This paper was presented as part of the main technical program at IEEE INFOCOM 2011

Exploiting Temporal Dependency for Opportunistic Forwarding in Urban Vehicular Networks

Hongzi Zhu[1, 2], Shan Chang[3], Minglu Li[2], Kshirasagar Naik[1] and Sherman Shen[1]
[1] Department of ECE, University of Waterloo, Canada
[2] Department of CSE, Shanghai Jiao Tong University, China
[3] Department of CS, Xi'an Jiaotong University, China
{hongzi, xshen}@bbcr.uwaterloo.ca, cschangshan.07@stu.xjtu.edu.cn, mlli@sjtu.edu.cn, snaik@uwaterloo.ca

Abstract—**Inter-contact times (ICTs) between moving vehicles are one of the key metrics in vehicular networks, and they are also central to forwarding algorithms and the end-to-end delay. Recent study on the tail distribution of ICTs based on theoretical mobility models and empirical trace data shows that the delay between two consecutive contact opportunities drops exponentially. While theoretical results facilitate problem analysis, how to design practical opportunistic forwarding protocols in vehicular networks, where messages are delivered in carry-and-forward fashion, is still unclear. In this paper, we study three large sets of Global Positioning System (GPS) traces of more than ten thousand public vehicles, collected from Shanghai and Shenzhen, two metropolises in China. By mining the temporal correlation and the evolution of ICTs between each pair of vehicles, we use higher order Markov chains to characterize urban vehicular mobility patterns, which adapt as ICTs between vehicles continuously get updated. Then, the next hop for message forwarding is determined based on the previous ICTs. With our message forwarding strategy, it can dramatically increase delivery ratio (up to 80%) and reduce end-to-end delay (up to 50%) while generating similar network traffic comparing to current strategies based on the delivery probability or the expected delay.**

Keywords-Inter-contact time; vehicular networks; temporal dependency; opportunistic forwarding; Markov chain

I. INTRODUCTION

Vehicular networks are emerging as a new landscape of mobile ad hoc networks, aiming to provide a wide spectrum of safety and comfort applications to drivers and passengers. In vehicular networks, vehicles equipped with wireless communication devices can transfer data with each other (inter-vehicle communications) as well as with the roadside infrastructure (vehicle-to-roadside communications). In order to successfully transfer data from a vehicle to another, the vehicle needs to first wait until it geographically "meets" other vehicles (i.e., within the communication range of each other) for data-relay. Data transfer, therefore, arises in a *store-carry-forward* fashion. Applications based on this type of data transfer strongly depend on vehicular mobility characteristics, especially on how often such contact opportunities take place and on how long they last. The delay between two consecutive contacts (referred to as *inter-contact time*) of the two vehicles is a major component of the end-to-end delay, as it represents how long it takes for this vehicle to encounter the other vehicle for data transmission. Larger inter-contact time (ICT) results in larger end-to-end delay.

In the literature, several opportunistic message forwarding protocols in intermittently connected mobile ad hoc networks (MANETs) and delay tolerant networks (DTNs) have been proposed. When the future node movement is known in advance, data forwarding path can be calculated based on the knowledge of network topology and workload of traffic [1]. In reality, it is often the case that the information of future movement is unavailable. However, when node mobility is not completely random, it is possible to make forwarding decision based on mobility history [2] [3] [4], such as past records of connection opportunities with other nodes, node motility patterns, rate of change of connectivity and friendship index with other nodes. Random walks [5] [6] and epidemic routing [7] [8] require no history information while conducting data forwarding. In a random walk, a node randomly selects a neighbor as the next hop to carry a message. Using random walks generates very moderate network traffic but tends to show very large end-to-end delay. On the contrary, epidemic routing can achieve the minimum end-to-end delay by flooding a message throughout the network, which introduces unacceptable network overhead.

Recently, there have been several studies on analyzing mobility characteristics based on empirical trace data collected from urban areas [9] [10] and public transportation systems [11] [12]. These studies mainly focus on the distribution of ICTs, having the observation that vehicles in urban environments tend to meet very frequently. They demonstrate the tail distribution of ICTs can decay exponentially fast. Although the exponential distribution facilitates the problem analysis on the performance bound of routing algorithms, it is not clear *how to design a practical opportunistic forwarding algorithm utilizing the characteristics of inter-contact times?*

In this paper, we take a data-driven approach in designing and evaluating our opportunistic forwarding algorithm in urban vehicular networks. Extensive GPS trace data collected from more than *ten thousand* public vehicles (taxies and buses) in Shanghai and Shenzhen, two metropolises in China, are mined. Specifically, we analyze more than 45 million pairwise contacts resolved from the trace to characterize the contact interaction among vehicles. By studying the distribution of ICTs, in addition to the exponential tail distribution, we find that the layout of ICTs also demonstrates an apparent pattern: if a vehicle meets another vehicle at certain time, the probability that the two vehicles meet again at the same time in the following days is very high. With this observation, we characterize the temporal correlation of ICTs and then capture those characteristics with higher order Markov chain models. We then design an opportunistic forwarding algorithm exploiting the temporal dependency of ICTs. In our algorithm, a vehicle estimates the expected delay between a neighboring vehicle and the destination of a message, based on their previous ICTs. If this vehicle has smaller estimation, it forwards the message for data-relay. The goal of our algorithm is twofold: first, we concern the delivery performance in vehicular networks, trying to minimize the end-to-end delay and maximize the delivery ratio; second, since vehicles communicate via wireless channels, we try to minimize the network overhead for data transmission. We verify the performance of our algorithm through extensive trace-driven simulations. The results show that we can achieve comparable delivery performance as epidemic routing in terms of end-to-end delay and delivery ratio with a very moderate network overhead. Compared with current message forwarding strategy based on the delivery probability or the expected delay, our algorithm can dramatically increase 84% delivery ratio and reduce 53% end-to-end delay while generating similar network traffic.

The remainder of this paper is organized as follows. Section II presents the related work. In Section III, we describe the characteristics of the GPS trace data and the distribution of the ICTs resolved. Section IV presents the temporal correlations of ICTs. We develop our opportunistic forwarding algorithm in Section V. Section VI describes the methodology to evaluate the performance our message forwarding algorithm and presents the results. Finally, we give concluding remarks and outline the directions for future work in Section VIII.

II. Related Work

In intermittently connected MANETs and DTNs, where moving objects communicate in a store-carry-forward fashion, the mobility characteristics of objects are central to forwarding algorithms and the ultimate performance in terms of end-to-end delay and delivery ratio. Based on the available knowledge about the movement of objects, data forwarding algorithms in these networks can be divided into two basic categories: knowledge-based and non-knowledge-based.

In the knowledge-based category, there are several methods available to estimate the end-to-end path delay when the future movement of nodes is known ahead of time. For example, S. Jain et al [1] discuss the path selection algorithms according to how much knowledge about the network topology and network traffic workload being known. The path delay can be calculated as the sum of the expected delay of each hop on this path. A recursive process is deployed in [13] to calculate the minimum end-to-end delivery delay, assuming that the tail distribution of ICTs is exponential and ICTs are independent. In reality, however, it is often the case that information about the future movement of moving objects is unavailable. A number of utility-based routing schemes [14] [15] have been proposed for data forwarding based on node history mobility information, such as the contact records, mobility patterns and the rate of connectivity change. In these schemes, a utility function is defined and measured for every other node in the network. If the current message carrier meets a node with a higher utility, the message is forwarded to this node. Our algorithm fits in this category. In our algorithm, instead of examining the node mobility or pairwise contact patterns, we study the effect of the temporal dependency of ICTs to the delivery performance in vehicular networks, which we believe is one of the first reports analyzing the characteristics of ICTs from thousands of real urban vehicles.

In the non-knowledge-based category, without requiring any information, random walks [5] [6] can be used for data-relay. For a random walk, a node randomly selects a neighbor as the next hop to carry a message. Using random walks generates very moderate network traffic but tends to have very large end-to-end delay. By performing multiple walks, both delivery ratio and end-to-end delay can be improved. An extreme case of this is epidemic routing [7] [8], where a message is flooded throughout the network. Using epidemic routing can achieve the minimum end-to-end delay and maximum delivery ratio but generates unacceptable network overhead at the same time. Techniques such as limiting the number of duplicated copies of a message, setting a living timeout for packets and forwarding to selected neighbors can be used to reduce the overhead of epidemic routing.

III. Empirical Vehicular Data

A. Collecting Urban Vehicular Trace Data

In order to design message forwarding algorithms between urban vehicles, it is of great importance to study the empirical data on the frequency, duration and temporal distribution of contacts between them. Ideally, such a data set would contain a large number of vehicles over a sufficient long period of time, as well as include all connection opportunity information twenty-four hours a day with a fine granularity measured in seconds.

We collect three sets of GPS traces (partially available at *http://www.cse.ust.hk/scrg*) of more than 10 thousands of public vehicles, namely taxies and buses, in Shanghai and Shenzhen, two metropolises in China. As a bus or a taxi commutes in the city, it periodically sends reports back to a data center via an on-board GPS-enabled device (an example taxi in Shanghai is shown in Fig. 1). The specific information contained in such a report includes: the vehicle's ID, the longitude and latitude coordinates of the current location, timestamp, moving speed, heading and operational status (i.e., whether a taxi has passengers onboard or a bus is arriving at a bus stop). Due to the GPRS communication cost for data transmission, reports are sent at a granularity of around one

Figure 1. A taxi with a commercial GPS device installed (highlighted in the inset), the location and operational information thus can be preriodically sent back via GPRS wireless channels.

minute. The specific characteristics of the three data sets are shown in Table I.

Definitely, there are various types of vehicles in urban environments, each of which has particular mobility patterns. We choose taxies and buses to study for two reasons. First, taxies and buses shows two distinct mobility patterns, namely, rather random and well scheduled, respectively. They are quite representative for a large variety of other vehicles in a city. For example, a private vehicle may follow a nearly precise route and schedule traveling back and forth between work and home on weekdays (bus-like behavior) whereas it is more interest-driven (e.g., towards a park for picnic, a friend's home for gathering, and a mall for shopping) on weekends or holidays (taxi-like behavior). Second, the privacy problem is less concerned since they are public vehicles. As privacy preservation schemes progress and more mobility data of private vehicles available, it is more practical to study private vehicles in the future.

B. Statistics of Inter-Contact Time

1) *Extraction of Inter-Contact Time from Trace Data:* Since GPS reports are sent in discrete time, usually on one minute, we use a sliding time window to check contacts between a pair of taxies as introduced in our pervous work [12]. Here we make the assumption that two vehicles would have a connection opportunity (called a *contact*) if their locations reported within a given time window are within the communication range. Althogh the inaccuracy may be introduced by this assumption and contact extraction algorithm, the essential vehicular mobility characteristics are

TABLE I. COMPARISON OF THERE DATA SETS

Data Set	*Shanghai Bus*	*Shanghai Taxi*	*Shenzhen Taxi*
Number of vehicles	2,501	2,109	8,291
From date	Feb. 19, 2007	Feb. 1, 2007	Oct. 1, 2009
Duration (day)	15	31	31
Granularity (second)	60	15*, 60**	60
Number of contacts	1,229,380	22,053,178	23,968,860
Mean ICT (minute)	31.8	47.6	30.5

*vacant, **passengers onboard

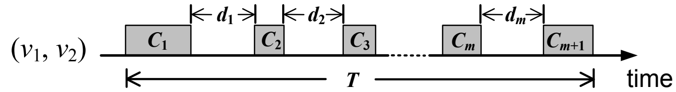

Figure 2. An example of contacts and inter-contact times between a pair of vehicles v_1 and v_2.

preserved and therefore the results are very valuable for study.

We refer to *inter-contact time* as the time elapsed between two successive contacts of the same vehicles as defined in [16] [17] [18]. Specifically, the inter-contact time is computed at the end of each contact, as the time period between the end of this contact and the beginning of the next contact between the same two vehicles. For example, in Fig. 2, inter-contact time d_1 can be computed as the starting time of contact C_2 minus the end time of its previous contact C_1.

2) *Inter-Contact Time Distribution Characteristics:* We apply the above contact extraction algorithm with a time window of one minute and a communication range of 100 meters to each pair of vehicles in all three data sets, respectively (basic statistics are shown in Table I). We plot the tail distribution (CCDF) of inter-contact time over time in linear-log scale in Fig. 3. The linear delay of all plots in linear-log scale indicates that the tail distribution of inter-contact time between vehicles drops exponentially. The reason that the ICT tail distribution of vehicles is exponential rather than power law as found in human mobility [16] might be that traffic tends to converge around certain areas in the urban settings, such as the underlying topology of road networks and distribution of residential areas, shopping centers and commercial zones, which enormously increases contact opportunities of vehicles [12]. The exponential distribution implies, to some extent, vehicles meet each other in urban settings very frequently. While exponential distribution is convenient for the problem analysis, we are athirst for the answer to the following question: *how to design a practical opportunistic forwarding algorithm utilizing inter-contact time distribution characteristics?*

To answer the question, it is not enough knowing only the frequency of connection opportunities but particularly the temporal layout or patterns between each inter-contact time within the distribution. Therefore, we examine the probability density function (PDF) of inter-contact time as shown in Fig. 4. It is easy to notice an apparent pattern that the probability reaches local maxima when the length of an inter-contact time equals an integral multiple of one day. This indicates that if a vehicle meets another vehicle at certain time the probability that the two vehicles meet again at the same time in the following days is very high. The reason can be explained as follows. Buses can constantly encounter with each other since they have dedicated routes and schedules. Intuitively, taxies behave rather randomly and have higher mobility than buses. Nevertheless, taxi drivers also have their own preferences in choosing serving areas and path planning. Evidence shows that other vehicles of different kinds in urban settings also

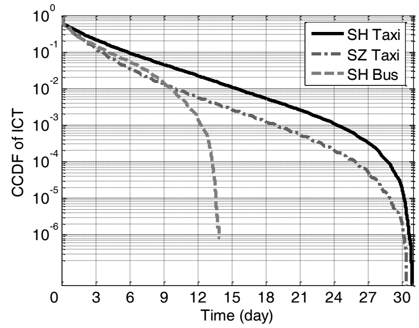

Figure 3. Tail distribution of the inter-contact time of urban public vehicles in Shanghai (SH) and Shenzhen (SZ).

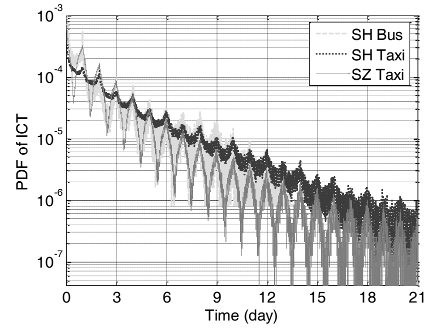

Figure 4. Probability density function of the inter-contact time of the same experimental vehicle sets.

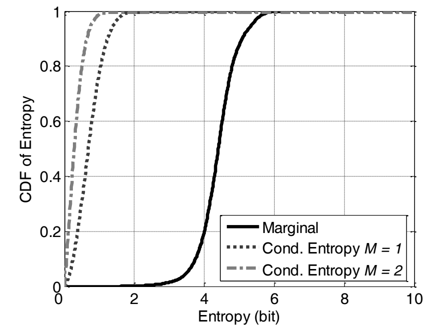

Figure 5. CDFs of marginal entropy and conditional entropy of inter-contact times between each pair of taxies in Shanghai data set.

demonstrate strong mobility patterns during daily routines [19], which constitute regular connection opportunities. In other words, *temporal dependency* of inter-contact time does exist between two vehicles in urban vehicular networks.

IV. Analyzing ICT Temporal Patterns

In this section, we examine two specific questions: 1) how historical inter-contact time information is related to the current inter-contact time; and 2) how inter-contact time patterns evolve over time and how much historical information we need to track to capture the inter-contact time patterns over time.

A. Characterizing Temporal Correlations of Successive ICTs

We examine the correlation between inter-contact times by computing the marginal entropy of inter-contact times between each pair of vehicles and the conditional entropy of the inter-contact times between a pair of vehicles given their previous M inter-contact times in all of the three data sets.

Although an inter-contact time can be infinitely long in time, due to the fast exponential decay of inter-contact time tail distribution, most inter-contact times are less than a relatively short period of time. For example, in Fig. 3, more than 90% inter-contact times are less than six days. Therefore, an inter-contact time $\mathcal{T}$ can be specialized into a discrete finite value space as,

$$\mathcal{T}' = \begin{cases} \lfloor \mathcal{T}/\lambda \rfloor, & if\ \mathcal{T} < \mathbb{T} \\ \lfloor \mathbb{T}/\lambda \rfloor, & otherwise \end{cases}, \tag{1}$$

where $\mathbb{T}$ is the maximum inter-contact time, and λ is the counting measure. In the rest part of this paper, without explicit specification, inter-contact times are referred to as their specialized counterparts.

Let X be the random variable representing the inter-contact times between a pair of vehicles. If we have observed N inter-contact times between this pair of vehicles, these inter-contact times can be presented by a vector $T = (t_0, t_1, \cdots, t_{N-1})$ where $t_i \in [0, \lfloor \mathbb{T}/\lambda \rfloor]$, $0 \le i \le N-1$ denotes the i^{th} inter-contact time. Assume each of these inter-contact times appeared x_j times in T, $0 \le j \le \lfloor \mathbb{T}/\lambda \rfloor$. Thus, the probability of the inter-contact time being j can be computed as x_j/N. Therefore, the entropy of T is:

$$H(X) = \sum_{j=0}^{\lfloor \mathbb{T}/\lambda \rfloor} (x_j/N) \log_2 \frac{1}{x_j/N}, \tag{2}$$

For $M = 1$, let X' be the random variable for the immediately previous inter-contact time between this pair of vehicles given the inter-contact time X. X' and X have the same distribution when N is large enough. The vector T can be written as $Q = \{(t_i, t_{i+1}): 0 \le i \le N-2\}$. Therefore, the joint entropy of X' and X can be computed as:

$$H(X', X) = \sum_{(x', x) \in Q} P(x', x) \log_2 \frac{1}{P(x', x)}, \tag{3}$$

where $P(x', x)$ is the number of times (x', x) appearing in Q divided by the total number of elements in Q. With $H(X)$ and $H(X', X)$, the conditional entropy of X given X' is:

$$H(X|X') = H(X', X) - H(X') = H(X', X) - H(X) \tag{4}$$

For $M = 2$, let X'' denote the random variable representing the distribution of the previous two ICTs given X. Similarly, the conditional entropy $H(X|X'')$ is:

$$\begin{aligned} H(X|X'') &= H(X'', X) - H(X'') \\ &= H(X'', X) - H(X', X), \end{aligned} \tag{5}$$

The joint entropy $H(X'', X)$ can be calculated similarly.

Fig. 5 shows the CDFs of the mean entropy and the mean conditional entropy, for $M = 1$ and 2, over each pair of taxies in the Shanghai data set. It can be seen that the conditional entropy for $M = 1$ is much smaller than the marginal entropy, and that the conditional entropy for $M = 2$ is smaller than that for $M = 1$. This implies that the uncertainty about the inter-contact time decreases when knowing the previous inter-contact times between the same pair of taxies.

We further examine the entropy and conditional entropy for vehicles in all data sets. Fig. 6 shows the results for marginal entropy and conditional entropy when $M = 1$. It is

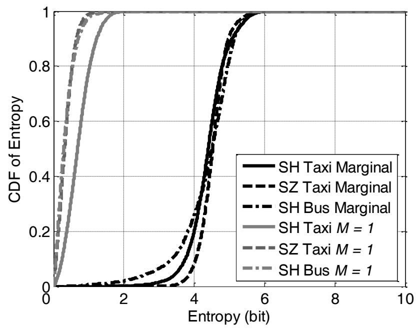

Figure 6. CDFs of marginal entropy and conditional entropy of inter-contact times between each pair of taxies in all data sets.

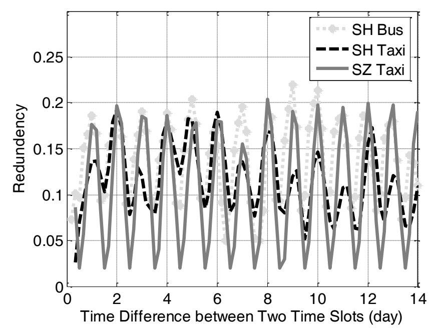

Figure 7. Mean redundancy of the layout of inter-contact times between two different time slots over all pairs of vehicles in the three data sets.

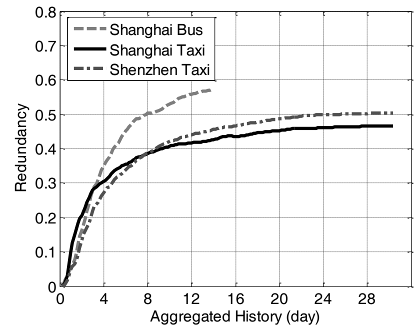

Figure 8. Mean redundancy of the layout of inter-contact times with aggregated history ICTs over all pairs of vehicles in the three data sets.

clear to see that the conditional entropy is much smaller than the marginal entropy for all types of vehicles. In addition, all entropy distributions are very close. Buses have much smaller conditional entropy than taxies in Shanghai. Therefore, although a pair of buses can have as many inter-contact times as a pair of taxies do, the inter-contact times between buses are more correlated than those between taxies. Interestingly, taxies in Shenzhen also have much smaller conditional entropy than taxies in Shanghai. This suggests that taxies in Shanghai operate more randomly with less interference of drivers than taxies in Shenzhen.

B. Evolution of ICT Patterns

In order to establish informed message forwarding strategy utilizing inter-contact time temporal patterns, we divide time into short time slots and examine the correlation between the distribution of inter-contact times between a pair of vehicles in time slot *t* and that in time slot *t* - *n*, increasing *n* from one to a large number. We use *redundancy* to quantify the correlation. Specifically, the inter-contact times between this pair of vehicles in time slot t forms a time series $T_t = (n_0, n_1, \cdots, n_{|t|-1})$, where $|t|$ is the length of a time slot and n_i is the number of inter-contact times occurred at time i $(0 \leq i \leq |t| - 1)$. We also have the time series of inter-contact times in time slot $t - n$, T_{t-n}. We compute the mutual information of T_t and T_{t-n}, $I(T_t, T_{t-n})$ via the joint entropy $H(T_t, T_{t-n})$ and the marginal entropy $H(T_t)$ and $H(T_{t-n})$ as follows:

$$I(T_t, T_{t-n}) = H(T_t) + H(T_{t-n}) - H(T_t, T_{t-n}). \tag{7}$$

We define the redundancy of $X_{r1}$and X_{r2} by

$$R(T_t, T_{t-n}) = \frac{I(T_t,T_{t-n})}{H(T_t)+H(T_{t-n})}. \tag{8}$$

We compute the mean redundancy averaged over all pairs of vehicles in Shanghai bus data set from March 5, Shanghai taxi data set from March 3 and Shenzhen taxi data set from October 31, respectively. Time is divided into time slots of four hours. Fig. 7 shows the result for *n* = 1 to 84 (a period of two weeks). It can be seen that the layout of inter-contact times in a period of time has higher correlation with historical information when the time difference is a multiple of one day for all types of vehicles. Buses have higher redundancy than taxies. Therefore, the inter-contact times between buses are more predictable. Interestingly, the redundancy with Shanghai taxies achieves higher values on even numbers of days than on odd ones whereas the redundancy with Shenzhen taxies is more homogeneous throughout the whole period of time, having larger variances. This should reflect the different shift rules of taxies in these two cities. In Shanghai, taxi drivers usually shift every 24 hours so a taxi behaves very differently on every day but very similarly on every other day. The case in Shenzhen, where drivers shift twice a day (e.g., 7am and 5pm), is that a taxi behaves differently during the daytime but similarly on every day.

To better understand how much history data should be considered in capturing the inter-contact time patterns, we examine the redundancy between the layout of inter-contact time in time slot *t* and the aggregated historical information from $t-1$ to $t-n$, i.e., $\sum_{i=1}^{n} T_{t-i}$. We plot the average redundancy over all pairs of vehicles in the three data sets shown in Fig. 8. It is clear that the redundancy increases until n reaches to about three weeks. This implies that information older than three weeks does not help in capturing inter-contact time temporal patterns.

V. Opportunistic Forwarding Algorithm Design

The analysis based on empirical vehicular trace data in Section III suggests that it is possible to predict when the next connection opportunity between a pair of vehicles will probably occur based on their recent inter-contact times. This enlightens the design of new opportunistic forwarding algorithms in urban vehicular networks. In this section, we first capture the inter-contact time temporal patterns between each pair of vehicles using higher order Markov chain models. Then, we describe our opportunistic forwarding strategy and discuss the algorithm parameter configuration in terms of system performance and memory cost.

A. Markov Chain Model of k-th order

The class of finite-state Markov processes (Markov chain models) is rich enough to capture a large variety of temporal dependencies. In Markov chain models, the current state of the process depends only on a certain number of previous values of the process, which is the order of the process. By (1), continuous values of inter-contact times can be specialized into finite state space, $\mathcal{S} = \{0,1,\cdots,\lfloor \mathbb{T}/\lambda \rfloor\}$. Thus, we can establish a k-th order Markov chain to represent the temporal dependency of inter-contact time between a pair of vehicles. The number of states is $(\lfloor \mathbb{T}/\lambda \rfloor + 1)^k$ and the number of conditional probabilities is $(\lfloor \mathbb{T}/\lambda \rfloor + 2)^k$.

More specifically, let $\{x_i\}_{i=1}^{n}$ be an observed sequence of inter-contact times between this pair of vehicles. The k-order state transition probabilities of the Markov chain can be estimated for all $a \in \mathcal{S}$ and $\underline{b} \in \mathcal{S}^k$, $\underline{b} = (b_1, b_2, \cdots, b_k)$ as follows. Let $n_{\underline{b}a}$ be the number of times that state $\underline{b}$ is followed by value a in the sample sequence. Let $n_{\underline{b}}$ be the number of times that state $\underline{b}$ is seen and let $p_{\underline{b};a}$ denote the estimation of the state transition probability from state $\underline{b}$ to state $(b_2, \cdots, b_k, a)$. The maximum likelihood estimators of the state transition probabilities of the k-th order Markov chain are

$$p_{\underline{b};a} = \begin{cases} n_{\underline{b}a}/n_{\underline{b}}, & if\ n_{\underline{b}} > 0 \\ 0, & otherwise \end{cases}. \tag{9}$$

B. Opportunistic Forwarding Strategy

In order to acquire the knowledge of inter-contact patterns, a vehicle first collects recent inter-contact times between itself and all other vehicles. Meantime, it establishes a k-th order Markov chain for each interested vehicle in the network by determining the state transition probabilities according to (9). As a new inter-contact time comes, the vehicle also updates the corresponding Markov chain. It then uses the established Markov chain model as guidance to conduct future message forwarding. Specifically, when a vehicle v_1 encounters vehicle v_2, v_1 examines all messages stored in the buffer of v_2. Suppose v_d is the destination of such a message. Let $\underline{b}_{v_1,v_d}$ denote the current state in the k-th order Markov chain between v_1 and v_d. The estimated delay of the next contact between v_1 and v_d, $\mathcal{E}_{delay}^{v_1,v_d}$ can be calculated as,

$$\mathcal{E}_{delay}^{v_1,v_d} = \sum_{a=0}^{\lfloor \mathbb{T}/\lambda \rfloor} p_{\underline{b}_{v_1,v_d};a} \cdot a. \tag{10}$$

Vehicle v_1 will act as the next relay for this message if one of the two following cases happens: 1) v_1 is the destination of this message, i.e., $v_1 = v_d$, and 2) v_1 is a better candidate for relaying this message if the estimated delay of the next contact between v_1 and v_d is shorter than that between v_2 and v_d, i.e., $\mathcal{E}_{delay}^{v_1,v_d} < \mathcal{E}_{delay}^{v_2,v_d}$. After transmitting the message to v_1, v_2 simply removes this message from its buffer. Similarly, v_2 will also check messages carried by v_1 and relay messages if needed.

C. Algorithm Parameter Configuration

In our opportunistic forwarding algorithm, there are four key parameters that impact the system performance, namely the maximum inter-contact time in consideration $\mathbb{T}$, the counting measure λ, the order of Markov chain models k and the length for learning stage. In addition, vehicles can have large but limited memory.

Given $\mathbb{T}$, a small counting measure λ will increase the number of states in the Markov chain models, preserving more detailed information at a price of larger memory consumption. On the other hand, if λ equals $\mathbb{T}$, there is only two states in the Markov chain. Thus, a pair of vehicles can only judge the probability that the delay of their next connection is larger than $\mathbb{T}$. This has less sense in helping message forwarding. Intuitively, with more detailed information, vehicles can predict more accurately about next communication opportunities. Therefore, there is a tradeoff between memory cost and system performance. Given the state space of a Markov chain model, simply increasing k will not help increase the number of state transition probabilities. The order of Markov chain models k can be determined by conducting AIC tests [20]. Due to the limitation of space, we omit the details.

Fig. 9 shows an example of the average number of state transition probabilities per pair of vehicles in Shanghai taxi trace data set. It can be seen that the number of state transition probabilities reaches the maximum when λ takes the minimum value (i.e., four hours in this example) and k equals six.

From the analysis in Section IV, it is clear that increasing the length of learning stage will definitely help improving the accuracy of estimation for next connections. It also suggest that history information that is old than about three weeks will not help. Note that all Markov chains are established along with the movement of vehicles in real time. The performance of the proposed opportunistic forwarding algorithm will gradually improve as more learning data becomes available. We will further examine the effect of λ, k and the length of learning stage through trace-driven simulations in the next section.

VI. Performance Evaluation

A. Methodology

In this section, we compare our opportunistic forwarding algorithm with several alternative schemes:

- **Epidemic.** In this scheme [7] [8], vehicles exchange every packet whenever they experience a contact. If vehicles have infinite buffer size, using epidemic routing will find the shortest path between the source and destination vehicles and therefore has the shortest end-to-end delay. On the other hand, since there is no control on data forwarding, it also generates a tremendously large volume of network traffic, overwhelming limited wireless bandwidth.
- **Minimum Expected Delay (MED).** This scheme [1] utilizes the *expected delay* metric to guide data forwarding. Expected delay is used to estimate the expected delay between two vehicles v_1 and v_2 based on contact records. Given the contact record shown in Fig. 2, expected delay can be calculated

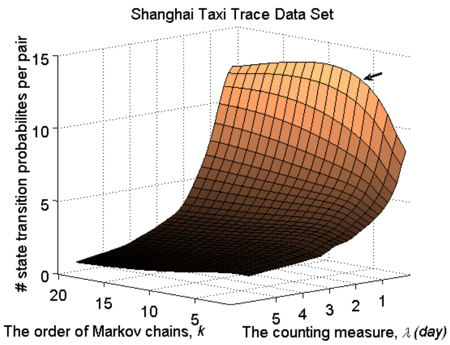

Figure 9. The memory cost vs. counting measure λ and the order of Markov models *k*.

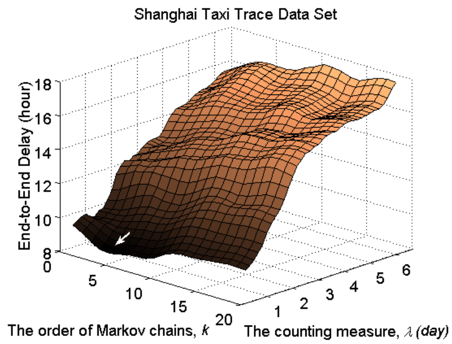

Figure 10. The end-to-end delay vs. counting measure λ and the order of Markov models *k*.

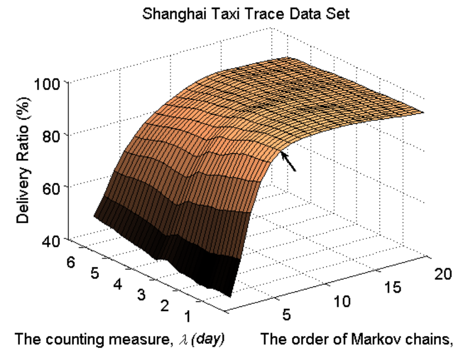

Figure 11. The delivery ratio vs. counting measure λ and the order of Markov models *k*.

as $D(v_1, v_2) = \frac{\sum_{i=1}^{m} d_i^2}{2T}$. When conducting packet forwarding, the vehicle with the minimum expected delay is chosen as the next hop.

- **Maximum Delivery Probability (MDP).** This scheme [2] [21] utilizes the *delivery probability* metric to guide data forwarding. Delivery probability is designed to reflect the contact frequency, i.e., how often two vehicles meet each other. For example, if the contact record between vehicles v_1 and v_2 is shown in Fig. 2, the delivery probability between vehicles v_1 and v_2 can be calculated as $P(v_1, v_2) = 1 - \frac{\sum_{i=1}^{m} d_i}{T}$. Upon selecting a next-hop vehicle to forward a packet, a vehicle prefers the neighbor with the maximum delivery probability.

We consider three important metrics to evaluate the performance of our algorithm and the above schemes:

*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) **End-to-end delay.*** It refers to the 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) **Network traffic per packet.*** It refers to the average network cost per packet, calculated by dividing the total number of data forwarding by the number of packets.

In the following simulations, we evaluate the performance of our opportunistic forwarding algorithm in terms of the above metrics, using real trace data from Shanghai taxies, Shenzhen taxies and Shanghai buses. From each data set, we randomly choose 500 vehicles. We then extract contact records between each pair of vehicles for all selected vehicles, using the algorithm described in Subsection III.B. We use the contact records in the first three weeks (one week for bus due to the limited available data) as the learning stage for all alternative schemes and use the last week for data transmission. At beginning of each experiment, we inject 100 packets using a Poisson packet generator with a mean interval of ten seconds. For each packet, the source and destination are randomly chosen among all vehicles in each data set. Here we assume that two vehicles can always successfully conduct all data transmission when they have a contact.

B. Effect of Algorithm Parameters

We first examine the effects of protocol parameters to the network delivery performance. The maximum inter-contact time $\mathbb{T}$ is set to be six days (90% confidence interval). We vary the counting measure λ from four hours to six days at an interval of four hours and vary the order of Markov chain *k* from one to 20 at an interval of one. For each value of λ and *k*, we run the experiment 50 times and measure the average results.

Fig. 10 shows the end-to-end delay based on Shanghai taxi. The minimum end-to-end delay can be achieved with the smallest λ equal to four hours and *k* equal to six in this case. It is clear that increasing λ will result in larger end-to-end delay. To some extent, increasing *k* will not reduce the end-to-end delay. Fig. 11 shows the delivery ratio as a function of λ and *k*. It can be seen the delivery ratio reaches the maximum with the smallest λ and k equal to six. The delivery ratio increases very fast as *k* increases in the beginning but after that it starts to decrease gradually. When λ varies from four hours to six days, the delivery ratio decreases. These results verify the conclusion described in Subsection V.C. We also check the effect of the configuration of λ and *k* to the delivery performance on Shanghai buses and Shenzhen Taxies. The result is similar, i.e., taking the smallest λ will get the best performance with k equal to five based on Shanghai bus data and six based on Shenzhen Taxi data.

C. Effect of Learning Stage

In this simulation scenario, we examine how much history information is essential for setting up our models. We apply a small λ and the corresponding optimal configuration of *k* and gradually increase the time for learning. For example, in Shanghai taxi trace data, we set λ = 30 minutes and *k* = 6 and use the trace in last week, from Feb. 25 to Mar. 3, for data transmission. We increase the time for learning from one day (i.e., Feb. 24), two days (i.e., Feb. 23 - Feb. 24) till 24 days (i.e., Feb. 1 - Feb. 24). For each training time, we run the experiment 50 times and measure the average results.

Fig. 12 shows the end-to-end delay as the length of learning stage grows. It can be seen that, with more history

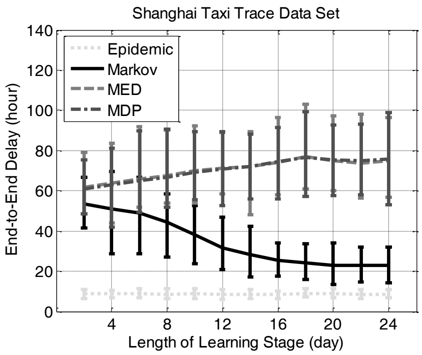

Figure 12. The end-to-end delay vs. the length of learning stage based on Shanghai taxi data.

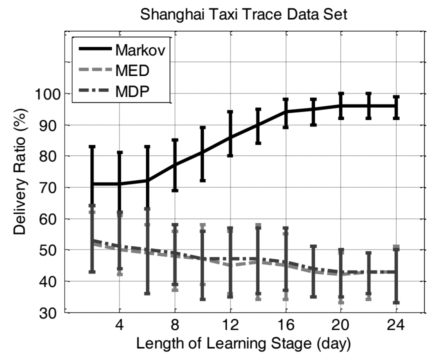

Figure 13. The delivery ratio vs. the length of learning stage based on Shanghai taxi data.

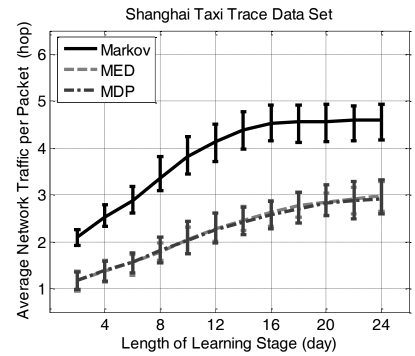

Figure 14. The network traffic per packet vs. the length of learning stage based on Shanghai taxi data.

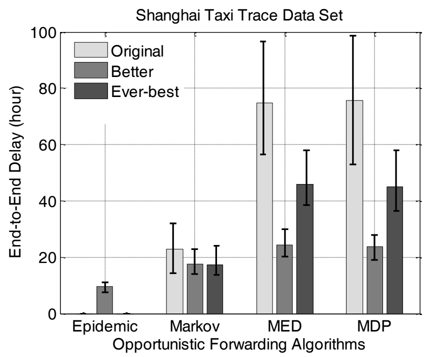

Figure 15. The end-to-end delay vs. opportunistic forwarding algorithms based on Shanghai taxi data.

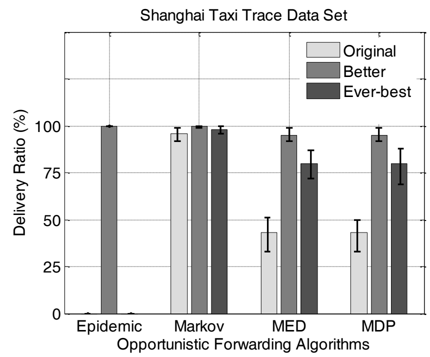

Figure 16. The delivery ratio vs. opportunistic forwarding algorithms based on Shanghai taxi data.

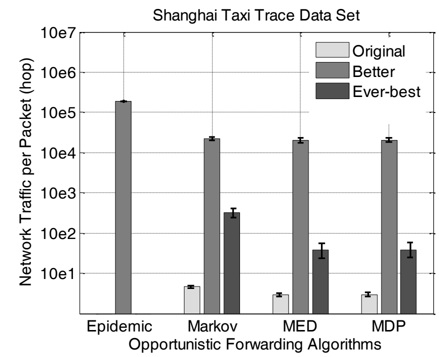

Figure 17. The network traffic per packet vs. opportunistic forwarding algorithms.

information available, our algorithm can dramatically reduce the average end-to-end delay from 53.62 hours to 22.87 hours. When the length of learning stage is larger than 19 days, the end-to-end delay hits a plateau and stabilizes. This is consistent with our observation in Subsection IV. B that history information older than three weeks will not contribute more. Surprisingly, as the learning time grows, both MED and the MDP schemes have larger end-to-end delay. Since these schemes are based on aggregated characteristics of inter-contact times, they cannot fully utilize the temporal dependency of vehicular mobility. The MED and the MDP schemes achieve the minimum end-to-end delay of 61.62 hours and 61.02 hours, respectively, using one day for learning. The epidemic scheme has the minimum end-to-end delay of 8.6 hours.

We plot the delivery ratio as a function of learning time shown in Fig. 13. We omit results from the epidemic scheme since it can always get a 100% delivery ratio in this setting. The Markov scheme can reach to a 96% delivery ratio when the length of learning stage is larger than three weeks. It can also delivery about 84% more packets, compared with the best performance of the MDP and the MED schemes (53% and 52%). Fig. 14 shows the average network traffic per packet generated in the network. It can be seen that it takes three more hops on average to deliver a packet using the Markov scheme than using the MED and the MDP schemes to achieve best performance. The epidemic scheme has the largest network cost of 1.87×10^5 hops. In summary, our scheme can achieve comparable delivery performance as the epidemic scheme with a conservative network cost. We also examine the effect of learning stage to the network performance based on Shenzhen taxi data and Shanghai bus data. Results are presented in Table II.

D. *Effect of Multiple Paths*

In previous simulations, each packet follows only one path, i.e., at any time, at most one copy of a packet exists in the network. In this simulation, we extend our algorithm to allow multiple copies of a packet, thus to improve delivery performance in terms of shorter delay and higher delivery ratio. We consider two multiple path forwarding strategies:

1) ***Better Candidate.*** In this strategy, instead of removing a packet from its buffer after message forwarding, a vehicle keeps a copy of a packet and can always forward the packet to other candidate vehicles in the future;

2) ***Ever-best Candidate.*** In this strategy, a vehicle also keeps a copy of a packet but only transmits the packet to a candidate that has the ever-best delay estimation among all previous candidates it has chosen.

We apply these two strategies to our Markov scheme, the MDP and the MED schemes, and conduct experiments with the same configuration as that in the above simulation except all available data in learning stage are used. The end-to-end delay, delivery ratio and the network traffic per packet based on Shanghai taxi data are shown in Fig. 15, Fig. 16 and Fig. 17,

TABLE II. PERFORMANCE COMPARISON OF ALL SCHEMES

Shenzhen Taxies	*Min. End-to-end Delay (hour)*	*Max. Delivery Ratio*	*Network Traffic (hop)*	*Shanghai Buses*	*Min. End-to-end Delay (hour)*	*Max. Delivery Ratio*	*Network Traffic (hop)*
Markov	23.68	83%	3.34	**Markov**	34.12	95%	2.33
MED	49.70	40%	1.82	**MED**	74.90	53%	1.47
MDP	48.81	41%	2.04	**MDP**	74.29	53%	1.47
Epidemic	3.34	100%	1.25×10^5	**Epidemic**	11.67	100%	2.06×10^5

respectively. It can be seen that the proposed scheme can achieve appealing delivery performance (22.87-hour end-to-end delay and 96% delivery ratio) even with one-path forwarding. By conducting multiple path forwarding, the MED and MDP schemes can achieve smaller end-to-end delay and larger deliver ratio but at a very high network cost.

VII. CONCLUSION AND FUTURE WORK

In this paper, we have demonstrated that urban vehicles show strong temporal dependency in terms of how they meet each other. Although our study based on two specific types of public vehicles, namely taxies and buses, they are representative with respect to mobility characteristics in urban settings. Buses have dedicated routes and fix schedules which make their connection opportunities more predictable. On the other hand, taxies with much random mobility still have strong temporal correlation between every pairwise contact. Therefore, we have developed an appealing opportunistic forwarding algorithm using higher order Markov chains, which can significantly improves the delivery ratio and reduce the end-to-end delay for data delivery.

We will extend our work in two directions. First, it is often assumed in the literature that data transfers can be done instantaneously as soon as two vehicles have a chance to meet. It is definitely not the case in reality since wireless link quality is very dynamic. Thus, we will investigate the end-to-end delay with limited wireless link bandwidth since the delay is influenced not only by ICTs but also by retransmissions if the data transfer fails in a contact. Second, we will validate our algorithm by conducting field tests and collecting trace data of more types of vehicles.

ACKNOWLEDGEMENT

This research was supported in part by NSERC Strategic Grant, Canada, ORF-RE grant, Ontario, Canada, the National Basic Research Program of China (973 Program) under Grant No. 2006CB303000, the National Natural Science Foundation of China under Grant No. 60970106, 60673166 and Key Program of National MIIT under Grant 2009ZX03006-004, China.

REFERENCES

[1] S. Jain, K. Fall, and R. Patra, "Routing in a Delay Tolerant Network," in *Proceedings of ACM SIGCOMM*, pp. 145-158, 2004.

[2] A. Lindgren, A. Doria, and O. Schelen, "Probabilistic Routing in Intermittently Connected Networks," *Mobile Computing and Comm. Rev.*, vol. 7, no. 3, pp. 19-20, 2003.

[3] J. Ghosh, S. Philip, and C. Qiao, "Sociological Orbit Aware Location Approximation and Routing in MANET," *Ad Hoc Networks*, vol. 5, no. 2, pp. 189-209, 2007.

[4] J. Leguay, T. Friedman, and V. Conan, "Evaluating Mobility Pattern Space Routing for DTNs," in *Proceedings of IEEE INFOCOM*, 2006.

[5] Qin Lv, Pei Cao, Edith Cohen, Kai Li, and Scott Shenker, "Search and Replication in Unstructured Peer-to-Peer Networks," in *Proceedings of the 16th international conference on Supercomputing*, 2002.

[6] Christos Gkantsidis, Milena Mihail, and Amin Saberi, "Random Walks in Peer-to-Peer Networks," in *Proceedings of IEEE INFOCOM*, 2004.

[7] T. Spyropoulos, K. Psounis, and C. Raghavendra, "Efficient Routing in Intermittently Connected Networks: The Multi-Copy Case," *ACM/IEEE Transactions on Networking*, vol. 16, no. 1, pp. 77-90, 2008.

[8] A. Balasubramanian, B.N. Levine, and A. Venkataramani, "DTN Routing as a Resource Allocation Problem," in *Proceedings of ACM SIGCOMM*, pp. 372-384, 2007.

[9] M. McNett, and G. M. Voelker, "Access and mobility of wireless pda user", in Technical Report, Computer Science and Engineering, UC San Diego, 2004.

[10] T. Karagiannis, J. Le Boudec, and M. Vojnovi_c, "Power Law and Exponential Decay of Inter Contact Times between Mobile Devices," in *Proceedings of ACM MOBICOM*, pp. 183-194, 2007.

[11] X. Zhang, J. Kurose, B. N. Levine, D. Towsley, and H. Zhang, "Study of a Bus-based Disruption-Tolerant Network: Mobility Modeling and Impact on Routing", in *Proceedings of ACM/IEEE MOBICOM*, 2007.

[12] H. Zhu, M. Li, L. Fu, G. Xue, Y. Zhu, and L. Ni, "Impact of Traffic Influxes: Revealing Exponential Inter-Contact Time in urban VANETs", to appear in *IEEE Transactions on Distributed and Parallel Systems*, 2010.

[13] V. Conan, J. Leguay, and T. Friedman, "Fixed Point Opportunistic Routing in Delay Tolerant Networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 5, pp. 773-782, 2008.

[14] V. Erramilli, A. Chaintreau, M. Crovella, and C. Diot, "Delegation Forwarding," *in Proceedings of ACM MobiHoc*, 2008.

[15] M. Shin, S. Hong, and I. Rhee, "DTN Routing Strategies Using Optimal Search Patterns," in *Proceedings of ACM SIGCOMM Workshop Challenged Networks* (*CHANTS '08*), 2008.

[16] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of Human Mobility on the Design of Opportunistic Forwarding Algorithms", *in proceedings of IEEE INFOCOM, 2006*.

[17] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network", in *Proceedings of ACM MOBICOM*, 2004.

[18] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, "Pocket switched networks and the consequences of human mobility in conference environments," in *Proc. of ACM SIGCOMM first workshop on delay tolerant networking and related topics* (*WDTN-05*), 2005.

[19] R. Lu, X. Lin, and X. Shen, "SPRING: A Social-based Privacy Preserving Packet Forwarding Protocol for Vehicular Delay Tolerant Networks", in *Proceedings of IEEE INFOCOM*, 2010.

[20] H. Tong, "Determination of the Order of a Markov Chain by Akaike's Information Criterion", Journal of Applied Probability, Vol. 12, No. 3, pp. 488-497, 1975.

[21] J. Burgess, B. Gallager, D. Jensen, and B. Levine, "MaxProp: Routing for Vehicle-Based Disruption-Tolerant Networks," in Proceedings of IEEE INFOCOM, 2006.