Impact of Traffic Influxes: Revealing Exponential Intercontact Time in Urban VANETs

Hongzi Zhu, *Member*, *IEEE*, Minglu Li, *Senior Member*, *IEEE*, Luoyi Fu, Guangtao Xue, *Member*, *IEEE*, Yanmin Zhu, *Member*, *IEEE*, and Lionel M. Ni, *Fellow*, *IEEE*

Abstract—Intercontact time between moving vehicles is one of the key metrics in vehicular ad hoc networks (VANETs) and central to forwarding algorithms and the end-to-end delay. Due to prohibitive costs, little work has conducted experimental study on intercontact time in urban vehicular environments. In this paper, we carry out an extensive experiment involving thousands of operational taxies in Shanghai city. Studying the taxi trace data on the frequency and duration of transfer opportunities between taxies, we observe that the tail distribution of the intercontact time, that is, the time gap separating two contacts of the same pair of taxies, exhibits an exponential decay, over a large range of timescale. This observation is in sharp contrast to recent empirical data studies based on human mobility, in which the distribution of the intercontact time obeys a power law. By analyzing a simplified mobility model that captures the effect of hot areas in the city, we rigorously prove that common traffic influxes, where large volume of traffic converges, play a major role in generating the exponential tail of the intercontact time. Our results thus provide fundamental guidelines on design of new vehicular mobility models in urban scenarios, new data forwarding protocols and their performance analysis.

Index Terms—Vehicular ad hoc networks, intercontact time, exponential tail, mobility model, traffic influx, empirical data analysis.

1 Introduction

Vehicular ad hoc networks (VANETs) are recognized as an important component in the next generation of intelligent transportation systems, to improve safety, security, and efficiency of transportation systems and enable new mobile services to the public. In VANETs, vehicles equipped with wireless communication devices can transfer data with each other (vehicle-to-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 (within the communication range of each other) for data relay. Applications based on this type of data transfer will strongly depend on vehicular mobility characteristics, especially on how often such communication opportunities take place and on how long they last. In this paper, we focus on studying the metric called *intercontact time* [1], [12], [13], [14], [16], which denotes the time elapsed between two successive contacts of the same two vehicles. Since data transfer arises in a *store-carry-forward* fashion, the intercontact time of the two vehicles is a major component of the end-to-end delay, as it presents how long it takes to encounter the other mobile vehicle to have any chance to forward/relay the data for communication. Larger intercontact time results in larger end-to-end delay.

In the literature, there have been many studies on the characteristics of the intercontact time in delay tolerant networks (DTNs) and mobile ad hoc networks (MANETs). Most of these results focused on theoretical models, such as random walk mobility models (RWMs) [2], [3], [4], random waypoint mobility models (RWPs) [6], [7], [8], and random direction mobility models (RDMs) [9]. While theoretical mobility models facilitate problem analysis, they are far beyond reality and not practical in designing networking protocols for real systems and their performance analysis. Recently, some empirical results [1], [13], [14] based on human mobility showed that the tail distribution of the intercontact time is far from being exponential, but can be approximated or lower bounded by a power law (heavy tail). This implies that it is quite likely that two have to wait a long period of time before they can meet and communicate with each other. In order to have a better understanding of practical constraints in opportunistic data transfer between vehicles, experiments involving thousands of vehicles over a long time span of months are in pressing demand. However, due to vast deployment costs, there is no existing work, to the best of our knowledge, studying vehicular intercontact time distribution in urban settings based on real experiments.

In this paper, we collect real motion traces from about 2,109 operational taxies for over one month in Shanghai

• H. Zhu, M. Li, G. Xue, and Y. Zhu are with the Department of Computer Science and Engineering, Shanghai Jiao Tong University and Shanghai Key Lab of Scalable Computing and Systems, Room A-303, SEIEE No. 5 Building, 800 Dong Chuan Road, Min Hang, Shanghai, P.R. China. E-mail: {hongzi, mlli, gt_xue, yzhu}@sjtu.edu.cn.
• L. Fu is with the Department of Electronic Engineering, Shanghai Jiao Tong University, 800 Dong Chuan Road, Min Hang, Shanghai, P.R. China. E-mail: yiluofu@sjtu.edu.cn.
• L.M. Ni is with the Department of Computer Science and Engineering, The Hong Kong University of Science and Technology and Shanghai Key Lab of Scalable Computing and Systems, Clear Water Bay, Kowloon, Hong Kong, P.R. China. E-mail: ni@cse.ust.hk.

Manuscript received 4 Jan. 2010; revised 19 Apr. 2010; accepted 5 June 2010; published online 6 Oct. 2010.
Recommended for acceptance by Y.-C. Tseng.
For information on obtaining reprints of this article, please send e-mail to: tpds@computer.org, and reference IEEECS Log Number TPDS-2010-01-0017.
Digital Object Identifier no. 10.1109/TPDS.2010.176.

city, the biggest metropolis in China. By analyzing the large volume of trace data, we surprisingly find that the tail distribution of the intercontact time between taxies follows an exponential distribution on a large range of timescale, other than a power law exhibited in human mobility [1], [13], [14]. This implies that, to some extent, taxies can frequently meet with each other. Thus, data delivery would experience smaller end-to-end delay. Beyond this, we want to know the answers to two fundamental questions: *what is the essential factor, if any, that generates the exponential tail behavior of the intercontact time of taxies? Can this factor affect the intercontact times of other vehicles?*

We further delve into the geographical distribution of all experimental taxies and find that a taxi spontaneously presents itself in popular areas (e.g., at large commercial centers or traffic conjunctions) most of the time, which largely increases the potential communication opportunities between two taxies. Inspired by this intuition, we establish a simplified mobility model that characterizes the effect of traffic gathering at certain locations. More specifically, we refer to such an area where traffic tends to converge as a *traffic influx*. In our model, a vehicle randomly walks in an infinitely large region. Besides, it also revisits a traffic influx at least once within a given time period. Therefore, the length of the revisiting time period indicates how popular this traffic influx is with respect to this vehicle. We rigorously prove that, for any two vehicles of whatever kind, as long as they visit a common traffic influx within finite-length time periods, the tail distribution of the intercontact time follows an exponential decay. Otherwise, the distribution of the intercontact time can be lower bounded by a power law. Simulations based on ideal infinite surface as well as on real road network topology verify the soundness of our conclusion. Our findings thus provide fundamental guidelines on design of new vehicular mobility models in urban scenarios, new data forwarding protocols and their performance analysis.

The original contributions that we have made in the paper are highlighted as follows:

- We take an experimental study on the intercontact time characteristics of taxies by collecting and analyzing real traces from more than two thousands of operational taxies. We discover that the tail behavior of the intercontact time of taxies follows an exponential distribution as opposite to a power law reported by previous empirical work.
- We design a simplified mobility model to mimic the traffic-gathering behavior in urban settings. By rigorously proving that the tail distribution of the intercontact time is exponential as long as two vehicles constantly revisit a common traffic influx, we generalize our conclusion to all urban vehicles in addition to taxies.
- We conduct extensive simulations based on both theoretical infinite region and real road networks of Shanghai. All results strongly support our empirical and theoretical findings.

The rest of this paper is structured as follows: Section 2 is dedicated to related work. Our experiment results are presented in Section 3. In Section 4, we introduce the simplified mobility model we use in our analysis and present our theoretical results. We have some discussion in Section 5. Simulation results are then presented in Section 6 to support our empirical and theoretical findings. Finally, we present concluding remarks and outline the directions for future work in Section 7.

2 Related Work

In the literature, a majority of research results have uncovered a common property of many theoretical mobility models that the tail of the intercontact distribution decays exponentially. For example, authors in [4], [10], [11] draw this conclusion through numerical simulations based on RWP mobility models. Furthermore, some theoretical results showed that the first and second moments of the intercontact time are bounded above under Brownian motion model on a sphere. In particular, authors in [12] proved that a finite boundary is a major factor that causes the exponential tail behavior under any RWP mobility model and any RWM mobility model. While using theoretical mobility models simplify problem analysis, they are inconsistent with the reality, and thus, impractical in designing networking protocols for real systems.

In recent years, there has emerged more research work taking experimental study on the characteristics of the intercontact time. For example, authors in [1], [13], [14] found that the tail behavior of the intercontact time based on human mobility is far from being exponential but is close to a power law instead. These results were based on real traces, such as human contacts, while at conferences [14], campus WiFi login records [13], [15] and a Bluetooth network containing hundreds of people in an office [14]. It is apparent that the mobility of vehicles is significantly different from that of human beings in terms of speed, constraints of road transportation systems, and travel distances. Although these empirical results based on human mobility depict different intercontact time distribution, the situation in vehicular environments is still left unknown.

DieselNet [16] at UMass consisting of 40 buses studies the aggregated intercontact time distribution at a granularity of bus route and finds a clear periodic structure in the intercontact times between two bus routes. For the lack of enough contact samples between two individual buses, the bus trace data are not sufficient for studying the distribution of the intercontact time between two individual buses. In the RAPID routing protocol [17], it is assumed that the distribution of bus intercontact times is exponential to make their problem tractable.

To the best of our knowledge, there is no existing work studying the empirical characteristics of the intercontact time of urban vehicles.

3 Empirical Data Analysis

In this section, we first give a brief description of the trace data collected in the SG project [18]. We then present the intercontact time characteristics embedded in the trace data of taxies. Finally, we discuss the possible reasons behind our key observation on the intercontact time distribution.

Fig. 1. A taxi with a commercial GPS device installed, the highlight area in the inset shows such a device.

3.1 Collecting Vehicular Trace Data

In the SG project, about 6,850 taxies and 3,620 buses in Shanghai city are equipped with a commercial GPS receiver and a GPRS wireless communication module (an example taxi is shown in Fig. 1). As a bus or a taxi commutes in the city, it periodically sends a report back to a data center via a GPRS channel. In this paper, we intend to survey the impact of urban environments on the statistical features of intercontact time. We choose to study trace data of taxies because taxies are more sensitive to urban environments (e.g., the underlying road network topology, traffic control, and urban planning) and have broader coverage in terms of space and operation time than buses.

It would be ideal if the collected trace data have recorded the complete motion of all taxies for a sufficiently long period of time so that all possible contacts between two taxies are preserved. Unfortunately, due to network partitions, we may loss some reports for a short period of time from a taxi from time to time. We then select 2,109 taxies out of the whole data set, whose trace data continuously cover the whole month of February in 2007 without any interruptions (available at http://www.cse.ust.hk/scrg). Due to the GPRS communication cost for data transmission, a taxi sends reports back on every 1 minute when it has passengers onboard but on every 15 seconds when it is vacant. Vacant taxies are set to send reports at a higher frequency is for the reason of real-time scheduling. The specific information contained in such a report includes: the taxi's ID, the longitude and latitude coordinates of the taxi's current location, timestamp, its instant speed, heading, and its status (i.e., whether the taxi has passengers onboard).

3.2 Computing Intercontact Times

3.2.1 *Contact Extraction from Trace Data*

Ideally, all connection opportunities, encountering 24 hours a day with a granularity measured in seconds, should be recorded in the data for study. Since we collect GPS reports in discrete time, we make the assumption that two taxies would be able to communicate (called a *contact*) if their locations reported are within a specific time window and within a given communication range at the same time. Note that due to the interference and signal loss of wireless links, two taxies within the communication range may still not be able to perform data transfer. Since we focus on the characteristics of the tail behavior of the intercontact time, we study the potential communication opportunities between two taxies and leave the successful data transfer rate with no further discussion.

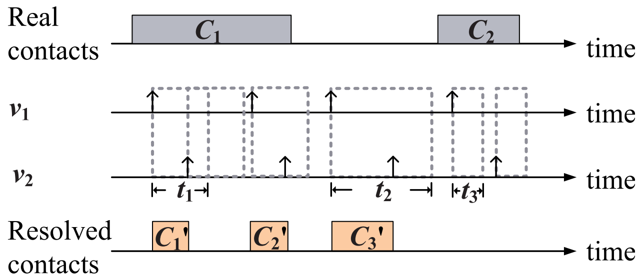

Fig. 2. Extract contacts from GPS reports of vehicle v1 and v2. Boxes in dotted line denote sliding time windows of different granularities used to check contacts. Individual GPS reports are presented by short arrow line segments.

The above assumption, though, can introduce inaccuracies in the following three cases.

First, if a relatively large time window is used, we may introduce false contacts into consideration. This is because two taxies may have already run far away from their reported locations. Therefore, the retrieved contact may never really exist. For example, in Fig. 2, suppose there are two real contacts C_1 and C_2 happening between vehicle v_1 and v_2. We may get a false contact C_3' if a large time window t_2 is used even though there is no real contact at all. The consequence of introducing false contacts is that it increases the weight of small values of intercontact times in the distribution since these false contacts cut large intercontact times into small pieces.

Second, if a small time window is used, we may omit real connection opportunities. This is because two taxies might indeed have a contact but did not send out reports simultaneously. In this case, we may not capture this contact due to the small size of the time window. For example, in Fig. 2, if a small time window t_3 is used to check contacts, it is unlikely to find the real contact C_2. The consequence of omitting real contacts is that it causes large values of intercontact times since two real small intercontact times are now considered as a single huge one.

Last, using a sliding time window to check contacts cannot resolve the duration of a contact. A real contact with a relatively long duration might be recognized as multiple short ones. For example, in Fig. 2, we will get C_1' and C_2' rather than the real one, C_1. To eliminate this effect, we examine the correlation between two consecutive contacts. Specifically, given the reported locations and speeds of each taxi, we calculate the remaining contact time of the first contact as the time these two taxies move along their trajectories at the same speeds before they are out of the communication range. If the second contact is contained within the remaining contact time of the first contact, we make a decision that these two contacts should be merged into one. Vice versa, we can also infer when the second contact started and further check whether the first contact can be merged.

Despite these inaccuracies, the taxi GPS trace data are very valuable to study vehicular mobility models since they cover thousands of vehicles and last for one month.

3.2.2 *Intercontact Time Computation*

We refer to intercontact time as the time elapsed between two successive contacts of the same vehicles as defined in [1], [13],

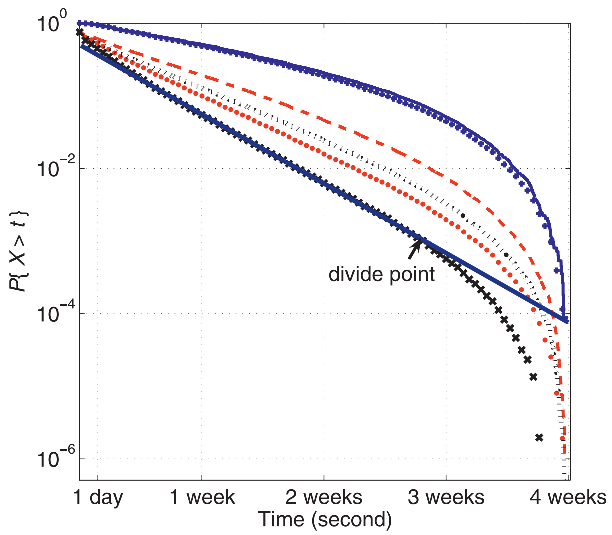

Fig. 3. Tail distribution of the intercontact time: data collected from 2,109 taxies in Shanghai city during the whole February of 2007. Contacts are collected under three different time windows of 1 second, 30 seconds, and 1 minute and two communication ranges of 50 and 100 meters.

[14]. Specifically, the intercontact time is computed at the end of each contact as the time period between the end of this contact and the start of the next contact between the same two vehicles. It should be noted that we do not take into consideration the intercontact starting after the last contacts.

3.3 Identifying Exponential Intercontact Time Tail

We plot the intercontact time distribution for the selected trace data in Fig. 3. The distribution of intercontact times is computed among all pairs of 2,109 taxies during the whole February in 2007. We get six different sets of intercontact times by combining different communication ranges and time window sizes used in the contact extraction. The time windows are set to 1 second, 30 seconds, and 1 minute, respectively, accompanied with two communication ranges of 50 and 100 meters [19]. All plots describe the tail distribution function, i.e., $P\{X > t\}$, in linear-log scale.

The most interesting part in Fig. 3 is that all plots exhibit a very clear exponential tail, i.e., $P\{X > t\} \sim e^{-\beta t}$. This can be indicated by the fact that all plots are almost straight lines with different negative slopes on linear-log scale from the very beginning of time and over a large range of timescale. An exponential decay means the tail distribution function decreases rapidly over this range. For example, in the lowest plot in Fig. 3, about 45 percent of intercontact times are greater than one day, and only 5 percent are greater than one week between taxies. Besides the exponential parts, we also notice that, gradually, all six distributions start to deviate from the exponential decay and drop faster till the end. This rapid cutoff is caused by the limited duration of the trace data, i.e., one month in our experiment. The reason is that intercontact times that last longer than the duration of the trace data cannot be observed and those ones with very large values close to the duration are less likely to be found. The effect of observation duration has also been noted in Chaintreau et al.'s study based on human mobility [1].

3.4 Establishing Exponent Constant

To identify the exponent constant β of the tail distribution in Shanghai city, we perform the least-square regression analysis to the resolved intercontact time. More specifically, we first need to identify the divide point from which the tail distribution function stops exponential decay. This can be achieved by seeking for the point from which the second derivatives (decay acceleration) of the log-scaled $P\{X > t\}$ are larger than a small positive value ε. The physical meaning of using the second derivatives is clear since a nonzero second derivatives of a plot indicates the plot is nonlinear. We then apply polynomial regression to the log-scaled $P\{X > t\}$ over the range from the first point to the divide point. The significance of the regression is measured by the coefficient of determination

$$r^2 = 1 - \frac{\sum_i (y_i - \bar{y})^2}{\sum_i (y_i - m_i)^2},$$

where y_i denotes the sample value with mean $\bar{y}$ whereas m_i is the modeled value. For example, we apply this exercise to the lowest plot in Fig. 3, where the tail distribution of intercontact time is very well approximated ($r^2 > 0.98$) by an exponential distribution $P\{X > t\} = e^{-3.71\times 10^{-6} t}$ when time is counted in seconds.

3.5 Recognizing Traffic Influxes

The finding of the exponential decay on the intercontact time tail distribution of taxies is in a sharp contrast to several recent empirical results on the intercontact time based on extensive human mobility traces [1], [13], [14]. These results indicate that the tail behavior of the intercontact time can be approximated or lower bounded by a power law, i.e., $P\{X > t\} \sim t^{-\alpha}$, for some constant $\alpha > 0$. The discrepancy between human mobility and mobility of taxies described above calls for answers to the following questions: *What is the key factor that makes taxies meet with each other frequently? Can this factor affect the intercontact times of other common vehicles?* Answers to these two questions are of great importance to many related studies in VANETs, such as capacity-delay trade-offs and design and performance analysis of data forwarding algorithms based on exponential intercontact time.

To gain an insight into these two questions, we further examine the geographical distribution of experimental taxies. We observe two facts as follows: First, a taxi spontaneously presents itself in popular areas (e.g., at large commercial centers, arterial roads and intersections) most of the time. Second, most of the taxies moves only within the city (as shown in Fig. 4). It is apparent to see that taxies do have certain mobility patterns as opposite to randomly walking in a finite field. Thus, it is not the boundary effect that generate the exponential tail distribution of intercontact time as claimed in [12].

We call such an area a *traffic influx* if traffic tends to converge around this area. Intuitively, a traffic influx has the effect of gathering vehicles from time to time, and hence, enormously increases contact opportunities of vehicles. Beyond this intuition, we take a penetrating study on the impact of traffic influxes from an analysis perspective seeking for answers to the above questions in the next section.

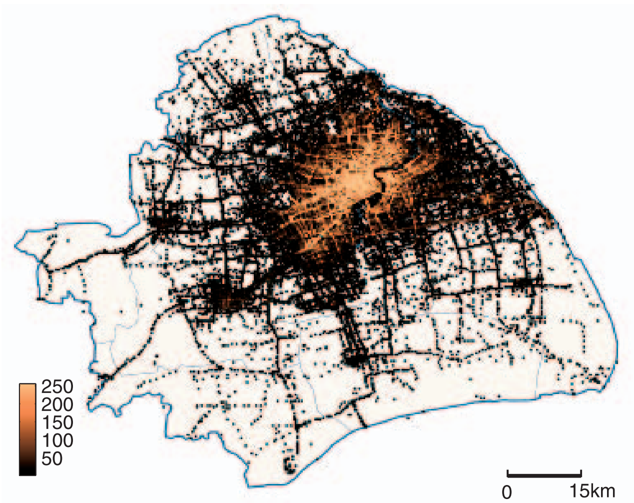

Fig. 4. The geographical distribution of all taxi GPS reports on Shanghai map. Every colored dot presents the average number of appearance per taxi per day in the corresponding $300\ \mathrm{m} \times 300\ \mathrm{m}$ square area on the map.

4 Model Analysis

In this section, we first give some basic definitions and preliminaries related to our analysis. Then we present our mobility model and major theoretical results.

4.1 Definitions and Preliminaries

We consider two arbitrary vehicles of any kind v_1 and v_2, each of which moves according to some mobility model in an infinitely large region Ω (i.e., $\Omega = \mathbb{R}^2$). Let $V_{v_1}(t), V_{v_2}(t) \in \Omega$ be the position of the vehicles v_1 and v_2 at time t, respectively. We assume that $V_{v_1}(t)$ and $V_{v_2}(t)$ are independent and two vehicles can communicate with each other whenever they are within the communication range R_c.

Definition 1. *The intercontact time T_I of vehicles v_1 and v_2 is defined as*

$$T_I \triangleq \inf\{t : \| V_{v_1}(t) - V_{v_2}(t) \| \leq R_c\}, \tag{1}$$

given that $\| V_{v_1}(0) - V_{v_2}(0) \| \leq R_c$ *and* $\| V_{v_1}(0^+) - V_{v_2}(0^+) \| > R_c$. *Here,* $\| \cdot \|$ *is the 2D Euclidian distance.*

Let X be a random variable that has Gaussian distribution with mean μ and variance σ^2, i.e., $X \sim N(\mu, \sigma^2)$.

If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ are two independent random Gaussian distributed variables, we have

Property 1. *The difference of X and Y is also Gaussian distributed*

$$U = X - Y \sim N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2). \tag{2}$$

4.2 Mobility Model with Traffic Influx

4.2.1 Model Description

With the constant observation of vehicle gathering in urban environments, we introduce a mobility model to characterize the gathering effect at a traffic influx.

In our model, a vehicle randomly walks in an infinitely large region to remove the boundary effects as examined by authors in [12]. During the process, it revisits a fixed location, called the *traffic influx*, at least once within a given time period T. More specifically, at the beginning of time, a vehicle randomly selects a direction uniformly from $[0, 2\pi]$, a random distance D chosen from $(0, \infty)$, and a speed randomly chosen from pre-defined $[\vartheta_{min}, \vartheta_{max}]$. Besides, it also sets up a timer $\mathcal{T}$ with the value of T. Then the vehicle starts to conduct a random walk until the timer $\mathcal{T}$ has expired. In that case, the vehicle stops the random walk and heads for the traffic influx. Once it reaches the traffic influx, the vehicle resets its timer and repeats the whole process. We call T the vehicle's maximum periodicity with respect to the traffic influx. The magnitude of T thus indicates how attractive this traffic influx is to this vehicle. A small T means the vehicle very often appears at the traffic influx. As T increases to infinity, chances are that the appearance of the vehicle at the traffic influx is less frequent. In this model, each vehicle can take its own maximum periodicity.

4.2.2 Analysis on Intercontact Time

Here, we present the following assumption and lemmas. Due to space limitation, we omit most proofs from this paper which can be accessed in the supplementary file.

Assumption 1. *At any time t, the mobility of each vehicle is independent, i.e., a vehicle chooses its destinations and routes according to the mobility model and its own preferences to the traffic influx.*

Lemma 1. *Given a specific maximum periodicity T of a vehicle, the region where the vehicle moves can be upper bounded by a disk with radius $r = \vartheta_{max}T$, where ϑ_{max} is the maximum velocity of the vehicle.*

Lemma 2. *Under the mobility model described above, the contact opportunities of a vehicle can be lower bounded by a mobility model that can yield independent symmetric Gaussian location distribution of vehicles.*

We now present our major result as follows:

Theorem 1. *For any two arbitrary vehicles moving according to the mobility model described above, there exists a constant c such that $P\{T_I > t\} \leq e^{-ct}$ for all sufficiently large t as long as they have finite maximum periodicities.*

Proof. Suppose the maximum periodicity of vehicles v_1 and v_2 are T_1 and T_2, respectively. According to Lemma 2, we choose a mobility model for v_1 that satisfies the condition of Lemma 2. The variances of the Gaussian location can be expressed by $\sigma_{X1}^2 = \sigma_{Y1}^2 = (n_1\vartheta_{\max}T_1)^2$, where n_1 is a positive number. Similarly, the variances of the Gaussian location for v_2 can be expressed by $\sigma_{X2}^2 = \sigma_{Y2}^2 = (n_2\vartheta_{\max}T_2)^2$.

Let $P_\Lambda\{T_I > t\}$ be the probability that v_1 never meets v_2 until time t under corresponding Gaussian distribution and $P_\Theta\{T_I > t\}$ be the probability that vehicles v_1 never meets v_2 until time t under our mobility model. The condition described above will guarantee $P_\Theta\{T_I > t\} \leq P_\Lambda\{T_I > t\}$. Next we prove that there exists a constant c such that $P_\Lambda\{T_I > t\} = e^{-ct}$.

Due to the independency and the same variances in each dimension, we can express the joint PDF $f(x, y)$ of two Gaussian distributions in each dimension as follows:

$$f(x,y) = f(x) \cdot f(y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right). \tag{3}$$

Note that the relative position of v_2 with reference to v_1 in each dimension is also independent. With Property 1, we have

$$f_{X_1-X_2}(x) = \frac{1}{\sqrt{2\pi} \cdot \sqrt{\sigma_1^2+\sigma_2^2}} \exp\left(-\frac{x^2}{2\cdot(\sigma_1^2+\sigma_2^2)}\right), \tag{4}$$

and

$$f_{Y_1-Y_2}(y) = \frac{1}{\sqrt{2\pi} \cdot \sqrt{\sigma_1^2+\sigma_2^2}} \exp\left(-\frac{y^2}{2\cdot(\sigma_1^2+\sigma_2^2)}\right). \tag{5}$$

For clarity of writing, we drop the index symbol $X_1 - X_2$ and $Y_1 - Y_2$ and simply write the joint PDF as

$$\begin{aligned} f(x,y) &= f(x) \cdot f(y) \\ &= \frac{1}{2\pi \cdot (\sigma_1^2+\sigma_2^2)} \exp\left(-\frac{x^2+y^2}{2\cdot(\sigma_1^2+\sigma_2^2)}\right). \end{aligned} \tag{6}$$

At any time t, the probability that v_1 and v_2 do not meet is the probability that they are out of the communication range R_c. Let E denote the event that two vehicles do not meet at any time t. Denote by the distance between the two vehicles at time t. We have

$$\begin{aligned} P\{E\} &= P\{\| V_{v_1}(t) - V_{v_2}(t) \| > R_c\} \\ &= \frac{1}{2\pi(\sigma_1^2+\sigma_2^2)} \int\int_{x^2+y^2>R_c^2} \exp\left(-\frac{x^2+y^2}{2\cdot(\sigma_1^2+\sigma_2^2)}\right) dxdy \\ &= \exp\left(-\frac{R_c^2}{2\cdot(\sigma_1^2+\sigma_2^2)}\right). \end{aligned} \tag{7}$$

Therefore, the intercontact time T_I can be expressed as

$$\{T_I > t\} \equiv \bigcap_{0^+}^{t} E. \tag{8}$$

The probability that the two vehicles meet at time t is equivalent to the probability that two vehicles never meet up to time t. Since in our model, the choice of a vehicle at any time t is independent of its previous behavior, E does not depend on t.

Let $\eta = \exp(-\frac{R_c^2}{2\cdot(\sigma_1^2+\sigma_2^2)})$. Thus, we have

$$P_\Lambda\{T_I > t\} = P\left\{\bigcap_{0^+}^{t} E\right\} = \eta^t = e^{-c\cdot t}, \tag{9}$$

where $c = \frac{R_c^2}{2\cdot(\sigma_1^2+\sigma_2^2)} > 0$. Since $P_\Theta\{T_I > t\} \leq P_\Lambda\{T_I > t\}$, we have $P_\Theta\{T_I > t\} \leq e^{-ct}$. We drop the index Θ and get $P\{T_I > t\} \leq e^{-ct}$, when t is sufficiently large. This completes the proof. □

Note that the exponent coefficient c of the tail distribution of the intercontact time depends on both the communication range R_c and variance σ_1^2 and σ_2^2. Given a communication range, larger variances indicate smaller c. However, since the maximum periodicity is finite, the corresponding σ_1^2 and σ_2^2 are constant, which will not change the nature of the tail distribution function of the intercontact time from being exponential.

4.3 Cross over from Exponential to Power Law

From the results in the above section, we can see that the maximum periodicity of different vehicles with regard to a traffic influx is crucial to the tail distribution of intercontact time. Considering the interaction among the maximum periodicity T_1, T_2 of two vehicles and the observation time t of an experiment, we have following conclusions:

1. $\max\{T_1, T_2\}$ *grows much faster than* t: Without loss of generality, we suppose $T_1 = \max\{T_1, T_2\}$. If T_1 grows much faster than the order of $O((\log(\frac{1}{1-\alpha \log t/t}))^{-1/2})$, where α is a constant, chances to observe v_1 revisiting the traffic influx are slim. This definitely prolongs the intercontact time and the power-law distribution arises.
2. *Both* T_1 *and* T_2 *grow slower than* t: Both T_1 and T_2 grow much slower than the order of $O((\log(\frac{1}{1-\alpha \log t/t}))^{-1/2})$. The vehicles can be treated as moving in "bounded" regions under the limited observation time, which implies that they will frequently revisit the traffic influx. This truly introduces an exponential intercontact time tail distribution.

5 Discussion

In urban environments, traffic influxes are mainly caused by the underlying infrastructure layout, such as the topology of road networks and distribution of residential areas, shopping centers, and commercial zones. Besides taxies, other vehicles of different kinds may have their own preference on choosing traffic influxes based on their own interests [20]. With Theorem 1 presented in Section 4, we generalize our conclusion to urban vehicles of any kinds that any two vehicles can have the tail distribution of the intercontact time that decays at least exponentially fast as long as they have at least one common traffic influx during their normal activities. For example, two vehicles that often traverse the same intersection or road segment can be regarded as sharing a common traffic influx. We further validate our results by simulations presented in Section 6.

In reality, a vehicle can have multiple traffic influxes included in its motion. Note that we are not trying to design a mobility model that can exactly depict the accurate motion of vehicles in urban environments. We claim that the tail distribution of intercontact time between two urban vehicles drops at least exponentially fast when they both share at least one common traffic influx.

In addition, there can be more complicated situations in urban environments that may increase the communication opportunities between vehicles. For example, the scheduling scheme of traffic lights can have a great impact on vehicular mobility. Two passing vehicles may gain a contact opportunity while waiting for a green light. Note that we do not claim that the traffic influxes in the urban environments are the only factor that can generate the observed exponential tail distribution of the intercontact time.

6 Simulations

In this section, we conduct simulations and present the results to support our theoretical conclusions in Section 4. In all simulations, the transmission range is set to 100 meters.

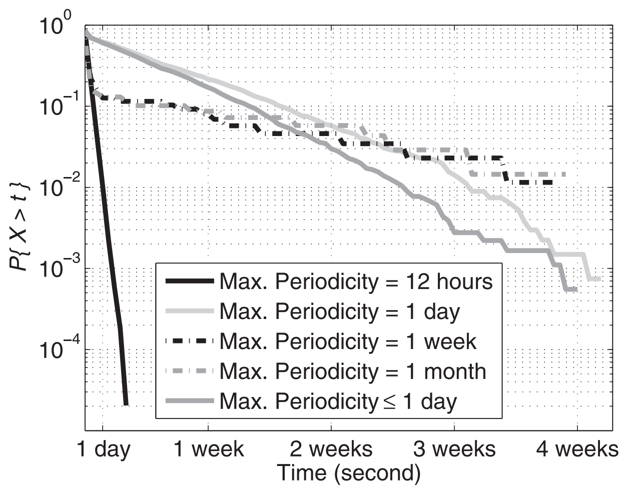

Fig. 5. Tail distribution of the intercontact time under different maximum periodicity on a linear-log scale.

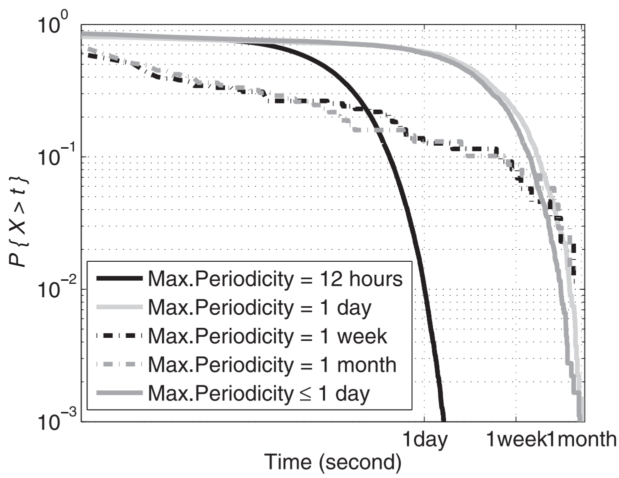

Fig. 6. Tail distribution of the intercontact time under different maximum periodicity on a log-log scale.

6.1 Impact of Maximum Periodicity

In this experiment, we generate 100 vehicles that run according to our mobility model on a sufficiently large region with one traffic influx. Each vehicle moves from its current location to a new location by randomly choosing a direction and speed from 60 to 80 km/h in which to travel. As mentioned before, the interaction between the timescale of the experiment and the timescale of the maximum periodicity of vehicles with respect to the traffic influx is essential in determining the tail distribution of the intercontact time. To clearly see this interaction, we first fix the experiment duration to one month and then increase the maximum periodicity from half of a day to one month to see the possible different types of tail distributions.

Fig. 5 shows the tail distribution of the intercontact time between all pairs of vehicles on a linear-log scale for our mobility model. It can be seen that the tail distribution can be approximated by a line in the front part of the curve when the maximum periodicity is set to 12 hours and one day, respectively. This indicates an exponential decay on the linear-log scale. As we take larger maximum periodicity, the tail behavior of the intercontact time starts to evolve from exponential decay to power law decay as these plots are not linear on the linear-log scale. This can be better observed from Fig. 6 on the log-log scale, where the tail distribution can be approximated by a line when the maximum periodicity equals to one week and one month.

We then conduct another set of simulations where each vehicle randomly chooses its maximum periodicity no longer than one day to revisit the traffic influx. The results are demonstrated in Figs. 5 and 6. As expected, using variable maximum periodicity within one day has generated faster CCDF drop of the intercontact time than using the fixed maximum periodicity of one day while the tail distribution still stays exponential.

6.2 Impact of Real Road Networks

Since urban vehicles always move on roads in the city, the road transportation system can put various constraints to the mobility of these vehicles. To study the impact of the real road networks to the tail distribution of the intercontact time, we conduct a simulation that involves 300 generated vehicles. A vehicle starts to perform random walks on the road networks from the central area. It randomly chooses a distance D and direction each time and changes its speed to the speed limitation marked on the current road segment. The period of the experiment is set for three days.

Since we take a relatively short experiment time and set vehicles to start from the central area. The majority of the vehicles have not reached the boundary of the city. Fig. 7 shows the tail distribution of the intercontact time with different settings of D. It can be seen that the tail behavior follows an exponential decay even the destinations are randomly chosen. It is because the road networks eventually have the effect of gathering mobile vehicles on roads.

7 Conclusion and Future Work

In this paper, we have demonstrated that the intercontact times of taxies have an exponential-like tail distribution by mining the real trace data in Shanghai. To understand the fundamental reason that generates such a tail behavior, we further re-examined the data and have identified the impact of traffic influxes existing in most urban environments by

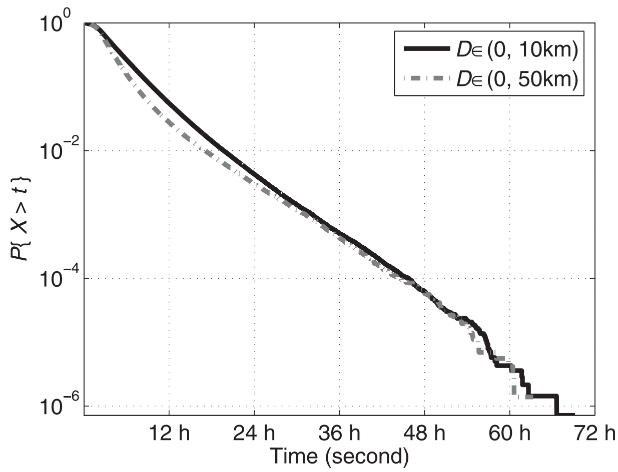

Fig. 7. Tail distribution of the intercontact time with real road networks in Shanghai on a linear-log scale.

theoretical analysis. We rigorously proved that the tail distribution of the intercontact time of any two vehicles follows an exponential decay as long as these vehicles have at least one constant traffic influx involved in their normal activities. Our results provide fundamental guidelines on design of new vehicular mobility models in urban scenarios, new data forwarding protocols and their performance analysis in VANETs.

There are still many aspects for us to investigate in the future. First, although we point out that urban vehicles have exponential tail distributions of intercontact time, it is hard to tell whether vehicles in different cities would follow the same distribution due to the complicated circumstances in urban settings. We will study this question based on trace data collected from different cities. Second, in reality, there can be multiple traffic influxes existing in the motion of a vehicle. The relationship between the geographical distribution of these traffic influxes and the tail distribution of intercontact time is uncertain and worth studying. Third, 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 where link quality shows very high dynamics. Thus, we will investigate the end-to-end delay since it is influenced not only by intercontact times, but also by retransmissions if the data transfer fails in a contact.

ACKNOWLEDGMENTS

This research was supported in part by China NSFC Grants 60933011, 60933012, 60903190, 60673166, 90612018, 60903190, 61027009, and 60970106, the National Basic Research Program of China (973 Program) under Grant No. 2006CB303000, the National Science and Technology Major Project of China under Grant No. 2009ZX03006-001 and Grant No. 2009ZX03006-004 from MIIT of China, Grant No. 2009AA012201 from China 863 program, the Science and Technology Planning Project of Guangdong Province, China under Grant No. 2009A080207002, the Science and Technology Commission of Shanghai Municipality under Grant No. 09511502603 and No.10dz1200204), Shanghai Pujiang Talents Grant 10PJ1405800, MIIT Grant 2009ZX03006-001-01, and STCSM Grant No. 05DZ15005.

REFERENCES

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

[2] A. Bar-Noy, I. Kessler, and M. Sidi, "Mobile Users: To Update or Not to Update?," *Proc. IEEE INFOCOM*, 1994.

[3] A.E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-delay Trade-Off in Wireless Networks," *Proc. IEEE INFOCOM*, 2004.

[4] G. Sharma and R. Mazumdar, "Scaling Laws for Capacity and Delay in Wireless Ad Hoc Networks with Random Mobility," *Proc. IEEE Int'l. Conf. Comm. (ICC)*, 2004.

[5] D. Johnson and D. Maltz, "Dynamic Source Routing in Ad Hoc Wireless Networks," *Mobile Computing*, T. Imelinsky and H. Korth, eds., pp. 153-181, Kluwer Academic Publishers, 1996.

[6] J. Broch, D. Maltz, D. Johnson, Y. Hu, and J. Jetcheva, "Multi-Hop Wireless Ad Hoc Network Routing Protocols," *Proc. ACM/IEEE MOBICOM*, 1998.

[7] C. Chiang and M. Gerla, "On-Demand Multicast in Mobile Wireless Networks," *Proc. IEEE Int'l Conf. Network Protocols (ICNP)*, 1998.

[8] P. Johansson, T. Larsson, N. Hedman, B. Mielczarek, and M. Degermark, "Routing Protocols for Mobile Ad-Hoc Networks—A Comparative Performance Analysis," *Proc. ACM/IEEE MOBICOM*, 1999.

[9] E. Royer, P.M. Melliar-Smith, and L. Moser, "An Analysis of the Optimum Node Density for Ad Hoc Mobile Networks," *Proc. IEEE Int'l Conf. Comm. (ICC)*, 2001.

[10] R. Groenevelt, P. Nain, and G. Koole, "Message Delay in MANET," *Proc. ACM SIGMETRICS*, 2004.

[11] G. Sharma and R.R. Mazumdar, "Delay and Capacity Trade-Off in Wireless Ad Hoc Networks with Random Mobility," *ACM/Kluwer J. Mobile Networks and Applications*, 2004.

[12] H. Cai and D.Y. Eun, "Crossing over the Bounded Domain: From Exponential to Power-Law Inter-Meeting Time in MANET," *Proc. ACM/IEEE MOBICOM*, 2007.

[13] T. Henderson, D. Kotz, and I. Abyzov, "The Changing Usage of a Mature Campus-Wide Wireless Network," *Proc. ACM Mobicom*, 2004.

[14] 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," *Proc. ACM SIGCOMM First Workshop Delay Tolerant Networking and Related Topics (WDTN '05)*, 2005.

[15] M. McNett and G.M. Voelker, "Access and Mobility of Wireless PDA User," technical report, Dept. of Computer Science and Eng., UC San Diego, 2004.

[16] 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," *Proc. ACM/IEEE MOBICOM*, 2007.

[17] A. Balasubramanian, B.N. Levine, and A. Venkataramani, "DTN Routing as a Resource Allocation Problem," *Proc. ACM SIGCOMM*, 2007.

[18] M. Li, M. Wu, Y. Li, J. Cao, L. Peng, Q. Deng, X. Lin, C. Jiang, W. Tong, Y. Gui, A. Zhou, X. Wu, and S. Jiang, "ShanghaiGrid: An Information Service Grid," *Concurrency and Computation: Practice & Experience*, vol. 18, no. 1, pp. 111-135, John Wiley and Sons Ltd., 2006.

[19] Y. Gunter and H.P. Großmann, "Usage of Wireless LAN for Inter-Vehicle Communication," *Proc. Eighth Int'l IEEE Conf. Intelligent Transportation Systems*, 2005.

[20] R. Lu, X. Lin, and X. Shen, "SPRING: A Social-Based Privacy-Preserving Packet Forwarding Protocol for Vehicular Delay Tolerant Networks," *Proc. IEEE INFOCOM*, 2010.

Hongzi Zhu received the MS degree in computer science from Ji Lin University in 2004 and the PhD degree from the Department of Computer Science and Engineering, Shanghai Jiao Tong University in 2009. He is a postdoctoral fellow in the Department of Electric and Computer Engineering at the University of Waterloo, Canada. His research interests include vehicular ad hoc Networks, wireless networks, distributed systems, and network security. He is a member of the IEEE, the IEEE Computer Society, and the IEEE Communication Society.

Minglu Li received the PhD degree in computer software from Shanghai Jiao Tong University in 1996. He is a full Professor and the vice dean in the School of Electronics Information and Electrical Engineering, the director in Grid Computing Center at the Shanghai Jiao Tong University. His research interests include grid computing, services computing, and sensor networks. He has published more than 100 papers in academic journals and international conferences. He is also a member of the Executive Committee of the Technical Committee on Services Computing of the IEEE Computer Society and a senior member of the IEEE.

Luoyi Fu is a postgraduate student in the Department of Electronic Engineering at the Shanghai Jiao Tong University. Her research interests include wireless ad hoc networks, Vehicular ad hoc Networks, and sensor networks.

Guangtao Xue received the PhD degree from the Department of Computer Science and Engineering, Shanghai Jiao Tong University in 2004. He is an associate professor in the Department of Computer Science and Engineering at the Shanghai Jiao Tong University. His research interests include peer-to-peer computing and mobile computing. He is a member of the IEEE and the IEEE Computer Society.

Yanmin Zhu received the BEng degree in computer science from the Xi'an Jiao Tong University in 2002 and the PhD degree in computer science from Hong Kong University of Science and Technology in 2007. He was a research associate in the Department of Computing, Imperial College London. Now, he is an associate professor in the Department of Computer Science and Engineering at the Shanghai Jiao Tong University. His research interests include ad-hoc sensor networks, mobile computing, grid computing, and resource management in distributed systems. He is a member of the IEEE and the IEEE Communication Society.

Lionel M. Ni received the PhD degree in electrical and computer engineering from Purdue University, West Lafayette, IN, in 1980. He is chair professor and was head in the Computer Science and Engineering Department at the Hong Kong University of Science and Technology. His research interests include parallel architectures, distributed systems, wireless sensor networks, high-speed networks, and pervasive computing. He has chaired many professional conferences and received a number of awards for authoring outstanding papers. He is a fellow of the IEEE.