This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2899627, IEEE Transactions on Vehicular Technology

Delay-Minimization Routing for Heterogeneous VANETs with Machine Learning based Mobility Prediction

Yujie Tang, *Member, IEEE*, Nan Cheng, *Member, IEEE*, Wen Wu, *Student Member, IEEE*, Miao Wang, *Member, IEEE*, Yanpeng Dai, *Student Member, IEEE*, and Xuemin (Sherman) Shen, *Fellow, IEEE*

***Abstract*—Establishing and maintaining end-to-end connections in a vehicular ad hoc network (VANET) is challenging due to the high vehicle mobility, dynamic inter-vehicle spacing, and variable vehicle density. Mobility prediction of vehicles can address the aforementioned challenge since it can provide a better routing planning and improve overall VANET performance in terms of continuous service availability. In this paper, a centralized routing scheme with mobility prediction (CRS-MP) is proposed for VANET assisted by an artificial intelligence-powered software-defined network (SDN) controller. Specifically, the SDN controller can perform accurate mobility prediction through advanced artificial neural network (ANN) technique. Then, based on the mobility prediction, the successful transmission probability and average delay of each vehicle's request under frequent network topology changes can be estimated by the roadside units (RSUs) or the base station (BS). The estimation is performed based on a stochastic urban traffic model in which the vehicle arrival follows a non-homogeneous Poisson process (NHPP). The SDN controller gathers network information from RSUs and BS which are considered as the switches. Based on the global network information, the SDN controller computes optimal routing paths for switches (i.e., BS and RSU). While the source vehicle and destination vehicle are located in the the coverage area of the same switch, further routing decision will be made by the RSUs or the BS independently to minimize the overall vehicular service delay. The RSUs or the BS schedules the requests of vehicles by either vehicle-to-vehicle (V2V) or vehicle-to-infrastructure (V2I) communication, from the source vehicle to the destination vehicle. Simulation results demonstrate that our proposed centralized routing scheme outperforms others in terms of transmission delay, and the transmission performance of our proposed routing scheme is more robust with varying vehicle velocity.**

***Index Terms*–Vehicular ad hoc network (VANET), routing, software-defined network (SDN), machine learning**

I. Introduction

Vehicular ad hoc network (VANET) is an important component of the intelligent transportation systems (ITS) and has emerged as a promising research topic in recent years to improve transportation efficiency, alleviate traffic congestion, enhance road safety, and provide entertainment service and internet access for both passengers and drivers [1]. In VANET, vehicles and roadside units (RSUs, e.g., roadside base stations or access points) are equipped with short range wireless transceivers. Each vehicle communicates with either an RSU through vehicle-to-infrastructure communication (V2I) or other vehicles through vehicle-to-vehicle communication (V2V) [2]. The dedicated short-range communication (DSRC) [3] is a special wireless technology aiming at supporting both V2I and V2V communications. DSRC refers to as a suite of standards including IEEE 802.11p, IEEE 1609.1/.2/.3/.4 protocol family and SAE J2735 message set dictionary [4]. In DSRC, the vehicle equipped with an on-board unit (OBU) can communicate with other vehicles and RSUs, while the RSUs are fixed infrastructures deployed along the road to provide applications and services for vehicles. In addition, existing Wi-Fi network and long term evolution (LTE) network can also be the options for the next-generation vehicular networks.

VANET has several challenging issues that distinguish it from other mobile ad hoc networks (MANETs). The amount of traffic under different traffic conditions such as rush hours and traffic jams results in a highly dynamic network topology. In addition, the high mobility characteristic of vehicles leads to an intermittent connectivity among the vehicles or between the vehicles and RSUs. Thus, communication links in VANET may have frequent outages when they are exchanging information. Therefore, traditional routing schemes such as optimized link state protocol (OLSR) [5], ad hoc on-demand distance vector (ADOV) [6] and dynamic source routing (DSR) [7] are not efficient or fully applicable to VANET. Aiming at addressing the challenges in VANET, various ad hoc routing protocols have been proposed [8]–[10]. However, those protocols adopt hello messages to detect neighbor nodes. The hello messages are used to carry the vehicle's geographic position, velocity, and moving direction information. This kind of information tends to exhaust the communication bandwidth due to periodic beaconing. Moreover, heavy control cost occurs under the rapidly changing network topology.

To this end, software-defined network (SDN) has emerged as a solution to control the networks with low communication cost. SDN is a paradigm in networking and computing, which separates the data communication and control planes to simplify the network management and expedite system evolution [11]. Therefore, SDN based VANET can provide lots of advantages for the management of mobility in VANET scenario. To sum up, there are three advantages of introducing SDN technology into VANET. First, the SDN controller can provide a promising opportunity to simplify the management and control of the network by decoupling the control and data

Y. Tang, N. Cheng, W. Wu, and X. Shen are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, N2L3G1, Canada (emails: {y59tang, n5cheng, w77wu, sshen}@uwaterloo.ca). N. Cheng is the corresponding author.

M. Wang is with the Department of Electrical and Computer Engineering, Miami University, Oxford, OH, USA (email: wangm64@miamioh.edu).

Y. Dai is with the School of Telecommunication Engineering, Xidian University, Xi'an, Shaanxi, China (email: yp_dai@stu.xidian.edu.cn).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2899627, IEEE Transactions on Vehicular Technology

planes, as well as providing elasticity for VANET. Second, with the help of SDN, switches do not need to exchange beacon information with each other periodically. The routing overhead can be reduced, and the SDN controller can acquire the global knowledge and then make the optimal routing decision. Third, the central SDN controller is aware of the connections among switches which makes the routing flexible, since routing decisions may change due to the failures of some switches. However, vehicles are moving fast and the topology changes dynamically. Therefore, the decisions made by the SDN controller need to be timely. In addition, the SDN controller can make the optimal routing decision based on the global information, which is not easy to achieve.

Before designing the routing algorithm, the vehicle traffic needs to be evaluated first. The vehicle traffic varies during the day, i.e., the VANET can be either fully connected (rush hour) or sparsely connected (non-rush hour) depending on the time of the day [1]. The majority of existing works model the vehicle arrival as a Poisson process [12], [13], but in reality there are multiple busy periods because of rush hours. Therefore, non-homogeneous Poisson process (NHPP) is more appropriate to model vehicle arrivals that exhibit in multiple busy periods. The stochastic monotonicity and related properties of the inter-arrival time of an NHPP have been studied by Kochar [14] and Pellerey *et al.* [15]. However, in NHPP, the arrival rate function, which is used for the probability and delay analysis, is very challenging to obtain due to the random vehicle arrivals and the dynamic traffic conditions.

In this paper, we propose a unicast routing protocol for VANET to minimize the delay without continuously monitoring vehicle locations. The continuous monitoring may cause a large amount of overhead, and the routing decision may not be accurate enough due to the high mobility of vehicles. Therefore, mobility prediction would be a better choice to estimate the vehicle connections. The proposed centralized routing scheme with mobility prediction (CRS-MP) can ensure more reliable and timely transmissions for VANET, in which the SDN controller can intelligently predict the vehicle arrivals and vehicle connections. Artificial neural network (ANN) is employed to learn and predict the vehicle arrival rate. Based on the prediction, the RSU/BS integrates a stochastic urban traffic model in the analysis to estimate the successful transmission probability and average delay. Whether the SDN controller decides the routing path among the RSUs and the BS or the RSU/BS makes the routing decision for the requests to minimize the delay is determined by the locations of source vehicle and destination vehicle. During a request scheduling period of each RSU/BS, vehicles in the V2I channel and the V2V channel can transmit data simultaneously due to the reason that V2I and V2V communications are using different spectrum access technologies. The motivation of the requests scheduling is to serve the vehicles which are far away from each other and to alleviate the burden of the RSU/BS in rush hours as well as to improve the vehicle connectivity by the assistance of RSU/BS in non-rush hours. In particular, the main contributions of this paper are outlined as follows.

- We propose a CRS-MP routing scheme, in which the controller does not continuously monitor vehicle locations. The controller uses the ANN to learn and predict the vehicle arrival rate, reaping the powerful learning capability of ANN. Based on the vehicle arrival rate function, the RSU/BS can do the traffic mobility estimation by building the statistical traffic model, and then make routing decisions to vehicles.
- The successful transmission probability and average time delay of not only the single-hop and multi-hop V2V transmissions, but also the V2I communication are analyzed in a string of vehicles based on the NHPP traffic model and the learnt arrival function.
- We formulate the routing problem aiming to minimize the overall vehicular service delay in VANET as an integer programming problem which is NP-hard.
- We solve the NP-hard routing problem by performing a bipartite matching algorithm, and design a request delivery algorithm to further improve the network performance. Meanwhile, we take the serving capability of RSU/BS into consideration.

The remainder of this paper is organized as follows. Section II reviews related work. The system model is described in Section III. In Section IV, the mobility prediction is presented in detail. In Section V, we propose the CRS-MP routing scheme. Numerical results are presented in Section VI. Finally, concluding remarks are given in Section VII.

II. Related Work

Routing schemes for VANET have been widely investigated in the literature [16]–[20]. They can be classified into two categories, topology-based and position-based [21]. In topology-based routing, packets are delivered based on the information of network links, while in geographic routing packets are forwarded based on locations of users' neighbors and destinations.

Nevertheless, most of existing works primarily focus on broadcasting routing protocols for VANET. In fact, message broadcasting increases the communication overhead and wastes the bandwidth. Therefore, we consider the unicast scenario [8], [9], [22] in which the messages are delivered by end-to-end transmission instead of broadcasting. Security issues are studied in [23]–[25] which may be integrated in our work, but they are beyond the scope of this work. Moreover, in the context of VANET, the establishment of routing path needs frequent negotiation since the topology may change very fast. Therefore, although the monitoring can be done by basic safety message, conventional routing protocols can still lead to network congestion. However, in our scheme, the routing-related control messages are sent through LTE, which avoids frequent rebroadcasting and makes the network less congested.

In order to establish the end-to-end route, reduce the information exchange cost, and transmit packets timely, several works focus on studying the routing based on SDN. In [26], within a given reconfiguration constraint, Destounis *et al.* provided a practical way of keeping the routing cost small. However, this SDN based routing scheme focuses on the routing policy among the nodes with wired connections. Duan *et al.* in [27] assumed that the SDN controller monitors the location of vehicles without the update information from vehicles. This kind of assumption reduces the communication cost but increases the complexity of the SDN controller. In [28], Liu *et al.* formulated the cooperative data scheduling algorithm. The RSU chooses source and destination vehicles and corresponding data for V2V communication. At the same time, the RSU broadcasts a message to vehicles that are informed to tune to the V2I channel. However, only one-hop V2V communication is considered in this work. In [29], the SDN controller sets up the routing paths of the periodic warning messages to the destination RSUs based on topological and geographical information. Nevertheless, it becomes difficult to manage the dynamically changing requests when network size increases. Therefore, though SDN based VANET improves the network efficiency by increasing throughput, reducing latency, and enhancing reliability of the network, it brings some new challenges.

The most challenging issue for SDN based VANET is that even though SDN can make decision on optimal routing path based on the global information, the network topology changes very fast in VANET, especially when the vehicle speed is high. Therefore, prediction becomes a crucial component in SDN controller, which can help the SDN controller make decision ahead of time based on the predicted information. Usually, machine learning can be used for prediction. In [30], Shen *et al.* proposed a novel adaptive fuzzy logic inference system to predict and estimate the probability information for wireless communication networks, and the prediction is performed using the recursive least square (RLS) approach. However, ANN is a powerful approach in machine learning which is widely used in solving various classification and forecasting problems. The authors in [31] proposed a spectrum detection method by adopting ANN. In [32], the authors used ANN to predict the traffic flow by developing an optimized structure through layer-by-layer feature granulation with a greedy-layer wise unsupervised learning algorithm. However, our work focuses on the vehicle mobility prediction by using ANN. A traffic flow prediction method is proposed in [33] by using a deep learning approach. Nevertheless, the prediction period is quite long which is not feasible for VANET due to the high mobility feature.

III. System Model

A. Network Model

We present the network model of a joint V2I and V2V communication system for VANET in urban area, in which the ubiquitous coverage can be achieved through cellular networks, i.e. BS, as shown in Fig. 1. A distributed SDN controller which copes with the distributed and heterogeneous nature of modern overlay networks is adopted in our work. Each controller is responsible for an SDN domain which refers to cities or communities. There exists a stable TCP/IP connection between distributed controllers, which minimizes the probability of unsuccessful transmission and packet loss. Each SDN controller connects all the BS and RSUs within its charged area, and the BS and the RSUs are considered as the switches which can communicate with each other. The RSUs are located in each intersection since deploying RSUs at intersections can reduce the need for non-line-of-sight (NLOS) transmissions along the road to provide better service coverage. The RSUs serve the vehicles within the RSUs' coverage area, while the BS serves the vehicles which are out of the coverage area of any RSU. The vehicles send their requests as well as the IP addresses of both themselves and their destinations to the RSU/BS. Afterwards, the RSU/BS delivers the information to the controller. Then, the SDN controller converts the IP addresses to the vehicle indexes according to the locations of both source vehicle and destination vehicle and makes the routing decision.

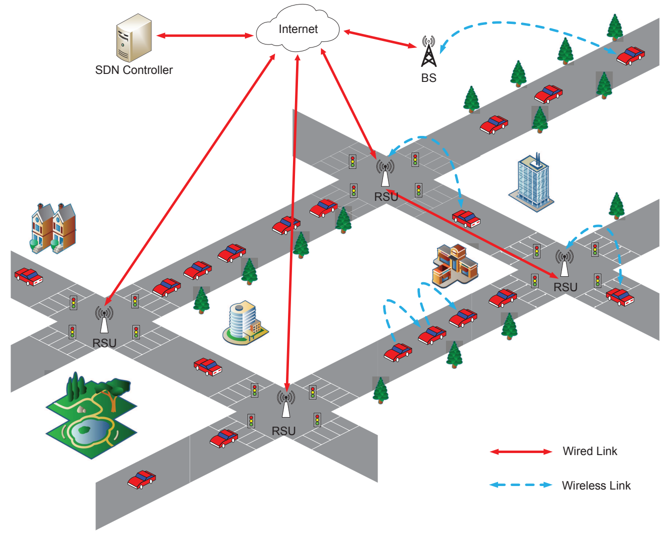


Fig. 1. Scenario in urban area.

Each vehicle is considered to be equipped with both IEEE 802.11p and LTE interfaces as this hybrid architecture has been recently adopted to exploit both the low cost of IEEE 802.11p and the wide-range low-latency communication of the cellular networks [34]. V2V communications use DSRC based on IEEE 802.11p protocol, where a 75 MHz band is allocated with 7 channels, each with 10 MHz bandwidth. Meanwhile, V2I communications operate through LTE, in which multiple vehicles can share the spectrum resources by orthogonal frequency division multiple access (OFDMA). The communications among the RSUs and the BS are via wired connections. The neighboring RSUs use different bandwidth to avoid the interference. For the DSRC, the control channel is for the standard usage such as the safety message broadcasting, while the other 6 channels are used for V2V communications. The routing-related messages are sent through the LTE system. Orthogonal signaling is assumed for LTE in each time slot (one request scheduling period), and each resource block is allocated to each request. Each RSU can serve up to C vehicles, and the transmission range of each RSU is D meters. The rest of the notations used in this paper are listed in Table I.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2899627, IEEE Transactions on Vehicular Technology

Table I
NOTATIONS TABLE

Notation	Description
$\lambda(t)$	Arrival rate function
$\Lambda(t)$	Mean value function
R_m^y	Uplink transmission rate from vehicle m to RSU y
R_n^y	Downlink transmission rate from RSU y to vehicle n
R_{V_k}	Transmission rate of per hop V2V communication
P_V	Transmission power of vehicles
P_y	Transmission power of RSU y to vehicles
N_0	Power of additive white Gaussian noise (AWGN)
W_m	Bandwidth allocated for uplink transmission
W_n	Bandwidth allocated for downlink transmission
W_V	Bandwidth allocated for V2V transmission
d_m	Distance from source vehicle m to RSU y
d_n	Distance from RSU y to destination vehicle n
d_k	Inter-vehicle spacing of kth hop transmission
X	Number of requests of vehicles
Y	Number of RSUs
Z	Size of request packet size
H	Number of hops transmitting from vehicle m to vehicle n

B. Statistical Mobility Model

An NHPP $\mathcal{A}$ counts the number of vehicles that enters the road, and $\mathcal{A}$ is characterized by its arrival rate function $\lambda(t), (t \geq 0)$ which is also referred to as the intensity function, and $\Lambda(t) = \int_0^t \lambda(u)du$ denotes the mean value function, which is the expected number of arrivals on the time interval $(0, t)$. Let $N(x, t)$ be the number of vehicles in the road segment $(0, x]$ at time t, and $n(x, t)$ be the density of vehicles in road segment $(0, x]$ at time t. The expectation of $N(x, t)$ is expressed by

$$\mathbb{E}[N(x, t)] = \int_0^t \lambda(u)du. \tag{1}$$

The relationship between $N(x, t)$ and $n(x, t)$ would be described by

$$n(x, t) = \frac{\partial N(x, t)}{\partial x}. \tag{2}$$

1) Inter-arrival Time Distribution: Given that the NHPP starts at $T_0 = 0$, the arrival time are denoted by $T_1, T_2, \ldots, T_n$, and the corresponding inter-arrival times are represented by $X_1 = T_1, X_2 = T_2 - T_1, \ldots, X_n = T_n - T_{n-1}(n = 1, 2, \ldots)$. According to the deduction from (3) in [35], the probability density function of T_n, i.e., $f_n(t)$, is

$$f_n(t) = \frac{\lambda(t)e^{-\Lambda(t)}\left[\Lambda(t)\right]^{n-1}}{(n-1)!}, t > 0, n \geq 1. \tag{3}$$

For $t > 0$, the probability density functions of X_1 and X_n, i.e., $g_1(t)$ and $g_n(t)$, are given respectively by (see (7) of [35])

$$g_1(t) = \lambda(t)e^{-\Lambda(t)}, \tag{4}$$

$$g_n(t) = \int_0^{+\infty} \lambda(x)\lambda(t + x)e^{-\Lambda(t+x)}\frac{[\Lambda(x)]^{n-2}}{(n-2)!}dx, n \geq 2. \tag{5}$$

2) Inter-vehicle Spacing Distribution: Since the vehicles move at an average speed v, the inter-vehicle spacing follows the NHPP with intensity function $\lambda_S(s)$. The inter-vehicle spacing $\phi_n(s)$ can be characterized by the following distribution which is given by

$$\phi_1(s) = \lambda_S(s)e^{-\Lambda_S(s)}, s > 0, \tag{6}$$

$$\phi_n(s) = \int_0^{+\infty} \lambda_S(x)\lambda_S(s + x)e^{-\Lambda_S(s+x)}\frac{[\Lambda_S(x)]^{n-2}}{(n-2)!}dx, \quad s > 0, n \geq 2 \tag{7}$$

where the parameter $\lambda_S(s)$ and $\Lambda_S(s)$ can be represented by

$$\lambda_S(s) = \lambda\left(\frac{s}{v}\right) \tag{8}$$

and

$$\Lambda_S(s) = \int_0^s \lambda_S(u)du \tag{9}$$

respectively.

3) Explanation by an Example: For simplification, we give an example of arrival rate function with a monotonic linear function:

$$\lambda(t) = ut, \quad t > 0, u > 0. \tag{10}$$

By $\lambda(t) = ut$ and $\Lambda(t) = \frac{ut^2}{2}$, for $t > 0$, and $n = 2, 3, \ldots$, from (10), we have

$$g_1(t) = ut \cdot e^{-\frac{ut^2}{2}} \tag{11}$$

and

$$\begin{aligned} g_n(t) &= \frac{u^n}{2^{n-2}(n-2)!}\int_0^{+\infty} x^{2n-3}(t + x)e^{-\frac{\lambda(t+x)^2}{2}}dx \\ &= \frac{\sqrt{u}e^{\frac{-at^2}{4}}(2n-3)!}{2^{n-2}(n-2)!}\left\{\sqrt{u}tD_{-2n+2}(\sqrt{u}t) + (2n-2)\right. \\ &\quad \left.\cdot D_{-2n+1}(\sqrt{u}t)\right\} \end{aligned} \tag{12}$$

where $D_\nu(x)$ denotes the parabolic cylinder function. Since $(2n-3)! = 2^{n-2}(n-2)!(2n-3)!!$, and

$$D_{\nu+1}(x) = xD_\nu(x) - \nu D_{\nu-1}(x), \tag{13}$$

we can obtain

$$g_n(t) = (2n-3)!!\sqrt{u}e^{-\frac{ut^2}{4}}D_{-(2n-3)}(\sqrt{u}t). \tag{14}$$

Therefore, we can see that after acquiring the vehicle arrival function $\lambda(t)$, the probability density function can be obtained which means the inter-arrival time and inter-vehicle spacing can be estimated at any given time.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2899627, IEEE Transactions on Vehicular Technology

C. Channel Model

The uplink transmission rate R_m^y and the downlink transmission rate R_n^y of RSU y for V2I communication are described by

$$R_m^y = W_m \cdot \log_2 \left(1 + \frac{P_V}{N_0} \delta d_m^{-\gamma} |h|_{m,y}^2 \right) \tag{15}$$

and

$$R_n^y = W_n \cdot \log_2 \left(1 + \frac{P_y}{N_0} \delta d_n^{-\gamma} |h|_{y,n}^2 \right) \tag{16}$$

respectively. P_V denotes the transmission power of the vehicles. P_y denotes the transmission power of RSU y, and it refers to the transmission power of BS when $y = 0$. W_m is the bandwidth allocated for uplink, while W_n is the bandwidth allocated for downlink. N_0 is the power of additive white Gaussian noise (AWGN), and δ is the log-normal shadowing component, with a mean of 0 dB and standard deviation σ_S. d_m and d_n are the distances from the source vehicle m to RSU y and from RSU y to the destination vehicle n, respectively. γ is the path fading exponent which is typically chosen within $[2, 4]$. $|h|_{m,y}$ and $|h|_{y,n}$ are the Rayleigh-distributed fading magnitude with $\mathbb{E}[|h|^2] = 1$.

The transmission rate of per hop V2V communication is presented by

$$R_{V_k} = W_V \cdot \log_2 \left(1 + \frac{P_V}{N_0} \delta d_k^{-\gamma} |h|_k^2 \right) \tag{17}$$

where W_V is the bandwidth allocated for V2V communication, and d_k is the inter-vehicle spacing of kth hop transmission which is a positive value since backward transmission is considered. $|h|_k$ is the Rayleigh-distributed fading magnitude with $\mathbb{E}[|h|^2] = 1$. d_m, d_n and d_k can be estimated through the probability density function of inter-vehicle spacing. Based on the estimation of the distance, we can obtain the transmission rates R_m^y, R_n^y and R_{V_k}.

D. ANN Model

ANN is a powerful technique which imitates the neural structure of the human brain. In ANN, knowledge can be obtained by the network through a learning procedure, and the knowledge can be stored. There are three main benefits of ANN. First, ANN can adapt itself according to the input data without knowing any information of the signal. Second, ANN can approximate any function with arbitrary accuracy. Third, ANN is a nonlinear model and can be applied to most of real world applications. As the aforementioned merits, ANN has a natural propensity for storing experiential knowledge and makes it available for use. Therefore, we use ANN to predict the vehicle arrival function and further estimate the successful transmission probability and average delay of multi-hop transmission.

There are many different types of ANN, and we focus on a simple and effective model of ANN, i.e., back-propagation neural network (BPNN), which has three layers: input layer, hidden layer and output layer. A two-layer neural network is selected in this work. There are 20 neurons in the hidden layer and the number the neurons is chosen according to the empirical value. The transfer function is "transig" in the hidden layer. The other layer is the output layer with one neuron, and the transfer function is "purelin". Function "transig" and "purelin" are the two transfer functions of BPNN. A training function "trainlm", one of the training functions of BPNN, is chosen.

IV. Mobility Prediction

In order to make the routing decision, the SDN controller needs to predict the vehicle arrival rate function. Based on the prediction, the successful transmission probability and average delay can be estimated ahead of time. In this section, we describe the proposed SDN based mobility prediction method in detail.

A. General Settings

All the vehicles are moving at the same speed, and the speed varies in different times of the day. The speed is determined by the SDN controller according to the vehicles' arrival rate. The vehicles send their requests together with their IP addresses to the RSU/BS through control channel, and then the RSU/BS delivers the received information to the controller. The controller sorts the source vehicles and destination vehicles associated with different RSUs or the BS according to the vehicle indexes. When the source vehicle and destination vehicle exist in the coverage area of the same RSU or in the same road segment but out of the RSUs' coverage, the SDN controller estimates the vehicle arrival rate function and sends the information to the RSU/BS to let it do its own routing by scheduling the vehicle requests based on the prediction and estimations. While the source vehicle and destination vehicle are associated with different RSUs or one of the vehicle is out of the coverage area of any RSU, the SDN controller makes the routing decisions among the RSUs and the BS.

Transmission ranges for RSUs and vehicles are denoted by D and D_V, respectively. When the communication distance is larger than D/D_V but smaller than $D_{max} = D + L$ or $D_{max} = D_V + L$, the vehicle will accelerate with a to decrease the communication distance, and let L be the acceleration distance. Afterwards, the vehicle decelerates to the average speed v once it moves into the transmission range D/D_V. This assumption aims to enlarge the coverage area of the RSUs and vehicles, and this feature may not be difficult to equip on the vehicles in the future VANET. The distances L_0 and L are denoted by $L_0 = vt_0 + \frac{1}{2}at_0^2$ and $L = L_0 + v_{max} \cdot (T - t_0)$ respectively, for V2I communication, and are represented by $L_0 = \frac{1}{2}at_0^2$ and $L = L_0 + (v_{max} - v_0)(T - t_0)$ respectively, for V2V communication. v is the average speed of the vehicles, t_0 is the acceleration time, and a is the acceleration, which is denoted by $a = \frac{v_{max} - v}{t_0}$. T is the maximum waiting time which means the RSU or the vehicle waits for the communication chance for up to T seconds. Otherwise, the transmission will not be successful.

B. Vehicle Arrival Rate Prediction by ANN

We use BPNN, which is a model used in ANN to calculate a gradient value used to estimate the weights in the network, for

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2899627, IEEE Transactions on Vehicular Technology

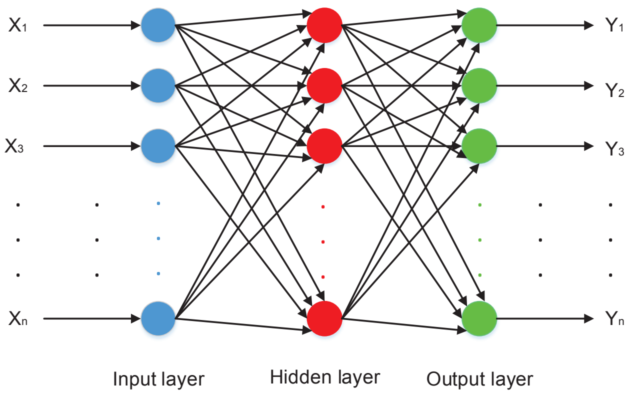


Fig. 2. Two-layer ANN diagram.

the vehicle arrival rate prediction. Fig. 2 gives an illustration of a two-layer ANN, which has one input layer, one hidden layer, and one output layer. $\mathcal{X} = \{X_1, X_2, \ldots, X_n\}$ and $\mathcal{Y} = \{Y_1, Y_2, \ldots, Y_n\}$ denote vectors in $\mathbb{R}^n$. The neural network corresponds to a function $Y = f_N(\omega, X)$, in which ω is the weight. Given a weight ω, the function f maps an input X to an output Y. The training process takes a sequence of training examples $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ as input and produces a sequence of weights $\{\omega_0, \omega_1, \ldots, \omega_n\}$ starting from some initial weight ω_0 which is usually chosen at random. We take the time $t = \{1, 2, \ldots\}$ as set $\mathcal{X}$ and the number of arrival vehicles in each unit time as set $\mathcal{Y}$. Then, we can obtain the vehicle arrival rate based on the ANN, which means that given any time t, we can predict the vehicle arrival rate. Based on the knowledge of continuous arrival rate, the arrival rate function can be obtained. Afterwards, the successful transmission rate and average delay can be estimated, which is deduced in detail in the following subsections.

C. Estimations based on Arrival Rate Function

After obtaining the arrival rate function by using ANN, the SDN controller sends it to the RSU/BS which is needed to make further routing decision.

1) V2I Communication: For uplink transmission, the source vehicle sends the packets to the RSU when it is within the extended transmission range D_{max} of RSU; otherwise the source vehicle transmits its packets to the BS. For the downlink, the successful transmission probability from RSU/BS to destination vehicle is denoted by

$$\mathbb{P}(0 < S \le D) = \int_0^D \lambda_S(x) e^{-\Lambda_S(x)} dx. \tag{18}$$

A transmission also has the probability to be successful with some time delay τ_I when the destination vehicle is in the location $S \in (D, D_{max}]$, and its probability is

$$\mathbb{P}(D < S \le D_{max}) = \int_D^{D_{max}} \lambda_S(x) e^{-\Lambda_S(x)} dx. \tag{19}$$

Accordingly, the successful transmission probability $\mathbb{P}_I$ and average delay τ_I of V2I communication are calculated by

$$\begin{aligned} \mathbb{P}_I &= \mathbb{P}(0 < S \le D) + \mathbb{P}(D < S \le D_{max}) \\ &= \int_0^{D+L} \lambda_S(x) e^{-\Lambda_S(x)} dx \end{aligned} \tag{20}$$

and

$$\begin{aligned} \tau_I = & \int_D^{D+L_0} \frac{\sqrt{v^2 + 2(x-D)} - v}{a} \cdot \lambda_S(x) e^{-\Lambda_S(x)} dx \\ & + \int_{D+L_0}^{D+L} \left(t_0 + \frac{x - (D+L_0)}{v_{max}} \right) \cdot \lambda_S(x) e^{-\Lambda_S(x)} dx \\ & + \int_{D+L}^{+\infty} T \cdot \lambda_S(x) e^{-\Lambda_S(x)} dx \end{aligned} \tag{21}$$

respectively.

2) V2V Communication: It is likely that the requests from the vehicles are beyond the serving capability of RSUs, but the vehicles are close with each other. Thus, the V2V communication can release the burden of RSUs. In V2V communication, the packets among vehicles can be delivered successfully in a multi-hop transmission manner. Transmission range of vehicles is represented by D_V, and the inter-vehicle spacings are denoted by $S_n (n = 1, 2, \ldots)$. If S_n between vehicle n and vehicle $n+1$ does not exceed transmission range D_V, the communication is immediate feasible, and vehicle $n+1$ moves at speed v. If $D_V < S_n \le D_V + L$, vehicle $n+1$ will choose to accelerate with acceleration a to catch up with the former vehicle for successful transmission. However, if $S_n > D_V + L$, transmission from vehicle n to vehicle $n+1$ will be failed.

Different from the V2I communication, we need to consider the safe driving distance for V2V communication in reality, which is denoted by L_s. Accordingly, the immediately transmission probability becomes $\mathbb{P}(L_s < S_n \le D_V)$.

The nth $(n = 1, 2, \ldots)$ hop successful transmission probability is represented by

$$\mathbb{P}_n = \begin{cases} \displaystyle\int_{L_s}^{D_V+L} \phi_1(x - L_s) dx, \quad n = 1 \\ \displaystyle\int_{L_s}^{D_V} \phi_1(x - L_s) dx \cdots \int_{L_s}^{D_V} \phi_{n-1}(x - L_s) dx \\ \displaystyle\cdot \int_{L_s}^{D_V+L} \phi_n(x - L_s) dx, \quad n \ge 2. \end{cases} \tag{22}$$

Accordingly, the average delay τ_n of the nth hop communication is calculated by

$$\begin{aligned} \tau_1 = & \int_{D_V}^{D_V+L_0} \sqrt{\frac{2(x - D_V)}{a}} \cdot \phi_1(x - L_s) dx + \int_{D_V+L_0}^{D_V+L} \Bigg(t_0 \\ & + \frac{x - (D + L_0)}{v_{max} - v} \Bigg) \cdot \phi_1(x - L_s) dx + \int_{D_V+L}^{+\infty} T \\ & \cdot \phi_1(x - L_s) dx \end{aligned} \tag{23}$$

when $n = 1$, and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2899627, IEEE Transactions on Vehicular Technology

$$\tau_n = \int_{L_s}^{D_V} \phi_{n-1}(x - L_s)dx \left[\int_{D_V}^{D_V+L_0} \sqrt{\frac{2(x - D_V)}{a}} \right.$$
$$\cdot\phi_n(x - L_s)dx + \int_{D_V+L_0}^{D_V+L} \left(t_0 + \frac{x - (D + L_0)}{v_{max} - v} \right)$$
$$\left. \cdot\phi_n(x - L_s)dx + \int_{D_V+L}^{+\infty} T \cdot \phi_n(x - L_s)dx \right] \tag{24}$$

when $n \geq 2$.

V. Proposed Routing Scheme

After receiving the requests of the vehicles, the SDN controller makes the routing decision in Algorithm 1 based on the road segment identification of the source vehicles and destination vehicles according to the vehicle indexes. The controller learns and predicts the vehicle arrival rate function using ANN and least quare fitting. Based on the function, the RSU/BS estimates the inter-vehicle spacing, and then calculates the successful transmission probability and the average time delay for each request ahead of time. By acquiring those information, the RSU/BS performs the routing algorithm in Algorithm 2, which is implemented to allocate the transmission mode (V2I or V2V) of vehicles. Bipartite matching algorithm in Algorithm 2 is a part of the routing algorithm which is adopted for further transmission mode selection that cannot be decided by the routing algorithm directly. The request delivery algorithm is introduced in detail in Algorithm 3.

A. Problem Formulation

The RSUs are located in each intersection, and the lanes are segmented by each intersection. Let each vehicle have an index $V_k(I_i, J_j)$, in which I_i and J_j denote the ith and the jth segment on each I and J lane respectively. I represents the label of westbound and eastbound lanes, and J denotes the label of northbound and southbound. When I is an odd number, the vehicle traffic is moving southbound; when I is an even number, the vehicle traffic is moving northbound. Similarly, when J is an odd number, the vehicle traffic is moving eastbound; when J is an even number, the vehicle traffic is moving westbound. k represents the order of the vehicles in each road segment. The vehicle updates its index $V_k(I_i, J_j)$ to the SDN controller via the RSU while passing the intersection, and thus the SDN controller knows the order of the vehicles on each road segment.

We denote the number of requests and RSUs by X and Y, respectively. $\mathbb{X} = \{1, 2, \ldots, X\}$ maps the elements in request set $\mathcal{K} = \{(V_S^1, V_D^1), (V_S^2, V_D^2), \ldots, (V_S^X, V_D^X)\}$. Set $V_S^x = \{I_m^x, i_m^x, J_m^x, j_m^x, k_m^x\}$ and set $V_D^x = \{I_n^x, i_n^x, J_n^x, j_n^x, k_n^x\}$, where k_m and k_n denote the sequence numbers of source vehicle and destination vehicle in each road segment. Thus, set V_S^x and V_D^x contain the road segment information of request x corresponding to the source vehicles and destination vehicles, respectively. $\mathbb{Y} = \{0, 1, 2, \ldots, Y\}$ denotes the BS and the RSUs, and the index is correlated to I_i and J_j. Set $\mathcal{K}$ is divided into two subsets: $\mathcal{V}_V$ and $\mathcal{V}_I$, which denote the set of vehicles in the V2V mode and V2I mode, respectively. $\mathcal{V}_V \bigcup \mathcal{V}_I = \mathcal{K}$ and $\mathcal{V}_I = \sum_{y=0}^{Y} \bigcup \mathcal{V}_{I_y}$.

For each request allocated to V2I communication, while the source vehicle and destination vehicle are located in the same road segment, the delay of V2I mode is represented by

$$T_{I_{m,n}} = \frac{Z}{R_m^y} + \frac{Z}{R_n^y} + \tau_m + \tau_n \tag{25}$$

where Z denotes the request packet size. R_m^y and R_n^y are the uplink transmission rate from vehicle m to RSU y and the downlink transmission rate from RSU y to vehicle n, which are given by (15) and (16), respectively. τ_m and τ_n are given by (21).

When the source vehicle and destination vehicle are not located in the same road segment, the delay of V2I mode is represented by

$$\widetilde{T}_{I_{m,n}} = \frac{Z}{R_m^y} + \frac{Z}{R_n^{\tilde{y}}} + \frac{Z}{R_y^{\tilde{y}}} + \tau_m + \tau_n \tag{26}$$

where $R_y^{\tilde{y}}$ denotes the fixed RSU-RSU or RSU-BS transmission rate through wired link.

The delay of V2V transmission mode is calculated by

$$T_{V_{m,n}} = \sum_{h=1}^{H} \left(\frac{Z}{R_{V_h}} + \tau_h \right) \tag{27}$$

where H is the number of hops transmitting from vehicle m to vehicle n, and H is determined by the successful transmission probability of the hth hop. R_{V_h} and τ_h are given in (17) and (24), respectively.

The objective of routing is to minimize the overall vehicular service delay, and the routing problem can be formulated as follows:

$$\begin{aligned}
(\mathcal{P}_1): \ \underset{\alpha_{x,y}, \beta_k}{\text{minimize}} \quad & \sum_{x=1}^{X} \sum_{y=0}^{Y} \sum_{k=1}^{2} \alpha_{x,y} \beta_k T_{x,y} \\
\text{subject to} \quad & \sum_{x=1}^{X} \alpha_{x,y} \beta_1 \leq C \\
& \sum_{y=0}^{Y} \alpha_{x,y} \beta_1 \leq 2 \\
& \sum_{k=1}^{2} \beta_k = 1 \\
& \alpha_{x,y} \in \{0, 1\} \\
& \beta_k \in \{0, 1\}
\end{aligned} \tag{28}$$

where $\alpha_{x,y}$ and β_k are binary variables. If $\alpha_{x,y} = 1$, it indicates that request x can be allocated to transmit with RSU y, which is the BS when $y = 0$. Otherwise, $\alpha_{x,y} = 0$. $\beta_1 = 1$ represents V2I transmission mode is selected, and $\beta_2 = 1$ denotes V2V transmission mode is selected. C is the maximum number of vehicles the RSU/BS can serve, and the first constraint means that each RSU/BS can not serve more than C vehicles. The second constraint denotes that the request can be associated with at most two RSUs or one BS and one RSU since the source vehicle and destination vehicle may be located

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2899627, IEEE Transactions on Vehicular Technology

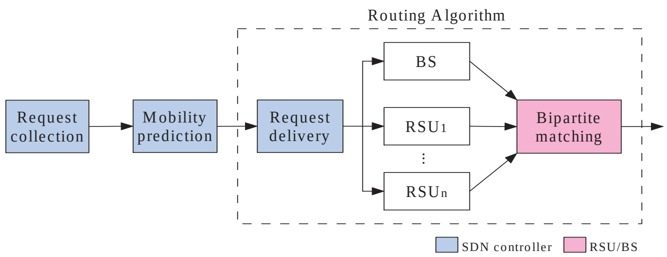


Fig. 3. Procedure used for solving the optimization problem in (28).

in different road segments. The third constraint represents only one communication mode, in which either V2I or V2V can be selected for each request. $T_{x,y} = \min\{T_{I_{m,n}}, T_{V_{m,n}}\}$ or $\widetilde{T}_{I_{m,n}}$.

The problem $\mathcal{P}_1$ is an integer programming problem which is NP-hard. To solve the problem $\mathcal{P}_1$, we propose a CRS-MP algorithm in which a routing algorithm is first performed to decide the transmission mode of the requests which can be allocated directly. Afterwards, the optimization problem is converted to $\mathcal{P}_2$, which can be solved by bipartite matching algorithm. The routing from one RSU to another can be completed by the request delivery algorithm. Fig. 3 summarizes the proposed framework that is used to solve the problem in (28). The actions in the blue boxes are performed in the SDN controller, while the action in the pink box is performed in the BS/RSU.

B. Routing Algorithm

When the SDN controller performs the road segment identification, there are five cases for the locations of the source vehicle and its destination.

1) Both source vehicle and its destination are located in the transmission range of the same RSU. The SDN controller informs the associated RSU to make the selection of the transmission mode either V2I or V2V for each vehicle by bipartite matching algorithm.
2) The source vehicle and its destination are located in the transmission range of different RSUs. V2I communication mode is allocated to the vehicles with the help of RSU-RSU communication. First, the source vehicle sends the data to its corresponding RSU. Then, the RSU transmits the received data to the RSU which is associated with the destination vehicle. Finally, the RSU transmits data to the destination vehicle.
3) The source vehicle is located in the transmission range of RSU, while the destination vehicle is out of the coverage area of any RSU. First, the source vehicle sends the data to its corresponding RSU. Then, the RSU transmits the received data to the BS. Finally, the BS transmits data to the destination vehicle.
4) The destination vehicle exists in the transmission range of RSU, while the source vehicle is out of the coverage area of any RSU. The source vehicle transmits data to the BS, and then the BS transmits data to the associated RSU of the destination vehicle. Afterwards, the RSU delivers data to the destination vehicle.
5) Both source vehicle and its destination are out of the transmission range of any RSU. When they are located in the same road segment, bipartite matching algorithm is performed for either V2V or V2I communication mode, which means selecting multi-hop transmissions from the source vehicle to its destination or the source vehicle and destination vehicle communicate through the BS.

Based on aforementioned five cases, the routing algorithm is described in Algorithm 1.

Algorithm 1 Routing algorithm.

Input:
X, Y, $\mathcal{K}$, N, H_{max}, t_u, t_d, L_{I_i,J_j};
// X is the number of requests, Y is the number of RSUs, $\mathcal{K}$ represents the index information of the vehicles, N denotes the number of vehicles in each road segment within the transmission range of RSUs, H_{max} is the maximum number of hops for V2V communication, t_u is the vehicle index update time, t_d is the time for decision making; L_{I_i,J_j} denotes the road segment length;
Output:
The routing path results $\mathcal{V}_I = \sum_{y=0}^{Y} \bigcup \mathcal{V}_{I_y}$ and $\mathcal{V}_V$;
Initially: $\mathcal{V}_I \leftarrow \emptyset$;
// The set of vehicles which choose V2I communication is initially empty;
Initially: $\mathcal{V}_V \leftarrow \emptyset$;
// The set of vehicles which choose V2V communication is initially empty;
while $\mathcal{K} \neq \emptyset$ **do**
  if $\{(I_m^x = I_n^x) \bigcap (i_m^x = i_n^x)\} \bigcup \{(J_m^x = J_n^x) \bigcap (j_m^x = j_n^x)\}$ **then**
    switch $((k_m^x \leq N) \bigcap (k_n^x \leq N))$
    case 0**:**
      if $k_n - k_m \leq H_{max}$ **then**
        $T_{x,y} = T_{V_{m,n}}$, $\mathcal{V}_V \leftarrow V_S^x, V_D^x$
      else
        $T_{x,y} = T_{I_{m,n}}$, $\mathcal{V}_{I_y} \leftarrow V_S^x, V_D^x$
      end if
    case 1**:**
      $T_{x,y} = \min\{T_{I_{m,n}}, T_{V_{m,n}}\}$, and go to Algorithm 2;
    end switch
  else
    $T_{x,y} = \widetilde{T}_{I_{m,n}}$, $\mathcal{V}_{I_y} \leftarrow V_S^x, V_D^x$
  end if
  Delete V_S^x and V_D^x from set $\mathcal{K}$.
end while
if $(t_d - t_u)v \leq D_{max}$ **then**
  $y = y$
else if $(t_d - t_u)v \geq L_{I_i,J_j} - D_{max}$ **then**
  $y = y + 1$
else
  $y = 0$
end if

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2899627, IEEE Transactions on Vehicular Technology

C. Bipartite Matching Algorithm

When the source vehicle and its destination are located in the same road segment, the transmission mode selection becomes more tricky since both V2V and V2I can be selected for the vehicles. Therefore, after implementing Algorithm 1, the optimization problem $\mathcal{P}_1$ is converted to

$$\begin{aligned}(\mathcal{P}_2): \underset{\alpha_x,\beta_k}{\text{minimize}} \quad & \sum_{x=1}^{X}\sum_{k=1}^{2}\alpha_x\beta_k T_x \\ \text{subject to} \quad & \sum_{x=1}^{X}\alpha_x\beta_1 \leq C \\ & \sum_{k=1}^{2}\beta_k = 1 \\ & \alpha_x \in \{0,1\} \\ & \beta_k \in \{0,1\}\end{aligned} \tag{29}$$

where α_x and β_k are the binary variables. If $\alpha_x = 1$, it means the request x needs to be allocated. Otherwise, $\alpha_x = 0$. $T_x = T_{x,y}$ and $T_{x,y}$ is given in line 13 of Algorithm 1.

The optimization problem of the transmission mode selection described in (29) can be rewritten as (30), which is given by

$$\begin{aligned}(\mathcal{P}_3): \underset{\alpha_x,\beta_k}{\text{maximize}} \quad & \sum_{x=1}^{X}\sum_{k=1}^{2}\alpha_x\beta_k\left(-T_x\right) \\ \text{subject to} \quad & \sum_{x=1}^{X}\alpha_x\beta_1 \leq C \\ & \sum_{k=1}^{2}\beta_k = 1 \\ & \alpha_x \in \{0,1\} \\ & \beta_k \in \{0,1\}.\end{aligned} \tag{30}$$

We convert (30) into a bipartite weighted matching problem on an $X \times M$ bipartite graph indicating X requests, M routing paths ($M = C + X$). This bipartite matching problem aims to find the best matching so that the sum of edges can be maximized, and the matching result can be obtained in polynomial time in a centralized network. An edge represents the transmission delay of the transmission mode when transmitting data from source vehicle to its destination. The weight of the edge is given by $-T_x$, which forms the weight matrix Φ.

The requests and transmission modes form a set V of vertexes in the graph, and E denotes the transmission mode selection in terms of transmission delay. $\phi(x,y) \in \Phi$ is a weight matrix for matching V and E in the bipartite graph, and the weights $\phi(x,y)$ denote the total delay per request. $G = (V,E)$ is a bipartite graph, where $x \in \mathcal{S}$, $y \in \mathcal{T}$, $V = \mathcal{S}\bigcup\mathcal{T}$, and $(x,y) \in E$. Given weights on the edges $w(x,y)$ from matrix Φ for each $(x,y) \in E$, we define the feasible labelling as a function:

$\ell(x) + \ell(y) \geqslant w(x,y)$;
$\ell(x) = \max_{y\in\mathcal{T}} w(x,y)$, if $x \in \mathcal{S}$;
$\ell(y) = 0$, if $y \in \mathcal{T}$;
$E_\ell = \{(x,y) \in E | \ell(x) + \ell(y) = w(x,y)\}$.

Hence, an equality subgraph of $G = G_\ell$ has been defined, i.e., $G_\ell = (V, E_\ell)$. By using equality subgraph, any maximum matching $\mathcal{M}$ is an optimal matching $\mathcal{M}^*$. If we do not have a maximum matching, we have to modify $G_\ell \to G_{\ell^*}$ by updating $\ell^*(v)$ and continue to find another matching $\mathcal{M}$ in G_{ℓ^*}. Then, the matching result with the minimum delay can be acquired. The bipartite matching algorithm is given in detail in Algorithm 2.

Algorithm 2 Bipartite matching algorithm.

Input:
A set of requests $\mathcal{S}$ ($\mathcal{S} \subseteq \mathcal{K}$), a set of transmission mode candidate $\mathcal{T}$, and a weight matrix Φ;
Output:
The routing path results, an optimal matching result $\mathcal{M}^*$;
1: Establish a weighted bipartite graph $G(V,E)$ and each edge $(x,y) \in E$ has a weight $\phi(x,y)$;
2: Randomly choose a matching $\mathcal{M}$ in $G(V,E)$;
3: **if** set $\mathcal{T}$ is saturated **then**
4: current $\mathcal{M}$ is the optimal matching $\mathcal{M}^*$;
5: **else**
6: flip current matching $\mathcal{M}$, and update $G(V,E)$;
7: **end if**
8: **repeat**
9: step 3 – step 7;
10: **until** find the maximum weighted matching $\mathcal{M}^*$.

The maximum weighted matching updated procedure can be stated by the following three steps.

1) Starting with $\ell(x)$, $\ell(y)$ as above, determine G_ℓ and choose Q in G_ℓ. Let u be an unsaturated vertex in X, $\mathcal{S} \leftarrow \{u\}$, $\mathcal{T} \leftarrow 0$.
2) If $N_{G_\ell(s)} \supset \mathcal{T}$, go to step 3); else $N_{G_\ell(s)} = \mathcal{T}$, $\nexists Q^*$ in G_ℓ, determine ℓ^* to modify G_ℓ and compute $\alpha_\ell = \min_{x\in\mathcal{S},y\notin\mathcal{T}}\{\ell(x) + \ell(y) - w(x,y)\}$. If $v \in \mathcal{S}$, $\ell^*(v) = \ell(v) - \alpha_\ell$; if $v \in \mathcal{T}$, $\ell^*(v) = \ell(v) + \alpha_\ell$; otherwise, $\ell^*(v) = \ell(v)$. Update $\ell(v)$ by $\ell^*(v)$, and G_ℓ by G_{ℓ^*}.
3) Expand matching, choose y in $N_{G_\ell(s)} \setminus \mathcal{T}$. If y is saturated, with $x, y \in Q$, replace $\mathcal{S}$ by $\mathcal{S} \cup \{x\}$, $\mathcal{T}$ by $\mathcal{T} \cup \{y\}$ and go to step 2); else let Q be augmenting (u,y) path in G_ℓ, and substitute $\mathcal{M}$ by Q.

D. Request Delivery

Each vehicle has an index $V_k(I_i, J_j)$, and the indexes are initialized immediately when the vehicles enter the lanes. As mentioned in the problem formulation, the parity of the index reflects the moving direction of each vehicle, and the index will be updated after the vehicles pass the intersections. At each intersection, the actions, such as turning left and turning right, adopted by each vehicle can be told through the changing of the vehicle index. For example, we assume that the vehicle's original index is $V_{old}(2,0)$. If the updated index is $V_{new}(0,3)$, the vehicle turns right; if the updated index is $V_{new}(0,2)$, the vehicle turns left.

Since the routing decision is made based on the index information of vehicles, there is a gap between index update

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2899627, IEEE Transactions on Vehicular Technology

time and decision making time. When the gap is nontrivial, the SDN controller needs to deliver the request to the corresponding RSU/BS who serves the vehicles at the decision time, and the delivery is based on the moving direction of the vehicles. According to the above statement, we specify the request delivery strategy in Algorithm 3.

Algorithm 3 Request delivery algorithm.

Input:
Vehicle index $V(I,i,J,j,k)$; Initialize $k=1$, $I=0$ and $J=0$;

Output:
The request delivery direction $d_k(\eta,\theta)$, i.e., east: $d_k(0,0)$, west: $d_k(0,1)$, south: $d_k(1,0)$ and north: $d_k(1,1)$.

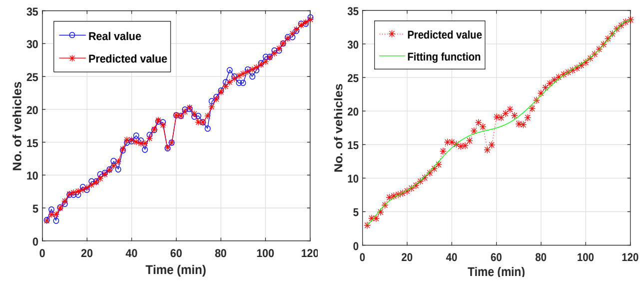
while Vehicle k passes the previous intersection Y_{I_i,J_j} **do**
  RSU Y_{I_i,J_j} has the updated index $V(I,i,J,j,k)$ of vehicle k;
  if $I=0$ **then**
    $\eta=0$, $\theta=0$ if J is an even number;
    $\eta=0$, $\theta=1$ if J is an odd number;
  else
    $\eta=1$, $\theta=0$ if I is an odd number;
    $\eta=1$, $\theta=1$ if I is an even number;
    $k=k+1$;
  end if
end while
All of the requests are delivered to the right RSU/BS.

VI. Numerical Results

The simulation model is built based on the system architecture described in Section III. We consider that all the vehicles drive in the same direction and the arrival rate of vehicles in each lane follows the NHPP. A wide range of vehicle speeds are simulated to evaluate the system performance under different scenarios. In each lane, the vehicle speed is selected by the given specific vehicle arrival rate function. The communication characteristics are simulated based on DSRC for V2V communication, and LTE for V2I and I2I communications. The scheduling period is the same as the maximum waiting time. Each vehicle can submit a request to the SDN controller at any time. The total number of submitted requests varies from 10% to 90% out of the total number of vehicles in each lane. Detailed vehicular system parameters are summarized in Table II.

Table II
Simulation Parameter

Parameter	Value
Mobility model	Non-homogeneous Poisson process
Transmission protocol	LTE for V2I & DSRC for V2V
RSU transmission range	500 m
Vehicle transmission range	200 m
Serving capability of RSU	10
Transmission power of RSU	20 dBm
Transmission power of vehicle	10 dBm
Bandwidth	10 MHz
Packet size	5 Mbit
Number of iterations	1000



(a) ANN prediction with 20 epochs. (b) Least square fitting.

Fig. 4. Mobility prediction.

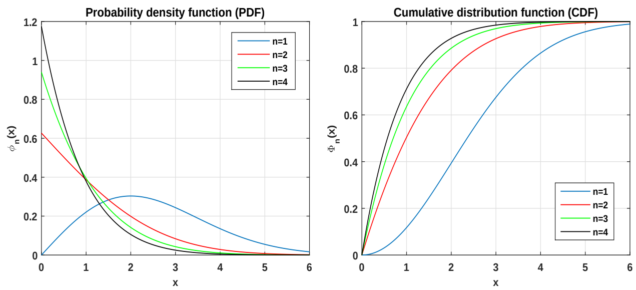
The proposed ANN model is applied to the data collected from the Coquitlam database as a numerical example. The traffic data are collected every 1 min from individual detectors along the roads. We adopt the data from the vehicle volumes in the intersection of Eagleridge Dr. and Lansdowne Dr., Coquitlam, BC, Canada. The traffic flow data collected in the weekdays of the year 2016 are divided into two groups: one for training and the other one for testing. Training parameters are set as follows: training goal is 10^{-1}, training epoch is 20, and learning rate is 0.01. During the training process, the thresholds and weights of each unit in the network are adjusted iteratively in such a way that the error between the actual output and the desired output is reduced. We apply ANN to predict the vehicle arrivals with a setting of 20 epochs, and then use the least square method to fit the arrival rate function.

Testing results and fitting function are shown in Fig. 4. The vehicle arrival rate is predicted by ANN as shown in Fig. 4(a) and the arrival rate function is fitted as shown in Fig. 4(b). The x-axis represents the time that varies within two hours, and the y-axis denotes the number of vehicle arrivals per minute. Blue circles are the real vehicle arrival rate, and red stars are the predicted vehicle arrival rate. The green line is the function we fit based on the predicted data. The iterations of least quare method is set to 20, and the obtained arrival rate function is $\lambda(t) = 1.6415\times10^{-13}x^{10} - 5.4755\times10^{-11}x^9 + 7.7203\times10^{-09}x^8 - 5.9885\times10^{-7}x^7 + 2.7844\times10^{-5}x^6 - 0.00079247x^5 + 0.013523x^4 - 0.12985x^3 + 0.61408x^2 - 0.49383x + 3.0104$.

Fig. 5 shows the probability density function (PDF) and cumulative distribution function (CDF) of the inter-vehicle spacing respectively. n represents the nth inter-vehicle spacing between vehicle n and vehicle $n+1$. In order to simplify the analysis, we set the vehicle arrival function as a linear function $\lambda(t) = a\cdot t + b$, in which $a = 6.25\times10^{-5}, b = 0.05$. Accordingly, the vehicle arrival rate varies from 0.05 (vehicle/sec) to 0.5 (vehicle/sec) within 2 hours. As shown in the figures, we can see that as n increases, the inver-vehicle spacing becomes smaller.

The successful transmission probability of V2V is shown in Fig. 6. It can be seen that, successful transmission probability decreases with the increase of the number of hops. Moreover, the higher vehicle arrival rate corresponds to higher successful transmission probability.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2899627, IEEE Transactions on Vehicular Technology



(a) PDF of the inter-vehicle spacing. (b) CDF of the inter-vehicle spacing

Fig. 5. PDF and CDF of the inter-vehicle spacing.

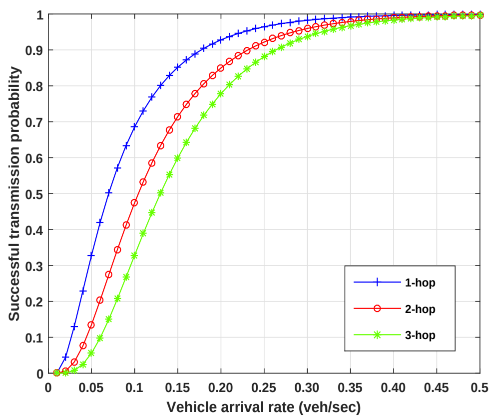


Fig. 6. V2V n-hop successful transmission probability.

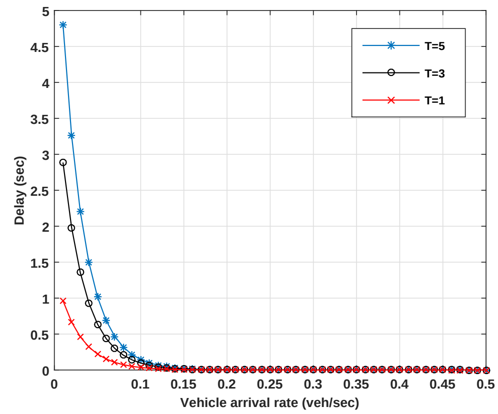


Fig. 7. V2I average delay.

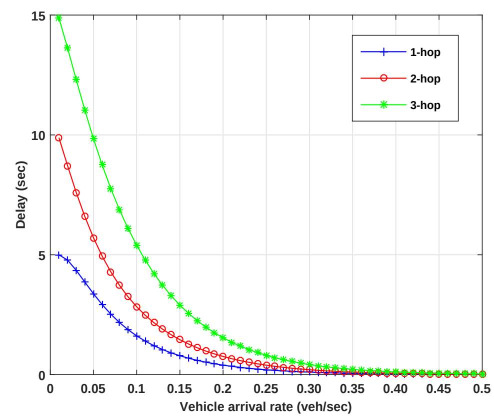


Fig. 8. V2V average delay.

The average delays of V2I and V2V transmission mode are shown in Fig. 7 and Fig. 8, respectively. In Fig. 7, we compare three cases with different maximum waiting time: $T = 1$, $T = 3$, and $T = 5$. A longer waiting time leads to a higher average delay. The V2I average delay decreases to approximately 0 as the vehicle arrival rate reaches 0.15 (veh/sec) for all the three cases. Moreover, when the vehicle arrival rate is smaller than 0.1, the maximum waiting time has more impact on the average delay. In this case, the lower the vehicle arrival rate and the larger the maximum waiting time, the higher the average delay. We set the maximum waiting time $T = 5$ in Fig. 8. The average delay increases with the increase of the number of hops, and the delay decreases to almost 0 when the vehicle arrival rate is lager than 0.25 (veh/sec), since it is the rush hour when the vehicle density should be high.

The parameter settings of our proposed CRS-MP algorithm simulation are listed in Table II. The simulation is repeated 1000 times, and the average delay is calculated. The number of vehicles is 40, and the maximum accessing number of vehicles (serving capacity) of RSU is 20. In the simulation, we assume that 60% of the vehicles will send requests and these vehicles are randomly selected. The simulation results are shown in Fig. 9. The objective of our proposed CRS-MP is to minimize the overall vehicular service delay in comparison with V2I mode and V2V mode. In V2I mode, all the vehicles communicate through the RSU with the help of controller, while all the vehicles communicate with each other through multi-hop transmission without the help of RSU in V2V mode. The performance of traditional routing schemes generally degrade when the speed increases [36]. However, the performance of our centralized routing scheme is more robust since it does not degrade much when the vehicle speed increases, as shown in Fig. 9 where we vary the speed from 13.9 m/s (50 km/h) to 33.3 m/s (120 km/h).

In Fig. 10, the average vehicle speed in the lane is set to 13.9 m/s (50 km/h), and the speed limit is 16.7 m/s (60 km/h). The number of vehicles is set to 40, 80 and 120 in each sub-figure, respectively. In each sub-figure, we vary the request fraction with 10%, 50% and 90%. The serving capacity of RSU is 100. Fig. 10 shows that our proposed CRS-MP scheme outperforms the routing scheme without mobility prediction as it achieves lower delay when the number of vehicles increases under different request fraction scenarios. Moreover, the delay of the scheme without mobility prediction increases faster than our proposed routing scheme when the request fraction becomes higher since the serving capacity of RSU is limited. When the number of requesting vehicles is close to the serving capacity of RSU, the subsequent vehicles have to wait to access the RSU. Therefore, the overall vehicular service delay increases faster without mobility prediction.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2899627, IEEE Transactions on Vehicular Technology

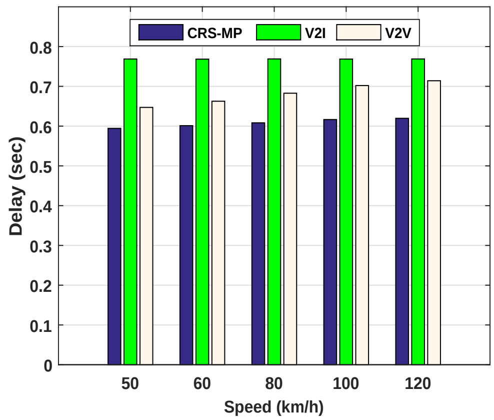


Fig. 9. Comparison of the delay among CRS-MP, V2I and V2V with different speeds.

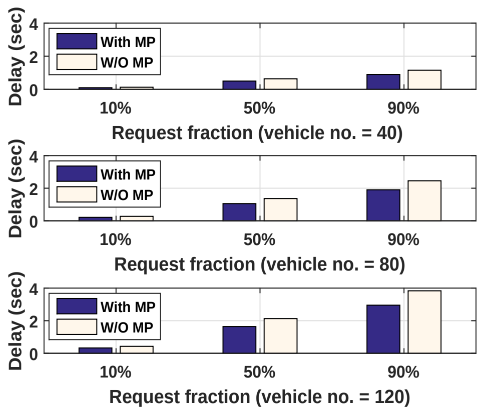


Fig. 10. Comparison of the delay of routing scheme with and without mobility prediction with different numbers of vehicles under different request fraction scenarios.

VII. Concluding Remarks

In this paper, we have proposed a centralized routing scheme for end-to-end unicast communication in VANET. The proposed routing scheme has the prediction capability and selects the optimal routing path based on the global information. To adapt to the dynamic changing network topology, the proposed routing scheme can choose either V2I or V2V communication. We have simulated our proposed routing scheme with NHPP, and compared it with other routing protocols (pure V2I and pure V2V) in VANET. Simulation results have shown that our proposed CRS-MP scheme outperforms other routing schemes in terms of overall vehicular service delay. Besides, the proposed scheme is more robust when the vehicle speeds vary. For our future work, we will implement practical vehicle arrival model in order to make the mobility prediction more accurate for enhancing the routing performance of VANET.

References

[1] W. Xu, H. Zhou, N. Cheng, F. Lyu, W. Shi, J. Chen, and X. Shen, "Internet of vehicles in big data era," *IEEE/CAA Journal of Automatica Sinica*, vol. 5, no. 1, pp. 18–35, Jan. 2018.

[2] N. Cheng, L. Feng, J. Chen, W. Xu, H. Zhou, S. Zhang, and X. Shen, "Big data driven vehicular networks," *IEEE Network*, vol. 32, no. 6, pp. 160–167, Dec. 2018.

[3] FCC, "Amendment of the Commission's rules regarding dedicated short-range communication services in the 5.850–5.925 GHz band," *FCC report and order 06–110*, Jul. 2006.

[4] Y. L. Morgan, "Notes on DSRC & WAVE standards suite: Its architecture, design, and characteristics," *IEEE Communication Surveys & Tutorials*, vol. 12, no. 4, pp. 504–518, May 2010.

[5] T. Clausen and P. Jacquet, "Optimized link state routing protocol (OLSR)," *RFC 3626 (Experimental)*, Oct. 2003.

[6] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc on-demand distance vector (AODV) routing," *RFC 3561 (Experimental)*, Jul. 2003.

[7] D. Johnson, Y. Hu, and D. Maltz, "The dynamic source routing protocol (DSR) for mobile ad hoc networks for IPv4," *RFC 4728 (Experimental)*, Feb. 2007.

[8] D. Tian, K. Zheng, J. Zhou, X. Duan, Y. Wang, Z. Sheng, and Q. Ni, "A microbial inspired routing protocol for VANETs," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2293–2303, Aug. 2018.

[9] N. Alsharif and X. Shen, "iCAR-II: infrastructure-based connectivity aware routing in vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 4231–4244, May 2017.

[10] J. He, L. Cai, J. Pan, and P. Cheng, "Delay analysis and routing for two-dimensional VANETs using carry-and-forward mechanism," *IEEE Transactions on Mobile Computing*, vol. 16, no. 7, pp. 1830–1841, Jul. 2017.

[11] ONF, "Software-defined networking: The new norm for networks," *Open Networking Foundation White Paper*, Apr. 2012.

[12] Q. Ye, J. Li, K. Qu, W. Zhuang, X. Shen, and X. Li, "End-to-end quality of service in 5G networks-Examining the effectiveness of a network slicing framework," *IEEE Vehicular Technology Magazine*, vol. 13, no. 3, pp. 65–74, Jun. 2018.

[13] L. Zhu, C. Li, Y. Wang, Z. Luo, Z. Liu, B. Li, and X. Wang, "On stochastic analysis of greedy routing in vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3353–3366, Dec. 2015.

[14] S.-C. Kochar, "Some results on interarrival times of nonhomogeneous Poisson processes," *Probability in the Engineering and Informational Sciences*, vol. 10, no. 1, pp. 75–85, Jan. 1996.

[15] F. Pellerey, M. Shaked, and J. Zinn, "Nonhomogeneous Poisson processes and logconcavity," *Probability in the Engineering and Informational Sciences*, vol. 14, no. 3, pp. 353–373, Jul. 2000.

[16] A. M. Mezher and M. A. Igartua, "Multimedia multimetric map-aware routing protocol to send video-reporting messages over VANETs in smart cities," *IEEE Transcations on Vehicular Technology*, vol. 66, no. 12, pp. 10611–10625, Dec. 2017.

[17] N. Li, J.-F. Martínez-Ortega, V. H. Díaz, and J. A. S. Fernandez, "Probability prediction-based reliable and efficient opportunistic routing algorithm for VANETs," *IEEE/ACM Transcations on Networking*, vol. 26, no. 4, pp. 1933–1947, Aug. 2018.

[18] M. H. Eiza, T. Owens, Q. Ni, and Q. Shi, "Situation-aware QoS routing algorithm for vehicular ad hoc networks," *IEEE Transcations on Vehicular Technology*, vol. 64, no. 12, pp. 5520–5535, Dec. 2015.

[19] M. A. Togou, A. Hafid, and L. Khoukhi, "SCRP: Stable CDS-based routing protocol for urban vehicular ad hoc networks," *IEEE Transcations on Intelligent Transportation Systems*, vol. 17, no. 5, pp. 1298–1307, May 2016.

[20] M. A. Salkuyeh and B. Abolhassani, "An adaptive multipath geographic routing for video transmission in urban VANETs," *IEEE Transcations on Intelligent Transportation Systems*, vol. 17, no. 10, pp. 2822–2831, Oct. 2016.

[21] H. Omar, W. Zhuang, and L. Li, "Gateway placement and packet routing for multihop in-vehicle internet access," *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 3, pp.335-351, Sept. 2015.

[22] D. Lin, J. Kang, A. Squicciarini, Y. Wu, S. Gurung, and O. Tonguz, "MoZo: a moving zone based routing protocol using pure V2V communication in VANETs," *IEEE Transactions on Mobile Computing*, vol. 16, no. 5, pp. 1357–1370, May 2017.

[23] H. Zhu, X. Lin, R. Lu, Y. Fan, and X. Shen, "SMART: a secure multilayer credit-based incentive scheme for delay-tolerant networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 8, pp. 4628–4639, Oct. 2009.

[24] H. Li, H. Zhu, S. Du, X. Liang, and X. Shen, "Privacy leakage of location sharing in mobile social networks: Attacks and defense," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 646–660, Aug. 2018.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2899627, IEEE Transactions on Vehicular Technology

[25] H. Zhu, C. Fang, Y. Liu, C. Chen, M. Li, and X. Shen, "You can jam but you cannot hide: Defending against jamming attacks for geo-location database driven spectrum sharing," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 10, pp. 2723–2737, Oct. 2016.

[26] A. Destounis, S. Paris, L. Maggi, G. S. Paschos, and J. Leguay, "Minimum cost SDN routing with reconfiguration frequency constraints," *IEEE/ACM Transcations on Networking*, vol. 26, no. 4, pp. 1577–1590, Aug. 2018.

[27] X. Duan, Y. Liu, and X. Wang, "SDN enabled 5G-VANET: Adaptive vehicle clustering and beamformed transmission for aggregated traffic," *IEEE Communications Magazine*, vol. 55, no. 7, pp. 120–127, Jul. 2017.

[28] K. Liu, J. K. Y. Ng, V. C.S. Lee, S. H. Son, and I. Stojmenovic, "Cooperative data scheduling in hybrid vehicular ad hoc networks: VANET as a software defined network," *IEEE/ACM Transactions on Networking*, vol. 24, no. 3, pp. 1759–1773, Jun. 2016.

[29] Y. Liu, C. Chen, and S. Chakraborty, "A software defined network architecture for geoBroadcast in VANETs," *IEEE/ACM Transactions on Networking*, pp. 6559–6564, Jun. 2015.

[30] X. Shen, J. W. Mark, and J. Ye, "User mobility profile prediction: An adaptive fuzzy inference approach," *Wireless Networks*, vol. 6, no. 5, pp. 363–374, Nov. 2000.

[31] Y. Tang, Q. Zhang, and W. Lin, "Artificial neural network based spectrum sensing method for cognitive radio," in *Proceedings of International Conference on Wireless Communications Networking and Mobile Computing*, Sept. 2010, pp. 1–4.

[32] H.-F. Yang, T. S. Dillon, and Y.-P. Chen, "Optimized structure of the traffic flow forecasting model with a deep learning approach," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2371–2381, Oct. 2017.

[33] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865–873, Apr. 2015.

[34] S. Ucar, S. C. Ergen, and O. Ozkasap, "Multihop-cluster-based IEEE 802.11p and LTE hybrid architecture for VANET safety message dissemination," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 4, pp. 2621–2636, Apr. 2016.

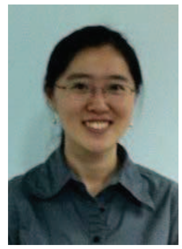
[35] L. Baxter, "Reliability applications of the relevation transform," *Naval Research Logistics Quart*, vol. 29, no. 2, pp. 323–329, Jun. 1982.

[36] P. Fazio, F. D. Rango, and C. Sottile, "A predictive cross-layered interference management in a multichannel MAC with reactive routing in VANET," *IEEE Transactions on Mobile Computing*, vol. 15, no. 8, pp. 1850–1862, Aug. 2016.

Yujie Tang (S'12, M'17) received the Ph.D. degree in Electrical and Computer Engineering from the University of Waterloo, Canada. She received the B.E. degree in Communications Engineering from Lanzhou Jiaotong University, and the M.Sc. degree in Information and Communications Engineering from Harbin Institute of Technology, China. She is currently a Post-doctoral fellow with the University of Waterloo. Her research interests include vehicular ad hoc networks, software-defined networks, machine learning, cognitive radio networks, cooperative networks, and resource management in heterogeneous networks.

Nan Cheng (S'12, M'16) received the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Waterloo, and B.E. degree and the M.S. degree from the Department of Electronics and Information Engineering, Tongji University. He is currently working as a joint Post-doctoral fellow with the Department of Electrical and Computer Engineering, University of Toronto and the Department of Electrical and Computer Engineering, University of Waterloo. His research interests include performance analysis, MAC, opportunistic communication for vehicular networks, unmanned aerial vehicles, and cellular traffic offloading.

Wen Wu (S'13) received the B.E. degree from South China University of Technology, Guangzhou, China, in 2012 and the M.A.Sc. degree in electrical engineering from University of Science and Technology of China, Hefei, China, in 2015. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo. His research interests include millimeter-wave wireless local area networks, interference alignment and management, and millimeter-wave fifth-generation cellular networks.



Miao Wang (M'15) is an Assistant Professor with the Department of Electrical and Computer Engineering, Miami University, Ohio, USA. She received her B.Sc. degree from Beijing University of Posts and Telecommunications and M.Sc. degree from Beihang University, Beijing, China, in 2007 and 2010, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2015. Her current research interests include electric vehicles charging/discharging strategy design in smart grid, traffic control, capacity and delay analysis, and routing protocol design for vehicular networks.

Yanpeng Dai received the B.S. degree in telecommunication engineering from Shandong Normal University, Jinan, China, in 2014. He is currently pursuing the Ph.D degree at the State Key Laboratory of Integrated Service Networks, Xidian university. He was a visiting student in University of Waterloo, Canada (Sept. 2017-Sept. 2018). His research interests include resource allocation and interference management for non-orthogonal multiple access system in wireless networks.

Xuemin (Sherman) Shen (M'97-SM'02-F'09) received the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990. He is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research focuses on resource management in interconnected wireless/wired networks, wireless network security, social networks, smart grid, and vehicular ad hoc and sensor networks. He is a registered Professional Engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, and a Distinguished Lecturer of the IEEE Vehicular Technology Society and Communications Society.

Dr. Shen received the James Evans Avant Garde Award in 2018 from the IEEE Vehicular Technology Society, the Joseph LoCicero Award in 2015 and the Education Award in 2017 from the IEEE Communications Society. He has also received the Excellent Graduate Supervision Award in 2006 and the Outstanding Performance Award in 2004, 2007, 2010, and 2014 from the University of Waterloo and the Premiers Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada. He served as the Technical Program Committee Chair/Co-Chair for the IEEE Globecom16, the IEEE Infocom14, the IEEE VTC10 Fall, the IEEE Globecom07, the Symposia Chair for the IEEE ICC10, the Tutorial Chair for the IEEE VTC11 Spring, the Chair for the IEEE Communications Society Technical Committee on Wireless Communications, and P2P Communications and Networking. He is the Editor-in-Chief of the IEEE INTERNET OF THINGS JOURNAL and the Vice President on Publications of the IEEE Communications Society.