

# Exponential and Power Law Distribution of Contact Duration in Urban Vehicular Ad Hoc Networks

Yong Li, *Student Member, IEEE*, Depeng Jin, *Member, IEEE*, Zhaocheng Wang, *Senior Member, IEEE*, Lieguang Zeng, and Sheng Chen, *Fellow, IEEE*

**Abstract**—Contact duration between moving vehicles is one of the key metrics in vehicular ad hoc networks (VANETs), that critically influences the design of routing schemes and network throughput. Due to prohibitive costs to collect enough realistic contact records, little experimental work has been conducted to study the contact duration in urban VANETs. 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 an invariant characteristic that there exists a characteristic time point, up to which the contact duration obeys an exponential distribution that includes at least 80% of the whole 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. Our observations thus provide fundamental guidelines for the design of new urban VANETs' routing protocols and their performance evaluation.

**Index Terms**—Contact duration, mobility modelling, urban vehicular ad hoc network.

## I. INTRODUCTION

URBAN vehicular ad hoc networks (VANETs) are recognised as an important element of the future intelligent transport system to alleviate serious problems, such as traffic jams and accidents, and to enable new mobile applications to the public [1]. In VANETs, network nodes equipped with wireless communications devices enable communication between vehicles and fixed roadside infrastructure or between vehicle and vehicle [2]. Since urban VANETs are highly mobile and sometimes sparse, it is hard to maintain a connected network to communicate. In order 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]. The performance of networks based on such type of data transmission depend strongly on the vehicular mobility characteristics, especially on how often such communication opportunities occur and how long they last. Recent work [4] focuses on studying the metric of *inter-contact time*, which is the time between two successive communication contacts of two vehicles, and it finds that the inter-contact time exhibits the exponential distribution over a large range of timescales. In this study, we focus on another key metric 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 range. It further influences the throughput and capacity of opportunistic VANETs. Longer contact duration results in higher network throughput and larger capacity. Therefore, revealing its fundamental laws and properties, validated in major city environments, can provide important hints on the characteristics of mobile nodes behaviours in VANETs, and will greatly benefit application designs, such as data dissemination schemes, in such network environments.

Empirical studies [5], [6] 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.* [7] investigated the influence of communication range and penetration rate on the inter-vehicle contact duration, using the 900-seconds data generated by a traffic simulator [7], which is far from reflecting the conditions of a real vehicular environment. Doering *et al.* [8] classified different types of contacts and examined the effects of radio range on contact duration, using a trace that includes 1600 bus nodes operating for 16 days. However, this study is only valid for bus based networks for obvious reasons, and it does not provide a realistic and accurate model to characterize the distribution of contact duration for a generic urban VANET. Due to lack of credible models for the urban vehicular environment, the current proposed forwarding protocols can only be investigated based on the models obtained from human mobility [9], [10], 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 contact duration distribution for opportunistic vehicular environments. However, different from inter-contact times, contact durations are usually short for highly mobile vehicles, and a large data record is required to obtain an objective characterisation. 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

Manuscript received August 28, 2012; revised November 13, 2012; accepted November 26, 2012. Date of publication December 03, 2012; date of current version December 07, 2012. This work was supported by National Nature Science Foundation of China under Grants 61171065 and 61021001, the Chinese National Major Scientific and Technological Specialized Project under Grant 2010ZX03004-002-02, and by the Program for Changjiang Scholars and Innovative Research Team in University. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Alireza Seyedi.

Y. Li, D. Jin, Z. Wang, and L. Zeng are with Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: yong-li07@mails.tsinghua.edu.cn; jindp@tsinghua.edu.cn; zcwang@tsinghua.edu.cn; zenglg@tsinghua.edu.cn).

S. Chen is with Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K., and also with King Abdulaziz University, Jeddah 21589, Saudi Arabia (e-mail: sqc@ecs.soton.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LSP.2012.2231412

to the costly investment for such experiments, there is no existing work studying the contact duration distribution in a realistic urban environment to the best of our knowledge.

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 [4], although this record is smaller than the new Beijing trace. By analysing a large volume of realistic urban vehicular mobility traces, 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 in contrast to a pure power law exhibited in human mobility [5], [6]. Our results of using an exponential and power law distribution to fit the empirical data in various scenarios for the two cities over different aggregate times support our hypothesis that this is an invariant and accurate property. Our findings provide the fundamental guidelines on design of new mobility models for urban vehicular environments, and new data forwarding and routing protocols as well as their accurate performance evaluation.

## II. DATA SETS AND DEFINITIONS

*Shanghai* trace [4] was collected by SG project [11], 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. *Shanghai* trace was the second largest available vehicular data record after *Beijing* trace, and it is a precious and valuable resource for investigating vehicular mobility, considering the huge costs and efforts involved in collecting this real vehicular contact data. It would be ideal however if vehicles involved in the original experiment were much larger than 2000, simply to better cover Shanghai's large urban environment.

In collecting *Beijing* trace, we used the mobility track logs obtained from 27 000 participating Beijing taxis carrying GPS receivers 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 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.

We now define the contact duration between vehicles. 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  $R$  of each other. Based on these notations, we formally define the contact duration 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,  $\|L_{v_i}(t_c^-) - L_{v_j}(t_c^-)\| > R$  and  $\|L_{v_i}(t_c) - L_{v_j}(t_c)\| = R$ , where  $t_c^-$  denotes the time before  $t_c$ . The contact duration of  $v_i$  and  $v_j$  is defined as the time 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  $\min_{t-t_c} \{t : \|L_{v_i}(t) - L_{v_j}(t)\| > R\}$ , where  $t$  and  $t_c$  are in the continuous-time scale.

## III. EMPIRICAL DATA PROCESSING

To obtain an accurate contact duration, we need to know the exact beginning and ending times of each contact. However, GPS reports were collected in discrete time at the time interval of 15 seconds. Therefore, the times that a contact starts and ends, respectively, may not be recorded in timestamps. Consequently, the trace need to be processed in order to extract the contact duration. In extracting the contact from 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 distribution of contact duration, 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 [4] uses a time window to measure the contact. However, as reported in [4], 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 contact duration, which is illustrated in Fig. 1. 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. In Fig. 1, we have selected  $t_1$ ,  $t_2$  and  $t_3$  as the three such time points recorded in the trace. The contact's beginning time  $t_c$  and ending time  $t_d$  can be estimated by the LI expressed as follows,

$$t_c = t_1 + \frac{\bar{L}(t_1) - R}{\bar{L}(t_1) - \bar{L}(t_2)}(t_2 - t_1),$$

$$t_d = t_2 + \frac{\bar{L}(t_3) - R}{\bar{L}(t_3) - \bar{L}(t_2)}(t_3 - t_2),$$

with  $\bar{L}(t_k) = \|L_{v_i}(t_k) - L_{v_j}(t_k)\|$ . Consequently, the contact duration can be obtained by  $t_d - t_c$ .

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

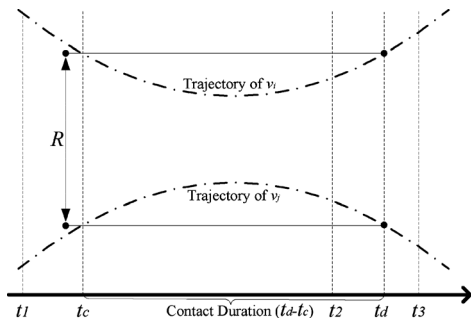


Fig. 1. Extract contacts from GPS reports of vehicles  $v_i$  and  $v_j$ . Two arc curves are the trajectories of  $v_i$  and  $v_j$ , and  $R$  is the communication range.  $t_1$ ,  $t_2$  and  $t_3$  are three timestamps in the GPS records, and the positions of the vehicles at  $t_1$ ,  $t_2$  and  $t_3$  are reported by GPS.  $t_c$  is the time that the contact begins, and  $t_d$  is the end of the contact. Since both  $t_c$  and  $t_d$  are not recorded, they have to be estimated to obtain the contact duration.

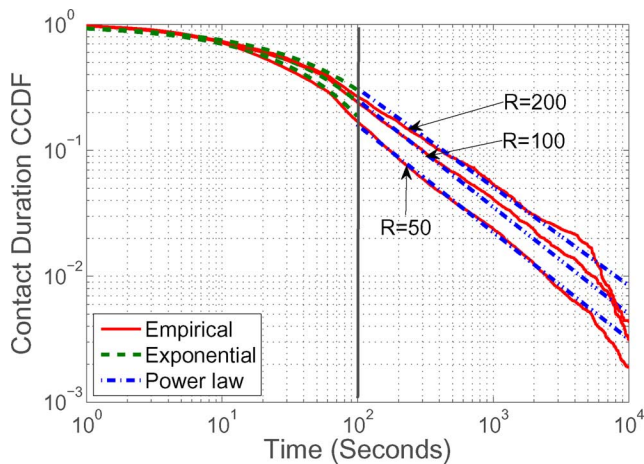


Fig. 2. CCDF of contact duration for *Beijing* trace, where data are collected from 27 000 taxis in Beijing city during the whole month of May 2010, and the three empirical curves are the aggregated contact durations of the whole month under three communication ranges of  $R = 50$ , 100 and 200 m, respectively.

nearby each other, e.g., one taxis is passing another one. Therefore, an LI may be sufficient for us to extract the contact. In order to obtain more accurate results, we use several different sets of times points and three 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 duration. We also consider three different communication ranges of 50, 100 and 200 m for both *Shanghai* and *Beijing* traces.

In this work, we investigate the complementary cumulative distribution function (CCDF) of contact duration. More specifically, we focus on the *aggregate* CCDF of the contact duration between all the devices, which is the CCDF of per contact sample of the contact duration over all the distinct pairs of devices. For simplicity, we often use the notation without explicitly mentioning “aggregate”. To investigate the influence of the different “aggregate” times, we study the two cases that the “aggregate” times cover the trace time of whole month and over one day, respectively.

#### IV. DICHOTOMY OF CONTACT DURATION

The empirical distributions of contact duration between taxis extracted from the two vehicular mobility traces introduced

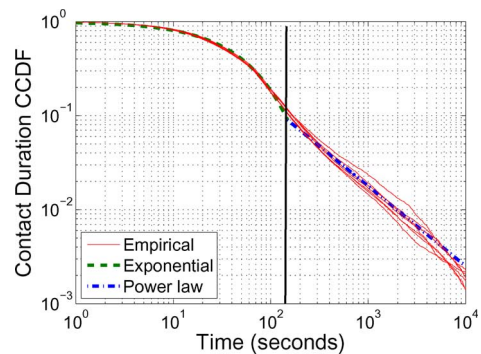


Fig. 3. CCDF of contact duration for *Beijing* trace, where data are collected from 27 000 taxis in Beijing city during the first week of May 2010, and the seven empirical curves are the aggregated contact durations of each day in the week, respectively, under communication ranges of  $R = 200$  m.

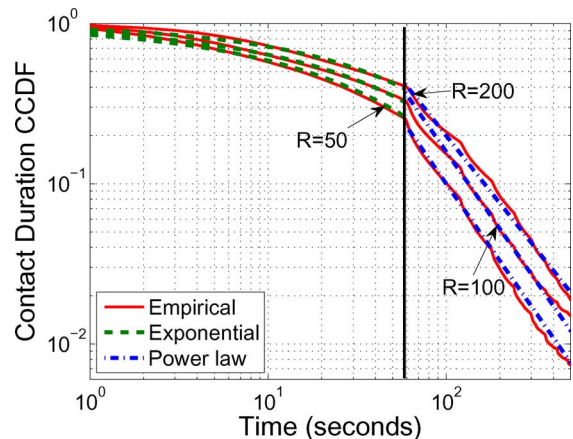


Fig. 4. CCDF of contact duration for *Shanghai* trace, where data are collected from 2,019 taxis in Shanghai city during the whole month of February 2007, and the three empirical curves are the aggregated contact durations of the whole month under three communication ranges of  $R = 50$ , 100 and 200 m, respectively.

above were examined, and the obtained CCDFs of contact duration, i.e.,  $P(X > t)$ , are plotted using log-log scale in Figs. 2–4, for different experimental conditions. In particular, Figs. 2 and 4 present the aggregate contact durations of all taxis under three different communication ranges,  $R = 50$  m,  $R = 100$  m and  $R = 200$  m, during the whole month of the trace collection time for *Beijing* and *Shanghai* traces, respectively, while Fig. 3 presents the distribution of aggregate contact duration in the time scale of one day with the transmission range of  $R = 200$  m for *Beijing* trace.

From all the empirical distributions of contact duration that are plotted by the red curves in the three figures, it can be observed that the tails of all the plots are almost straight lines with different negative slopes on log-log scale beyond approximately 100 and 50 seconds for *Beijing* and *Shanghai* traces, respectively, over a large range of timescale. 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 [5], [6] 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 of Figs. 2–4, namely,

TABLE I

THE ADJUSTED R-SQUARE STATISTICS OF THE EXPONENTIAL AND POWER LAW FITTINGS TO THE DATA OF CONTACT DURATION OBTAINED FROM THE TWO TRACES OF *BEIJING* AND *SHANGHAI* SHOWN IN FIGS. 2 AND 4.

Transmission range	$R = 50$	$R = 100$	$R = 200$
<i>Beijing</i> (Exponential Part)	99.70%	99.79%	97.12%
<i>Beijing</i> (Power law Part)	96.28%	96.39%	94.78%
<i>Shanghai</i> (Exponential Part)	99.66%	99.54%	99.61%
<i>Shanghai</i> (Power law Part)	97.72%	99.00%	98.77%

before the *characteristic time*, the CCDF exhibits an exponential distribution, which counts at least 80% of the entire distribution. By carefully examining the results for *Beijing* trace shown in Fig. 2, for example, it can be observed that at approximately 100 seconds, the CCDF has a knee before which the empirical CCDF fits well an exponential distribution but after which the distribution decays abruptly to a power law one. For the results of *Shanghai* trace depicted in Fig. 4, the same observation can be made but the characteristic time is now approximately 50 seconds. We note the different characteristic times for Beijing and Shanghai. Obviously, street grids, ring roads and bypasses, and other infrastructures are very different in the two cities, which may induce the different encounter properties between vehicles. By changing the aggregate time from the whole month to a day, whose results are represented in Fig. 3, we observe the same phenomenon, namely, the contact duration exhibits an exponential distribution before the characteristic time and a power law tail after the characteristic time.

In order to verify the accuracy of the dichotomy model of exponential distribution and power law tail for the contact duration of urban VANETs, we use the exponential distribution and power law to fit the first and second parts, partitioned by the characteristic time, of all the empirical curves in the three figures, respectively. The goodness of fit is measured quantitatively by the R-square statistics [12], which is defined as the percentage of the variation between the empirical CCDF and the fitted distribution. Table I lists the adjusted R-square statistics of the exponential and power law fittings, respectively, to the empirical data of contact duration, where the adjusted R-square statistics are computed with Matlab Curve Fitting Toolbox. It can be seen from Table I that the average adjusted R-square statistics is over 97% for the exponential fitting part, and it is over 94% for the power law fitting part. This confirms the accuracy of the dichotomy model of contact duration, namely, an exponential distribution before the characteristic time with a fast decaying power law tail.

## V. DISCUSSIONS AND CONCLUSIONS

We have revealed the dichotomy model of contact duration in vehicular mobility, which 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 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 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.

In this work, we focus on the *aggregate* contact duration distribution. Further research is warranted to investigate the pairwise distributions of contact duration and how they vary from each other in different time periods and regions. Our future work will also address the fundamental questions of what are the essential factors that generates the dichotomy behavior of the contact duration of taxis and how the characteristic time is determined by these factors as well as how to utilise the dichotomy based contact duration model for helping the routing and forwarding protocol design in urban VANETs.

## REFERENCES

- [1] M. Khabazian, S. Aissa, and M. Mehmet-Ali, "Performance modeling of message dissemination in vehicular ad hoc networks with priority," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 1, pp. 61–71, Jan. 2011.
- [2] J. J. Blum, A. Eskandarian, and L. J. Hoffman, "Challenges of intervehicle ad hoc networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 347–351, Dec. 2004.
- [3] J. Zhao and G. Cao, "VADD: Vehicle-assisted data delivery in vehicular ad hoc networks," in *Proc. 25th IEEE INFOCOM*, Barcelona, Spain, Apr. 23–29, 2006, pp. 1–12.
- [4] 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," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 8, pp. 1258–1266, Aug. 2011.
- [5] 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)*, Montreal, QC, Canada, Sep. 9–14, 2007, pp. 230–241.
- [6] 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*. Univ. Cambridge, Computer Lab, U.K., 2005, Tech. Rep. UCAM-CL-TR-617.
- [7] A. Rowstron and G. Pau, *Characteristics of a Vehicular Network*. Univ. California, Los Angeles, 2009, Tech. Rep. UCAM-CL-TR-617.
- [8] 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 Syst. and Services*, Bardonecchia, Italy, Jan. 26–28, 2011, pp. 195–202.
- [9] D. Camara, N. Frangiadakis, F. Filali, and C. Bonnet, "Vehicular delay tolerant networks," in *Handbook of Research on Mobility and Computing: Evolving Technologies and Ubiquitous Impacts*, M. M. Cruz-Cunha and F. Moreira, Eds. : IGI Global, 2011, pp. 356–367.
- [10] X. Zhuo, Q. Li, W. Gao, G. Cao, and Y. Dai, "Contact duration aware data replication in delay tolerant networks," in *Proc. 19th IEEE ICNP*, Vancouver, BC, Canada, Oct. 17–20, 2011, pp. 236–245.
- [11] M. Li, H. Zhu, Y. Zhu, and L. M. Ni, "Ants: Efficient vehicle locating based on ant search in shanghai grid," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4088–4097, Oct. 2009.
- [12] B. S. Everitt, *The Cambridge Dictionary of Statistics*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 2006.