When 3G Meets VANET: 3G-Assisted Data Delivery in VANETs

Qingwen Zhao, Yanmin Zhu, *Member, IEEE*, Chao Chen, Hongzi Zhu, *Member, IEEE*, and Bo Li, *Fellow, IEEE*

***Abstract*—In this paper, we consider a sensory data gathering application of a vehicular ad hoc network (VANET) in which vehicles produce sensory data, which should be gathered for data analysis and making decisions. Data delivery is particularly challenging because of the unique characteristics of VANETs, such as fast topology change, frequent disruptions, and rare contact opportunities. Through empirical study based on real vehicular traces, we find an important observation that a noticeable percentage of data packets cannot be delivered within time-to-live. In this paper, we explore the problem of 3G-assisted data delivery in a VANET with a budget constraint of 3G traffic. A packet can either be delivered via multihop transmissions in the VANET or via 3G. The main challenge for solving the problem is twofold. On the one hand, there is an intrinsic tradeoff between delivery ratio and delivery delay when using the 3G. On the other hand, it is difficult to decide which set of packets should be selected for 3G transmissions and when to deliver them via 3G. In this paper, we propose an approach called *3GDD* for 3G-assisted data delivery in a VANET. We construct a utility function to explore the tradeoff between delivery ratio and delivery delay, which provides a unified framework to reflect the two factors. We formulate the 3G-assisted data delivery as an optimization problem in which the objective is to maximize the overall utility under the 3G budget constraint. To circumvent the high complexity of this optimization problem, we further transition the original optimization problem as an integer linear programming problem (ILP). Solving this ILP, we derive the 3G allocation over different time stages. Given the 3G budget at each time stage, those packets that are most unlikely delivered via the VANET are selected for 3G transmissions. We comprehensively evaluate our *3GDD* using both synthetic vehicular traces and real vehicular 3G traces. Evaluation results show that our approach outperforms other schemes under a wide range of utility function deflations and network configurations.**

***Index Terms*—Vehicular ad hoc networks, 3G, packet allocation, data delivery, integer linear programming.**

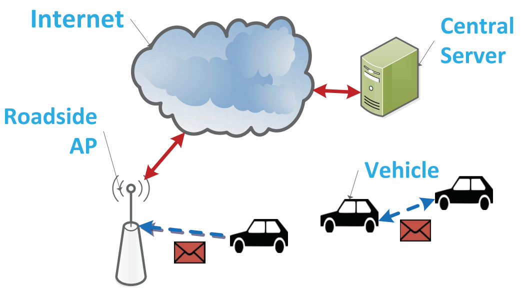


Fig. 1. The system architecture for gathering sensory data from mobile vehicles and processing at the central server.

I. INTRODUCTION

VEHICULAR ad hoc networks (VANETs) have recently attracted enormous attention. Equipped with sensors [1]–[3], every vehicle is a powerful mobile sensor and can exchange information with each other within the communication range. Many applications and services have been envisioned, such as vehicle collision avoidance and road safety [4], [5], P2P content sharing [6], and intelligent transportation [7]. Data delivery plays a key role in VANETs, since almost all applications require efficient data delivery. The performance metrics including delivery ratio, delay and throughput have been concerned in VANETs [8].

In this paper we consider data gathering [9], [10] in VANETs. Each vehicle may generate sensory data packets which are destined to a central server for data analysis and making decisions. We consider a VANET with an infrastructure with roadside access points (APs) that are connected to the Internet, as shown in Fig. 1. As the central server is also connected to the Internet, it is sufficient for each vehicle to deliver its packets to one of the APs. It is desirable for the data gathering application that the central server can collect data packets from the vehicles with high success rate and low latency.

However, data delivery in VANETs is particularly challenging due to the unique characteristics of VANETs, especially under the circumstance of sparsely connected VANETs [11]. *First*, the high mobility and uneven distribution of vehicles make it difficult to find a connected path between a pair of

Manuscript received February 1, 2013; revised April 23, 2013; accepted May 26, 2013. Date of publication May 31, 2013; date of current version August 21, 2013. This work was supported in part by the Shanghai Pu Jiang Talents Program under Grant 10PJ1405800, Shanghai Chen Guang Program under Grant 10CG11, the NSFC under Grants 61170238, 60903190, 61027009, 60933011, 61202375, and 61170237, MIIT of China under Grant 2009ZX03006-001-01, Doctoral Fund of Ministry of Education of China under Grant 20100073120021, the National 863 Program under Grants 2009AA012201 and 2011AA010500, HP IRP under Grant CW267311, SJTU SMC Project under Grant 201120, STCSM under Grants 08dz1501600 and 12ZR1414900, Singapore NRF under Grant CREATE E2S2, the NSFC/RGC under Grant N-HKUST610/11, HKUST under Grants RPC11EG29, SRFI11EG17-C, and SBI09/10.EG01-C, and Guang-Dong Bureau of Science and Technology under Grant GDST11EG06. The associate editor coordinating the review of this paper and approving it for publication was Prof. Honggang Wang.

Q. Zhao, Y. Zhu, C. Chen, and H. Zhu are with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: qwzhao@sjtu.edu.cn; yzhu@sjtu.edu.cn; chenchao@sjtu.edu.cn; hongzi@sjtu.edu.cn).

B. Li is with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Kowloon 999077, Hong Kong (e-mail: bli@cse.ust.hk).

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/JSEN.2013.2265304

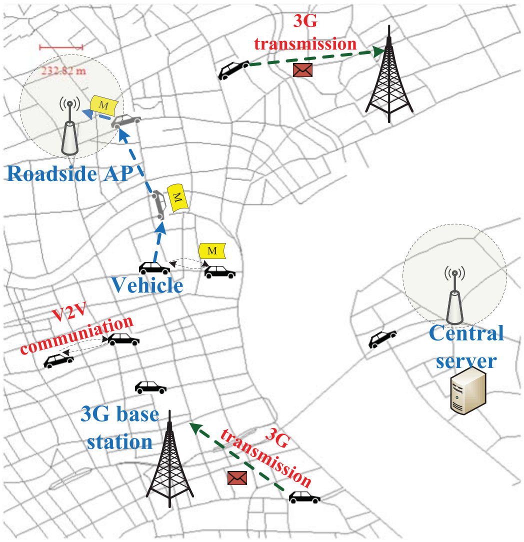


Fig. 2. Illustration of data delivery in vehicular networks. There are two ways for a node to deliver its data packets to the central server: via 3G and via inter-vehicle communications.

source and destination in VANETs. *Second*, the link capacity between two vehicles is typically very limited. Empirical study has shown that the duration time between two moving vehicles can be as short as 13 seconds on average [12], and thus the number of packets that can be transmitted within one encounter is limited.

Many routing algorithms have been proposed for data delivery in VANETs. The carry-and-forward paradigm has been exploited by several opportunistic forwarding schemes [13], [14]. Some approaches try to exploit vehicle trajectories [15], [16]. Some approaches also consider infrastructure-based VANETs, where the infrastructure refers to the roadside units (RSUs) or access points (APs). It has been shown that the use of the infrastructure can significantly improve packet delivery in a VANET [17], [18].

Unfortunately, through empirical study based on real vehicular traces we find that a noticeable percentage of data packets fail to be delivered within time-to-live (TTL). The main reason is that infrequent encountering opportunities in sparsely distributed VANETs pose a bottleneck for data delivery in VANETs. Even worse, vehicles exhibit very different mobility patterns. Some of the vehicles have less frequent encounters than other vehicles. Although the use of the infrastructure can help packet delivery in VANETs, it is usually insufficient to guarantee the delivery of all packets.

Nowadays, the third generation (3G) networks have been widely deployed and used. The 3G provides vehicles with ubiquitous Internet access and data communication with other vehicles. Thus, it is highly beneficial to exploit the 3G to assist data delivery in VANETs. However, the 3G cannot be used for free. The cost of the 3G traffic is noneligible. For instance, China Unicom, the telecommunication operator that provides W-CDMA coverage in China, charges several US dollars per MBytes. Thus, the 3G can be useful for improving the performance of data delivery in VANETs. As illustrated in Fig. 2, when 3G is available for data delivery in a VANET, a vehicle can either deliver its packets via 3G or via multihop transmissions of VANET. However, it is clear that the use of 3G must be minimized for cost reduction.

Therefore, this paper explores the data gathering problem in a VANET where each vehicle has an additional 3G data channel. More specifically, there is a budget constraint on the 3G traffic, which defines the maximum number of packets that can be delivered via 3G. The objective is to maximize the delivery ratio of packets and meanwhile to minimize the delivery delay. There are two challenges for solving the 3G-assisted data delivery problem in a VANET. *First*, there is an intrinsic tradeoff between delivery ratio and delivery delay when using the 3G. It is preferable to let all packets first try cost-free multihop transmissions in the VANET. A packet whose TTL is about to due can be selected to go through the 3G. By this mean, the limited 3G traffic can be well utilized (hence a better delivery ratio) by first directing the traffic demand to the VANET, but it may introduce large delivery delays. *Second*, it is difficult to decide which packets from the whole set of packets and when to deliver them through the 3G.

In this paper we propose an approach called *3GDD* for 3G-assisted data delivery in a VANET. We construct a utility function to explore the tradeoff between delivery ratio and delivery delay, which provides a unified framework to reflect the two factors. We formulate the 3G-assisted data delivery as an optimization problem in which the objective is to maximize the overall utility under the 3G budget constraint and the main decision variables characterize which packets at which time slots to delivery via 3G. *3GDD* employs the state-of-the-art routing algorithm to deliver the rest of data packets to reach one of the APs. To circumvent the high complexity of this optimization problem, we further transition the original optimization problem as an integer linear programming problem (ILP) by simplifying the contact graph model of the VANET. Solving this ILP, we derive the 3G allocation over different time stages. Considering the different delivery probability of each individual data packet via the VANET, we estimate the probability of each packet and sort the packets in terms of delivery probability. Given the 3G budget at each time stage, those packets that are most unlikely delivered via the VANET are selected for 3G transmissions.

We have comprehensively evaluated our *3GDD* by using both synthetic vehicular traces and real vehicular 3G traces. In addition, we compare our approach with other alternative schemes. Evaluation results show that our approach outperforms other schemes under a wide range of utility function deflations and network configurations.

The key technical contributions made in this paper are summarized as follows.

- This is the first attempt, to the best of our knowledge, to exploiting the 3G to assist data delivery in VANETs.
- We formally formulate the 3G-assisted data delivery problem as an optimization problem with constraints.
- We propose an approach called *3GDD* to solving the 3G-assisted data delivery in a VANET. This approach uses a utility function to explore the tradeoff between delivery ratio and delivery delay and adaptively allocates

the 3G traffic budget over the whole course of packet delivery.
- Comprehensive experiments based on both synthetic traces and real traces have been performed and the results show that our approach achieves better utility than the alternative schemes.

The rest of the paper is organized as follows. The next section shows our empirical study which demonstrates that a noticeable percentage of data packets fail to be delivered before their TTLs expire. Section II presents an overview of related work. Section IV gives the network model and then formally defines the problem. In Section V, we describe the design details of our approach. Evaluation results are presented in Section VI. The paper is concluded in Section VII.

II. RELATED WORK

In this section we review related work. First, we discuss related work on data delivery in infrastructureless VANETs. Then, we present some approaches for data delivery in infrastructure-based VANETs. And finally we look at a few studies which consider the integration of 3G and VANETs.

There has been a plethora of research work [13]–[16] on data routing in VANETs. [14] adopts the idea of carry and forward. A moving vehicle carries a data packet until it contacts another vehicle and forwards its data. In order to further improve the performance of data forwarding, trajectory-based routing algorithms have been designed [15], [16]. In [15], the trajectories of vehicles are predicted by using multiple order Markov chains. Based on the analytical model, the contact probabilities and forwarding metrics between different node pairs can be determined. There are also many other algorithms trying to exploit many other specific features such as geographic feature [19]–[21]. [22] considers data delivery from vehicles to APs and aims at minimizing the communication overhead under a delay bound. Vehicles leverage local or global knowledge of traffic to alternate between the data mining and forwarding. The network model is abstracted from digital map of the area.

To further improve data delivery performance of VANTEs, some routing algorithms assume the availability of the infrastructure consisting of roadside units (RSUs) or APs. In [23], some RSUs can be selected as relays which can help to forward data packets to the destination nodes. These roadside can be regarded as some static nodes in the network and they are not connected. In [17], roadside APs are employed to store and forward data packets by communicating with passing vehicles, which can enhance the connectivity of the network. The main design focus is on packet forwarding and buffer allocation at APs. Data dissemination problem is studied in [24]. RSUs at intersections can buffer packets from source nodes and then rebroadcast to other vehicles.

In recent years, some have started to explore the integration of 3G and VANET. In [25], vehicles are dynamically clustered according to different metrics. A minimum number of cluster heads are selected which serve as the gateway between a VANET and the 3G. A gateway has two radio interfaces: 3G and IEEE 802.11p. The authors assumed that the VANET has good network connectivity. A multicast protocol is presented in [26], which assumes that trajectories of vehicles and traffic demands are known. It solves the issues that how and when to download data through remote transmissions via 3G to cover traffic demands from vehicles while keeping remote transmission cost minimized.

In summary, most related work has not considered the use of 3G to assist data delivery in VANETs. A few existing studies notice that it is valuable to integrate 3G to a VANET, but they do not solve the problem how to make the full potential use of the limited 3G budget.

III. EMPIRICAL STUDY

This section presents our empirical study based on real vehicular GPS traces, through which we show that there is a noticeable percentage of data packets which cannot be delivered before their TTLs through multihop inter-vehicle transmissions.

A. Methodology and Setup

In our empirical study, the dataset of real vehicle traces were collected from around 12,000 taxis distributed in Shenzhen, China. Each of the taxis is equipped with a Global Positioning System (GPS) receiver and thus is able to know its location. Each trace records the consecutive locations of a taxi at a time interval of around 30 seconds to 1 minute. In simulation, we interpolate the discrete trace to get a continuous trajectory.

We use the digital map of Shenzhen, China to manually deploy APs on the map. As a result, most of the taxis can frequently encounter the APs. In simulation, five APs were deployed in the city. These APs are assumed to be connected via a wire network thus a packet is delivered as long as it reaches any one of the five APs. We select a subset of 500 vehicles from the whole set of 12,000 taxis in the trace dataset. 48 hour-long traces are adopted for simulations. The default TTL of the data packets is 60 minutes. In total 80 data packets were generated at randomly selected nodes.

It is well known that the routing algorithm has a great impact on the delivery performance packets. To show the significance of undelivered packets in a VANET, we use the aggressive routing algorithm called Epidemic. Using this algorithm, a node would actively replicate a data packet as it encounters another node without holding such packet. Apparently, the Epidemic routing algorithm is beneficial to the reduction of delivery delay and to the improvement of delivery ratio, but at a high cost of network traffic. It can only be useful when the traffic demand is very light.

B. Empirical Results

We next show the results from our empirical study. First, we show delivery delays of individual packets in Fig. 3. We can find that delivery delays of data packets differ significantly, which range from 0 to 3600 seconds. About 80% packets are successfully delivered within the TTL of 60 minutes, and nearly 20% of data packets fail to be delivered.

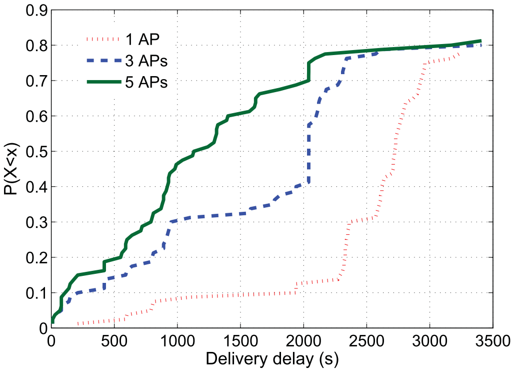


Fig. 3. CDF of delivery delays.

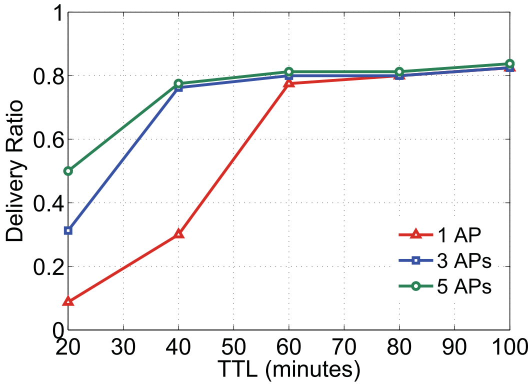


Fig. 4. Delivery ratio vs. TTL of data packets.

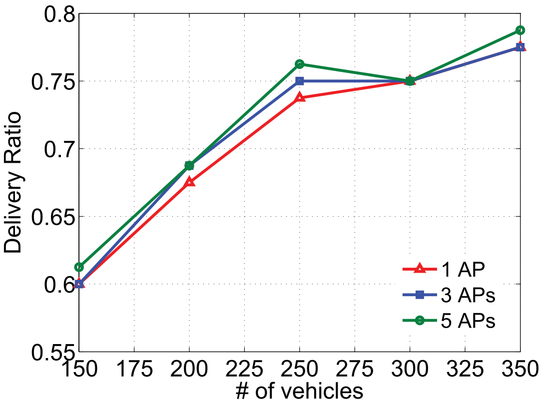


Fig. 5. Delivery ratio vs. number of vehicles.

We also study the impact of some important system parameters, including number of APs, number of vehicles and TTL on packet delivery performance.

To investigate the impact of TTL, we vary TTL from 20 to 100 minutes. From Fig. 4, we can see that the delivery ratio increases as the TTL becomes larger. However, when the TTL is longer than 80 minutes, the delivery ratio becomes stable at around 80%. It is easy to understand that the longer the TTL, the higher delivery probability a data packet has. However, some data packets cannot be delivered to one of the APs even if the TTL is long enough. The main reason is that some source vehicles may rarely encounter other vehicles.

To investigate the impact of number of vehicles, we vary the number of vehicles in the system from 100 to 300. In Fig. 5, we can find that the delivery ratio increases as the number of vehicles increases. However, when the number of vehicles is between 250 and 300, the delivery ratio is always round 75%. It is understandable that the increase in the number of vehicles can enhance the connectivity of the network. However, there are still some data packets which cannot be delivered in their TTL even there are a large number of vehicles.

In addition, we can also see from both Fig. 4 and Fig. 5 that a larger number of APs helps increase the delivery ratio in the network. This is apparent since the deployments of more APs increase the opportunities of vehicles to encounter APs and help the delivery of data packets.

In conclusion, we have observed that a noticeable percentage of data packets failed to be delivered even with a larger number of vehicles, more APs and a longer TTL. This strongly suggests that the use of 3G would greatly improve the packet delivery performance in a VANET.

IV. NETWORK MODEL AND PROBLEM FORMULATION

In this section, we first present the network model, and then formally define the problem of 3G-assisted data delivery in VANETs.

A. System Model

We consider an infrastructure-based VANET consisting of three parts: the set V of moving vehicles, the set Φ of APs, and the central server. The APs are connected via a wired network. Each vehicle may produce a packet containing sensory data, which should be gathered to a central server within its time-to-live (TTL) denoted by T. We assume all data packets share the same TTL. Data packets are of equal size. We use $\mathbb{P} = \{p_1, p_2, \ldots, p_m\}$ to denote the set of data packets which need to be gathered.

Equipped with both 3G radio and short-range radio for inter-vehicle and vehicle-to-AP communications (e.g., DSRC), each vehicle can directly deliver a packet via 3G or via multi-hop transmissions to one of the APs in the VANET. A short-range communication can only occur when two nodes are in the communication range of each other. When the 3G is used, the delivery delay is small and can be ignored. When multi-hop transmissions in the VANET are adopted, there is a significant delivery delay. Considering the cost of 3G traffic, we consider that there is a budget of 3G traffic, denoted by B.

We model the VANET as a *contact graph* $G(V, E)$, where V denotes the set of nodes in the network, and an edge $e_{ij} \in E$ represents the contact process between two nodes i, and j. According to some recent studies [27], [28], the contact process between two nodes i and j can be modelled as a Poisson process with the contact rate λ. Accordingly, the inter-contact time between two nodes is exponentially distributed. Such a model has also been experimentally validated by the study based on real vehicular traces [28].

In addition, the central server in the system not only gathers data from the network, but also it deals with 3G budget allocations. To this end, it needs to collect the current network information from vehicles via extra 3G communications. With this information, it can analyze the delivery probability of a data packet or decide the 3G budget allocation to each time slot. We assume the size of such information is small enough compared to the data packets. Thus the cost of 3G in this case is negligible.

B. Problem Formulation

The general objective of 3G-assisted data delivery in the VANET is to maximize the delivery ratio and minimize the delivery delay of the packets under the budget constraint of 3G traffic. To formally define the 3G-assisted data delivery problem, we first introduce some important notations.

In our model, time is divided into s equal slots with the length of τ, and $T = s \times \tau$. Let $\mathcal{T} = \{t_1, t_2, \ldots, t_s\}$ denote the set of all the slots. A data packet can go through 3G in any time slot before the TTL expires. We use an indicator variable x_i^k to denote whether a data packet goes through 3G or not,

$$x_i^k = \begin{cases} 1, & p_k \text{ goes through 3G in slot } t_i, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

The set of all data packets delivered in slot t_i can be divided into two subsets: one subset includes those packets which are delivered via 3G and the other subset include those packets delivered via multihop transmissions in the VANET. Let n_i denote the total number of data packets delivered via 3G in slot t_i, and we have

$$n_i = \sum_{k=1}^{m} x_i^k. \tag{2}$$

The number of data packets delivered via VANET is a random variable, dependent on both node mobility and the routing algorithm. We use a random variable Z_i to denote the number of data packets delivered in slot t_i, and the expected value of Z_i is z_i,, i.e., $E[Z_i] = z_i$. Thus, the total expected number of data packets delivered in slot t_i is $n_i + z_i$.

To combine the two performance metrics of delivery ratio and average delay, we define a utility function as follows.

Definition 1 (Utility): The utility U of the 3G-assisted data delivery is a function of the delivery ratio and the delivery delay, which is defined as,

$$\begin{aligned} U &= \alpha \frac{\sum_{i=1}^{s} (Z_i + n_i)}{m} + (1-\alpha) \frac{\sum_{i=1}^{s} (T - i\tau)(Z_i + n_i)}{mT} \\ &= \alpha \frac{\sum_{i=1}^{s} (Z_i + n_i)}{m} + (1-\alpha) \frac{\sum_{i=1}^{s} (s - i)(Z_i + n_i)}{ms}, \end{aligned} \tag{3}$$

where $i \times \tau$ is the delivery delay of each of the data packets delivered in slot t_i and the parameter $\alpha \in [0, 1]$.

From the definition of the utility, we can conclude that to improve the utility we should increase the delivery ratio and reduce the delivery delay of each individual packet. The parameter, α, controls the relative importance of the two metrics and can be customized by application developers. For those applications which prefer a short delivery delay, a smaller α can be selected; otherwise, a larger α should be used. The impact of α is further investigated in Section VI.

Next, we can formally define the problem of 3G-assisted data delivery problem.

Definition 2 (3G-assisted data delivery problem): The 3G-assisted data delivery problem is defined as follows. Given the budget B of 3G traffic and the set $\mathbb{P}$ of data packets, we should determine if a packet $p \in \mathbb{P}$ is delivered via 3G and when to deliver it via 3G, i.e., determine x_i^k, with the objective of maximizing the overall utility U.

$$\max \quad U, \tag{4}$$

s.t.,

$$\sum_{i=1}^{s} \sum_{k=1}^{m} x_i^k \leq B, \tag{5}$$

$$\sum_{i=1}^{s} \sum_{k=1}^{m} x_i^k + \sum_{i=1}^{s} Z_i \leq m, \tag{6}$$

$$x_i^k \in \{0, 1\}, i \in \{1, 2, \cdots, s\}, k \in \{1, 2, \cdots, m\}, \tag{7}$$

Constraint (5) guarantees that the total number of packets delivered via 3G is no more than B, and constraint (7) ensures that the total number of delivered packets is not greater than m.

This paper aims to investigate the potential increase of data delivery performances with the assistance of 3G. We assume the bandwidth is infinite in the network, and the epidemic routing algorithm is adopted in order to achieve the best routing performance in VANETs.

V. DESIGN OF *3GDD*

In this section we present the design details of our approach *3GDD*.

A. Overview

Having formally defined the 3G-assisted data delivery problem, one could solve it by determining all x_i^k. Unfortunately, this problem formulation is not constructive. The main reason is as follows. The set of packets delivered via VANET cannot be known in advance and therefore Z_i is unknown. Consequently, it becomes impossible to actually solve the optimization problem.

In this paper, we propose *3GDD* to approach the 3G-assisted data delivery problem. The main idea of *3GDD* is the following. First, it estimates an expected number of each Z_i based on the contact graph model of the VANET. Then, the original optimization problem is simplified as an integer linear programming (ILP) problem. Solving the ILP, we essentially allocate the 3G budget to different time slots. Finally, given the allocated 3G budget at each time slot, *3GDD* then selects those packets that are most unlikely delivered via VANET to deliver via 3G.

B. ILP Formulation

This subsection first presents how to estimate the expected number of packets delivered via VANET at each time slot, i.e., $z_i = E[Z_i]$, and then give the ILP formulation.

To estimate z_i, we have to simplify the network model. We assume that the contact process between any node i and j follows the same Poisson process with average contact rate λ.

Lemma 1: Given the set of APs Φ, the contacts between a node and one of the APs form a Poisson process with the contact rate is $|\Phi|\lambda$, where $|\Phi|$ is the total number of APs in the network.

Proof: As the contacts between a node and an AP in the network form a Poisson process with the contact rate λ and the contacts between a node and any one of the APs are independent, therefore the contact rate between a node and the total $|\Phi|$ APs forms a Poisson process with contact rate $|\Phi|\lambda$. ■

Let m_i' present the expected number of data packets which have not been delivered yet until time slot t_i, excluding the set of the data packets to be delivered via 3G in slot t_i. According to Lemma 1, it forms a Poisson process that m_i' node contacts the APs with the contact rate $|\Phi|\lambda m_i'$, and then the expected number of data packets delivered z_i in slot t_i is $|\Phi|\lambda m_i'\tau$. Clearly, $m_1' = m - n_1$, and $E[Z_1] = z_1 = m_1'|\Phi|\lambda\tau$. The expected value of m_i' at time slot t_i can be updated as follows:

$$m_i' = m - \sum_{j=1}^{i} n_j - \sum_{j=1}^{i-1} z_j, \tag{8}$$

We can deduce $z_i, i \in \{1, 2, \cdots, s\}$ from z_1. In this way, we can obtain the expression of the expected U, which is only the function of n_i. As n_i is required to be an integer, we can redefine the 3G-assist data delivery problem in VANETs as an integer programming problem:

Definition 3 (ILP formulation of 3G-assisted data delivery problem): Given the 3G-assisted data problem as defined in Definition 7, its simplified ILP formulation is as follows,

$$\max \quad E[U], \tag{9}$$

$$s.t. \quad \sum_{i=1}^{s} n_i \leq B, \tag{10}$$

$$\sum_{i=1}^{s} n_i + \sum_{i=1}^{s} z_i \leq m, \tag{11}$$

$$n_i \in \mathbb{N}, \tag{12}$$

Traditional data delivery problem which aims to maximize the delivery ratio can be reduced to the classical maximum disjoint edges problem, which has been proven NP-hard [17]. Our 3G-assisted data delivery problem is more complex since it has to decide when and which data packet should be transmitted via 3G, such that the delivery ratio is maximized and meanwhile the average delay is minimized. We formulate the problem as an integer linear programming problem, and solve it based on Branch and Bound, which implies implicit enumeration of solution space and unfortunately is NP-hard.

C. Heuristic Algorithm Design

To tackle the ILP formulation defined in Definition 3, we develop a heuristic algorithm based on the Tabu search algorithm [29], [30] which is very efficient for solving ILP. The main idea of the heuristic algorithm is to search the neighbor or local potential solutions starting from a feasible initial solution, in the hope of finding an improved solution which increases the total utility. Before introducing the algorithm, we first to define the concept of neighbor.

Algorithm 1: Deriving the Optimal Allocation of 3G Budget

```
Require: 3G budget $B$; time slots, $s$; the number of APs, $\Phi$; the
    number of data packets, $m$; TTL, $T$; contact rate, $\lambda$; the
    maximum number of iterations, $N_{max}$; the initial solution, $\mathcal{N}_0$;
    The minimum distance, $d_{min}$;
Ensure: 3G budget allocation, $\mathcal{N}^* = \{n_1, n_2, .., n_s\}$;
$\mathcal{N}^* \leftarrow \mathcal{N}_0$;
$tabList \leftarrow null$;
$i = 0$;
while $i < N_{max}$ do
    $candidates \leftarrow null$;
    for $nCandidate \in nNeighborhood$ do
        if $nCandidate \notin tablist$ then
            $candidates \leftarrow candidates + nCandidate$;
        end if
    end for
    $nCandidate \leftarrow FindBestCandidate(candidates)$;
    if $utiltiy(nCandidate) > utility(\mathcal{N}^*)$ then
        $tablist \leftarrow tablist + nCandidate$;
        $\mathcal{N}^* \leftarrow nCandidate$;
    end if
    $i \leftarrow i + 1$;
end while
```

Definition 4 (neighbor): For an integer a, b is a neighbor of a if $|b - a| < d_{min}$ is satisfied, where d_{min} is a given distance. For a vector $\boldsymbol{v}$, $\boldsymbol{u}$ is a neighbor of $\boldsymbol{v}$ if any element in $\boldsymbol{u}$ is the neighbor of the corresponding element in $\boldsymbol{v}$.

The detailed algorithm is presented in Algorithm 1. The algorithm starts from an initial solution $\mathcal{N}_0$ and search its local candidate solutions. To determine the initial solution $\mathcal{N}_0$, we can solve the solution to the LP relaxation of the original ILP, and then round the entries of the solution. The derived solution can be served as the initial solution. At first, the best solution is initialized as $\mathcal{N}_0$. Then, the algorithm searches the neighborhood of the current solution to find an improved solution which has a higher utility as defined in Section IV. Next, this improved solution will be further explored. The algorithm repeats this process and returns a current best 3G budget allocation solution after a maximum number of iterations N_{max} is reached. To improve the performance and avoid explore those solutions which have already been visited before, our algorithm use a list *tablist* to record visited solutions.

D. Packet Selection for 3G Transmissions

With the solution of the ILP, we can derive the allocation of the 3G budget to each time slot, with which we need to establish the set of packets to be delivered via 3G. Apparently, it is reasonable to select those packets that are most unlikely delivered via VANET. Therefore, the main challenge for such packet selection is to estimate the delivery probability of each packet before each time slot starts. In this subsection, we introduce how to estimate the delivery probability of each data packet at time t. For the simplicity of analysis, we assume that data packets are generated at time 0.

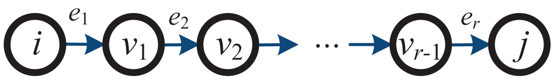


Fig. 6. Illustration of a route Υ from node i to node j via r hops.

For a one-hop path from node i to node j with the contact rate λ_i, as the inter-contact time between nodes is exponentially distributed, a data packet can be delivered successfully from nodes i to j within time x is $p_{ij}(x) = \lambda_{ij} e^{-\lambda_{ij} x}$, $(x > 0)$. Then for a route Υ from node i to node j with r hops (shown in Fig. 6), denoted by sequence $< e_1, \cdots, e_r >$ with the contact rate $< \lambda_1, \cdots, \lambda_r >$, the delivery probability at time t is

$$p_{ij}^{\Upsilon}(T-t) = \int_0^{T-t} \sum_{k=1}^{r} C_k^r p_k(x))dx \tag{13}$$

$$= \int_0^{T-t} (\sum_{k=1}^{r} C_k^r \lambda_k e^{-\lambda_k x})dx \tag{14}$$

$$= \sum_{k=1}^{r} C_k^r (1 - e^{-\lambda_k (T-t)}). \tag{15}$$

where $C_k^r = \prod_{s=1, s \neq k}^{r} \frac{\lambda_s}{\lambda_s - \lambda_k}$.

We define ϕ as the set of all possible routing routes from node i to j, and then the delivery probability of a data packet from node i to j at time t is

$$p_{ij}(T-t) = 1 - \prod_{\Upsilon \in \phi} p_{ij}^{\Upsilon}(T-t). \tag{16}$$

VI. PERFORMANCE EVALUATION

A. Methodology

In this section, we evaluate the performance of 3G-assisted data delivery algorithm. We mainly consider three performance metrics:

1) **Delivery ratio.** It refers to the success ratio of the number of successfully delivered packets to the total number of packets at the end of an experiment of certain time.
2) **Average delay.** It refers to the average delay for a packet to be received to its destination. It can be calculated by accumulating every delay of each hop. We only calculate end-to-end delay for successfully delivered packets.
3) **Utility.** It combines both delivery ratio and average delay, which can be calculated according to Definition 7.

To evaluate the performance of our algorithm, we compare the performance of our scheme with the following candidate routing schemes:

- **S_Random**: At the beginning, randomly select out B data packets and deliver them via 3G.
- **E_Random**: Randomly select those data packets which fail to be delivered at the last second and deliver them via 3G.
- **M_Average**: The 3G transmission budget is equally allocated to all time slots before a give TTL.

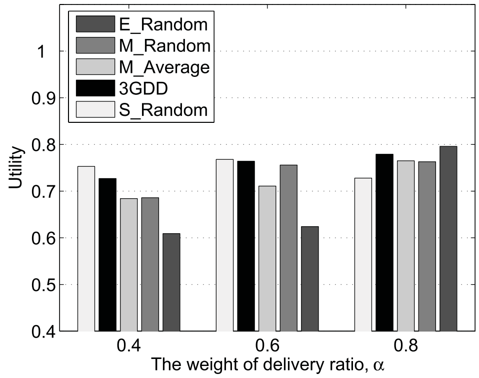


Fig. 7. Utility under different α values.

- **M_Random**: The 3G transmission budget is randomly allocated to all available time slots.

We use the Opportunistic Network Environment simulator (ONE) [31], which is specifically designed for performance evaluation of DTN protocols, to simulate these scenarios on both synthetic and realistic traces collected from 500 taxies in Shenzhen, a metropolis in China. For each simulation settings, we run the simulation twenty times and take the average.

B. Real Vehicular Traces

1) Simulation Settings: We evaluate the performances of all schemes using the real GPS traces collected from 500 taxis in Shenzhen on October 1, 2009. The average contact rate is nearly 0.58 per day. We set the communication range to be 300 meters. We generate 500 packets. For each data packet, we randomly select its source nodes. The number of APs and TTL are set to 50 and two hours, respectively.

2) Impact of Parameter α: At first, we investigate the impact of the weight of delivery ratio α in calculating utility defined in Definition 7. We random choose 400 taxies and vary α from 0.4 to 0.8 with an increment of 0.2.

Fig. 7 shows the utility as a function of α. It can be seen that our algorithm can achieve high utility over a wide range of α values while other schemes shows great instability with different α values.

3) Impact of 3G Budget: In this set of simulation, we random choose 400 taxies and vary the 3G budget from 100 to 200 with an increment of 25. The weight of delivery ratio α is set to 0.6 as suggested in the above simulation.

The simulation results are shown in Figs. 8, 9 and 10. From Fig. 8, we can see that the E_Random achieves the best delivery ratio while the S_Random has worst delivery ratio. It is reasonable that E_Random have such a high delivery ratio because it always let data packets go through 3G, and in this case, the 3G can be fully utilized but leading to long delivery delays. From Fig. 10, it is clear to see that the *3GDD* outperforms all of the other schemes with respect to the utility.

4) Impact of Number of Vehicles: To investigate the impact of number of vehicles on the performance of the data delivery schemes, we vary the number of vehicles in the network from 100 to 300 with an increment of 50 while the 3G budget is set to 150. The weight of delivery ratio α is set to 0.6.

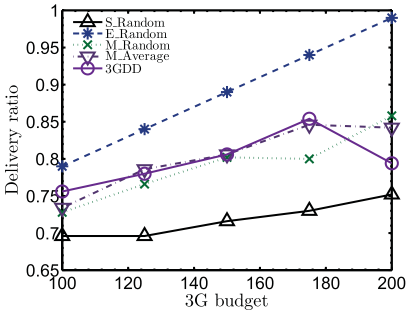


Fig. 8. Delivery ratio vs. 3G budget (Taxi).

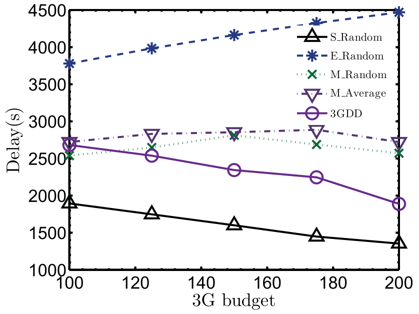


Fig. 9. Delay vs. 3G budget (Taxi).

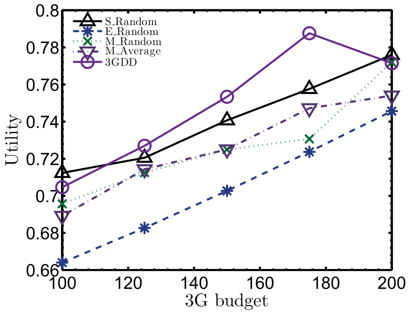


Fig. 10. Utility vs. 3G budget (Taxi).

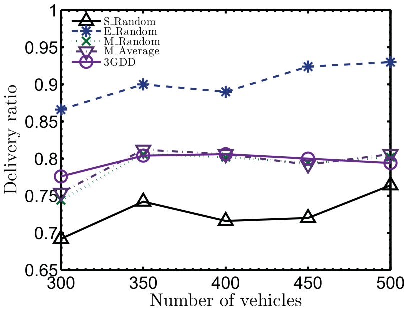


Fig. 11. Delivery ratio vs. number of vehicles (Taxi).

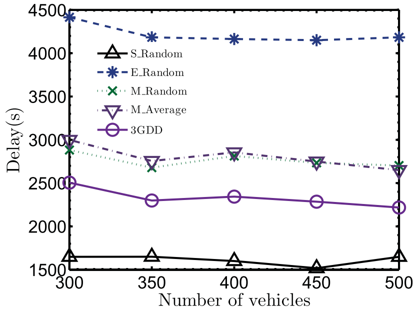


Fig. 12. Delay vs. number of vehicles (Taxi).

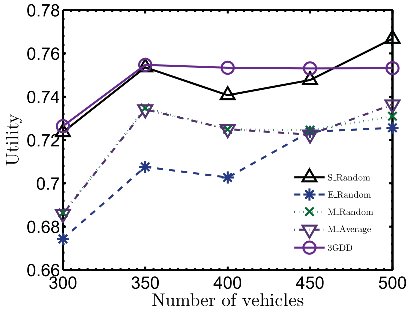


Fig. 13. Utility vs. number of vehicles (Taxi).

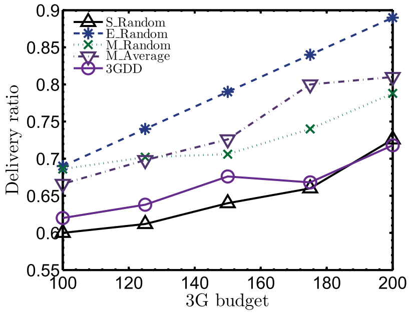


Fig. 14. Delivery ratio vs. 3G budget.

The simulation results are shown in Figs. 11, 12 and 13. It can be seen that, as the number of vehicles in the network increases, the delivery ratio slightly increases in general. Similar to the results in Fig. 8, E_Random has the best delivery ratio while S_Random is the worst. As expected, the total utility of the *3GDD* performs best in a wide range of number of vehicles. When the number of vehicles is larger than 450, S_Random outperforms *3GDD* by only 0.013% at most.

C. Synthetic Vehicular Traces

1) Simulation Settings: We generate five traces of different scales with the number of vehicles changing from 100 to 300. The contact rate is set to 0.002 per minute and the contacts between two nodes follows an exponential distribution. We generate 500 data packets randomly at time 0. The number of APs and TTL are set to five and two hours, respectively. The weight of delivery ratio α in calculating utility defined in Definition 7 is set to 0.6.

2) Impact of 3G Budget: To illustrate the impact of the 3G budget on the performance, we use the trace containing 200 vehicles and vary to the number of 3G budget from 100 from 200 with an increment of 25.

Figs. 14, 15 and 16 show the delivery ratio, average delay and the utility as the function of 3G budget, respectively. From Figs. 14 and 15, we can see the delivery ratio of data packets increases with the 3G budget. The E_Random achieves the best performance, while its delay is the largest among all the other schemes. Although our algorithm does not achieve as

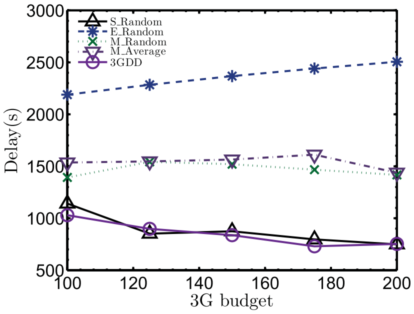


Fig. 15. Delay vs. 3G budget.

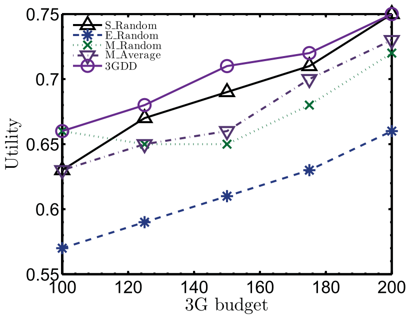


Fig. 16. Utility vs. 3G budget.

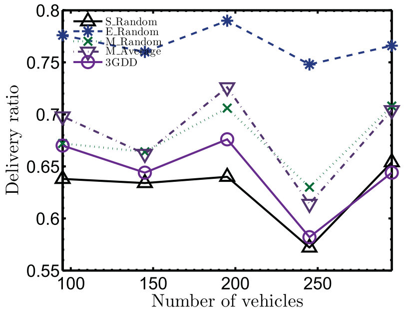


Fig. 17. Delivery ratio vs. number of vehicles.

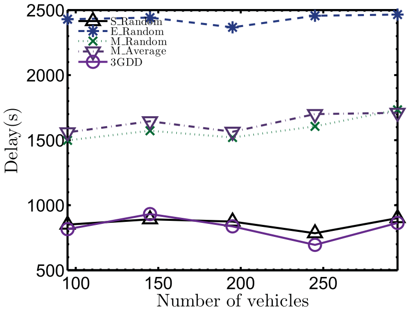


Fig. 18. Delay vs. number of vehicles.

high delivery ratio as the E_Random, but the delay is much smaller than the latter. From Fig. 16, we can find that the utility of our algorithm outperforms other schemes. This suggests our algorithm performs better when considering the delivery ratio and delay comprehensively.

3) Impact of Number of Vehicles: We change to the number of vehicles from the 100 to 300 to investigate the impacts on the algorithm performance. The 3G budget is set to 150.

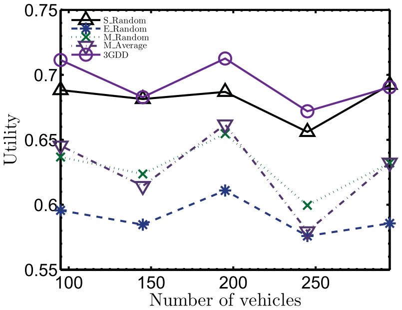


Fig. 19. Utility vs. number of vehicles.

The results are presented in Figs. 17, 18 and 19. In this set of simulation, the E_Random still achieves the highest delivery ratio, and M_Random comes the second. While their delay is much larger than our algorithm (see Fig. 18). The average delay of our algorithm is smaller the one third of the TTL. From Fig. 19, the utility of our algorithm still overcomes others. When the number of vehicles is 200, the utility of our algorithm is 16% higher than that of E_Random.

VII. CONCLUSION

Efficient data delivery is crucial for sensory data gathering applications of VANETs. However, we have observed that a noticeable percentage of data packets fail to be delivered even when there is a large number of vehicles in the network and many APs are deployed using the epidemic routing algorithm. Thus, it is highly useful to exploit 3G to further improve the data delivery performance in a VANET. In this paper we have made the first attempt to exploring the problem of 3G-assisted data delivery in VANETs. We have presented an approach called *3GDD* which first allocates the 3G budget to each time slot by solving the ILP formulation of the original optimization problem, and then selects those packets that are most unlikely delivered via VANET for 3G transmissions. The packet selection is performed before each time slot starts and this makes sure that the packet select can reflect the most update network status. Comprehensive simulations based on synthetic and real vehicular traces have been performed and comparative results show that our approach achieves better overall utility than the other alternative schemes.

REFERENCES

[1] M. Li and Y. Liu, "Rendered path: Range-free localization in anisotropic sensor networks with holes," *IEEE/ACM Trans. Netw.*, vol. 18, no. 1, pp. 320–332, Feb. 2010.

[2] K. Wu, H. Tan, H. Ngan, Y. Liu, and L. M. Ni, "Chip error pattern analysis in IEEE 802.15.4," *IEEE Trans. Mobile Comput.*, vol. 11, no. 4, pp. 543–552, Apr. 2012.

[3] F. Miao, L. Jiang, Y. Li, and Y. Zhang, "Biometrics based novel key distribution solution for body sensor networks," in *Proc. Annu. Int. Conf. Eng. Med. Biol. Soc.*, Sep. 2009, pp. 2458–2461.

[4] F. Farnoud and S. Valaee, "Reliable broadcast of safety messages in vehicular ad hoc networks," in *Proc. IEEE INFOCOM*, Apr. 2009, pp. 226–234.

[5] T. Taleb, A. Benslimane, and K. Ben Letaief, "Toward an effective risk-conscious and collaborative vehicular collision avoidance system," *IEEE Trans. Veh. Technol.*, vol. 59, no. 3, pp. 1474–1486, Mar. 2010.

[6] Y. Zhang, J. Zhao, and G. Cao, "Roadcast: A popularity aware content sharing scheme in VANETs," in *Proc. 29th IEEE Int. Conf. Distrib. Comput. Syst.*, Jun. 2009, pp. 223–230.

[7] N. Liu, M. Liu, J. Cao, G. Chen, and W. Lou, "When transportation meets communication: V2P over VANETs," in *Proc. 30th IEEE Int. Conf. Distrib. Comput. Syst.*, Jun. 2010, pp. 567–576.

[8] F. Bai, D. Stancil, and H. Krishnan, "Toward understanding characteristics of dedicated short range communications (DSRC) from a perspective of vehicular network engineers," in *Proc. 16th Annu. Int. Conf. Mobile Comput. Netw.*, 2010, pp. 329–340.

[9] Y. Zhu, Y. Bao, and B. Li, "On maximizing delay-constrained coverage of urban vehicular networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 804–817, May 2012.

[10] Z. Li, Y. Liu, M. Li, J. Wang, and Z. Cao, "Exploiting ubiquitous data collection for mobile users in wireless sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 2, pp. 312–326, Feb. 2013.

[11] N. Wisitpongphan, F. Bai, P. Mudalige, V. Sadekar, and O. Tonguz, "Routing in sparse vehicular ad hoc wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 8, pp. 1538–1556, Oct. 2007.

[12] V. Bychkovsky, B. Hull, A. Miu, H. Balakrishnan, and S. Madden, "A measurement study of vehicular internet access using in situ Wi-Fi networks," in *Proc. 12th Annu. Int. Conf. Mobile Comput. Netw.*, 2006, pp. 50–61.

[13] L. Pelusi, A. Passarella, and M. Conti, "Opportunistic networking: Data forwarding in disconnected mobile ad hoc networks," *IEEE Commun. Mag.*, vol. 44, no. 11, pp. 134–141, Nov. 2006.

[14] 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.

[15] Y. Wu, Y. Zhu, and B. Li, "Trajectory improves data delivery in vehicular networks," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 2183–2191.

[16] J. Jeong, S. Guo, Y. Gu, T. He, and D. Du, "TBD: Trajectory-based data forwarding for light-traffic vehicular networks," in *Proc. 29th IEEE Int. Conf. Distrib. Comput. Sys.*, Jun. 2009, pp. 231–238.

[17] Y. Wu, Y. Zhu, and B. Li, "Infrastructure-assisted routing in vehicular networks," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 1485–1493.

[18] I. Leontiadis, P. Costa, and C. Mascolo, "Extending access point connectivity through opportunistic routing in vehicular networks," in *Proc. IEEE INFOCOM*, Mar. 2010, pp. 1–5.

[19] I. Leontiadis and C. Mascolo, "GeOpps: Geographical opportunistic routing for vehicular networks," in *Proc. IEEE Int. Symp. World Wireless, Mobile Multimedia Netw.*, Jun. 2007, pp. 1–6.

[20] R. Ramanathan, R. Hansen, P. Basu, R. Rosales-Hain, and R. Krishnan, "Prioritized epidemic routing for opportunistic networks," in *Proc. 1st Int. MobiSys Workshop Mobile Opportunistic Netw.*, 2007, pp. 62–66.

[21] J. Nzouonta, N. Rajgure, G. Wang, and C. Borcea, "VANET routing on city roads using real-time vehicular traffic information," *IEEE Trans. Veh. Technol.*, vol. 58, no. 7, pp. 3609–3626, Sep. 2009.

[22] A. Skordylis and N. Trigoni, "Delay-bounded routing in vehicular ad-hoc networks," in *Proc. 9th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2008, pp. 341–350.

[23] J. Jeong, S. Guo, Y. Gu, T. He, and D. Du, "TSF: Trajectory-based statistical forwarding for infrastructure-to-vehicle data delivery in vehicular networks," in *Proc. IEEE 30th Int. Conf. Distrib. Comput. Syst.*, Jun. 2010, pp. 557–566.

[24] J. Zhao, Y. Zhang, and G. Cao, "Data pouring and buffering on the road: A new data dissemination paradigm for vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 56, no. 6, pp. 3266–3277, Nov. 2007.

[25] A. Benslimane, T. Taleb, and R. Sivaraj, "Dynamic clustering-based adaptive mobile gateway management in integrated VANET—3G heterogeneous wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 3, pp. 559–570, Mar. 2011.

[26] M. Mongiovi, A. Singh, X. Yan, B. Zong, and K. Psounis, "Efficient multicasting for delay tolerant networks using graph indexing," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 1386–1394.

[27] W. Gao and G. Cao, "User-centric data dissemination in disruption tolerant networks," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 3119–3127.

[28] H. Zhu, L. Fu, G. Xue, Y. Zhu, M. Li, and L. Ni, "Recognizing exponential inter-contact time in VANETs," in *Proc. IEEE INFOCOM*, Mar. 2010, pp. 1–5.

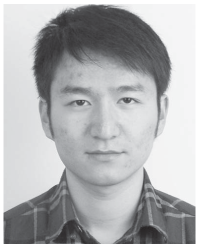
[29] F. Glover, "Tabu search—Part I," *ORSA J. Comput.*, vol. 1, no. 3, pp. 190–206, 1989.

[30] F. Glover, "Tabu search—Part II," *ORSA J. Comput.*, vol. 2, no. 1, pp. 4–32, 1990.

[31] A. Keränen, J. Ott, and T. Kärkkäinen, "The ONE simulator for DTN protocol evaluation," in *Proc. 2nd Int. Conf. Simul. Tools Tech.*, 2009, p. 55.

Qingwen Zhao received the B.S. degree from Xi'dian University, Xi'an, China, in 2012. She is currently working towards the M.S. degree with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. Her current research interests include data aggregation with vehicular networks.

Yanmin Zhu (M'08) is an Associate Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. His current research interests include vehicular networks, wireless sensor networks, and mobile computing. He is a member of the IEEE Communication Society. He received the Ph.D. degree from the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, in 2007. He was a Research Associate with the Department of Computing, Imperial College London, London, U.K.



Chao Chen received the B.S. degree from the South China University of Technology, Guangzhou, China, in 2011. He is a currently working towards the M.S. degree at the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. His current research interests include vehicular ad hoc networks and cloud computing.

Hongzi Zhu received the M.S. degree in computer science from Ji Lin University, Changchun, China, in 2004, and the Ph.D. degree from the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2009. He was a Post-Doctoral Fellow with the Department of Electric and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. He is an Assistant Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University. His current research interests include vehicular ad hoc networks, wireless networks, distributed systems, and network security. He is a member of the IEEE Computer and the IEEE Communication Society.

Bo Li (F'11) received the B.Eng. degree in computer science from Tsinghua University, Beijing, China, and the Ph.D. degree in the electrical and computer engineering from the University of Massachusetts, Amherst, MA, USA. He is a Professor with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong.