Real-Time Path Planning Based on Hybrid-VANET-Enhanced Transportation System

Miao Wang*, Hanguan Shan†, Rongxing Lu‡, Ran Zhang*, Xuemin (Sherman) Shen*, Fan Bai§

*Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada
†Department of ISEE, Zhejiang University, Hangzhou, China
‡Communication Engineering School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore
§ECI Lab, General Motors Global RD, Warren, MI, USA

Abstract—Real-time path planning can efficiently relieve traffic congestion in urban scenarios. However, how to design an efficient path planning algorithm to achieve a globally optimal vehicle-traffic control still remains a challenging problem, especially when we take drivers’ individual preferences into consideration. In this paper, we first establish a hybrid intelligent transportation system (ITS), i.e., a hybrid-VANET-enhanced ITS, which utilizes both vehicular ad hoc networks (VANETs) and cellular systems of the public transportation system to enable real-time communications among vehicles, road-side units (RSUs), and a vehicle-traffic server in an efficient way. Then, we propose a real-time path planning algorithm, which not only improves the overall spatial utilization of a road network but also reduces average vehicle travel cost for avoiding vehicles from getting stuck in congestion. Stochastic Lyapunov optimization technique is exploited to address the globally optimal path planning problem. Finally, the transmission delay of the hybrid VANET-enhanced ITS is evaluated in VISSIM to show the timeliness of the proposed communication framework. Besides, system-level simulations conducted in Java demonstrate that the proposed path planning algorithm outperforms the traditional distributed path planning in terms of balancing the spatial utilization and drivers’ travel cost.

I. INTRODUCTION

Traffic congestion, as an important societal problem, has received considerable attention. The 2007 Urban Mobility Report [1] stated that traffic congestion causes nearly 4.2 billion hours of extra travel every year in US; the extra travel almost accounts for 2.9 billion extra gallons of gasoline. Although many existing advanced Personal Navigation Devices (PNDs) have functionalities of providing an optimal end-to-end path [2] [3], traffic congestion problems in intelligent transportation systems (ITS) have not been fully resolved; on the contrary, conventional approaches still face a number of technical challenges. For example, Google Maps involve existing networks (e.g., Global Position System, Wi-Fi, cellular networks, etc) for individual path planning to avoid the traffic congestion. However, the provided services are very costly, and more importantly, they cannot make quick response to an emergency caused by an accident/incident. The essential reason for this imperfection lies in lack of real-time traffic information. Thus, to enhance the adaptability of path planning, it is indispensable to study how to efficiently collect and further exploit the real-time traffic information for path planning and traffic congestion avoidance.

First, to collect the real-time traffic information, the emerging vehicular ad hoc networks (VANETs) can provide an ITS system with enhanced communication capabilities for cost-effective and real-time traffic-information delivery [4]. Both vehicle-to-vehicle (V2V) and vehicle-to-road-side-unit (V2R) communications [6] are supported in VANETs to efficiently collect/report traffic updates from/to vehicles as well as road side units (RSUs) [7]. As a result, the collected real-time traffic information can be utilized for freeway-traffic-flow managements [8], individualized vehicle path planning [9], and vehicle localization [10]. However, most of the related works assume that the incorporated VANETs have sufficiently small delivery delay for real-time information collection. Actually, as VANETs rely on short-range multi-hop communications, the end-to-end transmission delay can be non-neglectable in some scenarios. Therefore, evaluations should be conducted to study how the end-to-end transmission performance of vehicular communications impacts on the performance of path planning in different scenarios and how to design the transmission mechanisms to reduce the delay when delay is not neglectable.

Second, to exploit the obtained real-time traffic information, many algorithms are designed to discover optimal paths for individual vehicles [11] [12]. But individual path planning may lead to new congestion if performed uncoordinatedly. To smooth the overall network flow, many works plan optimal paths from a global perspective for a group of vehicles simultaneously [13] [14]. However, most existing globally optimal path planning algorithms focus on the network-side performance improvement and neglect the drivers’ preferences (e.g., shorter travel length or time). Since the replanning decisions are made to avoid congestion and balance the traffic rather than discover optimal paths for individuals, some vehicles may

---

1On February 3rd, 2014, the U.S. Department of Transportation’s (DOT) National Highway Traffic Safety Administration (NHTSA) announced that it will begin taking steps to enable V2V for vehicles to talk to each other and ultimately avoid crashes altogether by exchanging basic safety data [5].
pay additional cost (e.g., a longer traveling length). Therefore, algorithms should be designed to jointly consider the balance of the network traffic and the reduction of average vehicle travel cost.

To this end, we propose a real-time global path planning algorithm which exploits VANET communication capabilities to avoid vehicles from congestion in an urban environment. Both the network spatial utilization and vehicle travel cost are considered to optimally balance the overall network smoothness and the drivers’ preferences. Specifically, the contributions of this paper are threefold:

• First, to facilitate the application of real-time path planning, we propose a hybrid-VANET-enhanced ITS framework, exploiting both the VANETs and the public transportation system. Based on the proposed hybrid ITS framework, a multi-hop message forwarding mechanism is designed to collect the real-time traffic information or the emergent warning messages, which usually have strict delay bounds. A theoretical analysis on the end-to-end transmission delay performance of the mechanism is presented as well;
• Second, we design a real-time global path planning algorithm to not only improve network spatial utilization but also reduce average vehicle travel cost per trip. A low-complexity algorithm is developed based on Lyapunov optimization to make real-time path planning decisions. With the proposed path planning algorithm, the tradeoff between the overall network spatial utilization and drivers’ preferences can be well balanced; and
• Finally, the transmission performance of the hybrid-VANETs is first evaluated under different vehicle densities via VISSIM, and then extensive simulations validate the effectiveness and efficiency of the proposed path planning algorithm. The results confirm that our proposed path planning algorithm is able to find alternative paths for vehicles to bypass congestion areas while reducing the average travel cost in an efficient, timely and coordinated way.

The remainder of this paper is organized as follows. Section II provides related works on path planning. The system model is discussed in Section III. Section IV presents the transmission mechanism in the proposed architecture and the corresponding performance analysis. A real-time path planning problem is formulated in Section V, followed by algorithm design in Section VI. Section VII demonstrates the performance of our proposed path planning algorithm by simulations. Finally, Section VIII concludes this paper.

II. RELATED WORKS

Traffic congestion, caused by unbalanced traffic flow or a sudden accident/incident, can cause late arrivals and additional cost for drivers and becomes a major problem in the transportation. However, this cost due to traffic congestion can be reduced by route navigation or path planning with congestion avoidance. For example, the paths of vehicles can be replanned with the shortest-path-based global positioning system (GPS) navigation [15], the accident duration prediction [16] and the route reservation in advance [17]. However, these approaches cannot make quick response to an emergency or congestion due to a sudden accident, since a timely update on the traffic-condition is lacking. Thus, the real-time traffic information becomes indispensable to support the vehicular real-time path planning algorithm.

To collect time-varying traffic-condition information, most existing works in conventional ITS usually rely on cellular systems or loop detectors. In [18]–[21], cellphones or mobile sensors with cellular access have been investigated to collect real-time traffic information for traffic forecast or reconstruction in experimental research. In [8], authors introduce a traffic management system with loop detectors for continuous traffic measurement and monitoring along arterials. However, inevitable drawbacks cast a shadow on the application of cellular systems and loop detectors. For cellular systems, as they are not dedicated for traffic data collection, the collection services can be highly costly; the high volume of traffic data may also cause congestion for other cellular services. For the loop detectors, the deployment expense can also be very high. Besides, the inaccuracy of position measurement becomes a problem for short-distance transmissions especially in dense networks, which will degrade the performance of path planning [22], [23].

Thanks to VANETs, V2V and V2R communications can make real-time message delivery much quicker, cheaper and more efficient than the existing systems even for short-distance transmissions in dense networks [24] [25]. More importantly, RSUs in VANETs can greatly enhance the timeliness of data collection and dissemination [26], which makes it possible to perform coordinated path planning for a group of vehicles. To improve the quality of experience (QoE), a point-to-point based vehicular network can be utilized to support the application of multimedia delivery [27] [28], which however may still experience large transmission delay. Hence, in the paper, to reduce the end-to-end transmission delay, taxis or buses are considered as super relays to help to deliver the information through the cellular network of public transportation system. On the other hand, in [27] [28], authors studied the media service applications, introducing heavy load to the involved cellular networks; however, in our work, the delivered information composes limited small-size packets, leading to a different transmission scenario with smaller data traffic load.

Many works have studied real-time vehicle path planning with the assist of VANETs. A distributed path planning method to avoid congestion is put forward in [11] using real-time traffic data collected from VANETs, with the increased traffic flow. Aiming to save gasoline for individual vehicle, a navigation system is designed in [12] to avoid congestion. However, the individual-user-optimal schemes may introduce additional traffic congestion due to human uncoordinated selfish behaviors. Thus, the paths of different vehicles should be jointly planned to balance the network traffic. The works [13] and [14] consider multi-vehicle path planning, but the average travel cost or the drivers’ preference is not considered.
Besides, how communications in VANETs can impact on the path planning algorithm is still not clear.

Therefore, in this paper, a globally optimal path planning algorithm is proposed for vehicles to avoid traffic congestion (including those caused by accidents) in a suburban scenario. With the real-time traffic information collection and decision delivery enabled by a hybrid VANET-enhanced network, the road network resources are fully utilized and the average travel cost of vehicles is significantly reduced. In addition, the impacts of VANETs on the path planning algorithm is further discussed.

### III. System Model

Aiming at providing real-time planned paths for vehicles from a global perspective, we first introduce the following network architecture. The traffic flow model is then elaborated, followed by the vehicle categorization and mobility model. A summary of the important mathematical notations used in the paper is given in Table I.

#### Table I: A summary of the important mathematical notations.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{V}$</td>
<td>The set of vehicles in the network</td>
</tr>
<tr>
<td>$\mathbb{I}$</td>
<td>The set of all intersections</td>
</tr>
<tr>
<td>$\mathbb{V}$</td>
<td>The set of all the roads in the network</td>
</tr>
<tr>
<td>$\mathbb{R}$</td>
<td>The set of RSUs in the network</td>
</tr>
<tr>
<td>$\lambda_{ij}(t)$</td>
<td>The inflow rate of road segment $(i, j)$ in time slot $t$</td>
</tr>
<tr>
<td>$\mu_{ij}(t)$</td>
<td>The outflow rate of road segment $(i, j)$ in time slot $t$</td>
</tr>
<tr>
<td>$c_{ij}(t)$</td>
<td>The maximum number of outflow vehicles of road segment $(i, j)$ in time slot $t$</td>
</tr>
<tr>
<td>$\delta(I_{ij})$</td>
<td>A congestion indicator of a warning message of congestion $I$ on road segment $(i, j)$</td>
</tr>
<tr>
<td>$Q^*_T(m)$</td>
<td>A virtual queue of intersection $i$ to represent the buffered vehicles destined to destination $d$</td>
</tr>
<tr>
<td>$\nu_\text{on}$</td>
<td>The capability of flexible turning for the vehicle $m$</td>
</tr>
<tr>
<td>$T_L$</td>
<td>A global message lifetime</td>
</tr>
<tr>
<td>$T_{\text{on}}$</td>
<td>“on” period of a vehicle</td>
</tr>
<tr>
<td>$T_{\text{off}}$</td>
<td>“off” period of a vehicle</td>
</tr>
<tr>
<td>$U_{\text{on}}$</td>
<td>The travel distances within an “on” period</td>
</tr>
<tr>
<td>$U_{\text{off}}$</td>
<td>The travel distances within an “off” period</td>
</tr>
<tr>
<td>$R$</td>
<td>The transmission range of a vehicle or an RSU</td>
</tr>
<tr>
<td>$M$</td>
<td>The average number of hops of an end-to-end transmission in pure VANETs</td>
</tr>
<tr>
<td>$M'$</td>
<td>The average number of hops of an end-to-end transmission in hybrid VANET-enhanced networks</td>
</tr>
<tr>
<td>$\psi$</td>
<td>The average transmission delay of a multi-hop transmission path in pure VANETs</td>
</tr>
<tr>
<td>$\psi'$</td>
<td>The average transmission delay of a multi-hop transmission path in hybrid ITS</td>
</tr>
<tr>
<td>$L_{r_{ij}}$</td>
<td>The changed path for vehicle $m$ at intersection $i$, routing from intersection $i$ to $j$</td>
</tr>
<tr>
<td>$L_{o_{ij}}$</td>
<td>The original path for vehicle $m$ at intersection $i$</td>
</tr>
<tr>
<td>$c_{r_{ij}}$</td>
<td>The cost of vehicle $m$ for a turning decision $r_{ij}$ towards destination $d$</td>
</tr>
<tr>
<td>$p_{r_{ij}}(T)$</td>
<td>The average cost factor for vehicles at intersection $i$</td>
</tr>
</tbody>
</table>

#### A. Hybrid-VANET-Enhanced Transportation System

Fig. 1 shows the architecture of the considered hybrid-VANET-enhanced transportation system, consisting of vehicles, RSUs, cellular base stations (BSs) and a vehicle-traffic server.

Vehicles are equipped with the on-board units (OBUs) which enable multi-hop V2V communication used in delivering the periodic vehicle information (e.g., vehicle velocity, density and location). When vehicles sense accident-related congestion, the warning message can be generated to alert the emergent accident information, and then be shared not only among vehicles but also with the nearest RSU via V2R communications. Besides pure VANETs, cellular communications, e.g., a global system for mobile communications (GSM) system which is set up for the functions such as mobile telemonitoring and management systems for inter-city public transportation [29], are also involved. Hence, the taxis or buses can directly upload the received warning message to the nearest cellular BS and the BS will deliver the message to the vehicle-traffic server.

RSUs deployed along the roads are assumed to be able to obtain vehicle-traffic statistical information (e.g., the vehicle arrival/departure rate on each road). We consider that taxis and buses are perfectly connected to the cellular system, and RSUs are well connected with each other through wireline. If RSUs are deployed at intersections, the traffic information can be detected by the equipped cameras or traffic flowmeters connected to RSUs directly [30]. Otherwise, the traffic flow can be predicted by the nearest RSUs based on the obtained vehicle information (e.g., periodically obtained vehicle density and velocity) from the VANETs [31]. An RSU can share its own collected information with other RSUs and the vehicle-traffic server. When an accident happens, based on all the collected information, the vehicle-traffic server is capable of performing real-time path planning to provide globally optimized travel paths for vehicles of interest.

We further define a road network into four main components (i.e., intersections, roads, vehicles, and RSUs) as $\zeta = (\mathbb{I}, \mathbb{V}, \mathbb{V}, \mathbb{R})$. The set of all intersections is denoted as $\mathbb{I}$. Let $\mathbb{V}$ be the set of all the roads in the network. Each road between two adjacent intersections is assumed bi-directional, possibly with multiple lanes in one direction. We refer to each of those lanes with the same direction in a road as a road segment, i.e., one normal bi-direction road between two adjacent intersections $i$ and $j$ has two different road segments with opposite directions, i.e., road segment $(i, j)$ and road segment $(j, i)$. The set of vehicles and that of RSUs are defined as $\mathbb{V}$ and $\mathbb{R}$, respectively.

#### B. Traffic Flow Model

To understand a vehicle-traffic flow more clearly, we model vehicle traffic as an “inflow/outflow” system [32]. Each vehicle is expected to follow a planned path from its starting point towards its destination. Here, the planned path can be referred to as a path preset in a GPS, according to the driver’s preferences and based on the locations of the starting and
ending points. The driver will keep following the preset path until the vehicle receives any information on congestion or accident. When an accident or congestion occurs, by running the path planning algorithm, the vehicle-traffic server will be in charge of finding an optimal alternative path or routing for the vehicles of interest. Specifically, in this paper, we refer to the road segments in which one vehicle’s starting point and destination are located as $s (\in \Gamma)$ and $d (\in \Gamma)$, respectively.

Let $J_i$ denote the set of neighboring crossings of intersection $i$. Define the inflow rate of road segment ($i, j$), $\lambda_{ij}(t)$, as the upstream-vehicle arrival rate from neighboring road segments in time slot $t$, where $j \in J_i$, as shown in Fig. 2. Let $\lambda^d_{ij}(t)$ ($j \in J_i$) denote the traffic flow rate on road segment ($i, j$) with the same destination $d$ in time slot $t$, and $\lambda_{ij}(t) = \sum_{d \in \Gamma} \lambda^d_{ij}(t)$.

In this paper, we consider each sample time duration (denoted as $\Delta$ and including a series of time slots) as a time unit, which is defined by sampling theorem to avoid information loss in the compressive sensing for traffic estimation in [33]. Within the $T^{th}$ sample time duration, based on the traffic flow rates of the involved time slots collected by RSUs, the average inflow rate of road segment ($i, j$) of the $T^{th}$ sample time duration is denoted as $\lambda_{ij}(T)$ and expressed as

$$\lambda_{ij}(T) = \frac{1}{\Delta} \sum_{t=(T-1)\Delta}^{T\Delta} \lambda_{ij}(t). \tag{1}$$

Similarly, the outflow rate $\mu_{ij}(T)$ of road segment ($i,j$) is the average departure rate of vehicles moving to neighboring road segments in the $T^{th}$ sample time. Note that all variables for the opposite directed road segment of ($i,j$), namely road segment ($j,i$), can be defined correspondingly, e.g., $\lambda_{ji}(T)$ and $\mu_{ji}(T)$.

Let $c_{ij}(T)$ denote the maximum number of outflow vehicles of road segment ($i,j$) in $T^{th}$ sample time, i.e., road capacity, which is determined by the road conditions, the number of lanes, the length of the road, and traffic congestion, etc. Due to fluctuating road conditions and traffic flow conditions, the road capacity can fluctuate with time, but is considered to remain constant within one sample time unit.

Basically, there are two kinds of traffic congestion: recurrent congestion and non-recurrent congestion [34]. The recurrent congestion is due to the tension between the current traffic flow situation (e.g., the traffic inflow $\lambda_{ij}(T)$) and the road conditions (e.g., the road capacity $c_{ij}(T)$), which is non-incident related. The non-recurrent congestion is caused by an accident or incident which can reduce the road capacity (to be introduced in Section V). We define a congestion indicator of a warning message, $\delta(I_{ij})(\in [0,1])$, to represent how the congestion type $I$ happening on road segment ($i,j$) impacts on the road capacity, where $\delta(I_{ij}) = 1$ means recurrent congestion and $\delta(I_{ij}) \in [0,1)$ implies non-recurrent congestion.

Each vehicle traveling from one intersection to the next is called routing in this paper. For each intersection (say intersection $i$), consider that the RSU nearest to the intersection maintains a virtual queue of length $Q^d_{ij}(T)$, representing the number of the buffered vehicles at this intersection specifically destined to destination $d (\in \Gamma)$ in sample time $T$. Then,
the total length of all virtual queues of intersection \( i \) for all destinations is 
\[ Q_i(T) = \sum_{d \in \Gamma} Q_i^d(T), \]
where
\[ Q_i^d(T) = \max\{Q_i^d(T-1) - \sum_{j \in j_i} \mu_{ij}^d(T-1), 0\} \]
\[ + \sum_{u \in J_i} \lambda_{ui}(T-1). \]
(2)

with \( \mu_{ij}^d(T-1) \) being the outflow rate of road segment \((i, j)\) with destination \( d \) in the \((T-1)\)th sample time, satisfying \( \mu_{ij}(T-1) = \sum_{d \in \Gamma} \mu_{ij}^d(T-1) \). Similarly, for road segment \((i, j)\), we define the leftover number of vehicles in sample time \( T \) as 
\[ Q_{ij}(T) = \max\{Q_{ij}(T-1) - \mu_{ij}(T-1), 0\} + \lambda_{ij}(T-1). \]

C. Vehicle Categorization and Mobility Model

Three types of vehicles are considered in this work, namely private cars, taxis, and buses. GPS devices are supposed to be deployed on all vehicles, and GPS devices have ordered the service of providing shortest paths. Compared with changeable paths of taxis or private cars, scheduled paths of buses are usually fixed. Let \( w_m \in \{0, 1\} (m \in \mathcal{V}) \) denote the capability of flexible turning for the vehicle \( m \) when the vehicle receives any information about congestion or accident, and take the value 1 if vehicle \( m \) is a taxi or a private car and 0 otherwise, since taxis or private cars can change their paths while buses have to wait until the traffic trap is cleaned up.

Furthermore, we refer to taxis and buses as super-nodes, connected to a control center through GSM systems. With a specially designed message transmission mechanism (to be introduced in Section IV), warning messages can be delivered to the vehicle-traffic server as efficiently as possible to facilitate real-time path planning.

The mobility of each vehicle can be characterized by two random variables \((V, D)\) [35]. Here, \( V \) represents the vehicle velocity which takes two possible values (i.e., a lower velocity \( v_L \) and a higher velocity \( v_H \)). The velocity transition is modeled as a two-state continuous time Markov chain (TCMC) with state transition rate \( \frac{1}{\tau} \). Under this model, a vehicle initially chooses \( v_L \) (or \( v_H \)), and after an exponentially distributed time interval with the mean of \( D \), the velocity changes to \( v_H \) (or \( v_L \)). The model can be exploited to describe the realistic driving behaviors, i.e., a driver usually drives at a constant velocity for a period and then changes to a higher/lower velocity based on his/her will and/or road conditions. Besides, when the vehicle density is low or medium (e.g., no larger than 30 vehicle/km/lane), vehicles can be considered to move independently [36] and the headway distance\(^2\) follows the exponential distribution with rate \( \zeta \) [37].

IV. Transmission Mechanism and Performance Analysis

Since the incident-related warning message is pivotal to the viability of a real-time path planning algorithm, we propose the following rapid message transmission mechanism and give corresponding analytical results on the end-to-end transmission performance.

\( ^2\)In this paper, the headway distance is defined as the distance between two neighboring vehicles in the same lane.

A. Outline of Transmission Mechanism

After sensing the congestion, vehicles in the vicinity of the congestion will generate and forward the warning message to other vehicles via multi-hop V2V relaying. If a supernode receives a warning message, it will upload the message to the nearest cell BS through cellular communication of the public transportation system; otherwise, the message will be transmitted all the way to one RSU via V2V and V2R transmissions. To reduce the redundancy of multi-hop relaying, the following relay node selection is adopted. If there is only one bus/taxi with the transmission range of a vehicle, the bus/taxi will be the next-hop receiver; otherwise, the farthest vehicle ahead in the same lane within the transmission range will be selected as the next relay [35]. Besides, we assume that a vehicle deletes the warning message once it has been transmitted. On the other hand, a global message lifetime \( T_L \) is preset for each warning message, at the end of which all the transmissions of the corresponding message will be terminated, thus to further control the redundancy in message delivery. Once an RSU or cellular BS receives a warning message, it forwards the message to the vehicle-traffic server via wireline. Upon receiving the warning message, the traffic server will operate the path planning algorithm based on the collected timely road-traffic information. By leveraging this transmission mechanism, emergent messages (e.g., congestion indicators) are promising to be disseminated more efficiently as compared with only utilizing VANETs or the cellular communication capabilities of the public transportation system. Finally, after the vehicle-traffic server finishing path planning, replanned paths are fed back to vehicles demanding path planning via a downlink transmission (i.e., the traffic server - an RSU/vehicle relay - the vehicle in need of path planning).

As shown in Fig. 1, the overall communications in the proposed VANET-enhanced ITS can be divided into three layers: V2V and V2R communications in VANETs, wireless communication between super-nodes and BSs via a cellular system, and wired communication between RSUs (or BSs) and the vehicle-traffic server. Thus, the main issues affecting the efficiency of the end-to-end message transmission comes to transmission delay in VANETs. By considering the following ideal medium access control (MAC) for V2V and V2R communications, we will analyze the transmission delay in VANETs in the next subsection. Specifically, for analytical simplicity, we assume that once a vehicle moves into the coverage range of an RSU or another vehicle, time slots can be scheduled with negligible delay for the corresponding V2R or V2V transmissions. Besides, the link rate of a V2V or V2R transmission is assumed to be constant, and the contact duration between each transmission pair is considered long enough to accomplish at least one packet delivery, which can be achieved by appropriately setting the packet size [38].

In general, the transmission delay in VANETs can be discussed under two cases: 1) when the vehicle density is very high (e.g., more than 56 vehicles/mile), the connections among vehicles can be found with high probability, considering that
the transmission range of a vehicle (e.g., more than 100m as shown in DSRC) is way more than the average headway distance. In this case, for a given connection path for example from a vehicle to an RSU, we consider neglectable transmission delay because of the assumption of the ideal MAC and small-size packet delivery; 2) for the medium or sparse vehicle density case, due to the intermittency of vehicle communications caused by high-speed mobility and/or node sparsity, the inter-contact time, namely, the waiting time of each hop for the receiver (vehicle or RSU) to fall into the transmission range of the transmitter dominates the end-to-end transmission delay. Notice that congestion may cause an unbalanced vehicle distribution on neighboring roads, and the traffic information report on a road of low node density can be the bottleneck of the VANET-assisted information collection. As such, in the following we analyze the impact of vehicle density on the inter-contact time of one-hop V2V or V2R transmission and further on the end-to-end transmission delay along the transmission path.

B. End-to-End Delay Analysis

In the following, we analyze the inter-contact time for the aforementioned transmission mechanism. The end-to-end delay analysis begins from the transmissions in pure VANETs, and then involves the public transportation system.

1) End-to-End Delay in Pure VANETs: First, consider an uplink with no taxis or buses, i.e., all the hops are based on V2V and V2R communications. We evaluate the transmission delay for the last hop of the V2R transmission. The transmission delay here is mainly due to the inter-contact time between a vehicle and an RSU. Similar to [35], we define the last hop as an “on-off” model, where a vehicle either directly connects to an RSU (i.e., during the “on” state) or is the only vehicle approaching the RSU and there is no other vehicle in the transmission range of the RSU (i.e., during the “off” state). According to the transmission model, the transmission delay of a packet during the “on” state should be way smaller than that during the “off” state. Therefore, the transmission delay of the last V2R hop is mainly due to the “off” period.

Denote the “on” period and the “off” period of a vehicle as $T_{on}$ and $T_{off}$, respectively. Accordingly, the travel distances within the two periods are defined as $U_{on}$ and $U_{off}$, respectively, with $T_{on} = \frac{U_{on}}{V}$ and $T_{off} = \frac{U_{off}}{V}$, where $V$ is the average velocity for a vehicle based on the “on-off” mobility model (see Section III-C). Similar to [35], the event that a vehicle moves a distance of at least $u$ during $T_{on}$ before being scheduled to communicate with an RSU should satisfy: 1) there is no other vehicle within the distance $u$ from the end of the RSU coverage ahead of the vehicle, and 2) there is at least one vehicle within the distance $2R - u$, which results in this vehicle moving at least $u$ distance to avoid the collision, with $R$ representing the transmission range of an RSU or a vehicle. Then, we have

$$P_r(U_{on} > u) = \frac{(e^{-\gamma} u)^{\gamma-1}[1-(e^{-\gamma} (2R+u))^\gamma]^{\gamma-1}}{1 -(e^{-\gamma} 2R)^\gamma} \tag{3}$$

where $b$ is the summation of all road lengths, and $\gamma$ is the average vehicle density on the roads. Since the vehicle headway distance follows an exponential distribution as mentioned in Section III-C, the probability that a headway distance is larger than $u$ is $e^{-\gamma u}$. Based on (3), we can obtain

$$E[U_{on}] = \int_0^R P_r(U_{on} > u) du. \tag{4}$$

Similarly, the event that a vehicle moves a distance of at least $u$ during $T_{off}$ should satisfy: 1) there is no vehicle within a distance of $2R + u$ from the end of the coverage range of the nearest RSU ahead of the vehicle, and 2) there is at least one vehicle within the distance $L - (u + 2R)$, where $L$ is the distance between the adjacent RSUs. Then, we have

$$P_r(U_{off} > u) = \frac{(e^{-\gamma} (2R+u))^\gamma-1[1-(e^{-\gamma} (L-(2R+u))^\gamma]^{\gamma-1}}{1 -(e^{-\gamma} 2R)^\gamma} \tag{5}$$

$$E[U_{off}] = \int_0^{L-2R} P_r(U_{off} > u) du. \tag{6}$$

In the previous hops between vehicles within a transmission path except the last hop can be characterized with the vehicle mobility model. The process of the relative velocity between two vehicles can be represented by a CTMC with a state space $\mathbb{H} = \{h_0, h_1, h_2\}$. Here, $h_0$ represents a negative relative velocity when the vehicle in front moves with $v_f$, while the vehicle behind moves with $v_H$; $h_1$ models a zero relative velocity (i.e., both vehicles move with the same velocity); $h_2$ represents a positive relative velocity. If each vehicle keeps the same velocity for an exponential time with an average of $D$, the transition rate between any two states of the Markov process is equal to $2/D$. Thus, from [35], the average number of hops $M$ of an end-to-end transmission path from a message source to an RSU in pure VANETs can be approximated as

$$M = \frac{4(\log(E[U_{on}]) - E[U_{off}])}{D(v_f+v_H)} \tag{7}$$

Then, based on the average number of hops, the transmission delay of such a transmission path can be shown as

$$\psi = (M - 1)E[T_{V2V}] + E[T_{off}] \tag{8}$$

where $E[T_{V2V}] = \frac{1}{1-e^{-\gamma V}}$ is the average transmission delay for a V2V hop, since the headway distance follows an exponential distribution. $E[T_{off}]$ is the average duration of the ”off” period as defined before. If we consider the downloading as a similar process with uploading, the total transmission delay can be approximated by $2\psi^3$. Note that this transmission delay is related to the parameters including vehicle mobility parameters ($V$ and $D$), vehicle density ($\gamma$), and RSU related parameters (the transmission range $R$ and the average distance between RSUs $L$). Then, the probability of a $M$-hop transmission path with all V2V and V2R communications equals the probability that there is neither taxi nor bus in any hop within the $M$-hop transmission path, i.e., $(1 - P_T - P_B)^M$, where $P_T$ ($P_B$) is the percentage of taxis (buses) in the traffic stream.

$^3$The approximation is valid if the end-to-end transmission delay can be well controlled to a small value in which the network topology changes little or the source vehicle only moves a relatively small distance.
2) End-to-End Delay in Hybrid-VANET-enhanced Network:
If the public transportation system is involved in delivering messages as aforementioned, the probability of a given number of hops from a private car to the nearest bus/taxi follows a geometric distribution. The average number of hops in the hybrid VANET-enhanced ITS, \( M' \), is as
\[
M' = M \cdot (1 - P_B - P_T)^M + \sum_{i=1}^{M} (i - 1) \cdot (1 - P_B - P_T)^{M-i} \cdot (P_B + P_T).
\]
(9)

Then, if we consider that the public transportation system are perfectly connected with no delay, the average transmission delay is dominated by the transmission delay in VANETs. Based on the probability of a given number of hops from a private car to the nearest bus/taxi, the transmission delay in a multihop message transmission path is rewritten as
\[
\psi' = \psi \cdot (1 - P_B - P_T)^M + \sum_{i=1}^{M} (i - 1) \cdot (1 - P_B - P_T)^{M-i} \cdot (P_B + P_T) \cdot E[T_{V2V}].
\]
(10)

From (10), the end-to-end transmission delay in hybrid ITS is related to 1) vehicle mobility parameters (i.e., \( V \) and \( D \)), 2) vehicle density and super-node percentage (i.e., \( \gamma \), \( P_B \), and \( P_T \)), and 3) RSU deployment in the network (e.g., the transmission range \( R \) and the average distance between RSUs \( L \)).

V. Problem Formulation
In this section, based on the traffic flow model defined in Section III-B, the traffic flow balance constraint of each intersection is first identified. The road capacity and congestion indicator are then discussed under different traffic conditions. Subsequently, considering the drivers’ travel-cost preferences in the path planning, the cost metric of path planning for individual vehicle is defined. In addition, the network stability constraint is shown. Finally, the real-time path planning problem is formulated to not only avoid the congestion but also reduce the average travel cost caused by path planning.

A. Intersection Flow Balance Constraint
For an intersection \( i \) (\( i \in \mathbb{I} \)), the following flow balance equation should be satisfied to guarantee that the aggregate vehicle arrival rate is equal to the aggregate vehicle departure rate
\[
\sum_{j \in J_i} \mu_{ji}(T) = \sum_{u \in A_i} \lambda_{iu}(T), \quad \forall i \in \mathbb{I},
\]
(11)
where the left and right side of the equation are respectively referred to as the aggregate vehicle arrival and departure rates.

B. Road Capacity and Congestion Indicator
For road segment \((i, j)\), the vehicle inflow rate for sample time \( T \) is \( \lambda_{i,j}(T) \). The average outflow rate changes with the inflow rate, but with some time delay (denoted as \( \Lambda \) seconds which is the travel time for a vehicle moving from intersection \( i \) to intersection \( j \)), i.e., \( \mu_{ij}(T) = \lambda_{ij}(T - \Lambda) \), until reaching the outflow rate limit, i.e., road capacity \( c_{ij}(T) \). Here, \( \Lambda \) is decided by the tension between the traffic inflow and road capacity. Once an incident/accident occurs, the outflow rate drops dramatically on one road segment. To illustrate the road capacity under different traffic conditions, we discuss the road capacity in two cases: 1) no incident-related congestion (i.e., recurrent congestion), and 2) the incident-related congestion (i.e., non-recurrent congestion). The road capacities under two cases will be illustrated respectively as follows.

1) When there is no incident-related congestion on \((i, j)\), according to [34], we have
\[
c_{ij}(T) = c_{ij}^N = N_{ij} \cdot c_{ij}^P \cdot F_{PH} \cdot \frac{1}{(1 + E_B \cdot P_B) \cdot A}
\]
(12)
where \( c_{ij}^N \) is the road capacity under no incident-related congestion case. \( N_{ij} \) is denoted as the number of lanes in road segment \((i, j)\). The ideal capacity per lane is \( c_{ij}^P \). \( F_{PH} \) is the peak-hour factor, i.e., the ratio of the peak 15-min flow rate in vehicles per hour (vph) to the average hourly flow rate (vph). \( E_B \) is the bus equivalent to private cars or taxis. \( P_B \) is the percentage of buses in the traffic stream. \( A \) is an adjustment factor to account for other factors with impact on road capacity. Under this case,
\[
\mu_{ij}(T) = \min\{\lambda_{ij}(T - \Lambda'), c_{ij}(T)\}
\]
(13)
with \( \Lambda' \) called recurrent delay [34] and satisfying
\[
\Lambda' = T_{ij}^0 + D_{ij}^2 + 0.25T[(\frac{ij}{ij} - 1)^2 + \frac{16T_{ij}^2 \cdot \lambda_{ij}(T)}{N_{ij}^2 \cdot T_{ij}^4 \cdot \lambda_{ij}(T)}].
\]
(14)
Here, \( T_{ij}^0 = L_{ij}/V_0 \) is the segment travel time measured at free flow speed \( V_0 \), with \( L_{ij} \) being the length of road segment \((i, j)\). \( J_{ij} = \frac{T_{ij}^0 - T_{ij}^0}{L_{ij}} \) is a calibration parameter, with \( T_{ij}^0 \) being the segment travel time measured when the traffic demand equals road capacity. \( D_{ij}^2 \) is the delay due to leftover queue from the prior sample time, i.e.,
\[
D_{ij}^2 = \frac{Q_{ij}(T)}{2 \cdot c_{ij}(T) \cdot T} \cdot \min\{T, \frac{Q_{ij}(T)}{c_{ij}(T) \cdot [1 - \min(1, \frac{\lambda_{ij}(T)}{c_{ij}(T)})]}\}.
\]

2) When there is an incident \( I_{ij} \) on road segment \((i, j)\), we still hold
\[
\mu_{ij}(T) = \min\{\lambda_{ij}(T - \Lambda'^r), c_{ij}(T)\}
\]
(15)
where \( \Lambda'^r \) is called non-recurrent delay and can also be calculated based on (14). However, in this case,
\[
c_{ij}(T) = c_{ij}^I = c_{ij}^N \cdot \delta(I_{ij}), \forall \delta(I_{ij}) \in [0, 1]
\]
(16)
where \( \delta(I_{ij}) \) is the percentage of remaining road capacity during incident type \( I \) on road segment \((i, j)\), i.e., congestion indicator. The value of \( \delta(I_{ij}) \) depends on the incident type \( I \) and is considered to be sensed by witness/victim vehicles and delivered to the nearest RSU or BS. \( c_{ij}^I \) is thus the road capacity under the incident \( I \). Take the case that a road segment has one lane in each direction as an example. When

The bus equivalent is the number of buses displaced by a single taxi or private car in a suburban area [39].
an accident $I$ happens, we may consider that $\delta(I_{ij}) = 0$ and $\mu_{ij}(T) = c_{ij}^T = 0$, since no vehicle-traffic flow will pass. On the other hand, in a case that a road segment has multiple lanes in each direction, the traffic flow will not be zero, but might still drop dramatically.

Furthermore, if there is no incident-related congestion on road $(i, j)$, $\delta(I_{ij}) = 1$. Then, we can extend the following relationship between the indicator and road capacity:

$$c_{ij}(T) = c_{ij}^N \cdot \delta(I_{ij}), \forall \delta(I_{ij}) \in [0, 1]$$

which implies that the road capacity drops once an accident happens on a certain segment until the accident is cleaned up. The outflow rate should be always no more than the according road capacity, i.e.,

$$\mu_{ij}(T) \leq c_{ij}(T).$$

C. Path Planning Cost Metric

The path planning algorithm is to avoid the congestion on the road, with considering the preference of drivers, e.g., the shortest path or the most familiar path. Here, we consider the path length as the driver’s first-order preference. Let $L_{r_{ij}}$ denote the changed path for vehicle $m$ (with destination $d$) at intersection $i$, where $r_{ij}$ is the initial path, according to the newly planned path, vehicle $m$ changes its path by going through road segment $(i, j)$ towards destination $d$, satisfying $j \in J_i$. Compared to current path length $L_{S_{ij}}$, the increased path length is $|L_{r_{ij}}| - |L_{S_{ij}}|$, where $S_i$ is the path choice before being replanned. Obviously, it is possible that the changed path leads to more travel time and more consumed fuel energy. Let $p_{r_{ij}}$ denote the cost of vehicle $m$ for a certain turning decision $r_{ij}$ towards destination $d$, given $S_i \neq r_{ij}$. If intersection $i$ is not in the current path of $m_d$, $p_{r_{ij}}$ is zero; otherwise, it is modeled with respect to the increased path length as follows

$$p_{r_{ij}} = \rho(|L_{r_{ij}}| - |L_{S_{ij}}|)$$

where $\rho(\cdot)$ is a non-negative increasing function to measure impacts of the increase of path length $|L_{r_{ij}}| - |L_{S_{ij}}|$ [40]. Then, the average cost of vehicles taking turning $r_{ij}$ on road segment $(i, j)$ can be calculated as

$$p_{ij}(T) = \left\{ \begin{array}{ll} \sum_{m \in V} w_m \sum_{d \in D} w_m \cdot p_{r_{ij}}, & \text{if } \sum_m w_m \neq 0; \\
\infty, & \text{otherwise.} \end{array} \right.$$  \hspace{1cm} (20)

For an intersection (say intersection $i$), since there may be several neighboring intersections as the candidates of the coming intersections, the average cost of vehicles belonging to intersection $i$ is defined as

$$p_{iJ_i}(T) = \left\{ \begin{array}{ll} \sum_{j \in J_i} \alpha_{ij}(T) \sum_{j \in J_i} \alpha_{ij}(T)p_{ij}(T), & \text{if } \sum_{j \in J_i} \alpha_{ij}(T) \neq 0; \\
0, & \text{otherwise.} \end{array} \right.$$  \hspace{1cm} (21)

where $\alpha_{ij}(T)$ is set as 1 in the first case of Eq. (20) (i.e., when $\sum_m w_m \neq 0$); otherwise 0.

D. Network Stability

The definition of Queue and Network Stability [41] is used to represent traffic congestion avoidance in our path planning optimization problem. For intersection $i$, $Q_i(T)$ is strongly stable if and only if

$$\lim_{T_0 \to \infty} \sup_{T_0} \frac{1}{T_0} \sum_{T_0} E[Q_i(T)] < \infty.$$  \hspace{1cm} (22)

The information on $Q_i(T)$ is required to identify whether an intersection is stable or not. If the traffic inflow and outflow information is detected by the cameras or traffic flowmeter connected to RSUs, $Q_i(T)$ is expected to be calculated directly. If the traffic information is relayed in VANETs as there is no RSU at the intersection, the relayed information is utilized in the vehicle-traffic server to predict the traffic flow information with a certain transmission delay. According to (10), this uploading transmission delay can be estimated as $\frac{T}{S}$, which here is mainly caused by the intermittent connections in VANETs. With this transmission delay, the proposed algorithm can utilize a more accurate virtual queue information for path planning in each sample time, i.e., $Q_i(T - \left\lfloor \frac{T}{S} \right\rfloor)$. Note that if and only if all queues in the network are strongly stable, vehicle traffic in the whole road network is strongly stable.

E. Utilization-Minus-Cost Maximization Problem

Taking account of both the traffic flows of the network and the path planning cost of vehicles, the objective of the path planning algorithm is considered to maximize the overall spatial-utilization-minus-planning-cost at the same time with the network congestion avoidance. This objective indicates that the total traffic flow improvement and the path planning cost reduction should be jointly considered and carefully balanced. Specifically, once the traffic server receives the traffic flow and accident warning messages collected from both RSUs and vehicles via VANETs (or cellular networks), a path planning algorithm is calculated to update and determine $\lambda_{ij}(T)$ according to the optimization problem, i.e., the number of vehicles dispatched over road segment $(i, j)$ in the $T^{th}$ sample time.

$$\max \sum_{i \in I} \sum_{j \in J_i} \lambda_{ij}(T) - \sum_{i \in I} \|p_{iJ_i}(T)\|_1$$

s.t. (11), (18), (22)  \hspace{1cm} (23)

This objective is to maximize the spatial utility while minimizing travel cost, under the constraints: 1) the flow balance of each intersection; 2) the limitation of outflow rate on each road segment; and 3) the congestion avoidance of each intersection. We exploit Lyapunov optimization process [41] to solve this problem (to be introduced in Section VI). Then, in the sample time $T$, based on the path planning algorithm, a vehicle with destination $d$ can be dispatched from one intersection
to another (say from intersection $i$ to intersection $j$ with contribution $\lambda_{ij}(T)$), in order to improve the spatial utility as well as to reduce travel cost. And this updated path will deliver to the GPS device to navigate the required vehicle. In other words, a turning decision, $r_{ij}$, for a taxi or a private car at intersection $i$, can be decided based on the corresponding $\lambda_{ij}(T)$ and $p_{ij}(T)$, and furthermore the replanned path can be calculated based on this turning decision. Note that if the traffic flow information is collected by VANETs (or cellular networks), the transmission delay in VANETs, i.e., $\frac{w}{x}$, should be considered in the third constraint as discussed in Section V-D.

VI. REAL-TIME OPTIMAL PATH PLANNING

In this section, the path planning algorithm is first proposed to help vehicles to bypass congestion and balance traffic evenly in the whole network. Then, the convergence and the computation complexity of the proposed algorithm are discussed.

A. Path Planning Algorithm Design

The optimization problem (23) can be solved by applying the drift-plus-penalty framework in Lyapunov optimization process [41]. By following dynamic algorithm at each sample time, we derive vehicles’ turning decisions for maximizing the lower bound of network throughput. According to the Lyapunov optimization process, let $W_{i,j}(T)$ denote the weight of intersection $i$ in sample time $T$,

$$W_{i,j}(T) = \sum_{j \in J} \alpha_{ij}(T) \min\{c_{ij}(T), \sum_{d \in D} \{Q_{ij}^d(T) - Q_{j}^d(T)\}\} - K p_{ij}(T)$$

(24)

where $K$ is a non-negative constant defined by vehicle traffic server used for all vehicles, with the same order of the reciprocal of travel cost (i.e., $p_{ij}(T)$) [41]. Equation (24) implies that the weight of an intersection (say intersection $i$) is realized to i) the differential queue backlog between intersection $i$ and its neighboring intersections; ii) average intersection travel cost. Vehicles at intersection with the largest weight are replanned first. Vehicles with destination $d$ stored at intersection $i$, should be dispatched to queue $Q_{ij}^d(T)$ of intersection $j$, where $j = \max\{Q_{ij}^d(T) - Q_{j}^d(T), c_{ij}^*(T)\}$, according to the largest differential queue backlog. The number of the vehicles with destination $d$ replanned to intersection $j$ is $\min\{Q_{ij}^d(T) - Q_{j}^d(T), c_{ij}^*(T)\}$. Then queues at all the remaining intersections are updated correspondingly. The same process continues until all intersections related are processed. The sketch of the proposed dynamic algorithm is summarized in Algorithm 1. The implication of path planning is to prioritize those vehicles in such an intersection with larger differential queue backlogs and shorter increased path lengths under new turning decisions (i.e., lower average travel cost).

B. Analysis of Algorithm Performance

For the network stability of the proposed path planning algorithm, we have the following Lemma 1.

**Lemma 1:** With the proposed path planning algorithm, network stability can be guaranteed.

**Proof:** To prove network stability, according to [41], we need to show that the summation of the average square of queue sizes of those intersections’ virtual queues does not increase with time. Consider the inter-flow exchange between any two intersections (say $i$ and $j$). Let $Q_i(T)$ ($Q_i(T + 1)$) and $Q_j(T)$ ($Q_j(T + 1)$) respectively denote the queue lengths of intersections $i$ and $j$ in sample time $T$ ($T + 1$). In specific, based on our path planning algorithm, between two neighboring intersections, vehicles are always dispatched from a long queue to a short queue. Assume that the change of the queue length of the two intersection is because $q_{ij}^d(T)$ vehicles, where $d \in \Gamma$, are dispatched from intersection $i$ to intersection $j$, i.e.,

$$Q_i(T + 1) = Q_i(T) - q_{ij}^d(T)$$

and

$$Q_j(T + 1) = Q_j(T) + q_{ij}^d(T).$$

Then, the consequence of $q_{ij}^d(T)$ dispatched vehicles is,

$$E\{(Q_i(T + 1))^2 + [Q_j(T + 1)]^2\} - E\{(Q_i(T))^2 + [Q_j(T)]^2\} = 2E\{(\sum_d q_{ij}^d(T) - Q_i(T) + Q_j(T)) \cdot \sum_d q_{ij}^d(T)\},$$

(25)

where $\sum_d q_{ij}^d(T)$ is the total number of vehicles, which are dispatched from intersection $i$ to intersection $j$ at time $T$. As we have $q_{ij}^d(T) = \min\{Q_{ij}^d(T) - Q_{j}^d(T), c_{ij}^*(T)\}$, $Q_i(T) = \sum_d Q_{ij}^d(T)$ and $Q_j(T) = \sum_d Q_{ij}^d(T)$, the following inequality holds

$$\sum_d q_{ij}^d(T) + Q_j(T) - Q_i(T) \leq 0.$$  

(26)

Thus, the right side of (25) is no more than zero. Then, the summation of average squares of queue size is satisfied as,

$$E\{(Q_i(T + 1))^2\} + E\{(Q_j(T + 1))^2\} \leq E\{(Q_i(T))^2\} + E\{(Q_j(T))^2\}.$$  

(27)
That is, the summation of average square of queue size of those intersections’ virtual queues does not increase with time. Under the cases with all destinations and multiple intersection-s, the similar results still hold, which implies the stability of network and the avoidance of traffic congestion in a network as discussed in [41].

Furthermore, the computational complexity of the proposed algorithm is given as following Lemma 2.

**Lemma 2:** The total computational complexity is proportional to the square of the number of intersections in the map times the upper bound of the number of neighboring intersections.

**Proof:** We first calculate the weight of each intersection, thus the complexity of this step is $O(|I|)$. Second, we schedule each intersection in $I_c$. For each intersection to be scheduled, we need to find the right neighboring intersection $d$ for each destination $d$. Therefore, the complexity of the second step is $O(|I_c|((1 + |I_c|)/2 + |Γ|U))$, where $U$ is the upper bound of the number of neighboring intersections of one intersection. As the $|I_c|$ and $|Γ|$ are in the same order, the overall complexity is given by

$$O(|I|) + O\left(\frac{|I| + |I|^2}{2} + |I||Γ|U\right).$$

(28)

Furthermore, as the number of roads $|Γ|$ and that of intersections $|I|$ have the relationship $2Γ/U \leq |I|$, the complexity can be further simplified as

$$O(|I|) + O\left(\frac{|I| + |I|^2}{2} + |I|^2U^2\right) = O(|I|^2U^2).$$

(29)

Thus, the total computational complexity is proportional to the square of the number of intersections in the map times the upper bound of the number of neighboring intersections.

The proposed path planning algorithm can perform better than the conventional path planning, because (1) the proposed path planning algorithm is updated based on real-time and accurate messages received from V2V/V2R communication, by which, for instance, a warning message of traffic jam can be delivered and impact timely on decisions of path planning; (2) furthermore, in hybrid VANET-enhanced networks, public transportation system can help to deliver the messages, leading to the reduced transmission delay for delay-sensitive real-time path planning; (3) the proposed path planning is designed to reduce traveling cost in a coordinated manner to avoid particular parts of the road network overloaded; and (4) the relatively low computational complexity of the proposed algorithm makes the path planning algorithm achieve better performance in a reasonable and realistic way.

**VII. PERFORMANCE EVALUATION**

In this section, we consider a realistic suburban scenario as shown in Fig. 3, which is the region around the campus of University of Waterloo (Waterloo, ON, Canada). To emulate the timeliness of the proposed communication framework, a highly realistic microscopic vehicle traffic simulator, VISSIM [44], is employed to generate vehicle trace files for recording the vehicle mobility characteristics, based on which the effectiveness of the hybrid communication in supporting real-time path planning is studied. However, since the pathes of vehicles cannot be changed or controlled by the external algorithm in VISSIM, we further develop a Java-based platform to investigate the performance of the proposed path planning algorithm. Specifically, average moving delay (AMD), defined as the average travel time per trip, is used as a metric in the evaluation.

**A. Simulation Setup**

1) **Simulation settings in VISSIM:** To simulate a VANET with VISSIM in Kitchener-Waterloo (K-W) downtown region, vehicles are pushed into the region of 6000m * 2800m, as shown in Fig. 3. At the beginning of the simulation, vehicles are set to enter the region from the preset entries (e.g., 9 entries at the ends of main roads), following a Poisson process at a rate 2500 vehicle/hour/entry. The proportion of a bus or a taxi in the traffic flow is set as 5%. After the duration of the first 240s, the vehicle pushing-in stops to reach an equivalent average density 30 vehicle/km/lane which represents a medium density scenario. Similarly, if the first duration is set to be 480s, the scenario becomes a high density one. In the VISSIM, vehicle information (e.g., location and velocity, etc) is recorded every 0.2s. The total simulation time lasts for 3000s. In addition, the velocity distribution for all vehicles follows the velocity model described in Subsection III-C with parameters $v_L = 30$km/hour, $v_H = 60$km/hour, and $D = 600$ s. The reduced speed areas can be set at any time during the simulation in VISSIM, to simulate different kind of incidents/accidents in the suburban scenarios.

2) **Simulation settings in Java:** To evaluate the performance of the path planning algorithm in Java, with the same region, 500 vehicular nodes with transmission radius of 150 meters are first randomly deployed to cover K-W downtown region, as shown in Fig. 3. In addition, 12 intersections are chosen as candidates for RSU deployment in the region. Further, each vehicle moves to its destination with a velocity of 60 km/h (or 30 km/h). The path planning can be performed at the beginning of a sample time, e.g., 10s. The lifetime of a warning message, $T_{L}$, is set as 300s. The duration for each simulation is set to be three hours, and the results are averaged over 100 runs. To illustrate the effect of different kinds of accidents on path

![Fig. 3: The simulation scenario of University of Waterloo region in VISSIM.](image-url)
planning, big accidents are set to last for 20 mins, while small accidents are set to last for only 10mins.

B. Evaluation of Transmissions in VISSIM

We first evaluate the transmission performance of VANETs in a high density scenario. The evaluated metrics are the connection probability of a vehicle to an RSU and the end-to-end transmission delay. As shown in Fig. 4(a), in a high density scenario, the connection probability is high even without the support of cellular network. For instance, when the vehicle transmission range is 120m (which is very easy to be reached as discussed in [45] and way larger than the average headway distance), the connection probability can be 80%. As the transmission range of vehicle increases, the connection probability increases, since the increased the transmission range supplies more chances to connect with other vehicles or RSUs. Furthermore, as shown in Fig. 4(a), in the high density case, the transmission delay is only around 5.5s, which is less than a sample time 10s. Notice that a short end-to-end transmission delay facilitates the implementation of real-time path planning, which needs traffic information update as timely and accurate as possible.

The inter-contact time is evaluated through the vehicle headway distance (i.e., V2V distance) and the last-hop V2R distance. Based on the trace files from VISSIM, Fig. 4(b) shows the probability density function (PDF) of vehicle headway distance. It is shown that the PDF of the headway distance matches well with an exponential distribution as shown in Fig. 4(b), which validates the premise in Subsection III-C. Based on the resultant headway-distance distribution, the average V2V inter-contact time, \( E[T_{c2v}] \), can be obtained, as shown in Section IV-B.

Besides, the PDF of the distance from the last-hop vehicle to the nearest RSU for one delivery is given in Fig. 4(c). The simulated PDF matches well with the theoretical PDF, which is calculated with the parameters in the simulation setup based on Eq. (5). According to Fig. 4(c), the average distance from a last-hop vehicle to its nearest RSU can be further calculated to be around 180m. Then the transmission delay incurred by the inter-contact time of the last-hop V2R transmission can be calculated as discussed in Section IV-B, i.e., \( E[T_{off}] = E[U_{off}]/V = E[Last - hop V2R distance - R]/V \).

We then investigate the end-to-end transmission performance in terms of the connection probability and transmission delay in the medium density scenario. Based on the proposed transmission mechanism, a hybrid VANET is utilized to reduce the transmission delay, making the path planning more efficient and timely. As shown in Fig. 4(d), via pure VANETs, the average end-to-end transmission delay decreases as the transmission range increases, since the increased transmission range gives higher possibilities for a transmitting vehicle to find an end-to-end path to an RSU (given negotiable transmission delay when two vehicles are within the transmission range of each other). Moreover, in hybrid VANETs, when the public transportation system is utilized, the increased transmission range can significantly create more chances to meet a bus or a taxi, thus leading to a smaller transmission delay. Notice that once any bus or taxi nodes receive the messages, they can help deliver the messages to the vehicle-traffic server directly via the cellular network, and the intermittent connections of the multi-hop VANET can be efficiently reduced. Especially, as the transmission range of vehicles becomes smaller (i.e., the problem of intermittent connections in VANETs is severer), the delay reduction comes to be bigger if the hybrid VANET-enhanced transportation system is involved. The reason is that with a smaller transmission range, an end-to-end transmission path is more difficult to be guaranteed by pure VANETs, leading to a larger delay gap compared to the one that utilizes the hybrid VANET-enhanced transportation system. In addition, the simulated results of transmission delay match well to the theoretical ones shown in Eq. (10). Hence, based on the proposed transmission mechanism, an efficient and timely message transmission for path planning can be achieved, which makes it possible to perform global real-time path planning.

C. Simulation of the Proposed Path Planning in Java

Fig. 5(a) shows the average moving delay with and without implementing the proposed path planning algorithm. We can observe that the AMD with the proposed path planning is much lower than that without path planning. For example, when accident number is two, AMD is reduced by 35%. Furthermore, with more accidents, AMD becomes longer; however, the ones utilizing the proposed path planning algorithm increase more slowly. The cost of path planning in terms of the increased path length is also shown in Fig. 5(a). When a vehicle wants to change its previous shortest path due to a sensed accident ahead, a novel smooth path is generated with less AMD at the cost of the increased path length. It shows that the average cost for users is still admissible when traffic environments are in terrible conditions.

In addition, Fig. 5(b) shows the AMD comparison between our proposed path planning algorithm and a distributed path planning algorithm proposed in [46]. In the distributed path planning, each individual vehicle re-searches a new path based on the known information of accidents when it receives any information on congestion or accidents, but neither with coordination among vehicles nor considering the individual cost of path planning. As shown in Fig. 5(b), AMD under our proposed path planning is reduced on average by 27% as compared with that of the distributed algorithm. Because each individual vehicle plans path only on its own interest, it is very possible that a number of vehicles swarm into the same road segment based on the same warning message information. Then new traffic jam can happen with high probability and result in the increased AMD. Fig. 5(b) shows a good adaptability of the proposed path planning algorithm to avoid introducing other traffic jam.

Fig. 6(a) illustrates the effect of different kinds of accidents on AMD. It is shown that when a big accident continues for a long duration (i.e., 20 minutes), AMD increases, compared to a small accident (i.e., lasting 10 minutes only). This is
Fig. 4: The performance evaluation of the proposed transmission mechanism in a medium vehicle-density scenario.

because that some vehicles have no capabilities to change their current paths (e.g., buses), AMD increases due to their longer trapped time in congestion. Similarly, when the number of accidents increases, AMD becomes longer, but not much. Thus, it implies that our proposed path planning algorithm is with a good adaptability to different accident durations. Besides, if the number of slow-speed vehicles increases, more vehicles slowed down to 30 km/h will introduce larger AMD as shown in Fig. 6(b). Since more slow vehicles on one road can result in a high vehicle density, Fig. 6(b) shows a good adaptability to vehicle densities. Furthermore, comparing this performance with the one in Fig. 5(a), AMD is a little longer than the case under few slow vehicles, since network vehicle-traffic throughput is diminished due to more vehicles with slow speed stranded on one road.

The sensitivity analyses in terms of both the vehicle number and the number of accidents on average moving delay (AMD) are discussed in Fig. 7. Here, we considerer that the accidents are big ones, lasting for 20 mins. First, we can see that the AMD increases with the increased number of vehicles under our algorithm in Fig. 7. The reason for this AMD increment is that more vehicles may result in a higher probability of introducing another traffic jam at crossings. However, taking the case with three accidents as an example, even when the number of vehicles increases to 800, AMD is relatively small, around 375s as shown in Fig. 7. This result shows a good adaptability of the proposed path planning algorithm to the total vehicle number. In addition, Fig. 7 shows that the AMD increases with the increased number of accidents with the similar trend as aforementioned.

VIII. CONCLUSIONS

In this paper, we have developed a hybrid-VANET-enhanced real-time path planning for vehicles to avoid congestion in an ITS. We first propose a hybrid-VANET-enhanced ITS framework with functionalities of real-time traffic information collection, involving both V2V and V2R communications in VANETs and cellular communications in public transportation system. Then, a globally optimal real-time path planning algorithm is designed to improve overall spatial utilization and reduce average vehicle travel cost, by means of Lyapunov optimization. Extensive simulations have been conducted to demonstrate that the proposed path planning algorithm can achieve better performance than that without real-time path planning in terms of average moving delay as well as the adaptability to different accident durations and traffic densities. In our future work, we intend to find large-scale real-world vehicle traffic traces to further validate benefits of the proposed
algorithm in practical scenarios.

REFERENCES


Rongxing Lu (IEEE S09-M11) received the Ph.D. degree in computer science from Shanghai Jiao Tong University, Shanghai, China in 2006 and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2012. He has been an assistant professor at the School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore, since May 2013. Before that, he worked as a Postdoctoral Fellow at the University of Waterloo from May 2012 to April 2013. His research interests include wireless network security, big data security and privacy, network coding security, and applied cryptography.

Ran Zhang received his B.E. degree (2010) from Tsinghua University (China) in Electronics Engineering. He is now working toward his Ph.D. degree in Broadband Communication Research Group from University of Waterloo (Canada). His current research interest includes resource management in heterogeneous wireless access networks, carrier aggregation in Long Term Evolution - Advanced (LTEA) systems and electrical vehicle charging control in smart grids.

Fan Bai (General Motors Global R&D) is a Senior Researcher in the Electrical & Control Integration Lab., Research & Development and Planning, General Motors Corporation, since Sep., 2005. Before joining General Motors research lab, he received the B.S. degree in automation engineering from Tsinghua University, Beijing, China, in 1999, and the M.S.E.E. and Ph.D. degrees in electrical engineering, from University of Southern California, Los Angeles, in 2005. His current research is focused on the discovery of fundamental principles and the analysis and design of protocols/systems for next-generation Vehicle Ad hoc Networks (VANET), for safety, telematics and infotainment applications. Dr. Bai has published about 40 book chapters, conference and journal papers, including Mobicom, INFOCOM, MobiHoc, SECON, ICC, Globecom, WCNC, JSAC, IEEE Wireless Communication Magazine, IEEE Communication Magazine and Elsevier AdHoc Networks Journal. In 2006, he received Charles L. McCuen Special Achievement Award from General Motors Corporation in recognition of extraordinary accomplishment in area of vehicle-to-vehicle communications for drive assistance & safety. He serves as Technical Program Co-Chairs for IEEE WIVec 2007 and IEEE MoVeNet 2008. He is an associate editor of IEEE Transaction on Vehicular Technology and IEEE Transaction on Mobile Computing, and serves as guest editors for IEEE Wireless Communication Magazine, IEEE Vehicular Technology Magazine and Elsevier AdHoc Networks Journal. He is also serving as a Ph.D. supervisory committee member at Carnegie Mellon University and University of Illinois-Urban Champaign.

Xueming (Sherman) Shen (IEEE M’97-SM’02-F09) received the B.Sc.(1982) degree from Dalian Maritime University (China) and the M.Sc. (1987) and Ph.D. degrees (1990) from Rutgers University, New Jersey (USA), all in electrical engineering. He is a Professor and University Research Chair, Department of Electrical and Computer Engineering, University of Waterloo, Canada. He was the Associate Chair for Graduate Studies from 2004 to 2008. Dr. Shen’s research focuses on resource management in interconnected wireless/wired networks, wireless network security, wireless body area networks, vehicular ad hoc and sensor networks. He is a co-author/editor of six books, and has published many papers and book chapters in wireless communications and networks, control and filtering. Dr. Shen served as the Technical Program Committee Chair for IEEE VTC’10 Fall, the Symposium Chair for IEEE ICC’10, the Tutorial Chair for IEEE VTC’11 Spring and IEEE ICC’08, the Technical Program Committee Chair for IEEE Globecom’07, the General Co-Chair for Chinacom’07 and QShine’06, the Chair for IEEE Communications Society Technical Committee on Wireless Communications, and P2P Communications and Networking. He also serves/served as the Editor-in-Chief for IEEE Network, Peer-to-Peer Networking and Application, and IET Communications; a Founding Area Editor for IEEE Transactions on Wireless Communications; an Associate Editor for IEEE Transactions on Vehicular Technology, Computer Networks, and ACM/Wireless Networks; and the Guest Editor for IEEE JSAC, IEEE Wireless Communications, IEEE Communications Magazine, and ACM Mobile Networks and Applications. Dr. Shen is a registered Professional Engineer of Ontario, Canada, an IEEE Fellow, a Fellow of the Canadian Academy of Engineering, a Fellow of Engineering Institute of Canada, and a Distinguished Lecturer of IEEE Vehicular Technology Society and Communications Society.