IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 25, 2019, accepted January 4, 2020, date of publication January 7, 2020, date of current version January 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2964707

Edge-Based V2X Communications With Big Data Intelligence

SIRI GULENG[1], CELIMUGE WU[2], (Senior Member, IEEE), ZHI LIU[3], (Senior Member, IEEE), AND XIANFU CHEN[4], (Member, IEEE)

[1]Hohhot Minzu College, Hohhot 010051, China
[2]Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo 182-8585, Japan
[3]Department of Mathematical and Systems Engineering, Shizuoka University, Johoku Hamamatsu 432-8561, Japan
[4]VTT Technical Research Centre of Finland, 90571 Oulu, Finland

Corresponding author: Celimuge Wu (celimuge@uec.ac.jp)

This work was supported in part by the Inner Mongolia Science and Technology Major Project, China, in part by the JST-Mirai Program Grant JPMJMI17B3, in part by the Telecommunications Advanced Foundation, and in part by the JSPS KAKENHI Grant 18KK0279 and Grant 19H04093, Japan.

ABSTRACT Vehicular Internet-of-Things applications require an efficient Vehicle-to-Everything (V2X) communication scheme. However, it is particularly challenging to achieve a high throughput and low latency with limited wireless resources in highly dynamic vehicular networks. In this article, we propose a scheme that enhances V2V communications through integration of vehicle edge-based forwarding and learning-based edge selection policy optimization. The proposed scheme has three main characteristics. First, the *Hierarchical edge-based preemptive route creation* is introduced to create hierarchical edges and conduct efficient packet forwarding as well as route aggregation. Second, *Two-stage learning* is introduced to select efficient edge nodes using big data driven traffic prediction and reinforcement learning-based edge node selection. Third, *Context-aware edge selection* is employed to improve the performance of edge-based forwarding in various contexts. We use real traffic big data and realistic vehicular network simulations to evaluate the performance of the proposed scheme and show the advantage over other baseline approaches.

INDEX TERMS Edge computing, traffic big data, vehicular ad hoc networks, V2X communications.

I. INTRODUCTION

Vehicle-to-everything (V2X) communications are required for emerging Internet-of-Things (IoT) applications, such as intelligent transport systems, autonomous driving, collision avoidance systems and so on [1]–[3]. Two main challenges exist in enforcing vehicular IoT. First, the mobility of vehicles results in a spatial change for network elements including communication node, available wireless resources and network density. Second, the temporal change of network environment makes the problem particularly complex. As vehicular IoT systems involve multiple types of network nodes and different wireless communication approaches, the network environment could change upon time frequently, and therefore we have to design a more intelligent communication protocol that has self-evolving capability. To cope with these challenges, we need a new scheme that takes into account the efficiency of V2X communications when facing various spatial and temporal dynamics, including vehicle mobility, vehicle distribution, and resource limitations. Meanwhile, the scheme should be able to conduct efficient prediction of environment and adapt quickly to the change of environment.

Three different types of communications are considered as the main technologies supporting V2X communications. First, as the default standard for V2V communications, IEEE 802.11p would continue to show its importance in V2X communications. Second, by using high frequency spectrum, millimeter wave (mmWave) communication [2] is able to provide up to several Gbps throughput, which could be an emerging solution to achieve high data rate information exchange between vehicles in vicinity. Third, cellular communication operating at Sub-6 GHz has special advantages over other approaches as it can provide long range Internet connectivity to the vehicles. Although mmWave communications could achieve high data rate, it only works in a line-of-sight

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

environment. While IEEE 802.11p supports non-line-of-sight scenarios, the throughput is limited. It is also unpreferable to rely fully on cellular communications as it could be difficult for supporting high-density scenarios, or too expensive in terms of infrastructure development. Moreover, in a collision avoidance system that requires broadcasting of safety messages, IEEE 802.11p is the best candidate. This is because the mmWave communication technology is only helpful for unicast applications, and cellular communications could increase the latency due to some signaling overheads. Therefore, a seamless integration of these three communication technologies would be a new trend in V2X communications.

In recent studies, many efforts have been made to solve the performance degradation problem in mobile and high-density scenarios. These studies include vehicular routing protocols [4], broadcasting protocols [5], an integration of different wireless spectrums [6], and clustering [7]. Recent studies on the temporal challenges include the prediction of vehicular communication channel [8], and online adaptive routing protocols [9]. However, none of these studies addresses the use of vehicular big data in the online decision making process. A real-time collection of vehicle information requires a large amount of wireless resources, which is unrealistic. Moreover, since the network topology changes with the vehicle movement, a simple collection of sensor information cannot meet the application requirement. As a result, a prediction is always required to provide a high quality-of-service (QoS) for V2X communications.

Mobile edge computing (MEC) has been attracting special interests in recent years due to its capability of conducting data caching and computing in the vicinity of end users [10], [11]. Although there have been many studies discussing about the advantages of using MEC in vehicular networks, most of them discuss how to conduct computation offloading and data caching at edge nodes, and do not seriously address how to use vehicles as edge nodes to improve the communication performance. In terms of the network layer perspective, V2X communications can be broadcast or unicast, which have different characteristics and require different levels of QoS. The context information, such as the communication type, should be considered in the edge node selection.

In this paper, we propose an intelligent edge-based scheme for V2X communications. First, we introduce a hierarchical vehicle edge-based architecture to solve the high-density and mobility issue in vehicular networks where some vehicles are used as edge nodes to improve the routing performance. The edge nodes are selected by considering vehicle velocity, vehicle distribution, and signal quality based on a fuzzy logic algorithm using a decentralized approach. We then use a deep neural network-based approach to predict the average vehicle velocity and traffic density. The prediction is conducted at the roadside unit (RSU) or cloud, and the prediction result is used to initialize the edge selection algorithm at each vehicle. Each vehicle employs a reinforcement learning to further tune the fuzzy parameters by interacting with the environment, which ensures that the proposed scheme is adaptable to the frequent change of vehicular environment. The proposed scheme also chooses different edge nodes for broadcast and unicast applications in order to consider different QoS requirements of these applications. This paper is an extension of our previous conference paper [12]. While [12] only discusses a multi-hop broadcast problem, this work generalizes the problem by including V2X unicast communication problem and also presents new simulation results.

The main contributions are briefly summarized in the following.

- We propose a V2X communication scheme that integrates vehicle edge-based forwarding and learning-based edge selection policy optimization.
- *Hierarchical edge-based preemptive route creation* is introduced to create hierarchical edges and conduct efficient packet forwarding.
- *Two-stage learning* is proposed to select efficient edge nodes using big data driven traffic prediction and reinforcement learning-based edge node selection.
- *Context-aware edge selection* is employed to improve the performance of edge-based forwarding according to different application requirements.
- We conduct exhaustive simulations to show the advantage of the proposed scheme over other baseline approaches solutions for both broadcast communications and unicast communications.

The remainder of this article is organized as follows. A brief introduction of related work is given in Section II. We describe our scheme in Section III. Simulation results are presented in Section IV. Finally, we point out future research directions and then draw our conclusions in sections V and VI, respectively.[1]

II. RELATED WORK

From the perspective of networking, the existing studies on V2X communications can be classified into two types, namely, broadcast protocols and unicast protocols. Broadcast protocols focus on solving the multi-hop information dissemination between vehicles and vehicles/infrastructures. The studies on unicast communications try to solve whether spatial challenges or temporal challenges for V2X communications.

A. BROADCAST PROTOCOLS

Vehicular Internet-of-Things applications [13], [14] require efficient broadcast communications where a traffic alert message should be delivered to multi-hop distances. However, due to the uncertainty of the receivers, the design of a multi-hop broadcast protocol is challenging [15]. Existing studies on multi-hop broadcast protocols are whether sender-oriented or receive-oriented. In the sender-oriented approaches, the sender nodes specify the next forwarder nodes [5], [16], [17]. In order to ensure efficient packet

[1] We use the words "vehicle" and "node" interchangeably in the rest of this paper.

forwarding, the forwarder node selection algorithm should consider vehicle mobility, signal quality, and multi-hop forwarding efficiency, which is a complicated problem.

In the receiver-oriented protocols, a packet forwarding decision is made at the receiver side [18]–[21]. The simplest receiver-based approach is to conduct packet rebroadcasting according to a certain probability. However, the choice of forwarding probability is a difficult problem as a low probability fails to provide a good reliability and a large forwarding probability makes the message overhead high. As an enhanced approach, the inter-vehicle distance and node density could be considered in the forwarding probability determination. In this paper, we will show that based on an accurate prediction of vehicle density, we can achieve much better multi-hop broadcasting. Meanwhile, some geocast protocols use position information to improve the performance of multi-hop broadcasting [22]–[24]. However, a location service is required to know the position of other vehicles, which is not seriously considered in the existing studies.

B. SPATIAL CHALLENGES FOR V2X COMMUNICATIONS

Spatial challenges in vehicular networks include mobility and high-density issue. Due to the mobility of vehicles, the route selection and quality-of-service guarantee problem become more difficult. Recently, mobile edge computing (MEC) based approaches attract great interest in solving the performance problem in dynamic vehicular networks by caching the content near the vehicles or conducting some processing in vicinity [25]–[31].

A complete survey on the recent studies about edge computing has been conducted by Wang et al. [28]. In addition to the definition, architecture, and advantages of MEC, [28] also describes the research issues on computing, caching, and communication techniques at the edge. The key enablers, possible applications, main technologies, and existing problems are also discussed. Liu et al. [29] have conducted a review on the mobile edge computing for vehicular networks in terms of system architecture, interesting applications, and security issues. Khattak et al. [30] have introduced an intelligent transport light control system where road-side-units are in charge of conducting computation and generating notifications based on edge computing. In [31], the edge computing and software defined networking technology are integrated to support network functions visualization in order to enable more enhanced services in vehicular networks. Computing tasks can be processed whether on the edge servers or cloud servers based on the computing, storage, and bandwidth resources.

A highly dense distribution of vehicles limits the available bandwidth for each vehicle, resulting in a difficulty in providing good performance. There have been some studies on handling the high-density issue in vehicular networks. In [32], a backbone approach has been proposed to reduce the MAC layer contention delay by using the same backbone nodes for different traffic flows. Since the backbone approach can reduce the number of sender nodes, a significant reduction of delay can be expected. Chen et al. [33] have proposed a geographic routing protocol based on a similar backbone approach. In [33], a backbone link composed of a series of backbone nodes is built in each road segment. Then, each road segment is evaluated by a weight considering the real-time link information and historical traffic information, which is used to select the best routing path.

C. TEMPORAL CHALLENGES FOR V2X COMMUNICATIONS

The studies on temporal challenges in vehicular networks cover two main research directions. The first type of study has been focusing on some predictions based on existing big data. The other type of research has put efforts into the use of online learning algorithms to improve the adaptability of communication protocols in a frequently changing network environment. Lin et al. [34] have proposed a social data based localization algorithm (SBL) for vehicular networks. They classified the vehicles into different social clusters by exploring the social relationship between them. Liu and Shoji [35] have proposed an edge-assisted vehicle mobility prediction approach that adopts hybrid architecture of convolutional and recurrent neural networks. However, in these studies, the use of prediction information in V2X communications is not discussed adequately.

In the recent years, the use of deep reinforcement learning in the resource management for vehicular networks has attracted a great interest. Chen et al. [36] have discussed the radio resource management problem for vehicular networks, and proposed an online algorithm to handle the difficulty incurred from the high-vehicle mobility and data traffic variations. In [37], the authors proposed a deep deterministic policy gradient approach to solve the joint optimization problem of content placement and content delivery in vehicular edge computing. These works do not discuss how to utilize prediction information in the resource allocation. Therefore, in this paper, we will discuss how to efficiently integrate the big data driven intelligence and online intelligence.

III. EDGE-BASED V2X COMMUNICATIONS WITH BIG DATA INTELLIGENCE

A. OVERVIEW OF THE PROPOSED SCHEME

We assume that each vehicle is equipped with a positioning device and three wireless interfaces, namely, cellular, IEEE 802.11p and mmWave. The proposed scheme first uses a hierarchical edge based forwarding approach to handle the spatial challenge. A decentralized clustering of vehicles is conducted in the approach, and the cluster head vehicles are used as edge nodes to improve the communication performance in a highly mobile or dense environment. We then tackle the temporal problem from two perspectives. First, big data based traffic prediction is conducted to estimate the traffic density and vehicle velocity, which can be used to enable a better clustering (better edge node selection). Second, an online reinforcement learning approach is used to adjust the fuzzy parameters in order to adapt quickly to the change of environment, ensuring the self-evolvability of the proposed scheme. Finally, we introduce a context-aware edge node selection approach that uses different criteria for broadcast and unicast communications which have totally different QoS requirements. On the network layer, an edge-based approach is used to achieve efficient forwarding, and on an upper layer, the number of edge node is adjusted based on the application context.

B. HIERARCHICAL EDGE-BASED PREEMPTIVE ROUTE CREATION

We introduce a *hierarchical edge-based preemptive route creation* approach to facilitate the packet forwarding process. First, we use two-tier edge nodes to enhance the packet forwarding process. As shown in Figure 1, the tier-1 edge nodes are connected to the Internet with cellular interface. Each tier-2 edge node is connected to a tier-1 edge node through mmWave communications. The tier-2 edge nodes are used to provide the end users with high throughput connectivity through multi-hop mmWave communications. By connecting tier-1 edge nodes, a backbone for data delivery is constructed, and the backbone is used to create routes or aggregate different communication flows. The packet forwarding process at the tier-1 edges is as follows: if the current node is a neighbor of the destination node, then the node transmits the packet directly to the destination node; else, the packet will be delivered to the tier-1 node that is closer to the destination node than the current node (see Algorithm 1 and Algorithm 2). By using hierarchical edges, different wireless spectrums, namely, cellular, IEEE 802.11p and mmWave, can be efficiently integrated. Algorithm 1 and Algorithm 2 conduct packet forwarding based on the hierarchical edge structure, which improves the possibility of conducting caching and computing at the edge nodes. The main purpose of using edge nodes in this paper is to achieve efficient packet forwarding by improving the wireless spectrum usage efficiency, and aggregating different traffic flows. The computing and caching can be done at the edge nodes to further improve the performance. However, the computing task offloading and data caching strategies are outside the scope of this paper. This paper focuses on how to select the edge nodes in a multi-access environment.

Before a packet send request is initialized, the packet forwarding path is basically determined. Therefore, we call this scheme as *hierarchical edge-based preemptive route creation*. As the same backbone is used for different traffic flows initiated from different source nodes, route aggregation is also possible by using the same tier-1 edge nodes for packet delivery. The *preemptive route creation* is different from proactive routing protocols in that it does not require creating routes between all possible communication pairs in advance, resulting in much lower communication and management overheads.

Algorithm 1 Algorithm at a Vehicle for Sending a Packet to the Cloud

```
Check the status of the current node.
if The current node is a tier-1 edge node then
    Directly send the packet to the cloud.
    Exit.
end if
if The current node is a tier-2 edge node then
    Send the packet to the tier-1 edge node that the current
    has a direct access.
    Exit.
end if
if The current node is an ordinary node then
    Send the packet to the tier-2 edge node (or tier-1 edge
    node) that the current has a direct access.
    Exit.
end if
```

The selection of edge nodes is as follows. The tier-1 edge nodes are created by exchanging hello messages among neighbors. Based on the clustering method of [9], each node evaluates its suitability of a tier-1 edge node by using a fuzzy logic-based approach. The vehicle velocity, vehicle distribution (leadership factor), and signal quality are calculated for each one-hop neighbor that is within a reference distance where this reference distance is dependent on the types of

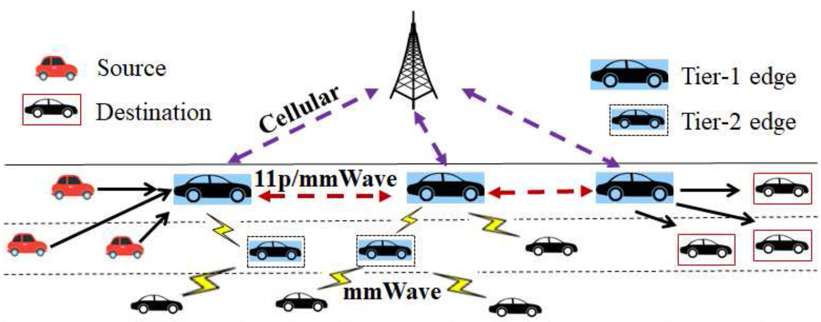

FIGURE 1. Two-tier vehicle edges.

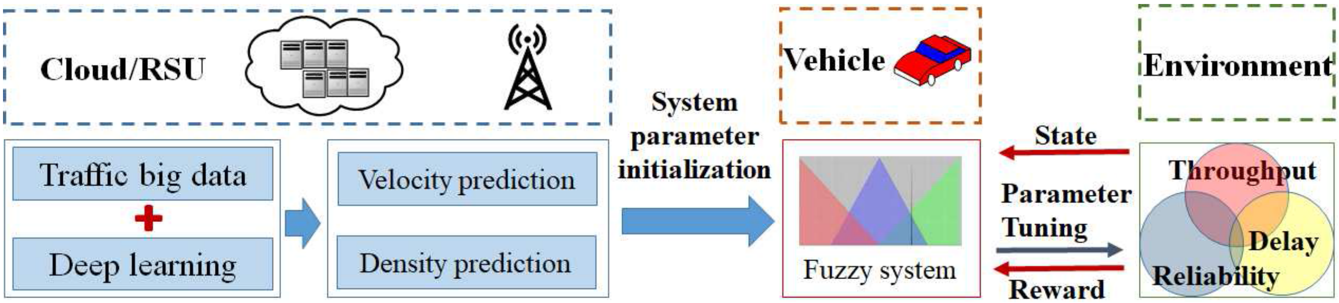

FIGURE 2. Two-stage learning-based parameter tuning.

communications.[2] If a node has the highest suitability value, the node will be a tier-1 edge node. If a node is not a tier-1 edge and the node has a direct mmWave communication link to a tier-1 node, then the node would be a tier-2 edge node.

C. TWO-STAGE LEARNING

Spatial challenges in V2X communications mainly come from high vehicle density or velocity. Temporal challenges incur from the difficulty in achieving a satisfactory performance due to the frequent changes of environment. An accurate estimation of vehicle density and velocity information could significantly improve the packet forwarding performance. However, it is difficult to collect this information in a real time manner due to the high communication cost. In order to solve these problems, we propose a two-state learning approach to enhance the V2X communications. As shown in Figure 2, based on traffic big data, deep neural networks are used to predict the vehicle velocity and density information. The prediction is conducted at the RSU or cloud, and the information is used to initialize the fuzzy parameters for the hierarchical edge selection process at each vehicle. Each vehicle then employs a reinforcement learning (Q-learning) algorithm to tune the fuzzy parameters online by evaluating the reward from the environment [38]. The deep neural network-based prediction approach could accurately estimate the vehicle density and velocity without incurring a high communication overhead. The reinforcement learning-based parameter tuning approach could improve the adaptability of the edge-based forwarding mechanism, making it self-evolvable.

The fuzzy logic-based approach for the edge selection jointly considers the stability factor, topology factor, and connectivity factor. These factors are defined as follows.

1) STABILITY FACTOR

Stability factor of node x is calculated as follows.

$$SF(x) = 1 - \frac{||\upsilon(x)| - \mathrm{avg}_{y\in N_x}|\upsilon(y)||}{\max_{y\in N_x}|\upsilon(y)|} \tag{1}$$

[2]Without any specific explanation, the reference distance is set as the half of the transmission range; see Subsection III-D for details.

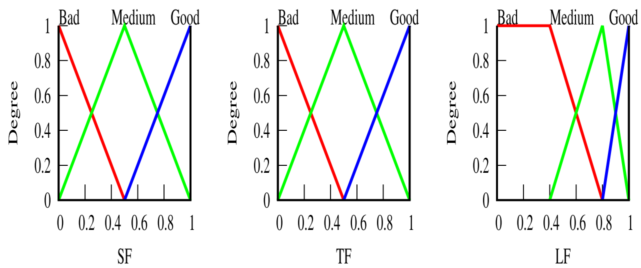

FIGURE 3. Fuzzy membership functions (left: ***SF***, middle: ***TF***, right: ***LF***).

where a higher value means a higher stability. Since hello messages are exchanged between neighbors, each vehicle can calculate its neighbors' SF. Here, $\mathrm{avg}_{y\in N_x}|\upsilon(y)|$ is predicted from the traffic big data. *SF* is updated periodically (one second interval) based on a weighted exponential moving average with smooth factor of 0.7.

2) TOPOLOGY FACTOR

Topology factor is calculated as follows.

$$TF(x) = min\left(1, \frac{c(x)}{\text{Number of neighbor vehicles}}\right) \tag{2}$$

where $c(x)$ denotes the number of vehicles traveling to the same direction as the node x in its neighbors. The higher the number, the higher chance the vehicle be elected as an edge head. Here, the number of vehicles in one-hop region is acquired from the traffic big data based prediction of vehicle density. *TF* is updated periodically (one second interval) based on a weighted exponential moving average with smooth factor of 0.7.

3) LINK FACTOR

Connectivity factor can be calculated by the ratio of "the number of hello messages received from all one-hop neighbors" to "the number of hello messages sent by all one-hop neighbors" as

$$LF(x) = \frac{\text{Num of hellos received from all NBs}}{\text{Num of hellos sent by from all NBs}}. \tag{3}$$

The fuzzy membership functions for the factors are defined as shown in Figure 3. Based on the fuzzy membership

Algorithm 2 Algorithm at a Vehicle for Sending a Packet to Another Vehicle

Check the status of the current node.
if The current node is a tier-1 edge node **then**
  if The destination vehicle is connected to the current node directly **then**
    Send the packet to the destination vehicle directly.
  else
    if The destination vehicle is connected to the current node through a tier-2 node **then**
      Send the packet through the tier-2 node.
    else
      Send the packet to a neighbor tier-1 edge node that is closer to the destination vehicle.
    end if
  end if
end if
if The current node is a tier-2 edge node **then**
  if The destination vehicle is connected to the current node directly **then**
    Send the packet directly to the destination vehicle.
  else
    Send the packet to the tier-1 edge node it connects.
  end if
end if
if The current node is an ordinary node **then**
  if The destination vehicle is connected to the current node directly **then**
    Send the packet directly to the destination vehicle.
  else
    Send the packet to the tier-2 edge node (or tier-1 edge node) it connects.
  end if
end if

TABLE 1. Fuzzy rules.

	Stability	Topology	Link	Rank
Rule1	Good	Good	Good	Perfect
Rule2	Good	Good	Medium	Good
Rule3	Good	Good	Bad	Unpreferable
Rule4	Good	Medium	Good	Good
Rule5	Good	Medium	Medium	Acceptable
Rule6	Good	Medium	Bad	Bad
Rule7	Good	Bad	Good	Unpreferable
Rule8	Good	Bad	Medium	Bad
Rule9	Good	Bad	Bad	VeryBad
Rule10	Medium	Good	Good	Good
Rule11	Medium	Good	Medium	Acceptable
Rule12	Medium	Good	Bad	Bad
Rule13	Medium	Medium	Good	Acceptable
Rule14	Medium	Medium	Medium	Unpreferable
Rule15	Medium	Medium	Bad	Bad
Rule16	Medium	Bad	Good	Bad
Rule17	Medium	Bad	Medium	Bad
Rule18	Medium	Bad	Bad	VeryBad
Rule19	Bad	Good	Good	Unpreferable
Rule20	Bad	Good	Medium	Bad
Rule21	Bad	Good	Bad	VeryBad
Rule22	Bad	Medium	Good	Bad
Rule23	Bad	Medium	Medium	Bad
Rule24	Bad	Medium	Bad	VeryBad
Rule25	Bad	Bad	Good	Bad
Rule26	Bad	Bad	Medium	VeryBad
Rule27	Bad	Bad	Bad	VeryBad

functions and the fuzzy rule defined in Table 1, the fitness value for each vehicle being an edge node can be calculated. The output membership function is the same as [38]. A vehicle that has a larger fitness value gets a higher chance for

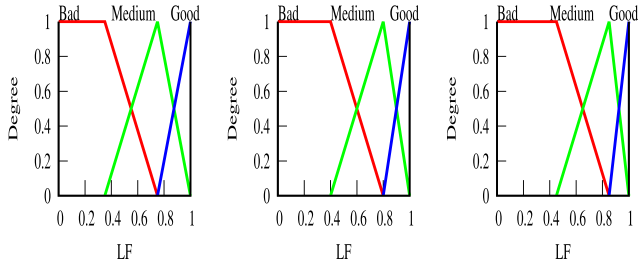

FIGURE 4. Action (left: Move left, middle: Maintain, right: Move right).

the edge selection. The number of edge node is dependent on the type of applications, which will be explained in the next subsection. In order to design an approach that could work for various situations, we also use a Q-learning algorithm to tune fuzzy parameters in an online manner.

For the sake of simplicity, we only adjust the fuzzy membership function for link factor. In our Q-learning algorithm, edge nodes are the agents, and the state space is the set of all possible actions. As shown in Figure 4, a maximum of three different actions, namely, ''Move left,'' ''Maintain'' and ''Move right,'' are available at each state. Here, ''Move left'' or ''Move right'' actions are limited to move horizontally the membership function by 0.05. By doing action ''Move left'' or ''Move right,'' we can tune the weight of link factor on the final fitness value, making the edge selection more suitable for dynamic vehicular environments.

Each node has to maintain a Q-value for each state and action. The Q-Table is updated periodically with hello interval. The initial value for each Q-value is 0. The Q-table is updated periodically for each hello interval as

$$Q(s_t, a) \leftarrow \alpha \times \left\{\hat{R} + \gamma \times \max_{y} Q_m(s_t, y)\right\} + (1-\alpha) \times Q(s_t, a) \tag{4}$$

The reward $\hat{R}$ is calculated as

$$\hat{R} = \begin{cases} 1, & \text{if } HRR \text{ satisfies the threshold} \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

where HRR is the average hello message reception ratio at the edge node from its members (non-edge nodes connected to this edge). Then each agent selects the next action (choose the action with the highest Q-value with probability $1-p$, and do exploration with probability p). Note that the threshold, α, γ and p are configurable parameters according to the application requirements, and set as 0.8, 0.7, 0.9 and 0.1 respectively by default. In order to ensure the synchronization between edge evaluation criteria, each edge nodes announces its best action periodically with hello messages. By using this Q-learning algorithm, the proposed criteria is able to select the best edge nodes in a time-varying vehicular environment.

D. CONTEXT-AWARE EDGE SELECTION

Different kinds of applications have different levels of QoS requirements. It is unrealistic to adjust the edge nodes

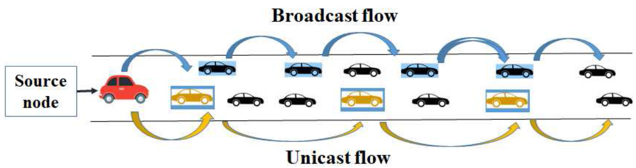

FIGURE 5. Context-aware edge selection.

TABLE 2. Simulation environment.

Topology	2000m, 4lanes
Number of nodes	100–500
Maximum velocity	100 km/h
Mobility generation	Ref. [41]
MAC	IEEE 802.11p MAC (3 Mbps, 27 Mbps)
Propagation model	Nakagami model
Simulation time	1500 s

according to a specific application. However, the applications in vehicular networks could be classified into two categories in terms of the routing requirements, specifically, broadcast and unicast applications. In the broadcast applications, since there is no reception status check for a MAC frame, we have to select more reliable forwarder nodes as compared to the unicast communications which can conduct retransmissions at the MAC layer. Thus, as shown in Figure 5, we introduce a context-aware edge selection approach that selects different edge nodes for broadcast and unicast communications.

More specifically, we define different reference distances for broadcast and unicast applications as they have different requirements on the link condition between a sender and receiver. The reference distance in unicast is larger than the broadcast communications. The reference distance values of the unicast and broadcast communications are set as $\frac{R}{2}$ and $\frac{R}{3}$, respectively, where R is the transmission range of IEEE 802.11p communication. By differentiating the edge nodes between broadcast and unicast communications, the proposed scheme could better fulfill the application needs.

IV. PERFORMANCE ANALYSIS

We conducted three different types of evaluations for the purpose of analyzing different functions of the proposed scheme. First, we used real vehicle traffic data in order to evaluate the performance of the prediction algorithm. Then, we evaluated the communication performance of the proposed scheme in the network simulator ns-2.34 by using both broadcast and unicast communications. We incorporated the result from the prediction algorithm into the network simulations. The Nakagami propagation model was integrated to simulate a realistic vehicular communication channel [39]. There were three different types of communications available for vehicles, namely, LTE (200 Mbps), IEEE 802.11p and mmWave. The average transmission range of IEEE 802.11p was 250m. In the following figures, the error bars show the 95% confidence intervals.

A. TRAFFIC FLOW PREDICTION ACCURACY

We used real road traffic data of road US101-N, District 4, Santa Clara Country, City of San Jose from Caltrans

TABLE 3. Parameters of nakagami model.

gamma0_	gamma1_	gamma2_	d0_gamma_	d1_gamma_
1.9	3.8	3.8	200	500
m0_	m1_	m2_	d0_m_	d1_m_
1.5	0.75	0.75	80	200

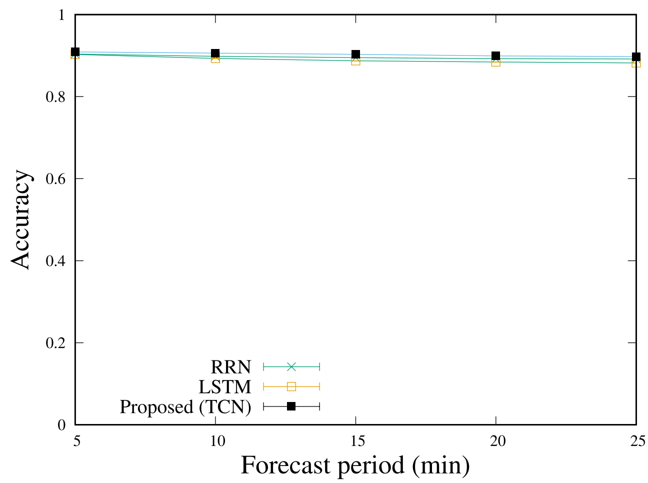

FIGURE 6. Prediction accuracy for different periods (traffic density).

Performance Measurement System (PeMS) [41]. The data were collected with five minutes intervals, and one month traffic data were used for our training. Experiments were conducted on a computer cluster that consists of 16 personal computers (each with 16GB memory and 6th gen Intel quad-core processor). In order to show the effect of learning algorithms in the prediction, we compared three deep learning algorithms, namely, LSTM [42], RRN [42], and TCN [43]. The comparison results of traffic density and vehicle velocity for different traffic forecast periods are shown in Figure 6 and Figure 7 respectively. We found that TCN outperforms other candidates in the prediction accuracy for various forecast periods. Moreover, as compared with the recurrent neural network approaches, such as LSTM and RRN, TCN is more suited for parallelism. This is the reason why we use TCN for the traffic data prediction.

B. BROADCAST COMMUNICATIONS

In order to simulate a realistic multi-hop broadcast scenario, we used a freeway with two lanes in each direction. In each broadcast process, two source nodes (which were neighbors) sent broadcast messages simultaneously. This is to simulate the case of two collided vehicles sending incident messages at the same time. The proposed scheme was compared with "EMPR" [17], "Weighted p-persistence" [17], and "Edge without learning". Here, "Edge without learning" denotes the edge-based forwarding without traffic flow prediction and online learning.

Figure 8 shows the data dissemination ratio comparison for various vehicle densities. It is clearly shown that the conventional sender-based broadcast ("EMPR") and probabilistic

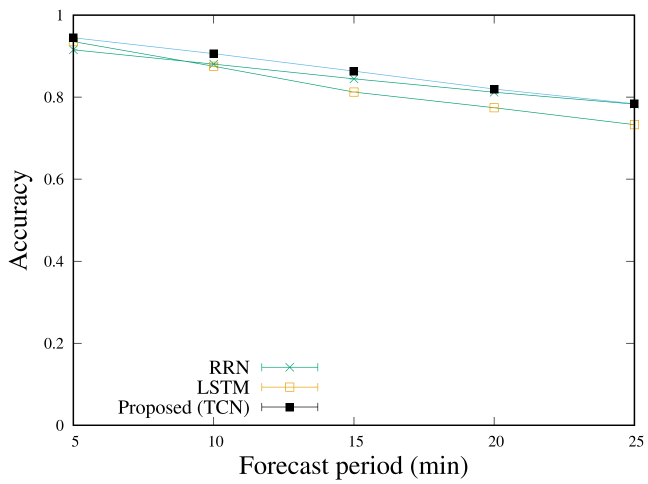

FIGURE 7. Prediction accuracy for different periods (vehicle velocity).

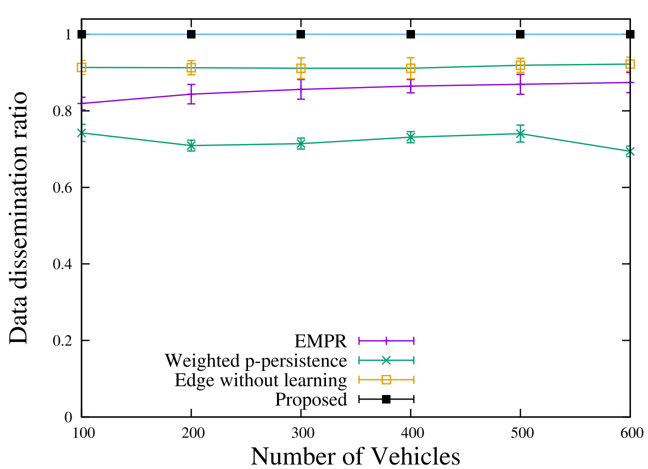

FIGURE 8. Data dissemination ratio for multi-hop broadcast communications.

broadcast ("Weighted p-persistence") are not capable of providing a high data dissemination ratio. This is due to the dynamic topology and limited wireless resources. By aggregating routes using edge nodes, "Edge without learning" can achieve a better performance. However, without any knowledge on the vehicle density and velocity, it is difficult to conduct a satisfactory edge node selection. The proposed scheme can achieve the best performance by combing the traffic flow prediction and online reinforcement learning-based parameter tuning. From the end-to-end delay comparison as shown in Figure 8, we can observe that the proposed scheme is able to provide a low latency by constructing efficient edge nodes that improve the wireless resource utilization efficiency.

C. UNICAST COMMUNICATIONS

In the simulation of unicast applications, we used a road as the same as in [39]. The proposed scheme was compared with "LTE-only", "Random edge (10%)", and "Edge without learning". Here, "LTE-only" denotes that every node uses LTE for the V2X transmissions. In "Random edge (10%)",

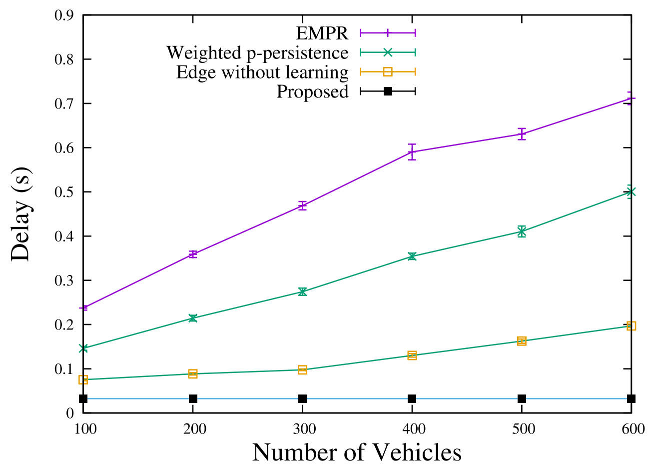

FIGURE 9. End-to-end delay for multi-hop broadcast communications.

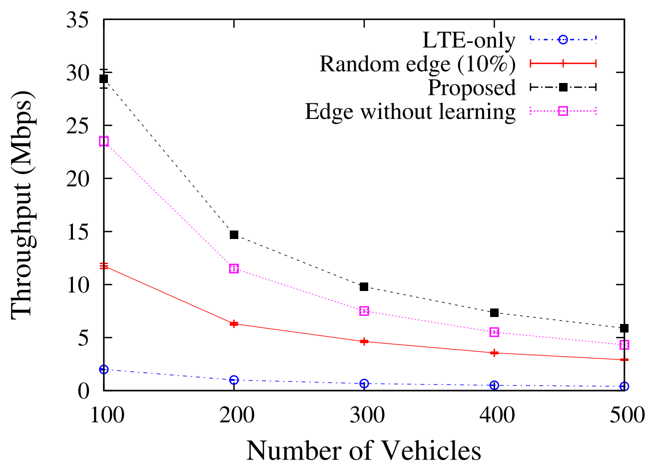

FIGURE 10. Throughput for various numbers of nodes.

ten percent of vehicles are randomly selected to work as edges and the others are connected to the edges through mmWave or IEEE 802.11p. The number of base stations (BS) was 1. The data traffics were from the BS to the vehicles, and all the vehicles were considered as intended receivers. We conducted the performance evaluation for various vehicle densities.

As shown in Figure 10, the poor performance of "LTE-only" shows the importance of utilizing different wireless resources and edge computing. As shown by "Random edge (10%)", the edge node selection algorithm could significantly affect the performance. The advantage of the proposed scheme over "Edge without learning" explains the effect of the proposed two-stage learning algorithm. By an accurate prediction of vehicle velocity and traffic density, the proposed scheme is able to select efficient edge nodes for packet forwarding, resulting in the best performance.

V. FUTURE RESEARCH DIRECTIONS

The fast varying topology of vehicular networks and the limited V2X communication resources could open up some interesting research topics about big data intelligent networking, such as:

- **Data driven optimization based on wireless resource usage big data–** In this article, we used traffic big data to predict the vehicle density and velocity. If we could predict the future user communication behavior, a much better performance would become possible. However, the V2X technology is still in the research stage, and the user traffic big data will be only available after the technology itself is widely used. Therefore, currently, although we can look at a user traffic big data-based optimization approach, it is difficult to rigorously analyze the approach. The optimization of V2X edge architecture based on wireless resource usage big data could be an important research topic in near future.
- **Optimizing the network performance by using deep neural networks–** In the reinforcement learning, an agent is able to find the best action by interacting with environment. However, most reinforcement learning approaches, such as Q-learning, require an explicit mapping between state and action. In vehicular networks, the overall communication performance is affected by many factors, including the user traffic pattern, vehicle density, vehicle mobility and available wireless resources, resulting in that the simple mapping approach is insufficient for representing the complex decision making. Therefore, the use of deep neural networks in the representation of complex state space, namely, deep reinforcement learning, could be a way to solve the resource allocation issue based on an online learning approach.
- **Edge-assisted distributed caching–** With the maturing of autonomous driving technology, customers require more entertainment applications on the drive. Therefore, the V2X communications are required to satisfy a large amount of content delivery with certain QoS. Efficient caching at edge nodes could have a remarkable improvement on the V2X communications, especially for the video data downloading or sharing applications. However, distributed caching and packet forwarding with different types of wireless spectrums are not discussed adequately. Distributed caching technologies that take into account user QoS and network connectivity would be an interesting research topic towards a better connected vehicle society.
- **Management of control messages based on accurate prediction–** The management of control messages is always an important topic for all network protocols. In vehicular networks, the control of beacon intervals could have a great impact on the networking overhead. In a decentralized system, the exchange of control messages among neighbor nodes is a main way to share the knowledge about the network topology. By predicting network density and traffic pattern, a better control message management strategy could be possible.
- **Intelligent cycle of prediction, control, sensing–** It is also interesting to explore an efficient cycle of prediction, control, and sensing. This article discussed the use of prediction information in the initialization of packet forwarding control where each agent employs a reinforcement learning algorithm to improve own behavior based on the results of control. If the knowledge about the control result can be shared between different agents, it could tremendously improve V2X communications. This requires a lightweight sensing mechanism that is able to measure user QoS requirements, network traffic, and wireless resource utilization ratio. The integration issue of prediction, control, and sensing should be carefully addressed as well.

VI. CONCLUSION

In the paradigm of big data driven intelligent solutions for V2X communications, this article addressed an integration of learning algorithms with edge-based forwarding mechanism, and proposed an intelligent edge-based scheme. The proposed scheme first introduces a hierarchical vehicle edge-based preemptive route creation framework where edge nodes are selected by using a fuzzy-logic algorithm. Then, a big data driven approach is used to estimate the traffic density and vehicle velocity which are used in the initialization process of the fuzzy edge selection algorithm. The fuzzy parameters are tuned online by employing a reinforcement learning-based algorithm. Finally, a context-aware approach is used to improve the packet forwarding based on the context information. We used realistic computer simulations and real traffic data to show that the proposed scheme achieves better performance than other baselines for both broadcast and unicast communications. Finally, we discussed some possible future research directions.

REFERENCES

[1] K.-L.-A. Yau, J. Qadir, C. Wu, M. A. Imran, and M. H. Ling, "Cognition-inspired 5G cellular networks: A review and the road ahead," *IEEE Access*, vol. 6, pp. 35072–35090, 2018.

[2] C. Wu, Z. Liu, D. Zhang, T. Yoshinaga, and Y. Ji, "Spatial intelligence toward trustworthy vehicular IoT," *IEEE Commun. Mag.*, vol. 56, no. 10, pp. 22–27, Oct. 2018.

[3] Z. Junhui, Y. Tao, G. Yi, W. Jiao, and F. Lei, "Power control algorithm of cognitive radio based on non-cooperative game theory," *China Commun.*, vol. 10, no. 11, pp. 143–154, Nov. 2013.

[4] G. Sun, L. Song, H. Yu, V. Chang, X. Du, and M. Guizani, "V2V routing in a VANET based on the autoregressive integrated moving average model," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 908–922, Jan. 2019.

[5] H. I. Abbasi, R. C. Voicu, J. Copeland, and Y. Chang, "Towards fast and reliable multi-hop routing in VANETs," *IEEE Trans. Mobile Comput.*, early access, 2019, doi: 10.1109/tmc.2019.2923230.

[6] A. Bazzi, B. M. Masini, A. Zanella, and I. Thibault, "On the performance of IEEE 802.11p and LTE-V2V for the cooperative awareness of connected vehicles," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10419–10432, Nov. 2017.

[7] 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, doi: 10.1109/tits.2019.2904953.

[8] J. Joo, M. C. Park, D. S. Han, and V. Pejovic, "Deep learning-based channel prediction in realistic vehicular communications," *IEEE Access*, vol. 7, pp. 27846–27858, 2019.

[9] C. Wu, T. Yoshinaga, X. Chen, L. Zhang, and Y. Ji, "Cluster-based content distribution integrating LTE and IEEE 802.11p with fuzzy logic and Q-learning," *IEEE Comput. Intell. Mag.*, vol. 13, no. 1, pp. 41–50, Feb. 2018.
[10] Y. Hui, Z. Su, T. H. Luan, and J. Cai, "Content in motion: An edge computing based relay scheme for content dissemination in urban vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 8, pp. 3115–3128, Aug. 2019, doi: 10.1109/tits.2018.2873096.
[11] J. Zhao, Q. Li, Y. Gong, and K. Zhang, "Computation offloading and resource allocation for cloud assisted mobile edge computing in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7944–7956, Aug. 2019.
[12] S. Guleng, C. Wu, T. Yoshinaga, and Y. Ji, "Traffic big data assisted broadcast in vehicular networks," in *Proc. ACM RACS*, Sep. 2019, pp. 236–240.
[13] J. Wu, L. Zou, L. Zhao, A. Al-Dubai, L. Mackenzie, and G. Min, "A multi-UAV clustering strategy for reducing insecure communication range," *Comput. Netw.*, vol. 158, pp. 132–142, Jul. 2019.
[14] C. Wu, T. Yoshinaga, Y. Ji, T. Murase, and Y. Zhang, "A reinforcement learning-based data storage scheme for vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6336–6348, Jul. 2017.
[15] F. Dressler, F. Klingler, C. Sommer, and R. Cohen, "Not all VANET broadcasts are the same: Context-aware class based broadcast," *IEEE/ACM Trans. Netw.*, vol. 26, no. 1, pp. 17–30, Feb. 2018.
[16] C. Wu, X. Chen, Y. Ji, F. Liu, S. Ohzahata, T. Yoshinaga, and T. Kato, "Packet size-aware broadcasting in VANETs with fuzzy logic and RL-based parameter adaptation," *IEEE Access*, vol. 3, pp. 2481–2491, 2015.
[17] C. Wu, S. Ohzahata, and T. Kato, "VANET broadcast protocol based on fuzzy logic and lightweight retransmission mechanism," *IEICE Trans. Commun.*, vol. E95-B, no. 2, pp. 415–425, Feb. 2012.
[18] N. Wisitpongphan, O. Tonguz, J. Parikh, P. Mudalige, F. Bai, and V. Sadekar, "Broadcast storm mitigation techniques in vehicular ad hoc networks," *IEEE Wireless Commun.*, vol. 14, no. 6, pp. 84–94, Dec. 2007.
[19] A. Tahmasbi-Sarvestani, Y. P. Fallah, and V. Kulathumani, "Network-aware double-layer distance-dependent broadcast protocol for VANETs," *IEEE Trans. Veh. Technol.*, vol. 64, no. 12, pp. 5536–5546, Dec. 2015.
[20] S. S. Shah, A. W. Malik, A. U. Rahman, S. Iqbal, and S. U. Khan, "Time barrier-based emergency message dissemination in vehicular ad-hoc networks," *IEEE Access*, vol. 7, pp. 16494–16503, 2019.
[21] K. Jia, Y. Hou, K. Niu, C. Dong, and Z. He, "The delay-constraint broadcast combined with resource reservation mechanism and field test in VANET," *IEEE Access*, vol. 7, pp. 59600–59612, 2019.
[22] S. Kumar, U. Dohare, K. Kumar, D. P. Dora, K. N. Qureshi, and R. Kharel, "Cybersecurity measures for geocasting in vehicular cyber physical system environments," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 5916–5926, Aug. 2019, doi: 10.1109/jiot.2018.2872474.
[23] P. Li, T. Zhang, C. Huang, X. Chen, and B. Fu, "RSU-assisted geocast in vehicular ad hoc networks," *IEEE Wireless Commun.*, vol. 24, no. 1, pp. 53–59, Feb. 2017.
[24] F. Zhang, B. Jin, Z. Wang, H. Liu, J. Hu, and L. Zhang, "On geocasting over urban bus-based networks by mining trajectories," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 6, pp. 1734–1747, Jun. 2016.
[25] Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo, and J. Zhang, "Edge intelligence: Paving the last mile of artificial intelligence with edge computing," *Proc. IEEE*, vol. 107, no. 8, pp. 1738–1762, Aug. 2019.
[26] X. Chen, H. Zhang, C. Wu, S. Mao, Y. Ji, and M. Bennis, "Optimized computation offloading performance in virtual edge computing systems via deep reinforcement learning," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4005–4018, Jun. 2019.
[27] N. Hassan, K.-L.-A. Yau, and C. Wu, "Edge computing in 5G: A review," *IEEE Access*, vol. 7, pp. 127276–127289, 2019.
[28] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang, and W. Wang, "A survey on mobile edge networks: Convergence of computing, caching and communications," *IEEE Access*, vol. 5, pp. 6757–6779, 2017.
[29] S. Liu, L. Liu, J. Tang, B. Yu, Y. Wang, and W. Shi, "Edge computing for autonomous driving: Opportunities and challenges," *Proc. IEEE*, vol. 107, no. 8, pp. 1697–1716, Aug. 2019.
[30] H. A. Khattak, S. U. Islam, I. U. Din, and M. Guizani, "Integrating fog computing with VANETs: A consumer perspective," *IEEE Commun. Standards Mag.*, vol. 3, no. 1, pp. 19–25, Mar. 2019.
[31] H. Peng, Q. Ye, and X. S. Shen, "SDN-based resource management for autonomous vehicular networks: A multi-access edge computing approach," *IEEE Wireless Commun.*, vol. 26, no. 4, pp. 156–162, Aug. 2019.
[32] C. Wu, S. Ohzahata, Y. Ji, and T. Kato, "How to utilize interflow network coding in VANETs: A backbone-based approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 8, pp. 2223–2237, Aug. 2016.
[33] C. Chen, L. Liu, T. Qiu, D. O. Wu, and Z. Ren, "Delay-aware grid-based geographic routing in urban VANETs: A backbone approach," *IEEE/ACM Trans. Netw.*, vol. 27, no. 6, pp. 2324–2337, Dec. 2019, doi: 10.1109/tnet.2019.2944595.
[34] K. Lin, J. Luo, L. Hu, M. S. Hossain, and A. Ghoneim, "Localization based on social big data analysis in the vehicular networks," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 1932–1940, Aug. 2017.
[35] W. Liu and Y. Shoji, "Edge-assisted vehicle mobility prediction to support V2X communications," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 10227–10238, Oct. 2019.
[36] X. Chen, C. Wu, M. Bennis, Z. Zhao, and Z. Han, "Learning to entangle radio resources in vehicular communications: An oblivious game-theoretic perspective," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4262–4274, May 2019.
[37] G. Qiao, S. Leng, S. Maharjan, Y. Zhang, and N. Ansari, "Deep reinforcement learning for cooperative content caching in vehicular edge computing and networks," *IEEE Internet Things J.*, early access, 2019, doi: 10.1109/jiot.2019.2945640.
[38] C. Wu, Y. Ji, X. Chen, S. Ohzahata, and T. Kato, "An intelligent broadcast protocol for VANETs based on transfer learning," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–6.
[39] Q. Hu, C. Wu, X. Zhao, X. Chen, Y. Ji, and T. Yoshinaga, "Vehicular multi-access edge computing with licensed sub-6 GHz, IEEE 802.11p and mmWave," *IEEE Access*, vol. 6, pp. 1995–2004, 2018.
[40] F. Bai, N. Sadagopan, and A. Helmy, "IMPORTANT: A framework to systematically analyze the impact of mobility on performance of routing protocols for adhoc networks," in *Proc. 22nd Annu. Joint Conf. IEEE Comput. Commun. Societies (INFOCOM)*, Mar. 2003, pp. 825–835.
[41] *Caltrans Performance Measurement System (PeMS)*. Accessed: Jun. 20, 2019. [Online]. Available: http://pems.dot.ca.gov/
[42] Y. Wang and F. Tian, "Recurrent residual learning for sequence classification," in *Proc. EMNLP*, 2016, pp. 938–943.
[43] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," Mar. 2018, *arXiv:1803.01271*. [Online]. Available: https://arxiv.org/abs/1803.01271

SIRI GULENG is currently a Professor with the Hohhot Minzu College, China. His current research interests include computer networks and cyber security.

CELIMUGE WU (Senior Member, IEEE) received the M.E. degree from the Beijing Institute of Technology, China, in 2006, and the Ph.D. degree from The University of Electro-Communications, Japan, in 2010. He is currently an Associate Professor with the Graduate School of Informatics and Engineering, The University of Electro-Communications. His current research interests include vehicular ad hoc networks, sensor networks, intelligent transport systems, the IoT, and mobile cloud computing. He is/has been a TPC Co-Chair of Wireless Days 2019, ICT-DM 2018, and a track Co-Chair of many international conferences, including ICCCN 2019 and IEEE PIMRC 2016. He is/has been serving as an Associate Editor of IEEE Access, the *IEICE Transactions on Communications*, the *International Journal of Distributed Sensor Networks*, *MDPI Sensors*, and a Guest Editor of the IEEE Transactions on Emerging Topics in Computational Intelligence, the *IEEE Computational Intelligence Magazine*, ACM/Springer MONET, and so on.

ZHI LIU (Senior Member, IEEE) received the B.E. degree from the University of Science and Technology of China, China, and the Ph.D. degree in informatics from the National Institute of Informatics. He is currently an Assistant Professor with Shizuoka University. His research interests include video network transmission, vehicular networks, and mobile edge computing. He has been serving as the Chair for a number of international conference and workshops. He has served as Guest Editor of journals, including the *Wireless Communications and Mobile Computing* and the *IEICE Transactions on Communications*.

XIANFU CHEN (Member, IEEE) received the Ph.D. degree in signal and information processing from the Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China, in March 2012. He is currently a Senior Scientist with the VTT Technical Research Centre of Finland, Oulu, Finland. His research interests include various aspects of wireless communications and networking, with the emphasis on software-defined networking, green communications, centralized and decentralized resource allocation, dynamic spectrum access, and the application of artificial intelligence to wireless communications.