Effective Urban Traffic Monitoring by Vehicular Sensor Networks

Rong Du, *Student Member, IEEE*, Cailian Chen, *Member, IEEE*, Bo Yang, *Member, IEEE*, Ning Lu, *Student Member, IEEE*, Xinping Guan, *Senior Member, IEEE*, and Xuemin Shen, *Fellow, IEEE*

***Abstract*—Traffic monitoring in urban transportation systems can be carried out based on vehicular sensor networks. Probe vehicles (PVs), such as taxis and buses, and floating cars (FCs), such as patrol cars for surveillance, can act as mobile sensors for sensing the urban traffic and send the reports to a traffic-monitoring center (TMC) for traffic estimation. In the TMC, sensing reports are aggregated to form a traffic matrix, which is used to extract traffic information. Since the sensing vehicles cannot cover all the roads all the time, the TMC needs to estimate the unsampled data in the traffic matrix. As this matrix can be approximated to be of low rank, matrix completion (MC) is an effective method to estimate the unsampled data. However, our previous analysis on the real traces of taxis in Shanghai reveals that MC methods do not work well due to the uneven samples of PVs, which is common in urban traffic. To exploit the intrinsic relationship between the unevenness of samples and traffic estimation error, we study the temporal and spatial entropies of samples and successfully define the important criterion, i.e., average entropy of the sampling process. A new sampling rule based on this relationship is proposed to improve the performance of estimation and monitoring. With the sampling rule, two new patrol algorithms are introduced to plan the paths of controllable FCs to proactively participate in traffic monitoring. By utilizing the patrol algorithms for real-data-set analysis, the estimation error reduces from 35% to about 10%, compared with the random patrol or interpolation method in traffic estimation. Both the validity of the exploited relationship and the effectiveness of the proposed patrol control algorithms are demonstrated.**

***Index Terms*—Matrix completion (MC), patrol control, traffic sensing, vehicular sensor network (VSN).**

I. Introduction

TRAFFIC congestion has become a severe problem in metropolises according to the data in "TTI's 2012 Urban Mobility Report" [1]. It is reported that 5.52 billion hours of delay and 2.88 billion gallons of fuel were wasted in the urban area of the United States due to traffic congestion in 2011. To reduce the traffic congestion, information on the traffic condition plays a vital role [2], [3]. With real-time traffic information, the drivers can plan their routes to avoid crowded areas toward their destinations in the urban area. The traffic-monitoring center (TMC) can also benefit from the information as it can provide timely traffic guidance by means of traffic light control to prevent traffic congestion. Moreover, it may benefit the usage of the vehicular network's spectrum by using some opportunistic spectrum access schemes, as mentioned in [4].

The objective of traffic monitoring is to achieve the traffic condition precisely and efficiently. In addition to static sensors such as digital camera and loop detector, the vehicular sensor network (VSN) [5], [6], which benefits from the mobility of the vehicles, is an effective and economical way to sense real-time traffic conditions. For example, in Shanghai, China, the taxis and buses are equipped with onboard units (OBUs) such as a GPS receiver, a speedometer, and an accelerometer to act as probe vehicles (PVs) such that traffic conditions of the roads can be sensed by the mobile vehicles. These PVs periodically generate traffic reports and send them through cellular networks to the TMC called the Shanghai Traffic Information Center (STIC). In the STIC, a map-matching algorithm is adopted to match each report to a certain road so that the mean traveling speed of the road during a period of time can be computed. The preprocessed data are then used to estimate the traffic condition of all the roads in the cloud server. Finally, the estimation results are published by the STIC on the Internet or Traffic Message Boards in the main streets and on elevated roads of Shanghai.

Generally, the traffic matrix is used to extract traffic information. In a traffic matrix $X = \{x_{ij}\}$, the entry x_{ij} represents the traffic condition (e.g., average speed based on all the reports from PVs on the road) of the ith road in the city at the jth duty cycle of a day. Thus, each row represents the traffic reports of a specific road during all the duty cycles of a day, and each column represents the traffic reports at a specific duty cycle for all roads. Unfortunately, the distribution of the PVs leads to the uneven sampling of the traffic conditions. By the analysis on the real traces of taxis in Shanghai from the STIC, we find that the distribution is so uneven that about 70% of the roads have traffic reports for less than 30% of the duty cycles in a day. Similar results are also given in [7] and [8]. Despite the mobility feature of the vehicles, the TMC cannot guarantee to get the traffic condition for most of the roads most of the time. It is very desirable to get the traffic estimation from aspects of

Manuscript received December 18, 2013; revised March 17, 2014; accepted April 10, 2014. Date of publication April 29, 2014; date of current version January 13, 2015. This work was supported in part by the National Basic Research Project of China under Grant 2010CB731803; by the National Natural Science Foundation of China under Grant 61221003, Grant 61273181, Grant 61290322, and Grant 61174127; by the Research Foundation for the Doctoral Program of Higher Education under Grant 20110073130005; by the Program of New Century Talents in University of China under Grant NCET-13-0358, by the Science and Technology Commission of Shanghai Municipal, China, under Grant 13QA1401900; and by the Cyber Joint Innovation Center, Hangzhou, China. This work was also supported by a research grant from the Natural Science and Engineering Research Council (NSERC), Canada.

R. Du, C. Chen, B. Yang, and X. Guan are with the Department of Automation, Shanghai Jiao Tong University, Shanghai 200030, China, and also with the Key Laboratory of System Control and Information Processing, Ministry of Education of China.

N. Lu and X. Shen are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

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/TVT.2014.2321010

location-based services for vehicles and traffic management for the TMC.

To estimate the traffic condition of these unsampled roads, matrix completion (MC) [9] technique [an extension of compressive sensing (CS)] can be developed to reconstruct the data in the empty entries of the traffic matrix (see, e.g., [8] and [10]). The advantages of MC-based estimation algorithms are the efficiency and high accuracy of estimation based on only a small portion of samples in the matrix under the conditions that the singular values of the matrix are sparse and that the samples distribute uniformly randomly in the matrix. However, the sampling reports of the PVs are uneven due to the uneven urban traffic distribution, and thus, the unevenness destroys the performance of MC-based estimation. Moreover, because the MC technique depends on the minimization of the rank of matrix, it is unapplicable for the case that there are any empty rows (all the entries of the row are empty) or empty columns of the matrix. By analyzing the real traffic data in Shanghai, it will be shown in Section IV for details that about 17% of the roads have no reports for a whole day. It results in a big challenge to utilize the MC technique for traffic estimation. How to achieve the traffic conditions of these unsampled roads is a vital problem for urban traffic management. One reason is that private cars may often pass these roads, although they are not frequently passed by PVs. More importantly, they are the alternatives for traffic guidance to avoid traffic congestion. One possible approach to facilitate MC-based estimation is to improve the evenness of samples. It is more effective to avoid empty columns or empty rows in the traffic matrix.

Inspired by the fact that utilizing sensors' mobility can improve the network performance such as coverage [11], [12] and connectivity [13], we intend to control the movement of some vehicles to avoid empty rows or empty columns in the sampled matrix and, thus, reduce the estimation error. Different from buses, taxis and private cars, which just passively participate in traffic monitoring, some controllable vehicles proactively participate in traffic monitoring. In Shanghai, these controllable vehicles could be the patrol cars from the TMC and traffic cars and surveillance cars of road policing units (RPUs) in Shanghai Municipal, which are called floating cars (FCs) in the rest of this paper. The traffic cars and surveillance cars are used to assist with their duties in patrolling and responding to incidents and enforcing traffic laws in certain areas to gather evidence of any criminal offenses. We can monitor the traffic more effectively by planning patrolling paths for FCs to improve the evenness of samples in the traffic matrix. Compared with the solution of equipping more static sensors such as digital cameras and loop detectors, utilizing the FCs in such a way is an economical solution without extra costs of installation and maintenance. According to our best knowledge, this is the first time to improve the MC-based estimation by controlling the mobility of some sensors.

To achieve the reasonable control for FCs, two issues are necessary to be addressed. One issue is to explore how the unevenness of samples affects the estimation error of MC-based algorithms. The other issue is to investigate effective sampling rules (i.e., control laws of FCs) such that the MC-based estimation can be improved.

For the first issue, it is necessary to find a new criterion to describe the evenness of samples, since the commonly used criterion, i.e., sampling ratio, is demonstrated to be not appropriate to represent the correlation of the data (detailed analysis will be given in Section V). It is known that entropy can be used to measure the uncertainty of the information and also the correlation of the data. It is a potential criterion to represent the performance of the sampling process. Consequently, we explore the relationship between the estimation error and samples of the MC-based method from the perspective of entropy in this paper. It turns out to be significant for designing the sampling rules. By analyzing the traffic data from the STIC in Shanghai, we introduce the criterion of average entropy of samples to evaluate the estimation. This kind of approach is not seen in the open literature.

For the second issue, we come up with a sampling rule based on the introduced average entropy so that the estimation error can be minimized with a small sampling ratio. According to the rule, we propose two patrol algorithms for FCs by minimizing the average entropy of each road.

The main contributions of this paper are as follows.

1) By analyzing the traffic data of STIC in Shanghai, it is found that the apparent uneven distribution of the PVs severely influences the performance of the MC-based estimation of the urban traffic conditions. Moreover, the new criterion of average entropy is introduced to describe the unevenness of samples.
2) The relationship between the estimation error of the MC-based method and average entropy of the samples is explored and modeled by a log function.
3) Based on the explored relationship between estimation error and average entropy, the control method of FCs are given in patrol ways. The number of required FCs can be determined to satisfy a predescribed estimation error.

The remainder of this paper is organized as follows. The related works are discussed in Section II. In Section III, we describe the system model for urban traffic estimation and show the preliminary of the MC technique. The analysis of the real traces is shown in Section IV. The relationship between the estimation error and entropy of samples is explored in Section V. Based on the proposed relationship, patrol algorithms for FCs are given in Section VI. Simulation studies are given in Section VII to evaluate the patrol algorithms, followed by the conclusion in Section VIII.

II. Related Works

Here, we discuss the related works on traffic estimation, mobility management, and patrol control of path planning and highlight the contributions of this paper.

A. Traffic Monitoring With VSN

There are some approaches for the estimation of the traffic conditions by using vehicles. In [16]–[18], PVs count the number of neighbor vehicles via broadcasting a beacon message and receiving responses through OBUs to estimate the traffic density. This way, traffic conditions can be estimated without equipping all vehicles with GPS. However, these methods, namely, Infrastructure-Free Traffic Information System in [16] and MobSampling in [17], can only estimate microscopic traffic density since they are not scalable for large-scale urban traffic.

Another way for infrastructure-free traffic estimation is CS. It is a signal processing technique for efficiently reconstructing a signal from relatively few measurements by taking advantage of the signal's sparseness or compressibility. There have been a lot of works on efficient data gathering for sensor networks with compression and recovery techniques in the CS methodology. Luo *et al.* [19] first proposed a compressive data gathering (CDG) for large-scale WSNs based on CS. The proposed method can reduce the global data traffic and balance the energy consumption of the sensor nodes. Similarly, the so-called compressed-sparse-function method is proposed in [20] to gather the data in a WSN. The sparse function is used to compress the original data of sensor nodes. Instead of utilizing the CS-based algorithm to recover the observations of sensors, Cheng *et al.* proposed spatiotemporal CDG in [21] to gather the data based on the MC technique. Inspired by the CS/MC-based data gathering scheme in WSNs, Liu *et al.* extended the CDG-based scheme to vehicular *ad hoc* networks (VANETs) in [22], where vehicles with OBUs need to report the spatial correlated data to the roadside units (RSUs) through multihop transmission. During the transmission, the data of vehicles are compressed with the similar technique as CDG and then recovered by RSUs. Taking the scarce communication resource among vehicles into account, Wang *et al.* proposed CS-based monitoring (CSM) in [23] for large-scale monitoring applications with vehicular networks. They used entropy analysis on the traffic data to show the strong correlation in the data readings of vehicles. In addition to CS, MC is also used in monitoring application with vehicular networks. In [8], it is reported that the traffic matrix can be well considered as low rank such that the MC technique can be applied.

The aforementioned CS/MC-based methods can be used to gather traffic data efficiently with sparsifying sensor (or vehicle in [22]) data; however, they are still unable to estimate the data in empty columns of the traffic matrix for urban traffic estimation. In one of our previous works in [10], the traffic estimation problem is also formulated into an MC problem by taking advantage of the low-rank feature, which represents sparsity in singular values. To achieve more reasonable estimations, we paid great attention to the temporal continuity of the traffic condition and the bounds of data in the traffic matrix and set a projection for a graded traffic matrix without empty columns. Then, the efficient Half Thresholding MC for Traffic Estimation and Monitoring (HaTTEM) method is proposed with low computational complexity. However, only those roads of TMC's interest can be estimated. In this paper, the HaTTEM method is used to estimate the traffic for all roads in the city. It means that much more empty columns are to be treated in the traffic matrix, and the method must be scalable. To the authors' best knowledge, few of the existing CS/MC works for WSNs and VANETs considered the influence of the sampling process. We will show in Section IV that in urban VSNs, the data-gathering cost dramatically increases if we only simply increase the number of PVs to collect more data. It is even worse that the estimation error may fail to obviously reduce with a higher sampling ratio in the traffic matrix, based on our analysis of a large real data set of bus and taxi traces in Section V. The features of samples (data in the traffic matrix) becomes significant for estimation effectiveness and efficiency, and they need to be exploited.

In this paper, by analyzing real traces of taxis in Shanghai, China, we reveal that the evenness of samples strongly relates to the MC-based estimation of the traffic matrix. Motivated by the important observation, we explore this intrinsic relationship by a new criterion of average entropy of samples. Different from the existing works, this paper focuses on proactively sensing the needed traffic data to improve the estimation performance with a lower sampling ratio. Thus, there is almost no extra cost for sensing and data gathering by only replanning the paths of FCs.

B. Mobility Management in Ad Hoc Networks

The mobility of the network nodes has been leveraged to improve the coverage and connectivity of the networks. Liu *et al.* [11] used mobile sensor nodes to cover a region for intrusion detection, so that the coverage of the network depends not only on the initial network configuration but on the sensors' mobility as well. He *et al.* [12] proposed that mobile nodes should move using the knowledge on the intruder to further improve the coverage of mobile sensor networks. To improve the connectivity, mobile nodes act as a data mule to deliver the data from the isolated area to the sink or other nodes [13], [14] so that the network can be deployed not as dense as the static sensor networks, and thus, it reduces deployment costs. Moreover, Ma *et al.* [15] proposed a mobile data-gathering scheme, in which the tour of the mobile sink is carefully planned to prolong the lifetime of the wireless sensor networks.

In this paper, we try to manage the mobility of FCs in VSN to improve the performance of the system. The objective is to change sampling to improve the evenness and, thus, reduce the traffic estimation error of the MC-based method.

Since FCs may be the surveillance cars of RPUs, patrolling control is an effective way to plan the paths of FCs. Chevaleyre gave some theoretical results for the patrolling problem in [24] and compared the performance of partition-based patrol and the cyclic patrol. In [25], the so-called multilevel subgraph patrolling algorithm is proposed for multiple robots to patrol. The area is modeled as a graph, and the robots need to maximize the visit frequency for every vertex in the graph. The patrol algorithm is based on Hamiltonian cycles and longest path. Different from them, the proposed patrol algorithms in this paper try to maximize the visit frequency for every edge in the graph.

III. System Models and Preliminaries

A. System Models

Consider a VSN for urban traffic monitoring, as shown in Fig. 1. PVs and FCs act as mobile sensor nodes to sense the traffic condition of the roads with assistance from on-board sensors such as a GPS receiver, a speedometer, and an

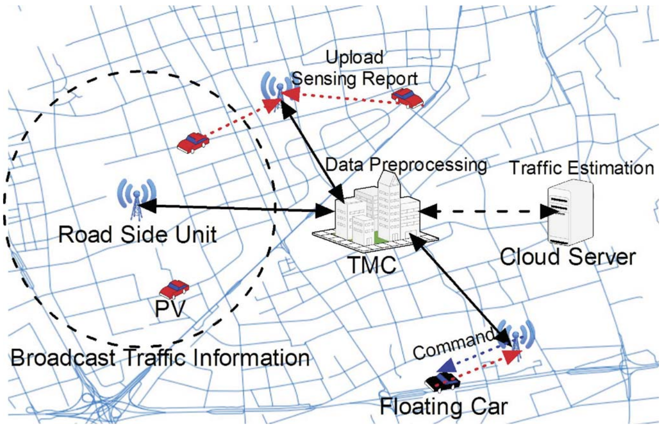

Fig. 1. Architecture of VSN for urban traffic estimation.

accelerometer. Generally, the PVs can be buses and taxis that passively participate in traffic monitoring. The moving paths of buses and taxis are predefined and determined by the drivers or the passengers, respectively. Hence, they are uncontrollable by the TMC. On the other hand, FCs, such as patrol vehicles from the TMC that are specially used for traffic monitoring, or surveillance cars from RPUs, proactively participate in traffic monitoring. They choose their paths autonomously or reactively according to the guidance from the TMC to travel a set of given roads. Consequently, their paths are controllable (or partially controllable) according to the requirement of traffic monitoring and other surveillance purposes.

In the considered system, PVs and FCs upload their sensing reports, including location, time of report, current speed, and headings, to the TMC via GPRS channel or RSU as reported in [7] and [26]. Then, the TMC collects all the traffic reports and preprocesses them for estimation. For example, the location of each report is matched to one road by the map-matching algorithm, data from different sensors are fused, and traffic matrix X can be formed. The MC-based estimation of the traffic condition could take place in the cloud server, as shown in Fig. 1. The estimation result is then sent back to the TMC and published to vehicles in the city.

The roads in the city are modeled as the links in the directed graph $G = \{V, E\}$, where V is the vertex set and represents the intersections, and E is the edge set. $\langle v_i, v_j\rangle \in E$ if and only if the vehicles can drive from the intersection v_i to the intersection v_j directly. We refer to a directed edge as a link, and a road is the undirected path connecting two intersections in the rest of the paper. n_1 is the number of links whose traffic conditions need to be estimated. The sensing reports of PVs in the past time duration T are used. Define the time granularity as ΔT, where the time duration T is partitioned into $n_2 \triangleq \lceil T/\Delta T\rceil$ duty cycles. The TMC constructs an indicate matrix $A = \{a_{ij}\}$, where $a_{ij} = 1$ for the case where at least one report of the ith link during the jth duty cycle is uploaded; otherwise, $a_{ij} = 0$. Hence, the samples of PVs can be represented by the indicate matrix A.

Define the traffic matrix to be estimated as $X = \{x_{ij}\} \in \mathbb{R}_+^{n_1\times n_2}$, where $\mathbb{R}_+^{n_1\times n_2}$ represents positive real space with the dimension of $n_1 \times n_2$. Let the sample matrix be $Y = \{y_{ij}\} \in \mathbb{R}_+^{n_1\times n_2}$, where y_{ij} denotes the average speed of all the PVs in the ith link during the jth duty cycle. Obviously, the entry y_{ij} is empty if $a_{ij} = 0$, and x_{ij} can be well estimated by $x_{ij} = y_{ij}$ for $a_{ij} = 1$. The objective of traffic estimation is to estimate all the entries of X based on the samples Y.

B. MC-Based Traffic Estimation

It has been shown in [8] that the traffic matrix can be approximated to a low-rank matrix. Hence, MC is used in this paper to estimate the unsampled values in the traffic matrix via solving the following general problem:

$$\hat{X} = \arg\min_{X} \mathbf{rank}(X)$$
$$s.t. \quad P_\Omega(Y) = P_\Omega(X) \tag{1}$$

where $\Omega = \{(i,j)|a_{ij} = 1\}$ is the set of the index of the sampled data of X, and $P_\Omega(X) : \mathbb{R}_+^{n_1\times n_2} \to \mathbb{R}_+^{n_1\times n_2}$ is the projection to let $x_{ij} = 0, \forall(i,j) \notin \Omega$. With the knowledge that all the values in the traffic matrix should be nonnegative and that the observations of each link usually do not rapidly change except for the peak hour, we modify problem (1) to the following problem for more realistic sake. Thus

$$\hat{X} = \arg\min_X \left\{\|\mathbf{svd}(X)\|_0 + \gamma\,\|T(X)\|_2^2 + \gamma\,\|B(X)\|_2^2\right\}$$
$$s.t. \quad P_\Omega(X) = P_\Omega(Y) \tag{2}$$

where $\|\cdot\|_0$ is the L_0-norm of a vector (returns the number of nonzeros in the vector), $\mathbf{svd}(X)$ denotes the singular values of X, $T(X)$ is the punishment if large changes exist among two consecutive samples at the same link, and $B(X)$ is the punishment of its negative entries. To solve problem (2), we adopt the HaTTEM algorithm proposed in our previous paper [10] by the following subproblems iteratively:

$$\begin{cases} u^{k+1} = \arg\min_u \left\{\lambda\gamma\|u\|_2^2 + \left\|u - T(X^k)\right\|_2^2\right\} \\ v^{k+1} = \arg\min_v \left\{\lambda\gamma\|v\|_2^2 + \left\|v - B(X^k)\right\|_2^2\right\} \\ X^{k+1} = \arg\min_X \Big\{\lambda\,\|\mathbf{svd}(X)\|_0 + \|P_\Omega(X) - P_\Omega(Y)\|_2^2 \\ \qquad\qquad + \left\|T(X) - u^{k+1}\right\|_2^2 + \left\|B(X) - v^{k+1}\right\|_2^2\Big\}. \end{cases} \tag{3}$$

The main idea of algorithm (3) is as follows. First, estimate some values in the unsampled indexes according to the temporal continuity and data bound of the traffic data, respectively. Second, use the sampled data, together with the estimated data, to solve the minimization problem of the rank of the traffic matrix. Due to space limitations, the detailed HaTTEM algorithm is not shown here (see [10] for more details). In the rest of this paper, $\mathrm{MC}(P_\Omega(X))$ represents the estimation result by using HaTTEM to recover the traffic matrix according to the sample matrix $Y = P_\Omega(X)$.

IV. Analysis of Unevenness of Samples in Real Vehicle Traces

For MC-based estimation methods, the data are required to be sampled uniformly randomly to achieve high estimation

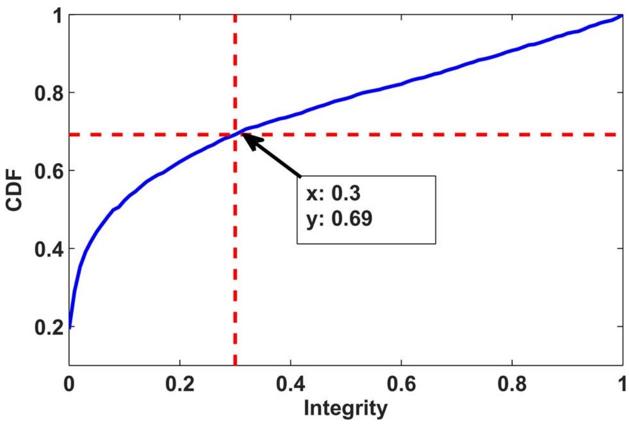

Fig. 2. CDF of integrity of links.

accuracy in data reconstruction. However, it has been shown that the distribution of the traffic is uneven [4], [7], which may be due to social features discussed in [27] and [28]. It leads to uneven sampling in the traffic matrix. We exploit the effect of unevenness on the estimation performance under real distribution of PVs in the city.

Here, we first use a real-trace data set from the STIC, Shanghai, to analyze the distribution of PVs. Each report in the data set includes information on the vehicle ID, the GPS reading of the vehicle, the time of record, the vehicle's speed, and the heading direction of the vehicle. The data that we use to analyze were collected on January 24, 2013, and there are about 65 million lines of records. Hence, there is about one record of each link every minute on average. Due to the uncertainties in GPS records, a "point-to-arc"-based map-matching method is applied in the analysis. We find the closest link according to a GPS record as the traveling link of a PV, subject to the requirement that the angle difference between the link's direction and the vehicle's heading direction is smaller than $\pi/4$.

In our analysis, we divide a day into 288 duty cycles with 5 min per duty cycle. Then, the integrity of link i is defined as $I_i = \sum_{j=1}^{288} a_{ij}/288$. The cumulative distribution function (cdf) of integrity is shown in Fig. 2. It indicates that about 70% of the links have reports for less than 30% of the time, and the average integrity of all links in the city is 24.26%. This result is similar to what is reported in [8] based on the data collected in 2007. For MC-based estimation methods, a matrix cannot be estimated if any row or column is empty. However, according to analysis of cdf of integrity, nearly 17% of the links have no reports for a whole day. It represents that 17% of the rows in the traffic matrix is empty, and the conditions of the corresponding links are hardly to be estimated even using the MC technique. One possible solution is to deploy static sensors, such as inductive loops and digital cameras, on these roads. To achieve the desired sensing coverage, the deployment and maintenance cost of static sensors are prohibitive [8], [29]. Comparatively, replanning some FC paths in a VSN is an economical solution, as has been noted in Section I.

To evaluate the performance of MC-based methods, we choose the links whose integrity is 100% to construct a new traffic matrix X_{full}. Then, X_{full} is sampled following the statistic integrity of the links, as shown in Fig. 2, and $\hat{X}_{\text{full}}$ is achieved by using algorithm (3). It results in relative estimation error $Err \triangleq \|\hat{X}_{\text{full}} - X_{\text{full}}\|_F / \|X_{\text{full}}\|_F = 35\%$ on average, where $\|\cdot\|_F$ is the Frobenius norm of a matrix. However, the estimation error could be reduced to about 10% if X_{full} is randomly sampled. It implies that the performance of algorithm (3) is strongly related to the distribution of samples and is greatly limited by the unevenness, particularly by the empty rows.

From the important observation, we exploit the explicit relationship between the estimation error and the unevenness of samples. We first give the following important definitions.

Definition 1: (**Empty row/column**) A row/column is called an **empty row/column** if and only if all the entries of the row/column are unsampled.

Definition 2: (**Coverage for Matrix Completion, MC-Coverage**) A matrix is of **MC coverage** if and only if neither empty columns nor empty rows exist in the matrix.

Notice that if a low-rank matrix is not MC-covered, it cannot be recovered from the samples by MC-based estimation. Hence, we estimate the number of required PVs to make the traffic matrix X MC-covered. For simplicity, suppose all PVs are independent [27] and identically distributed. Let $p_i(j)$ denote the probability of a PV on the link i reporting its reading during the jth duty cycle. The following theorem gives the probability P_{MC} of N PVs to **MC-cover** n_1 links in n_2 duty cycles.

Theorem 1: Given a selected link set E with the number of links $|E| = n_1$, the probability for N PVs to **MC-cover** all the links in E within n_2 duty cycles P_{MC} is

$$P_{\text{MC}} = \prod_{i=1}^{n_1} \left[1 - \prod_{j=1}^{n_2} (1 - p_i(j))^N \right]. \tag{4}$$

Proof: As PVs are on the road, there will be no empty columns in the traffic matrix X. Therefore, we only need to consider the cases of empty rows.

An empty row in the traffic matrix X implies that there is no PV reporting the traffic condition at the corresponding link. If the ith link is not sampled during the jth duty cycle, the probability can be calculated by the case that no PVs traverse the ith link during the jth duty cycle, which is equal to $(1 - p_i(j))^N$. Hence, the probability that the ith link is not sampled for all n_2 duty cycles is $\prod_{j=1}^{n_2} (1 - p_i(j))^N$. Since there are n_1 links to be MC-covered, (4) is obtained. ■

Due to the unevenness of city traffic, $p_i(j)$ is quite different for different links. Moreover, it also varies for different duty cycles. For example, the probability is obviously higher in the peak hours than at midnight. We need to estimate $p_i(j)$ according to the traffic distribution to determine the required number of PVs for MC coverage.

For simplicity, $p_i(j)$ is estimated by $\hat{p}_i(j) = c_{ij}/V_{\text{sum}}$, where c_{ij} is the number of reports of the ith link during the jth duty cycle, and V_{sum} is the number of PVs in the system. In the data set we analyzed, the number of different vehicle IDs is 32122. As a result, V_{sum} is set to be 32122.

Notice that there are some links with integrity 0, and it is not proper to set $\hat{p}$ of these links as zero according to

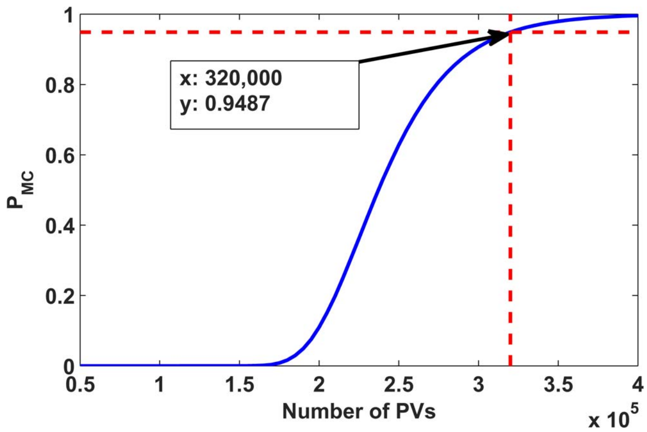

Fig. 3. Probability of MC coverage with respect to the number of PVs.

the aforementioned estimation method. This is because if the probabilities are set as zero, the corresponding rows in the indicate matrix will definitely be empty no matter how many PVs are deployed into the system. Hence, we set

$$\hat{p}_i(j) = \begin{cases} \frac{1}{n_2 V_{\text{sum}}}, & \text{if } \forall j \in [1, n_2], c_{ij} = 0 \\ \frac{c_{ij}}{V_{\text{sum}}}, & \text{otherwise.} \end{cases} \tag{5}$$

We then calculate P_{MC} according to (4) and (5), and the result is shown in Fig. 3. It can be seen that more than 320 000 PVs are required to achieve MC coverage with high probability of 95%. Considering the privacy issue pointed out in [30], private cars are not willing to use OBUs to upload their locations and running states periodically. It means that we still need about 300 000 taxis or buses acting as PVs, which is obviously unable to be fulfilled currently. It can be observed that simply increasing the number of PVs is not a pragmatic method for traffic monitoring based on the MC method in VSN. It is thus very necessary to find an efficient way to improve the evenness of PVs' samples so that the estimation error can be reduced from 35% to about 10%.

V. Relationship Between Estimation Error and Entropy of Traffic Samples

We have observed that simply increasing the number of PVs on the road cannot tackle the difficulty in the improvement of evenness of PVs' samples. It can be explained by the fact that the roads in the city seldom passed by taxis still have a small possibility of being reported although more PVs are deployed. As a result, we need to increase the reports on these links by replanning the paths of some controllable FCs. With this target, we first need to exploit the proper criterion to represent the evenness of samples. Then, we can further investigate the relationship between the estimation error of the MC-based method and the evenness of samples. Data correlation is used to explain such phenomenon in our previous work [31], but the relationship is still unexplored. It is significant for constructing control laws for FCs.

It is known that the disorder of samples can be expressed by entropy. The more random the samplings are, the more precise the estimation of the traffic matrix could be. Thus, we use entropy as the criterion to evaluate the sampling process and explore the connection between the entropy and the estimation error. For simplicity, we choose the links with 100% integrity to form X_{full} and then measure the entropy. Each entry in X_{full} is considered as a state, and we can analyze the temporal entropy and spatial entropy, followed by the relationship between entropy and estimation error.

A. Temporal Entropy

Suppose that there are K different states in X_{full}, and the temporal entropy $H_T(t)$ at time t is calculated as follows:

$$H_T(t) = \sum_{i=1}^{K} \Pr_T(i,t) \ln \frac{1}{\Pr_T(i,t)}$$

where $\Pr_T(i,t)$ is the probability that state i appears in the tth duty cycle and is achieved by taking the statistic of X_{full}. Let $k_{i,t}$ denote the number of times that state i appears in the tth column of X_{full} and n_1' the number of rows in X_{full}, $\Pr_T(i,t)$ is then estimated by $k_{i,t}/n_1'$. Based on the temporal entropy, we can calculate the mutual information between two traffic states of the same link at different times, which is defined as

$$I_T(t_1,t_2) = H_T(t_1) + H_T(t_2) - H_T(t_1,t_2) \tag{6}$$

where $H_T(t_1,t_2)$ is the joint temporal entropy of the t_1th and the t_2th duty cycle. Naturally, the closer the two duty cycles are, the stronger the correlation between the data is, and thus, the larger the mutual information is. To evaluate (6), we define the k-temporal mutual information (KTMI) as the averaged temporal mutual information of two duty cycles with the fixed time interval k as follows:

$$M_T(k) = \frac{1}{n_2 - k} \sum_{i=1}^{n_2-k} I_T(t_i, t_{i+k}).$$

Based on the curve-fitting tool in MATLAB, we plot the KTMI of X_{full} and find that it is well fitted by an exponential curve $f(k) = \alpha \exp(-\beta k) + \gamma$, as shown in Fig. 4, and the adjusted R-square of the fitting is 0.9864. Hence, KTMI can be estimated by $M_T(k) = 0.1226 \exp(-0.03462k) + 1.12782$, and $I_T(t_1,t_2) = M_T(|t_1 - t_2|)$.

Moreover, we can calculate the temporal conditional entropy as follows:

$$\begin{aligned} H_T(t_1|t_2) &= H_T(t_1,t_2) - H_T(t_2) \\ &= H_T(t_1) - I_T(t_1,t_2). \end{aligned}$$

As we have $I_T(t_1,t_2) = M_T(|t_1 - t_2|)$, the temporal conditional entropy can be calculated by

$$H_T(t_1|t_2) = H_T(t_1) - M_T(|t_1 - t_2|). \tag{7}$$

Let $x_\Omega = \{x_{ij} | (i,j) \in \Omega\}$ as the set of the sampled data. Then, we define the conditional temporal entropy (CTE) of the sampling process $H_T(x_{ij}|x_\Omega)$ as the temporal conditional entropy of x_{ij} based on the knowledge of x_Ω.

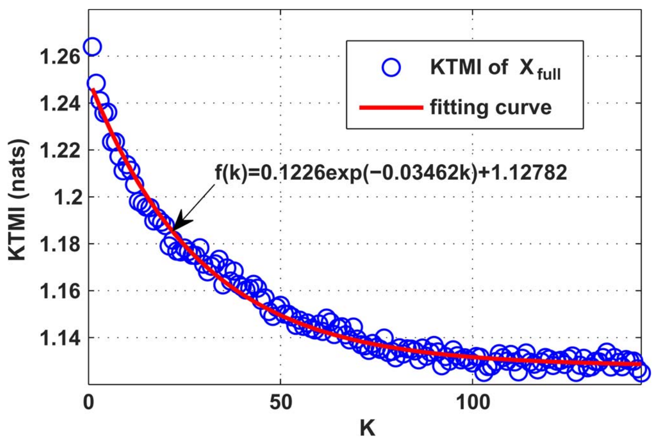

Fig. 4. Fitting curve of KTMI (1 nat corresponds to 1/ln 2 bits).

While calculating the CTE of an entry, e.g., x_{ij}, links are assumed to be independent of each other. Moreover, we only consider the mutual information of its adjacent sampled data at the same link. Consequently, the CTE of x_{ij} is approximated by the average of the conditional temporal entropies of x_{ij} with its neighboring sampled data, as shown in the following equation:

$$\hat{H}_T(x_{ij}|x_\Omega) = \frac{H_T\left(x_{ij}|x_{ik^l_T(i,j)}\right) + H_T\left(x_{ij}|x_{ik^u_T(i,j)}\right)}{2}, \tag{8}$$

where $k^l_T(i,j) = \max\{k|a_{ik} = 1 \wedge k \le j\}$ and $k^u_T(i,j) = \min\{k|a_{ik} = 1 \wedge k \ge j\}$ are the column indexes of x'_{ij} adjacent sampled entry. According to (7), we have

$$\hat{H}_T(x_{ij}|x_\Omega) = H_T(t_j) - \frac{1}{2}\left[M_T\left(|k^u_T(i,j) - j|\right) + M_T\left(\left|j - k^l_T(i,j)\right|\right)\right]. \tag{9}$$

B. Spatial Entropy

Similarly, if we let l_j denote the jth link, then the spatial entropy, spatial mutual information, and spatial conditional entropy are defined as follows:

$$H_S(l) = \sum_{i=1}^{K} \mathrm{Pr}_S(i,l) \ln \frac{1}{\mathrm{Pr}_S(i,l)}$$

$$I_S(l_1,l_2) = H_S(l_1) + H_S(l_2) - H_S(l_1,l_2)$$

$$H_S(l_1|l_2) = H_T(l_1) - I_S(l_1,l_2)$$

where $Pr_S(i,l)$ is the probability that the state i appears in the lth road and is estimated by taking the statistic of X_{full}.

Unlike temporal entropy, we fail to find a curve that can fit the spatial mutual information with the distance between two links. It can be explained that nearly all the roads are bidirectional and that the condition of one direction, e.g., $\langle v_1, v_2\rangle$, is quite irrelevant to its opposite direction, e.g., $\langle v_2, v_1\rangle$. As a result, the spatial mutual information of the two directions is small, although the distance between them is nearly zero. Moreover, the correlation between $\langle v_2, v_1\rangle$ and its predecessor, e.g., $\langle v_3, v_2\rangle$, is strong. While the correlation between $\langle v_3, v_2\rangle$ and $\langle v_1, v_2\rangle$ is not so strong since they only share the same end node. Notice that the spatial distance between $\langle v_2, v_1\rangle$ and $\langle v_3, v_2\rangle$ is equal to the distance between $\langle v_1, v_2\rangle$ and $\langle v_3, v_2\rangle$, the correlations greatly differ. Consequently, we cannot find an obvious relationship between the spatial mutual information and the distance of the two links. As a result, we need to calculate $I_S(l_1, l_2)$ for each (l_1, l_2) pair according to the traffic data instead of estimating by $|l_1 - l_2|$ as temporal mutual information does.

Similarly to CTE, the conditional spatial entropy (CSE) of the sampling process $H_S(x_{ij}|x_\Omega)$ is defined as the spatial conditional entropy of x_{ij} based on the knowledge of x_Ω. CSE is estimated by the following equation:

$$\hat{H}_S(x_{ij}|x_\Omega) = H_S(l_i) - I_S\left(l_{k_S(i,j)}, l_i\right) \tag{10}$$

where $k_S(i,j) = \arg\max_k\{I_S(l_k,l_i)|a_{kj} = 1\}$ is the index of the link most related to the ith link and sampled in the jth duty cycle.

C. Average Entropy of Samples

Here, we show how to evaluate a sampling process based on temporal entropy and spatial entropy.

Notice that the uncertainty of the estimation comes from the empty entries in the matrix, and the entropy of samples is defined as the entropy of the unsampled databased on the knowledge of the sampled data. It is shown as

$$H_\Omega(X) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(1-a_{ij})H(x_{ij}|x_\Omega),$$

where $H(x_{ij}|x_\Omega)$ is the conditional entropy of x_{ij} based on the knowledge of x_Ω. Although $H(x_{ij}|x_\Omega)$ can be calculated based on $H(x_{ij})$, it is hardly achieved due to the limited data. Therefore, we use temporal entropy and spatial entropy instead, and the entropy of the samples is modified into

$$H_\Omega(X) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(1-a_{ij})\left[H_T(x_{ij}|x_\Omega) + H_S(x_{ij}|x_\Omega)\right]. \tag{11}$$

Then, define the corresponding temporal entropy and spatial entropy as follows:

$$H_{T,\Omega}(X) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(1-a_{ij})H_T(x_{ij}|x_\Omega),$$

$$H_{S,\Omega}(X) = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(1-a_{ij})H_S(x_{ij}|x_\Omega).$$

Thus, $H_\Omega(X) = H_{T,\Omega}(X) + H_{S,\Omega(X)}$.

Moreover, define the average entropy of samples as

$$\bar{H}_\Omega(X) = \bar{H}_{T,\Omega}(X) + \bar{H}_{S,\Omega}(X) \tag{12}$$

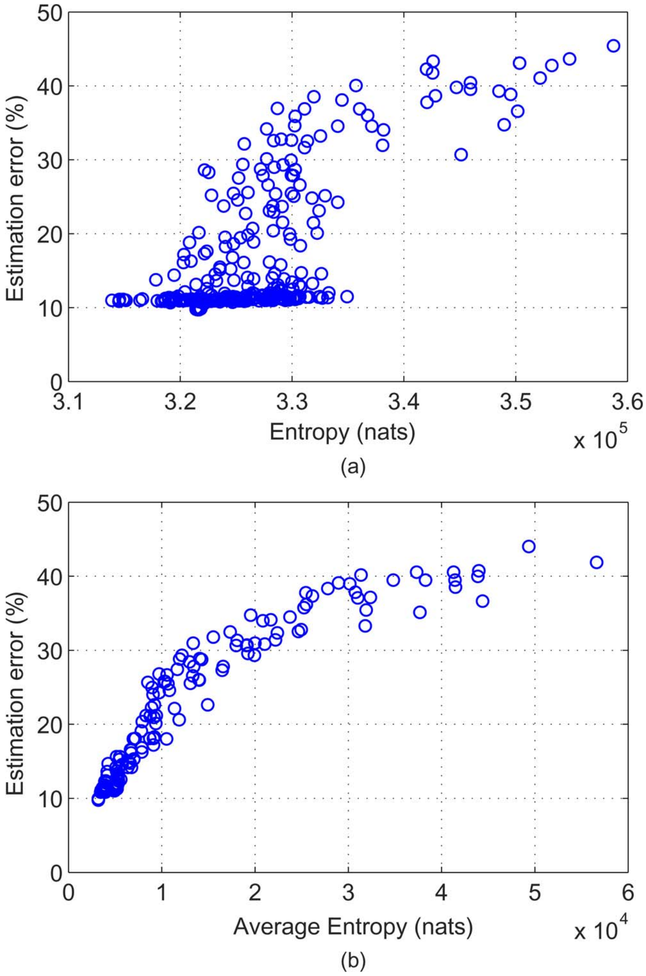

Fig. 5. (a) Estimation error versus the entropy of the sampling process (1 nat corresponds to 1/ln 2 bits). (b) Estimation error versus the average entropy of the sampling process.

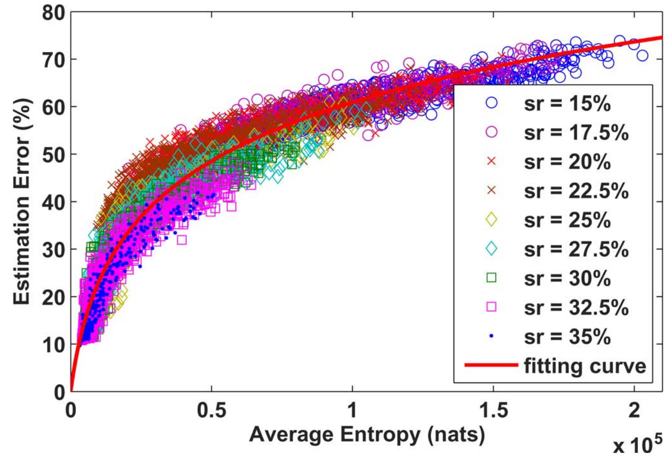

Fig. 6. Estimation error versus average entropy under different sample rates (1 nat corresponds to 1/ln 2 bits).

where

$$\bar{H}_{T,\Omega}(X) \triangleq \sum_{i=1}^{n_1} \frac{\sum_{j=1}^{n_2}(1-a_{ij})H_T(x_{ij}|x_\Omega)}{\sum_{j=1}^{n_2} a_{ij}}. \tag{13}$$

In (13), the averaged temporal entropy (ATE) of a link is given by getting the temporal entropy sum of a link, divided by the link's number of the duty cycles that is sampled. $\bar{H}_{T,\Omega}(X)$ is then the sum of all the links' ATE. Similarly, we have

$$\bar{H}_{S,\Omega}(X) \triangleq \sum_{j=1}^{n_2} \frac{\sum_{i=1}^{n_1}(1-a_{ij})H_S(x_{ij}|x_\Omega)}{\sum_{i=1}^{n_1} a_{ij}}. \tag{14}$$

To explore the relationship between the entropy and estimation error of the MC-based method, we generate a random sampling matrix Ω for X_{full} to get a new traffic matrix. Then, calculate the entropies (11) and (12), respectively. Then, we reconstruct X_{full} via the HaTTEM algorithm introduced in Section III-B. Since HaTTEM is an iterative algorithm, a constant maximum iteration is set in the test. Fig. 5 shows the results with the sampling ratio of 35%.

It is obviously seen in Fig. 5 that the reconstruction error is more related to average entropy than to entropy. More tests are given with different sampling ratios varying from 15% to 35% with the difference of 2.5%. For a given sampling ratio, we test 1800 times and give the estimation error in Fig. 6.

Interestingly, it is seen in Fig. 6 that the reconstruction error has a uniform relationship with the average entropy. Moreover, the estimation error can still be as low as 10% and 15% even if the sampling ratio is set as low as 15% and 25%, whereas the estimation error can be as bad as 35%, even if the sampling rate is 35%. It implies that the sampling ratio does not affect the estimation error directly. **What takes effect on the error of the MC-based method is the average entropy of samples**. Hence, the average entropy is chosen in this paper as the proper criterion for evenness of samples. Furthermore, it is also used to evaluate the relationship between the estimation error and the unevenness of samples.

It is desired to introduce an explicit function to show the relationship between the estimation error and the average entropy. It is obvious that if all the entries of a matrix are sampled, the average entropy is 0, as is estimation error. Consequently, the curve of function should pass the origin (0,0). We model the log-like function in Fig. 6 by the function $f(\bar{H}) = b\log(1+c\bar{H})$, where $\bar{H}$ is the average entropy, and $b>0$ and $c>0$ are the adjusting parameters to be determined. Then, we again use the curve-fitting tool in commercial software MATLAB to identify the parameters. The fitting result is $f(\bar{H}) = 0.1894\log(1+2.392\cdot 10^{-4}\bar{H})$, which is the red solid line in Fig. 6. It is found that the value of $\bar{H}_{T,\Omega}$ is very close to $\bar{H}_{S,\Omega}$. Then, we divide $\bar{H}_{T,\Omega}$ by the number of links and get the ATE per link, which is denoted by $\bar{H}_l$. The relationship of the estimation error of the MC-based method and the average entropy of samples is given in the following form:

$$\begin{aligned} Err\left(\bar{H}_\Omega(X)\right) &\triangleq \frac{\|\mathrm{MC}\left(P_\Omega(X)\right)-X\|_F}{\|X\|_F} \\ &= f\left(\bar{H}_\Omega(X)\right) = b\log\left(1+c\bar{H}_\Omega(X)\right). \end{aligned} \tag{15}$$

Moreover, it yields $Err_l = b_l\log(1+c_l\bar{H}_l)$ as the relationship of the estimation error of the MC-based method and the ATE per link.

VI. Sampling Rule Design for Floating Cars

Based on the relationship between estimation error and samples discussed in Section V, the main idea of this paper is to control the movement of FCs to minimize the average entropy of sampled traffic matrix Y.

In addition to PVs, suppose that there are n_f FCs, which can be controlled, running in the city to sense the traffic condition of the roads. Let $e(t_j, i)$ denote the index of the link where the ith FC is running at time t_j, and $\overrightarrow{E}_i = \{e(t_1, i), e(t_2, i), \ldots, e(t_f, i)\}$ denote the path of the ith FC during the time from t_1 to t_j. With the duty cycle of sampling, the resulting index set of FC's samples is denoted as

$$\Omega_i = R(\overrightarrow{E}_i) = \{R(e(t_1, i)), R(e(t_2, i)), \ldots, R(e(t_f, i))\}$$

where $R(e(t_j, i)) = (e(t_j, i), \lfloor t_j/\Delta T \rfloor)$ is the combinatory index of the link and duty cycle for the ith FC at time t_j.

Thus, the sampling process of all the FCs can be represented as $\Omega_f = \bigcup_{i=1}^{n_f} \Omega_i$. Let Ω_p be the sampling process of all PVs, and the index of all the sampled data is redefined as $\Omega = \Omega_f \cup \Omega_p$. Define $\overrightarrow{E}_f = \{\overrightarrow{E}_1, \overrightarrow{E}_2, \ldots, \overrightarrow{E}_{n_f}\}$ as the route vector of all FCs. The estimation problem is formulated as

$$\begin{aligned} \min_{\overrightarrow{E}_f} \quad & \|\hat{X} - X\|_F \\ \text{s.t.} \quad & \hat{X} = \mathrm{MC}\left(P_\Omega(X)\right) \\ & B(\overrightarrow{E}_f, X) = 1 \end{aligned} \tag{16}$$

where $B(\overrightarrow{E}_f, X)$ is a bool function that indicates whether the paths $\overrightarrow{E}_f$ of all the FCs satisfy the traffic condition X.

Note that it is not only difficult to construct the bool function $B(\cdot,\cdot)$ but also impossible to judge whether the paths of FCs satisfy the traffic condition before we have estimated X. That is why we try to reduce the average entropy of samples but not solve the minimization problem.

According to (15), we have

$$\|\mathrm{MC}\left(P_\Omega(X)\right) - X\|_F = \|X\|_F f\left(\bar{H}_\Omega(X)\right).$$

Noticing that $f(\cdot)$ is a monotonically increasing function, problem (16) can be equivalently formulated into

$$\begin{aligned} \min_{\overrightarrow{E}_f} \quad & \bar{H}_\Omega(X) \\ \text{s.t.} \quad & B(\overrightarrow{E}_f, X) = 1. \end{aligned} \tag{17}$$

Since ASE $\bar{H}_{S,\Omega}(X)$ is difficult to be minimized due to the unavailability of spatial mutual information, we focus on the minimization of ATE by the following theorem.

Theorem 2: Extract k samples from n_2 duty cycles $1, 2, \ldots, n_2$ of a link for traffic estimation, and form the new sampled duty cycle list $t_1, t_2, \ldots, t_k$. If $H_T(i) = H_T(j) = \zeta, \forall 1 \le i, j \le n_2$, the minimum of ATE is achieved for $t_1 = t_m - t_{m-1} = n_2 - t_k, (2 \le m \le k)$.

Proof: Notice that k sampled duty cycles partition the consecutive n_2 duty cycles into $k+1$ segments. Denote the length of the mth segment by $\omega_m, m = 1, 2, \ldots, k+1$. It is easy to get $\omega_1 = t_1 - 1$, $\omega_m = t_m - t_{m-1} - 1, (1 \le m \le k)$, $\omega_{k+1} = n_2 - t_k - 1$, and $\sum_{m=1}^{k+1} \omega_m = n_2 - k$.

The ATE of a link can be computed as

$$\bar{H}_{T,\Omega}(X) = \sum_{m=1}^{k+1} \frac{W(m)}{k}$$

where $W(m)$ is the temporal entropy at the mth segment, and it is given by

$$\begin{aligned} W(m) &= \sum_{j=1}^{\omega_m} \left\{ \zeta - \frac{1}{2}\left[M_T(j) + M_T(\omega_m - j + 1)\right] \right\} \\ &= \omega_m \zeta - \sum_{j=1}^{\omega_m} M_T(j) \\ &= \omega_m(\zeta - \gamma) - \alpha \frac{\exp(-\beta)\left[1 - \exp(-\beta\omega_m)\right]}{1 - \exp(-\beta)}. \end{aligned}$$

Hence, we have

$$\begin{aligned} \bar{H}_{T,\Omega}(X) &= \frac{1}{k}\left[(\zeta - \gamma)\sum_{m=1}^{k+1} \omega_m - \frac{\alpha \sum_{m=1}^{k+1}(1 - e^{-\beta\omega_m})}{e^\beta - 1}\right] \\ &= \frac{1}{k}\left[(\zeta - \gamma)(n_2 - k) - \frac{\alpha(k+1)}{e^\beta - 1}\right] \\ &\quad + \frac{\alpha}{k(e^\beta - 1)} \sum_{m=1}^{k+1} e^{-\beta\omega_m}. \end{aligned}$$

It follows from the fitting function $f(\cdot)$ in Section V-A that $\alpha > 0$. The ATE $\bar{H}_{T,\Omega}$ is minimized if $\sum_{m=1}^{k+1} e^{-\beta\omega_m}$ is minimized.

Based on Jensen's inequality, it yields

$$\sum_{m=1}^{k+1} e^{-\beta\omega_m} \ge (k+1)\exp\left\{-\frac{\beta\sum_{m=1}^{k+1}\omega_m}{k+1}\right\}$$

and the equivalence holds for $\omega_m = \omega_n, \forall m, n = 1, 2, \ldots, k+1, m \ne n$. Thus, the ATE of a link is minimized if $t_1 = t_m - t_{m-1} = n_2 - t_k, (2 \le m \le k)$. ∎

It is noted that the minimization of ATE reduces the average entropy and, thus, the estimation error. According to Theorem 2, the sampling rule can be given as follows: *Sampling the links with equal intervals reduces the estimation error based on the MC method.* This motivates us to propose a patrol algorithm to control the movement of FCs.

Patrol is the action of traveling around an area to supervise it at regular intervals. Different from the work in [32], where vertices are required to be traveled regularly, the links are desired to be traveled regularly in this paper.

A. Circuit Patrol

In circuit patrol, we need to find a circuit path for FCs (e.g., surveillance cars are desired in the way of circuit patrol). The required circuit path needs to be a Eulerian circuit that starts from a vertex, traverses all the links in the graph once, and returns to the starting point. Initially, FCs are deployed in the circuit, yielding an equal traveling time between every two consecutive FCs. All the FCs in the same cyclic path follow the same route.

Due to the uncontrollable moving of PVs, we divide the links into two categories: hot links, the integrity of which is larger than a threshold ρ_l; and cold links, the integrity of which is below ρ_l. The hot links are sampled by the PVs, whereas the cold links are patrolled by FCs. Construct a new directed graph G' whose edges are the cold links, and hence, the FCs need to travel the circuit path in G'.

This paper asks the following questions: 1) Does the Eulerian circuit exist in G'? 2) What should we do if the Eulerian circuit does not exist? The following lemma gives the answer.

Lemma 1: Given the strongly connected directed graph $G' = \{V', E'\}$, if $\langle v_i, v_j\rangle \in E'$ and $\langle v_j, v_i\rangle \in E'$ hold equivalently, then G' has the Eulerian circuit.

Proof: Lemma 1 is easy to prove. If $\langle v_i, v_j\rangle \in E'$ and $\langle v_j, v_i\rangle \in E'$ hold equivalently, the in-degree of every vertex is equal to its out-degree. Hence, G' has the Eulerian circuit if G' is strongly connected according to the property of the Eulerian circuit. ■

According to Lemma 1, the **circuit patrol** (C-Patrol) method is given as follows: *We first add the links whose direction are opposite to the cold links into the graph G', so that the in-degree of every vertex in $G'$is equal to its out-degree. Then, if any two parts of $G'$are not connected, they are separated into two subgraphs. For each subgraph, we find a Eulerian circuit and then assign FCs in each circuit based on the average traverse time of the circuit. The FCs move according to the Eulerian circuit.*

Next, we compute the number of FCs to be assigned to a given Eulerian circuit such that the predescribed estimation error request, i.e., Err, is satisfied. According to the relationship between the estimation error and ATE per link mentioned in Section V-C, the ATE per link, i.e., h_l, should satisfy

$$h_l \le \frac{\exp\left(\frac{Err}{b_l}\right) - 1}{c_l} \triangleq \bar{H}_l. \tag{18}$$

Based on Theorem 2, the sampling interval of a link should be the same for minimized ATE. The minimized ATE of a link under the interval w can be given according to (13) as follows:

$$\begin{aligned} h_l &= \frac{1}{\frac{n_2}{w+1} - 1}\left[\frac{n_2 w\zeta}{w+1} - \frac{n_2}{w+1}\sum_{k=1}^{w}(\alpha\exp(-\beta k) + \gamma)\right] \\ &= \frac{n_2}{n_2 - w - 1}\left[w(\zeta - \gamma) - \frac{\alpha(1 - e^{-\beta w})}{e^{\beta} - 1}\right]. \end{aligned} \tag{19}$$

From (18) and (19), we have

$$D_1 e^{-\beta w} + D_2 w \le D_3$$

where $D_1 = \alpha n_2$, $D_2 = (e^{\beta} - 1)(\bar{H}_l + n_2(\zeta - \gamma))$ and $D_3 = \bar{H}_l(e^{\beta - 1})(n_2 - 1) + \alpha n_2$. By solving the inequality, we have

$$w \le w_{\max} \triangleq \frac{D_2 W_0\left(-\frac{D_1\beta}{D_2}e^{-\frac{D_3\beta}{D_2}}\right) + D_3\beta}{D_2\beta} \tag{20}$$

where $W_0(\cdot)$ is the Lambert W-function.

Moreover, the required Err can be satisfied if the number of FCs for the Eulerian circuit is larger than the threshold $L/(\bar{v}\Delta T \cdot w_{\max})$, where L is the length of the Eulerian circuit, and $\bar{v}$ is the average moving speed in the circuit.

It is worth noting that the routes of the FCs in circuit patrol are preplanned, and the cars are not reactive to the real-time traffic condition. Thus, we try to improve the patrol control by the greedy strategy so that the FCs can choose the links to traverse according to the real-time traffic condition.

B. Greedy Patrol

The main idea of the greedy patrol is that every FC chooses the next link to traverse greedily according to the current and predicted idleness of the links in its area of interest.

The current idleness of the ith link, which is denoted by $\tau_{c,i}(t)$, is easy to get by counting the number of unsampled duty cycles. Hence, $\tau_{c,i}(t) = \lfloor t/\Delta T\rfloor - k_T^l(i, \lfloor t/\Delta T\rfloor)$. To predict future idleness, we assume that the movements of PVs follow the same random process. Let $p(i, j, t)$ be the probability of a PV starting from link i and arriving first at link j on t, which can be obtained by statistic from the trails of the PVs. Then, the expectation of the future idleness of the ith link is

$$\tau_{f,i} = \sum_{k=1}^{+\infty}(k-1)P(i, k\Delta T)$$

where $P(i, k\Delta T) = 1 - \prod_j (1 - p(j, y, k\Delta T))^{m_j}$, and m_j is the number of taxis currently in the jth link.

To reduce the computational complexity, $\tau_{f,i}$ is approximated by $\sum_{k=1}^{k_{\max}}(k-1)P(i, k\Delta T)$. Then, the idleness of link i can be calculated by $\tau_i = \tau_{c,i}(t) + \tau_{f,i}$, and the idleness map is $\boldsymbol{\tau} = \{\tau_1, \tau_2, \ldots, \tau_{n_1}\}$

The process of the **greedy patrol** (G-Patrol) is given as follows: *At the end of a duty cycle, the TMC calculates the current idleness τ_c of the links according to the reports from the PVs and FCs. Moreover, the TMC estimates the predicted idleness of the links. Then, it calculates the idleness map $\boldsymbol{\tau}$ and then report the result to FCs. Each FC greedily moves to the link with the highest idleness. To avoid choosing the common cold links, each FC decides the patrol path with its neighborhood FCs under the coordination of the TMC.*

VII. Evaluation

Here, we present the simulation results to evaluate the performance of the circuit patrol and the greedy patrol based on the real-world traffic data from the STIC, Shanghai, China.

A. Circuit Patrol

According to the circuit patrol, the FCs move in the Eulerian circuit that is predetermined according to the statistic of the links. We calculate the number of required FCs according to (20) so that the TMC can use MC-based estimation to recover the traffic matrix within a predescribed tolerant error. The simulation result is shown in Fig. 7(a).

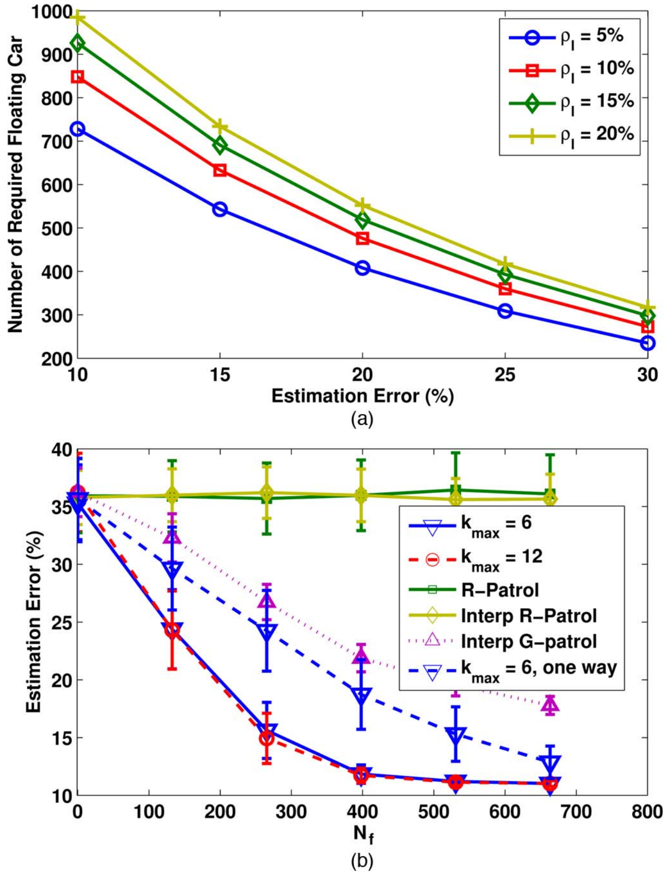

Fig. 7. (a) Number of required FCs of circuit patrol. (b) Estimation error versus number of FCs of greedy patrol.

For the integrity threshold ρ_l chosen as 5%, about 540 FCs are required for the whole city of Shanghai so that the estimation error is smaller than 15%. While only 230 FCs are required if the tolerant error is set as 30%. This is because when the tolerant error decreases, the sampling interval of a link decreases, and more FCs are required to guarantee the interval. For ρ_l to be increased to 20%, more links are labeled as cold links, and thus, the length of the Eulerian circuit increases. As a result, the numbers of required FCs increase to 730 and 310 for the tolerant error of 15% and 30%, respectively.

B. Greedy Patrol

In the greedy patrol, the movement of the FCs is reactive to the samples of PVs. They are expected to move to the links that PVs seldom pass. For simplicity, we choose the roads in a region whose integrity is 100% as the roads to be sampled and show the performance of greedy patrol by recovering the traffic matrix via the MC technique. The sampling matrix of PVs follows the uneven pattern shown in Fig. 2. N'_f FCs are deployed in the region to sense the traffic condition based on greedy patrol. Then, the number of FCs to monitor the whole city is $N_f = N'_f n_1/n_{\text{full}}$, where n_1 and n_{full} are the numbers of links in the city and the selected region, respectively. The estimation error and standard deviation is shown in Fig. 7(b).

Comparing the performance of greedy patrol with respect to different $k_{\max}$, it shows that $k_{\max}$ does not have much of an impact for $k_{\max} > 6$. It means that we only need to estimate the idlenss of the links by 30 min in the future. The simulation result also implies that we can achieve a good estimation with 15% error by controlling about 260 FCs. The performance of greedy patrol is better than circuit patrol because the corresponding routes are reactive to the PVs' samples although they cannot guarantee the same intervals between any two consecutive sampled duty cycles. To estimate the required FCs in the circuit patrol, we assume that the ATEs of the hot links are $\bar{h}_l$, although they are normally smaller than $\bar{h}_l$. Hence, the required estimation error can be satisfied although the number of FCs is actually smaller than the theoretical number.

To show the scalability of the proposed greedy patrol for different road topologies, we change the bidirectional roads of Shanghai to one-way roads. The estimation result has been shown in Fig. 7(b). It is observed that the estimation error becomes larger. This can be explained that for one-way roads, the FCs need to go across more roads to travel from one cold link to another, which reduces the sampling rate of the idle links. It is also shown that the greedy patrol algorithm benefits from the bidirectional-road topology.

Furthermore, we use the spline interpolation method to estimate the traffic matrix for comparison study. The resulting error shown in Fig. 7(b) is 5% more than that by the MC-based algorithm. Moreover, the deviation by the MC-based algorithm is smaller than that of interpolation. Compared with the interpolation method, fewer FCs are required for the MC algorithm to achieve the same estimation error, which demonstrates the advantage of the MC-based algorithm.

The performance comparison of greedy patrol and random patrol (R-Patrol) is shown in Fig. 7(b). It can be seen that for both MC algorithm and interpolation, there is no obvious improvement even for more FCs if they just randomly patrol. It proves that improvement in the estimation error from 35% to 12% is due to the increase in evenness rather than the number of FCs. On the other hand, the average entropy with greedy patrol control decreases from about 4.9×10^4 to 4.8×10^3 as the number of FCs increases from 0 to 650. However, the average entropy by random patrol remains nearly the same, although the number of FCs increases. It has then demonstrated that the traffic estimation benefits from the reduction of entropy for a better sampling process of patrolling control of FCs.

C. Further Discussion

For circuit patrol, FCs need to sample the traffic conditions of cold links with equal intervals. However, a cold link does not imply light traffic there. It means that the link is seldom passed or traveled by PVs, and its traffic condition is unknown. Consequently, the speed of traveling across cold links may be different. Thus, it may be difficult to guarantee that these links are sampled by the FCs with exact same intervals because of the unknown and varying traffic. We discuss the impact of traffic variation on estimation error by using circuit patrol as follows.

For the case that the cold links are with light traffic, FCs can try to keep almost equal travel time from one cold link to the

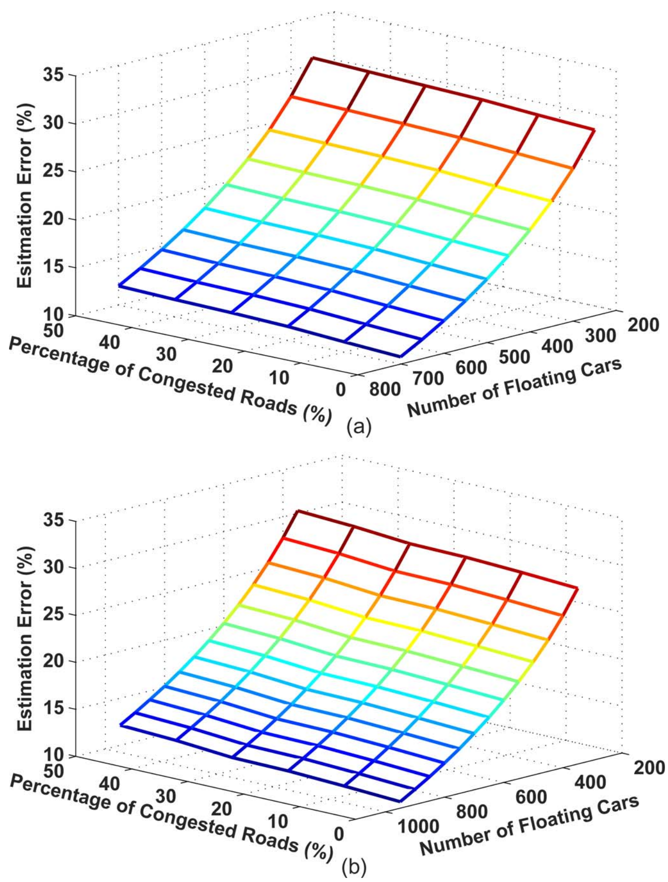

Fig. 8. Number of required FCs of circuit patrol with congested roads, sometimes with (a) $\rho_l = 5\%$ and (b) $\rho_l = 20\%$.

next by regulating their speed. In fact, the time granularity of estimation, i.e., ΔT, is set to be 5 min in our study. All the sensing reports uploaded in one duty cycle are considered as sampled at the same time. Suppose that the expected time difference between two consecutive FCs to sample the same cold link is Δt_s. It is acceptable as long as the real time difference, i.e., Δt_r, satisfies $|\Delta t_r - \Delta t_s| \leq 0.5\Delta T$. Consequently, the circuit patrol is robust against small variations in traffic conditions due to reasons such as the regulation of traffic lights.

For the case where the cold links are with heavy traffic, we test the effectiveness of circuit patrol under various traffic conditions. We randomly select 10%–50% of the cold links and reduce their average traversal speed during daytime by 20%. The result is shown in Fig. 8. The threshold of cold links in the circuit patrol, i.e., ρ_l, is 5% in Fig. 8(a) and 20% in Fig. 8(b). With $\rho_l = 5\%$, the estimation error slightly increases from 10.74% to 11.99% when the percentage of congested roads increases from 0 to about 50% with 700 FCs. For $\rho_l = 20\%$, the estimation error increases from 10.69% to 12.15% when 950 FCs are deployed. It has been shown that the estimation performance does not degrade much. Consequently, the circuit patrol is robust against traffic variation in cold links. If severe traffic congestion happens in cold links, FCs could change their predetermined routes by reporting the real-time traffic condition to the TMC and receiving guidance from the TMC. Under this circumstance, the cold links have turned into hot links. It is noted that Δt_s could also be adaptive according to the real-time traffic and the updated circuit routes.

As for greedy patrol, the time intervals of sampling cold links are not exactly the same due to traffic variation. However, we can achieve a good estimation result since the differences in time intervals are only one or two duty cycles from simulation. It is seen that the greedy patrol is adaptive to traffic variation in nature.

It implies from the discussion that the traffic estimation performance would not be deteriorated greatly if the variation in sampling time of cold links are not too large by the proposed patrol strategies.

VIII. Conclusion

In this paper, we have proposed patrol control algorithms to improve the performance of MC-based urban traffic monitoring. From the analysis of real-world traffic data from Shanghai, it has been shown that the performance of MC depends on the sampling process, and the uneven distribution of PVs leads to the uneven sampling of traffic matrix. Therefore, we explore the relationship between the estimation performance and the sampling process from the perspective of entropy. To reduce the average entropy of samples, two path-planning algorithms, namely, circuit patrol and greedy patrol, have been proposed to control the movement of the FCs, which proactively participate in traffic sensing. The circuit patrol algorithm preplans circuit paths for FCs that contain all the cold links (with low traffic), whereas the greedy patrol algorithm enables the FCs reactively to travel according to the distribution of the obtained samples. Simulation results have shown that our algorithms outperform the random patrol in terms of estimation error. By replanning the paths of about 400 FCs for the whole city of Shanghai, the estimation error of greedy patrol reduces from 35% to 12%.

For our future research, the distribution of PVs is to be explored in depth from the perspective of social pattern based on real traces of PVs (buses and taxis in this paper). Geographically, the mobility routes of buses are constant; thus, the average speed of the buses can be used to estimate the spatial distribution of traffic in the regions along the buses' routes. Meanwhile, taxis have restricted mobility regions around some specific social spots (e.g., business centers and financial centers) so that the spatial distribution of traffic in the regions sampled by taxis may follow a specific power law from the hot spots toward the border of the mobility regions. Taking this social pattern into account, the traffic estimation method proposed in this paper can be improved by more accurate spatial entropy of traffic samples. Another possible direction is to study traffic monitoring based on multiple sources of traffic samples, e.g., VSN, static traffic cameras, and smartphones. Then, a new relationship between estimation error and sampling process may be explored and then used to provide guidance on camera deployment. Moreover, studying the performance of the patrol algorithm with different road topologies might be useful in improving the strategies. It is also very interesting to incentive private vehicles to participate in traffic sensing by smartphones so that the traffic conditions in the whole city can be estimated in this efficient and economical way.

REFERENCES

[1] D. Schrank, B. Eisele, and T. Lomax, "TTI's 2012 Urban Mobility Report," Texas A&M Univ. System, College Station, TX, USA, 2012.

[2] I. Leontiadis *et al.*, "On the effectiveness of an opportunistic traffic management system for vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1537–1548, Dec. 2011.

[3] F. Knorr, D. Baselt, M. Schreckenberg, and M. Mauve, "Reducing traffic jams via VANETs," *IEEE Trans. Veh. Technol.*, vol. 61, no. 8, pp. 3490–3498, Oct. 2012.

[4] N. Cheng, "Opportunistic spectrum access for CR-VANETs: A game theoretic approach," *IEEE Trans. Veh. Technol.*, vol. 63, no. 1, pp. 237–251, Jan. 2013.

[5] U. Lee, E. Magistretti, M. Gerla, P. Bellavista, and A. Corradi, "Dissemination and harvesting of urban data using vehicular sensing platforms," *IEEE Trans. Veh. Technol.*, vol. 58, no. 2, pp. 882–901, Feb. 2009.

[6] R. Jiang, Y. Zhu, H. Wang, M. Gao, and L. M. Ni, "Compressive data retrieval with tunable accuracy in vehicular sensor networks," in *Proc. Wireless Algorithms, Syst., Appl.*, 2013, pp. 426–439.

[7] H. Zhu *et al.*, "Impact of traffic influxes: Revealing exponential intercontact time in urban VANETs," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 8, pp. 1258–1266, Aug. 2011.

[8] Z. Li, Y. Zhu, H. Zhu, and M. Li, "Compressive sensing approach to urban traffic sensing," in *Proc. IEEE ICDCS*, Minneapolis, MN, USA, Jun. 20–24, 2011, pp. 889–898.

[9] E. J. Candés and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, Dec. 2009.

[10] R. Du, C. Chen, B. Yang, and X. Guan, "Vanet based traffic estimation: A matrix completion approach," in *Proc. IEEE Global Commun. Conf.*, Atlanta, GA, USA, Dec. 9–13, 2013, pp. 52–57.

[11] B. Liu, O. Dousse, P. Nain, and D. Towsley, "Dynamic coverage of mobile sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 2, pp. 301–311, 2013.

[12] S. He, J. Chen, X. Li, X. Shen, and Y. Sun, "Mobility and intruder prior knowledge improving the barrier coverage of sensor networks," *IEEE Trans. Mobile Computing.* [Online]. Available: http://www.computer.org/csdl/trans/tm/preprint/06627887.pdf

[13] Y. Wang, F. Wu, and Y. Tseng, "Mobility management algorithms and applications for mobile sensor networks," *Wireless Commun. Mobile Comput.*, vol. 12, no. 1, pp. 7–21, 2012.

[14] M. Di Francesco, S. K. Das, and G. Anastasi, "Data collection in wireless sensor networks with mobile elements: A survey," *ACM Trans. Sensor Netw.*, vol. 8, no. 1, pp. 7:1–7:31, 2011.

[15] M. Ma, Y. Yang, and M. Zhao, "Tour planning for mobile data-gathering mechanisms in wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1472–1483, May 2013.

[16] M. Jerbi, S.-M. Senouci, T. Rasheed, and Y. Ghamri-Doudane, "An infrastructure-free traffic information system for vehicular networks," in *Proc. 66th IEEE VTC Fall*, Baltimore, MD, USA, Sep. 30–Oct. 3, 2007, pp. 2086–2090.

[17] L. Garelli, C. Casetti, C. F. Chiasserini, and M. Fiore, "Mobsampling: V2V communications for traffic density estimation," in *Proc. 73rd IEEE VTC Spring*, May 15–18, 2011, pp. 1–5.

[18] S. Panichpapiboon and W. Pattara-Atikom, "Exploiting wireless communication in vehicle density estimation," *IEEE Trans. Veh. Technol.*, vol. 60, no. 6, pp. 2742–2751, Jul. 2011.

[19] C. Luo, F. Wu, J. Sun, and C. W. Chen, "Compressive data gathering for large-scale wireless sensor networks," in *Proc. 15th Annu. Int. Conf. Mobicom*, Beijing, China, Sep. 20–25, 2009, pp. 145–156.

[20] L. Xu, X. Qi, Y. Wang, and T. Moscibroda, "Efficient data gathering using compressed sparse functions," in *Proc. IEEE INFOCOM*, Turin, Italy, Apr. 14–19, 2013, pp. 310–314.

[21] J. Cheng, Q. Ye, H. Jiang, D. Wang, and C. Wang, "STCDG: An efficient data gathering algorithm based on matrix completion for wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 850–861, Feb. 2013.

[22] C. Liu, C. Chigan, and C. Gao, "Compressive sensing based data collection in VANETs," in *Proc. IEEE WCNC*, Shanghai, China, Apr. 7–10, 2013, pp. 1774–1779.

[23] H. Wang, Y. Zhu, and Q. Zhang, "Compressive sening based monitoring with vehicular networks," in *Proc. IEEE INFOCOM*, Turin, Italy, Apr. 14–19, 2013, pp. 2923–2931.

[24] Y. Chevaleyre, "Theoretical analysis of the multi-agent patrolling problem," in *Proc. IEEE/WIC/ACM Int. Conf. IAT*, Beijing, China, Sep. 20–24, 2004, pp. 302–308.

[25] D. Portugal and R. Rocha, "MSP algorithm: Multi-robot patrolling based on territory allocation using balanced graph partitioning," in *Proc. ACM SAC*, Sierre, Switzerland, Mar. 22–26, 2010, pp. 1271–1276.

[26] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, "VeMail: A message handling system towards efficient transportation management," in *Proc. IEEE WCNC*, Shanghai, China, Apr. 7–10, 2013, pp. 4689–4693.

[27] N. Lu, T. Luan, M. Wang, X. Shen, and F. Bai, "Bounds of asymptotic performance limits of social-proximity vehicular networks," *IEEE/ACM Trans. Network.*, vol. 22, no. 3, pp. 812–825, Jun. 2014.

[28] N. Lu and X. Shen, *Capacity Analysis of Vehicular Communication Networks*. New York, NY, USA: Springer-Verlag, 2013.

[29] R. Rajagopal and P. Varaiya, Health of Californias loop detector system, Univ. of California, Berkeley, CA, USA, Tech. Rep., PATH, 2007. [Online]. Available: http://www.its.berkeley.edu/publications/UCB/2007/PRR/UCB-ITS-PRR-2007-13.pdf

[30] C. Y. Ma, D. K. Yau, N. K. Yip, and N. S. Rao, "Privacy vulnerability of published anonymous mobility traces," in *Proc. 16th ACM Annu. Inf. Conf. Mobicom*, Chicago, IL, USA, Sep. 20–24, 2010, pp. 185–196.

[31] Z. Zhou, R. Du, C. Chen, B. Yang, and X. Guan, "Packtrix: From data packing to reconstruction for a sensor network based structural health monitoring system," in *Proc. IEEE Global Commun. Conf.*, Atlanta, GA, USA, Dec. 9–13, 2013, pp. 472–477.

[32] Y. Elmaliach, N. Agmon, and G. A. Kaminka, "Multi-robot area patrol under frequency constraints," *Ann. Math. Artif. Intell.*, vol. 57, no. 3/4, pp. 293–320, 2009.

Rong Du (S'12) received the B.Sc. and M.Sc. degrees from Shanghai Jiao Tong University, Shanghai, China, in 2011 and 2014, respectively. He is currently working toward the Ph.D. degree with the Automatic Control Laboratory, School of Electrical Engineering, Royal Institute of Technology (KTH), Stockholm, Sweden.

His research interests include vehicular sensor networks, wireless sensor networks, and compressive sensing.

Cailian Chen (S'03–M'06) received the B.Eng. and M.Eng. degrees in automatic control from Yanshan University, Qinhuangdao, China, in 2000 and 2002, respectively, and the Ph.D. degree in control and systems from the City University of Hong Kong, Kowloon, Hong Kong, in 2006.

In 2008, she joined the Department of Automation, Shanghai Jiao Tong University, Shanghai, China, where she is currently a Full Professor. She has authored and/or coauthored two research monographs and over 100 referred international journal and conference papers in her areas of interest. Her current research interests include Internet of Vehicles and applications, wireless sensor and actuator networks, and cyberphysical systems.

Dr. Chen was one of the First Prize Winners of a Natural Science Award from the Ministry of Education of China in 2007. She received the IEEE TRANSACTIONS ON FUZZY SYSTEMS Outstanding Paper Award in 2008. She is on the editorial board of *Peer-to-Peer Networking and Applications* (Springer), *The Scientific World Journal: Computer Science* (Hindawi), and ISRN *Sensor Networks*. She serves as a Technical Program Committee Member for various conferences. She is a New Century Excellent Talent in University of China honored by the Ministry of Education of China, a "Pujiang Scholar," and a "Young Scientific Rising Star" honored by the Science and Technology Commission of Shanghai Municipal, China.

Bo Yang (M'09) received the Ph.D. degree in electrical engineering from the City University of Hong Kong, Kowloon, Hong Kong, in 2009.

Prior to joining Shanghai Jiao Tong University, Shanghai, China, as an Assistant Professor in 2010, he was a Postdoctoral Researcher with the Royal Institute of Technology (KTH), Stockholm, Sweden, from 2009 to 2010 and a Visiting Scholar with the Polytechnic Institute, Brooklyn, NY, USA, in 2007. He is currently an Associate Professor with Shanghai Jiao Tong University. He is also a Cyber Scholar with the Cyber Joint Innovation Center founded by Zhejiang University, Hangzhou, China; Tsinghua University, Beijing, China; and Shanghai Jiao Tong University. His current research interests include game theoretical analysis and nonlinear optimization of communication networks and smart grids.

Dr. Yang is on the editorial board of *Digital Signal Processing* (Elsevier) and on the Technical Program Committee of several international conferences. He has been a Principal Investigator/Coinvestigator on several research projects funded by the National Science Foundation of China, the Swedish Governmental Agency for Innovation Systems, and the United States Air Force.

Ning Lu (S'12) received the B.Sc. and M.Sc. degrees from Tongji University, Shanghai, China, in 2007 and 2010, respectively. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada.

His current research interests include capacity and delay analysis, media access control, and routing protocol design for vehicular networks.

Mr. Lu served as a Technical Program Committee Member for the IEEE 2012 International Symposium on Personal, Indoor, and Mobile Radio Communications.

Xinping Guan (M'02–SM'04) received the Ph.D. degree in control and systems from Harbin Institute of Technology, Harbin, China, in 1999.

In 2007, he joined the Department of Automation, Shanghai Jiao Tong University, Shanghai, China, where he is currently a Distinguished University Professor, the Vice Director of the University Research Management Office, and the Director of the Key Laboratory of Systems Control and Information Processing, Ministry of Education of China. Before that, he was a Professor and the Dean of Electrical Engineering, Yanshan University, Qinhuangdao, China. He has authored and/or coauthored three research monographs, more than 160 papers in IEEE TRANSACTIONS and other peer-reviewed journals, and numerous conference papers. His current research interests include cyberphysical systems, wireless networking and applications in smart cities and smart factories, and underwater sensor networks.

Dr. Guan has finished/been working on many national key projects as a Principal Investigator. He is the leader of the prestigious Innovative Research Team of the National Natural Science Foundation of China. He is a committee member of the Chinese Automation Association Council and the Chinese Artificial Intelligence Association Council. He is on the editorial board of the IEEE TRANSACTIONS ON SYSTEM, MAN, AND CYBERNETICS—PART C and several Chinese journals. He also serves as an IPC/TPC Member of various conferences. He received the First Prize from the Natural Science Award of the Ministry of Education of China in 2006 and Second Prize from the National Natural Science Award of China in 2008. He received the IEEE TRANSACTIONS ON FUZZY SYSTEMS Outstanding Paper Award in 2008. He is a "National Outstanding Youth" of the National Natural Science Foundation of China, a "Changjiang Scholar" of the Ministry of Education of China, and a "State-level Scholar" of the "New Century Bai Qianwan Talent Program" of China.

Xuemin (Sherman) Shen (M'97–SM'02–F'09) received the B.Sc. degree from Dalian Maritime University, Dalian, China, in 1982 and the M.Sc. and Ph.D. degrees from Rutgers University, New Brunswick, NJ, USA, in 1987 and 1990, all in electrical engineering.

He is a Professor and the University Research Chair with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. From 2004 to 2008, he was an Associate Chair for Graduate Studies. He is a coauthor/editor of six books and has published more than 600 papers and book chapters in wireless communications and networks, control, and filtering. His research focuses on resource management in interconnected wireless/wired networks, wireless network security, social networks, smart grid, and vehicular *ad hoc* and sensor networks.

Dr. Shen served as the Technical Program Committee Chair/Cochair for IEEE Infocom'14 and IEEE VTC'10 Fall, the Symposia Chair for IEEE ICC'10, the Tutorial Chair for IEEE VTC'11 Spring and IEEE ICC'08, the Technical Program Committee Chair for IEEE Globecom'07, the General Cochair for Chinacom'07 and QShine'06, and the Chair for the IEEE Communications Society Technical Committee on Wireless Communications and Peer-to-Peer Communications and Networking. He also serves/served as the Editor-in-Chief for the *IEEE Network, Peer-to-Peer Networking and Application, and IET Communications;* a Founding Area Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS; an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *Computer Networks*, and ACM *Wireless Networks*; and a Guest Editor for the IEEE JOURNAL ON SELECTED AREAS ON COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS, IEEE COMMUNICATIONS MAGAZINE, and the ACM *Mobile Networks and Applications*. He received the Excellent Graduate Supervision Award in 2006; the Outstanding Performance Award in 2004, 2007, and 2010 from the University of Waterloo; the Premiers Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada; and the Distinguished Performance Award in 2002 and 2007 from the Faculty of Engineering, University of Waterloo. He is a Registered Professional Engineer of Ontario, Canada. He is a Fellow of the Engineering Institute of Canada and the Canadian Academy of Engineering. He is a Distinguished Lecturer of the IEEE Vehicular Technology and the Communications Societies.