IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 14, 2019, accepted December 30, 2019, date of publication January 6, 2020, date of current version January 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2963850

RSU-Assisted Traffic-Aware Routing Based on Reinforcement Learning for Urban Vanets

JINQIAO WU, MIN FANG, HAIKUN LI, AND XIAO LI
School of Computer Science and Technology, Xidian University, Xi'an 710071, China

Corresponding author: Min Fang (fanglabtg@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61472305, Grant 61070143, and Grant 61806155, in part by the Science and Technology Project of Shaanxi Province, China, under Grant 2015GY027, in part by the Aeronautical Science Foundation of China under Grant 20151981009, in part by the China Postdoctoral Science Foundation Funded Project under Grant 2018M631125, and in part by the Fundamental Research Funds for the Central Universities under Grant XJS18037.

ABSTRACT In urban vehicular ad hoc networks (VANETs), the high mobility of vehicles along street roads poses daunting challenges to routing protocols and has a great impact on network performance. In addition, the frequent network partition caused by an uneven distribution of vehicles in an urban environment further places higher requirements on the routing protocols in VANETs. More importantly, the high vehicle density during the traffic peak hours and a variety of natural obstacles, such as tall buildings, other vehicles and trees, greatly increase the difficulty of protocol design for high quality communications. Considering these issues, in this paper, we introduce a novel routing protocol for urban VANETs called RSU-assisted Q-learning-based Traffic-Aware Routing (QTAR). Combining the advantages of geographic routing with the static road map information, QTAR learns the road segment traffic information based on the Q-learning algorithm. In QTAR, a routing path consists of multiple dynamically selected high reliability connection road segments that enable packets to reach their destination effectively. For packet forwarding within a road segment, distributed V2V Q-learning (Q-learning occurs between vehicles) integrated with QGGF (Q-greedy geographical forwarding) is adopted to reduce delivery delay and the effect of fast vehicle movements on path sensitivity, while distributed R2R Q-learning (Q-learning occurs between RSU units) is designed for packet forwarding at each intermediate intersection. In the case of a local optimum occurring in QGGF, SCF (store-carry-forward) is used to reduce the possibility of packet loss. Detailed simulation experimental results demonstrate that QTAR outperforms the existing traffic-aware routing protocols, in terms of 7.9% and 16.38% higher average packet delivery ratios than those of reliable traffic-aware routing (RTAR) and greedy traffic-aware routing (GyTAR) in high vehicular density scenarios and 30.96% and 46.19% lower average end-to-end delays with respect to RTAR and GyTAR in low vehicular density scenarios, respectively.

INDEX TERMS Mobile ad hoc networks (MANETs), vehicular ad hoc networks (VANETs), adaptive routing, reinforcement learning, Q-learning.

I. INTRODUCTION

With the rapid development of wireless communication technology, vehicular ad hoc networks (VANETs) have emerged as one of the most prospective solutions to enhance road traffic efficiency and decrease road traffic accidents in an intelligent transportation system (ITS). In addition, the significant progress in wireless communications technology and widespread use of mobile electronic terminal equipment have migrated VANETs from the realm of theory to a practical technology. However, message transmission in VANETs faces difficult challenges such as frequent changes of the network topology, intermittent connection, and nonuniformity of vehicle density [1]. These new challenges may greatly affect the experience of VANET-based applications that have a wide variety of quality of service (QoS) requirements such as low delay and high accessibility.

A considerable number of traditional routing protocols designed for MANETs have been proposed, among which previous studies have shown that they are not suitable for the VANETs environment. In addition, some conventional geographical routing protocols [2]–[4] are considered promising

The associate editor coordinating the review of this manuscript and approving it for publication was Maurizio Murroni.

approaches to forward packets in dynamic network environment. Despite routing simplicity and scalability, geographical greedy forwarding is still unable to achieve better performance in urban VANETs. Furthermore, the recovery strategies, such as perimeter forwarding, are also shown to be ineffective in urban VANETs due to the limits of the radio range, tall building obstacles and high vehicle mobility. More importantly, conventional geographical routing protocols do not take into account the real-time road traffic information which can help predict the occurrence of a local optimum and avoid unnecessary entry into the unreachable next forwarder.

To overcome the limitations of conventional geographical routing, a variety of traffic-aware routing protocols [5] have been proposed to improve the routing adaptability to urban VANETs. Unfortunately, many of the existing traffic-aware routing protocols select the next forwarder based on the greedy method both within road segments and in the intersection areas, neglecting the road structure and therefore obtaining lower routing performance. In addition, vehicles passing through intersections often change their speed and direction unexpectedly, which leads to high mobility and further results in poor forwarding performance. For this reason, a number of intersection-based traffic-aware routing protocols [6]–[11] have been proposed to make forwarding decisions at intersections. Nevertheless, these protocols strongly rely on accurate location information especially in the intersection areas. Therefore, a novel high mobility adaptive traffic-aware routing protocol suitable for urban VANETs based on the Q-learning algorithm is proposed, in which a routing path consisting of a succession of road segments and intersections is learned with high connection reliability and low average end-to-end delay in dense and sparse traffic cases, respectively. Combining the advantage of Q-learning-based geographic routing with the information of the static topology of road networks, the real-time road traffic information between two adjacent intersections is dynamically learned.

Reinforcement learning is increasingly being applied to solve dynamic routing problems [10]. The Q-learning [12] algorithm is one of the most common algorithms of reinforcement learning [13], which achieves optimal decisions through interaction with the environment without prior knowledge of the environment model. Through frequent exploration of the environment, the agents will continually attain and update the mapping from a set of environment states to a set of actions available in these states. In VANETs, the entire VANETs can be modeled as the environment. Each vehicle and packet in the VANETs can be regarded as a state and an agent, respectively. The packet forwarding process can be considered as the interaction between the agent and the environment. Each packet exchange, whether routing control packet or application data packet, means the learning of the newest state of the network.

The remainder of this paper is organized as follows. Section II presents an overview of the related works. Section III introduces the problem background and motivations of QTAR and is followed by a comprehensive presentation of QTAR in Section IV. In Section V, we evaluate QTAR with a detailed presentation of the simulation results. Finally, Section VI contains our concluding remarks and future works.

II. RELATED WORK

Traffic-aware routing is considered to be the most promising forwarding strategy in the urban VANETs environment. Many traffic-aware routing protocols have been proposed that make routing decisions by considering multiple traffic awareness-related metrics. Anchor-based street-traffic-aware routing (A-STAR [14]) was proposed based on GSR by assigning different weights to adjoining streets according to the probability of keeping vehicular connection within road streets. However, in the urban environment, only parts of the streets are for bus routes; thus, it may take a long forwarding delay for packets to reach their destination due to the lower density of anchor vehicles. Vehicle-assisted data delivery (VADD [15]) was proposed for sparse VANETs and aims to address delay-insensitive applications. However, when the vehicle density is sparse, the optimal next street may not be available. Thus, in this case, the packet should be forwarded through detoured streets. Furthermore, the estimation of the packet forwarding delay is based on statistical data such as the vehicle density. Since the vehicle density varies with time, the least-delay path selected based on the non-real-time statistical data cannot truly reflect the real situation. The static node-assisted data-dissemination protocol for vehicular networks (SADV [16]) was proposed based on VADD, where static nodes are arranged at intersections to deal with cases in which vehicle nodes are very sparse. SADV has three modules, namely, static node-assisted routing (SNAR), link delay update (LDU) and multi-path data dissemination (MPDD). Connectivity-aware routing (CAR [17]) was proposed which adapts the beaconing interval according to the number of one-hop neighbors of a node. However, the overhead introduced by the dynamic beaconing mechanism in the high vehicular density case is considerable.

Improved greedy traffic-aware routing (GyTAR [18]) is a vehicular traffic-adapted routing protocol designed for urban VANETs. In GyTAR, the cell data packet (CDP) is used to collect the real-time vehicular density information between adjacent intersections. However, the CDP may suffer from network partition of the inner street, resulting in difficulties in updating traffic information in a timely manner. This could lead to further inaccurate calculation of the score for the neighbor intersections. In addition, the CDP also introduces excessive extra overhead to the network. Road-based routing using vehicular traffic (RBVT [19]) was proposed to compute street-based routing paths by collecting real-time vehicular traffic information through proactive and reactive strategies, which is distinct from the traditional strategies adopted in most of the existing literature. Similar to A-STAR, the spatial and traffic-aware routing (STAR [20]) protocol was proposed to collect real-time vehicle traffic information on the street

and dynamically forward packets with the help of rated digital maps in a distributed manner. The intersection-based geographical routing protocol (IGRP [6]) was proposed to forward packets to the nearest fixed gateway station while satisfying specific quality of service (QoS) requirements. Zhang *et al.* [21] introduced a street-centric opportunistic routing protocol for urban VANETs combining a novel link correlation model with street-centric opportunistic routing. Zhang *et al.* [22] also proposed a spatial distribution-based connectivity-aware routing protocol that utilizes the uneven position distribution of vehicles moving on small-length road segments. Wu *et al.* [23] presented a vehicle-to-roadside communication protocol integrated with distributed clustering based on a coalitional game approach and a route selection strategy based on reinforcement learning. However, these protocols cannot take full advantage of combining dynamic traffic information within road segments with the global static road topology information to further improve network performance.

Intersection-based traffic-aware routing (iCar-II [7], [8]) was proposed to enable infotainment applications for urban VANETs and aims to improve the packet delivery ratio and reduce the end-to-end delay via LTE networks. However, iCar-II needs a real-time update of locations and mobility information at location centers. Furthermore, running the shortest-path algorithm between two arbitrary vehicles in a connected weighted graph is impossible since it involves unlimited unknown intermediate intersections, especially in large urban VANETs. A street-centric routing protocol based on the novel concept of microtopology (SRPMT [24]) was proposed for urban VANETs scenarios. However, the collection of dynamic characteristics of road segments for building the packet transfer graph in an MT can easily become invalid, especially for a long road segment. In [25], the authors proposed a reliable traffic-aware routing (RTAR) protocol, which introduces a reliable next-hop selection scheme within road segments and at intersections through road area reliable routing and intersection area reliable routing algorithms, respectively. However the real-time traffic and network status measurement (RTNSM) process for adjacent road evaluation only considers the adjacent road segments and ignores the segments from the adjacent roads to the destination roads. In addition, the extra overhead introduced in the different phases of RTNSM cannot be neglected, especially in the result announcement phase.

Many routing protocols based on reinforcement learning have been proposed in recent years [10]. Boyan and Littman proposed QRouting [26] for a wired network. Dowling et al. introduced a routing protocol called SAMPLE [27] for MANETs based on reinforcement learning. Celimuge WU et al. proposed QLAODV [28] and PFQ-AODV [29] to address adaptive routing in a highly dynamic network environment. However, the learning process is triggered based on the route discovery process, which cannot sense the dynamic changes of the network in time and introduces more overhead. The authors in [11], [30] proposed a Q-learning and grid-based routing protocol–QGrid. However, QGrid only focuses on the forwarding issue from the source vehicle to the fixed destination. In addition, it is difficult to determine the size of each grid for different network scenarios, and the greedy selection strategy for intragrid forwarding is inefficient, especially near or within the intersection areas. More importantly, the Q-table for intergrid forwarding is learned offline, which cannot adapt well to the dynamic characteristics of urban VANETs. To the best of our knowledge, this is the first work that studies traffic-aware routing based on reinforcement learning in urban VANETs. Table 1 presents the summary of routing strategy comparison between related existing routing protocols and our proposed work from the context of routing strategy perspective.

III. PROBLEM BACKGROUND AND MOTIVATIONS

Classical topology-based routing protocols [32]–[34] designed for MANETs depend on the distribution of network topology information between network nodes and are not suitable for VANETs due to the frequent topology changes. Geographic routing (GR) [2]–[4] is a promising alternative routing paradigm that utilizes only position information. Unfortunately, many of the existing GR protocols adopt the greedy forwarding strategy based on vehicle location information, which does not fully consider urban road network information. To overcome the shortcomings of GR, a variety of intersection-based traffic-aware routing protocols [5] have been proposed to further improve the adaptability to highly dynamic traffic conditions. Nevertheless, such traffic-aware approaches have no reliable next forwarder selection in urban intersection areas. Some related works [10] based on Q-learning exist that can learn and adapt to the dynamics of networks very well. However, Q-learning-based routing encounters scalability limitations for large highly dynamic networks because of the slow convergence of the learning algorithm; therefore, the forwarding decisions cannot keep up with the road traffic and network topology changes.

To this end, in this paper, we propose a novel RSU-assisted Q-learning-based Traffic Aware Routing (QTAR) protocol designed for urban VANETs to enhance the awareness of road traffic conditions and reduce the impact of the rapid mobility of vehicles on the network performance by providing an efficient packet forwarding mechanism for a variety of applications in scalable urban VANETs. The high-rate but short-range V2V communications within the road segments through the V2V channel are guided by low-rate but long-range R2R communications through the R2R channel. More specifically, the next forwarding vehicle selection within the road segments is implemented according to Q-greedy geographical forwarding based on V2V Q-learning, while at each intersection, it is completed according to Q-greedy intersection forwarding based on R2R Q-learning. In the case of a network fragment, store-carry-forward is adopted. The main contributions of this paper are as follows:

1) A novel high mobility adaptive traffic-aware routing protocol suitable for urban VANETs based on the

TABLE 1. Summary of routing strategy comparison between related existing routing protocols and QTAR.

Protocol	Forwarding within Road Segments	Forwarding at Intersections	Recovery Strategy	Traffic Aware Mechanism
LAR [31]	Along the path determined in advance through local flooding.	Same as Road Segment Forwarding.	Rediscover the path based on flooding in the request zone.	None
GPSR [2]	Location-based Greedy Forwarding.	Same as Road Segment Forwarding.	Perimeter Forwarding based on Right-Hand Rule.	None
GyTAR [18]	Improved Greedy Strategy based on position prediction.	Intersections are dynamically chosen considering both the local traffic density and the distance to the destination.	Carry-and-Forward	Local traffic information is computed using the Cell Density Packet (CDP).
iCar-II [7], [8]	Greedy-based next hop selection based on location and the latest received RSSI.	Shortest Path Algorithm (such as Dijkstra).	Sends a new path request to the location centers.	Road Segment Connectivity is evaluated by sending a unicast control packet (CP) that transverses the road segment.
RTISAR [9]	Improved greedy-based Forwarding based on the distance toward the next intersection and the transmission error rate.	Intersections are dynamically selected based on the road segment density, connectivity, load and the cumulative distance toward the destination.	Carry-and-Forward	Road Segment is evaluated based on the GFD-CP and the BG-CPR phase through the unicast of CPs and CPRs packets, respectively.
RTAR [25]	Selects the reliable next forwarder based on the average RSSI and predicted position of its available neighbors through the Road Area Reliable Routing (RARR).	Elects the best road based on Intersection Area Reliable Routing (IARR) which considers the neighbor's predicted position, average RSSI value, and mobility information recentness.	Carry-and-Forward	Exploit the Real-time Traffic and Network Status Measurement (RTNSM) process which consists of Collector Packet generation, updating and forwarding, results announcement and reply.
QTAR	Q-Greedy Geographical Forwarding based on the learning result of V2V Q-Learning considering link quality, expiration time and delay.	Intersections are dynamically determined based on the learning result of R2R Q-Learning.	Store-Carry-Forward	Road Segment status is dynamically learned based on the V2V Q-Learning and globally learned based on the R2R Q-Learning considering link quality, expiration time and delay through the periodic HELLO packets.

multilevel Q-learning algorithm is proposed, in which a routing path consists of a succession of road segments and intersections are learned with high connection reliability and low average end-to-end delay in dense and sparse traffic cases, respectively.

2) For packet forwarding within road segments, a novel distributed V2V Q-learning-based traffic-aware learning approach is proposed through exchange of V2V HELLO packets, which underlies the forwarding process at dynamic intersections.
3) For packet forwarding at intersections, an RSU-assisted dynamic adjacent intersection selection strategy based on distributed R2R Q-learning is proposed to reduce the possibility of packet loss and the effect of fast vehicle movements on routing sensitivity.

In the following sections, we first provide an elaborated description of QTAR and then present comprehensive experimental results compared with other existing related protocols.

IV. THE PROPOSED PROTOCOL

In this section, we first describe the network model and hypothesis. Then, we present the main functionality of QTAR, which mainly consists of the following components: first, deciding the first intersection to which packets are forwarded from the source vehicle V_s; second, packet forwarding at each intermediate intersection to the next adjacent intersection until reaching the last intersection that connects the road segment on which the destination vehicle V_d is moving; and finally, packet forwarding within the road segment from the last intersection to V_d.

A. NETWORK MODEL AND ASSUMPTIONS

We consider the urban road network as a directed graph $G = (V, E)$, in which V is the set of intersections and E is the set of road segments RS_{ij}, $i, j \in V$. An RS_{ij} begins at the intersection I_i, ends at I_j and has two lanes in each driving direction. The routing path in G from V_s to V_d consists of a sequence of road segments and intersections that connect these road segments.

In QTAR, we assume that each intersection I_i owns a static RSU node V_{RSU_i} to assist packet forwarding. Therefore, in the context of QTAR, the terms I_i and V_{RSU_i} are often used interchangeably to represent an intersection or an RSU node that resides on I_i statically. Each V_{RSU_i} provides partial coverage to road segments, and multihop forwarding is required to communicate with vehicle nodes that are not in range.

Each V_i knows its real-time position, direction and speed using a pre-installed GPS device, and vehicles communicate

with each other or with a radio-in-range RSU node through a pure V2V wireless channel. Furthermore, each V_i also knows its entered and upcoming intersection and the coordinates of each V_{RSU_i} in advance. Each V_i also maintains a table where each neighbor vehicle's mobility information, such as position, direction and velocity, is recorded and updated through the periodic exchange of V2V HELLO packets. Meanwhile, each V_{RSU_i} knows the real-time entered intersection $I_{enter}^{V_d}$ and corresponding upcoming intersection $I_{upcoming}^{V_d}$ of the V_d through GLS [35], and it communicates with vehicle nodes in its radio range through a V2V wireless channel and neighbor RSU nodes through an R2R wireless channel. Finally, each V_i maintains only one V2V Q-table for packet forwarding within the road segment to which V_i belongs, while each V_{RSU_i} stores a neighbor RSU's table and two Q-tables, in which one is a V2V Q-table for road segment traffic-aware forwarding and the other is an R2R Q-table for dynamic intersection forwarding.

B. QTAR OVERVIEW

Considering the specific characteristics of urban VANETs, QTAR is designed to deal with routing issues by combining the advantages of QGGF within road segments and Q-learning-based dynamic selection of the intermediate intersections through which packets will pass to reach their destinations. To forward packets effectively from V_s to V_d, QTAR includes three efficient steps:

1) Packet forwarding from V_s to the first intersection based on V2V Q-learning through exchange of HELLO packets between vehicles moving in the same road segment as V_s;
2) Packet forwarding in each intermediate intersection based on R2R Q-learning through exchange of HELLO packets between RSU nodes;
3) Packet forwarding from the last intersection to V_d based on V2V Q-learning.

Store-carry-forward is adopted to improve the forwarding reliability in the case of a local optimum to minimize the possibility of packet loss. Hence, in QTAR, packets can reach their destinations as fast as possible when there are enough vehicles providing connection. An example of the packet forwarding process from V_s to V_d through a routing path $V_s \rightarrow I_2 \rightarrow I_5 \rightarrow I_6 \rightarrow I_9 \rightarrow V_d$ is shown in Fig. 1. As mentioned above, the routing process in QTAR can be mainly divided into three steps, where $V_s \rightarrow I_2$ is the first step while $I_2 \rightarrow I_5 \rightarrow I_6 \rightarrow I_9$ is the second step and $I_9 \rightarrow V_d$ is the third step. It is worth noting that at I_2, $I_2 \rightarrow I_5 \rightarrow I_6$ is selected instead of $I_2 \rightarrow I_3 \rightarrow I_6$ as the next forwarding path. This is because the road segment RS_{23} is already in a congested state, and RS_{25} will have less delay than RS_{23} because of the channel collisions that occurred in the MAC layer. At I_5, the path $I_5 \rightarrow I_6 \rightarrow I_9$ is selected due to the shorter time of store-carry-forward caused by the network partition that occurs from V_i to V_d in Step 3 compared with

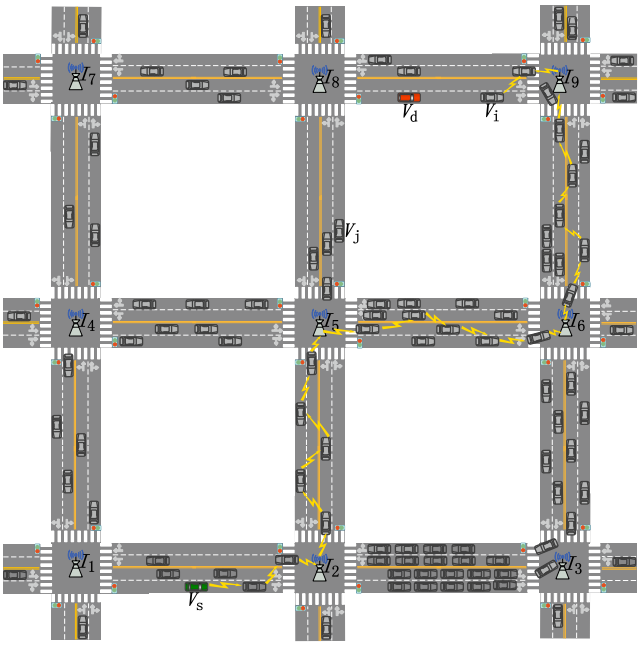


FIGURE 1. An example scenario of packet forwarding in QTAR.

that of RS_{58} from V_j to I_8 in Step 2 and then from I_8 to V_d in Step 3.

In the following sections, we first describe in detail the basic principles underlying the functions of QTAR; then, we elaborate the main novel packet forwarding mechanisms of QTAR.

C. ROUTING BASED ON Q-LEARNING

According to Q-learning [13], [29], [36], the unicast routing problem can be modeled and solved as follows. The entire VANETs can be considered as the environment in Q-learning. Each packet can be modeled as an agent, with the neighbors of V_i or V_{RSU_i} as the agent's available states. Specifically, for a V_i in V2V Q-learning, the set of its neighbor vehicles can be mapped to the available actions for V_i to be executed in the form of forwarding of packets to one of V_i's neighbors. For a V_{RSU_i} node in R2R Q-learning, the set of neighbor intersections of V_{RSU_i} is the available actions for V_{RSU_i}. The process of packet forwarding can be modeled as the interaction process in Q-learning. Therefore, the routing problem can be intuitively formalized as Eq. (1):

$$Q_c(d,x) \leftarrow (1-\alpha)\, Q_c(d,x) + \alpha \left[Reward_{c,x} + \gamma \max_{y \in N(x)} Q_x(d,y) \right] \quad (1)$$

where for V2V Q-learning, $Q_c(d,x)$ is the Q value of current vehicle node c for destination vehicle node d through one of c's neighbor vehicle nodes x. $N(x)$ is the one-hop neighbors of x. $Reward_{c,x}$ is the obtained reward of c from the action of packet forwarding to x. In R2R Q-learning, for current RSU node V_{RSU_c} residing at I_c, c denotes V_{RSU_c}, while d and x denote the destination intersection and one of the neighbor

intersections of I_c, respectively. α is the learning rate that determines the Q value update rate in each step. In other words, it reflects the adaptability ability of the Q-learning algorithm to the dynamic environment. The larger the value of α is, the stronger the learning ability, and the more suitable it is for the environment with severe dynamic characteristics. However, if α is too high, small fluctuations can cause large deviations in Q values, which cannot reflect the real state of the network. If α is too small, Q values cannot keep up with the change of the network. γ is the discount factor that determines the importance of multistep Q values. A larger value of γ means that more future steps are considered. For scenarios with less dynamics, a larger γ is reasonable, while for frequently changing scenarios, a smaller γ is more advisable due to the fast failure of multistep Q values.

D. HELLO PACKET FORMAT FOR V2V AND R2R Q-LEARNING

In QTAR for V2V Q-learning, each vehicle V_i moving in a road segment RS_{ij} maintains a Q-table reflecting the current traffic state of RS_{ij} via exchange of HELLO packets. Each HELLO packet in V2V Q-learning contains the following fields: the unique identifier vehicle/RSU ID, the broadcast timestamp, the coordinates X and Y, the velocity Vel, the entered intersection I_{enter} along with the corresponding optimal Q value Q_{MAX} to reach I_{enter} through the next hop vehicle NH, and the upcoming intersection $I_{upcoming}$ along with the corresponding optimal Q value Q_{MAX} to reach $I_{upcoming}$ through the next hop vehicle NH, as shown in Fig. 2. It is worth noting that each of the HELLO packets broadcast from an RSU node only includes the RSU ID and timestamp fields to reduce overheads and collisions in the intersection area. A vehicle node receiving the HELLO packet will update the I_{enter} or $I_{upcoming}$ Q value according to the driving direction relative to the RSU node.

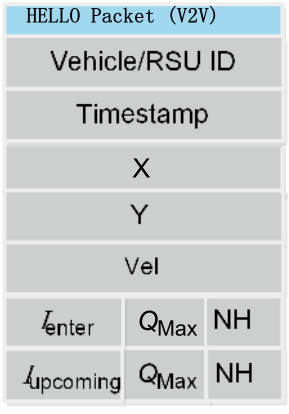


FIGURE 2. HELLO packet format for V2V Q-learning.

For R2R Q-learning, each HELLO packet consists of the fields as depicted in Fig. 3. As shown in Fig. 3, these fields include the sender RSU ID and the broadcast timestamp, the total number of QMax items and their corresponding content. Each QMax item includes three parts: the destination RSU – Dest RSU – and the corresponding optimal Q value to reach it through one of its neighbor RSUs – Next RSU.

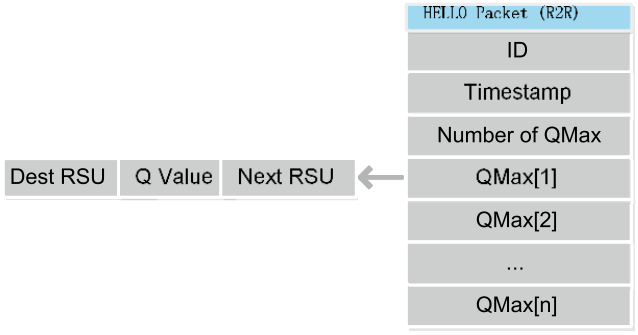


FIGURE 3. HELLO packet format for R2R Q-learning.

Algorithm 1 Packet FORWARDING Within Road Segments

Require:

P_k: A packet that is transmitting in the network.
V_{RSU_i}: The RSU node deployed at I_i.
V_i: A vehicle node.
V_s: The source vehicle of P_k.
V_d: The destination vehicle of P_k.
V_c: The current vehicle that is processing P_k.
I_i: An intersection that connects two or more road segments.
I_x: A set of I_i.
I_{temp}: The temporary destination intersection of P_k.
$N(V_i)$: The set of neighbor nodes of V_i.
$RS(V_i)$: The road segment on which V_i is moving.
$RSU(V_i)$: The two end-side intersections of $RS(V_i)$ or the RSU within radio range of V_i .
Upon V_i having a packet P_k to SEND/FORWARD to V_d

$V_c \leftarrow V_i$;
$I_x \leftarrow RSU\ (V_c)$
$I_{temp} \leftarrow$ Obtain the temporary destination intersection of V_c according to Eq. (2);
if $V_c == V_d$ **then**
  Deliver P_k to the upper layer;
else if $V_d \in N(V_c)$ **then**
  Send P_k directly to V_d;
else if $I_{temp} \in N(V_c)$ **then**
  Send P_k directly to I_{temp};
else
  Forward P_k to I_{temp} based on QGGF and SCF;
end if

E. V2V Q-LEARNING FORWARDING WITHIN ROAD SEGMENTS

When the source vehicle V_s has a packet P_k to send or an intermediate vehicle V_i receives P_k, it forwards P_k to the next hop based on V2V Q-learning until P_k reaches I_{temp} or its final destination vehicle V_d. For the sake of simplicity, here, we denote V_s or V_i as the current vehicle V_c that is processing P_k for further forwarding work if needed. Without loss of

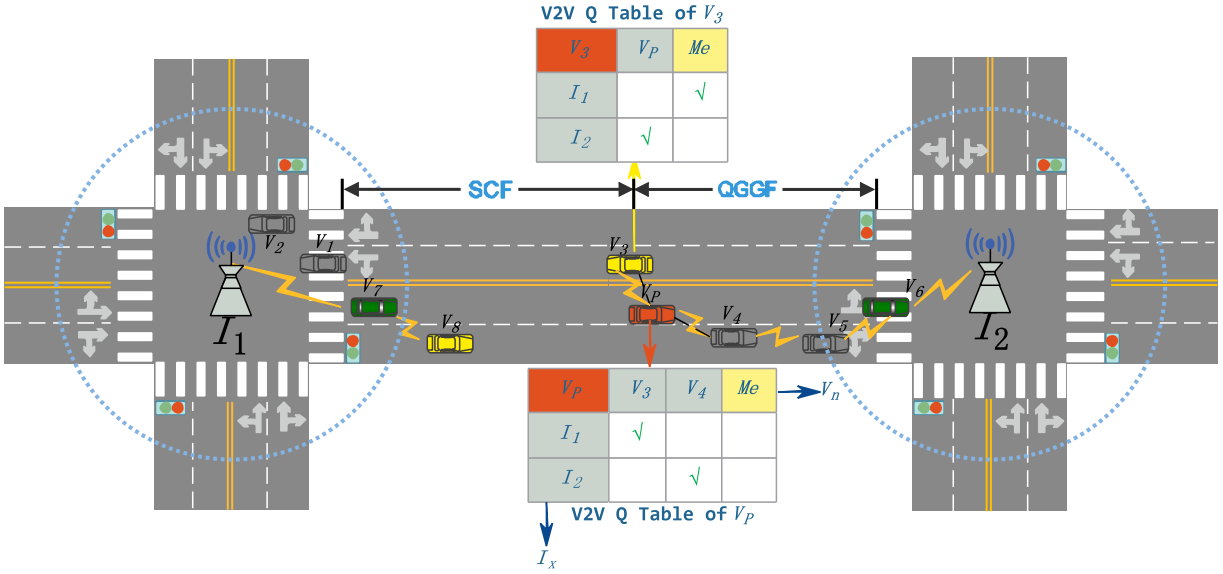


FIGURE 4. An example scenario of V2V Q-learning within RS_{12}.

generality, the current vehicle V_c (corresponding to vehicle V_p marked in red in Fig. 4) moving within a road segment RS_{ij} (referring to RS_{12} in Fig. 4) with two end-side intersections I_i and I_j (referring to I_1 and I_2 in Fig. 4) will forward its received packets P_k to the temporary destination intersection I_{temp} (referring to I_2 for V_P in Fig. 4), which is one of the end-side intersections of RS_{ij}. The first intersection in the routing path, denoted as $I_f = I_{temp}$, is determined as shown in Eq. (2):

$$I_{temp} \leftarrow \underset{I_x}{argmax}\, Q_{V_c}(I_x, V_n), \quad V_n \in N(V_c) \tag{2}$$

where I_x specifically denotes the $I_{enter}^{V_c}$ (referring to I_1 for V_p in Fig. 4) or $I_{upcoming}^{V_c}$ (referring to I_2 for V_p in Fig. 4). V_n is one of the neighbors of V_c (referring to V_3 or V_4 in Fig. 4), and $Q_{V_c}(I_x, V_n)$ indicates the V2V Q-table of V_c (referring to the V2V Q-table of V_p, shown at the middle bottom of RS_{12}). For the vehicles moving on the same road segment as V_c, QGGF or SCF is used until I_{temp} or V_d is reached.

The pseudocode of the forwarding process within road segments is given in Algorithm 1. As illustrated in Algorithm 1, Lines 4-5 mean that packet P_k is successfully forwarded to its destination V_d. The two lines 6 and 7 indicate that V_d is V_c's neighbor. Lines 8 and 9 indicate that P_k has arrived at I_{temp}. Lines 10 and 11 indicate that P_k needs to be forwarded to the temporary destination intersection I_{temp}.

To better understand the packet forwarding process within a specific road segment based on V2V Q-learning, Fig. 4 shows an example scenario that includes some local optimum cases in RS_{12}. As shown in Fig. 4, network partition has occurred in the V2V routing path from I_1 to I_2. In this case, the V2V packet forwarding process within RS_{12} consists of two parts: QGGF is employed when the next hop link exists, and SCF is employed when local optimum is reached. Take V_3 as an example. Its Q-table at the current moment in Fig. 4 is shown at the middle top of RS_{12} (indicated by the red cell in the upper left corner of the Q-table). From the Q-table of V_3, we can see that the next hop to I_2 is $V_P = \underset{V_n}{argmax}\, Q_{V_3}(I_2, V_n)$, while that to I_1 is $Me = \underset{V_n}{argmax}\, Q_{V_3}(I_1, V_n)$, as indicated in the last yellow column, which means that the next hop to I_1 from V_3 does not exist according to QGGF, and SCF is adopted in this case. Each optimal next hop for each intersection is marked in green in the Q-table cell. When there is a packet at I_2 that needs to be forwarded to I_1, V_6 is selected, and then, the QGGF forwarding path (denoted as $V_6 \rightarrow V_5 \rightarrow V_4 \rightarrow V_P \rightarrow V_3$) is selected based on V2V Q-learning. At V_3, the local optimum has occurred, and SCF is adopted to complete the final forwarding process from V_3 to I_1.

More generally, for the current vehicle V_c, its Q-table is initialized to 0 and updated based on the V2V Q-learning algorithm. When receiving a HELLO packet from $V_n \in N(V_c)$, the Q value $Q_{V_c}(I_x, V_n)$ is updated as Eq. (3):

$$\begin{aligned} Q_{V_c}(I_x, V_n) \leftarrow & (1-\alpha)\, Q_{V_c}(I_x, V_n) \\ & + \alpha \left[Reward_{V_c,V_n} + \gamma \cdot \max_{V_{n'} \in N(V_n)} Q_{V_n}(I_x, V_{n'}) \right] \end{aligned} \tag{3}$$

where I_x represents one of the two end-side intersections of road segment RS_{ij} on which V_c is moving, and the instant reward value $Reward_{V_c,V_n}$ is defined as Eq. (4):

$$\begin{aligned} & Reward_{V_c,V_n} \\ & = \omega_1 \cdot LQ_{V_c,V_n} + \omega_2 \cdot LET_{V_c,V_n} + \omega_3 \cdot Delay_{V_c,V_n} \end{aligned} \tag{4}$$

in which ω_1, ω_2 and ω_3 are weight factors that satisfy $\omega_1 + \omega_2 + \omega_3 = 1$ for corresponding parts of LQ (link quality), LET (link expiration time) and $Delay$, respectively. It can be found from Eq. (4) that if the next hop link selected by an action has good quality, a long survival time and a short delay, then

the reward is high, and vice versa. ω_1, ω_2 and ω_3 represent different QoS requirements for different applications. If the application requires high reliability, the values of ω_1 and ω_2 can each be 0.5. If the application is sensitive to delay, the value of ω_3 can be set to 1. If the application requires good overall performance, the values of ω_1, ω_2 and ω_3 can each be 0.333. LQ_{V_c,V_n} denotes the link quality in the form of the distance between the sender and receiver and is defined as Eq. (5):

$$LQ_{V_c,V_n} = 100 \cdot \left(1 - abs(\frac{dis(V_c, V_n)}{R} - \kappa)\right) \tag{5}$$

Here, $dis(V_c, V_n)$ means the two-dimensional Euclidean distance between V_c and V_n. R represents the wireless radio line-of-sight transmission range. The parameter κ represents the optimal normalized distance position with respect to R. The function $abs(\ )$ takes the absolute value of its parameter. LET_{V_c,V_n} denotes the link expiration time [37] and is defined as Eq. (6):

$$LET_{V_c,V_n} = \begin{cases} 100 \ a = 0, \ b = 0 \\ \min\left(100, \dfrac{-(ab+cd)+\sqrt{\left(a^2+c^2\right)R^2-(ad-bc)^2}}{a^2+b^2}\right) \end{cases} \tag{6}$$

where

$$\begin{aligned} a &= v_c \cos\left(\theta_{v_c}\right) - v_n \cos(\theta_{v_n}) \\ b &= x_c - x_n \\ c &= v_c \sin\left(\theta_{v_c}\right) - v_n \sin(\theta_{v_n}) \\ d &= y_c - y_n \end{aligned}$$

in which v_c and v_n represent the speeds of V_c and V_n with the velocity angles of θ_{v_c} and θ_{v_n} and the coordinates (x_c, y_c) and (x_n, y_n), respectively. $Delay_{V_c,V_n}$ is defined as Eq. (7):

$$Delay_{V_c,V_n} = 100 \ \cdot \ \frac{l_p/BW + dis(V_c, V_n)/C}{t_{V_c}^{recv} - t_{V_n}^{send}} \tag{7}$$

Here, l_p is the HELLO packet length. BW is the link available bandwidth. C is the electromagnetic radiation propagation speed. $t_{V_c}^{recv}$ and $t_{V_n}^{send}$ represent the sending and receiving timestamp of the HELLO packet at V_c and V_n, respectively.

There are two situations that need to be considered specifically in V2V Q-learning. One is the local optimum, as shown in Fig. 4 from V_3 and V_8 (the cars marked in yellow in Fig. 4) to I_1 and I_2, respectively. In this case, the next hop does not existing and SCF is used. At this point, let I_x denote a virtual vehicle node V_n in Eq. (3). Obviously, we have $LQ_{V_c,I_x} = 0$ and $LET_{V_c,I_x} = 0$ in Eq. (4), and finally the $Reward_{V_c,I_x}$ can be calculated as Eq. (8):

$$\begin{aligned} Reward_{V_c,I_x} &= Delay_{V_c,I_x} \\ &= 100 \cdot \left(1 - \frac{dis\left(V_c, I_x\right)}{L}\right) \end{aligned} \tag{8}$$

Algorithm 2 Packet FORWARDING at Intersections

Require:

P_k: A packet that is transmitting in the network.
V_{RSU_i}: The RSU node deployed at I_i.
V_d: The destination vehicle of P_k.
I_i: An intersection that connects two or more road segments.
I_c: The intersection where P_k is processing.
I_d: The destination intersection of V_d for P_k.
I_x: A set of I_i.
I_{temp}: The temporary destination intersection of P_k.
$N(V_i)$: The set of neighbor nodes of V_i.
$RS(V_i)$: The road segment on which V_i is moving.
$RSU(V_i)$: The two end-side intersections of $RS(V_i)$ or the RSU within radio range of V_i.
Upon V_{RSU_i} receiving a packet P_k

$I_c \leftarrow V_{RSU_i}$;
if $V_d \in N(I_c)$ **then**
  Send P_k directly to V_d;
else
  $I_x \leftarrow RSU\ (V_d)$
  $I_d \leftarrow$ Obtain the destination intersection according to Eq. (10);
  if $I_d == I_c$ **then**
    $I_{temp} \leftarrow \{I_i | \ I_i \neq I_d, I_i \in I_x\}$
  else
    $I_{temp} \leftarrow$ Select the next intersection according to Eq. (9);
  end if
  Forward P_k to I_{temp} based on V2V Q-learning;
end if

where L represents the length of RS_{ij}. The other situation involves those vehicles that are within the coverage of the RSU node, such as V_1 and V_7 in I_1 and V_5 and V_6 in I_2. In this circumstance, Eq. (3) is optimized by setting $\alpha = 1$ to boost the convergence of the V2V Q-learning algorithm.

F. R2R Q-LEARNING FORWARDING AT INTERSECTIONS

R2R Q-learning is adopted for each intermediate temporary destination intersection I_{temp} selection except for I_f, which is selected as described in section IV-E. Each I_{temp} is dynamically selected if needed based on Eq. (9):

$$I_{temp} \leftarrow \underset{I_n}{argmax}\ Q_{I_c}(I_d, I_n), \quad I_n \in N(I_c) \tag{9}$$

where

$$I_d \leftarrow \underset{I_x}{argmax}\ Q_{I_c}(I_x, \ I_n) \tag{10}$$

in which $I_x \in \left[I_{upcoming}^{V_d}, \ I_{enter}^{V_d}\right]$ and I_c is the current intersection where the packets are processing and will be forwarded to one neighbor intersection of I_c. When receiving a HELLO

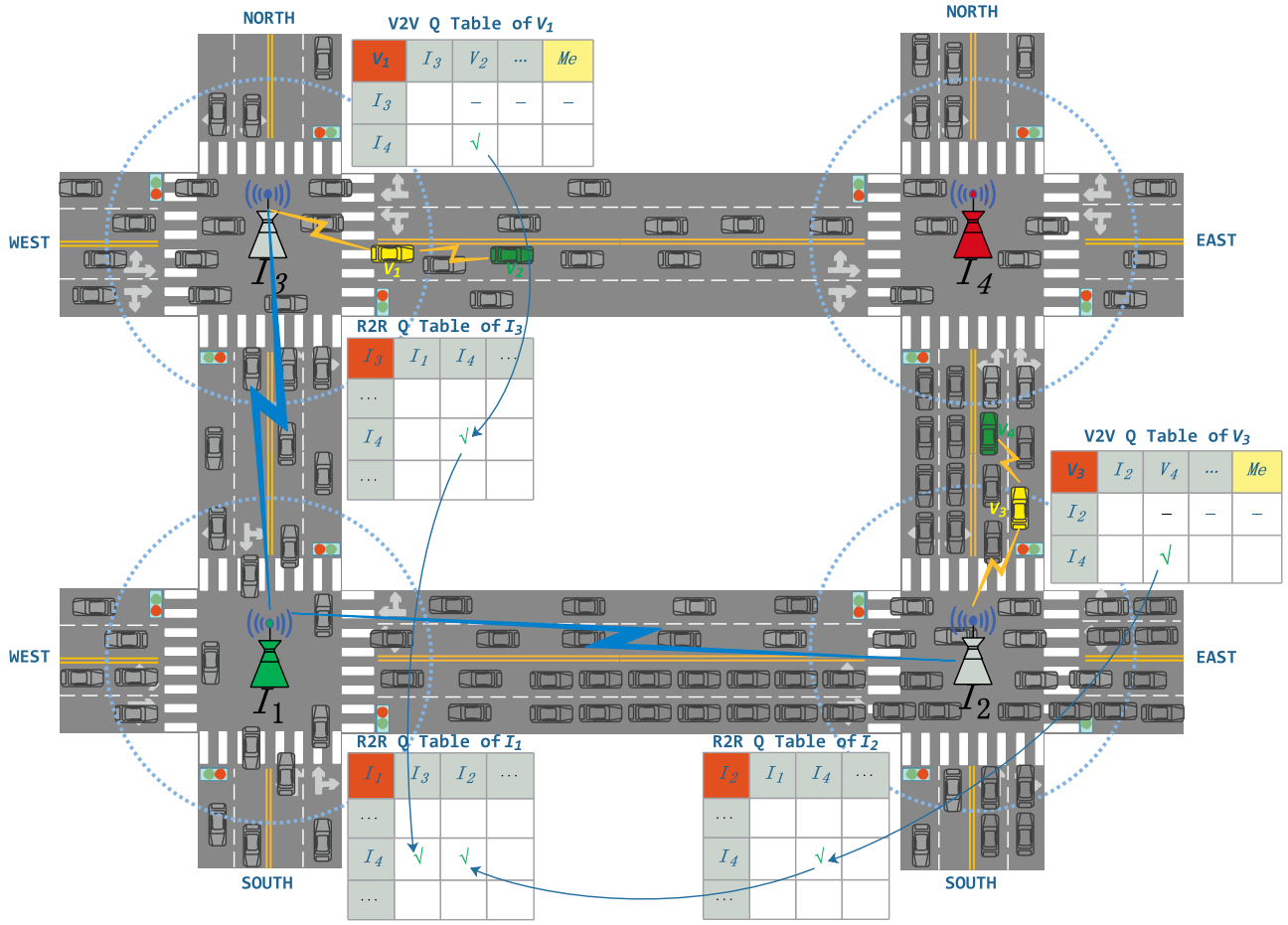


FIGURE 5. An example scenario of R2R Q-learning.

packet from I_n, I_c updates $Q_{I_c}(I_{V_d}, I_n)$ as Eq. (11):

$$Q_{I_c}(I_d, I_n) \leftarrow (1-\alpha)\, Q_{I_c}(I_d, I_n) + \alpha \left\{ Reward_{I_c,I_n} + \gamma \cdot \max_{I_{n'} \in N(I_n)} Q_{I_n}(I_d, I_{n'}) \right\} \quad (11)$$

in which $Reward_{I_c,I_n}$ is defined as Eq.(12):

$$Reward_{I_c,I_n} = \begin{cases} \max Q_{V_n}(I_n, V_{n'}), & I_n = I_x \\ 0, & otherwise \end{cases} \quad (12)$$

Here, $V_n \in N(I_c)$, $V_{n'} \in N(V_n)$.

The pseudocode of the forwarding process at each intermediate intersection is given in Algorithm 2. Lines 2-3 indicate that V_d is within the coverage of I_c. Lines 4-13 represent the forwarding process based on R2R Q-learning. In this case, lines 5-6 obtain the destination intersection I_d of P_k, while lines 7-11 select the next hop intersection, denoted as I_{temp}, from the adjacent intersections of I_c according to the R2R Q-table of I_c in the row indexed by I_d. The last step is to forward P_k to I_{temp}, as indicated in line 12.

Fig. 5 shows an example scenario of R2R Q-learning. In Fig. 5, we assume that the source vehicle V_s located in the west road segment of I_1 and the destination vehicle V_d located in the north road segment of I_4. For brevity, we focus only on the packet forwarding process from the assumed source intersection I_1 (marked as a green RSU node) to the assumed destination intersection I_4 (marked as a red RSU node). The main goal of R2R Q-learning is to choose the optimal next adjacent intersection (which denotes the intersection with the maximum Q value) to the destination intersection. As shown in Fig. 5, there are two paths from I_1 to I_4 (denoted as $I_1 \rightarrow I_2 \rightarrow I_4$ and $I_1 \rightarrow I_3 \rightarrow I_4$), and when a packet P_k arrives at intersection I_1, I_1 compares the Q values ($Q_{I_1}(I_4, I_2)$ and $Q_{I_1}(I_4, I_3)$) at row I_4 in its R2R Q-table, where $Q_{I_1}(I_4, I_2)$ and $Q_{I_1}(I_4, I_3)$ are learned and updated from those of I_2 and I_3 (denoted $Q_{I_2}(I_4, I_4)$ and $Q_{I_3}(I_4, I_4)$) through the R2R links $I_1 \leftrightarrow I_2$ and $I_1 \leftrightarrow I_3$, respectively, which are marked by directional arc lines as shown in Fig. 5. Obviously, $Q_{I_2}(I_4, I_4)$ and $Q_{I_3}(I_4, I_4)$ are learned and updated by I_2 and I_3 through V2V Q-learning in RS_{24} and RS_{34}, respectively. More specifically, for I_2, $Q_{I_2}(I_4, I_4)$ is updated based on the Q value $Q_{V_3}(I_4, V_4)$ from one of its neighbor vehicles V_3, while for I_3, $Q_{I_3}(I_4, I_4)$ is updated based on the Q value $Q_{V_1}(I_4, V_2)$ from its neighbor vehicle V_1.

V. EXPERIMENTAL RESULTS

In this section, we present simulation-based evaluation results of QTAR. The performance of QTAR is compared with

those of existing protocols such as GPSR [2], LAR [31], GyTAR [18], iCar-II [7], [8] and RTAR [25]. GPSR and LAR are the classic position-based ad hoc routing protocols commonly employed as performance benchmarks. GyTAR, iCar-II and RTAR are intersection-based traffic aware routing protocols designed for urban VANETs. iCar-II is the closest to our work and is modified to have the same network infrastructure hierarchy as QTAR for a fair comparison. Therefore, the LTE eNBs in iCar-II are ignored, while the location centers are replaced by the corresponding mobility-related APIs provided in QualNet [38] to obtain the real-time coordination of all vehicle nodes. We choose QualNet as our network performance simulation platform and VanetMobiSim [39], [40] as the urban traffic generator, for which the first 1000s of output of the mobility trace were ignored to reflect real movements of vehicles.

The performance of QTAR and the corresponding comparative protocols are evaluated based on the commonly used metrics of Average Packet Delivery Ratio (APDR) and Average End-to-End Delay (AEED). In addition, we have conducted a comprehensive performance evaluation through multiple group experiments to study the impact of different parameters on these protocols. In each group experiment, all of the vehicle's movements are randomly generated through VanetMobiSim. Each of the data points presented is the average value of five experiments, with error bars indicating the 95% confidence interval. In the following, we first present simulation settings and then analyze the simulation results.

A. SIMULATION SETTINGS

The simulation urban environment scenario map is configured as in [8], the vehicles' mobility traces are generated with VanetMobiSim, but only 4-lane roads remain, as shown in Fig. 6.

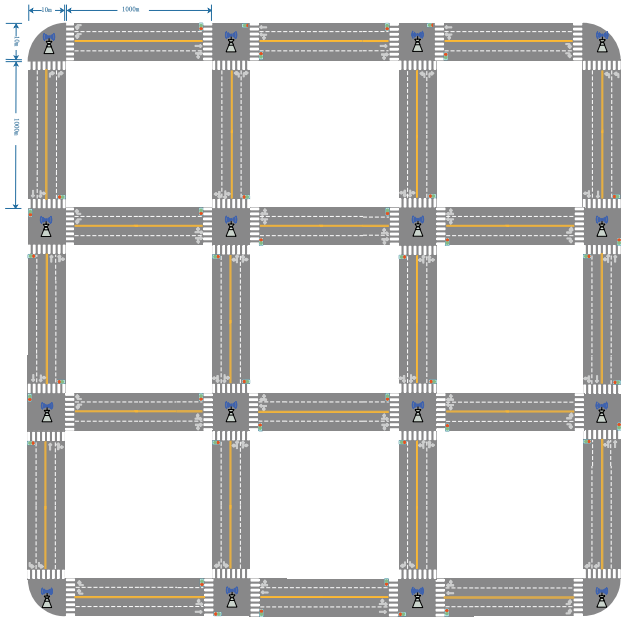


FIGURE 6. The simulation map of the urban VANET environment.

The initial location and destination of each vehicle is randomly selected, and the vehicle speed is uniformly set within the maximum allowable velocity. The data traffic patterns are generated by 20 randomly selected CBR flows. The MAC and PHY layer are configured according to the WAVE (wireless access in vehicular environments) protocol [41]. The other key simulation parameters are summarized in Table 2.

TABLE 2. Simulation parameters.

Parameter	Value
Network Simulator	QualNet(v8.2) [38]
Mobility Generator	VanetMobiSim [39], [40]
Simulation Time	600 s
Simulation Area	3040 m x 3040 m
Number of Vehicles	50 500
Allowable Maximal Velocity	5 35 m/s
Transmission Range	355 m
Number of CBR Flows	20 (Randomly Selected)
CBR Packet Interval	0.1 10 s
CBR Packet Size	512 bytes
MAC Protocol	802.11p
V2V PHY Protocol	802.11p Service Radio-SCH 172
R2R PHY Protocol	802.11p Control Radio-CCH 178
V2V Channel Bandwidth	12 Mbps
R2R Channel Bandwidth	3 Mbps
Pathloss Model	Street Microcell/LOS
Shadowing Model	Constant(4.0)
$\omega_1, \omega_2, \omega_3$	0.333,0.333,0.333
α, γ, κ	0.8,0.9,0.7

B. PERFORMANCE FOR VARYING α AND γ

In this section, we evaluated the performance sensitivity of the learning rate α and discount factor γ to obtain a good trade-off between them. We varied α and γ from 0.1 to 1 at a step size of 0.1 while fixing κ at 0.7. We also set the number of vehicles to 300, the maximum allowable velocity to 10 m/s, the number of CBR connections to 20, and the data generation interval to 1 s.

Fig. 7 shows the correlation between α and γ values and the QTAR routing performance. From Fig. 7a, we can observe that the APDR increases when α and γ increase from 0.1 to 0.8 and from 0.1 to 0.9, respectively, and then decreases as α and γ further increase to 1. This is expected because a suitable value of α can not only ensure the learning awareness of the dynamic characteristics of the urban VANETs but also achieve resistance to some local small fluctuations that can result in learning Q values with large deviations. It is worth noting that for $\gamma = 1$, the APDR drops drastically regardless of the value of α. This is because the largest value of $\gamma = 1$ denotes that future steps is considered equally and many redundant loops are learned.

Fig. 7b shows the trend of the AEED of QTAR with varying α and γ from 0.1 to 1.0. From Fig. 7b, we can see that the

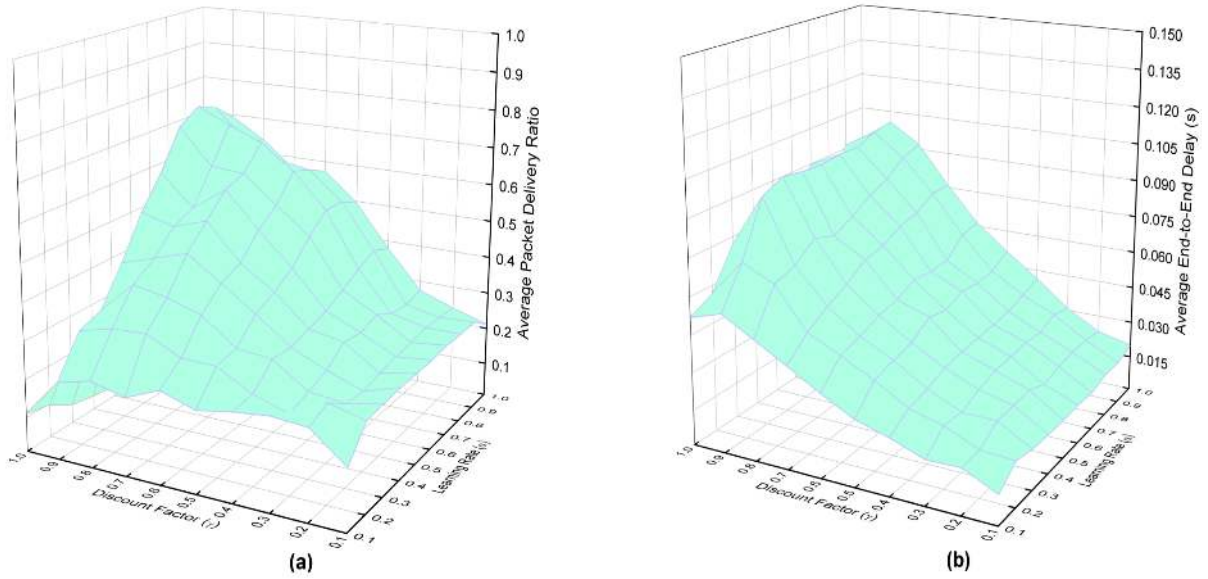


FIGURE 7. The effect of α and γ on the performance of QTAR. (a) Average packet delivery ratio. (b) Average end-to-end delay.

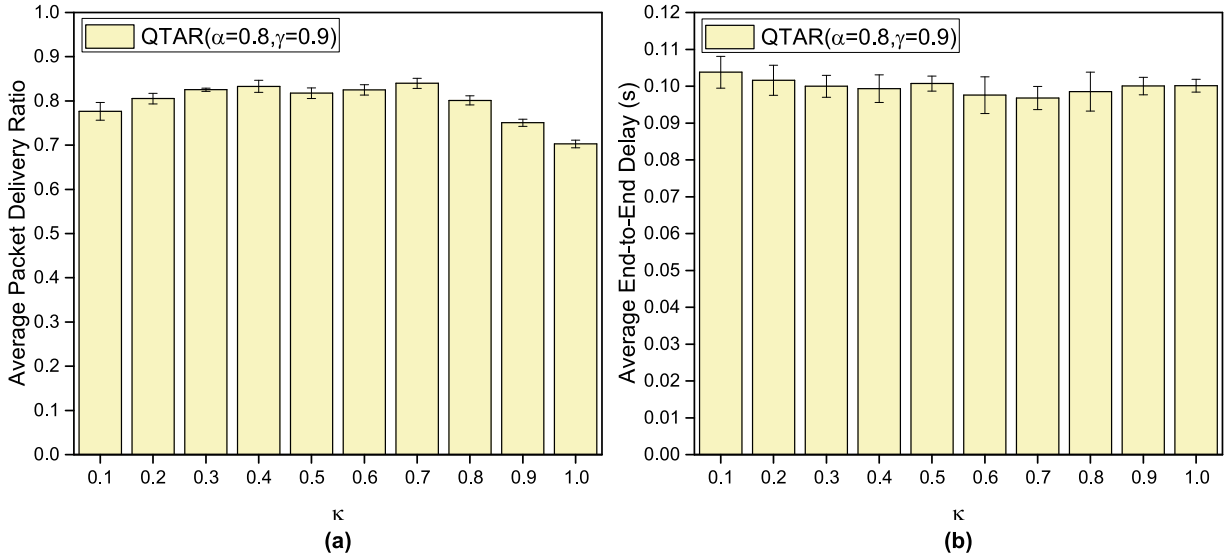


FIGURE 8. The effect of parameter κ on the performance of QTAR. (a) Average packet delivery ratio. (b) Average end-to-end delay.

AEED increases as α and γ increase in most cases. Moreover, it is interesting that APDR begins to decrease but AEED begins to increase when α increases from 0.8 to 1.0. This is because the unique characteristics of urban VANETs, such as tall concrete buildings, various trees, and vehicles of different sizes, further increase the uncertainty of learning, and a high value of α will result in a drastic change in the learning process that will further increase the likelihood of network loops occurring.

C. PERFORMANCE FOR VARYING κ

In this section, we evaluated the effect of parameter κ on the performance of QTAR. We varied κ from 0.1 to 1 at a step size of 0.1 while fixing α and γ to 0.8 and 0.9, respectively. We also set the number of vehicles to 300, the maximum allowable velocity to 10 m/s, the number of CBR connections to 20, and the data generation interval to 1. Fig. 8 depicts the trend of the QTAR routing performance with variation of the parameter κ from 0.1 to 1.0.

Fig. 8a shows the variation of APDR with κ. From Fig. 8a, we can see that the APDR increases in most cases as κ increases from 0.1 to 0.7. This is because the larger the value of κ is, the longer the optimal reward distance while the stability of the link can still be guaranteed. However, the APDR decreases to a minimum value as κ increases from 0.7 to 1.0. This means that the optimal reward distance has an increasingly significant impact on the APDR as κ varies from 0.7 to 1.0. When $\kappa = 1$, the optimal reward distance

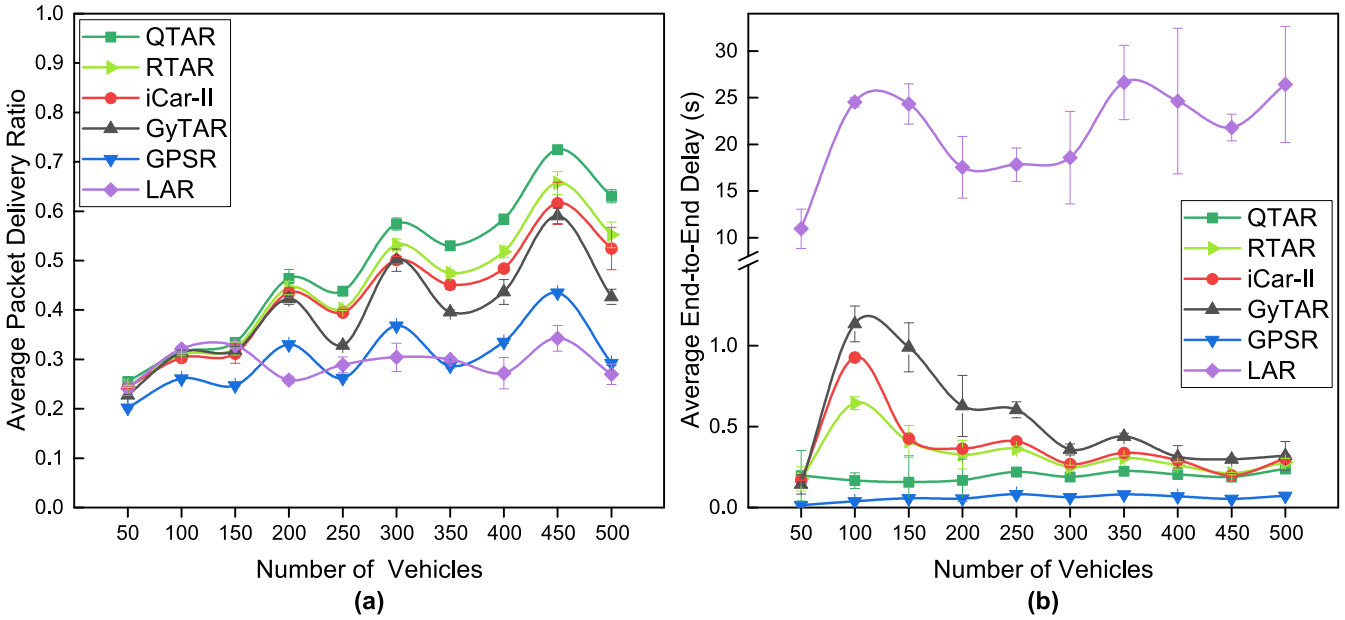


FIGURE 9. Performance of QTAR, RTAR, iCar-II, GyTAR, GPSR and LAR as the number of vehicles is varied from 50 to 500. (a) Average packet delivery ratio. (b) Average end-to-end delay.

is equal to the wireless communication range R, and the learned optimal next hop always preferentially selects the node whose distance is close to R, which will result in lower link reliability. Furthermore, $\kappa = 0.7$ improves the APDR by 19.5% (as shown in Fig. 8a) and reduces the AEED by 3.3% (as shown in Fig. 8b) compared with $\kappa = 1$. Therefore, we set κ to 0.7 in the subsequent experiments.

D. PERFORMANCE FOR VARYING DENSITY OF VEHICLES

To study the performance of the proposed QTAR and corresponding compared protocols under different node densities, in this section, we vary the number of vehicles in the network from 50 to 500 at a step size of 50. We also randomly select 20 CBR flows with the data generation interval of 1 s and the maximum allowable velocity fixed to 20 m/s, while α, γ and κ are set to 0.8, 0.9 and 0.7, respectively.

Fig. 9 demonstrates the performances of each routing protocol for the different densities of vehicles. Fig. 9a shows the APDR performance of each protocol as the number of vehicles is varied from 50 to 500. From Fig. 9a, it can be observed that the trend of the APDR of all five protocols increases in a zigzag manner as the number of vehicles increases from 50 to 450 and decreases as the number of vehicles increases from 450 to 500. This can be interpreted by the fact that the probability of the network connectivity increases with increasing number of vehicles, and the zigzag change is mainly caused by the complicated channel environment at the intersections in urban VANETs. In more detail, however, the APDR begins to decrease when the vehicle density is sufficiently high (450 or more). This is because the higher the vehicle density is, the higher the probability that a packet collision occurs in the MAC layer. In general, QTAR has a higher APDR than the other four protocols in all situations. The reason is that SCF is adopted to reduce the possibility of packet loss in the sparse vehicle density case, while QGGF is adopted in the high vehicle density case. In addition, QTAR achieves better performance in terms of APDR than that achieved by RTAR, especially in low and high vehicular density cases. This is due to the full consideration of the road traffic for each road segment in the path in QTAR. Furthermore, the RSU nodes deployed at each intersection can stably learn and distribute the traffic flow information of each road segment through V2V and R2R Q-learning, respectively. LAR and GPSR have the lowest APDR in most cases. This is because the local area flooding used by LAR cannot find an optimal path whether the vehicle density is sparse or dense, while GPSR only depends on the location information of the destination and its one-hop neighbors to find the path, which very easily falls into a local optimum, especially in the environment of urban VANETs. Overall, QTAR improves the APDR by 7.9%, 10.74% and 16.38% compared with that of RTAR, iCar-II and GyTAR, respectively.

Fig. 9b depicts the AEED performance of each protocol as the number of vehicles is varied from 50 to 500. From Fig. 9b, it can be observed that LAR has the highest AEED in all cases compared with the others because of the increasing degree of collisions and the consequent number of re-transmissions of the MAC layer incurred by the local flooding route discovery mechanism. Furthermore, the AEED of the LAR protocol also varies severely as the vehicle density increases. This is attributed to the instability of the channel condition in complex urban VANETs and the rapid fading of the signals

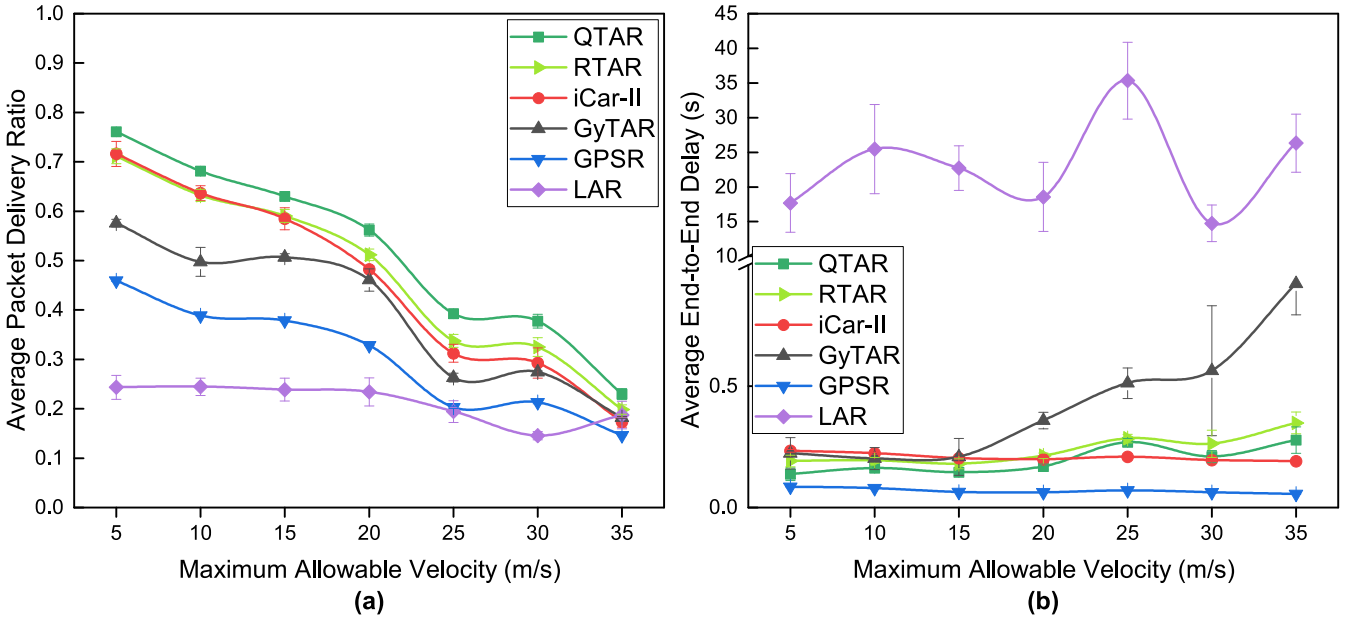


FIGURE 10. Performance of QTAR, RTAR, iCar-II, GyTAR, GPSR and LAR for varying maximum allowable velocity from 5 m/s to 35 m/s. (a) Average packet delivery ratio. (b) Average end-to-end delay.

caused by tall buildings, as well as the rapid movement of vehicles, which leads to high latency and a low delivery rate of the LAR protocol. GPSR achieves the lowest AEED at the expense of the lowest PDR because of the frequent occurrence of local optima. Regarding iCar-II and GyTAR, they have a much lower AEED than LAR in all cases. This is because packets are forwarded through intersections one by one and are dynamically selected according to the real-time traffic information on each adjacent road. In more detail, iCar-II has lower AEED than GyTAR in most configurations. This is due to the minimum one-hop transmission delay record in the CP packet of iCar-II, while only roads with higher vehicle density are preferred in GyTAR. Furthermore, RTAR achieves lower AEED in low and medium vehicular density cases than that of iCar-II because of the reliable next-hop selection scheme in the road and intersection area reliable routing. In general, QTAR achieves 30.96%, 34.78% and 46.19% lower AEED with respect to RTAR, iCar-II and GyTAR, respectively. This is mainly because QTAR not only considers the road segment forwarding delay when selecting the next hop adjacent intersection but also considers the delay from the next hop intersection to the destination intersection that the destination vehicle has just entered or is upcoming to through the R2R learning process.

E. PERFORMANCE FOR VARYING MAXIMUM ALLOWABLE VELOCITY

In this section, to evaluate the performance under different degrees of vehicle mobility, we vary the maximum allowable velocity from 5 to 35 m/s at a step size of 5 m/s while fixing the number of vehicles to 300, the number of CBR connections to 20, and the data generation interval to 1 s. α, γ and κ are set to 0.8, 0.9 and 0.7, respectively.

Fig. 10 demonstrates the performances of each protocol for varying MAV (maximum allowable velocity) from 5 m/s to 35 m/s. As shown in Fig. 10a, the APDR decreases as the MAV increases in most configurations for all six protocols. This is because an increase of the MAV will causes frequent network topology changes and increased instability of wireless link connections. In more detail, QTAR shows the best APDR performance, while LAR achieves the worst. The reason is that the R2R learning-based dynamic intersection forwarding strategy and V2V learning-based road segment forwarding in QTAR can effectively alleviate the impact of vehicle mobility on the APDR performance, while local flooding with poor mobility adaptability is the main cause of the lowest APDR performance of LAR. Moreover, RTAR shows slightly higher APDR than that of iCar-II in the high mobility case. This is mainly because RTAR selects the next forwarder based on multiple improved criteria and utilizes the traffic and network status measurement scheme for adjacent roads. Furthermore, iCar-II shows higher APDR than GyTAR and GPSR, especially when the MAV is less than 20 m/s. This is because iCar-II has global network connectivity awareness, while GyTAR and GPSR are not acutely aware of the full connected path. In general, QTAR improves the APDR by 7.92%, 14.44% and 24.24% compared with that of RTAR, iCar-II and GyTAR, respectively.

Fig. 10b depicts the AEED performance of each protocol for varying MAV from 5 m/s to 35 m/s. As shown in Fig. 10b, the AEED of each protocol basically remains constant as the MAV increases except for that of GyTAR. This is because

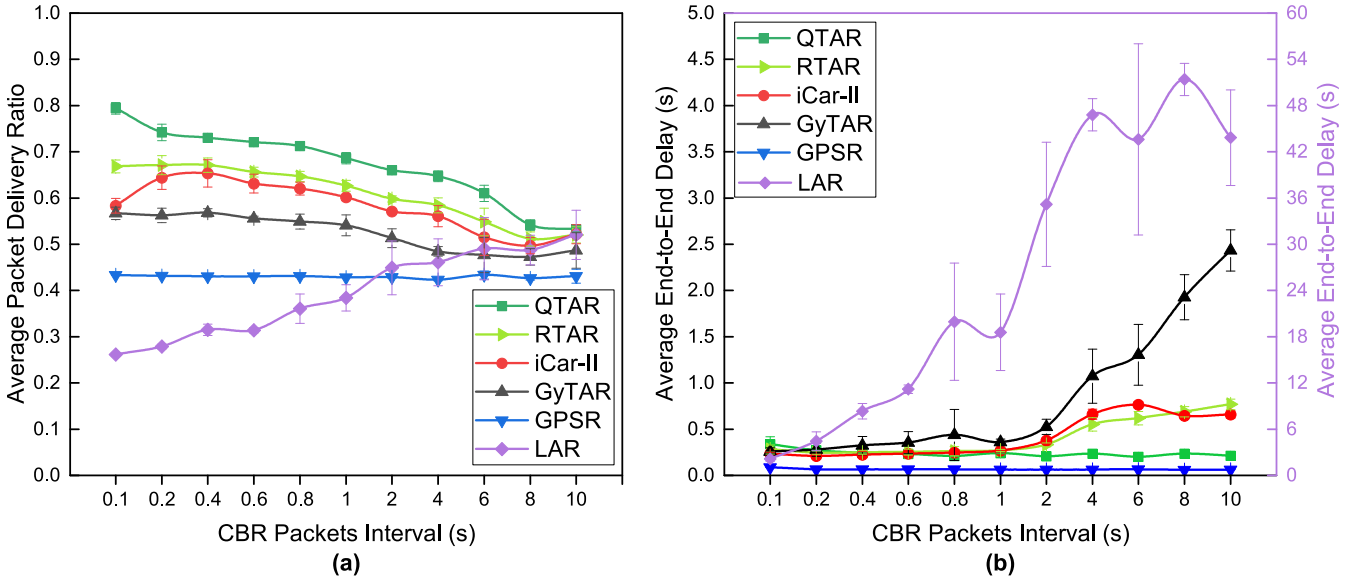


FIGURE 11. Performance of QTAR, RTAR, iCar-II, GyTAR, GPSR and LAR for varying CBR packet transmission interval from 0.1 s to 10 s. (a) Average packet delivery ratio. (b) Average end-to-end delay.

GyTAR cannot obtain the vehicle density information in time through the generation of CDP messages when the MAV exceeds 15 m/s, which will further lead to the possibility of adopting the storage-carry-forward strategy, and hence, the AEED of GyTAR increases with increasing MAV in the case of high-speed mobility. It is worth noting that GPSR shows the lowest AEED at the expense of a lower APDR because of the local optimum due to the frequency void occurrence in the complicated urban VANET environment. Furthermore, LAR has the highest AEED but still has the lowest APDR, as shown in Fig. 10a, because of the inefficient flood-based broadcast strategy. In more detail, in the case of low mobility, QTAR shows lower AEED than that of GyTAR in high mobility cases and slightly lower AEED than that of RTAR in high mobility cases. This is because the V2V learning algorithm deployed within each road segment can obtain the traffic information in a more timely manner compared with the generation of CP messages of RTAR and CDP messages of GyTAR while considering the experienced transmission delay in both the V2V and R2R learning processes. However, in the case of high mobility, the routing loop is still not completely avoided, which is why the delay of QTAR is slightly increased. On average, QTAR reduces the AEED by 4.22%, 18.68% and 45.94% compared with that of iCar-II, RTAR and GyTAR, respectively.

F. PERFORMANCE FOR VARYING CBR PACKET SEND INTERVAL

In this section, to evaluate the performance under different network payloads, we vary the data generation interval from 0.1 s to 10 s while fixing the number of vehicles to 300, the number of CBR flows to 20, and the maximum allowable velocity to 20 m/s. α, γ and κ are set to 0.8, 0.9 and 0.7, respectively.

Fig. 11 demonstrates the performance of each protocol for the different CBR transmission intervals. Fig. 11a shows the APDR performance of each protocol for varying CBR packet transmission interval from 0.1 s to 10 s. From Fig. 11a, we can see that the APDR of QTAR, RTAR and GyTAR increases as the data traffic load increases (the shorter the packet transmission interval is, the higher the data traffic load). This is because the packet delivery efficiency is improved with the increased data traffic loads. Moreover, the APDR of iCar-II increases and then decreases while that of RTAR remains almost constant (when the data transmission interval is less than 0.4 s) as the data traffic load increases. This is mainly because a large number of CP and CBR packets gradually cause channel congestion. On the other hand, the APDR of GPSR remains almost the same as the data traffic increases, while, unlike the other five protocols, the LAR APDR decreases as the data traffic load increases. This is expected because the frequency of route discovery based on flooding increases as the frequency of data transmission increases. In more detail, QTAR has a higher APDR than that of the other four protocols in all configurations. The reason is that the LET (link expired time) is considered while the R2R and V2V learning strategies jointly alleviate the effects of vehicle mobility on the APDR performance. In summary, QTAR improves the APDR by 9.85%, 12.8% and 21.14% compared with that of RTAR, iCar-II and GyTAR, respectively.

Fig. 11b shows the AEED performance of each protocol for varying CBR packet transmission interval from 0.1 s to 10 s. Since the delay of LAR is too different from the other five protocols, we considered the double Y-axis scale to

more clearly distinguish the AEED difference between them. LAR uses the Y-axis scale on the right, while the others the one on the left. From Fig. 11b, we can see that the AEED of LAR, GyTAR, iCar-II and RTAR increases as the data traffic load decreases (the shorter the packet transmission interval is, the higher the data traffic load). For LAR, this is because the longer the interval is, the more times route discovery is required, which increases the latency of the packet. For RTAR, GyTAR and iCar-II, this can be explained by the fact that most of the data packets are transmitted to their destination with a relatively low delay in the high data traffic load case in contrast to most of the packets having a relatively high delay in the low data traffic load case because of the contention of the MAC layer. The AEED of GPSR and QTAR remains the same in all configurations because routing information is updated by periodically broadcasting HELLO packets, and therefore, they are independent of data traffic conditions. In particular, QTAR achieves lower AEED than that of RTAR, GyTAR and iCar-II, especially in low data traffic load cases. This is because the road traffic learning process in QTAR is more adaptable to the dynamic urban environment. In more detail, QTAR integrates the traffic information into Q values by combining the V2V learning strategy within the road segments and the R2R learning strategy between the intersections, while iCar-II and GyTAR only consider the adjacent road segment traffic. In summary, QTAR reduces the AEED by 28.54%, 29.41% and 50.25% compared with that of RTAR, iCar-II and GyTAR, respectively.

VI. CONCLUSION AND FUTURE WORK

We have proposed a novel RSU-assisted Q-learning-based Traffic-Aware Routing (QTAR) protocol that improves the urban VANETs comprehensive routing performance through optimized Q-greedy geographical forwarding based on V2V Q-learning within the road segments and intersection forwarding based on R2R Q-learning. Simulation evaluation results have demonstrated that QTAR outperforms other existing related routing protocols in terms of a higher packet delivery ratio in sparse and dense traffic cases and a reduced packet delivery delay, with a negligible communication overhead in moderate traffic cases.

To further refine QTAR, our future work will mainly focus on the following aspects. First, an in-depth analysis of some key protocol parameters for adaption to more complex VANETs scenarios will be considered. Second, dynamically selecting anchor vehicle nodes at each intersection to remove the dependence on RSUs for packet forwarding at intersections can also be considered. Finally, in order to better adapt to the inconsistency of the length of the road segments in urban VANETs, we will consider merging multiple shorter and narrower road segments and deploying RSU nodes only at certain critical intersections for these crowded roads. At the same time, for long-length and spacious road segments, we will consider splitting these long and wide road segments into multiple shorter ones to adapt to the rapidly changing cases when traffic is sparse and the congestion cases when traffic is dense.

ACKNOWLEDGMENT

The authors appreciate the editors and anonymous reviewers for their helpful review comments and suggestions.

REFERENCES

[1] S. Al-Sultan, M. M. Al-Doori, A. H. Al-Bayatti, and H. Zedan, "A comprehensive survey on vehicular Ad Hoc network," *J. Netw. Comput. Appl.*, vol. 37, pp. 380–392, Jan. 2014.

[2] 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.

[3] C. Lochert, M. Mauve, H. Füler, and H. Hartenstein, "Geographic routing in city scenarios," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 9, no. 1, pp. 69–72, 2005.

[4] C. Lochert, H. Hartenstein, J. Tian, H. Fussler, D. Hermann, and M. Mauve, "A routing strategy for vehicular ad hoc networks in city environments," in *Proc. IEEE Intell. Vehicles Symp.*, Mar. 2003, pp. 156–161.

[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] 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.

[7] N. Alsharif and X. S. Shen, "iCARII: Intersection-based connectivity aware routing in vehicular networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 2731–2735.

[8] N. Alsharif and X. Shen, "iCAR-II: Infrastructure-based connectivity aware routing in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 4231–4244, Aug. 2017.

[9] Y. R. B. Al-Mayouf, N. F. Abdullah, O. A. Mahdi, S. Khan, M. Ismail, M. Guizani, and S. H. Ahmed, "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.

[10] G. Li, L. Boukhatem, and S. Martin, "An intersection-based QoS routing in vehicular ad hoc networks," *Mobile Netw. Appl.*, vol. 20, no. 2, pp. 268–284, 2015.

[11] T. Darwish and K. A. Bakar, "Lightweight intersection-based traffic aware routing in Urban vehicular networks," *Comput. Commun.*, vol. 87, pp. 60–75, Aug. 2016.

[12] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 279–292, 1992.

[13] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, vol. 1. Cambridge, MA, USA: MIT Press, 2016.

[14] B.-C. Seet, G. Liu, B.-S. Lee, C.-H. Foh, K.-J. Wong, and K.-K. Lee, "A-STAR: A mobile ad hoc routing strategy for metropolis vehicular communications," in *Proc. Int. Conf. Res. Netw.* Berlin, Germany: Springer, 2004, pp. 989–999.

[15] J. Zhao and G. Cao, "VADD: Vehicle–assisted data delivery in vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 3, pp. 1910–1922, May 2008.

[16] Y. Ding, C. Wang, and L. Xiao, "A static-node assisted adaptive routing protocol in vehicular networks," in *Proc. 5th ACM Int. Workshop Veh. Ad Hoc Netw.*, 2007, pp. 59–68.

[17] V. Naumov and T. R. Gross, "Connectivity-aware routing (CAR) in vehicular ad-hoc networks," in *Proc. INFOCOM 26th IEEE Int. Conf. Comput. Commun.*, 2007, pp. 1919–1927.

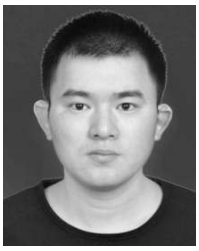
[18] M. Jerbi, S.-M. Senouci, T. Rasheed, and Y. Ghamri-Doudane, "Towards efficient geographic routing in urban vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 58, no. 9, pp. 5048–5059, Nov. 2009.

[19] J. Nzouonta, N. Rajgure, G. Wang, and C. Borcea, "VANET routing on city roads using real–time vehicular traffic information," *IEEE Trans. Veh. Technol.*, vol. 58, no. 7, pp. 3609–3626, Sep. 2009.

[20] J.-J. Chang, Y.-H. Li, W. Liao, and I.-C. Chang, "Intersection-based routing for urban vehicular communications with traffic-light considerations," *IEEE Wireless Commun.*, vol. 19, no. 1, pp. 82–88, Feb. 2012.

[21] 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.

[22] X. Zhang, Z. Wang, and X. Jiang, "A realistic spatial-distribution-based connectivity–aware routing protocol in multilevel scenarios of urban VANETs," *IEEE Commun. Lett.*, vol. 22, no. 9, pp. 1906–1909, Sep. 2018.
[23] 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.
[24] 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.
[25] T. S. 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.
[26] J. A. Boyan and M. L. Littman, "Packet routing in dynamically changing networks: A reinforcement learning approach," in *Proc. Adv. Neural Inf. Process. Syst.*, 1994, p. 671.
[27] J. Dowling, E. Curran, R. Cunningham, and V. Cahill, "Using feedback in collaborative reinforcement learning to adaptively optimize MANET routing," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 35, no. 3, pp. 360–372, May 2005.
[28] W. Celimuge, K. Kumekawa, and K. Toshihiko, "Distributed reinforcement learning approach for vehicular ad hoc networks," *IEICE Trans. Commun.*, vol. 93, no. 6, pp. 1431–1442, 2010.
[29] 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.
[30] 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.
[31] Y.-B. Ko and N. H. Vaidya, "Location-aided routing (LAR) in mobile ad hoc networks," *Wireless Netw.*, vol. 6, no. 4, pp. 307–321, 2000.
[32] C. Perkins, E. Belding-Royer, and S. Das, *Ad Hoc on Demand Distance Vector (AODV) Routing*, document RFC 3561, IETF MANET Working Group, 2003.
[33] T. Clausen and P. Jacquet, *Optimized Link State Routing Protocol (OLSR)*, document RFC 3626, IETF, Oct. 2003.
[34] D. Johnson, Y. Hu, and D. Maltz, *Dynamic Source Routing Protocol (DSR) for Mobile Ad Hoc Networks for IPv4*, document RFC 4728 and 2070-1721, 2007.
[35] J. Li, J. Jannotti, D. S. De Couto, D. R. Karger, and R. Morris, "A scalable location service for geographic ad hoc routing," in *Proc. 6th Annu. Int. Conf. Mobile Comput. Netw.*, 2000, pp. 120–130.
[36] C. Wu, Y. Ji, F. Liu, S. Ohzahata, and T. Kato, "Toward practical and intelligent routing in vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 12, pp. 5503–5519, Dec. 2015.
[37] T. Taleb, E. Sakhaee, A. Jamalipour, K. Hashimoto, N. Kato, and Y. Nemoto, "A stable routing protocol to support ITS services in VANET networks," *IEEE Trans. Veh. Technol.*, vol. 56, no. 6, pp. 3337–3347, Nov. 2007.
[38] Scalable Network Technologies. (2018). *Network Simulator*. [Online]. Available: https://web.scalable-networks.com/qualnet-network-simulator-software
[39] J. Harri and M. Fiore, "VanetMobiSim–vehicular ad hoc network mobility extension to the CanuMobiSim framework," *Inst. Eurécom Dept. Mobile Commun.*, vol. 6904, pp. 1–19, 2006.
[40] J. Härri, M. Fiore, F. Filali, and C. Bonnet, "Vehicular mobility simulation with VanetMobiSim," *Simulation*, vol. 87, no. 4, pp. 275–300, Apr. 2011.
[41] D. Jiang and L. Delgrossi, "IEEE 802.11 P: Towards an international standard for wireless access in vehicular environments," in *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, May 2008, pp. 2036–2040.



JINQIAO WU received the M.S. degree from the Xi'an University of Post and Telecommunications, Xi'an, China, in 2014. He is currently pursuing the Ph.D. degree in computer science with Xidian University, Xi'an. His research interests include machine learning, networking architectures, and routing protocols.

MIN FANG received the B.S. degree in computer control, the M.S. degree in computer software engineering, and the Ph.D. degree in computer application from Xidian University, Xi'an, China, in 1986, 1991, and 2004, respectively. She is currently a professor with Xidian University. Her research interests include intelligent information process, multiagent systems, and network technology.

HAIKUN LI received the M.S. degree in computer software and theory form Yunnan University, Kunming, China, in 2017. He is currently pursuing the Ph.D. degree in computer application with Xidian University, Xi'an, China. His research interests include pattern recognition, machine learning, and transfer learning.

XIAO LI received the B.S. degree from the Xi'an University of Finance and Economics, Xi'an, China, in 2012. She is currently pursuing the Ph.D. degree in computer science with Xidian University, Xi'an, China. Her research interests include pattern recognition, machine learning, and computer vision.