Exploiting Traveling Information for Data Forwarding in Community-Characterized Vehicular Networks

Zhong Li, Cheng Wang, Lu Shao, Chang-Jun Jiang, and Cheng-Xiang Wang

***Abstract*—In intelligent vehicular communication networks, a hybrid communication architecture is used which combines both centralized and ad hoc transmission schemes. In order to maximize the end-to-end delivery ratio while reducing the network overhead, one important problem is to efficiently design the data forwarding algorithm to guarantee the quality of data transmission. In this paper, by considering the traveling information and vehicular space-crossing community structure, two metrics, "space–time approachability" and "social approachability," are defined to provide the absolute and relative geographical information of the forthcoming contacts, respectively. Then, a novel data-forwarding algorithm, called approachability-based algorithm, is proposed, which utilizes two metrics together for better routing quality. We evaluate the proposed approachability-based algorithm utilizing San Francisco Cabspotting and Shanghai Taxi Movement datasets. Simulation results show that the approachability-based data forwarding algorithm can achieve better performance than the popular data forwarding algorithms ZOOM and BUBBLE RAP in all the interested scenarios.**

***Index Terms*—Data forwarding, social community, traveling information, vehicular networks.**

Manuscript received January 18, 2016; revised May 14, 2016 and September 13, 2016; accepted November 14, 2016. Date of publication November 29, 2016; date of current version July 14, 2017. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61602110, in part by the Shanghai Sailing Program under Grant 16YF1400200, in part by the National Science Foundation for Postdoctoral Scientists of China under Grant 2016M591575, in part by the Fundamental Research Funds for the Central Universities under Grant 2232016D-30, in part by the Fok Ying-Tong Education Foundation for Young Teachers in the Higher Education Institutions of China under Grant 151066, in part by the Shuguang Program from Shanghai Education Development Foundation under Grant 14SG20, in part by the Program for Young Excellent Talents in Tongji University under Grant 2013KJ048, in part by the Shanghai Key Basic Research Project under Grant JC1405800, in part by the Research Funds for Donghua Young Teachers 104-07-0053036, in part by the NSFC under Grant 71171045 and Grant 61472004, and in part by the International S&T Cooperation Program of Shanghai Science and Technology Commission under Grant 15220710600. The review of this paper was coordinated by Dr. J. Pan. *(Corresponding author: Chang-Jun Jiang.)*

Z. Li is with the College of Information Science and Technology, Donghua University, Shanghai 201620, China (e-mail: 007lizhong@gmail.com).

C. Wang, L. Shao, and C.-J. Jiang are with the Department of Computer Science, Tongji University, Shanghai 201804, China (e-mail: 3chengwang@gmail.com; 15275652962@163.com; cjjiang@tongji.edu.cn).

C.-X. Wang is with the Institute of Sensors, Signals and Systems, School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, U.K (e-mail: cheng-xiang.wang@hw.ac.uk).

This paper has supplementary downloadable multimedia material available at http://ieeexplore.ieee.org provided by the authors. The Supplementary Material contains the Appendix. This material is 2.14 MB in size.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2016.2633431

I. Introduction

Pushed by governments and car makers, vehicular communication networks have recently received increasing attention. In vehicular networks, one important problem is to efficiently design data forwarding algorithms to support data transmission tasks. On one hand, we rely on highly efficient transmissions by using centralized road side units (RSUs). On the other hand, we want to take advantage of moving vehicles to help data carrying in the physical space by using ad hoc inter vehicle communications, which are complementary to centralized RSU transmissions. Therefore, we carry on our data forwarding study under the hybrid communication architecture, i.e., centralized and ad hoc transmissions.

It is crucial to utilize various information obtained from a vehicular network to design an efficient data-forwarding scheme. Previous studies usually make use of various road information to predict the future forwarding direction to forward data, such as some geography- or trajectory-based methods [3], [6]–[12]. However, the computation on the road information/digital map has a very high cost since the knowledge of GPS, road segment, road length, and vehicle speed is required. Then, researchers turn to only using the vehicle social contact behaviors to predict the encounter probability to improve data-forwarding efficiency, such as some contact- or social-based methods [1], [2], [4], [13]–[18]. However, such methods do not sufficiently consider the differences between the vehicular networks and the common mobile networks in data forwarding. In vehicular networks, vehicles are moving on the physically constrained roads. The inter contact time between two vehicles is limited by the traffic condition. Moreover, the shortest communication paths do not always match the physical shortest paths [3]. Therefore, we can see the road information is still important for vehicles to forward data.

In fact, in daily life, a person has some information about at what time he/she will drive a car to some place. For example, a car will arrive at Hilton Hotel at 7:00 o'clock and will arrive at Pudong Airport at approximately 10:00 o'clock. We call this kind of information traveling information (i.e., the forthcoming position and the corresponding arrival time). With the help of the car navigator, given a destination, the traveling route can be calculated and adaptively adjusted according to the real-time traffic conditions. Therefore, the vehicle/driver can obtain some knowledge about the forthcoming traveling route information easily. The traveling information can reflect the road conditions to some extent.

In this work, combining the above *traveling information* and the *social information*, we design an efficient data-forwarding scheme under the hybrid communication architecture. In particular, the paper studies the *unicast* session among vehicles.

In this paper, first due to the stability of the road map, we define the traveling information (the position and the arrival time of a vehicle) as the absolute geographical information. Meanwhile, due to the dynamics of social behaviors, we define the social attribute of vehicles, i.e., community structure, as the relative geographical information. With respect to the traveling information, we give the definition of *space-time approachability* which is adopted to measure the probability of a vehicle approaching to the destination.

The space-time approachability is calculated based on all the shared traveling information when vehicles encounter each other. The larger space-time approachability a vehicle has, the faster the message will be carried to the destination. With respect to the vehicular social attribute, we use the community structure to characterize it in vehicular networks. From another perspective, we introduce social approachability to depict the probability of a vehicle approaching the destination.

Then, combining the space-time approachability with the social approachability, we design *an approachability-based data-forwarding algorithm* in vehicular networks. In particular, in our vehicular networks, RSUs do not cover the whole city area and some of the neighboring RSUs are interconnected via wired or wireless links, i.e., there are several RSU connected components. For different transmission phases (i.e., vehicle to vehicle, vehicle to RSU, or RSU to vehicle), the approachability-based measurement is utilized in different ways. In addition, in the proposed algorithm, RSUs do not passively send messages to all passing vehicles. According to the approachability-based measurement, they have "brains" to decide how to forward data.

At last, we extensively evaluate the approachability-based data forwarding algorithm on two different datasets: San Francisco Cabspotting and Shanghai Taxi Movement. The results show that the approachability-based data forwarding algorithm significantly outperforms several existing vehicular data-forwarding algorithms.

The rest of this paper is organized as follows. We review the related work in Section II. In Section III, we present the network model and some assumptions. In Section IV, we formalize the partially shared traveling information and introduce the space-crossing community detection method. In Section V, we define the space-time approachability and design an approachability-based algorithm to show how to use the partially shared traveling information in data forwarding. In Section VI, we introduce two experiment datasets and describe the issues of trace preprocessing, RSU deployment, and contact extraction. In Section VII, we conduct extensive experiments and analyze the results in detail. Finally, we conclude the paper in Section VIII.

II. Related Work

Some previous studies used the distance information calculated by GPS records to forward data, called *geography-based methods*. In these methods, the source vehicle transmits the message to one of its neighbors, in which the Euclidean GPS distance between the neighbor and the destination is smaller than that of other neighbors [6]. The same relay operation will be executed continuously until the message arrives at the destination. We can see this kind of method may result in a problem such as a dead-end road [19], since the shortest Euclidean communication paths do not always match the shortest physical road paths.

Therefore, in order to overcome this disadvantage, some improved geography-based data-forwarding schemes emerge, called *trajectory-based methods*. In these methods, besides the GPS information, other road information is added, such as the road intersection [7]–[9], the road length [3], [10], the destination direction [11], the traffic density, digital map [12], etc. The reader can refer to literature [21], [22]. Because the destination is always moving, in fact, we can see this kind of methods is a course of finding the shortest dynamic path from the source to the destination by using various kinds of information, just like the car navigator. Meanwhile, besides the path programming, most of these methods use the path-reversing scheme to guarantee the correct routing path in their algorithms. Thus, they need to store the passing roads and the relay nodes in the road intersection devices. Although we could not precisely estimate the quantity of the information and the storage they used, intuitively, we can see that the complexity of the trajectory-based methods is high.

Recently, people were enlightened by the node contact relationship and proposed some *contact-based methods*. In Prophet [13], each node maintains the encounter history. The routing decision is made based on the encounter probability. In OIA [4], the authors give a study of infrastructure/RSU-assisted routing in vehicular networks by using a Markov chain to predict the future encounter probability. In particular, they consider the problem of buffer limits. In GeoMob [14], the authors study the macroscopic mobility pattern and the microscopic mobility pattern to predict the future region where a vehicle will go. The message is relayed to a vehicle, which is most likely to go to the destination region. In addition, a storage-friendly routing scheme, called RENA [15], is proposed to improve the data-forwarding efficiency in vehicular networks. RENA is based on regional movement history with avoiding excessive storage for tracking encounter history.

Based on the above device contacts, researchers find that vehicle moving behaviors have some social characteristics that can be explored, such as community structure and node centrality. So, some more advanced methods emerge by exploiting the social relationship among vehicles, called *social-based methods* [1], [2], [16]–[18]. For example, the representative work ZOOM [2] is the first paper that uses the contact-level mobility (obtained by a Markov chain) together with the social-level mobility (obtained by calculating the social ego centrality) to forward data. The paper demonstrates that capturing the social-level mobility as a complementary counterpart of the contact-level priors can significantly improve the performance of the opportunistic data forwarding.

III. System Model and Assumptions

In this paper, we model the underlying network with infrastructure support as a *dynamic graph*, which can be defined as a time sequence of network graph, denoted by $\mathcal{G} = \{G_0, G_1, \ldots, G_t, \ldots\}$, where $G_t = (V_t, E_t)$ represents a time-dependent network snapshot recorded at time t, V_t denotes the set of nodes, including the set of vehicles and the set of stationary RSUs, and $E_t = \{(u, v)|u, v \in V_t\}$ denotes the edge set. Both node and edge sets can change over time. The edges in the network are aggregated by using the following *median-based sliding window mechanism*.

Let $l(u, v, t) = 1$ denote the start of a contact between node u and v at time t $(0 \le t < \infty)$. Then, we have $\sum_{t=t_{\rm now}-\Delta}^{t_{\rm now}} l(u, v, t)$ denote the overall numbers of contacts between node u and v from time $t_{\rm now} - \Delta$ to $t_{\rm now}$ and have $\sum_{t=t_{\rm now}-\Delta}^{t_{\rm now}} l_*(t)$ denote the overall numbers of contacts for all nodes from time $t_{\rm now} - \Delta$ to $t_{\rm now}$, $0 < \Delta < t_{\rm now}$. The window length Δ of the sliding window mechanism is usually empirically determined [1], [23]. We take the length of sliding window Δ as a constant, not a variable. We have Δ equal 6×3600 s, since the comparison algorithms [1], [2] used in our experiment are all set this value.

We define the *encounter ratio* between node u and v at current time $t_{\rm now}$ as

$$e(u, v, t_{\rm now}) = \frac{\sum_{t=t_{\rm now}-\Delta}^{t_{\rm now}} l(u, v, t)}{\sum_{t=t_{\rm now}-\Delta}^{t_{\rm now}} l_*(t)}.$$

Note that we assume $e(u, v, t_{\rm now}) = e(v, u, t_{\rm now})$ by assigning the larger value to the other. In the following sections, for brevity, we denote the current time $t_{\rm now}$ by t, without confusion.

In the median-based sliding window mechanism, according to the encounter ratios between any two nodes, we filter the edge between two nodes if the encounter ratio value of the two nodes is below the median of encounter ratio values among all nodes.

IV. Traveling Information Formalization in Community-Characterized Vehicular Networks

A. Partially Shared Traveling Information

Assume that the vehicle shares some traveling route information when it meets another vehicle. Let sequence $\mathcal{P}_u^k = \{< p_u(1), t_u(1) >, < p_u(2), t_u(2) >, \ldots, < p_u(i), t_u(i) >, \ldots, < p_u(k), t_u(k) >\}$ stand for the number of k partially shared traveling information of vehicle u at an encounter, where $p_u(i)$ denotes the ith key position that the vehicle will pass by in the future, and $t_u(i)$ denotes the corresponding arrival time at ith key position, $1 \le i \le k, k \in N^+$. The value of k is obtained through the following steps. First, for any source–destination pairs, we assume that there are L positions selected randomly on the basis of the traveling route. In the experiment, we set L as a constant equaling 10. Other values are also allowed. We vary the degree of sharing to see the different simulation results. Second, a driver can choose the degree of sharing, defined by a parameter $\alpha \in [0, 1]$, to release their traveling information, in which α means the sharing probability of one position in L positions. Note that, if a vehicle has the strong privacy demand, it can choose not to open its traveling route information totally. The partially shared traveling information can be seen as the absolute geographical trajectory, which is useful for data forwarding in vehicular networks. In the course of traveling information sharing, the problems of privacy preservation and incentive mechanism will be further studied in our future work.

B. Space-Crossing Community Structure

In vehicular networks, due to daily activities of drivers, vehicle movement behaviors show a characteristic of cluster or community. This social attribute reflects the law of object interactions in the underlying network. The social community can be seen as the relative geographical information, which has been proved useful in data forwarding in opportunistic networks [16], [17].

In different application scenarios, we can obtain different community structures through different community detection methods (see the recent review papers by Fortunato [24] and Lancichinetti *et al.* [25]). In our paper, we have three demands when doing the community detection. First, the detection method can be used in a dynamic environment. Second, the detection method has the ability of handling the hybrid underlying network with infrastructure support. Third, the detection results do help the data forwarding in opportunistic networks and can reflect the positive role of infrastructures. Based on above three requirements, we choose *space-crossing community detection* [20] to tackle our vehicular network. In the vehicular network, the *space-crossing community* is a subgraph. It comprises of moving vehicles and stationary RSUs that frequently transmit data to each other than to other devices. The internal edges in the space-crossing community represent the communication connectivity capability among these devices. The capability includes the *directly* frequent interactions among infrastructures-vehicles or vehicles-vehicles and the *indirectly* frequent interactions among vehicles–vehicles through some infrastructures.

In particular, many previous community detection methods simply put the infrastructures and the vehicles on an equal footing. These methods cannot reflect the true communication connectivity capability among those long-distance nodes in far areas through infrastructures. However, the space-crossing community detection method does well in this point. We provide a detailed description of the space-crossing community detection method in Appendix A. The Appendix file can be found in literature [29].

V. Approachability-Based Data Forwarding in Vehicular Networks

In this section, we will present how to use the shared traveling information to improve the data forwarding in community-characterized vehicular networks.

We give a concept of approachability. The approachability is defined for a node/vehicle. It reflects the capability of a vehicle to advance the messages toward the destination. First, in Section V-A, we give a definition of *space-time approachability* of a vehicle by using the partially shared travelling information.

Second, in Section V-B, we calculate the *social approachability* of a vehicle by using the space-crossing communities detected in vehicular networks. The space-time approachability and the social approachability both predict a vehicle's future movement and describe the capability of a vehicle to advance the messages towards the destination from two different perspectives. Finally, we use the product of both to form the approachability measurement and design an *approachability-based data forwarding algorithm* in vehicular networks in Section V-D. Using the approachability measurement, if the approachability of the message holder is lower than that of its encounter node, the algorithm will let the message holder transmit the data to the encounter node since the encounter node has stronger ability to forward the data to the destination than the current message holder.

A. Space-Time Approachability

Definition 1 (approximate destination set): Assuming that, for a destination vehicle d, from last Δ length of sliding window to current time t, vehicle d has passed several RSUs. Then, for every passed RSU r, if $e(r,d,t)$ is larger than the median of $\{e(r',d,t)|r' \in \{\text{all the passed RSUs by vehicle } d\}$ and $e(r',d,t) \neq 0\}$, we will put RSU r into a new set, called approximate destination set $R_t(d)$.

Definition 2 (RSU importance): Let $CC_t(r)$ denote the RSU connected component containing RSU r. Based on Definition 1, we define RSU importance for every $r \in R_t(d)$ as $w(r,d,t)$.

1) If $|R_t(d) \bigcap CC_t(r)| = 1$, we will have

$$w(r,d,t) = e(r,d,t).$$

2) If $|R_t(d) \bigcap CC_t(r)| > 1$, we will have

$$w(r,d,t) = \sum\nolimits_{r' \in R_t(d) \bigcap CC_t(r)} e(r',d,t)$$

where $e(r,d,t)$ and $e(r',d,t)$ denote the encounter ratio between $r-d$ and $r'-d$ at time t, respectively. For $r \notin R_t(d)$, we have $w(r,d,t) = 0$.

Definition 3 (space-time approachability): For a vehicle u, it shares its k partial traveling route information $\mathcal{P}_u^k = \{< p_u(1), t_u(1) >, < p_u(2), t_u(2) >, \ldots, < p_u(i), t_u(i) >, \ldots, < p_u(k), t_u(k) >\}$ at an encounter, as described in Section IV-A. Taking the ith position information $p_u(i)$ as a center point, we can find a set $I_u(i)$ which denotes the nearby RSUs within the vehicle-RSU communication range. Therefore, based on approximate destination set $R_t(d)$, we let $R_t(d) \bigcap I_u(i)$ denote the set of proximate destinations (RSUs) approached by vehicle u in its ith shared position.

We define the space-time approachability for vehicle u whose message destination is vehicle d as $\text{TSA}(u,d)$, having

$$\text{TSA}_t(u,d) = \sum_{i=1}^{k} \frac{\sum_{r \in R_t(d) \bigcap I_u(i)} w(r,d,t)}{t_u(i) - t}$$

where $w(r,d,t)$ denotes the RSU importance of r at current time t with the related destination d, and we assume that $t_u(i) - t > 0$. A larger TSA means the vehicle will approach the destination set using the high contact frequency (described in the numerator) and the short time (described in the denominator). Note that, we allow $\text{TSA}_t(u,d) = 0$, if vehicle u rejects to share traveling route information with the encountered vehicles.

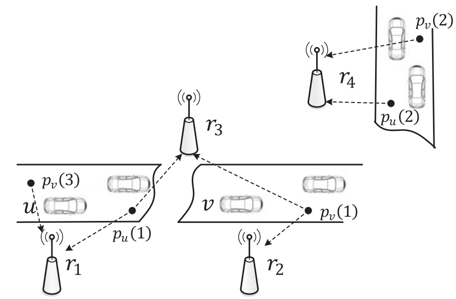


Fig. 1. Example of calculating space-time approachability.

Here, we give an example of calculating space-time approachability in Fig. 1. The black dot represents the shared position by each vehicle. The dashed arrow points to the communication RSU(s) associated with the shared position. Assume that there are two vehicles. The vehicle u shares its partial traveling route information as $\{< p_u(1), t_u(1) >, < p_u(2), t_u(2) >\}$, and the vehicle v shares its partial traveling route information as $\{< p_v(1), t_v(1) >, < p_v(2), t_v(2) >, < p_v(3), t_v(3) >\}$. RSUs r_1, r_2, r_3, r_4 are in the approximate destination set $R_t(d)$. Therefore, we have

$\text{TSA}_t(u,d) = \frac{w(r_1,d,t)}{t_u(1)-t} + \frac{w(r_3,d,t)}{t_u(1)-t} + \frac{w(r_4,d,t)}{t_u(2)-t}$

$\text{TSA}_t(v,d) = \frac{w(r_2,d,t)}{t_v(1)-t} + \frac{w(r_3,d,t)}{t_v(1)-t} + \frac{w(r_4,d,t)}{t_v(2)-t} + \frac{w(r_1,d,t)}{t_v(3)-t}.$

B. Social Approachability

Definition 4 (local activity): Case I: For moving vehicles:

Let $\text{lav}_t(i,u)$ denote the local activity of a moving vehicle u in ith space-crossing community $SC_t(i)$ at time t. Then, we have

$$\text{lav}_t(i,u) = \frac{\sum_{(u,v) \in SC_t(i)} e(u,v,t)}{\sum_{(v',v'') \in SC_t(i)} e(v',v'',t)}, 1 \leq i \leq q$$

where v' and v'' are any two nodes in $SC_t(i)$; $e(u,v,t)$ denotes the encounter ratio between node u and v at current time t; q represents the number of space-crossing communities. The numerator represents the sum of the encounter ratio between node u and other nodes in community $SC_t(i)$ and the denominator represents the sum of the encounter ratio between any two nodes in community $SC_t(i)$.

Case II: For stationary RSUs

Let $CC_t(i,j)$ denote the jth connected component of RSUs in ith space-crossing community $SC_t(i)$. We define the local activity of every RSU $r \in CC_t(i,j)$ as

$$\text{lap}_t(i,r) = \sum\nolimits_{r' \in CC_t(i,j)} \text{lav}_t(i,r')$$

where, for every RSU $r' \in CC_t(i,j)$, $\text{lav}_t(i,r')$ is obtained according to the method provided in *Case I*, with treating RSUs as ordinary vehicles.

The local activity of a node (including the moving vehicle and the stationary RSU) can represent the importance of a node in a certain space-crossing community. A larger local activity means that the node has more interactions with other members in the community.

Definition 5 (activity vector): For each moving vehicle u, we define the activity vector at time t as

$$A_t(u) = (\mathrm{lav}_t(1,u), \mathrm{lav}_t(2,u), \ldots, \mathrm{lav}_t(i,u), \ldots, \mathrm{lav}_t(q,u)).$$

For each RSU r, we define the activity vector at time t as

$$A_t(r) = (\mathrm{lap}_t(1,r), \mathrm{lap}_t(2,r), \ldots, \mathrm{lap}_t(i,r), \ldots, \mathrm{lap}_t(q,r))$$

where $\mathrm{lav}_t(i,u)$ and $\mathrm{lap}_t(i,r)$ denote the local activity of moving vehicle u and RSU r, respectively in space-crossing community $SC_t(i)$ at time t. q represents the number of communities after applying the space-crossing community detection method.

Definition 6 (social approachability [20]): Given two activity vectors $A_t(u)$ of node u and $A_t(v)$ of node v, we define social approachability between u and v at time t as $SA_t(u,v)$, having

$$SA_t(u,v) = A_t(u) \cdot A_t(v)$$

where the operator $\cdot$ denotes the inner product of vectors.

On the one hand, a node having a larger social approachability with the destination can guarantee this node has similar distribution of the belonging communities with the destination. On the other hand, it can guarantee this node has larger local activity in the activity vector correspondingly. Together, a node having a larger social approachability with the destination indicates that it has higher chance to approach the destination.

C. Approachability-Based Measurement

Definition 7 (approachability-based measurement): Assume that there is a session from source node u to destination node d, and now node u meets another node v. Then, we define a *fair* approachability-based measurement $M_t(u,d)$ and $M_t(v,d)$ for node u and node v at time t.

1) If $\mathrm{TSA}_t(u,d) \neq 0$, $SA_t(u,d) \neq 0$, $\mathrm{TSA}_t(v,d) \neq 0$, and $SA_t(v,d) \neq 0$, we will have $M_t(u,d) = \mathrm{TSA}_t(u,d) \times SA_t(u,d)$ and $M_t(v,d) = \mathrm{TSA}_t(v,d) \times SA_t(v,d)$.
2) If $\mathrm{TSA}_t(u,d) \times \mathrm{TSA}_t(v,d) = 0$, i.e., the vehicle(s) does(do) not like to share traveling route information, we will have $M_t(u,d) = SA_t(u,d)$ and $M_t(v,d) = SA_t(v,d)$.
3) If $SA_t(u,d) \times SA_t(v,d) = 0$, i.e., the social approachability between node u/v and destination d is totally irrelevant, we will have $M_t(u,d) = \mathrm{TSA}_t(u,d)$ and $M_t(v,d) = \mathrm{TSA}_t(v,d)$.
4) For other cases, we have $M_t(u,d) = M_t(v,d) = 0$.

D. Approachability-Based Data Forwarding Algorithm in Different Phases

Phase 1: Vehicle → Vehicle

When a vehicle holds a message, it will try to send the message to a moving vehicle with a larger approachability-based measurement than itself and will let the vehicle send the message to the destination consecutively. If the encountered vehicle has a smaller approachability-based measurement than the message holder, the message holder will inquire the current neighbors of the encountered vehicle. If the set of current neighbors contains the destination, then the message will be transmitted to the encountered vehicle.

Phase 2: Vehicle → RSU

When a vehicle holds a message and it meets an RSU, first the vehicle will detect whether the RSU is in the approximate destination set $R_t(d)$. If yes, the message is delivered to the RSU directly. If no, the vehicle will transmit the message to the encountered RSU with a larger approachability-based measurement than the vehicle itself. Otherwise, the message holder will inquire the current neighbors of the encountered RSU. If the set of current neighbors contains the destination, the message will be transmitted to the encountered RSU.

Phase 3: RSU → Vehicle

When a RSU holds a message, it first delivers the message to other RSUs in its common connected components. Then, the RSUs will deliver the message to the passing vehicles with larger approachability-based measurement than RSUs themselves. Otherwise, the message holder will inquire the current neighbors of the encountered vehicle. If the set of current neighbors contains the destination, then the message will be transmitted to the encounter vehicle.

In the above three phases, after the message holder transmits the packet to the encountered node, the holder removes the packet from its buffer.

Note that the implementation of the approachability-based data forwarding algorithm in the hybrid communication structure is provided in Appendix B.

VI. Vehicle Trace Data Analysis

We use two large GPS-based vehicle mobility traces to evaluate the efficiency of our algorithm. One is San Francisco Cabspotting in America, and the other is Shanghai Taxi Movement in China.

A. Datasets

1) San Francisco Cabspotting: The data can be downloaded from [26]. The data contains GPS coordinates of 536 taxis collected over a period of three consecutive weeks in the San Francisco Bay Area. Each taxi is equipped with a GPS receiver and sends a location update (timestamp, identifier, geo-coordinates) periodically. The location updates are quite fine-grained. The average time interval between two consecutive location updates is less than 60 s.

2) Shanghai Taxi Movement: Partial data can be downloaded from [27]. The dataset was collected by our research group in Shanghai, China, which was approximately 3000 taxis from January to September, 2006. The taxi periodically sends reports back to the data collector via an onboard GPS-enabled device. Each taxi reports every 60 s. The information in the dataset includes vehicle ID, location coordinates, timestamp, onboard, vehicle moving speed, and heading direction.

Fig. 2. RSUs deployed uniformly in San Francisco.

Fig. 3. RSUs deployed in dense traffic area in Shanghai.

B. Trace Preprocessing

GPS accuracy might be affected by many factors. In above two vehicle datasets, there are some noisy data, e.g., the locations of the data are wrong or fall outside of the map bounds. So, a noise suppression method, which uses map polygon clipping, is executed to filter the noisy data on the vehicle datasets. The results of the trace preprocessing on two datasets are provided in Appendix C.

C. RSU Deployment

In order to sufficiently evaluate our algorithm in different scenarios, in the paper, we discuss three configurations of RSU deployment (70 RSUs for San Francisco Cabspotting and 352 RSUs for Shanghai Taxi Movement). The first configuration has RSU locations selected in dense vehicle traffic area. The second configuration has RSU locations selected in sparse vehicle traffic area. The third configuration has RSU locations selected uniformly in the vehicular network. There are two illustrations of RSU deployment, shown in Figs. 2 and 3, with red triangles representing RSUs. Note that, efficiently deploying RSUs is crucial to improve packet forwarding efficiency in vehicular networks. For different network scenarios and research purposes, the optimal RSU deployment strategies are different. However, this problem is not the main focus of our paper.

Then based on above settings, we execute contact extraction to form community structure. The detailed procedure and the captured space-crossing communities are provided in Appendix D.

VII. Evaluation

A. Simulation Setup

We use an open-source simulator "The ONE" [28] for simulation. We import the vehicle mobility traces and the road map to simulate the node mobility. The source and destination pairs are chosen randomly among all vehicles. Each simulation is repeated 20 times with different random seeds. Besides the parameters described in Section VI-A, the other simulation settings are summarized in Table I. Note that, in real world, the transmission speed, transmission range, and buffer size of the RSUs are larger than our settings. However, in order to clearly test the tiny performance variations, we set the values as described in Table I. Since all the comparison algorithms use the same settings, it does not impact our simulation results.

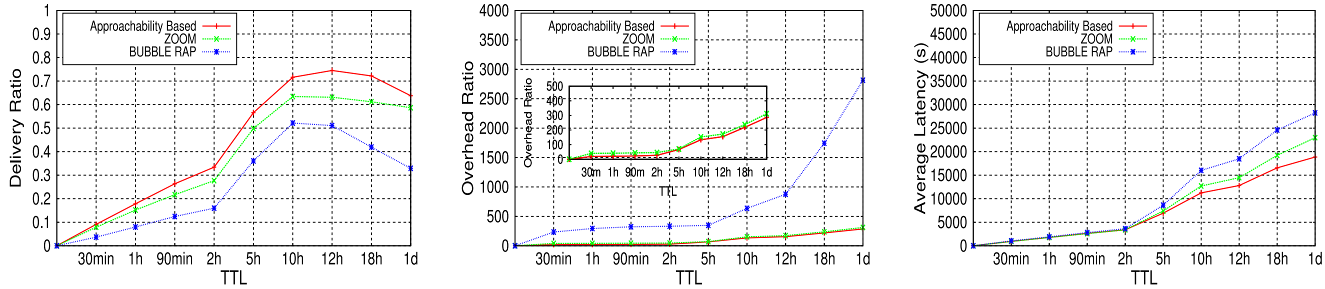
TABLE I
SIMULATION SETTINGS

Parameter	Settings
Vehicle transmission range	100 m
RSU transmission range	300 m
V2V transmission speed	250 Kb/s
V2I transmission speed	1 Mb/s
Vehicle buffer size	5 MB
RSU buffer size	5 MB
Packet generation interval	200–300 s randomly
Packet size	50–100 KB randomly
Sliding window size	6×3600s
Degree of sharing α	Cabspotting:0.6, Shanghai: 0.5
Simulation time	Cabspotting: 2071531s, Shanghai: 2073599s

B. Experiment Results and Analysis

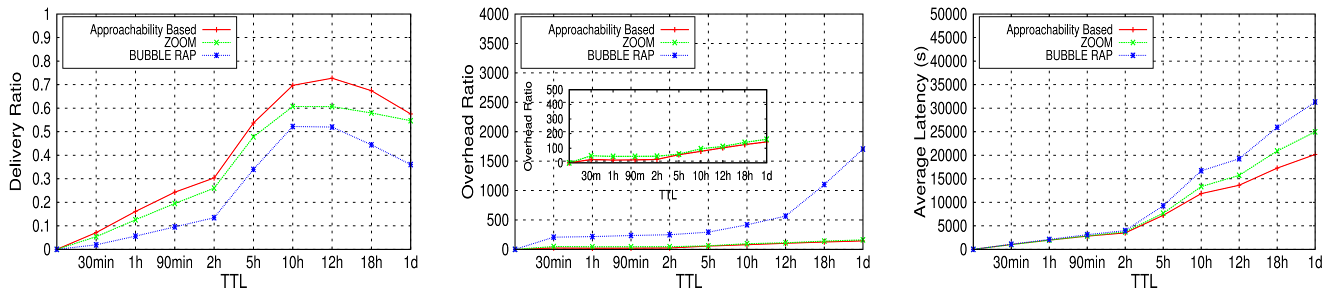
1) Comparison Fairness: In this section, we compare our algorithm with two popular social-based data forwarding algorithms (BUBBLE RAP [1] and ZOOM [2]) in vehicular networks. The descriptions of the two algorithms are provided in Appendix E.

Note that, for the sake of fairness, we select settings or parameters which bring about the best performances for above two comparison algorithms. *Additionally, since the comparison algorithms are not based on the underlying network with infrastructure support, we use the fair infrastructure strategy (spreading the messages in RSU connected component) for above*



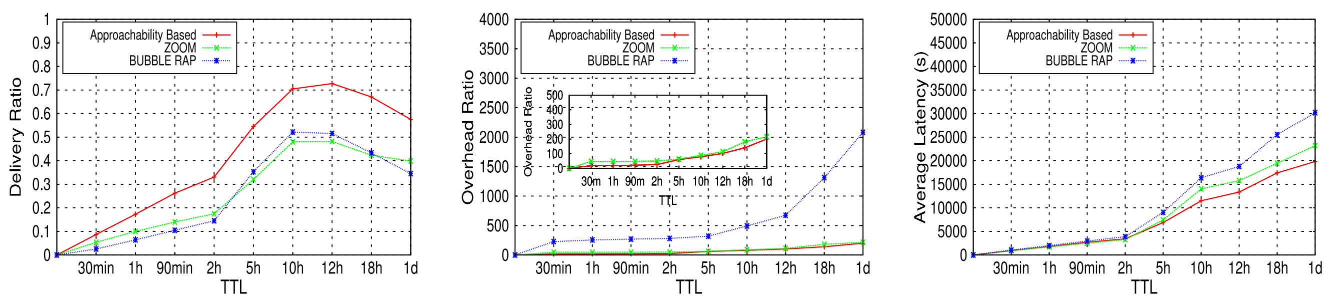
(a) (b) (c)

Fig. 4. Simulation results of (a) delivery ratio, (b) overhead ratio, and (c) average latency on San Francisco Cabspotting with RSUs being deployed in a dense traffic area.



(a) (b) (c)

Fig. 5. Simulation results of (a) delivery ratio, (b) overhead ratio, and (c) average latency on San Francisco Cabspotting with RSUs being deployed in a sparse traffic area.



(a) (b) (c)

Fig. 6. Simulation results of (a) delivery ratio, (b) overhead ratio, and (c) average latency on San Francisco Cabspotting with RSUs being deployed uniformly.

comparison algorithms. That is to say, ZOOM and BUBBLE RAP are also be applied in the environment with RSU support. ZOOM and BUBBLE RAP both have the same copies strategy on RSUs with our approachability-based algorithm.

2) Metrics: The performance of the proposed approachability-based algorithm is evaluated in the following metrics.

1) Delivery ratio: the ratio of *the number of successfully delivered messages* to *the total number of created messages*.
2) Average latency: the average message delay for all the successful sessions.
3) Overhead ratio: the proportion of *the difference between the number of relayed messages and successfully delivered messages* out of *the successfully delivered messages*.

3) General Comparison Experiment: From Figs. 4 to 9, we show the delivery ratio, overhead ratio, and average latency of our approachability-based algorithm, ZOOM, and BUBBLE RAP in San Francisco Cabspotting and Shanghai Taxi Movement for three kinds of RSU deployment. We can see, on average, the delivery ratio of approachability-based algorithm achieves best among these compared algorithms, and meanwhile the overhead ratio and average latency are lowest.

In terms of delivery ratio, for all kinds of scenarios, the peak value of our approachability-based algorithm is higher than ZOOM and BUBBLE RAP. The emerging time of the peak value of approachability-based algorithm is later than ZOOM and BUBBLE RAP. These phenomena all demonstrate the better performance of our algorithm. It is normal that the delivery ratios of the three algorithms all decrease as the time to live (TTL) increases. This is because the network capacity is limited and the buffer overflows the excessive messages. Note that,

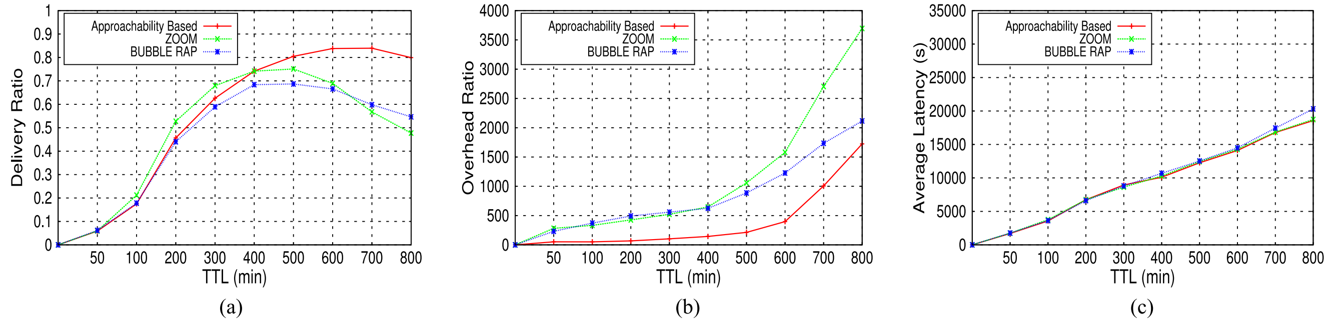


Fig. 7. Simulation results of (a) delivery ratio, (b) overhead ratio, and (c) average latency on Shanghai Taxi Movement with RSUs being deployed in a dense traffic area.

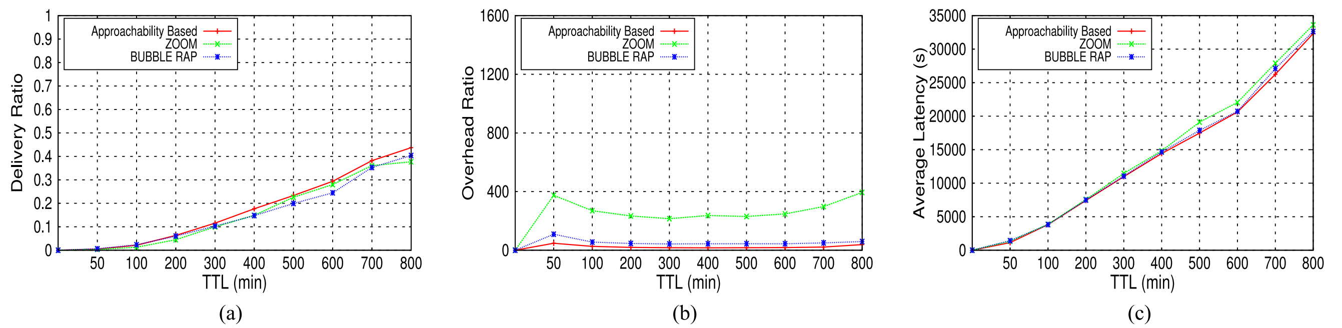


Fig. 8. Simulation results of (a) delivery ratio, (b) overhead ratio, and (c) average latency on Shanghai Taxi Movement with RSUs being deployed in a sparse traffic area.

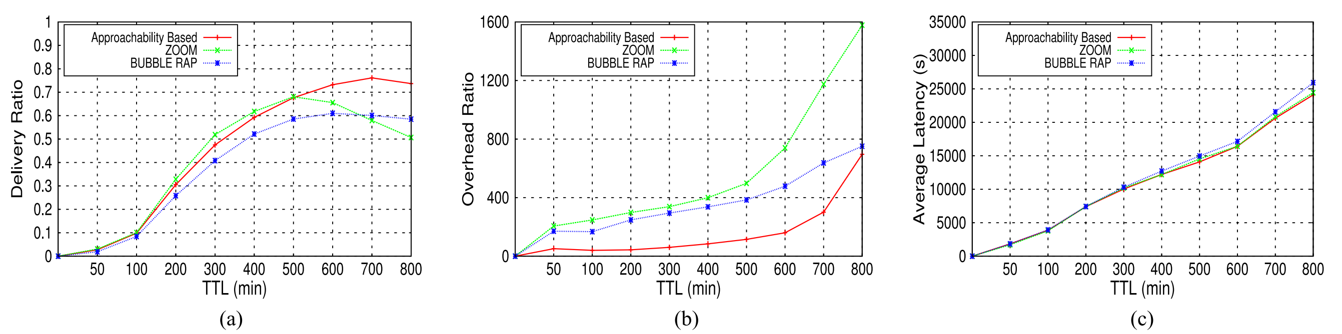


Fig. 9. Simulation results of (a) delivery ratio, (b) overhead ratio, and (c) average latency on Shanghai Taxi Movement with RSUs being deployed uniformly.

in Shanghai Taxi Movement [Figs. 7(a) and 9(a)], the delivery ratio of ZOOM is slighter higher than approachability-based algorithm between TTL 100 and 400 min, when the RSU locations are selected in dense traffic area and selected uniformly. However, the approachability-based algorithm still respectively outperforms ZOOM with 12.74% and 10.97% on average.

In terms of overhead ratio and average latency, the approachability-based algorithm keeps a low overhead ratio and a low average latency than the compared algorithms in both datasets. Note that in San Francisco Cabspotting, the overhead of ZOOM is smaller than BUBBLE RAP, but in Shanghai Taxi Movement, the opposite is true. This is because the vehicle contacts in Shanghai Taxi Movement are denser than San Francisco Cabspotting, which results in ZOOM, who uses the contact-level mobility in its first phase of data forwarding, generating more relays. All in all, together with the delivery ratio, we can see, the approachability-based algorithm does not use a *too long* delay to exchange for a good delivery ratio. Here, in Table II, we list the detailed performance contributions in a quantifiable way. The percents are the gains of our approachability-based algorithm compared with ZOOM and BUBBLE RAP. The data *are averaged on* the three kinds of RSU deployment scenarios.

Here, we give the detailed analysis based on above simulation results. In BUBBLE RAP, it uses the global betweenness as the relay criterion in its first phase of data forwarding. If it delivers the message to a node with a high global betweenness, although it indeed has high contact frequency with other nodes with respect to the entire network, it may be in a community which is irrelevant to or does not overlap with the destination community. As a relay, this node may increase the time of

TABLE II
PERFORMANCE CONTRIBUTIONS

Scenarios / Metrics	Delivery Ratio	Overhead Ratio	Average Latency
ZOOM in San Francisco Cabspotting	30.81%	31.27%	11.04%
BUBBLE RAP in San Francisco Cabspotting	64.67%	90.52%	27.84%
ZOOM in Shanghai Taxi Movement	12.45%	71.16%	4.09%
BUBBLE RAP in Shanghai Taxi Movement	18.61%	45.08%	4.42%

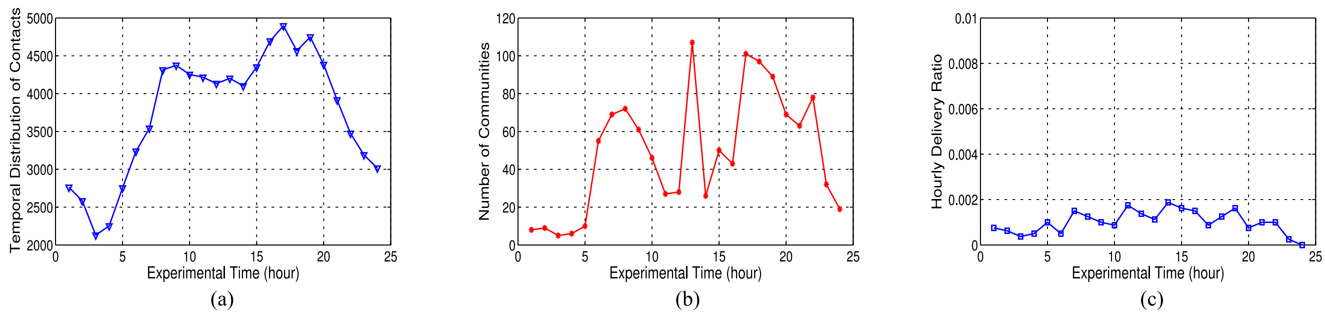


Fig. 10. Rush and nonrush hours tests of (a) distribution of contacts, (b) number of communities, and (c) hourly delivery ratio on San Francisco Cabspotting.

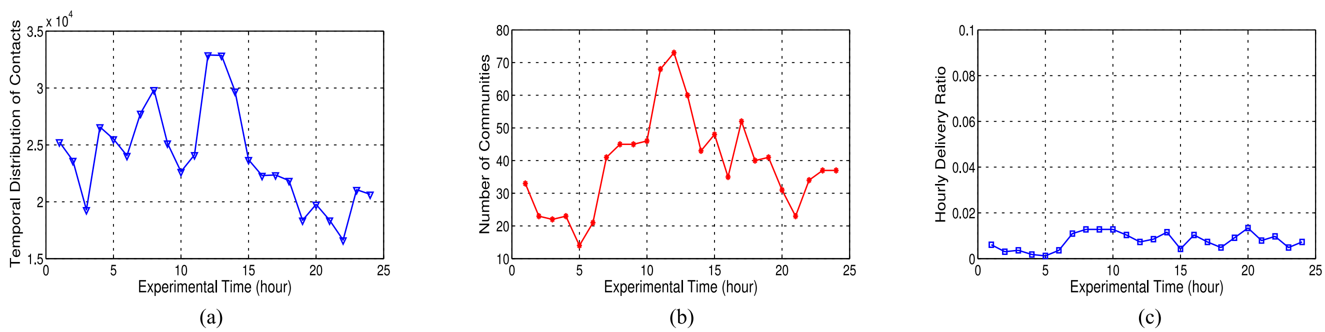


Fig. 11. Rush and nonrush hours tests of (a) distribution of contacts, (b) number of communities, and (c) hourly delivery ratio on Shanghai Taxi Movement.

reaching the destination. In ZOOM, it uses Markov chain to predict the next contact time based on the previous inter contact time between nodes. The message holder delivers the message to a node with the minimal next contact time with the destination. We can see the Markov-based method puts much emphasis on the nodes that are required to directly contact the destination in the future. The method is prone to losing many potential useful contacts which may *not* be *directly* linked with the destination. It is the key difference between the Markov-based method and our method. Our method concerns the probability of the next relay approaching to the destination, not the specific contact, which leads to a wider range of contacts than the Markov-based method.

4) Rush and Nonrush Hours Comparison Experiment: Especially, we do experiments to verify the effectiveness of the partially shared traveling information independently. In vehicle datasets, we observe a phenomenon: the traffic is different in different time period of a day, i.e., existing the rush hours and nonrush hours, as illustrated in Figs. 10(a) and 11(a).

In Figs. 10(a) and 11(a), we show the distribution of the number of contacts captured in San Francisco Cabspotting (26th May, 2008) and Shanghai Taxi Movement (5th August, 2006). In Figs. 10(b) and 11(b), we show the distribution of the number of space-crossing communities in San Francisco Cabspotting (26th May, 2008) and Shanghai Taxi Movement (5th August, 2006). We can see the time periods of 8:00–9:00, 11:00–13:00, and 16:00–20:00 are the rush hours approximately. The number of contacts in rush hours is more than that in nonrush hours. Similarly, the distribution of the number of space-crossing communities also shows the characteristic of the rush and nonrush hours.

In the rush hours, the social communities and the partially shared traveling information all work in data forwarding. In the nonrush hours, due to the decrease of social communities, the data forwarding mainly relies on the shared traveling information. Figs. 10(c) and 11(c) are the results of delivery ratio, which are evaluated on the rush and nonrush hours in San Francisco Cabspotting (26th May, 2008) and Shanghai Taxi Movement

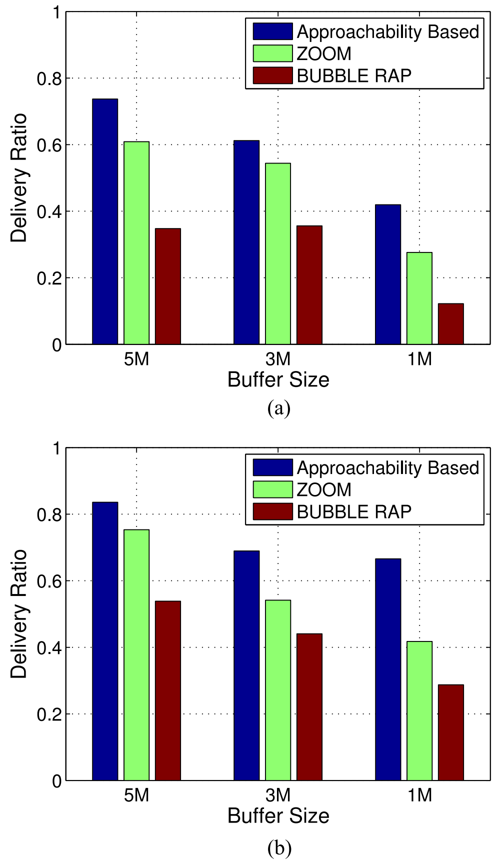


Fig. 12. Impact of buffer size on delivery ratio for (a) test on San Francisco Cabspotting and (b) test on Shanghai Taxi Movement.

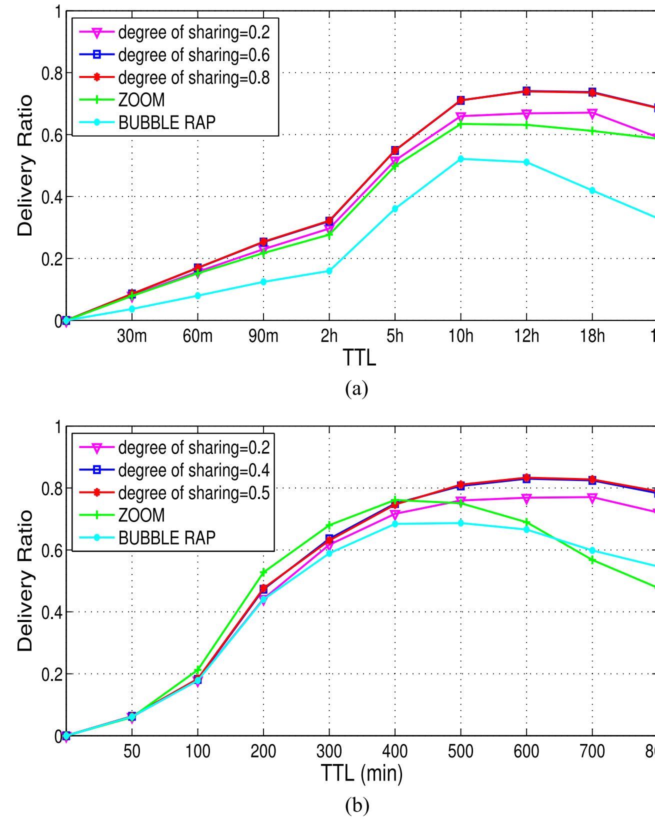


Fig. 13. Impact of degree of sharing on delivery ratio for (a) test on San Francisco Cabspotting and (b) test on Shanghai Taxi Movement.

(5th August, 2006). In order to avoid the cumulative effect, each hour is independently tested. From the results, we can see clearly that the delivery ratio do not fluctuate largely on rush and nonrush hours from time 5:00–20:00. This can demonstrate that the shared traveling information indeed helps the data forwarding in vehicular networks. For time periods of 0:00–5:00 and 20:00–24:00, since the number of contacts is few, the social communities and the shared traveling information both play a little role in data forwarding.

5) Different Parameters Comparison Experiments: For further performance study, we investigate the impact of buffer size and degree of sharing. We see how the performance of data forwarding reacts to these parameters in vehicular networks.

The RSU locations are selected in the dense traffic area. Due to space limit, we omit other cases. First, in San Francisco Cabspotting, setting TTL as 1080 s, we vary the buffer size from 5 to 1 MB. The results are shown in Fig. 12(a). When the buffer size decreases, all algorithms have low packet delivery ratio. However, at every freezing value of buffer size, our algorithm achieves better performance compared with ZOOM and BUBBLE RAP. It demonstrates the advantage of our algorithm. In Shanghai Taxi Movement, setting TTL as 600 min, we also vary the buffer size from 5 to 1 MB. The results are shown in Fig. 12(b). The similar trend also appears in Shanghai Taxi Movement.

Second, in San Francisco Cabspotting, we vary the value of degree of sharing α from 0.2, 0.6 to 0.8. In Shanghai Taxi Movement, we vary the value of degree of sharing α from 0.2, 0.4 to 0.5. The results are shown in Fig. 13(a) and (b). The delivery ratio of approachability-based Algorithm increases with the increasing value of degree of sharing. Since the more shared locations will provide more information about whether the vehicle will pass the destination in a short time. However, this increasing tendency will not always continue. See the curves labeled the degree of sharing α equaling 0.6 and α equaling 0.8 in San Francisco Cabspotting, and the curves labeled the degree of sharing α equaling 0.4 and α equaling 0.5 in Shanghai Taxi Movement. The curves respectively stated above are almost coincident. So we obtain, for different datasets, the degree of sharing has an upper bound. The bound can guarantee that the vehicle owns enough traveling information (compared with the whole map) to decide a good relay. Moreover, for Shanghai Taxi Movement, since the contacts in it are denser than San Francisco Cabspotting, a lower degree of sharing can satisfy the data forwarding requirement.

VIII. Conclusion

In this paper, we have investigated how to use a vehicle's traveling information to forward data in social community-characterized vehicular networks. We have proposed a very efficient approachability-based data-forwarding algorithm for vehicular networks under the hybrid communication architecture. Through comprehensive simulations, we have demonstrated that our algorithm outperforms some widely used vehicular data forwarding algorithms such as ZOOM and BUBBLE RAP. In terms of delivery ratio, our approachability-based algorithm is better than ZOOM with 30.81% and better than BUBBLE RAP with 64.67% on average using the San Francisco Cabspotting dataset. Using the Shanghai Taxi Movement dataset, the proposed algorithm is better than ZOOM with 12.45% and better than BUBBLE RAP with 18.61% on average. More importantly, the proposed algorithm also achieves better performance than ZOOM and BUBBLE RAP in terms of both the overhead ratio and latency. For our future work, we will further explore the issues of privacy preserving and incentive mechanism when sharing traveling information in vehicular networks.

Acknowledgment

The authors would like to thank the anonymous reviewers for their constructive comments.

References

[1] P. Hui, J. Crowcroft, and E. Yoneki, “BUBBLE RAP: Social-based forwarding in delay-tolerant networks,” *IEEE Trans. Mobile Comput.*, vol. 10, no. 11, pp. 1576–1589, Sep. 2011.
[2] H. Zhu, M. Dong, S. Chang, Y. Zhu, M. Li, and X. Sherman Shen, “Zoom: Scaling the mobility for fast opportunistic forwarding in vehicular networks,” in *Proc. IEEE INFOCOM*, Turin, Italy, Apr. 2013, pp. 2832–2840.
[3] J. Jeong, S. Guo, Y. Gu, T. He, and D. H. Du, “TSF: Trajectory-based statistical forwarding for infrastructure-to-vehicle data delivery in vehicular networks,” in *Proc. IEEE 30th Int. Conf. Distrib. Comput. Syst.*, Genova, Italy, Jun. 2010, pp. 557–566.
[4] Y. Wu, Y. Zhu, and B. Li, “Infrastructure-assisted routing in vehicular networks,” in *Proc. IEEE INFOCOM*, Orlando, FL, USA, Mar. 2012, pp. 1485–1493.
[5] X. Chen, W. Huang, X. Wang, and X. Lin, “Multicast capacity in mobile wireless ad hoc network with infrastructure support,” in *Proc. IEEE INFOCOM*, Orlando, FL, USA, Mar. 2012, pp. 271–279.
[6] B. Karp and H.-T. Kung, “Gpsr: Greedy perimeter stateless routing for wireless networks,” in *Proc. ACM MobiCom*, Boston, MA, USA, Aug. 2000, pp. 243–254.
[7] V. Naumov and T. R. Gross, “Connectivity-aware routing (CAR) in vehicular ad-hoc networks,” in *Proc. IEEE INFOCOM*, Anchorage, AK, USA, May 2007, pp. 1919–1927.
[8] K. C. Lee, M. Le, J. Härri, and M. Gerla, “Louvre: Landmark overlays for urban vehicular routing environments,” in *Proc. IEEE Veh. Technol. Conf.*, Calgary, Canada, Sep. 2008, pp. 1–5.
[9] M. Jerbi, S.-M. Senouci, T. Rasheed, and Y. Ghamri-Doudane, “Towards efficient geographic routing in urban vehicular networks,” *IEEE Trans. Veh. Technol.*, vol. 58, no. 9, pp. 5048–5059, Nov. 2009.
[10] J. Jeong, S. Guo, Y. Gu, T. He, and D. Du, “TBD: Trajectory-based data forwarding for light-traffic vehicular networks,” in *Proc. IEEE Int. Conf. Distrib. Comput. Syst.*, Montreal, Canada, Jun. 2009, pp. 231–238.
[11] J. Zhao and G.Cao, “Vadd: Vehicle-assisted data delivery in vehicular ad hoc networks,”*IEEE Trans. Veh. Technol.*, vol. 57, no. 3, pp. 1910–1922, May 2008.
[12] Y. Xiang, Z. Liu, R. Liu, W. Sun, and W. Wang, “Geosvr: A map-based stateless vanet routing,” *Ad Hoc Netw.*, vol. 11, no. 7, pp. 2125–2135, Sep. 2013.
[13] A. Lindgren, A. Doria, and O. Schelen, “Probabilistic routing in intermittently connected networks,” *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 7, no. 3, pp. 19–20, Jul. 2003.
[14] L. Zhang, B. Yu, and J. Pan, “Geomob: A mobility-aware geocast scheme in metropolitans via taxicabs and buses,” in *Proc. IEEE INFOCOM*, Toronto, Canada, Apr. 2014, pp. 1279–1787.
[15] H. Wen, F. Ren, J. Liu, C. Lin, P. Li, and Y. Fang, “A storage-friendly routing scheme in intermittently connected mobile network,” *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 1138–1149, Mar. 2011.
[16] N. P. Nguyen, T. N. Dinh, S. Tokala, and M. T. Thai, “Overlapping communities in dynamic networks: their detection and mobile applications,” in *Proc. ACM MobiCom*, Las Vegas, NV, USA, Sep. 2011, pp. 85–96.
[17] W. Gao, G. Cao, T. La Porta, and J. Han, “On exploiting transient social contact patterns for data forwarding in delaytolerant networks,” *IEEE Trans. Mobile Comput.*, vol. 12, no. 1, pp. 151–165, Nov. 2013.
[18] J. Qin, H. Zhu, Y. Zhu, L. Lu, G. Xue, and M. Li, “Post: Exploiting dynamic sociality for mobile advertising in vehicular networks,” in *Proc. IEEE INFOCOM*, Toronto, Canada, Apr. 2014, pp. 1761–1769.
[19] R. H. Khokhar, R. M. Noor, K. Z. Ghafoor, C.-H. Ke, and M. A. Ngadi, “Fuzzy-assisted social-based routing for urban vehicular environments,” *EURASIP J. Wireless Commun. Netw.*, vol. 2011, no. 1, pp. 1–15, Dec. 2011.
[20] Z. Li, C. Wang, S. Yang, C. Jiang, and I. Stojmenovic, “Space-crossing: Community-based data forwarding in mobile social networks under the hybrid communication architecture,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 4720–4727, Sep. 2015.
[21] A. Dua, N. Kumar, and S. Bawa, “A systematic review on routing protocols for vehicular ad hoc networks,” *Veh. Commun.*, vol. 1, no. 1, pp. 33–52, Jan. 2014.
[22] S. Basagni, M. Conti, S. Giordano, and I. Stojmenovic, *Mobile Ad Hoc Networking: The Cutting Edge Directions*, 2nd ed. Hoboken, NJ, USA: Wiley, 2013.
[23] T. Hossmann, T. Spyropoulos, and F. Legendre, “Know thy neighbor: Towards optimal mapping of contacts to social graphs for DTN routing,” in *Proc. IEEE INFOCOM*, San Diego, CA, USA, Mar. 2010, pp. 1–9.
[24] S. Fortunato, “Community detection in graphs,” *Phys. Rep.*, vol. 486, no. 3–5, pp. 75–174, Feb. 2010.
[25] A. Lancichinetti and S. Fortunato, “Community detection algorithms: A comparative analysis,” *Phys. Rev. E*, vol. 80, no. 5, Nov. 2009, Art. no. 056117.
[26] [Online]. Available: http://cabspotting.org.
[27] [Online]. Available: https://www.dropbox.com/s/xbx0iw954a9m7kk/Shanghai%20taxi%20data-08.zip.
[28] A. Keränen, J. Ott, and T. Kärkkäinen, “The one simulator for DTN protocol evaluation,” in *Proc. Int. Conf. Simul. Tools Techn.*, Rome, Italy, Mar. 2009, Art. no. 55.
[29] [Online]. Available: https://www.dropbox.com/s/qc6a7tc1wsfnlm4/appendix-TVT.pdf?dl=0.

Zhong Li received the Ph.D. degree from Tongji University, Shanghai, China, in 2015.

She is currently a Lecturer with DongHua University, Shanghai, China,. Her research interests include vehicular networks, wireless networking, and Internet of Things.

Cheng Wang received the Ph.D. degree from Tongji University, Shanghai, China, in 2011.

He is currently a Professor of computer science with Tongji University. His research interests include wireless networking, mobile social networks, and cloud computing.

Lu Shao, photograph and biography not available at the time of publication.

Chang-Jun Jiang received the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1995.

He is currently a Professor with the Department of Computer Science and Engineering, Tongji University, Shanghai, China. His research interests include concurrent theory, Petri net, and formal verification of software, wireless networks, concurrency processing, and intelligent transportation systems.

Cheng-Xiang Wang received the B.Sc. and M.Eng. degrees in communication and information systems from Shandong University, Shandong, China, in 1997 and 2000, respectively, and the Ph.D. degree in wireless communications from Aalborg University, Aalborg, Denmark, in 2004.

He has been with Heriot-Watt University, Edinburgh, U.K., since 2005, and became a Professor with the Department of Wireless Communications in 2011. He is also an Honorary Fellow of the University of Edinburgh, U.K., the Chair Professor of Shandong University, and a Guest Professor of Southeast University, China. He was a Research Fellow with the University of Agder, Grimstad, Norway, from 2001 to 2005, a Visiting Researcher with Siemens AG-Mobile Phones, Munich, Germany, in 2004, and a Research Assistant with Hamburg University of Technology, Hamburg, Germany, from 2000 to 2001. His current research interests focus on wireless channel modeling and 5G wireless communication networks, including green communications, cognitive radio networks, high mobility communication networks, massive MIMO, millimeter-wave communications, and visible light communications.