

Trajectory Improves Data Delivery in Vehicular Networks

Yuchen Wu^{*,†}, Yanmin Zhu^{*,†} and Bo Li^{*,‡}

[†] Shanghai Jiao Tong University

^{*} Shanghai Key Lab of Scalable Computing and Systems; [‡] Hong Kong University of Science and Technology
{eaufavor, yzhu}@sjtu.edu.cn; bli@cse.ust.hk

Abstract—Efficient data delivery is a great challenge in vehicular networks because of frequent network disruption, fast topological change and mobility uncertainty. The vehicular trajectory knowledge plays a key role in data delivery. Existing algorithms have largely made predictions on the trajectory with coarse-grained patterns such as spatial distribution or/and the inter-meeting time distribution, which has led to poor data delivery performance. In this paper, we mine the extensive trace datasets of vehicles in an urban environment through conditional entropy analysis, we find that there exists strong spatiotemporal regularity. By extracting mobile patterns from historical traces, we develop accurate trajectory predictions by using multiple order Markov chains. Based on an analytical model, we theoretically derive packet delivery probability with predicted trajectories. We then propose routing algorithms taking full advantage of predicted vehicle trajectories. Finally, we carry out extensive simulations based on real traces of vehicles. The results demonstrate that our proposed routing algorithms can achieve significantly higher delivery ratio at lower cost when compared with existing algorithms.

Keywords—trajectory, routing, vehicular networks, prediction, Markov chain

I. INTRODUCTION

A vehicular network is a network of vehicles which communicate with each other via wireless communications. Vehicular networks have many appealing applications such as driving safety [7], Internet access [10], intelligent transport, and infrastructure monitoring [9]. Efficient data delivery is of central importance to vehicular networks, which focuses on performance metrics including delivery ratio, delay and throughput.

Compared with traditional mobile ad hoc networks, data delivery in vehicular networks faces a set of new challenges. First, vehicular networks are subject to frequent disruptions. It is difficult to find a connected path between a pair of source and destination in vehicular networks. This is caused by high mobility and uneven distribution of vehicles over the network. Figure 1 illustrates the instantaneous network topology formed by 2500 taxis in Shanghai metropolitan city, China. Second, vehicles usually move at a high speed. Two vehicles can communicate only when they are within the communication range. Recent study has shown that the contact duration in case of a vehicle and a static access point is as short as 10 seconds on average [4].

More importantly, there is a great deal of uncertainty associated with vehicle mobility. Vehicles move at their own wills. It is difficult, if not impossible, to gain the complete knowledge about the vehicle trace of future movement. For routing in a vehicular network, a relay node must decide how long a packet should be kept and which node a given packet should be for-

warded to. Existing study [3, 11] shows that it is possible to find an optimal routing path when the knowledge of future node traces is available, which is NP-hard though. However, it is impractical to have prior knowledge about future traces of nodes.

A number of algorithms have been designed for data delivery in vehicular networks, which can loosely be divided into two categories. The first category simply assumes the availability of future movement, that is, the traces of nodes are fixed and can be known beforehand. GeOpps [15] assumes that the trace of a node can be obtained through the navigation system equipped onboard in the vehicle. Such algorithms are limited by the availability of navigation systems and the propensity of drivers [14].

The second category tries to obtain the knowledge of future mobility by predictions or estimations. It has been revealed [2] that the inter-meeting time is exponentially distributed based only on the historical meetings. Thus, the delivery delay of a packet based on the inter-meeting time distribution can be computed. MobySpace [13] maintains appearance frequencies of nodes in a given spatial place, and estimates meeting opportunities between any two nodes. In [12], Max-Contribution tries to jointly optimize links and packets scheduling, and also assumes the exponential distribution of inter-meeting times.

Thus, the knowledge of future vehicle traces plays a key role for optimal data delivery. Existing routing algorithms heavily rely on predictions of vehicle mobility. However, they have adopted only simple mobile patterns, such as the spatial distribution and inter-meeting time distribution, which support coarse-grained predictions of vehicle movements. Some algorithms [6, 13] assume random mobility in which vehicles move randomly in an open space or a road network. This model is

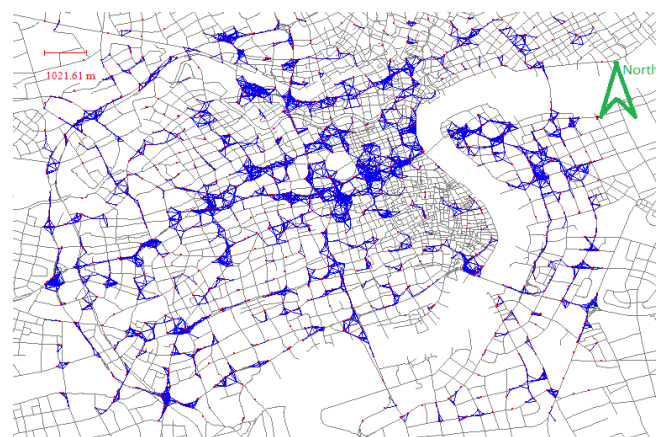


Figure 1: A snapshot of the topology formed by 2500 taxis, where the communication range is 250 meters, based on the real trace data collected in Shanghai, China on Feb 20, 2007.

simple but is far from the real mobility. Some other algorithms assume simple mobile patterns such as exponential inter-meeting times and regular spatial distributions. As a result, prediction results based on these simple patterns are of limited value to efficient data delivery in vehicular networks. In addition, many of existing algorithms ignore the fact that links in a vehicular network have unique characteristics [12]. On the one hand, a link is typically short-lived. This suggests that the capacity of the link is limited. Thus, the order for forwarding packets becomes important. On the other hand, links in a densely populated area may interfere with each other. This indicates that link scheduling becomes necessary.

To overcome the limitations in the existing algorithms, this paper proposes routing algorithms that take full advantage of trajectory predictions. By mining the extensive trace datasets of more than 4000 taxis over a duration of more than two years collected in Shanghai metropolitan city, China, we show that there is strong spatiotemporal regularity with vehicle mobility. More specifically, our results based on conditional entropy analysis demonstrate that the future trajectory of a vehicle is greatly correlated with its previous trajectory. Thus, we develop multiple order Markov chains for predicting future trajectories of vehicles. With the available future trajectories of vehicles, we propose an analytical model and theoretically derive the delivery probability of a packet.

Since the optimal routing problem with given vehicle trajectories is still NP-hard, we develop an efficient global algorithm for computing routing paths when predicted trajectories are available. For more practical situations, we develop a fully distributed algorithm which needs only localized information. The two algorithms jointly consider packet scheduling and link scheduling. We evaluate the algorithms with extensive trace driven simulations, based on the trace datasets collected in Shanghai. The results demonstrate that our algorithm considerably outperforms other algorithms in terms of delivery probability and delivery efficiency.

In summary, our major contributions in this paper are summarized as follows.

- By mining extensive trace datasets of taxis, we show that there is strong spatiotemporal regularity with vehicle mobility through conditional entropy analysis.
- We make effective trajectory predictions by developing multiple order Markov chains based on the extensive historical traces.
- We propose an analytical model and theoretically derive delivery probability with predicted vehicle trajectories.
- We propose both a global algorithm and a fully distributed algorithm. We demonstrate with trace-driven simulations that the algorithms achieve higher delivery ratio at lower cost compared with other algorithms.

The rest of the paper is organized as follows. The next section discusses related work. In Section III, we formally state the problem. In Section IV we reveal spatiotemporal regularity with vehicle mobility. Section V presents the analytical model and a global algorithm. Section VI details the design of our distributed algorithm. Performance results are presented in

Section VII. The paper is concluded in Section VIII.

II. RELATED WORK

Since the emergence of vehicular networks, data delivery in vehicular networks has been studied and a number of algorithms [5, 17, 18, 21, 23] have been proposed. Some algorithms [1, 16, 19] for delay tolerant networks may also be applied in vehicular networks. We classify existing routing algorithms into two categories. One category assumes the knowledge of pre-defined schedules or fixed traces of vehicles and the other assumes less information about future vehicle traces.

For first class of routing algorithms, there is no need to make estimation or prediction. They take advantage of navigation systems, in which vehicle traces are priori known before they begin to travel [15]. This model and assumption work well in traditional Delay Tolerant Networks, such as interplanetary networks or satellite networks. The nodes in such networks have simple and stable mobile patterns. These algorithms can only be applicable to those vehicular networks, in which drivers must tell the navigation system the destinations before their journeys and must follow the lead of the system [14].

The second class of routing algorithms makes estimations about routing metrics, since there is no further information about node future traces. Delegate Forwarding [8] demonstrates that forwarding with a metric of good quality can reduce network cost. This is done by the strategy of only forwarding packets to those nodes which lead the packet to the highest quality. RAPID [2] treats the routing of DTNs as a resource allocation problem and proposes a routing protocol to maximize the performance of a specific routing metric. The algorithms mentioned above care about calculation of metrics by estimating delay or inter-meeting time based on the simple model of exponential inter-meeting times. Max-Contribution [12] considers the joint optimization of link scheduling and packet forwarding. It illustrates that this optimization is better than former algorithms. MobySpace [13] characterizes mobile patterns with appearance frequencies of nodes over a space. The algorithm indicates that nodes with the similar frequencies of appearances have more opportunities to connect. This brings benefits to routing in DTNs.

In summary, existing algorithms either assume the availability of future traces of nodes or make coarse-grained prediction based on simple mobile patterns. We show the strong spatiotemporal regularity and predict the trajectory of a vehicle with a fine-grained mobile pattern. Routing algorithms are designed to take great advantage of predicted trajectories.

III. PROBLEM FORMULATION

A. System Model

The vehicular network is modeled as a set of nodes, $N = \{0, 1, 2, \dots, |N| - 1\}$. When two nodes, i and j , are in the communication range (denoted by D), i.e., $d_{ij} < D$, there is a link between the two nodes and they can communicate with each other while the link exists.

The position of node n at time τ is denoted by $p_n(\tau)$. The

time is slotted. Therefore, the trajectory of node n is a sequence of positions, denoted by

$$T_n = \langle p_n(0), p_n(1), \dots, p_n(\tau) \rangle. \quad (1)$$

The distance between two nodes, i and j , at time τ is denoted by $d_{ij}(\tau)$.

Links in the network are changing with the relative positions of the nodes. The set of all possible links at time τ is denoted by $L(\tau)$,

$$L(\tau) = \{l_{i,j} \mid d_{i,j}(\tau) \leq D; \forall i, j \in N\}. \quad (2)$$

Links are assumed to have the same capacity in theoretical analysis.

The vehicular network tries to deliver a set of packets, denoted by Φ , and the packets are of equal size. A packet has a source, $\delta(p)$, and a destination, $\psi(p)$. A packet is copied and forwarded from one node to another if there is a link between them. A packet group, $\theta(p)$, is introduced to denote all the copies of packet p in the network. The size of packet group, $|\theta(p)|$, increases when there is a new copy-forward process of packet p in the network.

B. Problem Formulation

The main task of data delivery is to move the packets from their sources to respective destinations. The delivery performance objectives include delivery ratio, delay and efficiency. In the following, we first analyze the properties of individual packets, and then show the objectives in a global view.

The delivery probability is the chance of a packet successfully delivered from its source to its destination. The delivery probability of packet p , denoted by ρ_p , depends on the routing strategy, Y , and the trajectories of nodes,

$$\rho_p = f_Y(\delta(p), \psi(p), \{T_n\}). \quad (3)$$

With given trajectories, ρ_p can be calculated at the beginning and it does not change over time. But in the real world where the future traces are uncertain, this delivery probability can be considered as a random variable.

The delay of a packet, denoted by σ_p , is defined as the total transmission time spent for the network to deliver the packet to its destination. Analysis of delay is similar to that of delivery probability. The delay consists of two parts. One part is the time already spent, and the other is time to spent,

$$\sigma_p(\tau) = \tau + \Gamma_Y(\delta(p), \psi(p), \{T_n\}). \quad (4)$$

When $\sigma_p(\tau) = \tau$, the delivery of the packet is complete.

The cost of the transmission of a packet, denoted by ζ_p , is defined as the total number of times that a packet is forwarded. A new copy is created when a packet is forwarded. Thus, the total number of copies indicates the delivery cost. The cost is therefore defined as

$$\zeta_p = |\theta(p)|. \quad (5)$$

One of the common objectives of vehicular networks is to maximize delivery ratio. The delivery ratio is defined as the proportion of successfully delivered packets to the total packets to be transmitted. The number of total packets is often

omitted in calculations because it is fixed and not affected by routing strategies. Thus, one objective can be given by

$$\max \sum_{p \in \Phi} \mathbb{E}[\rho_p]. \quad (6)$$

Thus, the problem is to find the optimal routing of the packets through the vehicular network that meets (6).

Two other common objectives are minimization of delay and minimization of total cost, which are given respectively by

$$\min \sum_{p \in \Phi} \mathbb{E}[\sigma_p], \text{ and } \min \sum_{p \in \Phi} \mathbb{E}[\zeta_p]. \quad (7)$$

The efficiency is defined as the ratio of total successfully delivered packets to the total cost, given by

$$\ell = \sum_{p \in \Phi} \mathbb{E}[\rho_p] / \sum_{p \in \Phi} \mathbb{E}[\zeta_p]. \quad (8)$$

A high efficiency indicates that one algorithm achieves high delivery ratio at a low cost.

C. NP Hardness

The optimal routing for achieving the objectives mentioned above is extremely difficult to design. For analysis simplification, we first assume that the routing algorithm have the complete knowledge of all vehicles and their future traces, and the set of packets to be transmitted. Note that this is impractical in the real world.

Without loss of generality, we take the objective of maximizing the delivery ratio for example. Objective (6) can be written as,

$$\max \sum_{p \in \Phi} \mathbb{E}[f_Y(\delta(p), \psi(p), \{T_n\})]. \quad (9)$$

Theorem 1: This routing problem with objective (9) when the vehicle traces and the set of packets to transmit are given is NP-hard.

The concise proof is as follows. The well studied edge-disjoint path (EDP) problem is known to be NP-hard, which tries to find the maximal number of edge-disjoint paths. By reducing the EDP problem to the ideal routing problem, we can prove the ideal routing problem is also NP-hard [2]. The basic reduction procedure is as follows. The vertices of EDP are mapped to the vehicles, and each edge is mapped to a contact of vehicles. The source destination pairs are mapped to the source and destination pairs of the packets to transfer. As a result, the routes are valid edge-disjoint paths.

In practice, the complete knowledge is hard to be obtained because the knowledge of the future is usually not available. Following this practical constraint, the trajectory of a node is divided into two parts: the past and the future, denoted by

$$\{T_n\} = \{T_n \mid_{\tau \leq t}\} \cup \{T_n' \mid_{\tau > t}\}. \quad (10)$$

Let $\lambda_t(p)$ to denote the current position of the packet. The objective at time t for the routing becomes

$$\begin{aligned} & \max \sum_{p \in \Phi} \mathbb{E}[\rho_p] \\ & = \max \sum_{p \in \Phi} \mathbb{E}[f_Y(\lambda_t(p), \psi(p), \{T_n\}, \{T_n'\})]. \end{aligned} \quad (11)$$

The second part of the trajectory is different from the first part. The first part has been fixed while the second part is not fixed and unknown at the time of being.

Since the objective of delay minimization has similar formation, we generalize (11) to the following,

$$\max \sum_{p \in \Phi, n \in N} \Theta(\lambda_i(p), \psi(p), \{T_n |_{\tau \leq t}\}, \{T_n^t\}). \quad (12)$$

where Θ denotes the global metric for the whole network. If we omit the inputs, the objective becomes

$$\max \sum_{p \in \Phi} \Theta_{p,N,t}. \quad (13)$$

This is the practical objective with the limited knowledge beyond time t .

IV. SPATIOTEMPORAL REGULARITY ANALYSIS

In this section, we quantitatively reveal the spatiotemporal regularity with vehicle mobility. The quantitative analysis is based on mining the large trace datasets of more than 4000 taxis in Shanghai, China, which have a duration of more than two years. In the datasets, a taxi periodically recorded its speed and position. The period varied from 10 seconds to several minutes on average. The taxis were operational in the whole vast area of Shanghai.

For simplification of discussion, the whole space is divided into Q small grids. Hence, the whole space is denoted by S ,

$$S = \{s_0, s_1, \dots, s_{Q-1} \mid s_i \cap s_j = \emptyset\}. \quad (14)$$

where Q is the total number of grids. The time is slotted. The location of a vehicle at a given time is considered as a random variable which takes state values from the grid space. Let S_i denote the random variable for vehicle i . We reveal the spatiotemporal regularity by computing the marginal and the conditional entropies of S_i given the previous K states.

For a vehicle i , suppose we have observed its states for L time slots. The state sequence of the vehicle can be denoted by a vector $T_i = \langle s_0, s_1, \dots, s_{L-1} \rangle$, where $s_j \in S$, $0 \leq j \leq L-1$ is the positional state of vehicle i at time slot j . Suppose that s_j appears o_j times in the vector of T_i , $0 \leq j \leq L-1$. Thus, the probability of vehicle i taking state s_j can be computed as o_j/L . Then, the marginal entropy of S_i is

$$H(S_i) = \sum_{j=0}^{Q-1} (o_j / L) \times \log_2 \frac{1}{o_j / L}. \quad (15)$$

Next, we compute the conditional entropy of S_i given its immediately previous state S_i^1 which has the same distribution with S_i . The conditional entropy is,

$$H(S_i \mid S_i^1) = H(S_i, S_i^1) - H(S_i). \quad (16)$$

To derive the conditional entropy, we have to derive the entropy of the joint random variable (S_i, S_i^1) . By using the state sequence T_i , we can obtain a derivative sequence of 2-tuples, $T_i^1 = \langle (s_0, s_1), (s_1, s_2), \dots, (s_{L-2}, s_{L-1}) \rangle$. By counting the occurrences of (s_{j-1}, s_j) , denoted by $o_{j-1,j}$, we can get its probability. Thus, the joint entropy of the two dimensional random variable of (S_i, S_i^1) is,

$$H(S_i, S_i^1) = \sum_{j=1}^{Q-1} \frac{o_{j-1,j}}{L-1} \times \log_2 \frac{1}{o_{j-1,j} / (L-1)}. \quad (17)$$

By generalizing the previous computation, we can compute the conditional entropy of S_i given its immediately previous K states $S_i^1, S_i^2, \dots, S_i^K$

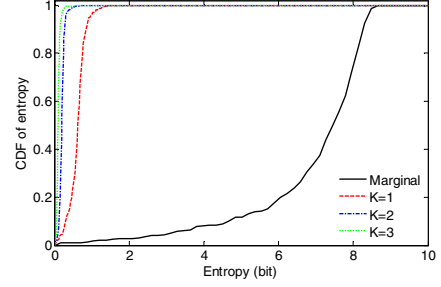


Figure 2: The CDFs of marginal entropy and conditional entropies of a vehicle positional state. Grid size is 200m×200m, and number of vehicles is 500.

$$H(S_i \mid S_i^1, S_i^2, \dots, S_i^K). \quad (18)$$

In essence, by showing the marginal entropy we reveal the spatial regularity, which characterizes the uncertainty of a vehicle residing a location in the space. By showing the conditional entropy given the previous states, we reveal the spatiotemporal regularity of vehicle mobility. This characterizes the uncertainty of a vehicle residing a specific location when its previous states are given. This demonstrates the correlation of a vehicle's positional states over different times.

In Figure 2, the cumulative distribution functions (CDFs) of the marginal and the conditional entropies for $K = 1, 2, 3$ are shown. We can see that the conditional entropies are significantly smaller than the marginal entropy. This implies that the uncertainty of the positional state becomes smaller when the previous states are known. We also find that when K becomes larger, the entropy continues to decrease. This suggests that more previous states help further reduce uncertainty. However, the improvement quickly stalls as K increases. This gives the guidance to the order selection for trajectory prediction using multiple order Markov chains.

V. GLOBAL ALGORITHM

Designing the optimal routing for achieving (13) is difficult, especially without any knowledge about future traces. In the following we propose a global routing algorithm with only predicted trajectories of vehicles.

The global metric of a packet changes after intermediate transfers over time. Consider that a packet, p , is forwarded from node i to j at time t . After the forwarding, the increment in the global metric becomes

$$\Delta \Theta_{p,N,t} = \Theta_{p,N,t}(j) - \Theta_{p,N,t}(i). \quad (19)$$

And the original objective becomes

$$\max \sum_{p \in \Phi} \left(\sum_{\tau=0}^{t-1} \Delta \Theta_{p,N,\tau} + \Delta \Theta_{p,N,t} \right). \quad (20)$$

Since the sum of increments before time t has been fixed, it can further be written as

$$\max \sum_{p \in \Phi} \Delta \Theta_{p,N,t}. \quad (21)$$

Item $\Delta \Theta$ can be considered as the current metric for packet p .

At time t , there may exist a number of links that may interfere with each other, and it is impossible for all the links to be active simultaneously. The routing algorithm must schedule

the links, i.e., choose a subset of links from all possible links to maximize the sum of current metrics. Thus, the objective becomes

$$\max \sum_{p \in \Phi, l \in L(t)} (\Delta \Theta_{p,N,t} \times l_{i,j}). \quad (22)$$

The optimal routing performs two tasks, i.e., link scheduling and packet scheduling. Before the routing decision can be made, we have to obtain the metric increment of every packet over all possible transfers. When all the increments are available, the optimal routing chooses such a set of links and a set of packet transfers that meet (22).

We should emphasize that even when all $\Delta \Theta_{p,N,t}$ are available, finding the best links and packet transfers is still an NP-hard problem. A brief proof is as follows. This problem can be considered as a weighted maximum independent set problem that has proved to be NP-hard. To solve this problem, a number of existing heuristic algorithms [22] can be used, in which $\Delta \Theta_{p,N,t}$ are considered as link weights.

As mentioned before, we take delivery probability maximization as example. The delivery probability of packet p , ρ_p , depends on the encounter probability of every two nodes, denoted by $\varepsilon_{i,j}, \forall i, j \in N$. In the following we derive the relationship between metric increment and encounter probability.

The probability for the packet to be delivered in no more than one hop is

$$\rho_p^1 = \varepsilon_{\lambda(p), \psi(p)}. \quad (23)$$

Then, for a two hop delivery with the two-hop route of $\langle \lambda(p), n_1, \psi(p) \rangle$, the probability is

$$\rho_p^{\langle \lambda(p), n_1, \psi(p) \rangle} = \varepsilon_{\lambda(p), n_1} \times \varepsilon_{n_1, \psi(p)}. \quad (24)$$

Thus, the probability for packet being delivered in two hops is

$$\begin{aligned} \rho_p^2 &= 1 - \prod_{n \in N, n \neq \lambda(p), \psi(p)} (1 - \rho_p^{\langle \lambda(p), n, \psi(p) \rangle}) \\ &= 1 - \prod_{n \in N, n \neq \lambda(p), \psi(p)} (1 - \varepsilon_{\lambda(p), n} \times \varepsilon_{n, \psi(p)}). \end{aligned} \quad (25)$$

The delivery probability of the packet with the h -hop route $\langle \lambda(p), n_1, \dots, n_{h-1}, \psi(p) \rangle$ is

$$\rho_p^{\langle \lambda(p), n_1, \dots, n_{h-1}, \psi(p) \rangle} = \varepsilon_{\lambda(p), n_1} \times \left(\prod_{i=1}^{h-2} \varepsilon_{n_i, n_{i+1}} \right) \times \varepsilon_{n_{h-1}, \psi(p)}. \quad (26)$$

Then the set of all h -hop routes (denoted by Ξ) which starts from λ and ends at ψ is denoted by

$$\begin{aligned} \Xi^h(\lambda, \psi) &= \{ \langle \lambda, n_1, \dots, n_{h-1}, \psi \rangle \mid n_i \in N; n_i \neq \lambda, \psi \}. \end{aligned} \quad (27)$$

Thus, the h -hop delivery probability is calculated by

$$\rho_p^h = 1 - \prod_{r \in \Xi^h(\lambda(p), \psi(p))} (1 - \rho_p^r). \quad (28)$$

Therefore, the total delivery probability is

$$\rho_p = 1 - \prod_{0 < h < H} (1 - \rho_p^h). \quad (29)$$

Constant H acts as the hop limit.

From the previous analysis, we can find that encounter probability is the key to computing the overall delivery probability of a packet. Since it is impossible to know future movements of vehicles, the knowledge of encounter probability of each pair of vehicles is not immediately available.

Fortunately, we have observed that there is strong spati-

otemporal regularity with vehicle mobility. Based on this observation, we propose mobile patterns to characterize this regularity. With the mobile pattern, we are enabled to predict the trajectory of a vehicle. With trajectories of vehicles, we can effectively compute the encounter probability of two vehicles.

We should stress that the encounter probability computed based on this method is more accurate than those derived from simple patterns, such as inter-meeting times and spatial distribution. Our prediction effectively makes use of both the current state information and the historical information of vehicle movement. In addition, the historical information is purely based on individual vehicles. This overcomes the problem with traditional methods that require historical information about any pair of vehicles, which introduces additional overhead.

A. Mobile Pattern

A mobile pattern of a node is to characterize the regularity of its mobility.

Definition 1. The mobile pattern of node n , denoted by MP_n , is defined as a pair of random variables, $MP_n = (\alpha, \beta)$. The probability distribution of the mobile pattern characterizes the regularity of the vehicle's mobility.

We develop the following mobile pattern for characterizing the spatiotemporal regularity of vehicle mobility. For vehicle i , its mobile pattern is defined as, $M_i = (H, F)$ where H is a vector of past positional states, $H = \langle H_0, H_1, \dots, H_{K-1} \rangle$, $K \geq 1$, and F is the future positional state. The mobile pattern characterizes the frequencies of F following H in the vehicle's traces by computing the probability distribution of the two-dimensional random variable M_i . This can be achieved by analyzing the historical traces of the vehicle, as discussed in Section III.

The notion of mobile pattern is general enough to cover the regularities that have been studied. For the study of inter-meeting times, the mobile pattern is

$$MP_n = (A, B), \quad (30)$$

where A is a random variable representing the inter-meeting time of vehicle n and a second vehicle, and B is the random variable denoting the second vehicle. Based on this mobile pattern, we are able to describe the distribution of inter-meeting times between n and any other vehicle.

For the study of spatial distribution of a vehicle, the mobile pattern is

$$MP_n = (S, \emptyset), \quad (31)$$

where S is the random variable representing the spatial state and \emptyset denotes a null random variable.

To predict the trajectory of a vehicle, existing simple patterns are inadequate. In the next subsection, we present our method of trajectory prediction by developing multiple order Markov chains based on the spatiotemporal mobile pattern.

B. Trajectory Prediction

The problem of predicting the trajectory of vehicle i at time t is to compute the future trajectory, $T_i |_{\tau > t}$, given the trajec-

tory before t , $T_i|_{\tau < t}$. A trajectory, T , can be described by a sequence of positional states,

$$T = \langle s^0, s^1, \dots, s^\tau, \dots \rangle. \quad (32)$$

Because of the uncertainty associated with vehicle mobility, there may exist a number of possible future trajectories.

Definition 2. The set of all possible trajectories of node n is defined as a trajectory bundle, denoted by TB_n , which can be characterized by

$$TB_n = \langle D^n(1), D^n(2), \dots, D^n(\tau), \dots \rangle. \quad (33)$$

where $D^n(\tau)$ is the probability distribution of spatial states at future time τ .

$$D^n(\tau) : P_s^n(\tau) | s \in S. \quad (34)$$

where $P_s^n(\tau)$ is the probability of the node n appearing in state s at time τ .

To predict the trajectory of a vehicle, it is essentially to calculate the trajectory bundle of the vehicle. For trajectory prediction, we develop multiple order Markov chains to make predictions. The key is to prepare the transition probabilities. Note that the transition probabilities are computed on an individual vehicle's basis, since different vehicle possesses different regularities.

For a vehicle, we can easily derive its transition matrix (denoted by X) from its mobile pattern. When we are using K -order Markov chain, an element, $x_{ij} \in X$, represents the transition probability from H_i to F_j , where H_i is a sequence of positional states, $H_i = \langle h_0^i, h_1^i, \dots, h_{K-1}^i \rangle$, and F_j is a single state. Then, $x_{ij} = Pr(H_i, F_j)$, where $Pr(\cdot)$ is the probability distribution function of the mobile pattern of the vehicle.

By applying the K -order Markov chain, we can compute the trajectory bundle as follows. Given the current trajectory of node n is $T_c^n = \langle s_{-K+1}, s_{-K+2}, \dots, s_0 \rangle$, the initial distributions for T_c^n are

$$D^n(\tau) : \begin{cases} P_{s_\tau}^n(\tau) = 1 \\ P_s^n(\tau) = 0, s \neq s_\tau \end{cases}, (\tau \leq 0). \quad (35)$$

The distributions of spatial state at future times can be iteratively calculated. For a single trajectory (denoted by $\langle s_1, \dots, s_K \rangle, s$), its probability is

$$P_{\langle s_1, \dots, s_K \rangle, s}^n(\tau) = x_{\langle s_1, \dots, s_K \rangle, s} \times \prod_{i=1}^K P_{s_i}^n(\tau - K + i) \quad (\tau > 0). \quad (36)$$

$P_{s_i}^n(t)$ are derived from the previous distributions and $x_{\langle s_1, \dots, s_K \rangle, s}$ is defined in the transition matrix. Then, $D^n(\tau)$ can be derived by

$$D^n(\tau) : P_s^n(\tau) = \sum_{all \langle s_i \rangle} P_{\langle s_i \rangle, s}^n(\tau), \quad \tau > 0. \quad (37)$$

C. Analysis of Delivery Probability

With the predicted trajectories, we are able to derive the encounter probabilities which are required for computing the eventual delivery probabilities.

Given the trajectory bundles of node i and j , $D^i(\tau)$ and $D^j(\tau)$ are known. Let $\varepsilon_{i,j}(\tau)$ denote the encounter probability

of the two nodes at time τ . It can be calculated by

$$\varepsilon_{i,j}(\tau) = \sum_{s \in S} P_s^i(\tau) \times P_s^j(\tau). \quad (38)$$

Then, the encounter probability, $\varepsilon_{i,j}$ is given by

$$\varepsilon_{i,j} = 1 - \prod_{\tau=t}^T (1 - \varepsilon_{i,j}(\tau)), \quad (39)$$

where T denotes the max prediction range of time.

If the overall objective is to minimize the delivery delay, the estimated encounter time for the two nodes can be derived by

$$\eta_{i,j} = \sum_{\tau=t}^T \tau \times \varepsilon_{i,j}(\tau) / \sum_{\tau=t}^T \varepsilon_{i,j}(\tau). \quad (40)$$

Therefore, the estimated delay of a packet can be calculated through trajectories in a similar way with delivery probability computation.

VI. DISTRIBUTED ALGORITHM

The global algorithm requires the complete knowledge of all packets and vehicles. More specifically, for each packet, the knowledge of its source and destination, and the current position must be known. For each vehicle, the knowledge of its position and the mobile pattern must be known. These pieces of knowledge are not available in a distributed setting. Thus, we design a practical, distributed algorithm with which a vehicle requires only limited and localized knowledge.

A. Overview

The distributed algorithm consists of two fundamental building blocks. In the first building block, the new objective for each individual node is designed, since it is impractical for individual vehicles to compute the global objective. In the second building block, we define the metadata for vehicles to exchange with each other at meetings.

In the distributed setting, the knowledge of nodes and packets is usually incomplete. Therefore, it is impossible to compute the global metric θ and the current metric $\Delta\theta$. In this case, we have to design a new metric based on which an individual vehicle makes routing decisions. Let the incomplete set of nodes and packets be denoted by Φ' and N' , respectively. Then, the new objective becomes

$$\max \sum_{p \in \Phi', p, N', t} \Delta\theta_{p, N', t}. \quad (41)$$

$\Delta\theta_{p, N', t}$ becomes the new metric, which is local and can be computed by individual vehicles for each packet. When making routing decisions, a node maximizes the sum of local metrics of all the packets it knows.

It is apparent that it would be better for a node to have more knowledge about nodes and packets. Since it is difficult for a node to have the complete knowledge, we design a distributed protocol for the nodes to exchanging information when they meet each other. By this way, the knowledge can be propagated throughout the network.

For exchanging information, we define metadata, which include two parts of information. The first part is about the mobile pattern and the most update position of each vehicle (time stamped). The second part is about the packets that the node carries. Note that for a relatively stable set of vehicles, the mobile pattern reflects the regularity of a vehicle's mobility.

Procedure 1 ($n, n', \Lambda_n, \mathcal{M}_n, \mathcal{M}_{n'}, t$)**Variables**

n : the node; n' : new neighbor; $\mathcal{M}_n, \mathcal{M}_{n'}$: metadata;
 Φ_n : packets of node n ; Λ_n : neighbors of n ;
 t : current time; N : total nodes

1. $\Lambda_n = \Lambda_n \cup \{n'\}$,
 2. n exchanges metadata with n' : $n \leftarrow \mathcal{M}_{n'}$, $n' \leftarrow \mathcal{M}_n$
 3. n calculates $\Delta\Phi_{n'} = \Phi_n - \Phi_{n'}$ from $\mathcal{M}_n, \mathcal{M}_{n'}$
 4. For every packet $p \in \Delta\Phi_{n'}$, calculate metric $\Delta\theta_{p,N,t}(n') = \rho_p(n') - \rho_p(n)$
 5. Sort all packets $p \in \Delta\Phi_i$ according to metric $\Delta\theta_{p,N,t}(i)$ for all $i \in \Lambda_n$ in the decreasing order
 6. Transmit the packets in this order
-

Figure 3: Pseudo code of the procedure of the distributed algorithm, which is executed by each vehicle each time a new neighbor is found.

Thus, it is relatively stable and therefore it is no need to update the mobile patterns frequently.

B. Algorithm Description

As a distributed algorithm, each vehicle executes the routing algorithm independently. For each vehicle, a routine procedure is invoked each time it finds a new communication neighbor. For neighbor discovery, it is required that every vehicle periodically broadcasts hello messages so that other vehicles can discover it when it enters their communication ranges.

In the following we describe the procedure, supposing that vehicle n finds another vehicle, n' , entering its communication range. The Pseudocode description of this procedure is shown in Figure 3. First, node n updates its neighbor set (Λ_n). Then, node n and n' exchange their metadata. The metadata of a node, i , denoted by \mathcal{M}_i , include two metadata sets, \mathcal{P}_i and \mathcal{V}_i . Set \mathcal{P}_i contains the metadata about the packets that node i carries, including identification and source-destination pair. Set \mathcal{V}_i contains the metadata about the vehicles known to node i , including ID, most update position and mobile pattern.

Next, it recalculates the metrics for all the packets it carries, and sort the packets respect to the new metrics in the decreasing order. The packets will then be transferred in the sorted order. Note that if the node has already been transferring a packet, the transmission of this packet is not interrupted. After its completion, the packets shall be transmitted according the new order. By ordering the packets, we are essentially doing the link scheduling in a distributed way. However, before a packet transfer can be started, a vehicle has to follow media access control protocols in order to avoid potential collisions.

As packets are transferred between vehicles, the packet set, \mathcal{P}_i , is updated whenever the vehicle receives a new packet from its neighbors.

C. Algorithm Analysis

The central part of the distributed algorithm is the procedure described above executed each time a node finds a new neighbor. The complexity of this procedure is given by the following theorem.

Theorem 2: the complexity of Procedure 1 is at most $\mathcal{O}(\phi\omega \log(\phi\omega) \times (\omega\pi^K\pi T + \omega^2\pi T + \omega^H))$, where H is the maximum hop, T is the maximum time-to-live, K is the order of Markov chain and π is the number of states ($\pi = |S|$), ω is the number of nodes (i.e., $\omega = |N|$), and ϕ is the size of the total packets.

Proof. The complexity of the procedure is mainly in calculation of the local metrics for each packet. The complexity of calculating the trajectory bundle of one node is $\mathcal{O}(\pi^K T)$, for all nodes requires $\mathcal{O}(\omega\pi^K T)$. The complexity of calculating encounter probability between the node and another node is $\mathcal{O}(\pi T)$, all node pairs require $\mathcal{O}(\omega^2\pi T)$. For all the nodes, the total complexity is $\mathcal{O}(\pi^K T + \omega^2\pi T)$. Calculating the delivery probability requires to use at most H nodes, and therefore this has a complexity of $\mathcal{O}(\omega^H)$. In total, the calculation has the complexity of $\mathcal{O}(\omega\pi^K T + \omega^2\pi T + \omega^H)$. The number of metrics that the procedure needs to compute depends on both the size of packet set and the size of neighbor set. The size of the packet set is at most ϕ , and the size of the neighbor set is at most ω . The sorting has a complexity of $\mathcal{O}(\phi\omega \log(\phi\omega))$. Thus, the total complexity is given by $\mathcal{O}(\phi\omega \log(\phi\omega) \times (\omega\pi^K T + \omega^2\pi T + \omega^H))$. \square

D. Discussions

The distributed algorithm requires the vehicles to exchange metadata about mobile patterns of vehicles. On the one hand, the set of vehicles may change over time. However, we observe that in an urban environment, the set of vehicles is relatively stable. In our system, for example, taxis in Shanghai form the vehicular network, which does not change often. On the other hand, the mobile pattern of a vehicle may also change over time. Nevertheless, the regularity of a vehicle is also stable. More importantly, our distributed algorithm supports updating of mobile patterns. This is achieved by each vehicle propagating its new pattern. Receiving an updated pattern, other vehicles can update the vehicle's pattern accordingly.

The distributed algorithm also requires the most update location of vehicles. The ideal case is that each vehicle can have the current locations of all vehicles, e.g., by a third channel. However, this may introduce a prohibitive cost. By propagating the most update locations in vehicular networks, the location knowledge about other vehicles maintained by a vehicle may be time lagged. The time lagged location may degrade the prediction quality, but predicted trajectories are still of value to routing decisions.

VII. PERFORMANCE EVALUATION

We perform trace-driven simulations and compare the performance of our algorithm with that of other algorithms.

A. Methodology and Settings

We evaluate our algorithms with the performance metric of delivery ratio, delay, cost and efficiency. These metrics have been defined in Section III. We compare our algorithms with several algorithms, which will be introduced shortly.

The simulations are conducted with the real trace data of more than 4000 taxis collected in Shanghai over a duration of

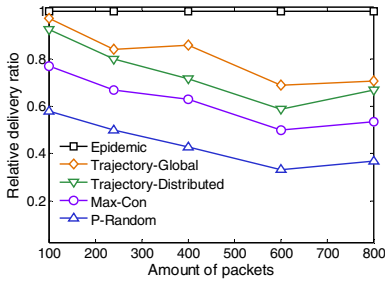


Figure 4: Delivery ratio vs. amount of packets

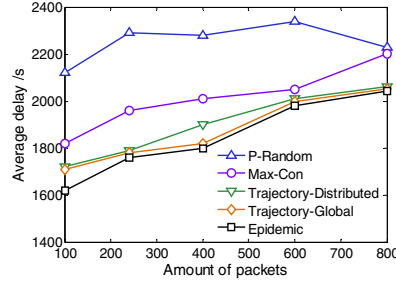


Figure 5: Average delay vs. amount of packets

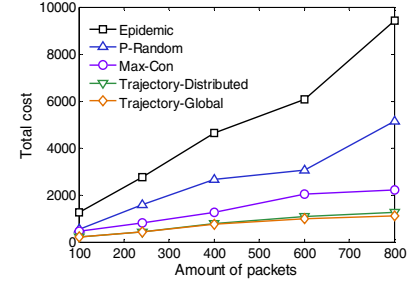


Figure 6: Total cost vs. amount of packets

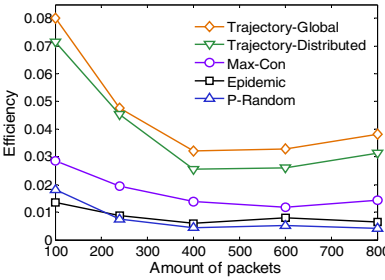


Figure 7: Efficiency vs. amount of packets

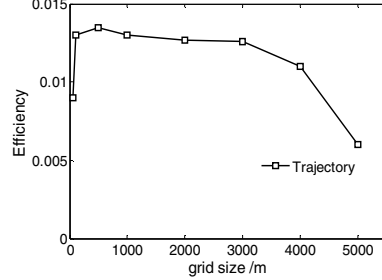


Figure 8: Efficiency vs. grid size

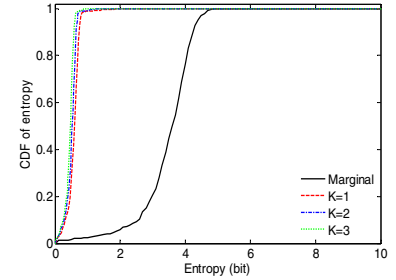


Figure 9: Entropy CDFs (grid size=3 km)

two years. The whole urban area of Shanghai is 133 kilometers in length and 69 kilometers in width. The whole space is divided into grids, and the grid size is three kilometers by default. The communication range is 300 meters and the link bandwidth is set according to IEEE 802.11b, dependent on distances. We consider link interfaces, and each node can communicate with one neighbor at any time. We select a subset of 400 taxis from the complete trace for simulations.

For each packet, we randomly select its source and destination. The packets are injected at different times. Every packet has the same size and priority. The maximum hop and the maximum time-to-live (TTL) of each packet is set to 20 and 2 hours, respectively. The number of packets is varied from 100 to 800 to study different loads of the network.

The order of Markov chains K is set to two. From conditional entropy analysis, we have already found that a higher order beyond two gives little reduction in uncertainty. In addition, a higher order results a higher algorithm complexity.

B. Compared Algorithms

Flooding [20], also known as epidemic routing, is a simple algorithm. Each node forwards all the packets it carries to any node it meets. This algorithm provides an upper bound on delivery ratio and a lower bound on delivery delay. It introduces very high cost, which is the major defect.

P-Random [11] is an opportunistic routing algorithm, which randomly decides whether to forward a packet to another node. In simulations, the probability is set to 0.4. This algorithm represents the algorithms without predictions.

Max-Contribution [12] is a routing algorithm using a simple mobile pattern that only considers the inter-meeting times of exponential distribution. This pattern makes prediction independent of the current location of a vehicle.

C. Results

We first present performance results of our algorithms against other algorithms. We then study the impact of grid size.

We vary the amount of packets and compare the five algorithms in terms of delivery ratio, average delay and total cost under different loads of network. For a given setting of amount of packets, the same set of packets and the set of nodes are used for all the five algorithms.

Since the delivery ratio is largely affected by the time length simulated, we use the metric of relative delivery ratio instead of bare delivery ratio, which is the delivery ratio of each algorithm normalized by that of Flooding. In Figure 4, the performance of the five algorithms in terms of relative delivery ratio is shown. We can see that our algorithms performs better than P-Random and Max-contribution. Among the five algorithms, Flooding performs the best, as expected. In general, the algorithms that use predictions produce better delivery ratios than the algorithms that make no predictions. Our algorithms are better than Max-Contribution because our algorithms use not only the historical trace information but also the current state and the previous states. Predicted trajectories of vehicles help to find better routing paths. The global algorithm is better than the distributed one since it has the global, complete network information. We can also find that when the amount of packets increases, the overall delivery ratios of all algorithms decrease. The reason is that the overall capacity of the network is limited. By injecting more packets, the packets may compete for network resources and as a result, fewer packets can be delivered in the end.

In Figure 5, average delay against amount of packets is plotted. Our algorithms have a lower delay than P-Random and Max-Contribution. As expected, Flooding has the smallest delay. The average delays of our algorithms are slightly larger

than that of flooding. This performance gain of our algorithms is mainly due to the fact that routing paths with high delivery probabilities usually lead to shorter delays. Since our algorithms can select routing paths of high delivery probabilities, the resultant delay is low.

Figure 6 shows the costs of the five algorithms. We can find that our algorithms have lower costs, better than all the other algorithms. The main reason is that by effectively predicting the trajectories of vehicles, we only consider the paths that lead to eventual delivery with high probability. Therefore, unnecessary packet transfers are greatly reduced.

In Figure 7, we compare the efficiencies of the five algorithms. We can see that our algorithms have higher efficiency, better than all the other algorithms. Flooding and P-Random have a similar efficiency and are much worse than the rest three. This is due to the blindness of Flooding and P-Random when they are making transfer decisions. Max-Contribution has a larger efficiency than Flooding and P-random, but has a lower efficiency than our algorithms, since Max-Contribution merely uses a simple mobile pattern of inter-meeting time.

The grid size influences prediction and in turn the algorithm performance. If we use a greater grid size, two vehicles in the same grid would be considered to encounter with each other but in reality they do not. But meanwhile, a greater grid size leads to a lower algorithm complexity. We study the impact of grid size on delivery performance. In Figure 8, the efficiency of the distributed algorithm is plotted against different grid sizes. We find that in general a smaller grid results in a better efficiency. When the grid size increases, the efficiency decreases. The reason is that when the grid size is larger, the trajectory prediction becomes worse. Figure 9 shows the entropy CDFs when the grid size is 3 km. By comparing them with the CDFs in Figure 2, we can find that a greater grid size results in a smaller gap between the marginal and the conditional CDFs. This suggests that a greater grid size reduces the efficacy of trajectory prediction.

VIII. CONCLUSION

Vehicular networks have received substantial attention recently. Although data delivery of vehicular networks has been studied, few exiting algorithms effectively exploit trajectories of vehicles. In this paper, we demonstrate the strong spatiotemporal regularity with vehicle mobility by entropy analysis. By developing multiple order Markov chains, we predict vehicle trajectories. Based on the analytical model, we derive delivery probability with the predicted trajectories. The proposed algorithms take full advantage of vehicle trajectories. Performance results verify that our algorithm outperforms other algorithms. This demonstrates that predicted trajectories do help data delivery in vehicular networks.

ACKNOWLEDGEMENT

This research is supported by NSFC (No. 60903190, 60970106, 60673166 & 60803124), Shanghai Pu Jiang Talents Program (10PJ1405800), Shanghai Chen Guang Program (10CG11), 973 Program (2005CB321901), MIIT of China (2009ZX03006-001-01 & 2009ZX03006-004), Doctoral Fund

of Ministry of Education of China (20100073120021), 863 Program (2009AA012201), HK RGC (615608) and Huawei Technologies Co. Ltd. (HUAW18-15L0181011/PN). In addition, it is partially supported by the Open Fund of the State Key Laboratory of Software Development Environment (Grant No. SKLSDE-2010KF-04), Beijing University of Aeronautics and Astronautics.

REFERENCES

- [1] E. Altman, F. De Pellegrini, and L. Sattatelli, "Dynamic control of coding in delay tolerant networks," *Proc. IEEE INFOCOM*, 2010.
- [2] A. Balasubramanian, B. Levine, and A. Venkataramani, "DTN routing as a resource allocation problem," *Proc. ACM SIGCOMM*, 2007.
- [3] J. Burgess, B. Gallagher, D. Jensen, and B. Levine, "Maxprop: Routing for vehicle-based disruption-tolerant networks," *Proc. IEEE INFOCOM*, 2006.
- [4] V. Bychkovsky, B. Hull, A. K. Miu, H. Balakrishnan, and S. Madden, "A measurement study of vehicular Internet access using In situ Wi-Fi networks," *Proc. ACM MobiCom*, 2006.
- [5] B. Chen and M. Chan, "MobTorrent: A framework for mobile Internet access from vehicles," *Proc. IEEE INFOCOM*, 2009.
- [6] Z. Chen, H. Kung, and D. Vlah, "Ad hoc relay wireless networks over moving vehicles on highways," *Proc. ACM Mobihoc*, 2001, p. 250.
- [7] L. Chisalita and N. Shahmehri, "A peer-to-peer approach to vehicular communication for the support of traffic safety applications," *Proc. 5th IEEE Conference on Intelligent Transportation Systems, Singapore*, 2002, pp. 336-341.
- [8] V. Erramilli, M. Crovella, A. Chaintreau, and C. Diot, "Delegation forwarding," *Proc. ACM Mobihoc*, 2008, pp. 251-260.
- [9] P. Gibbons, B. Karp, Y. Ke, S. Nath, and S. Seshan, "Irisnet: An architecture for a worldwide sensor web," *IEEE Pervasive Computing*, pp. 22-33, 2003.
- [10] H. Hartenstein, B. Bochow, A. Ebner, M. Lott, M. Radmirsch, and D. Vollmer, "Position-aware ad hoc wireless networks for inter-vehicle communications: the Fleetnet project," *Proc. ACM Mobihoc*, 2001, p. 262.
- [11] S. Jain, K. Fall, and R. Patra, "Routing in a delay tolerant network," *Proc. ACM SIGCOMM* 2004.
- [12] K. Lee, Y. Yi, J. Jeong, H. Won, I. Rhee, and S. Chong, "Max-contribution: On optimal resource allocation in delay tolerant networks," *Proc. IEEE INFOCOM*, 2010.
- [13] J. Leguay, T. Friedman, and V. Conan, "DTN routing in a mobility pattern space," *Proc. ACM SIGCOMM*, 2005, p. 283.
- [14] I. Leontiadis, P. Costa, and C. Mascolo, "Extending access point connectivity through opportunistic routing in vehicular networks," *Proc. IEEE INFOCOM*, 2010.
- [15] I. Leontiadis and C. Mascolo, "Geopps: Geographical opportunistic routing for vehicular networks," *Proc. IEEE WoWMoM*, 2007, pp. 1-6.
- [16] Q. Li, S. Zhu, and G. Cao, "Routing in socially selfish delay tolerant networks," *Proc. IEEE INFOCOM*, 2010.
- [17] B. Liu, B. Khorashadi, D. Ghosal, C. Chuah, and M. Zhang, "Assessing the VANETs local information storage capability under different traffic mobility," *Proc. IEEE INFOCOM*, 2010.
- [18] V. Naumov and T. Gross, "Connectivity-aware routing (car) in vehicular ad-hoc networks," *Proc. IEEE INFOCOM*, 2007, pp. 1919-1927.
- [19] S. Nelson, M. Bakht, R. Kravets, and A. Harris III, "Encounter: based routing in DTNs," *Proc. IEEE INFOCOM*, 2009.
- [20] R. Ramanathan, R. Hansen, P. Basu, R. Rosales-Hain, and R. Krishnan, "Prioritized epidemic routing for opportunistic networks," *Proc. ACM MobiSys workshop on Mobile Opportunistic Networks (MobiOpp)*, 2007, p. 66.
- [21] N. Wisitpongphan, F. Bai, P. Mudalige, and O. Tonguz, "On the routing problem in disconnected vehicular ad-hoc networks," *Proc. IEEE INFOCOM*, 2007.
- [22] Y. Yi, A. Proutière, and M. Chiang, "Complexity in wireless scheduling: Impact and tradeoffs," *Proc. ACM Mobihoc*, 2008, pp. 33-42.
- [23] Z. Zheng, Z. Lu, P. Sinha, and S. Kumar, "Maximizing the contact opportunity for vehicular Internet access," *Proc. IEEE INFOCOM*, 2010.