

Revealing Patterns of Opportunistic Contact Durations and Intervals for Large Scale Urban Vehicular Mobility

Yong Li*, Depeng Jin*, Lieguang Zeng*, Sheng Chen^{†‡},

*Tsinghua National Laboratory for Information Science and Technology (TNLIST),
Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

[†]Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K.

[‡]King Abdulaziz University, Jeddah 21589, Saudi Arabia.

E-mail: liyong07@tsinghua.edu.cn

Abstract—Opportunistic contact between moving vehicles is one of the key features in vehicular delay tolerant networks (VDTNs) that critically influences the design of routing schemes and the network throughput. Due to prohibitive costs to collect enough realistic contact recodes, to the best of our knowledge, little experiment work has been conducted to study the opportunistic contact patterns in large scale urban vehicular mobility environment. In this work, we carry out an extensive experiment involving tens of thousands of operational taxis in Beijing city. Based on studying this newly collected *Beijing* trace and the existing *Shanghai* trace, we find some invariant characteristics of the opportunistic contacts for large scale urban VDTN. Specifically, in terms of contact duration, we find that there exists a characteristic time point, up to which and including at least 80% of the distribution, the contact duration obeys an exponential distribution, while beyond which it decays as a power law one. This property is in sharp contrast to the recent empirical data studies based on human mobility, where the contact duration exhibits a power law distribution. In terms of contact interval, we find that its distribution can be modelled by a three-segmented distribution, and there exists a characteristic time point, up to which the contact interval obeys a power law distribution, while beyond which it decays as an exponential one. Our observations thus reveal fundamental patterns for large scale vehicular mobility, and further provide useful guidelines for the design of new urban VDTN' routing protocols and their performance evaluation.

Index Terms—Urban vehicular mobility, contact interval, contact duration, mobility modeling

I. INTRODUCTION

Recently, interests on large-scale vehicular ad hoc networks have grown significantly [1], [2], as more and more vehicles are equipped with devices to provide vehicular communication capacities. Many applications of vehicular networks are also emerging, include automatic collision warning, remote vehicle diagnostics, emergency management and assistance for safely driving, vehicle tracking, automobile high speed Internet access, and multimedia content sharing. In USA, Federal Communications Commission has allocated 75 MHz of spectrum for dedicated short-range communications in vehicular networks [3], and IEEE is also working on the related

standard specifications. Efficient content dissemination is a key issue for many vehicular network applications, such as content publishing for safety information and entertainment data [5]. These content dissemination applications may be supported within the existing wireless infrastructure, such as WiFi and 3G, but the coverage issue and economic consideration from both service providers and end users do not make such a solution efficient and viable. Another reason is that the wireless infrastructure may be congested, or even damaged in disaster. Consequently, content dissemination through vehicular ad hoc networks is highly desired. Since vehicular networks are highly mobile and sometimes sparse by nature, it is difficult to maintain a connected network to communicate [6]. To transfer data from the source to the destination, vehicles need to first wait until they meet some other vehicles, i.e. until they are within the communication ranges of other vehicles [3], which is referred to as the opportunistic contact [4]. Opportunistic contact between vehicles is capable of providing high-bandwidth communication capacity for content dissemination in what is known as the vehicular delay tolerant network (VDTN) [6].

The performance of the VDTN based on such type of data transmission depends strongly on the vehicular mobility characteristics, especially on how often such opportunities occur and how long they last. Recent work [8] focuses on studying the metric of *contact interval*, which is the time between two successive communication contacts of two vehicles, and it finds that the contact interval exhibits the exponential distribution over a large range of timescales. The exponential distribution of contact interval revealed in the reference [8] is based on the largest existing vehicular trace of *Shanghai* [8], collected from the experiment involving 2,019 operational taxis for Shanghai city. Since this trace only involves about two thousands of vehicles for a huge urban city, it may not be sufficient for the purpose of studying the large scale vehicular system in modern urban environment which may include tens of millions vehicles.

Apart from *contact interval*, another key metric is known as *contact duration*, which is how long a contact lasts. The

contact duration is one of the most important factors directly affecting the amount of data that can be transferred between vehicles when they come into the communication range. It further influences the throughput and capacity of VDTN. Longer contact duration results in higher network throughput and larger capacity. Currently, empirical studies [9], [10] in human mobility environments have validated that the contact duration of human mobility follows a power law distribution, and the Pareto distribution can be used to model it. Rowstron *et al.* [12] investigated the influence of communication range and penetration rate on the inter-vehicle contact duration. They used the artificial data of 900-seconds long, generated by a traffic simulator [12], which is far from a real vehicular environment. Doering *et al.* [13] classified different types of contacts and examined the effects of radio range on contact duration. They found that the angle of contact is an appropriate criterion for the contact classification. They used a trace that includes 1600 bus nodes for 16 days, which may not be sufficient for extracting realistic and accurate model to characterize the distribution of contact duration. Contact durations are usually short for highly mobile vehicles, and a large data record is required to obtain an objective characterisation.

Revealing fundamental laws and properties of opportunistic contacts, validated in major city environments, can provide important hints on the characteristics of mobile nodes' behaviours in VDTN and will greatly benefit application designs, such as data dissemination schemes, in such network environments [14]. Due to lack of credible models for the opportunistic contacts, the current proposed forwarding protocols can only be investigated based on the models obtained from human mobility [6], [7], [15], which may lead to misleading results in the vehicular environment due to the significant differences between human and vehicular mobilities. Therefore, it is critical to have a profound understanding of the opportunistic contact distribution for the large scale urban vehicular mobility environments. Specifically, data collected from tens of thousands of vehicles over a long time period are needed in order to perform statistically significant and meaningful analysis. Due to the costly investment for such experiments, there is no existing work studying the opportunistic contact distribution in a realistic urban environment to the best of our knowledge.

Against this background, in this paper, we have collected real mobility traces from about 27,000 operational taxis during one month in Beijing city, which records the key mobility features of urban vehicles in a large city. We also use the existing mobility traces recorded in Shanghai city [8], although this record is smaller than the new Beijing trace. By analysing large volume of realistic urban vehicular mobility traces, in terms of contact duration, we find that there exists a characteristic time, before which the contact duration distribution is exponential, and beyond which the distribution decays as a power law. This is contrast to a pure power law exhibited in human mobility [9], [10]. By using the exponential and power law distribution to fit the empirical data, we demonstrate that this is an invariant and accurate property, which exists in the various scenarios investigated, including different cities and

different aggregate times. In terms of contact intervals, we find that it can be modelled by a three-segmented distribution, and there exists a characteristic time point, up to which the contact interval obeys a power law distribution, while beyond which it decays as an exponential one. This property is in sharp contrast to the recent empirical data studies vehicular mobility [8], where the contact interval exhibits exponential distribution. Our findings provide the fundamental guidelines on the design of new mobility models for urban vehicular environments, and new data forwarding and routing protocols as well as their accurate performance evaluation.

The rest of this contribution is organized as follows. After introducing the mobility trace collection and the data sets to be analyzed in Section II, we describe the method for the empirical data processing in Section III. In Sections IV, we present the obtained patterns of opportunistic contact duration and contact interval, respectively. Finally, we conclude the paper in Section V.

II. DATA SETS AND DEFINITIONS

Shanghai trace was collected by SG project [16], in which 2,019 operational taxis continuously covered the whole month of February 2007 without any interruptions in Shanghai city. In this trace, a taxi sends its position report by GPRS to the central database every 1 minute when it has passengers onboard but every 15 seconds when it is vacant for the reason of real-time scheduling. However, the different intervals of reporting may distort the records of the physical movements of the taxis, since most of taxis are not vacant most of the time. Another drawback of this trace is that the number of taxis is limited. Indeed, even 2000 taxis and 1 minute duration may not be sufficient to record the statistical features of contact duration in a high-speed large urban environment.

In collecting *Beijing* trace, we used the mobility track logs obtained from 27,000 participating Beijing taxis carrying GPS receivers as shown in Fig. 1 during the whole May month in 2010. The reason for us to also choose taxis as vehicular devices is that taxis are more sensitive to urban environments



Fig. 1. The illustration of the deployed vehicular mobility trace collection system.

in terms of underlying road topology, traffic control and urban planning, and they have broader coverage in terms of space and operation time than buses and private cars. Specifically, we utilized the GPS devices to collect the taxis locations and timestamps and GPRS modules to report the records every 15 seconds for moving taxis. The specific information contained in such a report includes: the taxi's ID, the longitude and latitude coordinates of the taxi's location, timestamps, instant speed and heading. *Beijing* trace is the largest vehicular data trace available.

Consider two arbitrary vehicles, v_i and v_j , each of which moves according to the collected trace that covers the region $\Omega \subset \mathbb{R}^2$. Let $L_{v_i}(t) \in \Omega$ and $L_{v_j}(t) \in \Omega$ be the positions of vehicular v_i and v_j at time t , respectively, where t is in continuous-time scale. We assume that any two vehicles can communicate with each other whenever they are within the communication range $R_{\text{Tx}}^{v_i, v_j}$. Based on these notations, we now formally define the contact duration and the contact interval of vehicles v_i and v_j as follows.

Definition 1 (Contact Duration): Assume that v_i and v_j come within the range of each other at time t_c , that is, $\|L_{v_i}(t_c^-) - L_{v_j}(t_c^-)\| > R_{\text{Tx}}^{v_i, v_j}$ and $\|L_{v_i}(t_c) - L_{v_j}(t_c)\| = R_{\text{Tx}}^{v_i, v_j}$, where t_c^- denotes the time before t_c . The contact duration of v_i and v_j is defined as the time duration during which they are in contact with each other before moving out of the range of each other, that is, $CD_{v_i, v_j} = t - t_c$ with $\min\{t : \|L_{v_i}(t) - L_{v_j}(t)\| > R_{\text{Tx}}^{v_i, v_j}\}$, where t and t_c are in the continuous-time scale.

Definition 2 (Contact Interval): The contact interval of v_i and node v_j , denoted as CI_{v_i, v_j} , is defined as the time interval that takes them to come within the range of each other again from the last time, denoted as t_0 , when they were moving out of the range of each other, that is,

$$CI_{v_i, v_j} = \min_t \{(t - t_0) : \|L_{v_i}(t) - L_{v_j}(t)\| \leq R_{\text{Tx}}^{v_i, v_j}, t > t_0\}.$$

III. EMPIRICAL DATA PROCESSING

To obtain accurate contact duration and contact interval, we need to know the exact beginning and ending times of each opportunistic contact. However, GPS reports were collected in discrete time at the time interval of 15 seconds. Therefore, the times that an opportunistic contact starts and ends, respectively, may not be recorded in the timestamps. Consequently, the trace need to be processed in order to extract the contact duration and contact interval. In extracting the contact from the GPS trace, it is assumed that two taxis are able to communicate with each other if their estimated locations are within a given communication range and within the same specific time duration, which is called a contact. We note that in reality even if two taxis are within the communication range, they may not be able to successfully transfer data due to the interference and signal loss of the wireless link. Since we focus on the distributions of contact duration and contact interval, we only investigate the potential communication opportunities between any two taxis, and we will not discuss the issues related to how to ensure the successful data transmissions during contacts.

To extract contacts from the GPS traces, the existing work [8] uses a time window to measure the contact. However, as reported in [8], it is difficult to select a proper time window, since a large time window may introduce false contacts while a small window may miss some real connection opportunities. To overcome this drawback, in this work, we use the method of linear interpolation (LI) to extract the opportunistic contact. First, we select the time points near and inside the chosen contact, which should at least include one time point during the contact, one point before the contact and one after the contact. Then, we use the selected time points to estimate the contact beginning and ending times. The details of this method can be found in Ref. [11].

We note that the LI method is based on the assumption that the distance between two taxis changes linearly closed to and during the communication contact. Since in real urban vehicular environments, a contact occurs often when two taxis move nearby each other, e.g. one taxis is passing another one when the road is straight. Therefore, an LI is sufficient for us to extract the contact. In order to obtain more accurate results, we use several different sets of times points and different communication ranges in the LI based estimation. More specifically, in the contact extraction, we usually choose three sets of time points near and within the contact concerned to obtain three LI estimates, and use the average value as the final contact estimate. We also consider four different communication ranges of 50, 100, 200 and 300 meters for the experiment.

In this work, we investigate the complementary cumulative distribution functions (CCDFs) of contact duration and contact interval. More specifically, we focus on the *aggregate* CCDFs of the contact durations and contact intervals between all the devices, which are the CCDFs of per contact sample of the contact durations and contact intervals over all the distinct pairs of devices. For simplicity, we often use the notation without explicitly mentioning ‘‘aggregate’’.

IV. RESULTS AND ANALYSIS

A. Dichotomy of Contact Duration

The empirical distributions of contact duration between taxis extracted from the two vehicular mobility traces introduced in Section II were examined, and the CCDFs of contact duration obtained are reported using the log-log scale in Figs. 2 to 4 in Ref. [11], for different experimental conditions. From all the empirical distributions of contact duration, it can be observed that the tails of all the plots are almost straight lines with different negative slopes on the log-log scale beyond approximately 100 and 50 seconds for *Beijing* and *Shanghai* traces, respectively, over two very different timescales of one month and one day. This ‘‘knee’’ in a distribution plot will be referred to as the *characteristic time*. Therefore, the contact duration exhibits a very clear power law tail. Note that the power law distribution was reported in the previous works [9], [10] that studied human mobility. It is seen that the same property also holds for the distribution tail under vehicular

mobility. A new and more important information, however, can be confidently deduced from the empirical results, namely, before the *characteristic time*, the CCDF exhibits an exponential distribution, which counts at least 80% of the entire distribution.

B. Contact Interval Patterns

The aggregated empirical distributions of contact interval between vehicles were next examined, and the CCDFs of contact interval obtained are plotted using the log-log scale in Figs. 2 and 3 for *Beijing* and *Shanghai* traces, respectively, where part (a) of the figures presents the aggregate contact intervals of all vehicles under the two different communication ranges of $R = 100$ and $R = 300$ metres, respectively, during the whole trace collection time, while part (b) of the figures presents the distribution of aggregate contact interval in the time scale of one day under the communications range of $R = 200$ metres. By observing the empirical distributions shown in Figs. 2 and 3 as the red solid curves, it can be seen that each empirical distribution can be divided into the

three segments over the intervals $[0, 100) s$, $[100, 3000) s$ and $[3000, 60000) s$, respectively. Specifically, we can use the compound function of $y = b * e^{-x^a}$ to fit the first empirical distribution segment, as well as use a power law distribution and an exponential distribution to fit the middle and tail parts, respectively. Furthermore, from the empirical distributions of contact intervals, it can be observed that between the middle and tail parts, there exists a “knee” or *characteristic time* in the distribution plot. In particular, the contact interval exhibits a power law distribution before the characteristic time and an exponential tail after the characteristic time, which is in sharp contrast to the results given in the reference [8].

In order to verify the accuracy of the three-segmented distribution model for the contact interval of urban vehicular mobility, we use this three-segmented distribution model to fit the first, middle and tail parts of all the empirical curves, partitioned by the characteristic time, in Figs. 2 and 3, respectively. The goodness of fit is again measured quantitatively by the R-square statistics, and they are listed in Table I. It can be

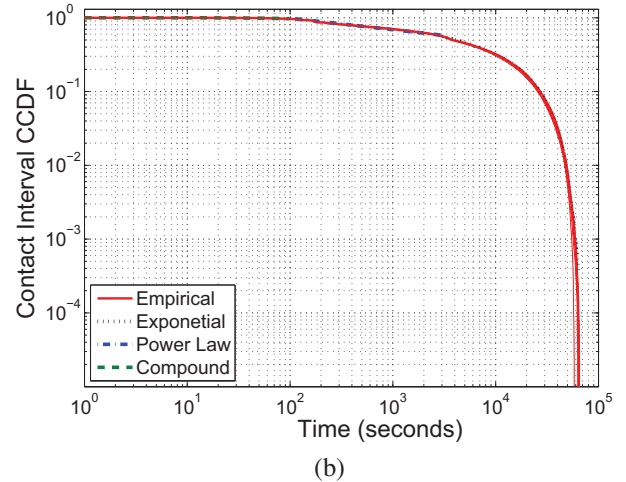
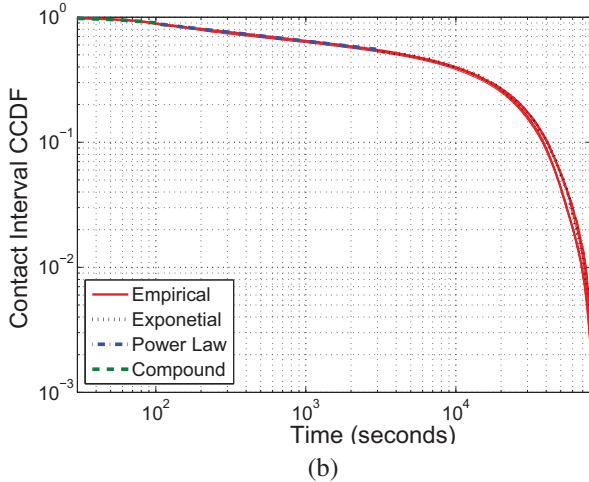
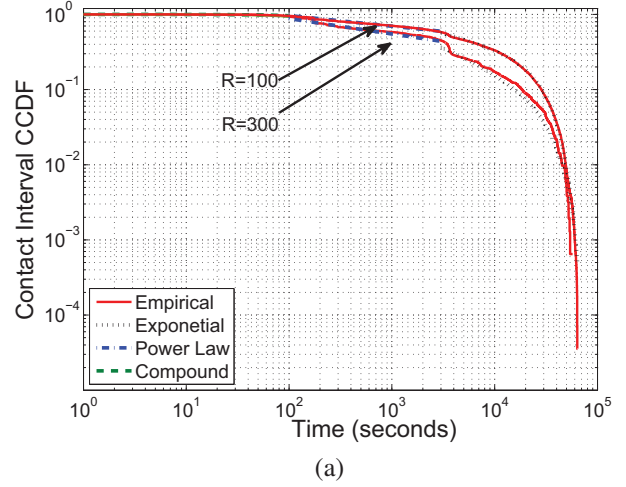
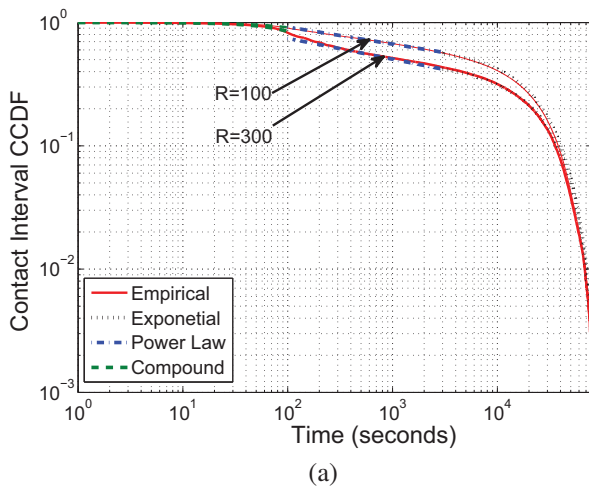


Fig. 2. Contact rates patterns of *Beijing* trace: (a) CCDF of contact interval for the whole month ($R = 100$ and $R = 300$), and (b) CCDF of contact interval for each day in a week ($R = 200$)

Fig. 3. Contact rates patterns of *Shanghai* trace: (a) CCDF of contact interval for the whole month ($R = 100$ and $R = 300$), and (b) CCDF of contact interval for each day in a week ($R = 200$)

TABLE I

THE ADJUSTED R-SQUARE STATISTICS OF THE EXPONENTIAL, POWER LAW AND COMPOUND FITTINGS TO THE DATA OF CONTACT INTERVAL OBTAINED FROM THE TWO TRACES OF *Beijing* AND *Shanghai* SHOWN IN FIGS. 2 AND 3.

Transmission range	$R = 100$	$R = 200$	$R = 300$
<i>Beijing</i> (Exponential Part)	99.61%	99.99%	99.99%
<i>Beijing</i> (Power law Part)	99.32%	99.69%	99.46%
<i>Beijing</i> (Compound Part)	99.29%	99.41%	99.68%
<i>Shanghai</i> (Exponential Part)	97.92%	99.66%	99.54%
<i>Shanghai</i> (Power law Part)	98.04%	98.78%	98.38%
<i>Shanghai</i> (Compound Part)	99.81%	99.95%	99.88%

seen from Table I that the average adjusted R-square statistics are all over 98% for the three parts. This confirms the accuracy of the three-segmented distribution model for contact interval.

V. CONCLUSIONS

We have revealed the dichotomy model of contact duration and the three-segmented model of contact interval in vehicular mobility. The contact duration model consists of an exponential distribution up to the point of characteristic time and a decaying power law tail beyond this point. This dichotomy hypothesis has been validated with two very different large-scale vehicular mobility traces of *Beijing* and *Shanghai*. Different from the existing results found in human mobility, in vehicular mobility, the first part of contact duration distribution, which counts for at least 80% of the total distribution, obeys an exponential distribution. This means that the contact duration in vehicular networks decays more quickly, which suggests that the existing forwarding schemes based on the power law distribution may be overly pessimistic. We can infer that the basic reason for faster decaying is because vehicles usually move faster than human, which reduce the contact duration. We also note that the characteristic times are different for different cities, which may reflect different road and street structures. Our three-segmented distribution model of contact interval, namely, the first compound model, the middle power-law distribution and the exponential tail distribution, separated by the characteristic time, is in sharp contrast to the existing results of the exponential distribution of contact interval for vehicular mobility.

In this work, we focus on the *aggregate* contact distribution. Further research is warranted to investigate the pairwise distributions of contact duration and contact interval, and how they vary for different time periods and regions. Our future work will also address the fundamental questions of what are the essential factors that generates the dichotomy and three-segmented behaviours of the opportunistic contact of taxis and how the characteristic time is determined by these factors as well as how to utilise these opportunistic contact models for helping the routing and security protocol design in VDTN [18].

VI. ACKNOWLEDGMENT

This work is supported by National Basic Research Program of China (973 Program) (No. 2013CB3291005), National Nature Science Foundation of China (No. 61171065, No.

61021001, No. 61133015), National High Technology Research and Development Program (No. 2013AA010601 and No. 2013AA010605), and Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT). The authors acknowledge the contributions of Professor Zhaocheng Wang to the work reported in this paper.

REFERENCES

- [1] M. Khabazian, S. Aissa, and M. Mehmet-Ali, "Performance modeling of message dissemination in vehicular ad hoc networks with priority," *IEEE J. Selected Areas in Communications*, vol. 29, no. 1, pp. 61–71, Jan. 2011.
- [2] Y. Li, Z. Wang, D. Jin, L. Zeng, and S. Chen, Collaborative vehicular content dissemination with directional antennas, *IEEE Transactions on Wireless Communications*, vol. 11, no. 4, pp. 1301–1306, 2012.
- [3] J. Zhao and G. Cao, "VADD: Vehicle-assisted data delivery in vehicular ad hoc networks," in *Proc. 25th IEEE INFOCOM* (Barcelona, Spain), April 23–29, 2006, pp. 1–12.
- [4] Y. Li, P. Hui, D. Jin, L. Su, and L. Zeng, Evaluating the impact of social selfishness on the epidemic routing in delay tolerant networks, *IEEE Communications Letters*, vol. 14, no. 11, pp. 1026–1028, 2010.
- [5] M. Johnson, L. De Nardis, and K. Ramchandran, "Collaborative content distribution for vehicular ad hoc networks," in *Proc. 44th Allerton Conf. Communication, Control, and Computing* (Monticello, IL), Sept. 27–29, 2006, pp. 751–760.
- [6] D. Camara, N. Frangiadakis, F. Filali, and C. Bonnet, "Vehicular delay tolerant networks," In: M. M. Cruz-Cunha and F. Moreira, Eds., *Handbook of Research on Mobility and Computing: Evolving Technologies and Ubiquitous Impacts*. IGI Global, 2011, pp. 356–367.
- [7] Y. Li, Y. Jiang, D. Jin, L. Su, L. Zeng, and D. Wu, Energy-efficient optimal opportunistic forwarding for delay-tolerant networks, *IEEE Transactions on Vehicular Technology*, vol. 59, no. 9, pp. 4500–4512, 2010.
- [8] H. Zhu, M. Li, L. Fu, G. Xue, Y. Zhu, and L. M. Ni, "Impact of traffic influxes: Revealing exponential inter-contact time in urban VANETs," *IEEE Trans. Parallel and Distributed Systems*, vol. 22, no. 8, pp. 1258–1266, Aug. 2011.
- [9] W. Wang, V. Srinivasan, and M. Motani, "Adaptive contact probing mechanisms for delay tolerant applications," in *Proc. 13th ACM Int. Conf. Mobile Computing and Networking (MOBICOM)* (Montréal, Canada), Sept. 9–14, 2007, pp. 230–241.
- [10] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Pocket switched networks: Real-world mobility and its consequences for opportunistic forwarding," *Tech. Rep. UCAM-CL-TR-617*, University of Cambridge, Computer Lab, Feb. 2005.
- [11] Y. Li, D. Jin, Z. Wang, L. Zeng, and S. Chen, Exponential and Power Law Distribution of Contact Duration in Urban Vehicular Ad Hoc Networks, *IEEE Signal Processing Letters*, vol. 20, no. 1, pp. 110–113, 2013.
- [12] A. Rowstron and G. Pau, "Characteristics of a vehicular network," *Tech. Rep.*, University of California at Los Angeles, Feb. 2009.
- [13] M. Doering, W. Pöttner, T. Pögel, and L. Wolf, "Impact of radio range on contact characteristics in bus-based delay tolerant networks," in *Proc. 8th Int. Conf. Wireless On-Demand Network Systems and Services* (Bardonecchia, Italy), Jan. 26–28, 2011, pp. 195–202.
- [14] Y. Li, G. Su, D. Wu, D. Jin, L. Su, and L. Zeng, The impact of node selfishness on multicasting in delay tolerant networks, *IEEE Transactions on Vehicular Technology*, vol. 60, no. 5, pp. 2224–2238, 2011.
- [15] X. Zhuo, Q. Li, W. Gao, G. Cao, and Y. Dai, "Contact duration aware data replication in delay tolerant networks," in *Proc. 19th IEEE Int. Conf. Network Protocols* (Vancouver, Canada), Oct. 17–20, 2011, pp. 236–245.
- [16] M. Li, H. Zhu, Y. Zhu, and L. M. Ni, "Ants: Efficient vehicle locating based on ant search in shanghai-grid," *IEEE Trans. Vehicular Technology*, vol. 58, no. 8, pp. 4088–4097, Oct. 2009.
- [17] B. S. Everitt, *The Cambridge Dictionary of Statistics* (3rd Edition). Cambridge, UK: Cambridge University Press, 2006.
- [18] Y. Li, P. Hui, D. Jin, L. Su, and L. Zeng, Optimal distributed malware defense in mobile networks with heterogeneous devices, *IEEE Transactions on Mobile Computing*, 13 Dec. 2012. IEEE computer Society Digital Library. IEEE Computer Society, <http://doi.ieeecomputersociety.org/10.1109/TMC.2012.255>