and loss can be caused by signal interference in the high vehicle density scenario. Considering the actual situation, we train a multidimensional Q-table based on traffic flow on urban road segments between intersections. The Q-value indicates the tendency of a packet to be forwarded to each segment, thus ensuring that the most suitable path for packet transmission is chosen. In the segment, we design a location-based V2X algorithm to select the best relay node. The hierarchical routing algorithm has shown considerable advantages over other simple location-based algorithms and ML or RL methods. The main contributions of this paper are as follows:

- We design a hierarchical routing framework that combines routing path planning via Q-learning with relay hop selection via a location-based algorithm. First, the packets are forwarded to the road with the best communication conditions, and then the most appropriate relay is selected on the road, which improves the reliability of transmission.
- We compute a multidimensional Q-table via an improved Q-learning algorithm to implement routing path selections between any two intersections. Traffic flow characteristic, such as the vehicle density, is used as the parameter in Q-learning. The agent makes routing decisions according to the Q-table. Traffic aware overhead and delay are reduced greatly.
- We introduce a congestion control mechanism to prevent potential network congestion. By monitoring the link status in real time, an alternate road segment will be activated once the network is congested. The success rate of packet transmission is improved.

The rest of this paper is organized as follows. Section II reviews the existing research work on routing protocols in VANETs. Section III provides the problem statement and the motivation of the proposed routing scheme. Section IV introduces some preliminary preparations and describes the framework of our model. Section V introduces the Q-learning algorithm and presents our routing protocol in detail. Section VI illustrates the simulation settings and analyzes the simulation results. Finally, Section VII concludes this paper.

II. Related Work

A. Routing Based on Geographic Location in VANETs

Compared with the existing routing protocols in VANETs, location-based geographic routing is considered one of the best solutions to dynamic network topology [8]. Vehicles obtain their own positions through GPS technology and then broadcast their status information between neighboring nodes. Routing decision-making is based on the position of the next hop and destination [9]. In protocols that adopt a greedy strategy, vehicles forward the packet to the next hop closest to the destination node in the neighbor table. Greedy perimeter stateless routing (GPSR) is a classical location-based routing protocol that combines a greedy strategy and perimeter forwarding [10]. However, the transmission of beacon packets causes a certain delay in updating neighbor information, which increases the possibility of dropping packets. The author of [11] proposed a predictive geographic routing protocol for VANETs to predict the future position of a vehicle according to its current location and velocity status at a beacon interval. The prediction of mobility can mitigate the effects of a highly dynamic vehicular network. In addition to accurate positioning, location-based routing protocols face the challenge of local optima. In a network with sparse vehicle density, the source node may not find the appropriate next hop within its

communication range. In [8], the vehicle stores and carries the packet until meeting an available relay; however, if the relay node is unreachable for a long time, then the delay will increase and the communication quality will degrade. The lifetime (LT) field is given in the packet header to indicate the maximum tolerable time a packet can be buffered until it reaches the destination in [12].

In [13]–[15], the authors consider the difference in message transmission on roads and intersections and propose intersection-based routing protocols. Vehicles may change their movement directions suddenly at intersections, leading to packet loss. Reliable traffic-aware routing (RTAR) presented in [15] distinguishes between packet forwarding in road and intersection areas. Data packets are first forwarded to the closest intersection to the destination via road area reliable routing (RARR); then, intersection area reliable routing (IARR) is performed to forward the packets from the intersection to the destination vehicle. Packets may pass through multiple intersections to reach the destination, although RTAR in [15] takes only the last intersection of the path into consideration. In [16], the author proposed real-time intersection-based segment aware routing (RTISAR) to select an optimal road segment as the packet transmission path by assessing the link quality at each intersection. The CPs collect information about link connectivity, distance to destination and vehicle density to evaluate the segment status. In [17], vehicles waiting at traffic lights at intersections establish multihop links with the adjacent intersections to estimate the link qualities on the road segments. However, the use of CPs results in the problems of broadcast storms and information redundancy. Some bioinspired routing protocols have been proposed to reduce the number of CPs by imitating the intelligent behavior of creatures in nature. In [18], the authors used the microbe-inspired cellular attractor selection mechanism (CASM) to select the next hops. CASM mimics the process of protein synthesis and degradation to adaptively find and maintain the routing path. The author of [19] proposed a traffic-aware routing protocol based on ant colony optimization. The source node finds the optimal routing path by broadcasting forward and backward ants and collecting pheromone messages. The protocols in [18], [19] can reduce the communication overhead to a certain extent via the self-adaptability characteristic of biology. However, when the link states change frequently, a number of feedback packets are still required to update information in a timely manner. The clustering approach to disseminate the information in VANETs is considered to reduce some information redundancy, although the stability of the cluster is still a challenge because of the rapid topological changes of the VANETs network [20].

Intersection-based routing has been considered to improve the performance of the geographic routing protocol. The widely used traffic aware routing obtains the traffic status by broadcasting collector packets periodically. In this paper, we adopt the reinforcement learning method to evaluate the link conditions on road segments adjacent to the intersection, which leads to considerable communication cost savings.

B. Routing Based on Machine Learning in VANETs

As an emerging technology, ML has been applied in various studies. In daily life, the trajectories of vehicles show a certain regularity. For example, people drive from home to work at a similar time in the morning following the same trajectory due to the constraints of the road structure. In [21], researchers predict the future short-term route of vehicles and their packet transmission probability for a specific destination based on historical trajectories and a hidden Markov model. The possible trip sequence and forwarding capability of nodes are the metrics chosen to determine the next hop. However, a large number of RSUs are deployed to prevent the vehicle density from affecting the transmission performance. In [22], the authors model vehicle arrival as a nonhomogeneous Poisson process and adopt an artificial neural network (ANN) model to learn and predict traffic flow on the road. Routing is designed for software-defined network (SDN)-assisted heterogeneous VANETs for the purpose of delay minimization. However, other network performance improvements, such as the packet delivery ratio and overhead ratio, are not reflected in the simulation results. The authors of [23] introduced the support vector machine (SVM) algorithm to select the routing metrics and their weight factors by classifying nodes into outstanding next hops and ordinary next hops. SVM is a supervised ML method that requires a large amount of classified data, including the velocity, moving direction, acceleration, delay, and packet delivery ratio. Vehicle status is continuous and variable; therefore, the training dataset is large and the algorithm is complex.

RL is a branch of ML. Compared with supervised learning and unsupervised learning, RL obtains the final output result through a series of decision-making processes to make inferences about the external environment [24]. In [25], [26], the authors defined nodes in a VANET as the states of the agent and the possible next hop in the neighbor table as the set of actions of the Q-learning algorithm. Each node maintains a Q-table and chooses the neighbor node with the largest Q-value as the best next hop. When the number of vehicles increases, the updating and maintenance of the Q-table entails considerable communication overhead and computing resource expenditures. The authors of [27] divide geographical areas into grids and use the Q-learning algorithm to learn the traffic flow characteristics in the grids. First, the optimal next grid is selected based on the Q-values in the Q-table; then, a greedy method or a Markov prediction method is applied to choose the relay node in the selected grid. The protocol addresses the issue of delivering packets from mobile vehicles to a stationary destination [27]. However, communication in the real world usually requires packets to be sent to multiple destinations. In addition, the protocol always chooses the grid sequence with the largest Q-value and vehicle nodes in these grids may be overloaded when the network contains a large number of packets.

The routing protocols via ML select the next hop by predicting the micro behaviors of the vehicles. The dynamic environment of the VANETs and the complex road topology lead to a vehicular action space with a large size and high

uncertainty. In the proposed protocol, we select the optimal routing path based on the macro traffic flow characteristic. The size of state space and action space has been reduced considerably; therefore, the Q-learning algorithm converges at a faster speed.

C. Congestion Control Mechanism in VANETs

Packet routing in VANETs is often accompanied by network congestion problems. Periodic traffic-aware packets and beacon packets occupy a large amount of bandwidth resources, which limits the load capacity of the network. An overloaded network has poor performance with a large delay, low throughput and frequent data collisions [28]. In [28], the authors predict the future short-term surrounding vehicle density via an altruistic prediction algorithm and then adjust the transmission parameters of the node, such as the signal transmitting power and rate. The adaptive transmit parameters can prevent peaks in channel load. This method has substantial hardware requirements and ignores shadow fading caused by obstacles. The author of [29] used vehicles parked on the roadside to alleviate the shadowing effect. To avoid broadcast storms, the nodes are sorted by link quality, and the packets are sent to only the first k nodes. The premise of this protocol is that the communication equipment remains running even when the vehicle has stopped.

Routing protocols via ML or RL prevent information redundancy caused by broadcasting traffic-aware packets but cannot obtain the real-time link status of a highly dynamic network. If the network is congested but undetected, then there is a high possibility of packet loss and large delay. The authors of [30] proposed the SDN-assisted routing protocol to make routing decisions related to load balancing. The SDN monitors the global network load in real time, finds two low-cost transmission paths based on the path loss function, and selects light-loaded relay nodes based on the utility function on these two paths [30]. When one of the two paths is congested, another path is activated as an alternate transmission path to balance the network load. SDN controls the network and plans the routing path in a centralized manner, and the computing resources of vehicles as simple forwarding nodes are not fully utilized. In [31], the authors proposed hybrid routing by combining dedicated short-range communication (DSRC) between vehicles and wired communication between RSUs. The wired RSU network improves the reliability of transmission and reduces the communication load of vehicle nodes. However, the structuring and running of wired networks is expensive, and the author intended to employ ML to achieve environmental awareness in their future work [31]. The authors of [32] added a Q-learning module to each RSU to explore the environment, and vehicles with lower queuing delays were preferentially selected as relays. However, this approach focuses on the issue of V2I message delivery.

Routing via ML based on historical data is implemented with a lag. In our protocol, the RSUs not only forward packets at the intersections but also monitor the real-time link status of road segments to prevent network congestion.

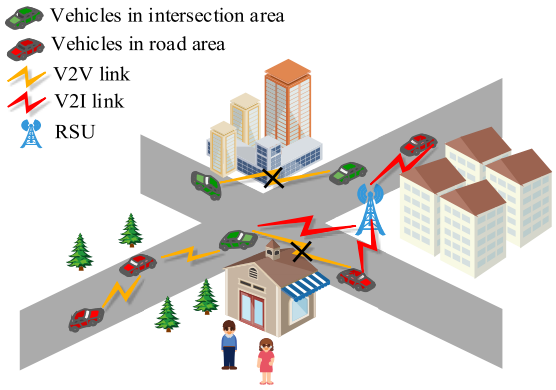


Fig. 1. Using an RSU as a relay at an intersection.

III. Problem Statement and Motivation

A. Problem Statement

The main research problems considered in this paper can be summarized as follows.

- Traffic-aware routing protocols have high overhead due to the generation and transmission of a large number of traffic-aware packets.
- The existing routing algorithms based on ML or RL are complex. The states of vehicles are variable and highly uncertain under the complex structure of city roads.
- Pretrained routing methods based on ML or RL cannot perceive dynamic networks and prevent network congestion. A good solution to network congestion could substantially improve network performance.

B. Motivation

In response to the limitations of the research mentioned above, we propose the intersection-based V2X routing protocol based on Q-learning and monitor the segment conditions by RSUs in real-time to prevent possible network congestion. By extracting the characteristics of historical traffic flow on roads and training the Q-learning algorithm to obtain the Q-table, the optimal routing paths can be found by selecting the road segment with the maximum Q-value. No traffic-aware packets are broadcast to obtain the real-time link information before the transmission of data packets; therefore, the routing has low overhead and delay. We take intersections and road segments as states and actions of Q-learning, respectively, which greatly decreases the number of states and makes the algorithm easy to implement. In a previous RL routing protocol that regarded the geographical environment as the state, the geographic area was divided into grids [27]. First, the optimal grid was chosen and then packets were forwarded to the next hop in the grid. However, intersections and buildings in the grid affect the signal transmission, especially in urban scenes. In our work, RSUs are introduced as relays at intersections to alleviate the problems caused by sudden changes in vehicle direction and signal attenuation due to building occlusion as shown in Fig. 1.

The Q-table in [27] is computed off-line and remains static during the routing process; therefore, it cannot observe the network load to adjust routing decisions. The load of a vehicle

depends on the length of the buffer queue [16]. We install monitoring modules on RSUs at intersections to obtain the total number of packets and estimate the average load of the local network on road segments. If the network on a path is overloaded, the packet will be forwarded to an alternate road segment.

IV. Preliminary Preparation and Framework Description

A. Preliminary

Although vehicles move quickly and have highly uncertain action patterns, the trajectories of vehicles depend on the owners' social behavior and the road structure. The traffic flow in different sections presents a high degree of regularity. The closer a road is to a busy block, the greater the traffic flow. The density of vehicles on main roads is higher than that on secondary roads and branch roads. As the number of personal cars increases, we have entered the era of transportation big data [33]. The authors of [34] measured the traffic volume of multiple urban roads adjacent to intersections in Huainan, China. The results show that the vehicle densities on different road segments are relatively stable, the average similarity coefficient of traffic flow on different days is high, and the fluctuation coefficient is low.

The node density in a network is one of the most important factors affecting communication quality. In low-vehicle-density scenarios, the source node is less likely to find an appropriate relay for forwarding packets, which leads to frequent link disconnections [27]. When the vehicle density is too high, extensive communication needs are observed; however, the nodes' resources and capacities are limited. As a result, the network becomes congested and suffers a high packet drop rate and low throughput [35]. In [28], the theoretical optimal local density is calculated according to the channel capacity of a node. The network has the best performance when there are 28 neighboring vehicles within the communication range. In this paper, we conduct simulations on the packet delivery ratio under various vehicle densities using the OMNeT++ tool and find that the packet delivery ratio is highest at a vehicle spacing of 100 m. As shown in Fig. 2, vehicular intervals that deviate further from 100 m present more obvious performance degradation. In the subsequent process of algorithm design, road segments with vehicle density close to the optimum value are selected with higher priority.

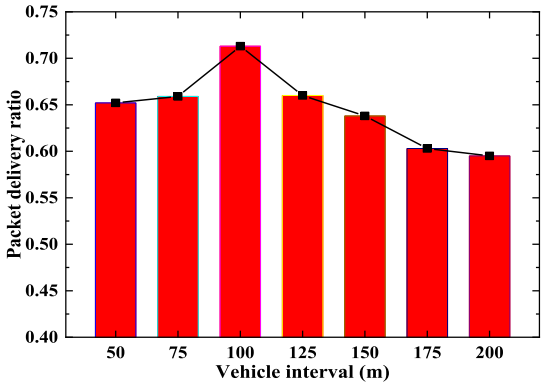


Fig. 2. Packet delivery ratio with vehicle density.

B. System Framework

We propose a hierarchical system framework to realize the routing of packets in urban scenarios. In our scenario, two-way road segments are adjacent to intersections. RSUs connected to the upper server are deployed at intersections to communicate with the vehicle for reliable packet forwarding in intersection areas. Vehicles are equipped with a GPS system and digital map so that they can obtain the real-time geographic positions of themselves and RSUs.

The server trains the Q-learning algorithm based on historical traffic flow information of segments and sends the Q-table to the RSUs at intersections. When the packets reach the intersections, the RSUs decide which road segment to forward packets to. The intersection sequence is determined by selecting the path with the maximum Q-value in the Q-table. The improved location-based greedy forwarding strategy is adopted to select relays on the path between two adjacent intersections. The server estimates the resource utilization status based on statistical information of packets entering and leaving the road segment from intersections of two ends. When the resource utilization exceeds the threshold, the segment with the second largest Q-value is selected as the alternate routing path to avoid a congested network section. The system structure is shown in Fig. 3.

To achieve the described framework, several issues must be addressed.

1) The nodes in the network need to exchange status information periodically. Timely updating of neighbor tables is critical to finding a suitable relay in location-based routing schemes. Thus, vehicles and RSUs broadcast beacon packets, including the location, velocity, and acceleration information, at regular intervals.
2) The multidimensional Q-table must be established. The conventional two-dimensional Q-table only provides a route to a fixed destination. To satisfy the demands of real situations, communication between any two nodes should be allowed.
3) The packets need to be forwarded to the path with the optimal routing performance, which depends on the vehicle density and network load. Thus, a method that combines RL and real-time monitoring is selected.
4) The forwarding strategy in road segments must be determined to realize V2V and V2I communications to ensure that packet routing is reliable and efficient.

V. Algorithm Design

In this section, we will explain the details of the intersection-based V2X routing via Q-learning (IV2XQ). The algorithm consists of three parts: i) a multidimensional Q-learning model, ii) a routing path selection model, and iii) a V2X communication model.

A. Multidimensional Q-Learning Model

Q-learning is a model-free RL method that evaluates the final result of a series of decision-making processes to

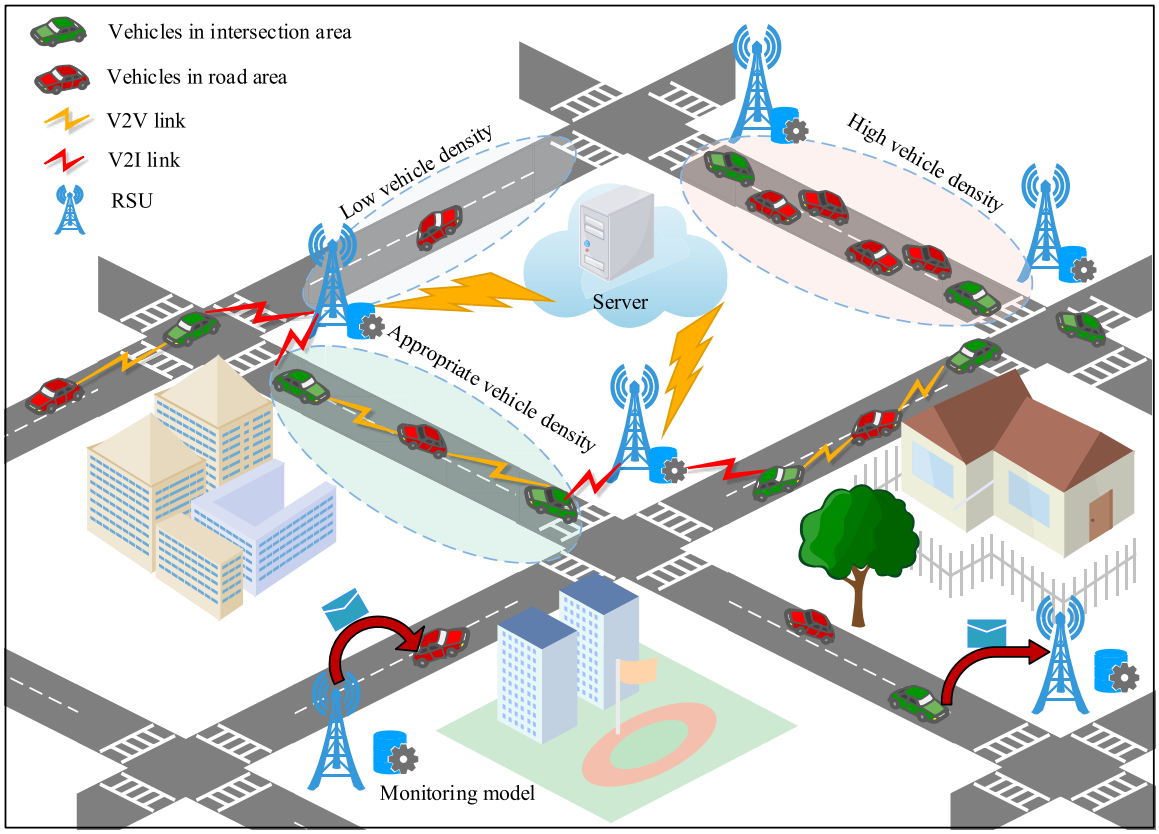


Fig. 3. System structure.

continuously explore an unknown environment. To maximize the cumulative rewards, the algorithm provides the optimal control strategy through trial and error [36]. The process of producing the optimal control strategy of Q-learning coincides with the idea of selecting the optimal routing path. Therefore, we apply the Q-learning algorithm to the routing model. Q-learning is based on the Markov Decision Process (MDP), which can be described by the tuple (S, A, P, R, γ) [37], where S denotes the finite state space set; A denotes the finite action space set; P means the state transition probability, with $P(s_{t+1}|s_t, a_t)$; R is the immediate reward; and γ is the discount factor.

The agent is the server with historical traffic flow information, and it explores the VANET network and implements the routing path planning for packet forwarding. The agent is the learner and decision maker in Q-learning and learns control strategies by constantly interacting with the environment.

The environment is the entire VANET scenario and includes all objects that interact with the agent. It provides the appropriate response to actions and changes states accordingly.

The state space (denoted as S) is the set of all intersections within the area. $s_t \in S$ can be depicted as I_i^t, and it indicates that the packet reaches intersection I_i at time t, and i is the identifier of the intersection.

The action space (denoted as A) is the set of road segments adjacent to an intersection and includes the north, south, east and west roads. $a_t \in A$ is the action taken by the agent at time t, and it determines the road segment for message forwarding. The action of selecting a road segment to forward packets makes the packets at I_i^t transition to I_j^{t+1}.

The immediate reward (denoted as R) is given to the agent after conducting the action of forwarding packets to one of the road segments. R_t is the reward that the agent receives for transitioning from the current state s_t to the next state s_{t+1} after performing action a_t. If the packet is forwarded to the destination intersection D_k, the reward is $\varphi, \varphi > 0$; otherwise, the reward is 0. The reward is assigned according to Formula (1).

$$R_t = \begin{cases} \varphi & \text{if reaching } D_k; \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

To explore the environment, the agent takes actions and transitions among states. Each time the agent conducts an action, the environment gives an immediate reward. When the agent achieves its goal after passing through a series of states, then the environment rewards the agent. If the agent fails, then the environment does not give a reward or even punishes the agent to avoid repetition of the error. The agent strengthens its memory through trial and error to learn the best strategy. The optimal control decision is achieved by selecting the action with the largest reward in each state.

The Q-table stores Q-values for state-action pairs. The value $Q(s_t, a_t)$ represents the expected reward of taking action a_t in current state s_t. The agent performs the action to gain the maximum revenue in each state based on the Q-values. By exploring the environment, the agent aims to learn a policy π

to maximize its cumulative reward. The policy π is evaluated by the function $Q^{\pi}(s,a)$ in (2).

$$Q^{\pi}(s,a) = E\left\{r_t(s_t,a_t) + \gamma\, Q^{\pi}(s_{t+1},a_{t+1})\right\} \tag{2}$$

The goal of training in Q-learning is to make the Q-matrix convergent, that is, the values of the matrix no longer change as the number of iteration increases. The training formula of Q-learning is as follows:

$$Q(s_t,a_t) \leftarrow (1-\alpha)\times Q(s_t,a_t) + \alpha \times \left\{R_t + \max_{a'} Q(s_{t+1},a')\times\gamma\right\} \tag{3}$$

where s_{t+1} is the next state the agent transitions to after conducting action a_t; a' represents the possible actions the agent can take in state s_{t+1}; and α is the learning rate that determines the extent to which new information overrides old information. If $\alpha = 0$, then the agent cannot learn any new information; if $\alpha = 1$, then the agent only considers the latest information. The discount factor γ is used to determine the scale of future reward. The agent pays more attention to previous experience when the discount factor is large, whereas the agent is more concerned about immediate interests when the discount factor is small.

In this paper, we obtain the multidimensional Q-table by training the Q-learning algorithm with historical traffic flow on road segments in an urban scenario. Q-values in the Q-table are used for path selection and packet forwarding. Packets travel along the road to one intersection after another until reaching the destination. Routing decisions regarding which road segment to follow was determined at each intersection. We regard the current intersection I_i^t as the state at time t. The action a_r^t of selecting a road segment r for forwarding packets makes the state I_i^t transition to I_j^{t+1}. r is the element of the set of roads adjacent to the intersection, usually in four adjoining directions, that is, $r \in Road = \{North, South, East, West\}$. For example, if the RSU takes action a_{North}^t, then the packet will be forwarded to the northbound road at the intersection. The original two-dimensional Q-matrix is improved to a three-dimensional matrix, including dimensions of the current intersection, the destination intersection and the action, which are denoted by I, D and a, respectively. The dimension of the state in the original matrix is replaced by the current intersection and the destination intersection. $Q(I_i^t, D_k, a_r^t)$ indicates the Q-value of taking action a_r^t at the current intersection I_i^t to reach destination intersection D_k. By increasing the dimension, the states of the agent become easier to observe. All the intersections can be reachable through the optimal routing path from the source intersections based on the Q-table. The action-utility function used to update the Q-table iteratively is defined as follows:

$$Q(I_i^t, D_k, a_r^t) \leftarrow (1-\alpha)\times Q(I_i^t, D_k, a_r^t) + \alpha \times \left\{R_t(I_i^t, D_k, a_r^t) + \max_{a^{t+1}} Q(I_j^{t+1}, D_k, a^{t+1})\times\gamma_t\right\} \tag{4}$$

The learning rate α is set to 0.1 according to the empirical value. In the traditional Q-learning algorithm, the discount factor γ is a fixed number between 0 and 1. To select the path with link reliability and continuity, we make the discount factor a dynamic parameter related to the vehicle density and distance to the destination. The definition of γ is as follows:

$$\gamma_t = \beta \times SP_t + (1-\beta)\times DP_t \tag{5}$$

where β is the weight factor, with $0 \le \beta \le 1$; SP_t is the impact factor of the vehicle density; and DP_t is the distance progress. SP_t and DP_t are defined according to Formulas (6) and (7), respectively:

$$SP_t = 1 - \frac{|s_{ij} - s_{best}|}{s_{\max} - s_{\min}} \tag{6}$$

$$DP_t = \max(0, 1 - \frac{dist_{t+1}}{dist_t}) \tag{7}$$

$$dist = \sqrt{(x_I - x_D)^2 + (y_I - y_D)^2} \tag{8}$$

In Formula (6), s_{max} is the maximum vehicle density; s_{min} is the minimum vehicle density on road segments; s_{best} is the optimal vehicle density calculated in Section IV; and s_{ij} is the vehicle density of the road segment between intersection I_i^t and intersection I_j^{t+1}. In Formula (7), $dist_t$ is the distance between the current intersection I_i^t and the destination intersection D_k; and $dist_{t+1}$ is the distance between the next intersection I_j^{t+1} and the destination intersection D_k. If the next intersection is further than the current intersection from the destination intersection, then DP_t will be 0. The distance is calculated according to Formula (8). When s_{ij} is close to the optimal vehicle density and the next intersection is closer to the destination intersection, the discount factor γ is large. The historical flow data of the road from the intersection I_i^t to the intersection I_j^{t+1} are used to update the Q-value. By implementing the dynamic parameter γ, we ensure that packets are transmitted on the path with a vehicle density as close to the optimal value as possible. Although adding one dimension to the Q-table increases the number of states, the introduction of distance progress in Formula (5) promotes the convergence rate of Q-learning.

During the training process of the Q-learning algorithm, the agent needs to explore the environment to discover states and use the values of state-action pairs to select the optimal action. In this paper, we adopt an ε- greedy strategy to realize exploration-exploitation. The agent selects an action from action space A randomly and uniformly with probability ε and selects the optimal action under the current state with probability 1 - ε. The improved Q-learning algorithm is shown in Algorithm 1.

B. Routing Path Selection Model

The agent obtains the Q-table from the above algorithm and distributes the Q-table to the RSUs at the intersections. The source vehicle obtains its own position and the destination vehicle's position by GPS and obtains the road topology from a digital map. Then, the source intersection and the destination intersection can be determined. There are usually candidate intersections at both ends of the roads on which the source vehicle and destination vehicle are located. We compare the distance between two candidate source intersections and

Algorithm 1 Improved Q-Learning Algorithm

Input: ε, α, β, S, A
Output: three-dimensional Q-table
Initialize the Q-matrix and R-matrix to all 0 matrices
for all D_k in the set of intersections S, **do**
  $R[D_k] = \varphi$;
  for *episode* $= 1$ to M, **do**
    select an intersection as the initial state;
    for $t = 1$ to N, **do**
    generate a random number $rn \in (0, 1)$;
    if $rn \leq \varepsilon$ **then:**
      choose an action in A randomly;
    else if $rn > \varepsilon$ **then:**
      choose the action with maximum Q-value;
    end if
    calculate γ_t with Formula (5);
    update Q-value with Formula (4);
    if I^{t+1} is D_k, **then:**
      break;
    end if
  end for
end for
$R[D_k] = 0$;
end for

the destination vehicle and choose the one with a shorter distance as the source intersection. The destination intersection is selected based on the distance to the destination vehicle and the cosine of the vehicle motion angle as shown in Formula (9). The candidate intersection with larger weight$w$ is chosen as the destination intersection.

$$w_i = \omega \times (1 - \frac{dist_i}{c}) + (1 - \omega) \times \cos(\overrightarrow{dV_i}, \overrightarrow{pV_i}) \quad (9)$$

where

$$\overrightarrow{dV} = (a_{dV}, b_{dV}) = (v_x t, v_y t) \quad (10)$$

$$\overrightarrow{pV} = (a_{pV}, b_{pV}) = (x_D - x_{vd}, y_D - y_{vd}) \quad (11)$$

$$\cos(\overrightarrow{dV}, \overrightarrow{pV}) = \frac{a_{dV} \times a_{pV} + b_{dV} \times b_{pV}}{\sqrt{a_{dV}^2 + b_{dV}^2} \times \sqrt{a_{pV}^2 + b_{pV}^2}} \quad (12)$$

where $dist_i$ is the distance between candidate destination intersection D_i and destination vehicle vd; c is a constant to normalize the distance; $\overrightarrow{dV}$ is the displacement vector of the destination vehicle at beacon interval Δt; and $\overrightarrow{pV}$ is a position vector starting at the destination vehicle and ending at the candidate destination intersection. The latter term of Formula (9) is the cosine of the angle formed by $\overrightarrow{dV}$ and $\overrightarrow{pV}$, which represents the relationship between the direction of vehicle movement and the location of the candidate destination intersection. If the vehicle is moving towards the intersection, then the value of the cosine in Formula (9) is positive; otherwise, it is negative. ω is a weight factor between 0 and 1.

At the beginning of routing, data packets are forwarded to the source intersection. The RSU at the intersection forwards packets to the appropriate road segment according to the Q-value. First, the RSU determines the state by checking the current intersection and the destination intersection. Then, the RSU looks up the Q-table to choose the action with the maximum state-action value in the current state. The RSU takes the action to forward packets to the selected path, and packets arrive at an intersection adjacent to the road through V2V communications on the path. The above process is repeated until the packets reach the destination intersection.

TABLE I
RECORD OF PACKET NUMBERS

	North	*South*	*East*	*West*
in	*North_in*	*South_in*	*East_in*	*West_in*
out	*North_out*	*South_out*	*East_in*	*West_out*

Routing via the Q-table fully exploits the historical traffic flow information; however, link conditions cannot be learned in real time, which increases the possibility of network congestion. Using the above routing process, all packets passing through intersections will be forwarded by RSUs. Assume that vehicles on each road generate packets at the same rate: the resource utilization on the road can then be estimated by counting the number of incoming and outgoing packets. Each RSU holds a table that records the number of incoming and outgoing packets from roads adjacent to the intersection. The record table is shown in Table I. Within the period of a beacon interval, when the RSU receives packets from road $r \in \{North, South, East, West\}$, *r_out* increases; and when the RSU forward packets to road r, *r_in* increases. The values in the table are updated periodically.

When the RSU is about to forward packets along one of the roads, the server extracts the data from the RSUs at both ends of the road and calculates the link average load as shown in Formula (13).

$$avg\_load = \frac{r\_in_i + r\_in_j - r\_out_i - r\_out_j}{Num_r} \quad (13)$$

where r_in_i is the number of packets entering road r from intersection I_i and r_out_i is the number of packets leaving road r from intersection I_i. Intersections I_i and I_j are at opposite ends of the road r to be evaluated. Num_r is the total number of vehicles on road r. The average load of each vehicle node can be obtained by Formula (13). If the value exceeds the preset threshold θ, the nodes' buffer queues are close to full and the network on this road is congested. θ is set to 0.5 in this paper. We choose the road segment with the second largest Q-value as the alternate route path. In each intersection, the RSU forwards packets to the path with great distance progress and good communication conditions; therefore, the routing performance is guaranteed. The process of selecting the routing path is shown in Algorithm 2.

C. V2X Communication Model

To forward packets on the selected routing path, we adopt the improved greedy strategy to select the relay. Vehicles maintain neighbor tables containing the neighbor nodes' status information. The neighbor tables are updated periodically by exchanging beacon packets to adapt to a highly dynamic network, and the routing path is partitioned according to

Algorithm 2 Routing Path Selection Model

Input: $Q\text{-}table$, θ
Output: Action a_r^t
for each packet received by RSU, **do**
  if the packet comes from r, **then:**
    $r\_out{+}{+}$;
  end if
  determine the current state (I_i^t, D_k);
  for $r \in Road = \{North, South, East, West\}$, **do**
    choose action a_r^t with the maximum $Q(I_i^t, D_k, a_r^t)$;
  end for
  calculate the avg_load of the selected road segment;
  if $avg\_load > \theta$, **then:**
    choose a_r^t with the second largest $Q(I_i^t, D_k, a_r^t)$;
  end if
  $r\_in{+}{+}$;
end for
if the timer reaches a beacon interval, **then:**
  update Table I;
  for $r \in Road = \{North, South, East, West\}$, **do**
    $r\_in = 0$;
    $r\_out = 0$;
  end for
end if
restart the timer

Algorithm 3 V2V Forwarding Strategy

Input: **vd**
Output: Relay
  if *vs* and *vd* are on the same road, **then:**
    TD is *vd*;
  else
    if the vehicle is *vs*, **then:**
    TD is the source intersection;
    else *TD* is I_j^{t+1} in algorithm 2;
    end if
  end if
  if *TD* is in the neighbor table, **then:**
    forward the packet to *TD*;
  else choose a relay the closest to *TD* in neighbor table;
  end if
  if step 11 fails, **then:**
    store the packet to buffer queue of the vehicle;
  else forward the packet to relay;
  end if
  while the buffer queue is not empty, **do:**
    check buffer and neighbor table periodically;
    if a neighbor can be used as the relay, **then:**
      forward the packet to relay;
    end if
end while

the position of intersections. The greedy forwarding strategy is implemented on path sections between two contiguous intersections. We employ the carry-and-forward strategy to recover from the local optima [38]. To achieve the goal of transmitting packets from the source vehicle to the destination vehicle, we introduce two algorithms: the V2V forwarding strategy in road areas and the V2I forwarding strategy in intersection areas.

1) V2V Forwarding Strategy: Vehicles on roads adopt the V2V forwarding strategy to select relays and transmit packets. The source vehicle *vs* generates data packets and determines the position of the destination vehicle. If the vehicle carrying the packets and the destination vehicle are on the same road, the target destination *TD* is the destination vehicle *vd* and the vehicle forwards packets to the destination vehicle by a greedy strategy directly. Otherwise, packets are forwarded on the path to the source intersection first. When the packets are on paths between the source intersection and the destination intersection, the *TD* is the next intersection determined at each intersection by Algorithm 2. The greedy forwarding strategy is executed on road segments between intersections to forward packets to the target destination. The vehicle node chooses the neighbor closest to the target intersection or target vehicle as the next hop to relay packets. The local optimum problem occurs when there is no available neighbor node within the communication range or no neighbor node is closer to the *TD* than the node itself. The vehicle node stores the packet to buffer the queue and keeps moving until it meets an appropriate relay and forwards the packet to the next hop. The V2V forwarding strategy is shown in Algorithm 3.

2) V2I Forwarding Strategy: Once data packets arrive at intersection areas via V2V communication on paths, the RSUs at intersections forward them to the optimal road segment selected by Algorithm 2 and choose a reliable relay from the road. The RSU determines routing path r and the next intersection I_j^{t+1} adjacent to the road segment based on the selected action a_r^t in Algorithm 2. Then, the RSU chooses a neighbor vehicle as the relay on path r. If the current intersection I_i^t is the destination intersection D, the RSU will forward the packet to a node closest to the destination vehicle. When available neighbor vehicles are not observed in the intersection area, the RSU carries the packet until the appropriate relay appears. The V2I forwarding strategy is shown in Algorithm 4.

Algorithm 4 V2I Forwarding Strategy

Input: vd, D, a_r^t
Output: Relay
if I_i^t is D, **then:**
  TD is the destination vehicle *vd*;
else *TD* is I_j^{t+1};
end if
if *vd* is in neighbor table, **then:**
  forward packet to *vd*;
else choose a relay the closest to *TD* in neighbor table;
if step 7 fails, **then:**
  store the packet to buffer queue of RSU;
else forward the packet to relay;
execute step 17 to step 22 in **Algorithm 3**

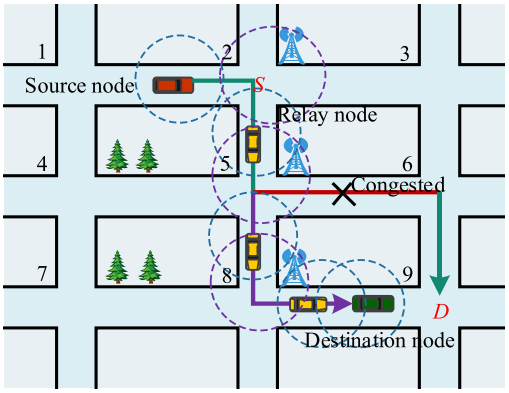


Fig. 4. Example of IV2XQ.

D. Example of IV2XQ

Fig. 4 shows an example of the IV2XQ routing protocol. The red vehicle between intersections 1 and 2 is the source vehicle and generates data packets. The green vehicle between intersections 8 and 9 is the destination vehicle. Intersection 2 is closer to the destination vehicle than intersection 1; therefore, it is chosen as the source intersection *S*. The destination vehicle is moving towards intersection 9 and has a shorter distance to 9; thus, intersection 9 is chosen as the destination intersection *D*. The source vehicle first forwards packets to the RSU at source intersection *S*. The RSU determines the current state of (2, 9) and finds road segment 2-5 with the maximum Q-value in the Q-table. Packets are forwarded to intersection 5 through relay vehicles on the road, such as the yellow vehicles in the figure. The RSU at intersection 5 finds that road segment 5-6 with the maximum value in the current state is congested; therefore, it selects 5-8 as an alternative path. When packets reach intersection 8, the RSU forwards them to destination intersection *D*. If the destination vehicle is in the neighbor table, then the packets are forwarded to the destination vehicle directly and the routing is successful.

VI. Simulation Results and Analysis

A. Simulation Environment and Parameters

We conduct extensive simulations using the tools SUMO, Veins and OMNeT++. SUMO is a traffic simulator that integrates vehicular driving patterns, vehicular driving behavior, path settings, etc.; therefore, realistic traffic scenarios can be generated and used in other external programs. The Veins framework is responsible for establishing the protocols of the underlying model and node mobility, ensuring the correct execution of the simulation, and collecting experimental results. OMNeT++ is an object-oriented discrete event network simulator that can simulate wireless network communication. Under the framework of Veins, OMNeT++ and SUMO can have intersections with real-time information. Our simulation scenario is an urban area covering 3 km × 3 km, and it consists of 38 two-way road segments and 24 intersections. The vehicle density varies by road. The key parameters used in the simulation are shown in Table II.

We obtain the Q-table for routing by training the Q-learning algorithm. The learning rate α of the Q-learning is 0.1, the probability of the ε - greedy strategy is 0.2, and the reward $\emptyset$ is set to 5 in our algorithm. To assess the convergence performance of the Q-learning, we show the relationship between the cumulative reward and the iterations in Fig. 5. As the training iterations increase, the cumulative reward increases. When the Q-learning is convergent, the cumulative reward remains stable. Fig. 5 illustrates that the proposed algorithm is effective and the server has low computation cost as the Q-learning converges at a fast speed.

The optimal vehicular density is a parameter used to update the Q-value function. The action of selecting the road segment with vehicular density close to the optimal value in section IV has larger expected reward. Fig. 6 shows the result of the algorithm using different vehicular density as the parameter, where the distance between vehicles is from 50 m to 150 m, and the vehicular speed varies from 40 km/h to 60 km/h. From the figure, the routing with the vehicular interval of 100 m has the best performance irrespective of the traffic model. Therefore, we set vehicular interval as 100m in the following simulations.

TABLE II
Parameters Used in Our Simulations

Parameters	Values
Simulation area	3 km × 3 km
Number of intersections	24
Number of segments	38
Vehicular speed	14 m/s
Vehicular density	0.005 ~ 0.02 vehicles/m
Number of vehicles	450
Signal transmission radius of a vehicle	250 ~ 300 m
Signal transmission radius of an RSU	300 m
Packet size	512 bytes
Packet sending rate	1 ~ 6 packets/sec.
Beacon interval	1 s
Simulation duration	1000 s

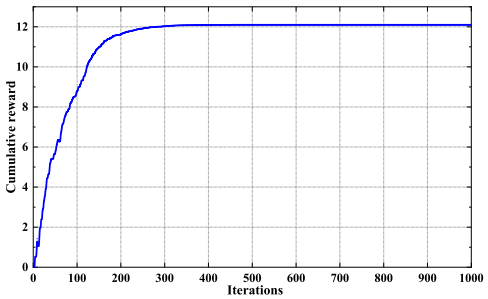


Fig. 5. Cumulative reward with iterations.

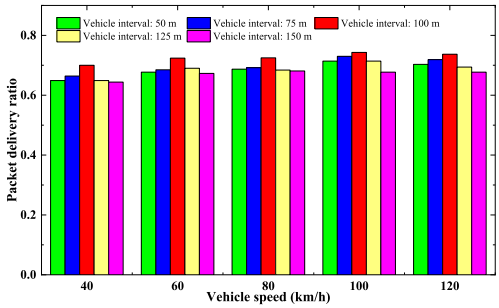


Fig. 6. Optimal vehicular density of different vehicular speeds.

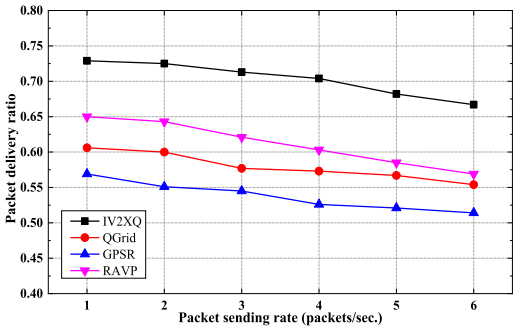


Fig. 7. Packet delivery ratio with the packet sending rate.

We compare IV2XQ with GPSR from [39], QGrid from [27] and RAVP from [40] and analyze the simulation results. The following metrics are used to evaluate the performance of the routing protocols.

- Packet delivery ratio: The ratio of the total number of packets successfully received by destination vehicles to the total number of packets sent by source vehicles within the simulation time.
- Average delay: The average end-to-end time that a packet takes to be forwarded from the source node to the destination node through the routing path.
- Average hop count: The average number of nodes that a packet passes through on the routing path from the source node to destination node.
- Overhead ratio: The ratio of the difference between the total number of packets generated by source vehicles and the total number of packets successfully received by destination vehicles to the total number of packets successfully received by destination vehicles as shown in Formula (14).

$$Overhead\ ratio = \frac{total\_Num_{generated} - total\_Num_{received}}{total\_Num_{received}} \tag{14}$$

B. Simulation Results and Analysis

1) Packet Delivery Ratio: We analyze the relationships between the packet delivery ratio and the packet sending rate and signal transmission radius. The results are shown in Fig. 7 and Fig. 8. The packet delivery ratio of IV2XQ is higher than that of the compared algorithms. IV2XQ selects the optimal routing path and implements the improved greedy strategy on road segments, while GPSR and QGrid implement the greedy strategy on the entire area or the divided grid area, where intersections and buildings may cause signal attenuation and packet loss. In addition, QGrid always chooses the grid with the maximum vehicle density as the next grid. When there are too many vehicles in an area, signal collisions may lead to dropped packets. IV2XQ forwards packets to road segments with the optimal vehicle density for packet transmission; therefore, IV2XQ has the best network performance. RAVP predicts vehicles' positions and broadcasts request packets to discover the routing path before sending a data packet. This approach ensures the success rate of routing, although the method of discovering and maintaining routes is inherently inefficient.

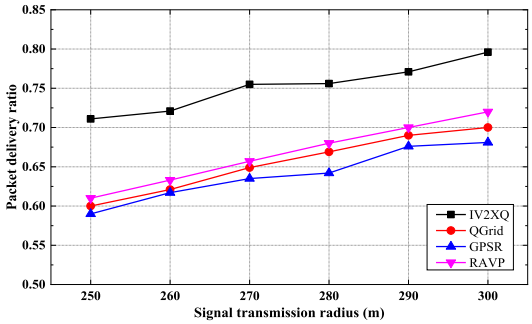


Fig. 8. Packet delivery ratio with the signal transmission radius.

Fig. 7 shows the packet delivery ratio with different packet sending rates. As the packet sending rate increases, the number of packets generated in a certain time increases. As the simulation time passes, the increasing number of packets makes the network heavily loaded, and the vehicles in the network are close to full communication capacity. Therefore, the packet delivery ratio degrades gradually and steadily. In Fig. 8, the improvement in the signal transmission radius increases the packet delivery ratio, and vehicles have a higher possibility of finding appropriate relays over a wider range of communications.

2) Average Delay: RAVP and GPSR have higher average delays than IV2XQ and QGrid. The process of discovering and updating routes before forwarding data packets increases the communication time of RAVP. Furthermore, the local optimum problem of GPSR may make a relay node unreachable for a long time and cause a large end-to-end delay. With the routing path planned in advance, IV2XQ and QGrid achieve low delays. The routing of IV2XQ is restricted to roads, where the greedy strategy works better without barriers; thus, IV2XQ has the lowest delay.

As shown in Fig. 9, as the packet sending rate increases, the average delay increases. The number of packets increases rapidly at a high packet sending rate; therefore, the network can more easily become congested and the waiting time of a packet in the buffer queue is longer. When the packet sending rate increases beyond a certain value, the average delay of the compared algorithms increases sharply; however, IV2XQ is not significantly affected due to the introduction of the congestion mechanism. When the average length of vehicle buffer queues on a road segment reaches the threshold value,

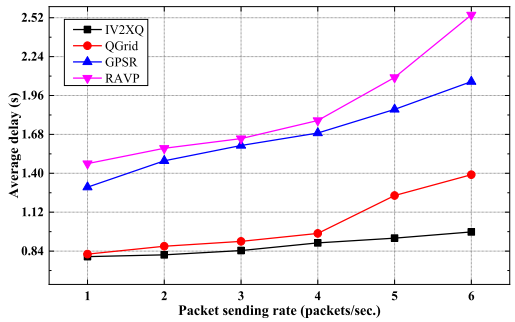


Fig. 9. Average delay with the packet sending rate.

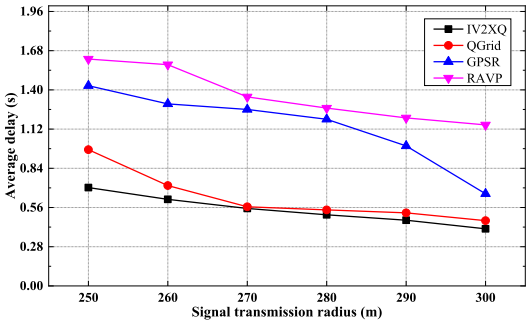


Fig. 10. Average delay with the signal transmission radius.

the alternate road segment will be used to balance the network load.

In Fig. 10, as the signal transmission radius increases, the average delay decreases. There are more candidate relays in the neighbor table when the communication range is longer. The greedy forwarding strategy selects the neighbor node closest to the destination vehicle. When the packet passes through a relay node, it consumes the processing time of the physical, MAC and application layers and the waiting time in the buffer queue, which is much longer than the signal transmission time. The increased signal transmission range allows the destination node to be reached with fewer hops, thereby greatly decreasing the delay. The downward trend is particularly obvious for GPSR.

3) Average Hop Count: RAVP has the highest hop count, and the packets pass through more hops in the IV2XQ protocol than in GPSR and QGrid. The simulation result is logical. Each time GPSR and QGrid select the next hop, they forward the packets to the neighbor nodes closest to the destination vehicle. QGrid chooses the optimal grid areas for packet transmission based on GPSR. Therefore, QGrid has the fewest hops and GPSR has the second fewest hops. Although IV2XQ also adopts a greedy strategy, packets must be relayed by RSUs at intersections, thereby increasing the hop count. RAVP broadcasts request packets to find a route from the source node to the destination node, the request packets record nodes passed during the routing discovery process, and data packets are forwarded to the recorded nodes to reach the destination. However, the request packets may be retransmitted between two nodes repeatedly when there are no other available nodes; therefore, the average hop count is large.

Fig. 11 shows the relationship between the average hop count and packet sending rate. As the packet sending rate

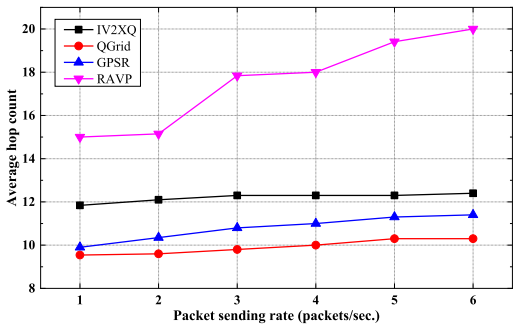


Fig. 11. Average hop count with the packet sending rate.

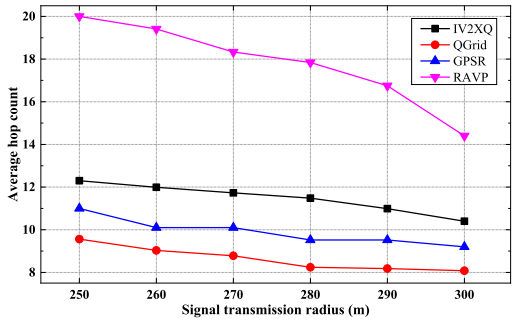


Fig. 12. Average hop count with the signal transmission rate.

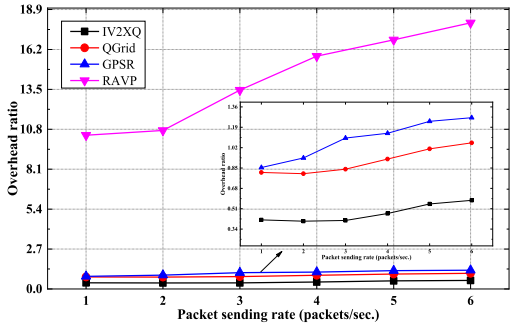


Fig. 13. Overhead ratio with the packet sending rate.

increases, the average hop counts of GPSR and QGrid increase slightly, and that of IV2XQ remains stable. An increasing trend is observed for the average hop count of RAVP because as the number of packets increases, retransmission caused by broadcasting also increases. In Fig. 12, a negative correlation is observed between the average hop count and the signal transmission radius: when the distance between the source node and destination node is unchanged, a longer signal transmission radius results in fewer relay hops.

4) Overhead Ratio: As shown in Fig. 13 and Fig. 14, IV2XQ has the lowest overhead ratio and RAVP has a much higher communication overhead ratio than the other three routing protocols. The overhead ratio is defined as the number of data packets failing transmission and other packets used to discover and maintain the route to the total number of packets in the network. IV2XQ has the highest packet delivery ratio and does not require additional packets to establish routes. Therefore, IV2XQ is the most efficient routing protocol and

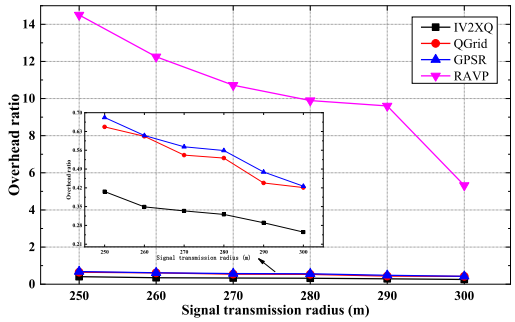


Fig. 14. Overhead ratio with the signal transmission rate.

has extremely low communication overhead. Similarly, GPSR and QGrid also have low routing overhead because they do not require the establishment and maintenance of a routing table. The routing process of RAVP consists of route discovery and route maintenance, which both consume a large number of extra packets. In the RAVP protocol, each vehicle stores a two-dimensional table that preserves the information of all the nodes in the network. The size of the table depends on the number of vehicle nodes, and the overhead ratio increases as the simulation scene expands.

Fig. 13 illustrates the relationship between the overhead ratio and packet sending rate: the overhead ratio decreases with increasing packet sending rate. The main reason for this result is that the more packets there are to be transmitted in the network, the more routes there are to be established and maintained. As the packet sending rate increases, the packet delivery ratio decreases, and the number of discarded packets increases. Dropped packets occupy some network resources in the transmission process before being dropped but do not achieve the goal of conveying useful information, which increases the communication overhead. In Fig. 14, the overhead ratio decreases as the signal transmission radius increases. Again, a major reason is the increase in the packet delivery ratio.

VII. Conclusion

In this paper, we proposed the intersection-based V2X routing algorithm via Q-learning (IV2XQ) based on historical traffic flow and real-time network conditions. The routing table is replaced by a multidimensional Q-table; therefore, the routing overhead decreases greatly without the process of route discovery and maintenance. RSUs at intersections choose the optimal routing path and forward packets. Although the introduction of RSUs increases the hop count, it prevents packet loss caused by signal shadow fading and vehicular direction changes. V2X communication is implemented via an improved greedy forwarding strategy. The hierarchical routing protocol improves the packet delivery ratio and decreases end-to-end latency. The use of alternate routing paths balances the network load and controls network congestion. The simulation results present the advantages of our algorithm in terms of network performance indicators, such as the packet delivery ratio, average delay and overhead ratio.

With the development of 5G technology, spectrum resources are becoming increasingly precious, and wired networks are expensive. Optimizing the utilization of network resources in the IoV has attracted considerable attention. Our future work will focus on the management and scheduling of multidimensional resources for intelligent routing in VANETs.

References

[1] D. Zhang, F. Richard Yu, Z. Wei, and A. Boukerche, "Trust-based secure routing in software-defined vehicular ad hoc networks," 2016, *arXiv:1611.04012*. [Online]. Available: http://arxiv.org/abs/1611.04012

[2] B. Goswami and S. Asadollahi, "Novel approach to improvise congestion control over vehicular ad hoc networks (VANET)," in *Proc. Int. Conf. Comput. Sustain. Global Develop.*, 2016, pp. 3567–3571.

[3] G. Sun, Y. Zhang, D. Liao, H. Yu, X. Du, and M. Guizani, "Bus-trajectory-based street-centric routing for message delivery in urban vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7550–7563, Aug. 2018.

[4] Y. R. B. Al-Mayouf, N. F. Abdullah, M. Ismail, S. M. Al-Qaraawi, O. A. Mahdi, and S. Khan, "Evaluation of efficient vehicular ad hoc networks based on a maximum distance routing algorithm," *EURASIP J. Wireless Commun. Netw.*, vol. 2016, no. 1, pp. 1–11, Dec. 2016.

[5] T. Darwish and K. A. Bakar, "Traffic aware routing in vehicular ad hoc networks: Characteristics and challenges," *Telecommun. Syst.*, vol. 61, no. 3, pp. 489–513, Mar. 2016.

[6] G. Sun, L. Song, H. Yu, X. Du, and M. Guizani, "A two-tier collection and processing scheme for fog-based mobile crowdsensing in the Internet of vehicles," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1971–1984, Feb. 2021.

[7] A. Srivastava, A. Prakash, and R. Tripathi, "Location based routing protocols in VANET: Issues and existing solutions," *Veh. Commun.*, vol. 23, no. 100231, pp. 1–30, 2020.

[8] T. Wang, Y. Cao, Y. Zhou, and P. Li, "A survey on geographic routing protocols in delay/disruption tolerant networks," *Int. J. Distrib. Sensor Netw.*, vol. 12, no. 2, pp. 1–12, 2016.

[9] J. Liu, J. Wan, Q. Wang, P. Deng, K. Zhou, and Y. Qiao, "A survey on position-based routing for vehicular ad hoc networks," *Telecommun. Syst.*, vol. 62, no. 1, pp. 15–30, May 2016.

[10] J. Li, P. Wang, and C. Wang, "Comprehensive GPSR routing in VANET communications with adaptive beacon interval," in *Proc. Int. Conf. Internet Things (iThings)*, Dec. 2016, pp. 1–6.

[11] R. Karimi and S. Shokrollahi, "PGRP: Predictive geographic routing protocol for VANETs," *Comput. Netw.*, vol. 141, pp. 67–81, Aug. 2018.

[12] *Vehicular Communications; GeoNetworking; Part 4: Geographical Addressing and Forwarding for Point-to-Point and Point-to-Multipoint Communications; Sub—Part 2: Media-Dependent Functionalities for ITS-G5*, document REN/ITS-0030035, ETSI TS, Intelligent Transport Systems (ITS); 2020.

[13] D. Das, "Distributed algorithm for geographic opportunistic routing in VANETs at road intersection," in *Proc. Int. Conf. Dependable, Autonomic Secure Comput.*, 2017, pp. 1202–1209.

[14] Z. Li, Y. Song, and J. Bi, "CADD: Connectivity-aware data dissemination using node forwarding capability estimation in partially connected VANETs," *Wireless Netw.*, vol. 25, no. 1, pp. 379–398, Jan. 2019.

[15] T. S. J. Darwish, K. A. Bakar, and K. Haseeb, "Reliable intersection-based traffic aware routing protocol for urban areas vehicular ad hoc networks," *IEEE Intell. Transp. Syst. Mag.*, vol. 10, no. 1, pp. 60–73, Jan. 2018.

[16] Y. R. B. Al-Mayouf *et al.*, "Real-time intersection-based segment aware routing algorithm for urban vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 7, pp. 2125–2141, Jul. 2018.

[17] G. Sun, Y. Zhang, H. Yu, X. Du, and M. Guizani, "Intersection fog-based distributed routing for V2 V communication in urban vehicular ad hoc networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 6, pp. 2409–2426, Jun. 2020.

[18] D. Tian *et al.*, "A microbial inspired routing protocol for VANETs," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2293–2303, Aug. 2018.

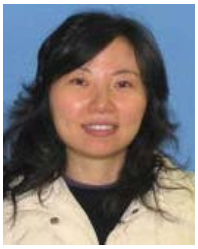
[19] F. Goudarzi, H. Asgari, and H. S. Al-Raweshidy, "Traffic-aware VANET routing for city environments—A protocol based on ant colony optimization," *IEEE Syst. J.*, vol. 13, no. 1, pp. 571–581, Mar. 2018.

[20] M. A. Mujahid, K. A. Bakar, T. S. J. Darwish, and F. T. Zuhra, "Cluster-based location service schemes in VANETs: Current state, challenges and future directions," *Telecommun. Syst.*, vol. 30, pp. 1–9, Oct. 2020.

[21] L. Yao, J. Wang, X. Wang, A. Chen, and Y. Wang, "V2X routing in a VANET based on the hidden Markov model," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 889–899, Mar. 2018.
[22] Y. Tang, N. Cheng, W. Wu, M. Wang, Y. Dai, and X. Shen, "Delay-minimization routing for heterogeneous VANETs with machine learning based mobility prediction," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3967–3979, Apr. 2019.
[23] L. Zhao, Y. Li, C. Meng, C. Gong, and X. Tang, "A SVM based routing scheme in VANETs," in *Proc. 16th Int. Symp. Commun. Inf. Technol. (ISCIT)*, Sep. 2016, pp. 380–383.
[24] P. Mars, *Learning Algorithms: Theory and Applications in Signal Processing, Control and Communications*. Boca Raton, FL, USA: CRC Press, 2018, pp. 285–290.
[25] C. Wu, S. Ohzahata, and T. Kato, "Flexible, portable, and practicable solution for routing in VANETs: A fuzzy constraint Q-Learning approach," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4251–4263, Nov. 2013.
[26] D. Zhang, T. Zhang, and X. Liu, "Novel self-adaptive routing service algorithm for application in VANET," *Int. J. Speech Technol.*, vol. 49, no. 5, pp. 1866–1879, May 2019.
[27] F. Li, X. Song, H. Chen, X. Li, and Y. Wang, "Hierarchical routing for vehicular ad hoc networks via reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 68, no. 2, pp. 1852–1865, Feb. 2019.
[28] S. Zemouri, S. Djahel, and J. Murphy, "An altruistic prediction-based congestion control for strict beaconing requirements in urban VANETs," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 12, pp. 2582–2597, Dec. 2019.
[29] G. Sun, L. Song, H. Yu, V. Chang, X. Du, and M. Guizani, "V2 V routing in a VANET based on the autoregressive integrated moving average model," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 908–922, Jan. 2019.
[30] Y. Gao, Z. Zhang, D. Zhao, Y. Zhang, and T. Luo, "A hierarchical routing scheme with load balancing in software defined vehicular ad hoc networks," *IEEE Access*, vol. 6, pp. 73774–73785, 2018.
[31] H. Gao, C. Liu, Y. Li, and X. Yang, "V2VR: Reliable hybrid-network-oriented V2V data transmission and routing considering RSUs and connectivity probability," *IEEE Trans. Intell. Transp. Syst.*, early access, Apr. 13, 2020, doi: 10.1109/TITS.2020.2983835.
[32] F. Khan and S. K. Nguang, "Location-based data delivery between vehicles and infrastructure," *IET Intell. Transp. Syst.*, vol. 14, no. 5, pp. 288–296, 2020.
[33] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.
[34] J. R. Wu and P. Yuan, "Statistics analysis and optimization design for intersection traffic volume on urban road," *Adv. Mater. Res.*, vol. 243–249, pp. 4422–4425, May 2011.
[35] Z. Cao, K. Shi, Q. Song, and J. Wang, "Analysis of correlation between vehicle density and network congestion in VANETs," in *Proc. 7th IEEE Int. Conf. Electron. Inf. Emergency Commun. (ICEIEC)*, Jul. 2017, pp. 409–412.
[36] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *J. Artif. Intell. Res.*, vol. 4, no. 1, pp. 237–285, Jan. 1996.
[37] R. S. Sutton and A. G. Barto, *Introduction to Reinforcement Learning*. Cambridge, MA, USA: MIT Press, 1998.
[38] S. M. Bilal, S. A. Madani, and I. A. Khan, "Enhanced junction selection mechanism for routing protocol in VANETs," *Int. Arab J. Inf. Technol.*, vol. 8, no. 4, pp. 422–429, 2011.
[39] B. Karp and H. T. Kung, "GPSR: Greedy perimeter stateless routing for wireless networks," in *Proc. Annu. Int. Conf. Mobile Comput. Netw.*, 2000, pp. 243–254.
[40] L.-L. Wang, J.-S. Gui, X.-H. Deng, F. Zeng, and Z. Kuang, "Routing algorithm based on vehicle position analysis for Internet of vehicle," *IEEE Internet Things J.*, vol. 7, no. 12, pp. 11701–11712, Dec. 2020.

Long Luo (Member, IEEE) received the B.S. degree in communication engineering from the Xi'an University of Technology in 2012 and the M.S. and Ph.D. degrees in communication engineering from the University of Electronic Science and Technology of China (UESTC) in 2015 and 2020, respectively. She is currently a Post-Doctoral Researcher with UESTC. Her research interests include networking and distributed systems.

Li Sheng is currently pursuing the master's degree in communication and information system with the University of Electronic Science and Technology of China. Her research interests include VANET routing and algorithms.



Hongfang Yu (Member, IEEE) received the B.S. degree in electrical engineering from Xidian University in 1996 and the M.S. and Ph.D. degrees in communication and information engineering from the University of Electronic Science and Technology of China in 1999 and 2006, respectively. From 2009 to 2010, she was a Visiting Scholar with the Department of Computer Science and Engineering, University at Buffalo (SUNY). Her research interests include network survivability, network security, and next generation internet.

Gang Sun (Member, IEEE) is currently a Professor of computer science with the University of Electronic Science and Technology of China (UESTC). His research interests include vehicular ad hoc networks, cloud computing, parallel and distributed systems, ubiquitous/pervasive computing and intelligence, and cyber security.