

Limits of Predictability for Large-Scale Urban Vehicular Mobility

Yong Li, *Member, IEEE*, Depeng Jin, *Member, IEEE*, Pan Hui, *Member, IEEE*,
Zhaocheng Wang, *Senior Member, IEEE*, and Sheng Chen, *Fellow, IEEE*

Abstract—Key challenges in vehicular transportation and communication systems are understanding vehicular mobility and utilizing mobility prediction, which are vital for both solving the congestion problem and helping to build efficient vehicular communication networking. Most of the existing works mainly focus on designing algorithms for mobility prediction and exploring utilization of these algorithms. However, the crucial questions of how much the mobility is predictable and how the mobility predictability can be used to enhance the system performance are still the open and unsolved problems. In this paper, we consider the fundamental problem of the predictability limits of vehicular mobility. By using two large-scale urban city vehicular traces, we propose an intuitive but effective model of areas transition to describe the vehicular mobility among the areas divided by the city intersections. Based on this model, we examine the predictability limits of large-scale urban vehicular networks and obtain the maximal predictability based on the methodology of entropy theory. Our study finds that about 78%–99% of the location and above 70% of the staying time, respectively, are predicible. Our findings thus reveal that there is strong regularity in the daily vehicular mobility, which can be exploited in practical prediction algorithm design.

Index Terms—Mobility modeling, mobility prediction, predictability limits, vehicular networks.

I. INTRODUCTION

URBAN vehicular traffic congestion is an increasingly serious problem that is significantly affecting many aspects of the quality of metropolis life around the world [1]. Scientific traffic engineering, which aims to achieve efficient management

Manuscript received August 20, 2013; revised March 12, 2014; accepted May 3, 2014. Date of publication June 17, 2014; date of current version December 1, 2014. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2013CB329001; by the National Nature Science Foundation of China under Grant 61301080, Grant 61171065, and Grant 61273214; by the National High Technology Research and Development Program under Grant 2013AA013501 and Grant 2013AA013505; by the Chinese National Major Scientific and Technological Specialized Project under Grant 2013ZX03002001; and by the China Next Generation Internet under Grant CNGI-12-03-007. The Associate Editor for this paper was R. Liu.

Y. Li, D. Jin, and Z. Wang are with the State Key Laboratory on Microwave and Digital Communications, Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: liyong07@tsinghua.edu.cn; jindp@tsinghua.edu.cn; zcwang@tsinghua.edu.cn).

P. Hui is with the Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Kowloon, Hong Kong, with Telekom Innovation Laboratories, 10587 Berlin, Germany, and also with Aalto University, 02150 Espoo, Finland (e-mail: panhui@cse.ust.hk).

S. Chen is with the School of 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/TITS.2014.2325395

of the resource of networks of roads and transportation systems, becomes a hot research topic that attracts broad interests [2]. On the other hand, newly emerged vehicular communication networks are seen as a key technology to help in relieving the traffic congestion and improving road safety, by building intelligent transportation systems [3]–[5]. Thus, conscious efforts in developing intelligent transportation systems to deal with the transportation problems of urban cities have simultaneously come from both research fronts on the networks of vehicles and the vehicular communication networks. Critical issue for transportation systems, which are networks formed by vehicles, is how to handle the vehicular dynamics in terms of mobility within the capacity of the existing road system by predicting and guiding the vehicular traffics [1], [9]. On the other hand, the capability of predicting the vehicular mobility can play a significant role in various communication and networking functions from bandwidth reservation to service provisioning [4], [10]–[12].

From the preceding discussion, we have identified that vehicular mobility prediction is vital both to solving the problems of vehicular transportation systems and to building efficient vehicular communication networks. After the recent years of intensive research, a large number of mobility models are available, which can be classified into three different classes: synthetic models, which are obtained by mathematical modeling; survey models, which are obtained by extracting mobility patterns from surveys; and trace-based models, which are obtained by generating mobility patterns from real mobility traces [12]. In addition, many works [13], [14] have exploited the design of prediction algorithms based on the existing mobility models. Currently, further works are exploring the applications of mobility prediction to help with the urban traffic prediction, driving guidance, vehicular communications, etc. However, the vehicular dynamics in a transportation system or vehicular communication network are highly complex. For example, the movement schedules of individual vehicles are typically unknown to the “network,” and vehicular trajectories can easily appear random and unpredictable. In particular, how much the mobility can be predicted is an open and unsolved problem. Therefore, there are more fundamental questions to be addressed before designing prediction algorithm and utilizing prediction, which include what is the role of the randomness playing in the vehicular mobility, is there any regularity in the daily vehicular movement, and to what degree is the mobility predictable.

Against this background, in this paper, we consider the fundamental problem of the predictability limit for the vehicular

mobility in large-scale cities. Specifically, we explore the use of areas transition to model the vehicular mobility among the areas divided by the intersections of city roads. In this mobility model, vehicles stay in an area for some time and then move to another area by making the next location choice, which is the most intuitive description of mobility. The questions we addressed include that are there regularities governing the vehicular mobility, in terms of location transition between areas and the staying time in each area, and how can these regularities influence the predictability of vehicular mobility. Most importantly, we answer the fundamental question of what are the limits of predictability existing in the seemingly random vehicular mobility. Our contributions are summarized as follows.

- 1) We collect a large vehicular trace, i.e., the *Beijing* trace, which includes more than 20 000 vehicles for the duration of one month to study the mobility of vehicles. This is the largest urban city vehicular data trace available.
- 2) We propose an intuitive but effective model of areas transition to describe the vehicular mobility among the areas divided by the key intersections of city. Based on this mobility model, we utilize the method for predictability limits based on entropy theory to quantify how much the vehicular mobility can be predicted, in terms of location choice and staying time.
- 3) We demonstrate our proposed predictability limits of vehicular mobility by experimentally verifying that about 78%–99% of the location above and 70% of the staying time, respectively, can be predicted based on the Beijing trace and another existing large urban vehicular trace, i.e., the *Shanghai* trace [15]. Thus, our real-mobility-trace-based results reveal that there exists a strong regularity in the daily vehicular mobility, in terms of both temporal and spatial dimensions, which can potentially be exploited in prediction algorithm design.

The rest of this paper is organized as follows. After introducing the related work in Section II, we present the two vehicular data traces in Section III, which are used to illustrate and experimentally verify our proposed methodology. Section IV provides our mobility model. With this mobility model, Section V presents the method for predictability limits based on entropy theory. In Sections VI and VII, we present our experimental results for the predictability limits of the location and staying time, respectively, based on the Beijing and Shanghai traces. We conclude this paper in Section VIII.

II. RELATED WORK

As previously mentioned, the existing mobility models can be classified into three different classes: synthetic models, survey models, and trace-based models [12]. The synthetic models [16], [17] and the survey-based models [18], [19] are often unable to provide realistic modeling of motion patterns, although they can be very complex. The trace-based approach attempts to extract mobility models from real mobility traces by approximating the movements based on the observed movement patterns [20], [21]. All the aforementioned mobility models are often too complex to be concisely described by mathematical equations. In contrast to these existing models, we use an intu-

itive but effective model of areas transition to describe the vehicular mobility for urban city. Furthermore, instead of modeling the individual mobility, as most of the existing works do, our work focuses on the fundamental problem of the predictability limits for vehicular mobility and experimentally verifies the obtained predictability limits based on real mobility trace data.

The related work in vehicular mobility prediction comes from the two main communities, i.e., the transportation system community [1], [2], [9], [22] and the vehicular networking community [13], [14]. From the first community, some works investigate the vehicular behavior prediction in terms of trajectories and routes [1], [6]–[9], [22]. For examples, Wu *et al.* [9] applied the support vector regression to predict the travel time of vehicles, and Stathopoulos and Karlaftis [1] concentrated on developing the multivariate time-series state-space method to predict the urban traffic for solving the traffic congestion problem, whereas Kindzerske and Ni [22] applied the nonparametric regression approach to predict traffic conditions. From the second community, Zhang *et al.* [13] proposed an instant traffic clustering algorithm to partition the road points into the time-variant clusters by deriving the application-specific message update rules for affinity propagation to aid vehicular network design, whereas Namboodiri and Gao [14] explored the predictability of the location of mobility and use the prediction to aid the routing algorithm design for vehicular ad hoc networks. All these works focus on predication algorithm design and apply the prediction for the transportation system and vehicular network design. In contrast to these studies that investigate specific prediction algorithms, our work considers the fundamental problem of what is the limit of predictability for vehicular mobility in large-scale urban cities.

Rather than studying vehicular mobility prediction, some interesting studies investigate human mobility and prediction. For example, [23] examines the predictability limit of human mobility, whereas [11] explores the temporal predictability of human mobility and uses the prediction to assist the bandwidth reservation in advance.

III. DATA TRACES AND PREPROCESSING

Since we use two large-scale urban vehicular traces, i.e., the Shanghai and Beijing traces, to study the vehicular mobility predictability and to verify our proposed predictability limits, we first provide a brief description of these two traces.

A. Data Sets

The Shanghai trace [15] was collected by SG project [24], in which 2019 operational taxis continuously covered the whole month of February 2007 in Shanghai city. In this trace, a taxi sends its position report by general packet radio service (GPRS) to the central database every 1 min when it has passengers on board but every 15 s 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 taxis are not vacant most of the time. Another potential drawback of this trace is that the number of taxis is limited, as 2000 taxis and 1 min duration may not be sufficient to record

the statistical features of mobility in a high-speed large urban environment.

In collecting the Beijing trace, we used the mobility track logs obtained from 27 000 participating Beijing taxis carrying GPS receivers during May 2010. The reason for us to also choose taxis, instead of other 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 space and operation time than buses and private cars. Specifically, we utilized the GPS devices to collect the taxis' locations and time stamps and GPRS modules to report the records every 15 s for moving taxis. The specific information contained in such a report includes the following: the taxi's ID, the longitude and latitude coordinates of the taxi's location, time stamps, instant speed, and heading.

B. Data Preprocessing

By collecting the GPS information of longitude and latitude coordinates, we obtain the taxis' moving traces that indicate the taxis' locations varying with the time. Since these locations are measured by GPS devices, the collected data are noisy due to the inaccuracy of the GPS device. Furthermore, the taxis may not all report their locations at the same time slot with the same fixed frequency, as in the case of the Shanghai trace. Therefore, it is necessary to process the data trace to obtain the accurate locations of all the taxis in the same time slots and with the same frequency. In order to achieve these goals, we first use the city maps of Shanghai and Beijing for the respective traces to correct the taxis' locations so that they are on the related city roads. Then, we delete the sequent records of GPS positions that do not change during time interval of 2 min to cleanse the long stopping of vehicles due to reasons of waiting for passengers, parking, etc. Finally, we use the method of linear interpolation (LI) to insert location points so that all the taxis have location information at every 15-s interval. To illustrate how this LI method works, consider the location information of one taxi in the original trace with the locations l_1, l_2, \dots, l_n recorded at the time points $t_1 < t_2 < \dots < t_n$, and we want to insert the location information l_t at the time point t , which is calculated according to the 15-s frequency. We need to find t_m that satisfies $t_m \leq t < t_{m+1}$ and then to estimate the location l_t by the following LI:

$$l_t = \frac{t_{m+1} - t}{t_{m+1} - t_m} \cdot l_m + \frac{t - t_m}{t_{m+1} - t_m} \cdot l_{m+1}.$$

In order to verify that the preceding data preprocessing approach does not introduce artificial and inaccurate information into the original data trace, we use the data obtained by this preprocessing method for the one-day taxi locations to plot the trajectories of all the taxis, and the results show that the data sets are sufficiently large, and even using one-day data can recover the whole city maps. By comparing the recovered maps of Beijing and Shanghai with the true Beijing and Shanghai maps, we find that all the taxis' trajectories determined by the preprocessing are on the related city roads, and the two city maps drawn by these one-day trajectories are very similar to the corresponding true city maps.

IV. MOBILITY PREDICTION MODEL

A. Motivation

Two of the most important properties that characterize the mobility are the temporal and spatial parameters. The spatial property deals with the positions of vehicles, such as in which region and which road that a vehicle is traveling, as well as its moving direction, etc. When vehicles are traveling on the city roads, they usually transfer from one road to another at almost stationary velocity when the roads are free. Therefore, the transitions from one road to another are important observations of the vehicular mobility. The temporal property describes the vehicular staying time in each region, which depends on the vehicular moving speed and the traffic conditions. An accurate vehicular mobility model should have the ability to faithfully describe these basic temporal and spatial properties.

Consider a vehicle moving in the roads of a city. It will travel along a road and comes across an intersection. It may wait at the traffic light for some time and choose the direction at the intersection and then travels to another road to drive on. In the downtown of a large city, the roads are usually very crowded, and the intersections are very dense, which lead to very long waiting time at intersections and relatively short driving time along roads. Therefore, intersection is an important factor in modeling the urban vehicular mobility. Imagine viewing from the sky above the city; we observe a crowd of vehicles waiting at the area of each intersection and streams of traffic moving from one area to another area. Thus, in order to describe the vehicular distribution, we should pay particular attention to the areas around intersections.

By dividing the whole city into different areas, each including at least one intersection, we can model the vehicles moving from one area to an adjacent area and therefore model the vehicular traffic transiting from one area to another. Then, the mobility prediction is about how and when the vehicles will transit among these areas/locations. The spatial parameter describes the next area that each vehicle will transit to based on its historical transition information, whereas the temporal property characterizes how long each vehicle will stay in the current area and how long it will stay after it moves into the next area. Although vehicular mobility can be described by a variety of different models [12], this simple approach appropriately characterizes its temporal and spatial parameters, which enables the analysis of predictability limits for large-scale urban vehicular mobility.

B. Area Partition

In order to divide the vehicular mobility system into the areas that vehicles transit between, we need to take the roads and city structure into consideration. As intersections are the most important factor in modeling the urban vehicular mobility and distributions, we divide the system according to the positions of the key intersections in the city roads. More specifically, we use the key intersections as the centers of the partitioned areas and employ the Voronoi diagram, which is a frequently used method of decomposing a given space [25], to achieve the actual partitioning. In this approaches, the Voronoi diagram

facilitates dividing the city by the positions of intersections, which cannot be achieved by other simple model of grid shape meshes, for instance.

In a Voronoi diagram, a finite set of sites $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$ is given in the Euclidean plane. The Voronoi cell A_n corresponding to the site v_n consists of all the points whose distances to v_n are not greater than their distances to any other site v_m , $m \neq n$. In using the Voronoi diagram to partition the system region, the set of all the key intersections is \mathcal{V} . Each site $v_n \in \mathcal{V}$ is a selected intersection that is the center of the site or area. If we denote all the points in the system region as the set \mathbb{X} , the distance between the point $x \in \mathbb{X}$ and the site $v_n \in \mathcal{V}$ is defined as $d_{x,v_n} = \|x - v_n\|$. The area A_n for site v_n can be then formally expressed as

$$A_n = \{x \in \mathbb{X} | d_{x,v_n} \leq d_{x,v_m}, \forall v_m \in \mathcal{V} \setminus \{v_n\}\}. \quad (1)$$

Based on this Voronoi diagram, we obtain all the boundaries that partition the system region into the N different areas. After the region has been divided into the different areas, we can then decide which area a vehicle belongs to according to its longitude and latitude information. Consequently, the vehicular mobility is described by the transitions of vehicles from one area to another.

Let us consider the 2-D vehicular mobility defined by a sequence of steps that a vehicle travels in the city, which is described or partitioned as the areas around the intersections, as previously discussed. Step n is defined by the tuple (t_n, A_n, t_n^d) , which records that the vehicle enters area A_n at time t_n , and it stays in A_n for the duration of t_n^d . Every vehicle moves in step by step by transiting from one area to another area, which depicts each vehicle's mobility. Thus, the traffic flows of the whole system can be described by combining all the vehicles' mobility patterns and all the intersections together as a system.

From the historical information, we can obtain the sequence of the areas/locations that each vehicle has traveled, which is denoted by $\mathcal{A}_{n-1} = \{A_1, A_2, \dots, A_{n-1}\}$. Then, the task of location prediction becomes finding the most possible next area A_n that the vehicle will travel to according to the historical transitions \mathcal{A}_{n-1} . Furthermore, with the continuous record of the areas traveled by the vehicles, we can calculate the degree of location predictability by analyzing the location sequences. For the staying time in each area, we can also obtain the history distribution of $\mathcal{S}_{n-1} = \{t_1^d, t_2^d, \dots, t_{n-1}^d\}$. Then, the staying-time prediction problem becomes deciding t_n^d given \mathcal{S}_{n-1} . Similarly, the degree of staying-time predictability can be calculated by analyzing all the staying-time sequences. From the preceding analysis, we can see that the proposed mobility prediction model is capable of analyzing the predictability based on the vehicular historical mobility information.

C. Basic Statistic Properties of Mobility

The area partition results or the number of areas and the sizes of the areas divided depend on the intersection selection. Therefore, we need to be able to select different numbers of intersections to partition the areas for different applications. For

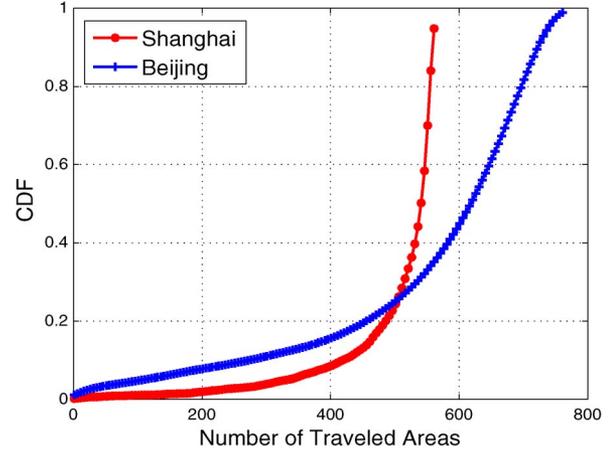


Fig. 1. Distributions of the aggregated number of traveled areas over all the vehicles.

this purpose, we count all the intersections for the two cities, which are 845 intersections in Beijing and 622 intersections in Shanghai. We then sort all the intersections of a city by the traffic volumes from the largest to the smallest. This allows us to select some desired percentage of the intersections for city area partitioning and mobility studying.

When we select 90% of the intersections (760 for Beijing and 559 for Shanghai) for our model, we obtain the statistics of the aggregated number of traveled areas over all the vehicles and depict its cumulative distribution function (cdf) in Fig. 1. From the results, we observe that 80% of the Shanghai vehicles travel about 500–560 areas, whereas 80% of the Beijing vehicles travel about 500–760 areas. This difference in the number of areas traveled reflects the underlying different scales of the two cities. In addition, the percentage of the Beijing taxis, which travel less than 500 areas, is larger than that of the Shanghai taxis, which travel less than 500 areas. This indicates that relatively more Beijing vehicles stay in small local areas. The area distributions thus demonstrate that most of the vehicles in both Beijing and Shanghai cover the areas of the whole city, and only a small number of taxis travel in some local areas.

To get a visual view of our area partition method, we select 50% of the intersections with largest traffic through sorting all the intersections by the traffic volumes from the largest to the smallest and plot the area partition results for Beijing and Shanghai traces in Figs. 2 and 3, respectively, where we divide the whole city, shown in the figures' left part, into the given number of areas, shown in the figures' right part. In these two area partitions, we mark the selected intersections as the red points and plot the boundaries of each area by the blue curves. From these results, we can see that the main underlying city structures of both Shanghai and Beijing are captured by selecting only 50% of the intersections.

To illustrate diverse behaviors of individual vehicles, we choose 10% of the intersections, which are 84 and 62 intersections for Beijing and Shanghai, respectively, to partition the two cities, and select two representative vehicles with very different mobility patterns from each trace for observation. Specifically, we plot the areas visited by the vehicle as nodes and use the size of a node to indicate the percentage of the time that the

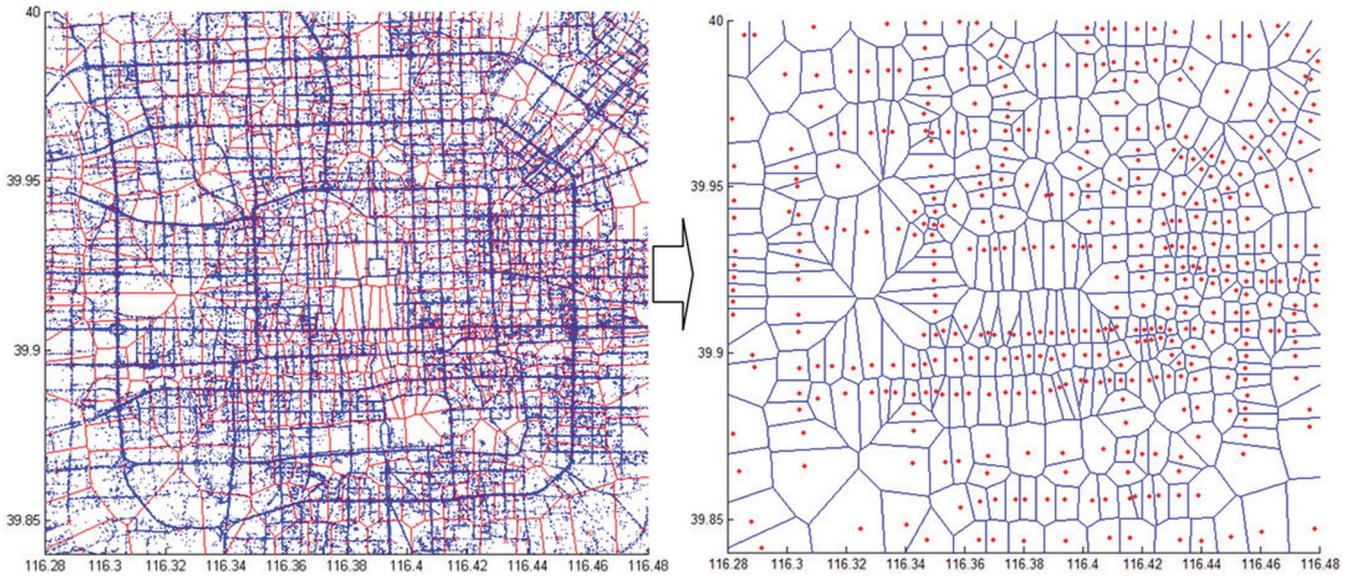


Fig. 2. Illustration of area partition for Beijing city by selecting 50% of the intersections.

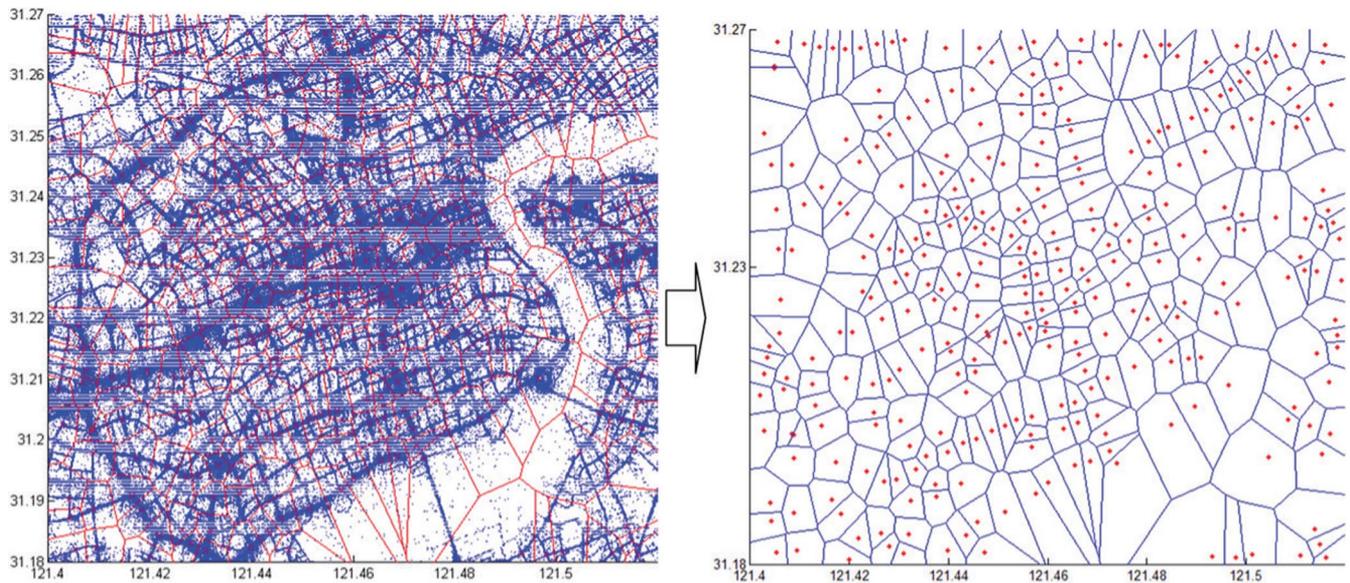


Fig. 3. Illustration of area partition for Shanghai city by selecting 50% of the intersections.

vehicle spends in the area, whereas we use the links between the nodes to represent the observed movements between these areas. The results obtained for Beijing and Shanghai traces are shown in Figs. 4 and 5, respectively. For the Beijing trace, the first selected vehicle moves in the vicinity of ten areas, whereas the second visits as many as 72 areas. For the Shanghai trace, one taxi selected travels in about 12 areas, and the other covers all the areas. These results clearly indicate the very diverse vehicular mobility patterns. The results also show that individual taxis tend to spend most of their staying times in a few particular areas.

V. METHODOLOGY FOR PREDICTABILITY

We use the entropy that quantifies the uncertainty to obtain the fundamental limits of vehicular predictability.

A. Entropy Theory for Predictability

The basic concept of entropy is originally defined in the context of thermodynamics. It measures the degree of randomness in a set of configurations [26]. For a discrete random variable X that takes the value from the set $\{x_1, x_2, \dots, x_N\}$, the entropy is defined as

$$H(X) = - \sum_{i=1}^N p(x_i) \log_2(p(x_i))$$

where $p(x_i)$ denotes the probability of $X = x_i$. Entropy is measured in bits, and it indicates how predictable a variable is. Low entropy implies high degree of predictability.

Under the mobility model introduced in Section IV, both the prediction problems for location and staying time can be

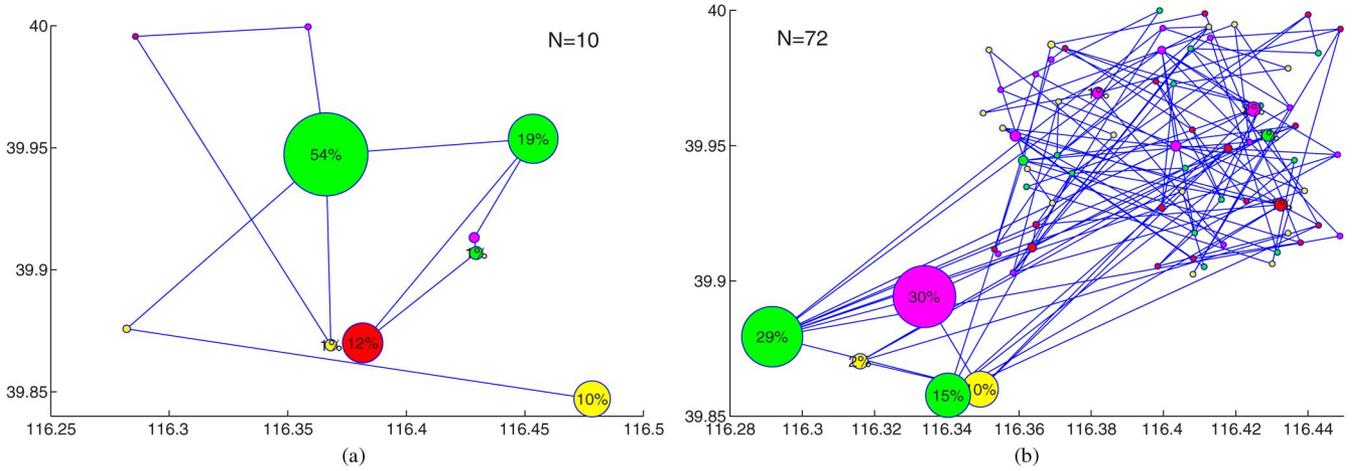


Fig. 4. Travel locations and staying times of the two selected representative vehicles from the Beijing trace. (a) Vehicle traveling a small number of locations. (b) Vehicle traveling a large number of locations.

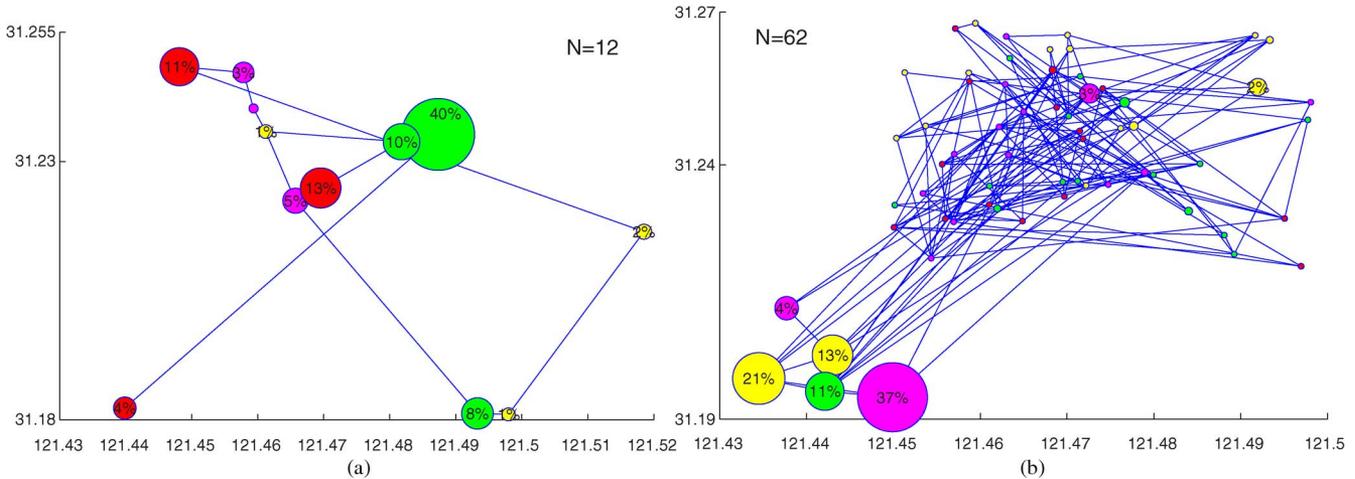


Fig. 5. Travel locations and staying times of the two selected representative vehicles from the Shanghai trace. (a) Vehicle traveling a small number of locations. (b) Vehicle traveling a large number of locations.

abstracted as the following unified problem. The prediction problem is underpinned by the stochastic process $\mathbf{X} = \{X_1, X_2, \dots, X_n, \dots\}$, and the task at the n th step of prediction is to predict the next symbol X_n given the historical information $\mathcal{X}_{n-1} = \{X_1^o, X_2^o, \dots, X_{n-1}^o\}$, where \mathcal{X}_{n-1} denotes the past observation of the sequence $\mathbf{X}_{n-1} = \{X_1, X_2, \dots, X_{n-1}\}$. More specifically, for location prediction, $\mathcal{X}_{n-1} = \mathcal{A}_{n-1}$, and each X_i is a discrete random variable that takes the value from the set of N values. Similarly, for staying time prediction, $\mathcal{X}_{n-1} = \mathcal{S}_{n-1}$, but each X_i is a continuous-valued random variable. For the purpose of a unified treatment of both the location and staying time predictions, we assume that the continuous-valued staying time has been quantized into the set of N values. Therefore, each X_i in the case of staying time prediction is turned into a discrete N -valued random variable. Here, N is used for notational purpose, and it does not imply that the size of the area set is equal to the size of the discrete staying time set.

For the process $\mathbf{X} = \{X_1, X_2, \dots, X_n, \dots\}$, where each X_i is a discrete N -valued random variable, the entropy rate measures the uncertainty that remains in the next information

symbol produced by the process given the complete knowledge of the past. It is a natural measure of the predictability in predicting the evolution of the process [27]. For a stationary process \mathbf{X} , its entropy rate can be written as

$$H \equiv \lim_{n \rightarrow \infty} \frac{1}{n} H(\mathbf{X}_n) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n H(X_n | \mathbf{X}_{n-1}) \quad (2)$$

where $H(X_n | \mathbf{X}_{n-1})$ is the conditional entropy at the n th prediction step, and it can be obtained by the chain rule for entropy. The entropy rate (2) is defined under the generic condition that the process has memory or the series $\{X_1, X_2, \dots, X_n, \dots\}$ has temporal correlation, so that the probability of the next symbol depends on the current and past ones. Otherwise, the entropy rate of an uncorrelated series is simply defined as

$$H^u = H(X_n) = - \sum_{i=1}^N p_i \log_2(p_i) \quad (3)$$

where p_i denotes the probability of X_n taking the i th value in the set of N values.

To calculate the entropy rate from the vehicular historical information series \mathcal{X}_n , we use an estimation based on Lempel–Ziv data compression [28], which is proved to rapidly converge to the true entropy of the series. For the series \mathbf{X}_n with size n , the entropy rate is estimated as

$$H^e = \left(\frac{1}{n} \sum_{i=1}^n \Lambda_i \right)^{-1} \ln(n) \quad (4)$$

where Λ_i is the length of the shortest substrings starting at position X_i that does not appear as a contiguous substring of the previous $i - 1$ symbols X_1, \dots, X_{i-1} . It has been proven that H^e converges to the true entropy of \mathbf{X} when n approaches infinity [28].

Thus, we have the following three cases of entropy [23].

- 1) The random entropy $H^r = \log_2(N)$ for the purely random sequence \mathbf{X} , whose element X_n takes the i th value with equal probability of $p_i = 1/N$ for $1 \leq i \leq N$. It measures how random the vehicles are traveling when the location visiting probability and temporally correlated series are not considered, which directly indicates how many locations the vehicle has traveled in the modeled areas.
- 2) The uncorrelated entropy $H^u = -\sum_{i=1}^N p_i \log_2(p_i)$ for the temporally uncorrelated series \mathbf{X} .
- 3) The entropy H for the generic temporally correlated series \mathbf{X} , whose estimate is H^e given in (4).

B. Fundamental Limits of Predictability

If the sequence \mathbf{X} of one vehicle has the entropy rate of $H = 0$, then the vehicular mobility location or staying time is completely regular and its trajectory is fully predictable. On the other hand, if it has the entropy rate of $H = H^r = \log_2(N)$, which is the maximum value of entropy, then its mobility follows a purely random pattern and we cannot predict its next move with the accuracy exceeding $1/N$. However, for most of the vehicles, their mobility patterns are governed by a certain amount of randomness and some degree of regularity in their movements or staying times, which can be exploited for prediction. In other words, their entropy rate values H lie between 0 and H^r .

Related to the limit of predictability for the location or staying time given the historical series, we obtain from [23] that the upper bound of the predictability limit, i.e., Ψ^{\max} , for the series \mathbf{X} can be calculated from

$$H = -(\Psi^{\max} \log_2(\Psi^{\max}) + (1 - \Psi^{\max}) \log_2(1 - \Psi^{\max})) + (1 - \Psi^{\max}) \log_2(N - 1) \quad (5)$$

where H is the entropy rate of \mathbf{X} .

VI. LOCATION PREDICTION LIMITS

To analyze the predictability of travelling locations across the vehicles based on the vehicular mobility, we combine the entropy and maximal predictability limit. Specifically, we first determine the entropy H and the predictability limit Ψ of each

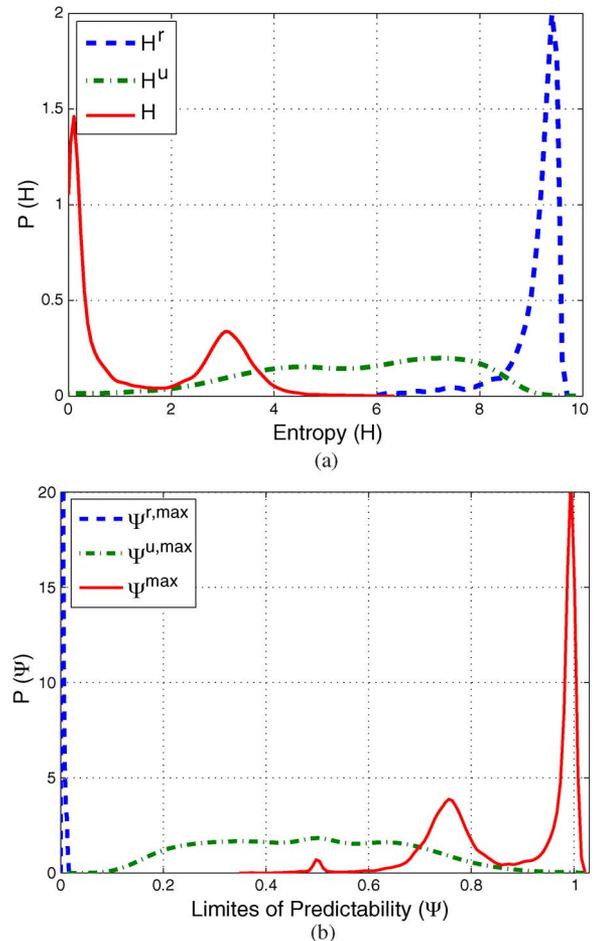


Fig. 6. Location prediction properties of the Beijing trace. (a) Distribution of entropy. (b) Distribution of maximal predictability limit.

vehicle according to its travelling area records during the whole trace collection time, as well as its maximal predictability limit Ψ^{\max} according to (5). Then, we obtain the distributions $p(H)$ and $p(\Psi^{\max})$ over all the vehicles. We are also interested to know what would be the statistics if the individual vehicular sequence of the area record were to be a temporal uncorrelated sequence or a purely random equiprobable sequence. Thus, we also calculate H^u and H^r and the corresponding maximum predictability limits $\Psi^{u,\max}$ and $\Psi^{r,\max}$ under these two assumptions, respectively. This allows us to obtain the distributions of $p(H^u)$ and $p(H^r)$ and $p(\Psi^{u,\max})$ and $p(\Psi^{r,\max})$ for comparison purpose.

A. Results of Predictability Limits

The distributions of the entropy and the maximal predictability limit so obtained for Beijing and Shanghai traces are shown in Figs. 6 and 7, respectively, where Figs. 6(a) and 7(a) depict the entropy distributions $p(H^r)$, $p(H^u)$ and $p(H)$; whereas Figs. 6(b) and 7(b) show the distributions of maximal predictability limits $p(\Psi^{r,\max})$, $p(\Psi^{u,\max})$, and $p(\Psi^{\max})$.

We observe that the distributions $p(H^r)$ peak at $H^r = 9.5$ and $H^r = 9.1$ for Beijing and Shanghai traces, respectively. This implies that each update of the vehicles' distribution would represent 9.5 and 9.1 bits of the new information on

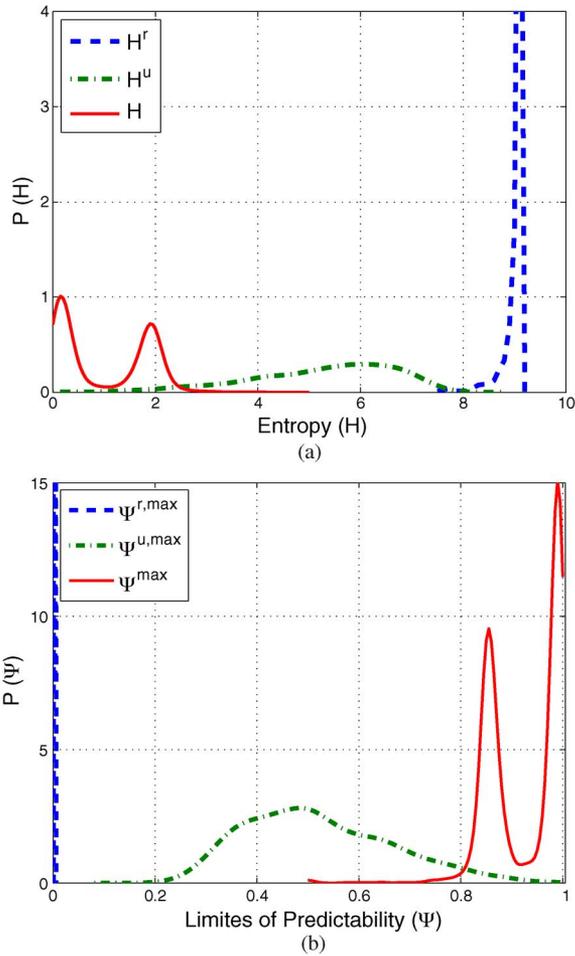


Fig. 7. Location prediction properties of the Shanghai trace. (a) Distribution of entropy. (b) Distribution of maximal predictability limit.

average for the two cities, respectively. In other words, a vehicle would choose its next area randomly from the $\hat{N} = 2^{9.5} \approx 724$ and $\hat{N} = 2^{9.1} \approx 549$ next locations on average in the two cities, respectively. Recall that the numbers of the total areas partitioned for Beijing and Shanghai are $N \approx 760$ and $N \approx 559$, respectively, which agree with the values provided by the two distributions of random entropy H^r . In contrast, the true entropy distributions $p(H)$ for both Shanghai and Beijing traces peak at two much smaller entropy values. Specifically, for the Beijing trace, $p(H)$ peaks at $H = 0.2$ and 3.1 , whereas for the Shanghai trace, $p(H)$ peaks at $H = 0.2$ and 1.9 . It is seen that the vehicles in both cities can roughly be divided into the two groups according to entropy distribution: one group's entropy is approximately around the high peak, and the other group's entropy is approximately around the low peak. The vehicles in the first group have low uncertainty of $H = 0.2$ in their location trajectories, that is, these vehicles will choose their next locations from the $2^{0.2} = 1.15 \approx 1$ alternative areas on average, which is an almost certain decision. The vehicles in the high peak group will choose their next locations from the alternative $2^{1.9} \approx 4$ and $2^{3.1} \approx 9$ locations on average for Shanghai and Beijing, respectively. Note that these are much smaller than \hat{N} suggested by the distributions of random entropy. Similarly, using the distribution of $p(H^u)$ also gives the misleadingly higher uncertainty level in the vehicular location patterns.

The limit in the probability that any algorithm can correctly predict the vehicle's next location is the predictability limit Ψ , and the upper bound of Ψ is given by Ψ^{max} . From the distributions of Ψ^{max} shown in Figs. 6(b) and 7(b) for Beijing and Shanghai traces, respectively, we can see that $p(\Psi^{max})$ is narrowly distributed with the two peaks for both Shanghai and Beijing. Specifically, for the Beijing trace, the two peaks occur at $\Psi^{max} = 0.76$ and $\Psi^{max} = 0.99$, whereas for the Shanghai trace, they are at $\Psi^{max} = 0.85$ and $\Psi^{max} = 0.99$. Obviously, these two peaks of $p(\Psi^{max})$ correspond to the two peaks of $p(H)$. The high peak of $p(\Psi^{max})$ indicates that the mobility of this group of vehicles can potentially be correctly predicted with probability almost equal to 1, whereas the low peak of $p(\Psi^{max})$ indicates that the mobility of this group of vehicles can be correctly predicted with probability approximately equal to 0.76 and 0.85 for Beijing and Shanghai, respectively. It is seen that the predictability of the Shanghai trace is slightly higher than that of the Beijing trace. These results confirm that, despite the apparent randomness of the individuals' trajectories, the historical record of the daily vehicular mobility patterns contains a surprisingly high degree of potential predictability. This is a far cry from the almost complete unpredictability suggested by the distributions of $p(\Psi^{r,max})$, which peak at $\Psi^{r,max} \approx 0$ for both Shanghai and Beijing traces. The distribution $p(\Psi^{r,max})$ simply shows that the future location becomes almost completely unpredictable if only the number of the areas visited is used to predict the next location. Observe the distributions $p(\Psi^{u,max})$; they are much more widely distributed with the inconspicuous peak at $\Psi^{u,max} \approx 0.5$ for both Beijing and Shanghai traces. Thus, if an algorithm only relies on the different visiting probabilities for individual areas in predicting and does not exploit the inherent temporal correlation in the mobility patterns, it cannot achieve the full potential prediction power.

To further analyze the results of maximal predictability limit shown in Figs. 6 and 7, we choose two typical vehicles that represent the low peak and high peak groups of $p(\Psi^{max})$, respectively, for the Beijing trace. We vary the number of the areas for partitioning Beijing city by changing the percentage of the selected intersections from 5% to 100% and obtain the maximal predictability limits Ψ^{max} of these two vehicles as the function of the percentage of the selected intersections. The results are given in Fig. 8(a) and (b) for the high-predictability group and low-predictability group representatives, respectively. We observe that the predictability limit increases with the increase in the intersections or areas used. The larger the number of areas used for partitioning, the smaller the individual areas become, and more details related to the mobility regularity appear, which increases the predictability. With more than 40% of the total intersections, the predictability limit of the high-predictability group is higher than 99% and becomes almost constant. For the low-predictability group, when using more than 60% of the intersections, the predictability probability is above 80% to 82%.

Next, we divide all the vehicles into the two groups of high predictability and low predictability, respectively, for both Shanghai and Beijing traces, and analyze the average location predictability limit of each group. The results of the average

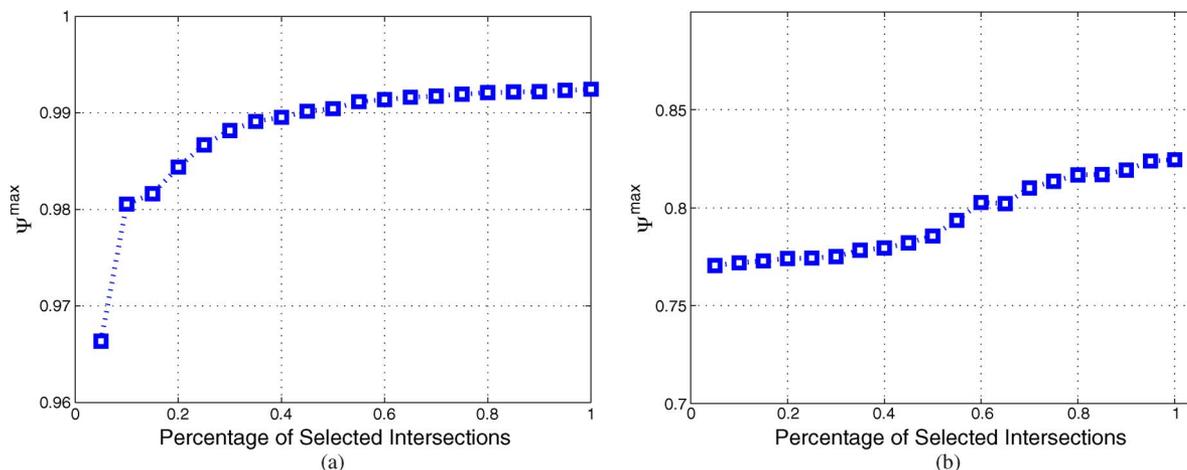


Fig. 8. Maximal predictability limits of the low- and high-predictability vehicle groups in the Beijing trace. (a) High-predictability group representative. (b) Low-predictability group representative.

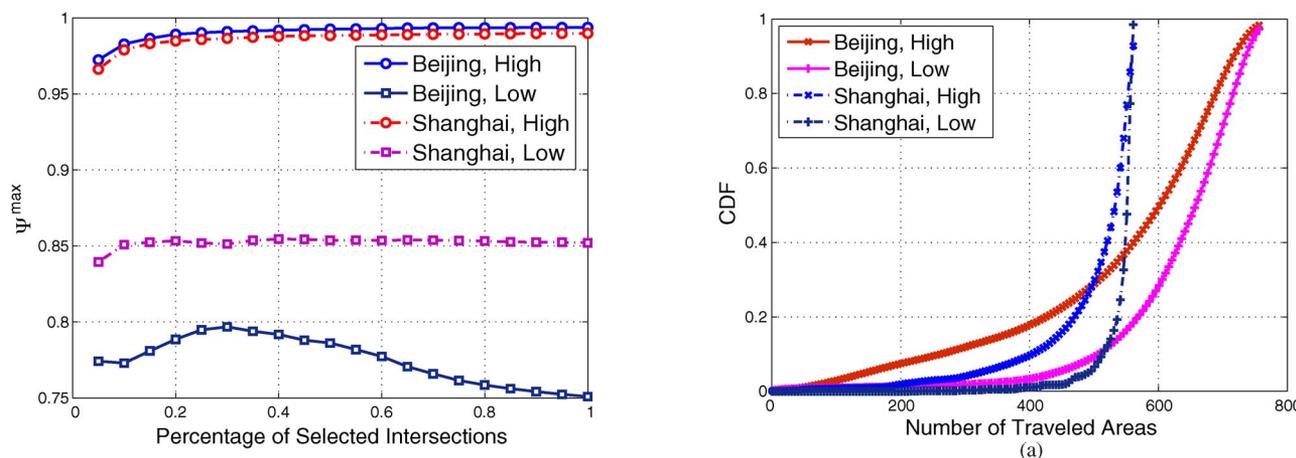


Fig. 9. Limits of average location predictability for Beijing and Shanghai traces.

maximal predictability limit of each group as the function of the percentage of the selected intersections are depicted in Fig. 9. We observe that the average location predictability limits of the high-predictability groups for Beijing and Shanghai are almost the same: they both reach around 99% when selecting more than 20% of the intersections for partition, and the limits remain very stable. For the two low-predictability groups, the average maximal predictability limit of Beijing is around 78%, which is lower than that of Shanghai. In addition, the average location predictability limit for Shanghai’s low-predictability group is more stable than that for Beijing’s low-predictability group. In summary, the location predictability results obtained reveal that the taxis in both Shanghai and Beijing can be divided into the two groups with the predictability limits of about 99% and 80%, respectively.

B. Predictability Validation

The results of location predictability reveal that there is a high degree of regularity embedded in vehicular daily travelling mobility. In order to analyze the variations of predictability limits in the high- and low-predictability groups of Beijing and Shanghai traces, the cdfs of the numbers of areas traveled are

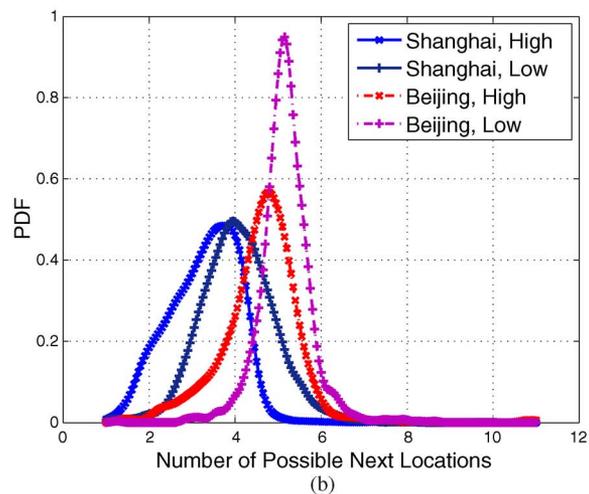


Fig. 10. (a) cdf of the number of areas traveled and (b) pdf of the number of possible next locations, for the low- and high-predictability groups in Beijing and Shanghai traces.

plotted in Fig. 10(a), whereas the probability density functions (pdfs) of the numbers of next possible locations are depicted in Fig. 10(b), for both groups of the two cities. The reason to use the cdf, instead of the pdf, in Fig. 10(a) is for better visualization. In terms of the number of traveled locations, as expected, the vehicles in the high-predictability group of a city travel

less locations than those of the corresponding low-predictability group. From the distributions in Fig. 10(a), we observe that, generally, Shanghai’s taxis travel less locations than Beijing’s taxis. In terms of the number of next possible locations, as expected, the pdf’s peak of the high-predictability group of a city is in the left of the pdf’s peak of the corresponding low-predictability group. From the pdfs shown in Fig. 10(b), we can see that, in general, Shanghai’s taxis are more predictable than Beijing’s taxis in choosing next locations. The pdf width of Beijing’s low-predictability group is narrower than the pdf width of the high-predictability group, indicating that the variation in the individual vehicles’ predictability limits of the high-predictability group is larger. In contrast, the pdf widths of the two Shanghai groups are approximately the same, which suggest that the variations in the individual vehicles’ predictability limits of the two groups are approximately the same.

VII. STAYING TIME PREDICTION LIMITS

Different from the location prediction that mainly depends on the regularity of individual vehicle, the staying time of a vehicle in one area depends on the traffic status of the area. Therefore, to predict the possible staying time for a particular vehicle that enters the area, the historical information of all the other vehicles’ staying times in this area plays a much more important role than the historical information of this vehicle’s staying time in this area. In other words, the sequence of an individual vehicle’s staying time in one area can be regarded as an uncorrelated sequence. In addition, as previously mentioned, the staying time is a continuous value, and we need first to quantize it with certain quantization precision (QP).

A. Initial Investigation of Prediction Limits

For each area, we want to obtain the distribution of the staying times of all the vehicles that pass through the area. We first need to select a QP to quantize the continuous staying time into a discrete-valued set. Since each series of single vehicle’s staying time is uncorrelated, we can use the uncorrelated entropy H^u to measure the uncertainty of individual series, which is denoted by H_t^u here for emphasizing that we are dealing with the staying time. The corresponding maximal predictability limit for staying time obtained using the method described in Section V is denoted by $\Psi_t^{u,max}$. For the Beijing trace, with the QP of 10, 30, and 50 s, respectively, we obtain the distributions $p(H_t^u)$ and $p(\Psi_t^{u,max})$ of each area and depict the average distributions of $p(H_t^u)$ and $p(\Psi_t^{u,max})$ over all the areas in Fig. 11(a) and (b), respectively. From the results in Fig. 11, we observe that, given the QP of 10 s, the entropy H_t^u of the staying time is in the range of [6, 9], and the predictability limit is distributed in the range of [0.1, 0.2]. In addition, the chosen QP influences the distributions of entropy and predictability limit. The higher the quantization accuracy (the smaller the QP value), the higher the entropy and, hence, the lower the predictability limit. However, even with QP = 50 s, the entropy distribution peaks at $H_t^u = 6.8$, and the maximal predictability limit peaks at $\Psi_t^{u,max} = 0.16$, which means that, if we want to predict how many 50 s the vehicle will stay in the area, we have

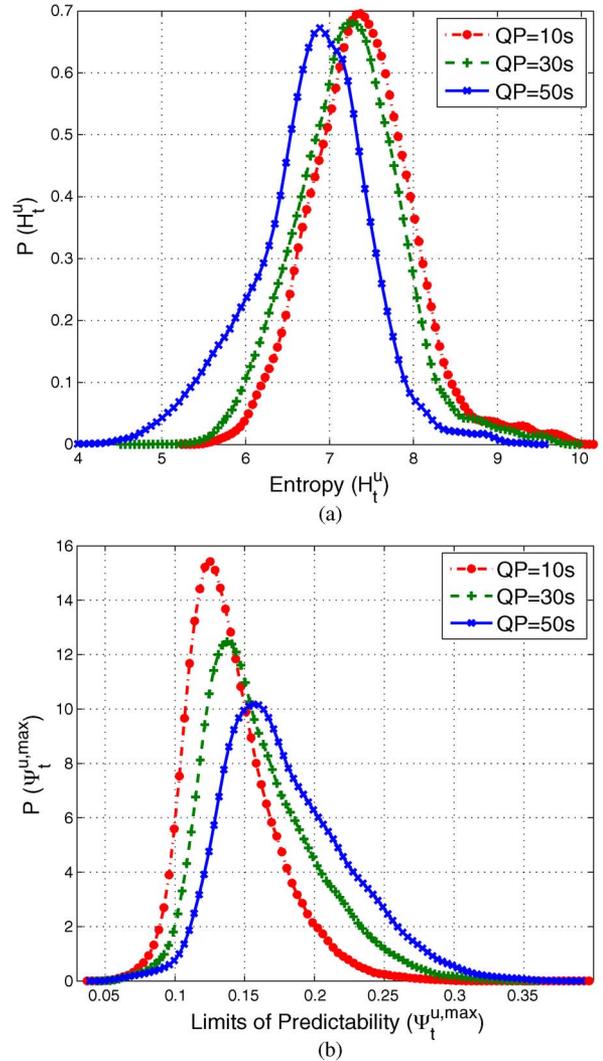


Fig. 11. Staying time prediction properties of the Beijing trace. (a) Distribution of entropy. (b) Distribution of maximal predictability limit.

at most 16% of the probability to get it right. This indicates that it is difficult to forecast the individual vehicle’s staying time by its historical information in spatial dimension alone. A practical prediction algorithm must exploit other vehicular properties for a better prediction of the individual vehicle’s staying time.

B. Limits of Predictability Exploiting Temporal Regularity

Based on the preceding investigation and analysis, it is clear that relying on the spatial historical distribution to predict the future staying time in an area will result in poor predictability. It is also obvious that the historical distribution of stay time in an area significantly varies in the different time of a day. In order to utilize the regularity existed in the temporal dimension, we slot the whole day by every 2 h and obtain the staying time distribution for every 2 h of the whole day. In the prediction, we then utilize the corresponding distribution that locates in the same time slot when the vehicle enters into the area. Again, in order to quantify the predictability of the staying time with the aid of temporal regularity, we investigate the limits of predictability of these distributions.

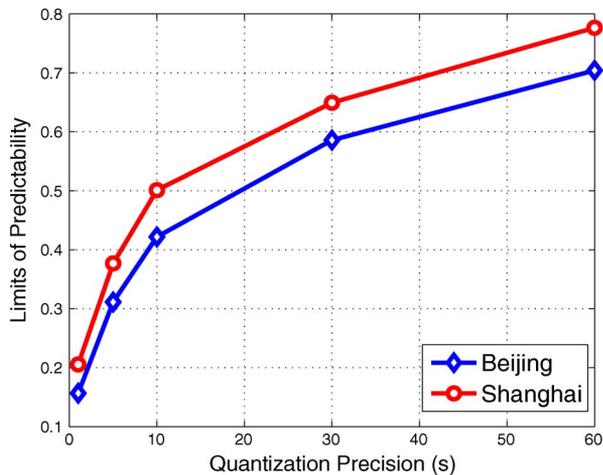


Fig. 12. Limit of staying-time predictability as a function of QP.

In order to obtain the limit of predictability for staying time as a function of QP, we vary the value of QP from 5 to 60 s, and for each QP value, we plot the peak point of the maximum predictability limit distribution. The results are shown in Fig. 12. We observe that when QP varies from 5 to 30 s, the predictability limit rapidly increases. With QP = 30 s, the predictability limits are around 59% and 65% for Beijing and Shanghai traces, respectively. However, when QP increases to 60 s, the maximum predictability limit $\Psi_t^{u,\max}$ only increases to 77% and 70% for Shanghai and Beijing traces, respectively. The results obtained clearly confirm that, by exploiting the regularity in the temporal dimension, the prediction accuracy associated with the QP of 30 s is sufficient for most practical applications of predicting the vehicular staying time. With the QP of 50–60 s, about 70% of the predictability limit can be achieved.

VIII. CONCLUSION

In this paper, we have extensively investigated the predictability limits of vehicular mobility based on the two realistic large-scale urban city vehicular mobility traces. The main results obtained show that there exists some stronger regularity in the daily vehicular mobility in both the temporal and spatial dimensions, which can be exploited to predict the vehicular mobility with a high degree of prediction accuracy. Specifically, for both Shanghai and Beijing traces, the location predictability limit of 80%–99% can be achieved, whereas above 70% of the staying-time predictability limit can be reached with appropriate QP by exploiting the temporal regularity.

The most important finding in this study is that the development of accurate predictive models and algorithms is possible and has solid scientific foundation, and this provides the fundamental guiding principle for solving large-scale problems and benefiting many potential applications, from urban transportation system planning, transportation safe and traffic control, to the vehicular network algorithm design and system deployment. Although using specific prediction algorithm to make explicit prediction based on vehicular historical mobility trajectory is beyond the scope of this paper, appropriate data-mining- or Kalman-filter-based algorithms could turn the predictability limit identified in this study into the actual mobility

prediction reality. Therefore, our future work will investigate whether the existing or any new prediction algorithms can indeed achieve the predictability limit founded in this study.

REFERENCES

- [1] A. Stathopoulos and M. G. Karlaftis, "A multivariate state space approach for urban traffic flow modeling and prediction," *Transp. Res. C, Emerging Technol.*, vol. 11, no. 2, pp. 121–135, Apr. 2003.
- [2] Y. Kamarianakis and P. Prastacos, "Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1857, pp. 74–84, Jan. 2003.
- [3] G. Dimitrakopoulos and P. Demestichas, "Intelligent transportation systems," *IEEE Veh. Technol. Mag.*, vol. 5, no. 1, pp. 77–84, Mar. 2010.
- [4] 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.
- [5] 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.
- [6] R. Horiguchi, M. Kuwahara, and H. Akahane, "The optimal arrangement of infrared beacons on a road network to collect vehicle trajectories—Pattern analysis using schema theory," *J. Eastern Asia Soc. Transp. Studies*, vol. 4, no. 4, pp. 219–228, Oct. 2001.
- [7] J. Liu, O. Wolfson, and H. Yin, "Extracting semantic location from outdoor positioning systems," in *Proc. 7th Int. Conf. Mobile Data Manag.*, Nara, Japan, 2006, pp. 73–80.
- [8] J. Froehlich and J. Krumm, "Route prediction from trip observations," *Soc. Autom. Eng.*, vol. 2193, p. 53, 2008.
- [9] C.-H. Wu, J.-M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 276–281, Dec. 2004.
- [10] F. Li and Y. Wang, "Routing in vehicular *ad hoc* networks: A survey," *IEEE Veh. Technol. Mag.*, vol. 2, no. 2, pp. 12–22, Jun. 2007.
- [11] L. Song, U. Deshpande, U. C. Kozat, D. Kotz, and R. Jain, "Predictability of WLAN mobility and its effects on bandwidth provisioning," in *Proc. 25th IEEE INFOCOM*, Barcelona, Spain, Apr. 23–29, 2006, pp. 1–13.
- [12] J. Harri, F. Filali, and C. Bonnet, "Mobility models for vehicular *ad hoc* networks: A survey and taxonomy," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 4, pp. 19–41, 2009.
- [13] B. Zhang, K. Xing, X. Cheng, L. Huang, and R. Bie, "Traffic clustering and online traffic prediction in vehicle networks: A social influence perspective," in *Proc. 31th IEEE INFOCOM*, Orlando, FL, USA, Mar. 25–30, 2012, pp. 495–503.
- [14] V. Nambodiri and L. Gao, "Prediction-based routing for vehicular *ad hoc* networks," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 2332–2345, Jul. 2007.
- [15] H. Zhu *et al.*, "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.
- [16] A. Rojas, P. Branch, and G. Armitage, "Experimental validation of the random waypoint mobility model through a real world mobility trace for large geographical areas," in *Proc. 8th ACM Int. Symp. Model., Anal. Simul. Wireless Mobile Syst.*, Montreal, QC, Canada, Oct. 10–13, 2005, pp. 174–177.
- [17] W.-J. Hsu, K. Merchant, H.-W. Shu, C.-H. Hsu, and A. Helmy, "Weighted waypoint mobility model and its impact on *ad hoc* networks," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 9, no. 1, pp. 59–63, Jan. 2005.
- [18] Q. Zheng, X. Hong, and J. Liu, "An agenda based mobility model21," in *Proc. 39th Annu. Symp. Simul.*, Huntsville, AL, USA, Apr. 2–6, 2006, pp. 188–195.
- [19] M. Musolesi and C. Mascolo, "A community based mobility model for *ad hoc* network research," in *Proc. 2nd Int. Workshop Multi-hop Ad Hoc Netw., Theory Reality*, Florence, Italy, May 26, 2006, pp. 31–38.
- [20] M. Kim, D. Kotz, and S. Kim, "Extracting a mobility model from real user traces," in *Proc. 25th IEEE INFOCOM*, Barcelona, Spain, Apr. 23–29, 2006, pp. 1–13.
- [21] J. Yoon, B. D. Noble, M. Liu, and M. Kim, "Building realistic mobility models from coarse-grained traces," in *Proc. 4th Int. Conf. Mobile Syst., Appl. Serv.*, Uppsala, Sweden, Jun. 19–22, 2006, pp. 177–190.
- [22] M. D. Kindzerske and D. Ni, "Composite nearest neighbor nonparametric regression to improve traffic prediction," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1993, pp. 30–35, Oct. 2007.

- [23] C. Song, Z. Qu, N. Blumm, and A. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, Feb. 2010.
- [24] M. Li, H. Zhu, Y. Zhu, and L. M. Ni, "Ants: Efficient vehicle locating based on ant search in ShanghaiGrid," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4088–4097, Oct. 2009.
- [25] K. Kise, A. Sato, and M. Iwata, "Segmentation of page images using the area Voronoi diagram," *Comput. Vis. Image Understanding*, vol. 70, no. 3, pp. 370–382, Jun. 1998.
- [26] E. Maasoumi and J. Racine, "Entropy and predictability of stock market returns," *J. Econom.*, vol. 107, no. 1/2, pp. 291–312, Mar. 2002.
- [27] N. Navet and S.-H. Chen, "On predictability and profitability: Would GP induced trading rules be sensitive to the observed entropy of time series?" in *Natural Computing in Computational Finance*, A. Brabazon and M. O'Neill, Eds. Berlin, Germany: Springer-Verlag, 2008, pp. 197–210.
- [28] A. D. Wyner and J. Ziv, "Fixed data base version of the Lempel–Ziv data compression algorithm," *IEEE Trans. Inf. Theory*, vol. 37, no. 3, pp. 878–880, May 1991.



Yong Li (M'09) received the B.S. degree in electronics and information engineering from Huazhong University of Science and Technology, Wuhan, China, in 2007 and the Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China, in 2012.

During July to August 2012 and 2013, he was a Visiting Research Associate with Telekom Innovation Laboratories and The Hong Kong University of Science and Technology, respectively. During December 2013 to March 2014, he was a Visiting

Scientist with the University of Miami, Coral Gables, FL, USA. He is currently a Faculty Member of the Department of Electronic Engineering, Tsinghua University. He has authored or coauthored over 100 research papers and is a holder of ten granted and pending Chinese and international patents. His Google Scholar Citation is about 460 with H-index of 11, as well as over 120 total citations without self-citations in Web of Science. His research interests are in the areas of networking and communications, including mobile opportunistic networks, device-to-device communication, software-defined networks, network virtualization, and future Internet.

Prof. Li has served as a Technical Program Committee (TPC) Chair for the WWW workshop of Simplex 2013 and the TPC for several international workshops and conferences. He is a Guest Editor of ACM/Springer Mobile Networks and Applications, Special Issue on Software-Defined and Virtualized Future Wireless Networks. He is currently the Associate Editor of *EURASIP Journal on Wireless Communications and Networking*. He was a recipient of the Outstanding Postdoctoral Researcher, Outstanding Ph.D. Graduates, and Outstanding Doctoral Thesis awards from Tsinghua University. His research is granted by the Young Scientist Fund of the Natural Science Foundation of China, the Postdoctoral Special Fund of China, and industry companies of Hitachi, ZET, etc.



Depeng Jin (M'11) received the B.S. and Ph.D. degrees in electronics engineering from Tsinghua University, Beijing, China, in 1995 and 1999, respectively.

He is currently an Associate Professor and the Vice-Chair of the Department of Electronic Engineering with Tsinghua University. His research fields include telecommunications, high-speed networks, application-specific integrated circuit design, and future Internet architecture.

Dr. Jin was a recipient of the National Scientific and Technological Innovation Prize (Second Class) in 2002.



Pan Hui (M'09) received the B.Eng. and M.Phil. degrees from The University of Hong Kong, Pokfulam, Hong Kong, and the Ph.D. degree from the University of Cambridge, Cambridge, U.K.

He is currently a Faculty Member of the Department of Computer Science and Engineering with The Hong Kong University of Science and Technology, Kowloon, Hong Kong, where he directs the System and Media Laboratory. He also serves as a Distinguished Scientist with Telekom Innovation Laboratories (T-Labs), Berlin, Germany, and an Adjunct

Professor of social computing and networking with Aalto University, Espoo, Finland. Before returning to Hong Kong, he has spent several years in T-labs and Intel Research Cambridge. He has authored or coauthored over 100 research papers and has several granted and pending European patents.

Prof. Hui has founded and chaired several IEEE/ACM conferences/workshops and served on the technical program committee of numerous international conferences and workshops, including IEEE Infocom, SECON, MASS, Globecom, WCNC, and ITC.



Zhaocheng Wang (M'09–SM'11) received the B.S., M.S., and Ph.D. degrees from Tsinghua University, Beijing, China, in 1991, 1993, and 1996, respectively.

From 1996 to 1997, he was a Postdoctoral Fellow with Nanyang Technological University, Singapore. From 1997 to 1999, he was with OKI Techno Centre (Singapore) Pte. Ltd., first as a Research Engineer and then as a Senior Engineer. From 1999 to 2009, he was with Sony Deutschland GmbH, Stuttgart, Germany, first as a Senior Engineer and then as a

Principal Engineer. He is currently a Professor with the Department of Electronic Engineering, Tsinghua University. He has authored or coauthored over 60 technical papers and is a holder of 22 granted U.S./EU patents. His research areas include wireless communications, digital broadcasting, and millimeter-wave communications.

Prof. Wang is a Fellow of The Institution of Engineering and Technology. He has served as a Technical Program Committee Cochair/Member for many international conferences.



Sheng Chen (M'90–SM'97–F'08) received the B.Eng. degree in control engineering from East China Petroleum Institute, Dongying, China, in 1982, the Ph.D. degree in control engineering from City University, London, U.K., in 1986, and the D.Sc. degree from the University of Southampton, Southampton, U.K., in 2005.

From 1986 to 1999, he held research and academic appointments with The University of Sheffield, The University of Edinburgh, and the University of Portsmouth, all in U.K. Since 1999, he has been with

the School of Electronics and Computer Science, University of Southampton, where he is currently a Professor of intelligent systems and signal processing. He is also a Distinguished Adjunct Professor with King Abdulaziz University, Jeddah, Saudi Arabia. He has authored or coauthored over 470 research papers. His recent research interests include adaptive signal processing, wireless communications, modeling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, and evolutionary computation methods and optimization.

Dr. Chen is a Fellow of The Institution of Engineering and Technology. He was an ISI highly cited researcher in the engineering category (March 2004). He is a Chartered Engineer.