An Intersection-Based QoS Routing for Vehicular Ad Hoc Networks With Reinforcement Learning

Lanlan Rui, *Member, IEEE*, Zhibo Yan, Zuoyan Tan, Zhipeng Gao, *Member, IEEE*, Yang Yang, *Member, IEEE*, Xingyu Chen, *Member, IEEE*, and Huiyong Liu, *Member, IEEE*

***Abstract*— Vehicular ad hoc networks (VANETs) have the characteristics of high mobility, frequently changing topology and uneven distribution, which made it a challenge to design an efficient and robust routing protocol with low latency and high packet delivery rate. Currently, intersection-based routing method and full-path based routing method are two popular solutions for the packet routing in VANETs. Although the intersection-based routing method has better real-time performance, it has the problem of local optimization, making the routing results not global optimal. Although the full-path based method can obtain the global optimal solution, it is weak in dealing with the dynamically changing topological network. Aiming at solving the above problems, this paper designs an intersection-based QoS routing (IQRRL) algorithm, which mainly includes two crutial steps: the next intersection selection and the next hop vehicle selection. In the selection of the next intersection, this paper uses an improved intersection-based routing protocol. In addition to considering connectivity and delay, IQRRL also considers the communication quality from the neighbor's road to the destination node while evaluating the quality of the neighbor's road, which minimizes the problem of local optimization. When the next intersection is determined, a road is then determined, and then the next hop vehicle within the road should be chosen to relay the packet forward. In the next-hop vehicle selection step, this paper adopts multi-hop evaluation technology based on reinforcement learning. In addition to using "greedy decision-making" to select the next-hop vehicle, it also comprehensively evaluates whether the next-hop vehicle is still optimal in the future, so that the stability and reliability of data forwarding are improved and local optimal problems are avoided. Besides, this article uses a simulation system to compare IQRRL with other routing algorithms. The result reveals that IQRRL outperforms in terms of packet delivery ratio and transmission delay.**

***Index Terms*— Quality of service, routing protocol, reinforcement learning, vehicular ad hoc networks.**

I. Introduction

AS A KEY technology in intelligent transportation systems (ITSs), vehicular ad hoc networks (VANETs) have received considerable critical attention, and they are expected to play a vital role in the smart cities [1]. Information can be exchanged via vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) [2] wireless communications in VANETs. For V2V communication, vehicle nodes act not only a communication node but also a router to communicate with each other without the help of the infrastructure. One goal of the VANET is to support the traffic safety and infotainment applications. An efficient routing algorithm can deliver packets in multi-hops data transmission with high reliability and low delay for these applications.

Currently, MANET [3] is another famous ad hoc network, which can be deployed quickly and conveniently. However, MANET can suffer from limited network bandwidth and weak performance in real-time communication. Although MANETs and VANETs both have mobile devices and changing network topologies, the mobility of the devices the changing frequency in VANETs is much higher than those in MANETs. As a result, compared to MANETs, VANETs have the characteristics of high mobility, frequently changing topology, uneven distribution, road restriction, signal block, and limited wireless resources in the urban environment. The above characteristics may cause frequent broken links and short link maintenance between vehicles, which can result in high delays and low packet delivery rates due to queuing in the buffer. Also, in a scenario where the vehicles move fast, the destination node and the intermediate node are likely to deviate from the original road and cause the previously selected path to fail. Therefore, we have proposed an Intersection-Based QoS Routing (IQRRL) in VANET routing, to solve the problems of dynamically changing network, high mobility and uneven distribution.

The first key process of IQRRL routing is the selection of the next intersection. Besides the consideration of street connectivity, delay and the distance away from the destination node, this process also focuses on the reduction of the local optimal defects in terms of the intersection-based routing method. Specifically, the state of the neighboring road and the state of the entire shortest path from the neighbor intersection to the destination node are considered by using Dijkstra algorithm. The whole process of the intersection selection can be simplified as follows: Three QoS metrics, street connection probability, street transmission delay, and the distance from the neighbor intersection to the destination node, are used to score the neighbor intersection, and the

Manuscript received 16 June 2021; revised 17 August 2022 and 27 November 2022; accepted 24 April 2023. Date of publication 9 May 2023; date of current version 30 August 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFB1807800. The Associate Editor for this article was S. Sacone. *(Corresponding authors: Lanlan Rui; Huiyong Liu.)*

The authors are with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: llrui@bupt.edu.cn; yanzhibo_sdclone@bupt.edu.cn; 541629067@qq.com; gaozhipeng@bupt.edu.cn; yyang@bupt.edu.cn; chenxy@bupt.edu.cn; hyliu@bupt.edu.cn).

Digital Object Identifier 10.1109/TITS.2023.3271456

intersection node with the highest score is as chosen the next intersection.

The second key process of IQRRL routing is the selection of the next hop vehicle within the road. After an intersection node (or a road) is determined in the first process, the next hop vehicle should be chosen so that the packet can be relayed forward within the determined road segment. Currently, "Greedy forwarding" method is popular in reducing the number of hops, but it may encounter a high packet loss ratio because of high mobility [6]. In recent researches, reinforcement learning [15] is used to solve the frequent changing topology and the choice of the optimal path. Upon applying such method, the agents can sense the rapid change of topology and obtain the optimal actions by periodic exploration of the VANET environment. As a result, this process of IQRRL routing is a combination of "Greedy forwarding" and Q-Learning [16], a form of reinforcement learning. More specifically, HELLO messages are used as the control packets to learn the environment, and each vehicle updates its Q-Table when they receive HELLO packets. Then the system will choose the vehicle with the max Q value. Hence, the second process of IQRRL comprehensively evaluates whether the next-hop vehicle is still optimal in the future, so that the stability and reliability of data forwarding are improved and local optimal problems can be avoided.

It is noted that different types of services in the Internet of Vehicles put forward diversified service quality requirements, including reliability, packet transmission rate, delay, throughput, etc. There are two main types of business in the Internet of Vehicles, which are security service and entertainment service. Security services have higher requirements on latency, while entertainment services are less sensitive to latency than security services, but pay more attention to data transmission efficiency and throughput. In addition, because entertainment services are not sensitive to delays and often require information beyond a wider range, messages are usually sent from the source node to the destination node in a unicast and multi-hop manner and it is very important to ensure the QoS of various services to successfully send the message from the source node to the destination node in this process. This paper mainly focuses on the entertainment services in the Internet of Vehicles, and proposes the IQRRL algorithm to optimize the transmission and improve the QoS of the service.

In summary, the main contributions of this paper are as follows:

1) In order to solve the problem of message transmission and routing in VANET environment, an intersection-based QoS routing protocol (IQRRL) is proposed, which can solve the problems of the routing algorithms based only on intersection or based only on full path. IQRRL routing consists of the selection of the intersection and the selection of the next-hop within the selected road, both of which ensure the road performance while avoiding the disadvantage that the package is getting farther away from the destination node.
2) In the first process of the IQRRL routing (selection of the next intersection), IQRRL considers the basic indicators such as delay and connection probability. Besides, IQRRL not only considers the communication quality of the neighboring road, but also considers the communication quality of the path from the neighboring road to the destination node by using Dijkstra algorithm, so as to avoid the problem that the next intersection is getting farther and farther from the end point due to the local optimization.
3) In the second stop of IQRRL routing (the selection of the next hop vehicle), we have proposed using reinforcement learning to establish a network between vehicle nodes within the road. With the self learning of each node, the Q values from the current node to the neighboring node can be achieved and the system will choose the node that has the max Q value and is closest to the next intersection. In this way, whether the next hop is still optimal in the future is considered, which greatly increases the stability and reliability during packet transmission. Also, it can avoid the problem of local optimum caused by mere greedy choosing.

The remainder of this paper is organized as following. Section II describes related works. Section III discusses assumptions, metrics and system model. Section IV illustrates the intersection-based QoS routing algorithm with reinforcement learning in detai. Section V introduces complexity analysis. Section VI gives the performance evaluation of IQRRL compared with other routing algorithms.

II. Related Work

Currently, there are some existing VANET routing protocols, and they are discussed below.

A. Geographic Routing Protocol

Geographic routing protocols are considered to be the most suitable protocols to deliver packets in VANET [21]. Because they use the geographic location of the vehicle nodes to send packets from the source node to the destination node, which can adapt to the mobility of VANET. GPSR [4] protocol is the well-known geographic routing protocol, which consists of greedy forwarding and perimeter forwarding. Greedy forwarding method is to find the one-hop neighbor node which is closest to the destination location, and then forward the packet to this node. And the perimeter forwarding method is a supplement to greedy forwarding mode, which solves the dilemma of data grouping forwarding in the area where greedy forwarding fails. In paper [33], PA-GPSR, an improved GPSR-based algorithm, uses an additional extension table in neighbors' table to select the best path. However, due to the feature of "greedy search", GPSR may suffer from "routing hole" caused by local optimum. Although GPSR has provided a "right hand law" for the problem, the cost of this solution is high and thus the system adaptivity and fault tolerance can be quite low.

B. Traffic Aware Routing Protocol

Considering the problem of GPSR that is discussed above, a lot of research has begun to focus on the traffic and network states that are regarded as the essential factor in VANET routing protocol. This new type of routing protocol is called as the traffic aware routing (TAR) protocol [8]. The traffic aware routing protocol's core idea is to consider both the vehicles positions and traffic status. There are basically two types of TAR protocol: intersection-based routing protocol and full-path based routing protocol, which are introduced as follows.

1) Intersection-Base Routing Protocol: Intersection-based routing protocol is one category of the TAR protocol [8]. In recent papers, intersection-based connectivity aware geocast routing (IB-CAGR) [34] designs a connectivity-aware geocast routing protocol which analyses the impact of traffic signal on connectivity among vehicular nodes. Though it improves the packet delivery ratio to some extent, the problem of local minimum problem still exists. GyTAR [22] considers two parameters to choose intersection: traffic density of roads and distance to destination. And the intersection that is the shortest distance and highest density toward the destination will be chosen. However, the GyTAR may easily encounter the local maximum problem and its parameters cannot reflect the network status very well. IGRP [5] chooses the intersection based on the connectivity while satisfying quality-of-service (QoS) constraints. The QoS is constrained by delay, hop count and bit error rate (BER). In order to reduce path's sensitivity, the geographic forwarding is used between two selected intersection. However, in order to optimize problem, IGRP uses the genetic algorithm whose convergence speed and computation complexity need to be discussed further. Like IGRP, AQRV [14] also selects the route while satisfying the QoS constraints. Delay, packet delivery ratio, connectivity are the QoS metrics. But unlike IGRP, AQRV solves the optimization problem using ant colony optimization (ACO) algorithm. However, AQRV adopts greedy carry-and-forward within the selected road segment. RTISAR [23] presents new formulas and considers the segment's density, load, connectivity and the distance toward the destination when selecting the intersection. In order to solve the frequency of the measurement process, RTISAR proposes a verity period mechanism which can be calculated for the collector packet. But it also does not solve the problem of local optimum. In general, the common advantage of intersection-based routing protocols is that they can adapt to real-time changing vehicle networks and have a small packet overhead. But the frequent measurement process of the road segment may result in high network overhead and delay.

2) Full Path Routing Protocol: Full path routing protocol is another category of the TAR protocol [8]. ACAR [9] addresses the multi-hop communication and rapidly changing topology problems by selecting a route based on the transmission quality and the network connectivity. However, it cannot be adjusted according to the network status. EGSR [24], an enhanced version of the geographical source routing, uses a control packet to update the connectivity of the road segment, and then exploits the Dijkstra algorithm to compute the shortest path and adds the list into the packets. However, EGSR only considers connectivity and cannot fully reflect the network status. QFRG [25] designs a road weight estimation (REW) to get the information of the entire network topology and divides the total network into some small grid zones. And then, QFGR selects the best path by considering quality of forwarding (QoF) and link reliability. Although full-path routing solves the problem of frequent measurement process in the intersection-based routing, they also have their limitations: 1) the routing path is static, so it cannot be adjusted according to the traffic and network status. 2) and the list of routing path added into the packets introduces high overhead.

C. Routing Protocol Based on Reinforcement Learning

Reinforcement learning, as one of the paradigms and methodologies of machine learning, is used to solve the problem that agents learn strategies to maximize returns or achieve specific goals in the process of interacting with the environment. A common model of reinforcement learning is the standard Markov Decision Process (MDP). MDP is a cyclic process in which an agent takes actions to change its state, obtain rewards and interact with the Environment.

In the second step, selection of the next hop vehicle, of the algorithm in this paper, IQRRL combines "greedy learning" and reinforcement learning to evaluate a vehicle node not only by channel quality, but also by the possibility that it may still be the optimal node in the future. Recently, many researchers are using reinforcement learning (RL) to deal with the dynamic VANET routing problem. Q-learning is a specific type of reinforcement learning, which maps the state of the external input environment to output action through continuous learning, hoping to make the corresponding reward value the maximum. Paper [42] uses Q-learning to complete the adjustment of fuzzy rule strategy in unknown environment. Paper [44] uses Basic Q-Learning Reinforcement Learning agents who will learn market patterns to trade in financial assets to maximize the fund value, using portfolio returns net of transaction costs as the learning criteria. Paper [43] proposes Multi-Q-Table Q-learning to address a problem of low average sum of rewards and the proposed method constructs a new Q-table whenever a sub-goal is reached. PFQ-AODV [26] combines fuzzy constraint and Q-learning to learn the optimal routing path in VANET. (Fuzzy constraints conclude link quality, bandwidth, and vehicle movement). And they use route quest (RREQ) message to discover the route. Celimuge Wu et al. [19] propose a vehicle-to-roadside routing protocol. They use game theory to guide the vehicle to choose the cluster header to relay packets, and then uses RL to select the optimal route toward the roadside. However, cluster head selection is only based-on the partitioning of the network. ARPRL [17] is also a heuristic routing protocol based on the RL for VANET. ARPRL continuously learns in the form of Q-table updated by the Q-learning algorithm and actively acquires new network link states with periodic HELLO packets.

III. System Model And Routing-Related Metrics

In this section, we introduce the system model and routing-related metrics.

TABLE I
COMPARISON OF DIFFERENT INTERSECTION BASED PROTOCOLS AND FULL-PATH BASED PROTOCOLS

Protocols	Metrics	Relaying strategy (within selected segment)	Types	Traffic aware	Common problems
TA-GPSR [33] (2019)	Distance	Greedy forwarding based on distance	GPSR-based routing protocol	No	suffer from "routing hole" and low adaptivity and fault tolerance
IB-CAGR [35] (2016)	Distance, density	Improved intersection-based TAR protocol	Intersection-based routing protocol	Yes	Problem of local optimum still exists
AQRV [14] (2017)	Delay, packet delivery ratio, connectivity	Greedy carry-and-forward			
RTISAR [23] (2018)	Density, load, connectivity, distance	Improved greedy carry-and-forward mechanism			
EGSR [24] (2019)	Connectivity	The minimum total weight for the complete route	Full path routing protocol	Yes	Only considers connectivity and cannot fully reflect the network status
QFRG [2] (2019)	Transmission cost, packet delivery ratio, link reliability	The best route based on the road weight evaluation scheme			
IQRRL (proposed)	Connectivity probability, delay, distance	The best next hop based on the Q-learning algorithm	Intersection-based routing protocol	Yes	Try to avoid local maximum and be aware of road conditions

A. System Model

We assume that each vehicle is equipped with a digital map, a Global Position System (GPS), and a navigation system [14], [25], [27] in the urban scenario. Based on these assumptions, vehicles can get the length of road segments and the positions of intersections. And some location services also enable vehicles to obtain their locations. Moreover, static node [28] at the intersection is used to help relay data.

In the process of selection of the next intersection of IQRRL, there are some assumptions before the connectivity probability of the road segment can be estimated. Firstly, we assume there are U platoon and K ordinary vehicles in the road segment, and the platoon members in each platoon can be connected with each other, and each platoon is a single vehicle. So, there are N = U + K single vehicles in the road segment, as shown in Fig. 1. It's assumed that the distance between two consecutive vehicles is X, and Xi (i = 1, 2, ..., N-1) denotes the intervehicle distance between two consecutive vehicles. Secondly, we suppose the vehicles on the road follow the Poisson distribution. Therefore, distance between vehicles follows exponential distribution. And the traffic density is λ (vehicles per meter). Thirdly, it's assumed that the vehicle velocity is constant.

In our system, we suppose that vehicles are moving on the two-way lanes whose directions are opposite. From two perspectives, the entire VANET network can be abstracted into two graphs, as shown in Fig. 2. On one hand, we use $G(I, L)$ to represent the street graph, where I represents the intersection and L represents the road segment. A road segment is usually connected by two intersections. We define the backbone route from the source vehicle V_s to the destination vehicle Vd consisting of a sequence of roads segment $\{L_1, L_2, \ldots, L_k\}$. In order to reduce the route search time, we first find the terminal source intersection (TSI) and the terminal destination intersection (TDI). Note that the road segment from V_s to TSI is L_1 and the road segment from TDI to V_d is L_k. On the other hand, a transmission graph is formed between vehicles within a road segment. We define that the $G'(V, E)$ represents the transmission graph, where V is the vehicle and E is the transmission link between two vehicles. All the key notations used in this article are summarized in Table II.

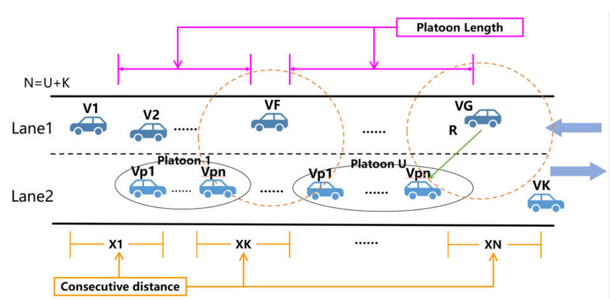


Fig. 1. Two-way VANET scenario.

B. Routing-Related Metrics

1) Connectivity Probability in the Two-Way Lanes: Based on the above assumptions, the probability that we can find k

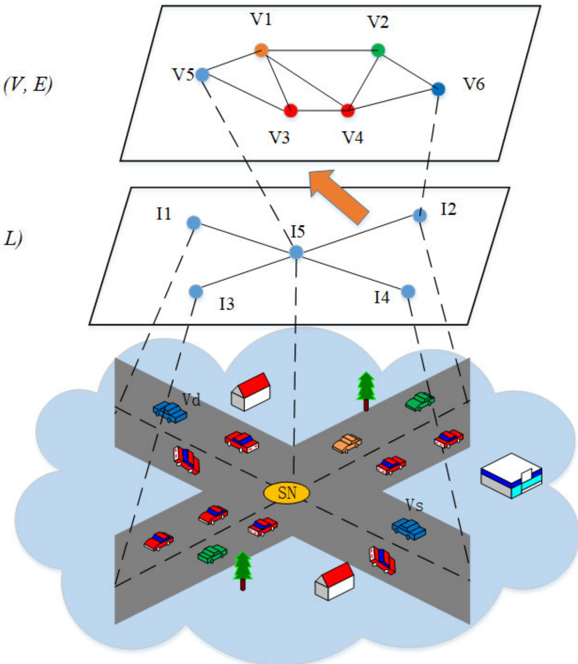


Fig. 2. System model.

TABLE II

NOTATIONS USED IN THE PAPER

Notation	Description
$G(I, L)$	The street graph
$G'(V, E)$	The transmission graph
TSI	The terminal source intersection
TDI	The terminal destination intersection
TS	The temporary source intersection
TD	The temporary destination intersection
CP	The connectivity probability
D	The delay
DE	The distance from the next intersection to the destination
$LScore(L_{mn}, V_d)$	The total score toward the destination vehicle through the road segment $L_{m,n}$
I_{wait}	The candidate intersection
$LRTP$	The learning request packet
$LRYP$	The learning reply packet
N	The number of single vehicles in the road segment
P_{whole}	The connectivity possibility of the road segment
D_{whole}	The delay of the road segment
$Q_c(d, nr)$	The Q-value from the current node c through its neighbor node nr to the destination node d

vehicles in a distance of x is

$$P(K = k) = \frac{(\lambda x)^k e^{-\lambda x}}{k!} \tag{1}$$

And the probability that the distance between two consecutive single vehicles is small than x is given by

$$P(X \le x) = 1 - e^{-\lambda x} \tag{2}$$

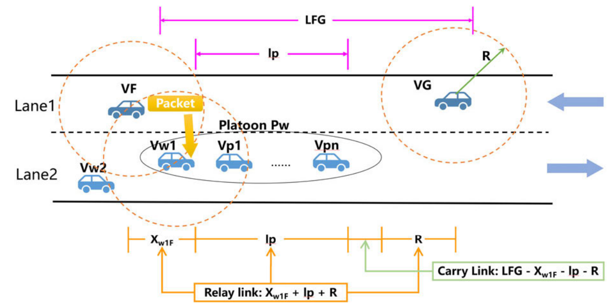


Fig. 3. The first case of modeling the connectivity.

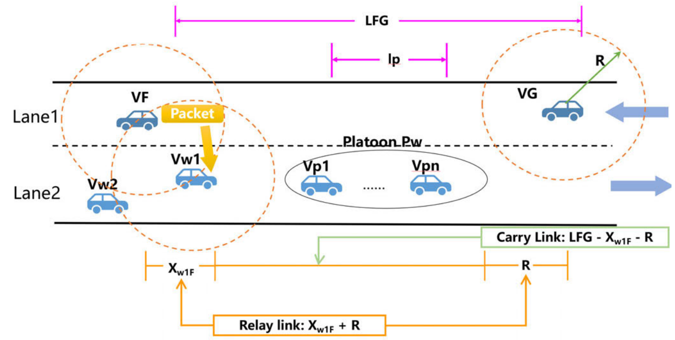


Fig. 4. The second case of modeling the connectivity.

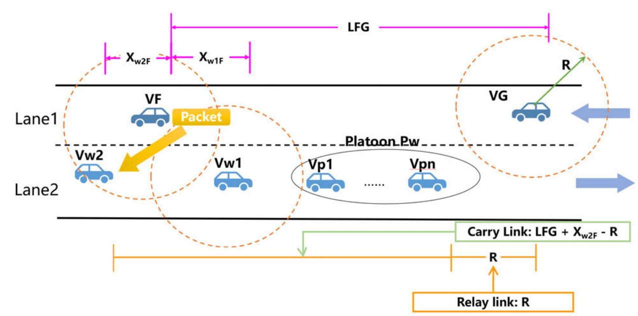


Fig. 5. The third case of modeling the connectivity.

If the distance between two consecutive single vehicles is larger than R (transmission range) in the one-way road, it's defined as a broken link. The probability of the broken link is given by

$$P_{bl} = P(X > R) = 1 - P(X \le R) = e^{-\lambda R} \tag{3}$$

There is a broken link between vehicle V_F and vehicle V_G in lane 2, as shown in Fig. 5. We use Z to denote the number of broken links on a one-way road. The probability that Z links are disconnected in *N-1* satisfies the binomial distribution, which is given by

$$P_Z(z) = \binom{N-1}{z} \cdot P_{bl}^z \cdot (1 - P_{bl})^{N-1-z} \tag{4}$$

However, in two-way road, the broken link can be repaired by the vehicles on the opposite lane. And the act of reconnecting the broken link with the help of the opposite vehicles is called repair behavior.

In the following part, the probability of link repair on a one-way lane if first described, and then the connectivity possibility of total two-way road segment will be discussed.

a) Repair probability of the broken link: As shown in Fig. 1, packets can be forwarded between vehicle V_F and vehicle V_G if the vehicle V_F and V_{p1}, vehicle V_{p1} and V_{pn}, vehicle V_{pn} and V_G are all connectable to each other. Therefore, the possibility that the broken link is repaired is given by [29]

$$P_{repair} = P_{F,p1} \cdot P_{p1,pn} \cdot P_{pn,G} \tag{5}$$

where

$$\begin{cases} P_{F,p1} = P_{pn,G} = P(X \le R) = 1 - e^{-\lambda R} \\ P_{p1,pn} = \int_R^{\infty} \sum_{k=1}^{M} \dfrac{(\lambda x)^k \cdot e^{-\lambda x} \cdot P_{kc}}{k! \int_R^{\infty} y^k \cdot e^{-\lambda y} dy} dx \end{cases} \tag{6}$$

where $P_{\mathrm{F},p1}$ and $P_{pn,G}$ represent the probability connection between vehicle V_F and vehicle V_{p1}, vehicle V_{pn} and vehicle V_G, respectively. $P_{p1,pn}$ is the probability that vehicle V_{p1} is connected with vehicle V_{pn}. It is also the possibility of the platoon one connection. M is the number of connected vehicles in platoon one. When $M = 1, P_{p1,pn} = \lambda$. P_{kc} in $P_{p1,pn}$ is expressed as

$$P_{kc} = (1 - e^{-\lambda R})^{k-1} \tag{7}$$

Among it, we assume that there are k vehicles between the vehicle V_{p1} and vehicle V_{pn}. We use P_{kc} to represent the probability that the k vehicles are connected with each other.

b) Connectivity possibility of the whole road segment: The road is connected only when all the broken links are connected. The possibility of z broken link connections is given by

$$P_{zc} = P_{repair}^{z}, \quad z = 0, 1, 2, \ldots, N-1 \tag{8}$$

Therefore, the connectivity possibility of whole two-way road segment is

$$P_{whole} = \sum_{z=0}^{N-1} P_{zc} \cdot P_z(z) = \sum_{z=0}^{N-1} P_{repair}^{z} \cdot \binom{N-1}{z} \cdot P_{bl}^{z} \cdot (1 - P_{bl})^{N-1-z} \tag{9}$$

When there is no broken link (z = 0), the road is fully connected. The possibility of full connection is

$$P_{whole} = \sum_{z=0}^{0} (1 - P_{bl})^{N-1} = (1 - e^{-\lambda R})^{N-1} \tag{10}$$

It is consistent with the theoretical analysis.

2) Delay in the Two-Way Lanes: In this part, like the above analysis of connectivity probability, we first analyze the repair time of the broken link between two vehicles in the one-way lane and then analyze the transmission delay of the whole road segment under the help of the opposite vehicles.

a) Repair time of the broken link: In order to more accurately analyze the repair time, we divide the broken link between vehicle V_F and vehicle V_G into three cases. As shown in Fig. 2, X_{w1F} and X_{w1p1} are the distance between vehicle V_{w1} and vehicle V_F, vehicle V_{w1} and vehicle V_{p1}, respectively. P_w represents the platoon close to vehicle V_{w1}, the whose length is l_p. To simplify the calculation, we set l_p and L_{FG} as an average value, which are given by [30].

$$\begin{cases} l_p = (e^{\lambda R} - 1) \cdot (\dfrac{1}{\lambda} - \dfrac{R \cdot e^{-\lambda R}}{1 - e^{-\lambda R}}) \\ L_{FG} = R + \dfrac{1}{\lambda} \end{cases} \tag{11}$$

Therefore, the first case is $X_{w1F} <$ R and $X_{w1p1} < R$. The two case is $X_{w1F} <$ R and $X_{w1p1} > R$. The third case is only the $X_{w1F} > R$. The reason is that in the third case, the package cannot be forwarded to the vehicle V_{w1} first and can only be sent to another vehicle within the scope of the vehicle V_F, such as the vehicle V_{w2}. The repair time in these three cases is:

i) The first case: As shown in Fig. 3, in this case, vehicle V_{w1} becomes the member of the platoon P_w because of $X_{w1p1} < R$. And the packet can be relayed by vehicle V_{w1} and platoon P_w. So the relay link is $L_{r1} = X_{w1F} + l_p$+R, the carry link is $L_{c1}(X_{w1F}) = L_{FG} - L_{r1} = L_{FG} - X_{w1F} - l_p -$R. If the $L_{c1} \le 0$, the link between V_F and V_G is fully connected. The relay time compared with repairing time between V_F and V_G is negligible according to [13]. However, if the $L_{c1} > 0$, vehicle V_{pn} has to carry the packet until it reaches the $V_G^{\prime}s$ transmission range. Since the vehicle V_{pn} and the vehicle V_G are moving in opposite directions, the vehicle V_{Pn} and the vehicle V_G go together to complete the carrying distance.

Based on the above analysis, the probability of this case is

$$P_{case1} = P(X_{w_1F} < R) \cdot P(X_{w_1p_1} < R) = (1 - e^{-\lambda R})^2 \tag{12}$$

And the delay in this case between V_F and V_G is given as [13]

$$\begin{aligned} D_{case1} &= P_{case1} \cdot \int_0^R \frac{L_{c1}(x) \cdot \lambda e^{-\lambda R}}{(v_{pn} + v_G) \cdot (1 - e^{-\lambda R})} dx \\ &= \frac{1 - e^{-\lambda R}}{(v_{pn} + v_G)} \cdot \int_0^R L_{c1}(x) \cdot \lambda e^{-\lambda x} dx \\ &= \frac{1 - e^{-\lambda R}}{(v_{pn} + v_G)} \cdot \int_0^R (L_{FG} - x - l_p - R) \cdot \lambda e^{-\lambda x} dx \\ &= \frac{1 - e^{-\lambda R}}{(v_{pn} + v_G)} \cdot \{(e^{-\lambda R} - 1) \cdot l_p + R \cdot e^{-\lambda R}\} \\ &= \frac{1 - e^{-\lambda R}}{(v_{pn} + v_G)} \cdot \{(e^{-\lambda R} - 1) \cdot (e^{\lambda R} - 1) \\ &\quad \cdot (\frac{1}{\lambda} - \frac{R \cdot e^{-\lambda R}}{1 - e^{-\lambda R}}) + R \cdot e^{-\lambda R}\} \\ &= \frac{1 - e^{-\lambda R}}{(v_{pn} + v_G)} \cdot \{(e^{\lambda R} - 1) \\ &\quad \cdot (\frac{e^{-\lambda R} - 1}{\lambda} + R \cdot e^{-\lambda R}) + R \cdot e^{-\lambda R}\} \end{aligned} \tag{13}$$

ii) The second case: As shown in Fig. 4, vehicle V_{w1} is not a member of the platoon P_w because of $X_{w1p1} > R$. Therefore, the packet has to be carried by vehicle V_{w1} until it reaches the $V_G^{\prime}s$ transmission range. And the relay link is $L_{r2} = X_{w1F}$+R, the carry link is $L_{c2}(X_{w1F}) = L_{FG} - L_{r2} = L_{FG} - X_{w1F} -$R. As in the first case, the vehicle V_{w1} and the vehicle V_{G} go together to complete the carrying distance.

Based on the above analysis, the probability of this case is

$$P_{case2} = P(X_{w_1F} < R) \cdot P(X_{w_1p_1} > R) = (1 - e^{-\lambda R}) \cdot e^{-\lambda R} \tag{14}$$

And the delay in this case between V_F and V_G is given as

$$\begin{aligned} D_{case2} &= P_{case2} \cdot \int_0^R \frac{L_{c2}(x) \cdot \lambda e^{-\lambda x}}{(v_{pn} + v_G) \cdot (1 - e^{-\lambda R})} dx \\ &= \frac{e^{-\lambda R}}{(v_{pn} + v_G)} \cdot \int_0^R L_{c2}(x) \cdot \lambda \cdot e^{-\lambda x} dx \\ &= \frac{e^{-\lambda R}}{(v_{pn} + v_G)} \cdot \int_0^R (L_{FG} - x - R) \cdot \lambda e^{-\lambda x} dx \\ &= \frac{e^{-\lambda R}}{(v_{pn} + v_G)} \cdot \{(1 - e^{-\lambda R}) \cdot (L_{FG} - R - \frac{1}{\lambda}) \\ &\quad + R \cdot e^{-\lambda R}\} \\ &= \frac{e^{-2\lambda R}}{(v_{pn} + v_G)} \cdot R \end{aligned} \tag{15}$$

iii) The third case: As shown in Fig. 5, vehicle V_{w1} is out of the transmission range of vehicle V_F. Therefore, the packet need to relayed by another vehicle in $V_F's$ transmission range. Since vehicle V_{W2} and vehicle V_F are moving in the opposite direction, the vehicle V_{w2} can reach the $V_F's$ transmission range, and then V_{w2} carry the packet until it reaches the $V_G's$ transmission range. Therefore, the carry link is $L_{c3}(X_{w2F}) = L_{FG} + X_{w2F}$−R. As in the first case, the vehicle V_{w2} and the vehicle V_G go together to complete the carrying distance.

Based on the above analysis, the probability of this case is

$$P_{case3} = P(X_{w_1F} > R) = 1 - P(X_{w_1F} < R) = e^{-\lambda R} \tag{16}$$

And the delay in this case between V_F and V_G is given as

$$\begin{aligned} D_{case3} &= P_{case3} \cdot \int_R^{+\infty} \frac{L_{c3}(x) \cdot \lambda e^{-\lambda x}}{(v_{pn} + v_G)} dx \\ &= \frac{e^{-\lambda R}}{(v_{pn} + v_G)} \cdot \int_R^{+\infty} L_{c3}(x) \cdot \lambda e^{-\lambda x} dx \\ &= \frac{e^{-\lambda R}}{(v_{pn} + v_G)} \cdot \int_R^{+\infty} (L_{FG} + x - R) \cdot \lambda e^{-\lambda x} dx \\ &= \frac{(e^{-\lambda R})^2}{(v_{pn} + v_G)} \cdot (L_{FG} + \frac{1}{\lambda}) = \frac{e^{-2\lambda R}}{(v_{pn} + v_G)} \cdot R \end{aligned} \tag{17}$$

Based on the above analysis, the repair time of the broken link between vehicle V_F and vehicle V_G is

$$D_{repair} = D_{case1} + D_{case2} + D_{case3} \tag{18}$$

b) Delay of the whole road segment: As with the analysis of connectivity probability, assume there are Z broken links on a one-way lane. The time it takes for the packet to pass through the whole segment is

$$D_Z(z) = \sum_Z D_{repair} \tag{19}$$

Therefore, the delay of whole two-way road segment is

$$D_{whole} = \frac{L}{l_p + L_{FG}} \cdot D_{repair} \tag{20}$$

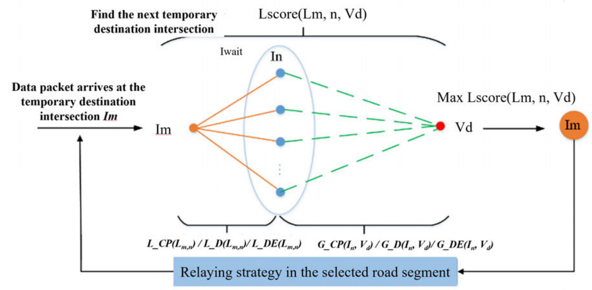


Fig. 6. The score counting process of the next intersection.

According to (19), the road length L, the transmission distance R, vehicle velocity v, and vehicle density λ ultimately determine the delay of the road segment.

IV. Intersection-Based QoS Routing Algorithm With Reinforcement Learning

IQRRL is an intersection-based geographic routing with the help of reinforcement learning algorithm to find the optimal route. QoS models are used to evaluate the next intersection. And the QoS is measured in terms of the connectivity probability, the transmission delay and the distance from the neighbor intersection to the destination node. IQRRL is mainly composed of two parts: the selection of the next intersection and the selection of a next hop within the road.

Due to the highly dynamic mobility of vehicles and the rapidly changing topology, the links between vehicles are easily disconnected. Therefore, it is very difficult to obtain instant information of all vehicles and maintain the effectiveness of the selected optimal route in the network [1], [27]. In order to address the above issue, we adopt real-time traffic-aware approach. Our routing protocol mainly includes two aspects: the choice of the next intersection in G(I, L) and the selection of the next hop within the road of $G'(V, E)$. The next intersections are chosen dynamically one by one according to the segment status. After the selection of the next intersecion, Q-learning algorithm is used to estimate the long-term outcome of the next hop's decision and then the next hop is chosen within the road segment according to the Q-value.

A. Selection of the Next Intersection

In order to find the next intersection to form the optimal route path, we consider three QoS metrics of the connectivity probability (CP), delay (D) and distance to the destination (DE) to limit the choice of the next intersection. When the package arrives at an intersection, the vehicle can obtain a set of candidate intersections, which we define as I_{wait}. In order to avoid loops, the intersections already recorded in the packet header should be deleted from the candidate intersections. Note that one (except TSI) of the two intersections where the source node is located should be marked as the passed intersection. Then, vehicle scores each candidate intersection, which is shown in Fig. 6. The score expression is given as

$$\begin{aligned} \mathrm{LScore}(\mathrm{L_{m,n}}, \mathrm{V_d}) &= \delta_1 \cdot \mathrm{CP}(\mathrm{L_{m,n}}, \mathrm{V_d}) + \delta_2 \cdot \frac{1}{(1+\mathrm{D}(\mathrm{L_{m,n}}, \mathrm{V_d}))} \\ &\quad + \delta_3 \cdot \frac{1000}{\mathrm{DE}(\mathrm{L_{m,n}}, \mathrm{V_d})} \end{aligned} \tag{21}$$

where

$$\begin{cases} \mathrm{CP(L_{m,n},V_d) = L\_CP(L_{m,n}) + G\_CP(I_n,V_d)} \\ \mathrm{D(L_{m,n},V_d) = L\_D(L_{m,n}) + G\_D(I_n,V_d)} \\ \mathrm{DE(L_{m,n},V_d) = L\_DE(L_{m,n}) + G\_DE(I_n,V_d)} \end{cases} \tag{22}$$

where LScore (L_{mn}, V_d) is the total score toward the destination vehicle through the road segment $L_{m,n}$. CP $(L_{m,n}, V_d)$, D$(L_{m,n}, V_d)$ and $DE\left(L_{m,n}, V_d\right)$ denote the connectivity probability, delay and distance toward the destination vehicle through the road segment $L_{m,n}$, respectively. Note that the intersection I_n represents the neighbor intersections of the intersection I_m. δ_1, δ_2 and δ_3 are the weight parameters. $L\_CP\left(L_{m,n}\right)$, $L\_D\left(L_{m,n}\right)$, and $L\_DE\left(L_{m,n}\right)$ stand for the local connectivity probability, delay and distance of the road segment $L_{m,n}$, respectively. $G\_CP(I_n, V_d)$, $G\_D(I_n, V_d)$ and $G\_DE(I_n, V_d)$ are the global connectivity probability, delay and distance toward the destination vehicle through the intersection I_n of $L_{m,n}$. We use the Dijkstra's algorithm to get the shortest path from I_n to V_d. The connectivity probability, delay and distance of this shortest path are computed as follows:

$$\begin{cases} G\_CP(I_n, V_d) = \prod_{i=1}^{t} G\_CP(L_{ij}) \\ G\_D(I_n, V_d) = \sum_{i=1}^{t} G\_D(L_{ij}) \\ G\_DE(I_n, V_d) = \sum_{i=1}^{t} G\_DE(L_{ij}) \end{cases} \tag{23}$$

where t is the number of road segments on the shortest path, $G\_CP(L_{ij})$, $G\_D(L_{ij})$ and $G\_DE\left(L_{ij}\right)$ are the connectivity probability, delay and distance of the road segment L_{ij}.Since the distance is 1000 times the delay in magnitude, we multiply it by 1000 before the distance to increase the effect of the distance variable.

Therefore, the one with the higher score among the candidate intersections will become the next temporary destination intersection (TD), and be stored in the packet. Note that before building the optimal route between the source node and the destination node, we first need to choose the TSI and the TDI to determine the corresponding communication pairs. In order to shorten the forwarding distance, the two intersections of the road where the source / destination node is located, which is closer to the destination / source node is the TSI / TDI. TSI and TDI also need to be stored in the data packet. So, in the beginning, TSI is the temporary destination intersection. In addition, we need to update the TDI regularly, because the destination node may depart from the original road in the process of packet forwarding.

B. Selection of the Next Hop Within One Road Segment

After the next intersection is determined, a specific road segment is determined, and then the next-hop vehicle within the road segment should be chosen to relay the packet forward.

In this process, Q-learning model, a specific reinforcement learning model, is used to evaluate the future reward of the next hop decision.

1) The Q-Learning Model: The Q-learning model is shown in Table III. As Table III shows, the environment of the Q-learning model is the total vehicle ad hoc network. The agent is each packet in VANET. All the network nodes form a state space. And agent's action is to select a one-hop neighbor as its next hop.

TABLE III
THE BASIC COMPONENTS OF Q-LEARNING MODEL

Environment	VANET
Agent	Each packet
State of the agent	Each node
State space	All the network nodes
Actions	The selection of the next hop from one-hop neighbors

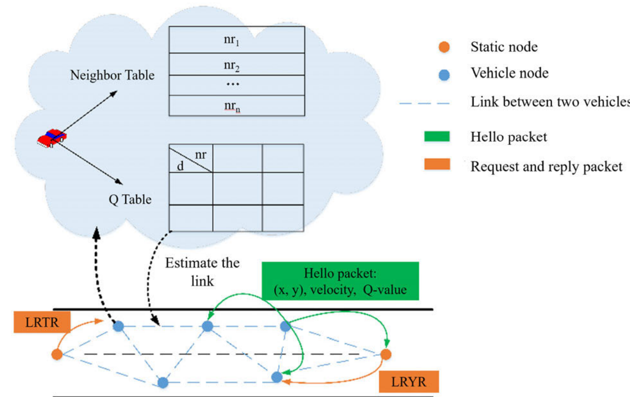


Fig. 7. Update of the Q-Table when receiving Hello message.

2) Update of Q-Table and Relaying Strategy: In the road segment, each vehicle node maintains a two-dimensional Q-Table, as shown in Fig. 7. Each row in the Q-Table represents the destination node, and each column represents the neighbor node. When vehicles receive Hello messages, they not only update neighbor tables, but also update Q-Tables. The initialization value of each Q-value is 0. Vehicle nodes broadcast the Q-value via Hello messages. When node c receives a hello message from its neighbor node nr, node c updates the Q-Table according to the learning function as follows

$$Q_c(d, nr) \leftarrow (1-\alpha) \cdot Q_c(d, nr) + \alpha \cdot [\text{Reward} + \gamma \cdot \max_{y \in N_n} Q_n(d, y)] \tag{24}$$

where Reward is

$$Reward = \begin{cases} 1, & if\ nr = d \\ 0, & otherwise \end{cases} \tag{25}$$

where $Q_c\left(d, nr\right)$ is the Q-value from the current node c through its neighbor node nr to the destination node d. α controls the learning rate. It can neither be too large nor too small [17]. If α is too large, agent considers only the most recent information which can result in the misleading of the packet forwarding. But if α is too small, the convergence speed will be slow. Similarly, the discount value γ should not be too large or too small. Too small value will lead to immediate rewards dominated, and too large value will lead to over reliance on long-term rewards. And $max_{y \in N_n} Q_n\left(d, y\right)$ is the maximal Q-value of node nr to destination node d through node y.

After such a Q-Table update process, a microscopic transmission graph within a road segment is established. But in order to speed up the Q-Learning convergence, before sending a packet, temporary source intersection broadcasts a routing request packet (LRTP) until the packet reaches the temporary destination intersection. Then, temporary destination intersection unicasts a route reply packet (LRYP) until arriving at the temporary source intersection. After that, the node can select its one-hop neighboring nodes with the largest Q-value to forward the packet.

C. Detailed Routing Algorithm and A Routing Example

The IQRRL routing algorithm steps are shown in the following two tables.

Algorithm 1 IQRRL Routing Algorithm

```
Input: source vehicle V_s, destination vehicle the data V_d, packet dp
if V_s and V_d are in the same road then
    Find the best next hop to Vd by Algorithm 2
    return
end if
Select TSI and TDT.
set TSI as the first temporary destination intersection (TD)
set V_s as the first temporary source intersection (TS)
while dp does not reach Vd do
    while dp does not reach the TD do
        Find the best next hop to TD by Algorithm 2
        Record the number of vehicles and the total speed from TS to TD
    end while
    Update the road's number of vehicles and the average speed according to the above records by using on-the-fly method [9].
    Update TDI if V_d departs from the original road.
    Set TS = TD
    Find the best next intersection as the new TD from I_wait
end while
```

An example of routing decision is shown in Fig. 8 to better explain IQRRL in Algorithm 1 table above. The source node and the destination node obtain I_6 and I_7 as TSI and TDI, respectively. At first, TSI is considered as the temporary destination intersection. When the packet reaches I_6 through $L_{6,vs}$,we can get $I_{wait} = \{I_3, I_5, I_7, I_9\}$. After counting their scores, I_7 is selected as the next intersection. By repeating the above process, a backbone route path $\{L_{VS,6}, L_{6,7}, L_{7,Vd}\}$ from V_s to V_d is obtained. It is worth noting that every time we determine the next temporary destination intersection, we can choose the best next hop within the road segment based on the results of reinforcement learning, which is talked in the section VI. For example, the source vehicle will pass the packet to I_6 through N_1 and N_2. What's more, in the case of network partition, vehicle has to adopt the carry-and-forward strategy.

Algorithm 2 Next-Hop Vechicle Selection within One Road Segment

```
Input: Temporary source intersection (TS), temporary destination intersection (TD), the data packet dp.
When receiving hello messages from one-hop neighbor nodes, vehicles update their Q-Tables according to (23).
TS broadcasts a LRTP, terminal nodes rebroadcast LRTP and updates their Q-Tables according to (23) until TD receives LRTP.
TD unicasts a LRYP, terminal nodes rebroadcast LRTP and update their Q-Tables according to (23) until TS receives LRYP.
while dp does not reach TD do
    Vehicle finds the next hop from its neighbor node with the largest Q-value, and send dp to the next hop.
end while
```

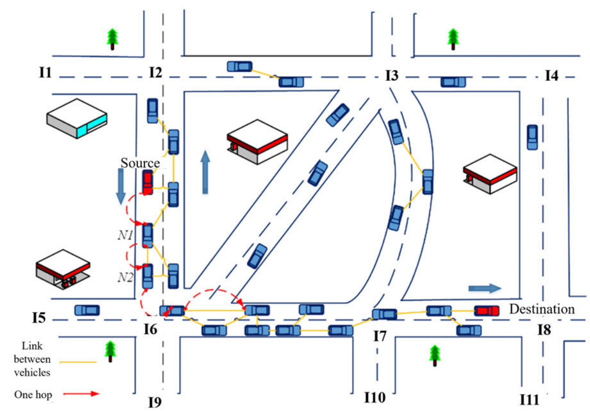


Fig. 8. An example of a routing process.

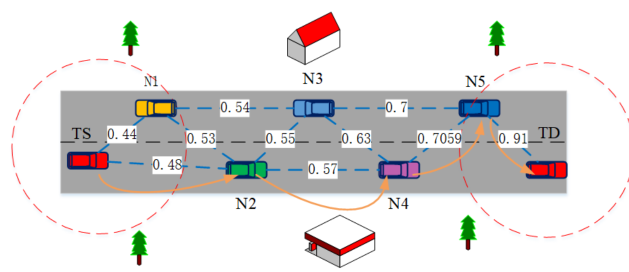


Fig. 9. An example of route in the micrograph.

In order to more clearly describe the packet relay strategy within the road segment, or the selection of the next-hop vehicle within the road, in Algorithm 2 table, we present an example of forwarding packet in the micrograph. As shown in Fig. 9, a microtopography is formed between vehicles after vehicles update their Q-Table. TS broadcasts an LRTP packet firstly, and the intermediate nodes forward it until TD receives LRTP. Then, TD unicasts an LRYP packet until it is received by the TS. In this way, the vehicles on the road have updated the Q-value again. The weights on the link edge represent Q-value between the current node and its one-hop neighbor node. As shown in Fig. 9, when TS wants to send a packet to TD, it only needs to find the next hop from its neighbor node with the largest Q-value. Therefore, the optimal path in Fig. 9 is TS− > N2− > N4− > N5− > TD.

D. Routing Maintenance

IQRRL routing algorithm performs well when facing high mobility and frequently changing network states. However, in the first step of the algorithm (the selection of the next intersection), if the destination vehicle happens to move out of the current road during the packet transmitting process, the sender vehicle needs to reselect the TDI. Every vehicle has two tables: the neighbor table and Q-Table. Every time the vehicle receives a hello message, it updates the neighbor table and Q-Table at the same time, and deletes outdated information periodically.

E. Complexity Analysis

In order to select the optimal route path, there are two steps. One step is to choose the optimal intersection. The time complexity of this step is O(I·L). I is the number of intersections in the digital map, L is the number of the edge. Although we only need to consider the neighboring roads, the number of neighboring roads of the temporary intersection will reach L in the worst case. The second step is to choose the optimal next hop within the road. The time complexity of this progress is O(V). V is the number of vehicle nodes in the whole network, which is much larger than the number of vehicles in the selected road in actual. Therefore, the total complexity of our algorithm is O(I· L·V).

F. Summary

Due to the high mobility of the vehicular network, the topology changes very quickly. Therefore, IQRRL in this paper adopts the real-time approach and defines two main processes in packet routing, which are the selection of the next intersection and the selection of the next-hop vehicle within the road segment. In the first process, the intersection selection, the optimal intersection is selected by using above the process of “optimal route establishment”. In the second process, the next-hop vehicle selection to relay the packet within the road segment, we consider the two nodes at both ends of the road segment as TS and TD. The TS first broadcasts a learning request packet (LRTP), and the nodes that receive the LRTP on the way rebroadcast it and update their Q-values until the TD is reached. Finally, the TD unicasts the learning reply packet (LRYP). And the nodes on the path update the Q-values again. After the above process, the initial routing path in the micrograph is established.

V. Performance Evaluation

In this section, we implement our proposed IQRRL protocol in NS-2 in its version ns-2.35. To analyze the performance of IQRRL, we compare it with other three protocols, namely ARPRL [17], GyTAR [22], and SRPMT [27]. ARPRL is a heuristic routing protocol, which chooses the next hop according to the Q-Table. GyTAR is an intersection-based geographical protocol, which selects the intersection according to the distance to the destination and the traffic density. Inside the road, GyTAR adopts greedy carry-and-forward. SRPMT is a street-centric routing protocol, which considers the delay when choosing the next intersection.

(a) Selected area from openstreetmap (b) Edited streets map

Fig. 10. Digital maps of los angeles.

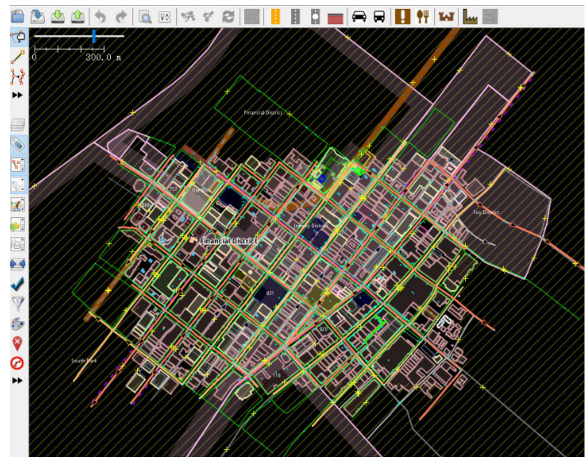


Fig. 11. An example of using JOSM to edit the osm file.

A. Experimental Environment

We first load a street map, an osm file, from the OpenStreetMap.[a] And then we use the JOSM,[a] a Java OpenStreetMap editor, to delete some non-primary edges and nodes to make the street map clearer. Fig. 10(a) is the selected area in Los Angeles from OpenStreetMap, and Fig. 10(b) is the edited street map. Fig. 11 is an example of using JOSM to edit the osm file and Fig. 12 is an example of using VanetMobisim to generate the moving file. The mobility model we used is Intelligent Driver Model with IntersectionManagement (IDM_IM). The constant bit rate (CBR) is used to generate the traffic data of the source node. The source node and destination node are selected randomly. The other simulation parameters are indicated in Table IV. Each simulation is repeated 50 times to reduce the effect of the random seed, and the confidence interval is set as 95%.

B. Metrics to Evaluate The Performance

We considers three metrics to evaluate our proposed routing protocol.

[a] *OpenStreetMap:https://www.openstreetmap.org*

[a] *JOSM: https://josm.openstreetmap.de/*

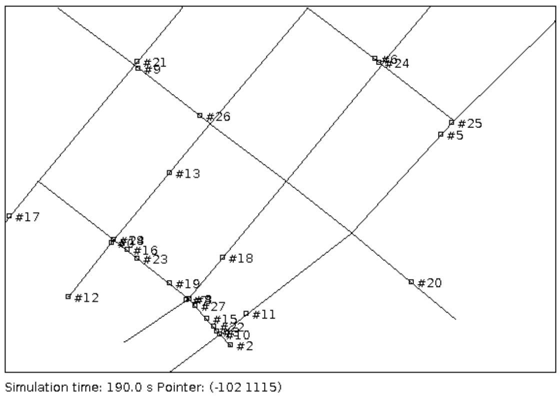


Fig. 12. An example of using VanetMobisim to generate moving file.

TABLE IV
SIMULATION PARAMETERS

Simulation Parameter	Value
Simulation area	1500 m * 1500 m
Simulation time	300 s
Number of vehicles	50, 100, 150, 200, 250, 300
Mac protocol	IEEE 802.11p
Propagation	TwoRayGround
Interface queue	PriQueue
Interface queue length	50
Communication range	250 m
CBR packet speed	1 ~ 0.1
CBR packet size	512 bytes
Hello period	5 s
Min. speed \ Max. speed	8 m/s \ 20 m/s
$\delta_1, \delta_2, \delta_3$	0.6, 0.2, 0.2
α, γ	0.7, 0.9

- Metric1: End-to-End Packet delivery ratio (PDR)

$$PDR = Num\_DReceived/Num\_SGenerated \quad (26)$$

Num_DReceived is the number of data packets that are successfully received by the destination node. And the *Num_SGenerated* is the number of data packets that are generated by the source node.

- Metric2: Average End-to-End Delay (AD)

$$AD = Time\_DReceived - Time\_SGenerated \quad (27)$$

Time_DReceived is time that destination node successfully receives packets. And the *Time_SGenerated* is the time that the source node generates the packets.

- Metric3: Overhead (OD)

$$OD = Size\_ConPackets/Size\_PacReceived \quad (28)$$

Size_ConPackets is the total size of control packets. And the *Size_Pac Received* is the total size of data packets that are successfully received by the destination node.

C. Validation and Analysis of QoS Model

In chapter 3, we choose connectivity probability model and delay model to determine the communication quality of the road. The following part describes the validation of these two judging metrices.

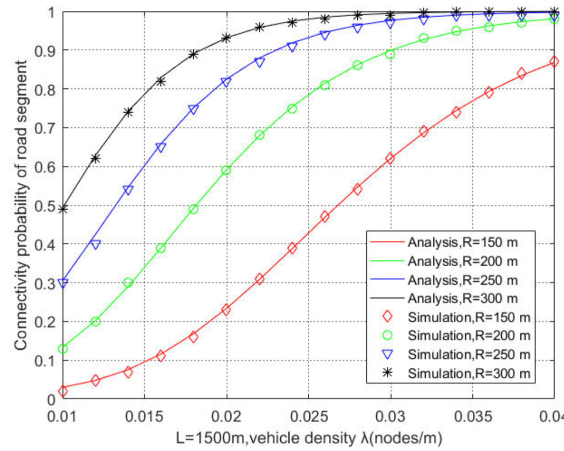


Fig. 13. Validation of the connectivity probability.

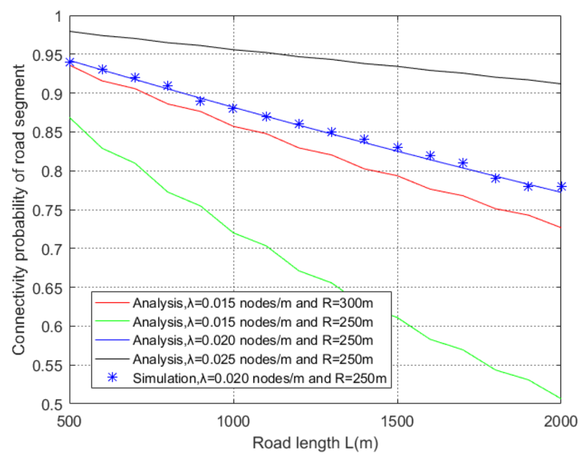


Fig. 14. Validation of the connectivity probability.

1) Validation of the Connectivity Probability Model: The connectivity probability is compared with the data collected through the VanetMobiSim. Fig. 13 and Fig. 14 show the connectivity probability at different vehicular scenarios, which is affected by the transmission range R, the road density λ, and the length of the road segment L. In Fig. 13, we set $L = 1500$ m, $R = 150$, 200, 250, 300 m and change the road density. Fig. 13 shows that theoretical analysis results and simulation results are consistent. In addition, Fig. 13 shows that the larger the transmission range and the greater the density are more conducive to repair the broken link. Moreover, it can be seen from Fig. 13 that the network connectivity is good enough when R = 250m. In order to reduce the direct interference of nodes, it is not appropriate to set the transmission range too large. So in our simulation experiment, we set the transmission range R = 250m.

In Fig. 14, we set $\lambda = 0.015$, 0.02, 0.025 nodes/m, $R = 250$, 300 m, and change the length of the road segment. Number of nodes changes accordingly. When the $\lambda = 0.02$ nodes/m and $R = 250$ m, we compare the simulation results with the analytical results. Fig. 14 also shows that the theoretical analysis results and simulation results are consistent. What's more, when $\lambda = 0.015$ nodes/m, we compare the results

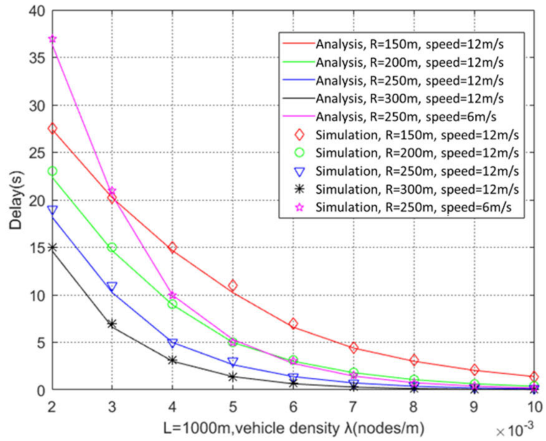


Fig. 15. Validation of the delay model.

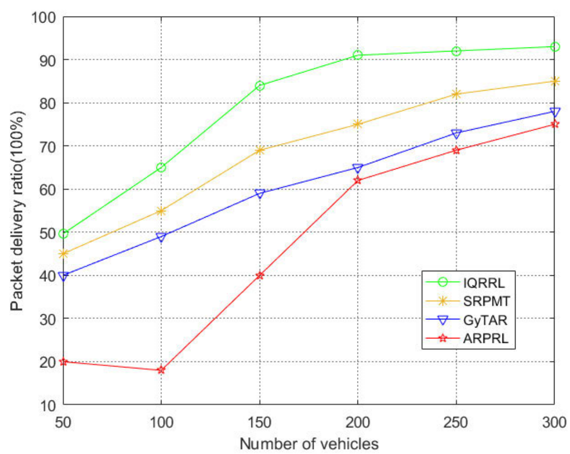


Fig. 16. Packet delivery ratio vs. Number of vehicles.

of $R = 250$m and $R = 300$m, which shows that a larger transmission range helps to repair the broken links. In the case of the same road density, the shorter the road, the better the connectivity. The reason is that the number of broken links will be less on shorter road segment.

2) Validation of the Delay Model: Fig. 15 shows the delay at different vehicular scenarios, which is affected by the transmission range R, the road density λ, vehicle velocity v and the length of the road segment L. In Fig. 15, we set $L =$ 1000 m, $R = 150$, 200, 250, 300 m, $v = 12$ m/s, and change the road density. The theoretical analysis results and simulation results shown in Fig. 15 are in good agreement, which proves that our delay model is correct. In addition, Fig. 15 shows that the delay decreases as the density increases. This is because the broken link can be repaired as the number of vehicles increases. In order to obtain the effect of vehicle speed on the delay, we fix $R = 250$ m and change the velocity from 12 m/s to 6 m/s. The results show that the smaller the speed, the greater the delay. The repair time of a single broken link is fixed when the transmission range, road density, and speed are fixed. Therefore, the total road delay increases linearly with the increase of the road length, which is obviously not be verified here.

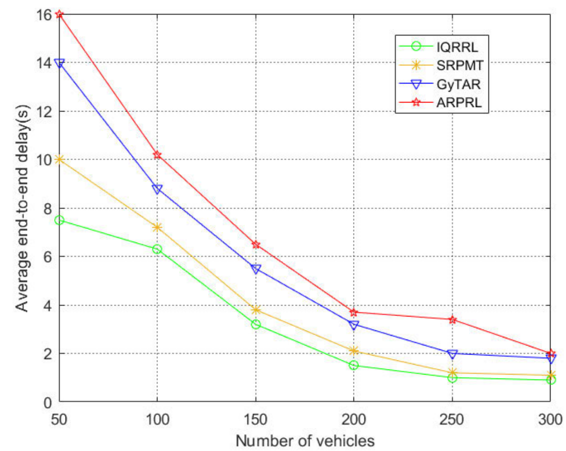


Fig. 17. Average end-to-end delay vs. Number of vehicles.

D. Performance Analysis of Proposed Routing Protocol

1) The Performance of Varying the Number of Vehicles: Fig. 16 shows the packet delivery ratio for different number of vehicles, where we set packet generation speed as 1 packets/s. We can see that the PDR of all protocols increases as the number of vehicles increases. The reason is that an increase in the number of vehicles can increase the network connectivity, improving the possibility for the packets to find the next hop. In addition, we find that our protocol has the best performance among the three protocols. The reasons may come from the following points. First of all, when choosing the intersection, IQRRL chooses the backbone path with the best QoS performance, which includes the connectivity. Second, the next hop has the best long-term returns inside the road, which also can improve the PDR. When choosing an intersection, GyTAR only considered the delay and density performance of neighboring intersections. This local optimum problem may results in the rest of routes encountering the network partitions and packet loss. SRPMT outperforms GyTAR because SRPMT takes into account the characteristic of the candidate road and remote end side to the destination. But the packet may be transmitted through a poor link by greedy carry-and-forward within the road. Before selecting the next hop, ARPRL learns the vehicle mobility information from the source to the destination node, but it ignored the traffic information and may lead to packet loss.

Fig. 17 represents the average end-to-end delay of the four protocols for different number of vehicles. For this figure, we find that with the increase of the number of vehicles, the average end-to-end delays of all protocols are decreasing. The reasons are as follows. The increase in the number of vehicles reduces network partitions, which allow packets to be transmitted over wireless channels instead of being carried by vehicles. Note that the transmission of the wireless channel is much smaller than the delay of carrying. Compared with other three protocols, IQRRL has the lowest average end-to-end delay. The reasons area as follows. Firstly, the delay model of IQRRL considers not only the delay of wireless transmission but also the repair time of carrying, which can more accurately assess the delay of the road. Secondly, due to reinforcement

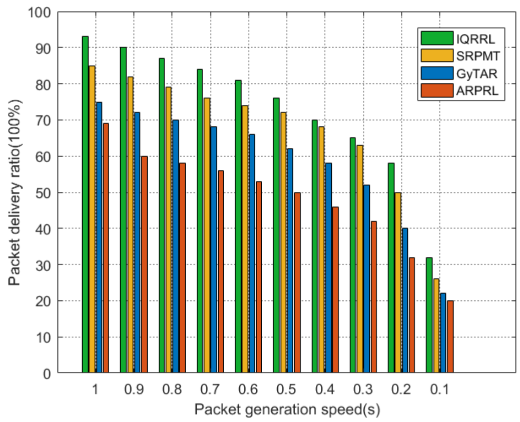


Fig. 18. Packet delivery ratio vs. Packet generation speed.

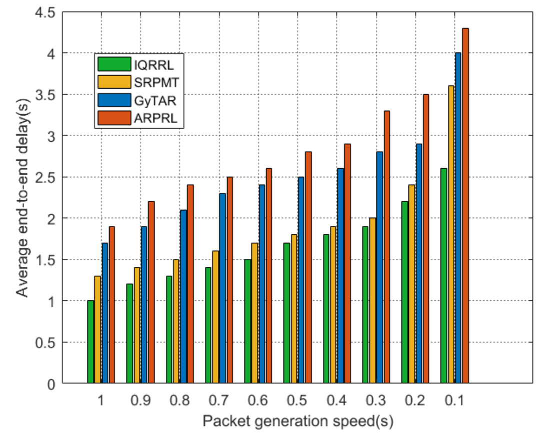


Fig. 19. Average end-to-end delay vs. Packet generation speed.

learning, IQRRL can update available routing information between two intersections, which is conducive to coping with rapid network topology changes and avoid carrying packets. Although SRPMT also estimates the local and global delay, it ignores the distance to the destination. In extreme cases, the intersection selected by SRPMT may be getting farther and farther from the destination node. Therefore, average end-to-end delay does not performance the best. The increase in the number of vehicles will reduce the delay estimation error of GyTAR, but the local optimum problem still exists. In addition, ARPRL route detection from the source node to the destination node will increase the delay.

It is noted that IQRRL is mainly oriented entertainment messages transmitted based on multi-hop unicast mode. Different from safety messages, entertainment messages focus more on data transmission rates rather than latency, so it is acceptable that the end-to-end delay is 1s or more.

2) The Performance of Varying the Packet Generation Speed: Fig. 18 shows the packet delivery ratio for different packet generation speed, where the number of vehicles is set to 150. When the packet generation speed increases, the PDR of all protocols decreases rapidly. This is because the size of vehicles' buffer is limited and there are many network partitions caused by uneven distribution of vehicles. When vehicles encounter network partitions, they need to carry and forward packets. But, the time and the buffer size are limited. Therefore, the more packets that are generated on the sender per second, the more probability it is that packets will be lost due to buffer overflows or timeouts. From Fig. 18, we can notice that IQRRL reach the highest PDR. The reasons are that IQRRL is required to dynamically consider the connectivity of neighboring roads and the remaining shortest path, so that packets can be routed more accurately. In addition, dynamically selecting the next hop based on the results of reinforcement learning inside the road can also reduce the possibility of carrying packets. SRPMT outperforms than GyTAR, because SRPMT, like our solution, dynamically evaluates the performance of the next road and remaining possible roads. But it only considers delay, which is no more accurate than ours. The increase of data packets in the network increases the possibility of collision with the control packet of ARPLR, so the PDR of ARPLR decreases quickly with the increase of the packet generation speed.

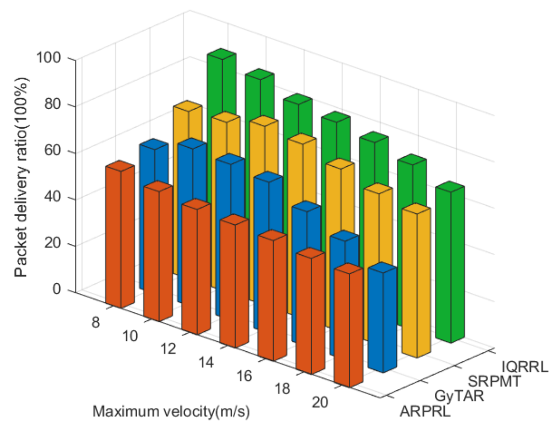


Fig. 20. Packet delivery ratio vs. Maximum velocity.

Fig. 19 compares the average end-to-end delay for different protocols. As shown in the Fig. 19, the average end-to-end delay of three protocols tends to increase as the packets generation speed increases. This is reasonable since more packets result in a larger average transmission delay, which is due to more retransmitted packets and delay extension caused by queueing, buffer overflows or timeouts. Considering local and global QoS including delay, IQRRL can ensure the optimal and real-time selection of roads and effectively avoid network partitions. SRPMT makes the decision depending on the remaining delay, so the possibility of store-and-carrying is smaller than GyTAR. The Q-Table update of ARPRL is not very timely due to packet collision, so the delay performance is relatively poor.

3) The Performance of Varying the Maximum Velocity: Fig. 20 indicates the PDR with the different maximum velocities. We can see that the PDR decreases as the maximum velocity increases. The reason behind this is that the high-speed mobility of vehicles will cause the topology of the network to change very quickly, resulting in shorter routing validity periods. Since IQRRL not only dynamically selects the next intersection, but also considers the mobility of the vehicle inside the road, IQRRL outperforms than other protocols.

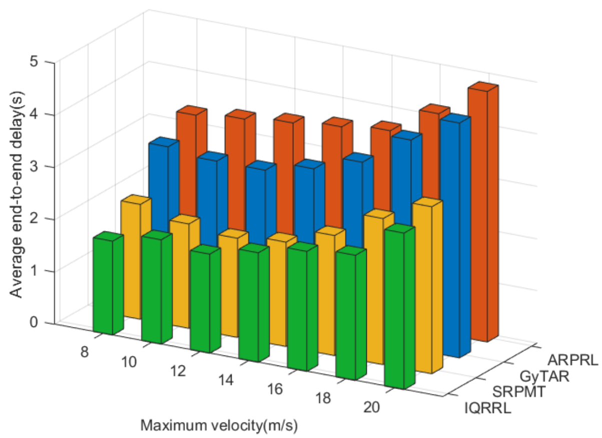


Fig. 21. Average end-to-end delay vs. Maximum velocity.

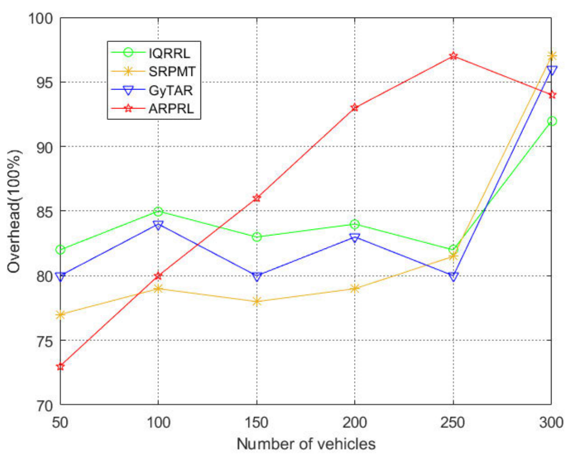


Fig. 22. Overhead vs. Number of vehicles.

Although GyTAR and SRPMT both dynamically select the next intersection, they both adopt greedy carry-and-forward inside the road, which ignores the possibility of vehicle mobility and leads to high packet loss. ARPRL takes full account of vehicle mobility, but it ignores network's QoS. This does not make its PDR perform best.

Fig. 21 describes the average end-to-end delay with variation of maximum velocity. We can know that average-end-to-end delay of all the protocols increases as the maximum velocity increases. This is because high mobility increases the probability of network partitions, which results in packet carrying and increases latency. When the speed is faster, the repair time of IQRRL will be reduced accordingly because of the delay model taking into account the repair time of the network partition. This is why the average end-to-end delay of IQRRL is smaller than other protocols. GyTAR is more likely to encounter network partitions when the speed is fast. SRPMT performs better than GyTAR due to consider local and global factors. However, ARPRL performs better than SRPMT and GyTAR because ARPRL fully considers vehicle mobility.

4) The Performance of Overhead: Fig. 22 shows the overhead of each protocol with varying the number of vehicles. From the figure, it is found that when the number of vehicles increases, the overhead of all routing protocols increases, because the hello control packets occupy a big proportion at

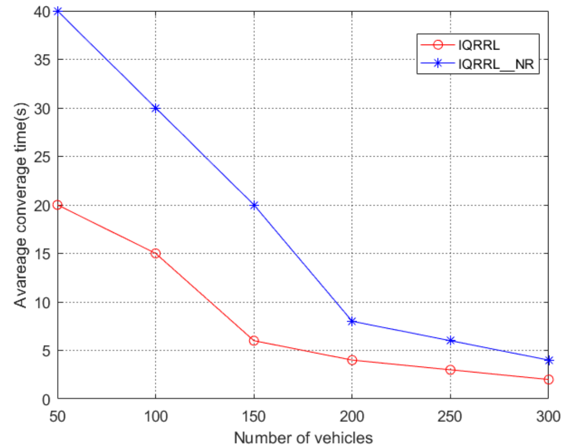


Fig. 23. Average converge time vs. Number of vehicles.

this time. Our IQRRL protocol shows higher overhead than GyTAR and SRPMT protocols, but lower than ARPRL. This is because the Q-Table information of the vehicles is added to the hello packet, and the request packet and the feedback packet also increase overhead. However, since we only make request between two intersections instead of between the source and the destination node, overhead is lower than ARPRL. In addition, it is undeniable that the reinforcement learning process improves the overall performance. How to effectively further reduce the load of IQRRL will be considered as our future work.

5) The Analysis of the Route Convergence Time: Fig. 23 shows the Q-Table's converge time for IQRRL and IQRRL_NR. IQRRL_NR is a version of IQRRL, which does not send LRTP and LRYP before finding the best next hop within the selected road segment. From this figure, we can observe that the average converge time of Q-Table in IQRRL is shorter than that in IQRRL_NR. Obviously, request packets and feedback packets provide vehicle nodes with more routing information, which speeds up the process of reinforcement learning. Therefore, they help to reduce the converge time.

VI. Conclusion

In this paper, a new routing protocol IQRRL is proposed for selecting an optimal route path in the urban environment, which is used to support commercial or entertainment applications for VANETs. IQRRL has two main steps: the selection of the next intersection and the selection of the next-hop vehicle within the road segment.

In the first step, a big challenge is how to use more accurate QoS parameters to dynamically select the next intersection in two-lane road segment scenarios, and the solution in this paper is considering three metrics, the connectivity probability, delay and the distance to the destination. Besides, another challenge is the problem of local optimum in the selection of the intersection, and this paper considers the performance of neighbor roads and remaining shortest roads to minimize this problem.

In the second step, after the road segment is determined by the first step, packets need to be relayed forward through vehicles on the road. However, vehicle mobility is a big challenge for the choice of the next-hop of the packet, and this problem is solved by reinforcement learning in this paper.

Finally, in order to verify the accuracy of the QoS model and the performance of IQRRL, we completed the simulation experiment on NS2. The simulation results show that our protocol improves the packet delivery rate, reduce the average end-to-end delay. In addition, the QoS model and simulation results are consistent.

REFERENCES

[1] M. Hashem Eiza, T. Owens, Q. Ni, and Q. Shi, "Situation-aware QoS routing algorithm for vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 12, pp. 5520–5535, Dec. 2015.

[2] H. I. Abbasi, R. C. Voicu, J. A. Copeland, and Y. Chang, "Towards fast and reliable multihop routing in VANETs," *IEEE Trans. Mobile Comput.*, vol. 19, no. 10, pp. 2461–2474, Oct. 2020.

[3] S. Harrabi, I. B. Jaafar, and K. Ghedira, "An enhanced AODV routing protocol for vehicular adhoc networks," in *Proc. Int. Conf. Big Data Adv. Wireless Technol.*, Nov. 2016, pp. 1–6.

[4] B. Karp and H. T. Kung, "GPSR: Greedy perimeter stateless routing for wireless networks," in *Proc. 6th Annu. Int. Conf. Mobile Comput. Netw.*, 2000, pp. 243–254.

[5] H. Saleet, R. Langar, K. Naik, R. Boutaba, A. Nayak, and N. Goel, "Intersection-based geographical routing protocol for VANETs: A proposal and analysis," *IEEE Trans. Veh. Technol.*, vol. 60, no. 9, pp. 4560–4574, Nov. 2011.

[6] 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.

[7] X. Zhang, X. Cao, L. Yan, and D. K. Sung, "A street-centric opportunistic routing protocol based on link correlation for urban VANETs," *IEEE Trans. Mobile Comput.*, vol. 15, no. 7, pp. 1586–1599, Jul. 2016.

[8] D. Tasneem and A. B. Kamalrulnizam, "Traffic aware routing in vehicular ad hoc networks: Characteristics and challenges," *Telecommun. Syst.*, vol. 61, no. 3, pp. 489–516, Mar. 2016.

[9] Q. Yang, A. Lim, S. Li, J. Fang, and P. Agrawal, "ACAR: Adaptive connectivity aware routing protocol for vehicular ad hoc networks," *Mobile Netw. Appl.*, vol. 15, no. 1, pp. 36–60, Feb. 2010.

[10] S. Panichpapiboon and W. Pattara-atikom, "Connectivity requirements for self-organizing traffic information systems," *IEEE Trans. Veh. Technol.*, vol. 57, no. 6, pp. 3333–3340, Nov. 2008.

[11] Y. Xiang, Z. Liu, R. Liu, W. Sun, and W. Wang, "GeoSVR: A map-based stateless VANET routing," *Ad Hoc Netw.*, vol. 11, no. 7, pp. 2125–2135, Sep. 2013.

[12] H. Yu, J. Yoo, and S. Ahn, "A VANET routing based on the real-time road vehicle density in the city environment," in *Proc. 5th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Jul. 2013, pp. 333–337.

[13] C. Chen, L. Liu, T. Qiu, K. Yang, F. Gong, and H. Song, "ASGR: An artificial spider-web-based geographic routing in heterogeneous vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 5, pp. 1604–1620, May 2019.

[14] G. Li, L. Boukhatem, and J. Wu, "Adaptive quality-of-service-based routing for vehicular ad hoc networks with ant colony optimization," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3249–3264, Apr. 2017.

[15] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, vol. 1. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[16] S. Wang, X. Chai, X. Song, and X. Liang, "Deep Q-learning enabled wireless resource allocation for 5G network based vehicle-to-vehicle communications," in *Proc. IEEE 6th Int. Conf. Signal Image Process. (ICSIP)*, Nanjing, China, 2021, pp. 903–907.

[17] J. Wu, M. Fang, and X. Li, "Reinforcement learning based mobility adaptive routing for vehicular ad-hoc networks," *Wireless Pers. Commun.*, vol. 101, no. 4, pp. 2143–2171, Aug. 2018.

[18] S. Bitam, A. Mellouk, and S. Zeadally, "Bio-inspired routing algorithms survey for vehicular ad hoc networks," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 843–867, 2nd Quart., 2015.

[19] C. Wu, T. Yoshinaga, Y. Ji, and Y. Zhang, "Computational intelligence inspired data delivery for vehicle-to-roadside communications," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 12038–12048, Dec. 2018.

[20] 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.

[21] A. Fonseca and T. Vazão, "Applicability of position-based routing for VANET in highways and urban environment," *J. Netw. Comput. Appl.*, vol. 36, no. 3, pp. 961–973, May 2013.

[22] M. Jerbi, S.-M. Senouci, T. Rasheed, and Y. Ghamri-Doudane, "Towards efficient geographic routing in urban vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 58, no. 9, pp. 5048–5059, Nov. 2009.

[23] 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.

[24] 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. 2019.

[25] L. Liu, C. Chen, B. Wang, Y. Zhou, and Q. Pei, "An efficient and reliable QoF routing for urban VANETs with backbone nodes," *IEEE Access*, vol. 7, pp. 38273–38286, 2019.

[26] 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.

[27] X. M. Zhang, K. H. Chen, X. L. Cao, and D. K. Sung, "A street-centric routing protocol based on microtopology in vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5680–5694, Jul. 2016.

[28] Y. Ding and L. Xiao, "SADV: Static-node-assisted adaptive data dissemination in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 5, pp. 2445–2455, Jun. 2010.

[29] C. Shao, S. Leng, Y. Zhang, A. Vinel, and M. Jonsson, "Performance analysis of connectivity probability and connectivity-aware MAC protocol design for platoon-based VANETs," *IEEE Trans. Veh. Technol.*, vol. 64, no. 12, pp. 5596–5609, Dec. 2015.

[30] A. B. Reis, S. Sargento, F. Neves, and O. K. Tonguz, "Deploying roadside units in sparse vehicular networks: What really works and what does not," *IEEE Trans. Veh. Technol.*, vol. 63, no. 6, pp. 2794–2806, Jul. 2014.

[31] V. Vijayakumar et al., "Research on QoS aware dynamic reconfiguration and performance measures in VANET," in *Proc. Global Conf. Commun. Technol. (GCCT)*, Thuckalay, India, 2015, pp. 829–833, doi: 10.1109/GCCT.2015.7342777.

[32] J. Härri, M. Fiore, F. Fethi, and C. Bonnet, "VanetMobiSim: Generating realistic mobility patterns for VANETs," in *Proc. 3rd Int. Workshop Veh. Ad Hoc Netw.*, 2006, pp. 96–97.

[33] A. Silva, N. Reza, and A. Oliveira, "Improvement and performance evaluation of GPSR-based routing techniques for vehicular ad hoc networks," *IEEE Access*, vol. 7, pp. 21722–21733, 2019.

[34] D. P. Dora, S. Kumar, and M. Joshi, "Impact of traffic signal on connectivity in intersection based connectivity aware geocast routing (IB-CAGR) in VANETs," in *Proc. 3rd Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Feb. 2016, pp. 4–8.

[35] D. B. Rawat, D. C. Popescu, G. Yan, and S. Olariu, "Enhancing VANET performance by joint adaptation of transmission power and contention window size," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 9, pp. 1528–1535, Sep. 2011.

[36] M. M. E. A. Mahmoud, X. Lin, and X. Shen, "Secure and reliable routing protocols for heterogeneous multihop wireless networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 4, pp. 1140–1153, Apr. 2015.

[37] Z. Khan, P. Fan, S. Fang, and F. Abbas, "An unsupervised cluster-based VANET-oriented evolving graph (CVoEG) model and associated reliable routing scheme," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 10, pp. 3844–3859, Oct. 2019.

[38] C. Chen, L. Liu, T. Qiu, J. Jiang, Q. Pei, and H. Song, "Routing with traffic awareness and link preference in Internet of Vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 1, pp. 200–214, Jan. 2022.

[39] Y. Ding, Y. Hu, K. Hao, and L. Cheng, "MPSICA: An intelligent routing recovery scheme for heterogeneous wireless sensor networks," *Inf. Sci.*, vol. 308, pp. 49–60, Jul. 2015.

[40] Y. Hu, Y. Zheng, X. Wu, and H. Liu, "A rendezvous node selection and routing algorithm for mobile wireless sensor network," *KSII Trans. Internet Inf. Syst. (TIIS)*, vol. 12, no. 10, pp. 4738–4753, 2018.

[41] Y. Zhang, J. Zhou, S. Ma, and M. Shu, "A routing enhanced localisation algorithm for wireless sensor networks," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 15, nos. 1–3, pp. 6–11, 2014.

[42] B. Tan, Y. Peng, and J. Lin, “A local path planning method based on Q-learning,” in *Proc. Int. Conf. Signal Process. Mach. Learn. (CONF-SPML)*, Nov. 2021, pp. 80–84, doi: 10.1109/CONF-SPML54095.2021.00024.

[43] N. Kantasewi, S. Marukatat, S. Thainimit, and O. Manabu, “Multi Q-table Q-learning,” in *Proc. 10th Int. Conf. Inf. Commun. Technol. Embedded Syst. (IC-ICTES)*, Mar. 2019, pp. 1–7, doi: 10.1109/ICTEmSys.2019.8695963.

[44] N. Darapaneni et al., “Automated portfolio rebalancing using Q-learning,” in *Proc. 11th IEEE Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON)*, Oct. 2020, pp. 596–602, doi: 10.1109/UEMCON51285.2020.9298035.

Lanlan Rui (Member, IEEE) received the Ph.D. degree in computer science technology from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2010.

She is currently an Associate Professor with the State Key Laboratory of Networking and Switching Technology, BUPT. Her research interests include the IoT, MEC, content-based measurement and analysis, quality of service (QoS) and intelligent theory, and technology of network services. As a result of the standardization in network management.

Zhibo Yan is currently pursuing the master's degree in computer science and technology with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China.

His research interests include edge computing and vehicular ad-hoc networks (VANET).

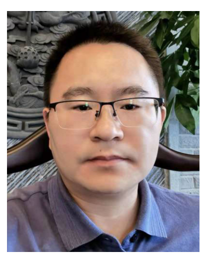
Zuoyan Tan is currently pursuing the master's degree with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China. Her current research interests include vehicular ad-hoc networks (VANET) and quality of service (QoS).

Zhipeng Gao (Member, IEEE) received the Ph.D. degree in computer science technology from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2007. He is currently a Professor and the Vice-President of the International School, BUPT. He has presided over a series of key research projects on network and service management, including projects supported by the National Natural Science Foundation and 863 Programs. He has received eight provincial scientific and technical awards.

Yang Yang (Member, IEEE) received the Ph.D. degree in computer science technology from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2011.

She is currently an Associate Professor with the State Key Laboratory of Networking and Switching Technology, BUPT. Her research interests include data mining and artificial intelligence.

Xingyu Chen (Member, IEEE) received the Ph.D. degree in computer science technology from Northeastern University in 1998. He is currently a Professor with the Beijing University of Posts and Telecommunications. As the project chief researcher, he has participated in and completed more than ten major projects, including National 863, National Key Problems, National Natural Science Foundation of China, and international cooperation. His current research interests include network management and communication software.



Huiyong Liu (Member, IEEE) received the Ph.D. degree. He is currently an Associate Professor with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications. His main research interests include network management, big data, and industrial internet.