Data-driven Vehicular Communications in Urban Vehicular Network

Wen Zhang
School of Computer Science and Technology
University of Science and Technology of China
Hefei, China
e-mail: zw666@mail.ustc.edu.cn

Zhengyang Zhou
School of Computer Science and Technology
University of Science and Technology of China
Hefei, China
e-mail: zzy0929@mail.ustc.edu.cn

Chuancai Ge
School of Computer Science and Technology
University of Science and Technology of China
Hefei, China
e-mail: gcc810@mail.ustc.edu.cn

Pengkun Wang
School of Software Engineering
University of Science and Technology of China
Hefei, China
e-mail: pengkun@mail.ustc.edu.cn

***Abstract*—Vehicular communications, referring to data packets exchange among vehicles and infrastructures, have attracted a lot of attention recently because of its contributions in intelligent transportation systems. Due to the complexity of large-scale network topology and dynamic of mobile nodes, vehicular communication is difficult to be established in an ultra-reliable and low-latency way. Existing efforts of vehicular communications mainly focused on historical trajectories and perform experiments on simulated road networks with free moving vehicles, or just analyze a limited scope of traffic pattern over public vehicles. However, previous studies ignored the traffic pattern of urban private vehicles, leading to a narrow perception for data transmission. To make up for this blind spot, we propose a data-driven method for vehicular communications by analyzing the traffic pattern of urban vehicles. We model the urban vehicles traffic pattern by deep neural networks and clustering technique is applied to enhance the ability of collecting data for vehicles. Additionally, to deal with the unbalanced traffic load in road networks, we devise a novel method to detect the appropriate location for deploying auxiliary WiFi-spots to help data transmission in low traffic load area. Therefore, our solution enables the data packets to be transmitted in an optimal way thus comprehensive and valuable information is collected for guiding the transmission. Our proposed method outperforms the state-of-the-art vehicular communication methods in terms of diffusion speed and range in urban vehicular networks.**

Keywords-data-driven; urban vehicles; vehicular communication; urban vehicular networks

I. INTRODUCTION

Automated driving has been regarded as an interesting research topic in past few years. Recently, the advancements in artificial intelligence and information technologies enables automated driving to be true [1]. An automated driving vehicle should be capable of sensing its surrounding environment with onboard surveillance system, understanding the driving scenarios, and making appropriate driving decisions. In addition, information collected by nearby vehicles can assist the automated driving vehicle to make more intelligent decisions. Vehicular communication networks are the medium that connect vehicles, infrastructures, clouds, as well as other devices with communication modules, thus automated driving vehicles can obtain local and global information. Automated driving vehicles communicate directly with other urban vehicles and infrastructures to extend their perception beyond the range offered by their integrated sensors. As illustrated in Fig. 1, an automated driving vehicle can acquire local information by surveillance system equipped in itself, and collect global traffic information by nearby urban vehicles and infrastructures through vehicular communications. The global traffic information help automated driving vehicle make intelligent decisions for driving.

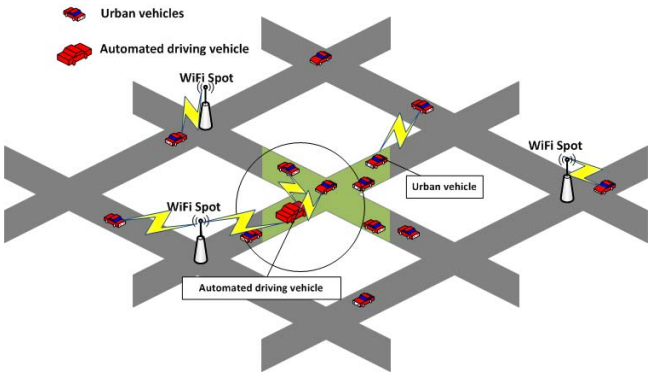

Figure 1. Automated driving vehicle with urban vehicles

However, existing vehicular communication methods, which assumes simulated road networks [2][3] or in a limited scope of traffic pattern analysis [4][5], thus they can't obtain traffic information that carried by urban private vehicles for the automated driving vehicle. It remains challenging for automated driving vehicle to collect and model helpful information from nearby urban vehicles globally due to high mobility of private vehicles and frequent interrupted connection in the network. To address the limitations above,

978-1-7281-2184-0/19/$31.00 ©2019 IEEE

we propose a novel data transmission algorithm to obtain more efficient and reliable real-time information from the urban vehicles and infrastructures. Firstly, we analyze the urban vehicles traffic pattern from the historical trajectories of urban vehicles in different time intervals. Specifically, assistant ability, which quantifies the meaningful information collected for automated driving vehicles, will be calculated through the clustering of urban vehicles. Secondly, we differentiate the traffic load in urban road networks, and select intersections on low traffic load to deploy WiFi-spots, which can transmit data packets to automated driving vehicle to make smart decisions. Finally, a novel communication algorithm is proposed to select the most significant and useful data packets to transmit to automated driving vehicles.

For performance evaluation, we compare our Data-Driven Vehicular Communications (DDVC) algorithm with First In First Out (FIFO) strategy, the classic photoNet [6] algorithm, and the Realistic and Optimized V2V Communication (ROVC) [5] algorithm on real-world Suzhou Industry Park (SIP) datasets. The results demonstrate the superior performance of our algorithm in terms of the average diffusion speed and range.

In summary, the main contributions of our work are as follows:

- To the best of our knowledge, this is the first paper to solve the vehicular communications with urban private vehicles traffic data. We proposed a novel data-driven vehicular communication method to achieve efficient and reliable real-time data transition in urban vehicular networks for automated driving.
- We analyze the urban vehicles traffic pattern from the historical trajectories and surveillance systems by deep learning method. Then we calculate the assistant ability for each urban vehicle participating in large-scale urban vehicular networks.
- A novel intersection selection algorithm is devised to choose the best adaptive intersections for deploying WiFi spots, making up for the low density of traffic load in vehicular networks.
- We evaluate our solutions on real-world datasets, and the results demonstrate that the proposed algorithms outperform other alternative approaches.

The rest of this paper is organized as follows: we first introduce some related work for packet transition methods in vehicular networks in section II. Then we formulate the problem in Section III. We will give a more elaborate explanation of our algorithms in section IV. Numerical studies are presented in Section V and Section VI concludes the paper.

II. Related Work

Previous efforts on traditional Vehicle-to-Vehicle (V2V) communications, most of them are constructed based on vehicle trajectories [2][7][8][9], considering geographical information [10] and network connectivity [11][12][13]. We focus mainly on trajectories analysis recently.

In detail, majority of previous studies assume a simulated road network with free moving vehicles [2][3][7]. Reference [2] proposes a communication protocol, which requires the future information of a vehicle to be broadcasted in addition to current information, and then the efficiency of the vehicles can be improved in a simulated four-way intersection. Reference [3] sheds light on the need for mechanisms that take vehicles' trajectories into account. It presents a characterization that shows the spatiotemporal regularity of the vehicle movement. Then it gives a new methodology for identifying the spatiotemporal relationship between vehicle trajectories. Reference [7] is a fully distributed and lightweight scheme, requiring only a minimum hardware deployment, and does not need synchronization between vehicles or any pre-constructed signal maps. However, they all fail to involve the datasets of private cars and cannot capture the global conditions of the whole road network.

In [4] and [5], the authors analyze the traffic patterns of taxicabs. Reference [4] informs us an apparent scale and city-invariant law of temporal reachability in vehicular networks. It evaluates the proposed model with urban taxicab trajectories that exhibit an unprecedented combination of dependability, scale and generality of vehicular mobility. Reference [5] takes both the habit-driven individual patterns of single taxicabs and the urban-layout-driven patterns of crowds of taxicabs into consideration. It proposes a novel method to calculate the expected probability that a single taxicab obtains the information, and proves that the time complexity of real-time optimizing is also linear. Notwithstanding both studies enable promising results, but they also analyze traffic patterns in a limited scope of public vehicles.

In conclusion, all these algorithms target at a simulated road network or a limited scope of public vehicles. We argue that a data-driven perspective should be taken into understanding traffic patterns of all urban vehicles. To distinguish our work from others, this is the first paper that involves the spatiotemporal patterns of urban vehicles for data transmission via deep learning techniques. Meanwhile, our method for location selection to deploy WiFi spots is proposed to address the unbalanced traffic load, and then we devise a novel data-driven vehicular communication to transmit data packets to automated driving vehicles.

III. Problem Defination

A. Facts

The study area of SIP, which contains 1276 intersections and 2414 roads, covering 288 square kilometers. As illustrated in Fig. 2 (a), a series of fixed video cameras are pre-deployed on 130 key intersections to record urban vehicles passed by intersections. We utilize this surveillance system to reconstruct the trajectories of urban private vehicles, and then mine the traffic patterns of urban vehicles. There are 299,818 local private vehicles running independently and constantly in the SIP road network. As shown in Fig. 2 (b), there are 4305 taxicabs equipped with GPS devices running in the SIP road network every day. We

can collect GPS trajectories datasets of taxicabs and the sample rate of GPS points is every 30 seconds.

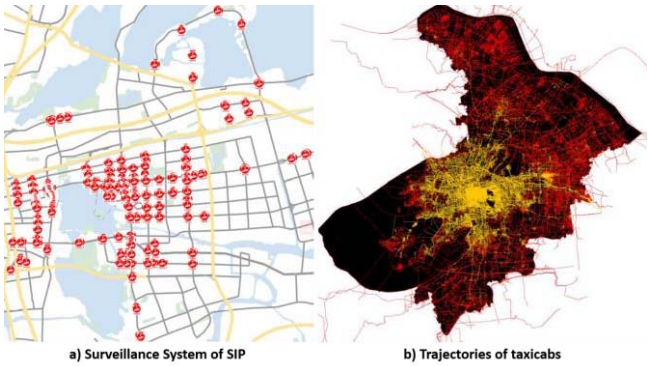

Figure 2. Surveillance of SIP and trajectories of taxicabs

B. Assumption

Vehicles can communicate with each other within the communication range via its equipped device such as Dedicated Short Range Communications (DSRC) and WiFi, to participate in a vehicular network. The wireless device can be IEEE 802.11p [14], IEEE 1609.x [15] or DSRC [16] device, and the communication radius is R. Moreover, each vehicle can communicate with K vehicles at the same time, and transmit N data packets in one communication period. WiFi spots, a device equipped in intersections, can communicate with urban vehicles by WiFi or DSRC, and the communication radius is R_1. In our experiment, the taxicabs are regarded as automated driving vehicles, and urban private vehicles represent the urban vehicles that participate in urban vehicular networks. Automated driving vehicles can make more intelligent control decisions if they collect more information of road networks. Notice that, we assume that the data packets are delivered to target vehicles once two vehicles establish connection successfully.

C. Problem Definition

Definition 1 (ROAD NETWORK). Given the intersections and road segments between intersections, the road network of the main urban area of SIP can be modeled as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where vertex set $\mathcal{V}$ denotes all intersections and edge set $\mathcal{E}$ corresponds to all road segments. $\forall\ i, j \in [1, |\mathcal{V}|]$, intersection $I_i \in \mathcal{V}$ and $r_{ij} \in \mathcal{E}$ indicates the road segment from intersection I_i to I_j. Note that road segments r_{ij} and r_{ji} represent two different road segments between intersection I_i and I_j with opposite directions.

Definition 2 (TRAJECTORY). Let $TR_{v_i}(t)$ be a collection of trajectories of vehicle v_i in one specific time interval t, and it can be formulated as $TR_{v_i}(t) = \{r_{k_1k_2}(t_1,t_2), r_{k_2k_3}(t_2,t_3), \ldots, r_{k_{e-1}k_e}(t_{e-1},t_e)\}$ which means vehicle v_i starts its trip at road segment $r_{k_1k_2}$ at time t_1, then ends at road segment $r_{k_{e-1}k_e}$ and time t_e. For one element $r_{k_{j-1}k_j}(t_{j-1},t_j)$, $2 \le j \le e$ in this trajectory, t_{j-1} is the time stamp that vehicle v_i enters road segment $r_{k_{j-1}k_j}$ and time stamp t_j corresponds to the time it leaves this road segment.

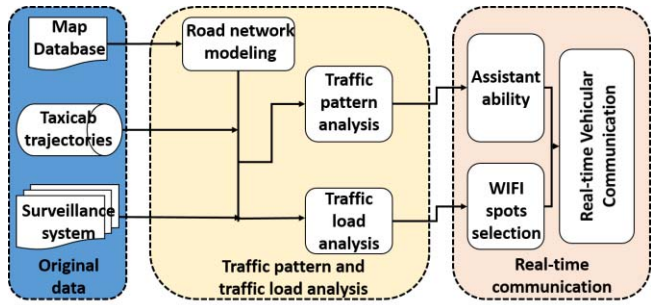

Figure 3. Solution overview

Definition 3 (DATA PACKETS). A data packet can be denoted as a tuple (r_{ij}, s, t). One data packet is corresponding to road segment r_{ij} at time t, and s is the road condition in r_{ij} consists of average speed and other dynamic features.

Definition 4(ROAD TRAIFFIC LOAD). Traffic load of road network, referring to number of the vehicles pass by road segments, can be formulated by $F = \{f_1, f_2, \ldots, f_{|\mathcal{E}|}\}$, where f_i is the traffic volumes of road segment $r_{ij} \in \mathcal{E}$.

Problem (V2V COMMUNICATION OPTIMIZATION PROBLEM). Optimize the transmissions between two encountered vehicles, and then maximize the valid data packets including traffic information of road segments that an automated driving vehicle can obtain.

IV. Data-driven Urban Vehicular Communications

A. Solution Overview

In this section, we give an overview of our data-driven vehicular communication solution. Fig. 3 shows the overall architecture of our solution, which consists of three major parts: a) original data preprocess; b) traffic pattern and traffic load analysis; and c) optimization module: maximize the transmitting efficiency with assistant ability of each cluster and priority of data packets. The first part of the scheme extracts the deep representation of urban vehicle trajectories by an encoder-decoder network, and then maps all vehicles into different clusters with a clustering algorithm. We devise a novel method to evaluate the assistant ability for each vehicle by clustering of urban vehicles. Next, we analyze the traffic load of road network and devise a method to select most important intersection to deploy WiFi spots. Based on these preliminary results, we propose a novel method to optimize the data packet transmissions between two encountered vehicles. The proposed model can maximize the quantity of traffic information obtained by automated driving vehicles. We elaborate each step of our algorithm in the following sections.

B. Traffic Pattern Analysis of Urban Vehicles

Previous studies on traffic pattern analysis, such as [17] and [18] have already suggests that traffic patterns of urban vehicles have both strong spatial and temporal patterns. Fig.

4 illustrates the traffic volumes and directions of road segments during 8:00a.m. to 8:30a.m. and 5:30p.m. to 6:00p.m. in SIP road network respectively.

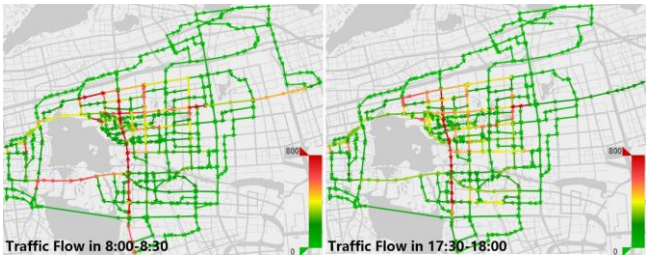

Figure 4. Traffic patterns of urban vehicles

Obviously, the traffic flows of most road segments during the rush hours of morning and afternoon are nearly opposite. This is in line with people's travel patterns such as commuting. Fig. 4 confirms the spatial and temporal characteristics of urban vehicle traffic flows in SIP. Based on them, we conclude that a group of vehicles that have similar origins and destinations at same time interval may share similar trajectories. Specifically, we infer that the traffic flow of urban private vehicles must follow the intragroup characteristics.

C. *Seq2Seq Based Trajectories Clustering*

Trajectory clustering is a traditional method of traffic pattern analysis, which can be employed to detect groups of similar trajectories. However, when it combines with deep learning techniques, it will produce amazing results. In this work, we extract the features of urban vehicles moving behaviors by seq2seq model [19]. Considering the urban vehicle trajectory $TR_{v_i}(t)$, we extract the attributes of road segment and then construct as a sequence $BR_{v_i} = \{b_1, b_2, \cdots, b_n\}$, where b_i consists of speed, time interval and lanes of road segment. Human mobility changes can be reflected by the different attributes of road segments between two consecutive records. Notice that, we utilize an embedding method for categorical attributes such as road segment, lanes of road segment. Finally, we put this extract sequence into a multi-layer encoder-decoder network and get a fixed length vector to represent the urban vehicle trajectory.

We describe a model that uses multi-layer Bidirectional-Long Short Term Memory (Bi-LSTM) and multi-layer Long Short Term Memory (LSTM) to extract the moving behavior sequence and generate a fixed-length deep representation of the trajectory. The multi-layer encoder-decoder model is composed of two modules - the encoder network is shown on the left of Fig. 5 and the decoder network is illustrated on the right part. The input of the model is a vehicle moving behavior sequence BR_{v_i}. The input sequence is feed into the encoder network, a multi-layer Bi-LSTM, sequentially and the hidden state h_t is updated subsequently. Each layer of the encoder Bi-LSTM is updated by:

$$h_t = f_{bi-lstm}\left(h_{t-1}, b_t\right) \tag{1}$$

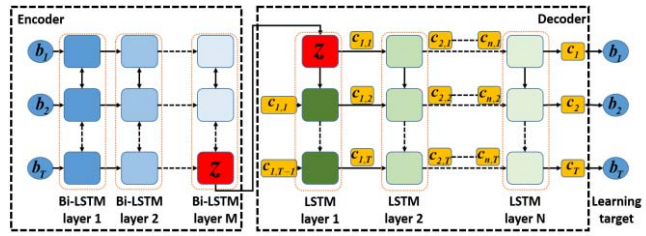

Figure 5. Deep representation of trajectory by Seq2Seq

After the last b_T is processed, the hidden state h_T is used as the representation for the whole sequence, formulated by z. Then, the decoder first generates the output $c_{1,1}$ by taking z as the initialized hidden state of the decoder LSTM, and then further generate (c_1, c_2, …, c_T), as illustrated in Fig. 5. Similarly, there is a multi-layer LSTM in decoder network. The decoder LSTM is updated by:

$$h_t^d = f_{lstm}\left(h_{t-1}^d, c_{t-1}^d, h_T^{d-1}\right) \tag{2}$$

where h_t^d refers to the hidden state of d layer LSTM.

The target of the decoder is to reconstruct the input sequence $BR_{v_i} = \{b_1, b_2, \cdots, b_n\}$. In other words, the encoder Bi-LSTM and decoder LSTM are trained together by minimizing the reconstruction error, measured by the general mean squared error $\frac{1}{T}\sum_i \|b_i - c_i\|^2$. As the output is compelled to be as close as possible to the input sequence by the network, the training process can be unsupervised. The fixed-length moving behavior vector z is a meaningful representation for the input behavior sequence BR_{v_i}, because the completely input sequence can be reconstructed from z by the LSTM decoder. After this procedure, we get the moving behavior vector set $Z = \{z_1, z_2, \ldots, z_n\}$. Then, we feed them in a classic clustering algorithm, such as K-means.

The encoder-decoder network contains 3 Bi-LSTMs in encoder and 3 LSTMs in decoder. The numbers of hidden units in Bi-LSTM and LSTM are both set to 1024. The results of trajectories clustering as illustrated in Fig. 6. To show the efficiency of vehicle clustering intuitively, we select 4 random vehicle clusters and visualize trajectories of these 4 clusters from 8:00a.m. to 8:30a.m. and 5:30p.m. to 6:00p.m. on workdays and weekends. As shown in this figure, there exists significant differences among the trajectories of the corresponding 4 vehicle clusters during different time periods on different types of days.

D. *Traffic Load Analysis of SIP Road Network*

Due to the distribution of commercial, residential and industrial areas, the traffic load of road segment in road network is unbalanced, as illustrated in Fig. 4. The red road segments refer to high traffic load roads while the green roads correspond to low traffic load. The traffic load of SIP road network is unbalanced.

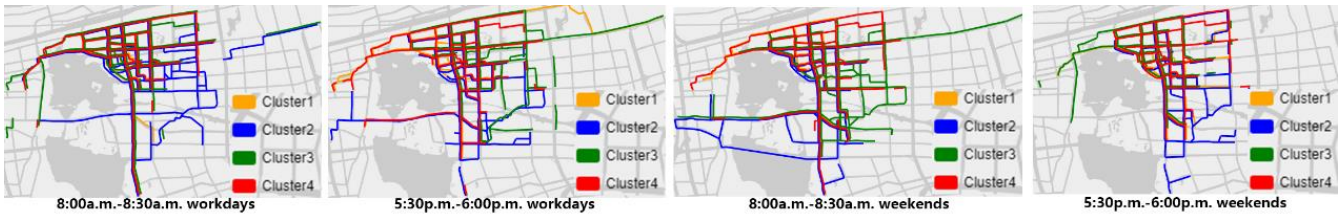

Figure 6. The clusters of different time period

Automated driving vehicles can meet up with urban vehicles on roads with high traffic load easily. However, there is rarely urban vehicles on roads with low traffic load, and thus the automated driving vehicles cannot obtain information from other vehicles easily. Therefore, in order to help transmission of data packets, we deploy WiFi spots to collect information from urban vehicles on low traffic load roads. We devise a novel method to select the fewest number of intersections to realize the most effective results of data collection. We calculate weights of road segments by the traffic load, and then select 5 intersections to build WiFi spots. We elaborate this method as follows.

First, given the traffic load of all road segments in SIP, we define the weight of road segment in road network through the traffic volume that represents the traffic load by:

$$w_i = \frac{1}{\ln(1+f_i)+1}, f_i \in F, i \in \left[1, |\mathcal{E}|\right] \tag{3}$$

where f_i refers to the traffic volume of road segment r_i. Here we take the advantage of logarithmic function due to the distribution of the traffic volumes. We get weights of all road segments in SIP and reconstruct matrix A as:

$$A = \begin{bmatrix} 0 & a_{12} & \cdots & a_{1|\mathcal{V}|} \\ a_{21} & 0 & \cdots & a_{2|\mathcal{V}|} \\ \vdots & \vdots & \ddots & \vdots \\ a_{|\mathcal{V}|1} & a_{|\mathcal{V}|2} & \cdots & 0 \end{bmatrix} \tag{4}$$

where $a_{ij} = \begin{cases} w_k, & if \quad r_k \in \mathcal{E} \\ 0, & otherwise \end{cases}$. Next, we calculate the weights of all intersections by:

$$AI = A * E \tag{5}$$

where E is a matrix with size $|\mathcal{V}|$ *1, and all elements in E are 1.

Eventually, we rank AI in ascending order, and select top-5 intersections to build WiFi spots.

E. Vehicular Communications

In this section, we explain the proposed vehicular communication algorithm in detail. The novel proposed method consists of two parts: I) V2V communications, II) Vehicle-to-Infrastructure (V2I) communications. Automated driving vehicles exchange data packets with urban vehicles and collect traffic information for intelligent control decisions by V2V communications. Meanwhile, automated driving vehicles obtain meaningful information from WiFi spots by V2I communications.

First, we calculate the assistant ability for each urban vehicle, which quantifies the useful information carried in urban vehicles. Considering an automated vehicle v_i, its future trajectory formulated by road segments vector R = $\{r_1, r_2, \ldots, r_k\}$. We directly map the urban vehicles into different clusters, which is formulated by C = $\{cl_1, cl_2, \ldots, cl_M\}$. We count the number of occurrences of road segments appearing in vehicle cluster cl_i and get a vector $RFI = \{rfi_1, rfi_2, \ldots, rfi_{|\mathcal{E}|}\}$, where rfi_j corresponds to the appearing frequency of road segment r_j in cluster cl_i. Then we evaluate assistant ability of each cluster cl_i by:

$$ab_i = \frac{1}{|\mathcal{E}|}\sum_{m=1}^{|\mathcal{E}|} \ln\left(1+rfi_m\right) where \quad r_m \notin R \tag{6}$$

here, we calculate the assistant ability by the road segments in clusters cl_i but not in the future trajectories of vehicle v_i. We also utilize a logarithmic function as its nice property to model distribution of the road segment frequency in clustering.

Based on that, we obtain all clustering assistant abilities $AB = \{ab_1, ab_2, \ldots, ab_M\}$. An automated driving vehicle v_i meets up with urban vehicles, select the most appropriate vehicle to communicate via clustering assistant ability.

Next, we derive a formula to measure the priority of each data packet for automated driving vehicle. For an automated driving vehicle v_i in time interval t, if a packet isn't involved in its trajectory or v_i cannot receive this packet from other vehicles, the priority of this packet is higher. To be specific, we calculate the priority for each packet by:

$$p = \lambda_j * e^{-\frac{1}{dis(r_j, v_i)}}, where\, \lambda_j = \begin{cases} 1, & if \quad r_j \notin v_i \\ 0, & otherwise \end{cases} \tag{7}$$

In order to optimize the transmission efficiency, we propose our vehicular communications algorithm.

First, we describe our V2V algorithm. Once automated driving vehicle v_i and some urban vehicles $UV = \{uv_1, uv_2, \ldots, uv_n\}$ encounter in the same road segment, v_i should select a most appropriate urban vehicle uv_j to establish connection and exchange the packet lists. Then uv_j send packets that vehicle v_i doesn't have but can be found in uv_j. The central goal of this algorithm is to determine which packet should be transferred to v_i in a relatively short period, and this optimization is based on the priority of packets calculated by (7). The exact details are presented as follows:

- Regarding the automated driving vehicle v_i, we calculate the assistant ability for urban vehicles UV by (6). Then we select top-K maximal assistant ability vehicle uv_j to establish connection.

- For urban vehicle uv_j, we calculate priority of all packets for automated driving vehicle v_i by (7).
- We rank all packets by the priority values in ascending order.
- Send top - N packets to v_i.

Next, we demonstrate our V2I algorithm. Once automated driving vehicle v_i enters the communication range of a WiFi spot *ws*, v_i establishes the communication with *ws* and obtains traffic information from *ws*. Meanwhile, once an urban vehicle uv_j is connected with the *ws*, uv_j uploads the traffic information to the *ws*. We explain the process of an automated driving vehicle v_i to obtain information from WiFi spots as follows:

- For a WiFi spot *ws*, we first calculate the priority of all packets by (7).
- Rank all packets by the priority values in ascending order.
- Send top - N packets to v_i.

We will evaluate our method in a real-world dataset in the next part.

V. EXPERIMENT

A. Datasets

Both taxicab trajectories and surveillance information are of the same time range, 1st.Jan - 31th Mar.2017. The automated driving vehicles, which can communicate with urban private vehicles, are simulated with taxicabs. The road network of SIP contains 1276 intersections and 2414 road segments. The detailed descriptions of datasets are shown in Table I. We simulate and evaluate our proposed V2V and V2I communications in real-world SIP datasets from multiple aspects.

B. Evaluation

Instead of conducting experiments with simulation software, we perform comprehensive experiments to evaluate our proposed method on real-world datasets of SIP. We perform our experiments of traffic pattern analysis on GPU Tesla V100-PCIE-16GB with Python 3.6. Communication experiments are performed under Spark cluster. To be specific for evaluation setup, we make assumptions as follows. For all vehicles, they can obtain their own real-time location. For urban vehicles, they have the ability to get local traffic information such as speed, width of road segment. Thus, each vehicle carries real-time local traffic condition. In our experiment, we consider taxicabs as automated driving vehicles, and then we evaluate the percentage of global information obtained by automated driving vehicles through V2V and V2I during the same time. They can exchange the global information when the distance between automated driving vehicles and urban vehicles or WiFi spots in the communication range of R and R_1 respectively. Due to the highly speed mobility of the vehicle, we assume that each vehicle can transmit N data packets at one time. In order to avoid congestion, we assume that the maximal number of vehicles that can communicate simultaneously is K. We update the position and datagram of each vehicle carries according to the period T.

TABLE I. THE SIP DATASETS

Dataset	Description	Time | Content
Urban vehicles	The detailed GPS position of urban vehicels in SIP	1st.Jan - 31th Mar.2017
Taxicabs	The detailed GPS position of taxicabs in SIP	1st.Jan - 31th Mar.2017
Road network	The road network of SIP	1276 intersections and 2414 road segments

TABLE II. THE DEFAULT SETTINGS OF SIMULATED PARAMETERS

Symbol	Description	Value
R	Communication range of V2V	300m
R1	Communication range of V2I	500m
K	Number of connection in once time	1
N	Number of transition packets once time	3
T	Update time period	5 seconds
TTL	The lifetime of packet	15 minutes

Due to the information of road conditions is time-sensitive, the time period is set as 15 minutes, which meets the real-time requirements of road condition information. More details of settings in terms of simulated parameters are shown in Table II. The default communication range of V2V is 300m while the communication range of V2I is 500m. There are 472 automated driving vehicles are simulated in SIP road network. The Time To Live (TTL) of packet is 15 minutes, and the update period is 5 seconds. The automated driving vehicles can communicate with 1 urban vehicle and transmit 3 packets at one time of communication. And if no otherwise specified, we set the values of parameters to the default values.

The baselines are demonstrated as follows.

- FIFO: It is a classic algorithm of datagram exchange, which sends out the datagram firstly when comes first.
- photoNet: This method derives the priority of transmission for each packet based on TTL and a packet with smaller TTL has a higher priority to be transmitted first.
- ROVC: By analyzing the travelling patterns of taxicabs, this method predicts the near future's trajectories and derives the priority with predicted trajectories and TTL. The vehicle that has a higher possibility to fail to obtain data packets and has less TTL are selected to be transmitted first.

We evaluate the proposed model and baselines in the metric of average percentage of traffic information obtained in 15 minutes. In other words, it is the average percentage of real-time global traffic information obtained by vehicular communications for all automated driving vehicles within TTL. We denote the number of road segments of real-time traffic information obtained by automated driving vehicles as *nrs*. Then the average percentage of traffic information in the road network can be formulated as,

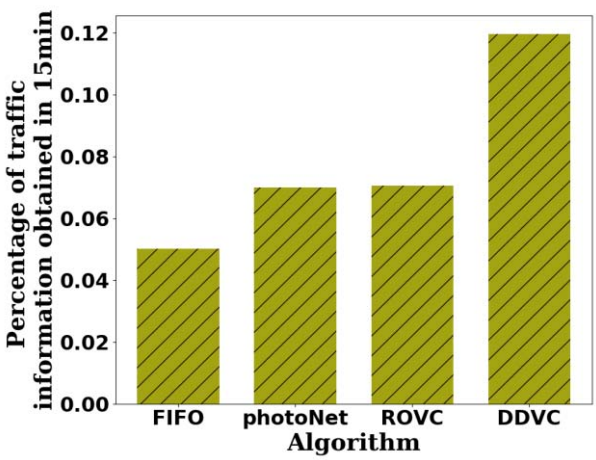

Figure 7. Percentage of traffic information obtained in 15 minutes.

$$\text{percentage of traffic information} = \frac{nrs}{|\mathcal{E}|} \tag{8}$$

Next, let us evaluate the performance of DDVC compared to state-of-the-art methods. As illustrated in Fig. 7, DDVC has best performance in terms of percentage of traffic information obtained in 15 minutes in real-world datasets. Our DDVC collects 12% road traffic information of SIP in 15 minutes, it means that we can obtain traffic information of 290 road segments in all 2414 roads. The performance of DDVC improves by 7%, 5%, 5% compared with FIFO, photoNet and ROVC. As a consequence, DDVC can obtain more global road traffic information in road network.

We can observe in Fig. 8, our DDVC collects more useful traffic information in terms of the maximal numbers of roads. DDVC can obtain 1908 roads traffic information in 15 minutes at most, and FIFO can only collect 413 roads. Thus, benefit from the urban vehicle traffic patterns analysis, and the speed and range for collecting traffic information of our DDVC have the best performance.

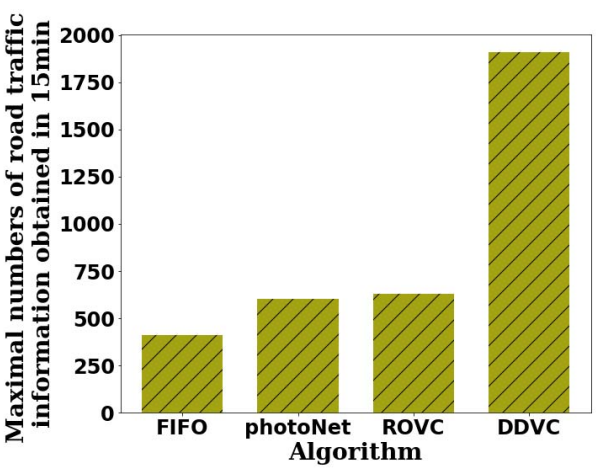

Figure 8. Maximal numbers of traffic information obtained in 15 minutes

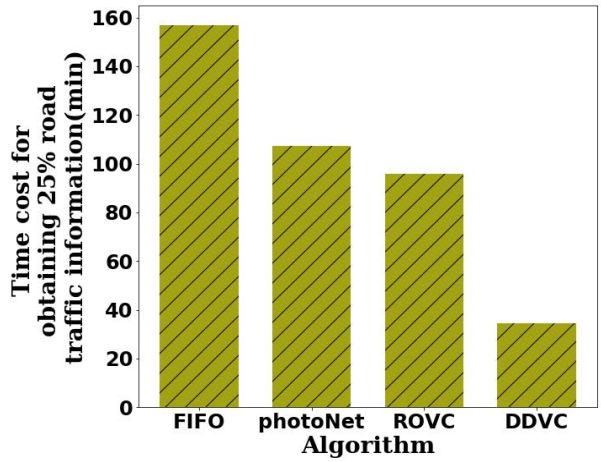

Figure 9. Time cost for obtaining 25% traffic information of SIP road networks

We evaluate the performance of DDVC from another aspect, and results are shown in Fig. 9. We calculate the time cost for obtaining 25% road traffic information in SIP road network. Our DDVC spends 35 minutes on collecting 25% roads in real-world datasets, and the FIFO needs to spend 157 minutes on obtaining 25% roads of all SIP road networks. It fully demonstrates that the proposed method significantly defeats the baselines.

We compare our solution with baselines for varieties of communication range of V2V in Fig. 10. With the increase of the communication range, the more traffic information obtained for automated driving vehicles. Our DDVC performs best in terms of different communication range compared with state-of-the-art methods.

The transition rate of V2V also plays a significant role in our simulated experiment. With the increase of transition rate, we can get more useful traffic information of SIP road network. As illustrated in Fig. 11, our DDVC obtains the most traffic information regardless of transition rate compared with alternative methods.

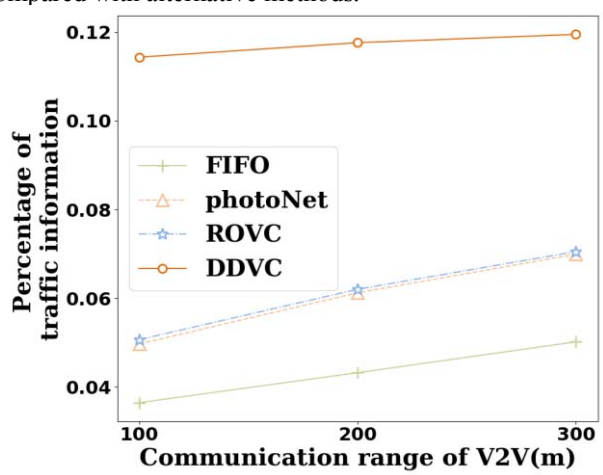

Figure 10. Impact of communication range

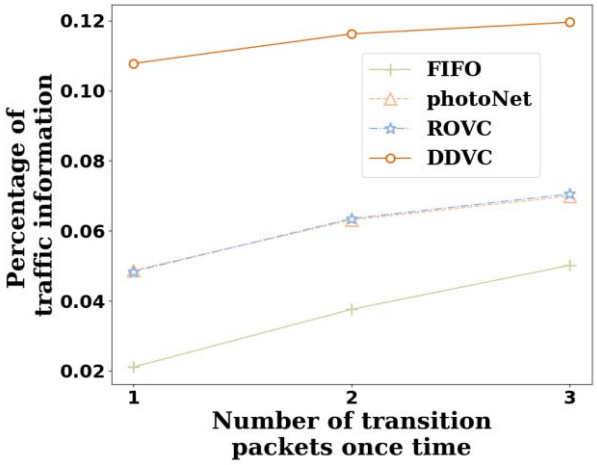

Figure 11. Impact of transition rate.

In summary, the proposed DDVC method outperforms other alternative approaches in real-world SIP road network. Our experiments suggest the performance of our proposed method, which can obtained real-time efficient and reliable traffic information for automated driving vehicles in an effective way.

VI. Conclusion

We proposed a novel data transition algorithm to obtain more efficient and reliable real-time information from the urban vehicles and infrastructures for automated driving. We analyze the urban vehicles and traffic load in urban road networks, and devise a method to select low traffic load intersections to build WiFi-spots. A novel communication algorithm is proposed to select the most significant and useful data packets to transmit to automated driving vehicles. We address the issue of vehicular communications with a data-driven perspective. With our solution, the automated driving vehicles can communicate with urban vehicles and infrastructures to make more intelligent decisions with real-time traffic information.

Acknowledgment

We gratefully acknowledge anonymous reviewers for reading this paper and giving valuable comments. Thanks for my supervisor and classmate.

References

[1] Zhang S, Chen J, Lyu F, et al. Vehicular Communication Networks in the Automated Driving Era[J]. IEEE Communications Magazine, 2018, 56(9): 26-32.

[2] Liu C, Lin C W, Shiraishi S, et al. Improving Efficiency of Autonomous Vehicles by V2V Communication[C]//2018 Annual American Control Conference (ACC). IEEE, 2018: 4778-4783.

[3] Celes C, Boukerche A, Braga R B, et al. Exploiting Daily Trajectories for Efficient Routing in Vehicular Ad Hoc Networks[C]//2018 IEEE International Conference on Communications (ICC). IEEE, 2018: 1-6.

[4] Bedogni L, Fiore M, Glacet C. Temporal Reachability in Vehicular Networks[C]//IEEE INFOCOM 2018-IEEE Conference on Computer Communications. IEEE, 2018: 81-89.

[5] Wang Y, Yang E, Zheng W, et al. A realistic and optimized v2v communication system for taxicabs[C]//2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS). IEEE, 2016: 139-148.

[6] Uddin M Y S, Wang H, Saremi F, et al. Photonet: a similarity-aware picture delivery service for situation awareness[C]//2011 IEEE 32nd Real-Time Systems Symposium. IEEE, 2011: 317-326.

[7] Zhu H, Chang S, Zhang W, et al. Online Vehicle Front–Rear Distance Estimation With Urban Context-Aware Trajectories[J]. IEEE Transactions on Vehicular Technology, 2018, 67(2): 1063-1074.

[8] Lee K, Yi Y, Jeong J, et al. Max-contribution: On optimal resource allocation in delay tolerant networks[C]//2010 Proceedings IEEE INFOCOM. IEEE, 2010: 1-9.

[9] Zhang L, Yu B, Pan J. GeoMob: A mobility-aware geocast scheme in metropolitans via taxicabs and buses[C]//IEEE INFOCOM 2014-IEEE Conference on Computer Communications. IEEE, 2014: 1279-1787.

[10] Chen K H, Dow C R, Chen S C, et al. HarpiaGrid: A Geography-Aware Grid-based Routing Protocol for Vehicular Ad Hoc Networks[J]. J. Inf. Sci. Eng., 2010, 26(3): 817-832.

[11] Nzouonta J, Rajgure N, Wang G, et al. VANET routing on city roads using real-time vehicular traffic information[J]. IEEE Transactions on Vehicular technology, 2009, 58(7): 3609-3626.

[12] Santini S, Salvi A, Valente A S, et al. A consensus-based approach for platooning with inter-vehicular communications[C]//2015 IEEE Conference on Computer Communications (INFOCOM). IEEE, 2015: 1158-1166.

[13] Naumov V, Gross T R. Connectivity-aware routing (CAR) in vehicular ad-hoc networks[C]//IEEE INFOCOM 2007-26th IEEE International Conference on Computer Communications. IEEE, 2007: 1919-1927.

[14] Stibor L, Zang Y, Reumerman H J. Evaluation of communication distance of broadcast messages in a vehicular ad-hoc network using IEEE 802.11 p[C]//2007 IEEE Wireless Communications and Networking Conference. IEEE, 2007: 254-257.

[15] Veeraraghavan M. Vehicular Networking[J]. Tutorial at IEEE ICC 2010, 2010.

[16] Kenney J B. Dedicated short-range communications (DSRC) standards in the United States[J]. Proceedings of the IEEE, 2011, 99(7): 1162-1182.

[17] Zheng Y, Liu Y, Yuan J, et al. Urban computing with taxicabs[C]//Proceedings of the 13th international conference on Ubiquitous computing. ACM, 2011: 89-98.

[18] Wang Y, Lv Z, Chen W, et al. Wi-Eye: Tracking Urban Private Vehicles with Inter-Vehicle Communications and Sparse Video Surveillance Cameras[C]//2018 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON). IEEE, 2018: 1-9.

[19] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473, 2014.