RISC: Resource-Constrained Urban Sensing Task Scheduling Based on Commercial Fleets

XIAOYANG XIE, Rutgers University, USA
ZHIHAN FANG, Rutgers University, USA
YANG WANG, University of Science and Technology of China, China
FAN ZHANG*, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China
DESHENG ZHANG, Rutgers University, USA

With the trend of vehicles becoming increasingly connected and potentially autonomous, vehicles are being equipped with rich sensing and communication devices. Various vehicular services based on shared real-time sensor data of vehicles from a fleet have been proposed to improve the urban efficiency, e.g., HD-live map, and traffic accident recovery. However, due to the high cost of data uploading (e.g., monthly fees for a cellular network), it would be impractical to make all well-equipped vehicles to upload real-time sensor data constantly. To better utilize these limited uploading resources and achieve an optimal road segment sensing coverage, we present a real-time sensing task scheduling framework, i.e., RISC, for Resource-Constraint modeling for urban sensing by scheduling sensing tasks of commercial vehicles with sensors based on the predictability of vehicles' mobility patterns. In particular, we utilize the commercial vehicles, including taxicabs, buses, and logistics trucks as mobile sensors to sense urban phenomena, e.g., traffic, by using the equipped vehicular sensors, e.g., dash-cam, lidar, automotive radar, etc. We implement RISC on a Chinese city Shenzhen with one-month real-world data from (i) a taxi fleet with 14 thousand vehicles; (ii) a bus fleet with 13 thousand vehicles; (iii) a truck fleet with 4 thousand vehicles. Further, we design an application, i.e., track suspect vehicles (e.g., hit-and-run vehicles), to evaluate the performance of RISC on the urban sensing aspect based on the data from a regular vehicle (i.e., personal car) fleet with 11 thousand vehicles. The evaluation results show that compared to the state-of-the-art solutions, we improved sensing coverage (i.e., the number of road segments covered by sensing vehicles) by 10% on average.

CCS Concepts: • **Networks** → **Sensor networks**; • **Information systems** → *Location based services.*

Additional Key Words and Phrases: Heterogeneous Fleets, Vehicle Sensing, Mobility Patterns

ACM Reference Format:
Xiaoyang Xie, Zhihan Fang, Yang Wang, Fan Zhang, and Desheng Zhang. 2020. RISC: Resource-Constrained Urban Sensing Task Scheduling Based on Commercial Fleets. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 2, Article 62 (June 2020), 20 pages. https://doi.org/10.1145/3397337

*Also with Beidou Intelligent Technology Co., Ltd., Shenzhen, China

Authors' addresses: Xiaoyang Xie, Rutgers University, New Brunswick, New Jersey, USA, xx88@cs.rutgers.edu; Zhihan Fang, Rutgers University, New Brunswick, New Jersey, USA, zhihan.fang@cs.rutgers.edu; Yang Wang, University of Science and Technology of China, Hefei, Anhui, China, angyan@ustc.edu.cn; Fan Zhang, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, China, zhangfan@siat.ac.cn; Desheng Zhang, Rutgers University, New Brunswick, New Jersey, USA, desheng.zhang@cs.rutgers.edu.

2474-9567/2020/6-ART62 $15.00
https://doi.org/10.1145/3397337

1 INTRODUCTION

There were more than 2 billion vehicles on the road by 2019 [4], and it is well expected that this increasing trend will keep going in the next 20 years[1]. For example, in some largest cities in the world, vehicles have been increasing significantly, e.g., in Beijing, the vehicles have been increasing from 3.50 million to 5.64 million in last ten years [7]; in Los Angels, the vehicles have been rising from 5.86 million to 6.49 million in the same period [2]. Such a surge of vehicles leads to various urban challenges, e.g., traffic jams, prolonged commuting time, tailpipe emission, traffic accidents, etc. To address these challenges, many efforts have been made from both technical aspects (e.g., real-time ride-sharing) and policy aspects (e.g., congestion fees in Singapore). In these efforts, a critical fundamental approach is real-time vehicular sensing[27]. Vehicular sensing utilizes the sensors equipped in the vehicles, e.g., dash-cam, lidar, automotive radar, RFID, etc., to sense some urban phenomenon[11, 12], e.g., the traffic speed, the traffic flow, the locations of the vehicles on the road, etc, to various applications, e.g., autonomous driving, anomaly detection, etc.

Given the importance of urban sensing, many techniques have been developed and implemented to model large-scale vehicular mobility patterns with various sensing devices[36, 39, 40]. For example, Wang *et al.* proposed an approach to calculate possible locations and time distribution of the hit-and-run vehicles in parallel [33] based on camera networks. However, given the stationary features of these sensing solutions, e.g., cameras or loop sensors, or the limited number of equipped cars[37], they cannot achieve urban-scale real-time vehicular sensing.

In this paper, we argue that there are tens of thousands of potential mobile sensors moving across cities around 24 hours a day on almost all the road segments, including commercial vehicles, e.g., taxis, buses, and trucks. More importantly, given recent efforts of smart transportation, many cities have their commercial vehicles equipped with sensing (e.g., GPS and cameras) and communication devices (e.g., Dedicated Short-Range Communications DSRC and cellular devices with a monthly service plan) for the management purposes, e.g., uploading vehicle sensing data to operating centers for security and accounting in real time. These new efforts provided us with an unprecedented opportunity to achieve a mobile sensing solution in real time at the urban scale, which includes, but not limited to, vehicle to vehicle sensing, i.e., sense other vehicles.

However, it is unclear if commercial vehicles can accomplish this objective sufficiently since their mobility patterns might be biased against their services (e.g., buses with a fixed route). More importantly, their sensing data uploading capabilities are limited since their current uploading purposes were designed for simple low data-rate tasks, e.g., reporting GPS locations for accounting [17], and occasionally uploading images for security [13].

The key question we want to explore is if it is possible to transparently use existing commercial vehicles (i.e., no changes to their existing mobility patterns) with existing data uploading infrastructures to enable real-time sensing to cover all road segments? We emphasize the keyword **transparently** because, in practice, it is difficult (i) to change mobility patterns of commercial vehicles just for vehicular sensing given their primary objective is to provide services (e.g., taxis picking up passengers, buses serving routes, trucks dropping off packages) and (ii) to ask for more resources in terms of a higher budget for real-time data uploading to perform such a task.

To answer this question, we conduct a case study in the most crowded city in China, i.e., Shenzhen, with 12 million population and 3.2 million personal vehicles. In this particular case, we use a combination of the taxi fleet, the bus fleet, and the truck fleet to provide urban-scale real-time vehicular sensing, by uploading sensing data from the selected commercial vehicles in real time to task scheduling centers to cover all road segments in a given city for complete urban sensing.

Without optimization, a straightforward solution is to constantly upload real-time sensing data from commercial vehicles with redundant information, e.g., the same sensing information is uploaded multiple times by vehicles on the same roads. However, in practice, every vehicle is uploading its data based on cellular communication with a fixed amount of data uploading, e.g., 1GB per car per month, so continuous uploading will quickly exceed the data uploading quota. More importantly, many commercial vehicles on the same road segments will upload the redundant sensing data. As a result, we design a vehicular sensing task scheduling framework RISC, which aims to achieve optimal large-scale Real-tIme Sensing from Commercial vehicle fleets with limited data uploading capability. In particular, RISC provides a real-time task scheduling assignment for commercial vehicles based on the predictability of their mobility patterns and achieves the real-time local optimal road segment coverage.

In particular, the contributions of the paper are as follows:

- To the best of our knowledge, we conduct the first study on the resource-constrained real-time vehicular sensing with heterogeneous fleets. In particular, our study is based on (i) a taxi fleet with 14 thousand vehicles, (ii) a bus fleet with 13 thousand vehicles, and (iii) a truck fleet with 4 thousand vehicles. Our mobility pattern study enables us to comprehensively compare these vehicle fleets in terms of real-time sensing capabilities and analyze their data for valuable urban mobility insights, which is difficult to be obtained by previous studies on a single fleet and their data.
- We design a vehicular sensing task scheduling system called RISC to quantify the mobility patterns of various commercial vehicles with a unified framework and schedule their vehicular sensing under their sensing and uploading constraints. Theoretically, we formulate our urban-scale vehicular sensing as a Markov Decision Process (MDP) problem and design an online framework to schedule dynamic sensing data uploading in real time. In the standard MDP, the transition matrix is calculated based on historical data, and the impact of future actions is typically not considered, and the rewards are defined by empirical observations. Instead, we design a convolutional neural network model to predict the future locations of vehicles, and dynamically incorporate the future actions in the state transmission. Our rewards are also dynamically updated based on the current distribution of vehicles on roads. Such a combination of the MDP and CNN components has not been considered by the previous work in a vehicular sensing scenario.
- We implement our RISC system in Shenzhen based on one-month of detailed data from three commercial fleets. Further, based on RISC we design a potential application to track suspect vehicles, e.g., hit-and-run vehicles, in real time. More importantly, we use a separate dataset from 11 thousand regular vehicles as the ground truth to evaluate this application. We rigorously evaluate the impacts of various factors (e.g., the sensing area, the limitation of uploading capability, the sensing angle of sensors, the sensing radius of sensors) on the performance of our sensing system. The evaluation results show that compared to a state-of-the-art solution, we improved sensing coverage (i.e., number of road segments covered by sensing vehicles) by 10% on average. In addition, our study results reveal some valuable insights for transparent urban vehicle sensing with commercial vehicles. Based on these insights, we provide a few lessons learned, which have the potential to offer some guidance for some real-world applications based on urban-scale real-time vehicular sensing.

2 RELATED WORK

We organize the current vehicular sensing work into four categories from on two aspects in Table 1. **(1)** The first aspect includes *participatory sensing* and *opportunistic sensing*, which is divided based on how the sensor participates during the sensing. In particular, *participatory sensing* methods sense a target by dispatching mobile sensors, e.g., vehicles, while *opportunistic sensing* methods sense a target without changing the vehicles' mobility pattern. In particular, many systems install sensors in the vehicles themselves to sense the mobility of the equipped vehicles are considered opportunistic sensing since this kind of works does not change vehicle traces. **(2)** The second aspect is the diversity of mobility patterns of the vehicles involved in the sensing, i.e., *homogeneous sensing* and *heterogeneous sensing*. *Homogeneous sensing* means these sensing systems only utilize the vehicles belonging to the same types of vehicles, e.g., bus, taxi, truck, which has similar mobility pattern features, while *heterogeneous sensing* indicates these sensing systems utilize the vehicles from multiple vehicles types, which contains diverse mobility patterns.

Table 1. Categories of Related Work

	Participatory	Opportunistic
Homogeneous	[19, 37, 48]	[10, 15, 24, 46]
Heterogeneous	[20, 43, 45, 49]	[34],RISC

2.1 Participatory Sensing

2.1.1 Homogeneous Sensing. Participatory sensing has been widely studied by many researchers, but most of them are based on the homogeneous fleets. [19] proposes a system to recover the situation of traffic accidents based on the shared dash-cam data from the vehicles, with the preservation of privacy. [48] design a bus travel time estimation system based on the interaction between the bus and the mobile phone. [37] proposes a dispatching system to dispatch taxis to obtain optimal sensing coverage with a limited budget.

2.1.2 Heterogeneous Sensing. [20] provides a greedy optimal heterogeneous vehicle selection algorithm to collect comprehensive spatial-temporal sensing data. [44] infers traffic conditions by utilizing two heterogeneous nationwide vehicles based on real-world contexts and multi-view learning. [45] develops a multi-view learning framework to iterative obtain mutually-reinforced knowledge for real-time human mobility at the urban scale based on a massive dataset including data about the taxi, bus, and subway passengers along with cellphone users. [49] design a crowdsensing method that selectively chooses heterogeneous sensors from participators to collect data. However, the participatory sensing might cause a heavy burden from the participation, e.g., dispatching commercial vehicles, or provide an incentive mechanism to participators, which leads to a higher cost than opportunistic ones.

2.2 Opportunistic Sensing

2.2.1 Homogeneous Sensing. There is also some existing work on opportunistic sensing. Most of them only utilize the homogeneous fleet. [46] utilizes taxis for urban sensing based on a minimum number of selected vehicles with the guarantee of specific coverage quality requirements. [15] deploys air quality monitoring nodes on city buses to monitor the air quality of a city. [10] provides a systematic study of three large-scale data sets of taxi GPS traces to understand spatial patterns in vehicle motions and how such patterns can support information dissemination. [24] presents a study of the instantaneous topology of the vehicular network in Cologne, Germany, and it also unveils the underlying structure of the vehicular network in this city. But, these only consider one homogeneous fleet, so it is hard for them to consider various mobility patterns of different fleets.

2.2.2 Heterogeneous Sensing. Some works utilize different fleets for urban sensing, which improves the capability of sensing. [34] infers and predicts the trace of regular vehicles based on the interactions between the heterogeneous fleets without any constraints. Different from previous work, RISC utilizes the equipped sensing devices from the heterogeneous fleets under the constraint of data uploading capability and addresses the bias problem caused by the specific mobility patterns of the homogeneous fleet and the expensive cost of real-time data uploading problem. Besides, RISC utilizes the devices equipped in commercial vehicles to preprocess the collected data, which reduces the burden of the cloud server. To our best knowledge, *RISC* is the first study on the resource-constrained real-time sensing feasibility for multiple generic fleets based on real-world heterogeneous fleets' data.

3 MOTIVATION

Based on a recent survey [3], the total number of connected vehicles is expected to be more than 200 million by 2025, leading a potentially huge market for Vehicles-to-Vehicles (V2V) applications and Vehicles-to-Infrastructure (V2I) applications for connected vehicles[32]. In some of the connected vehicle applications, e.g., route selection, safety warning, or assistant driving[28], real-time data uploading are necessary due to the short-latency requirement of these applications. For example, real-time route selection or safety warning based on traffics is dependent on the freshness of the traffic sensing data, and some outdated data could provide a wrong route or miss important warnings. Therefore in this section, we investigate the opportunities of utilizing commercial vehicles for vehicular sensing and the challenge of uploading all the real-time sensor data from commercial vehicles.

3.1 Opportunities for Vehicular Sensing

Most city-scale urban sensing applications assume a dense sensor distribution and to full coverage of road network sensing. However, the sparsity of stationary road sensing camera distribution can hardly be avoided in real scenarios, due to the high deployment overheads and the dynamic nature of urban road networks. For instance, in Shenzhen, a leading city in deploying urban sensing systems, there are 463 intersections in the industrial park, while only 3.2% intersections are equipped with traffic sensing systems. In contrast, commercial fleets potentially provide a mobile sensing opportunity. To explore the capability of the commercial fleets on the city-scale urban sensing, we show the percentage of the road segments sensed by at least one vehicle from a particular fleet in 1-minute slots in Figure 1. We show the coverage of using each fleet as sensors compared with the heterogeneous fleets. We found that these three commercial fleets have their particular features of coverage of road segments. For example, (i) the taxi fleet has the highest coverage compared with the bus fleet and the truck fleet, (ii) the difference on the coverage between daytime and nighttime for bus fleets is larger than that of other two fleets, (iii) the truck fleet has the lowest coverage compared to others, and it is also lower than the traffic sensing systems. Compared to the heterogeneous fleets and the traffic sensing systems, we found that heterogeneous fleets improve the coverage nearly four times the traffic sensing systems.

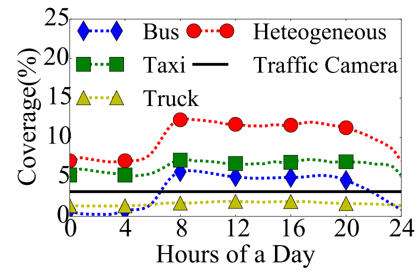


Fig. 1. Road Coverage

3.2 Challenge of Sensor Data Uploading

In the real-world, many commercial vehicles are passing the same regions, which will create a lot of duplicated real-time sensor data if all uploaded, increasing the burden of the cloud server and generating unnecessary cost, i.e., the duplicated sensor data collected by different commercial vehicles which will make the cloud server to process and store redundant data, in the data uploading.

To quantify this drawback, we study how many commercial vehicles will be in the same region in a one-minute slot and use extra costs as a metric. The extra cost indicates the percentage of duplicated data uploaded in real time if asking all the commercial vehicles to upload their sensor data. The extra cost is calculated as $\frac{M}{N}$, where M is the total number of regions visited by the commercial vehicles including duplicated regions; N is the number of distinct regions visited by commercial vehicles. As shown in Figure 2, during the early morning, using all commercial vehicles together to upload data will generate about 110% of duplicated data since most of the commercial vehicles in the early morning are concentrated, which generate a large amount of

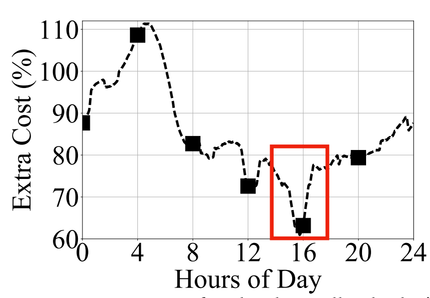


Fig. 2. Extra Cost of Uploading All vehicles' Data

duplicated data. Also, even in the lowest valley around 4 PM (highlighted in the box), this number would also be more than 60%. That is because, during the afternoon, commercial vehicles are distributed in the city more evenly, especially the taxis, which make less duplicated data.

3.3 Summary

We explore the sensing coverage of road segments by different fleets based on their spatiotemporal patterns. Further, we identify the unnecessary cost of uploading data for using all commercial fleets for sensing. The cost is in terms of data transmission cost. In a real-world setting, many commercial vehicles are passing through the same regions, which will cost a lot of duplicated real-time sensor data if all uploaded, increasing the burden of the cloud server and generating unnecessary transmission and processing cost such as cellular uploading fee, data storage, and processing cycles. As follows, we aim to address these issues to reduce the cost of urban vehicle sensing. It motivates us to design a system to provide an efficient assignment for real-time vehicular sensing. In the real-world, many commercial vehicles are passing the same regions, which will cost a lot of duplicated real-time sensor data if all uploaded, increasing the burden of the cloud server and generating unnecessary cost, i.e., the duplicated sensor data collected by different commercial vehicles which will make the cloud server to process and store redundant data, in the data uploading.

4 VEHICULAR DATASETS

	Length	# of Vehicles	Daily Records	Total Size
Taxi	one month	14K	65M	38.5GB
Bus		13K	44M	41.3 GB
Truck		4K	7M	7.9 GB
Regular Vehicle		11K	13M	21.8 GB

Format				
ID	Date&Time	GPS	Speed	Direction

Fig. 3. Fleets and Their Data

We get access to the vehicular system datasets provided by the Shenzhen Committee of Transportation (SCT), with which we collaborate for better urban transportation management. As in Figure 3, we introduce four types of vehicle fleets in Shenzhen, i.e., a taxi fleet, a bus fleet, a regular vehicle fleet, and a truck fleet. For each fleet, we show the length of days, the number of vehicles, the daily uploaded records, and the total size of the data. Please note that these datasets are the basic GPS data collected by the onboard devices, while the data that RISC aims to collect and schedule the uploading are higher volume of sensor data from more abundant and diverse sensors to detect environments. For example, RISC could be utilized to collect the image data from the dashcam, which cannot be fully supported by existing real-time uploading infrastructures for cellular networks within budget. It motivates us to design a resource-constrained sensing task scheduling approach.

- **Taxi Fleet:** The taxi fleet in Shenzhen contains over 14 thousand taxis where each taxi uploads its record every 30 seconds in real time, including GPS locations, time, speed, etc. The taxi fleet has random mobility patterns, and most of them have a frequent stopping place.
- **Bus Fleet:** The bus fleet in Shenzhen contains 13 thousand buses where each bus uploads record every 30 seconds in real time, including GPS locations, time, speed, etc. In contrast to the random mobility patterns of the taxi fleet, the buses have regular mobility patterns, and they have routine routes between the two terminals every day.

- **Truck Fleet:** We leverage a truck fleet's data from a logistics company in Shenzhen. All trucks upload their data every 15 seconds to the company's servers for real-time monitoring. The trucks have semi-random mobility patterns, and most of them would stay in their company if they are not delivering cargos and goods.
- **Regular vehicle Fleet:** The regular vehicle data are used as the ground truth in our evaluation, but not in our design. We access the data of the regular vehicle fleet through an insurance company, where regular vehicles upload their status to a server of the insurance company for reducing their insurance fee. This fleet in Shenzhen contains 11 thousand vehicles where each vehicle generates one record every 10 seconds. Regular vehicles have semi-random mobility patterns, and most of them commute between their homes and offices.

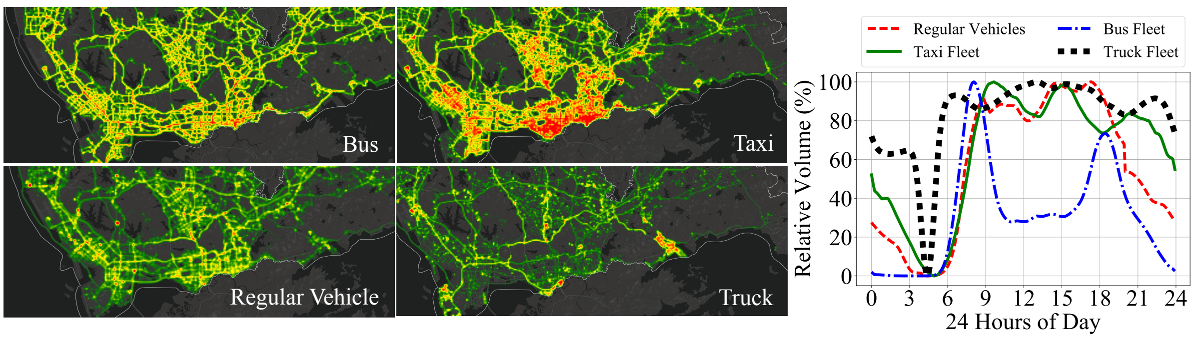


Fig. 4. Fleet Visualization

Our fleet access with heterogeneous mobility patterns enables fine-grained urban traffic sensing as shown in Figure 3. Given the diverse mobility pattern, each of fleets should provide some unique mobility coverage. For example, Figure 4 gives a heatmap visualization of these four fleets based on their one-day data and the relative volume of them in each hour in a day. We found that each fleet has its unique mobility pattern shown by the circles, e.g., (i) the truck fleet mostly focus on highways and a few industrial areas; (ii) the taxi fleet covers most urban areas; (iii) the bus fleet focuses on trunk road segments; (iv) the private fleet has similar patterns with taxis but with some exceptions at a few residential areas. Besides, from the volume of these four fleets, we found that there are more taxis and trucks during the late night and early morning than that of buses and regular vehicles; whereas there are more regular vehicles and buses during rush hours. Those insights motivate us to design a better real-time sensing scheduling approach under a resource-constrained scenario.

Entropy: Entropy is regarded as the most fundamental quantity capturing the degree of predictability [29][8]. For each vehicle, we extract its origin and destination in each trip then obtain its a sequence of visited locations L. We calculate the entropy S as $S = -\sum_{L_i'} p(L_i') log_2(p(L_i'))$, where L_i' is a sub-sequence of L and $p(L_i')$ is the probability of L_i' appearing in L. We use entropy to measure the randomness of the trip origins and destinations. The cumulative distribution is in Fig. 5. The regular vehicles have the lowest randomness on the origin and destination pairs in the four fleets. This is because the regular vehicles move between several most frequent locations, which has been proved by previous works [18]. Compared with taxis and buses, the randomness of the truck is lower. One possible reason is that the trucks are operated by logistics companies and move among warehouses intra-cities or

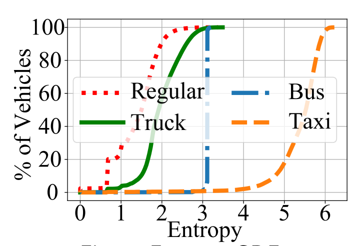


Fig. 5. Entropy CDF

inter-cities. Buses have a constant route arrangement, and the number of stations of different buses is similar, which leads to close entropy around 3.

Top-N locations: Human's mobility can be generally modeled as the movement between several essential locations, such as home, office, etc [18]. To determine how much of the mobility is rooted in the visitation patterns of the top locations, we calculated the probability that the vehicle is in one of the top n most visited locations in a given moment. Based on the mobility data, we study the number of locations a vehicle passes on every weekday, which is defined as the number of daily locations. Figure 6 presents the number of daily distribution in four fleets. In particular, for the bus fleet, there are more buses with entropies smaller than or equal to 2. The possible reason is that some bus lines may provide the express shuttle services from the airport or the train station to the downtown area[5], e.g., line 330B or 330C only has two terminals. Due to the high demand for this kind of service, there may be more buses to serve these lines. We found most vehicles have one or two most frequently visited stay locations, which may be home, work locations, or the logistic centers. For the bus fleet, buses have a fixed route arrangement. Therefore, it shows a more significant number of stay locations, which are mostly bus stations. In short, based on the entropy and the Top-N location analysis in Fig. 5 and 6, we found even though the numbers of important visited locations are similar in heterogeneous vehicular fleets (as shown in Top-N location distribution), their mobility patterns differ significantly (as shown in the entropy analysis). Therefore, these diverse (potentially complementary) mobility patterns enable heterogeneous vehicular fleets to have better spatial coverage on urban sensing compared with a single vehicular fleet.

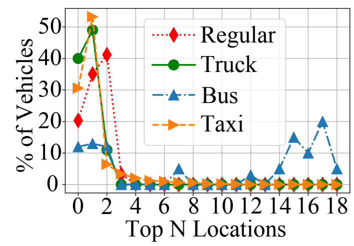


Fig. 6. Daily Locations

5 REAL-TIME SENSING TASK SCHEDULING DESIGN

In this section, we present the framework of RISC for real-time sensing task scheduling. RISC generates an optimal assignment based on Markov Decision Process (MDP) to assign a subset of the commercial vehicles to be the "mobile sensors" in given time duration (e.g., 5 mins) to upload their data captured by their local sensors to a cloud server based on the estimated rewards, i.e., the number of covered road segments of their sensing data. The higher the coverage, the higher the reward.

5.1 Scheduling Framework

Overview: Given the real-world constraints of commercial vehicle sensing and communication capability, the objective of this paper is to select minimal commercial vehicles, i.e., an assignment, to cover as many road segments as possible for urban traffic sensing in a given duration with the limitation of data uploading budget, e.g., 1GB per month. Due to the dynamic features of human mobility, Markov Decision Process is especially suitable for vehicle dispatching and scheduling since a vehicle's current location is highly dependent on its locations of previous time slots [41]. MDP has been successfully used in many problems such as stochastic planning problems and game playing problems. Actually, MDP has been a popular solution for sequential decision problems due to its internal design structure [38][35]. In our setting, the decisions for future actions are not related to the previous states, and commercial vehicles are highly dynamic. Thus, MDP is appropriate for such a stochastic decision process problem in our setting. On the other hand, the computation time for solving MDP models is much shorter compared with solving Markov models [30]. This is crucial for complicated problems with many states and actions such as urban-scale vehicular sensing. Due to the unpredictable events in cities, RISC improves the MDP framework to execute the first optimal assignment in the current slot but considers future actions at the same time. We formulate this problem as a Markov Decision Process problem and use a policy iteration algorithm [22] (PIA) to choose an optimal subset of commercial vehicles for sensing. In our framework, a policy is a sequence of assignments. We divide time into different slots. For the one-time slot, we have one corresponding bipartite graph where the vertices of one side are commercial vehicles; the vertices of the other side are the road segments (an example is given in Figure 7). In one slot, if a commercial vehicle passes some road segments, the vertex of this vehicle would have edges connecting the vertices for these road segments. Each slot many several optimal assignments with the same maximum reward. However, since each commercial vehicle has limited data uploading capability, and some of them will visit the road segments solely in the future time slot, which leads to a potentially higher future reward, e.g., a taxi going to a remote area where no other commercial vehicles nearby. Therefore, when to choose this kind of commercial vehicle with high future reward will impact the total spatial-temporal coverage in the long run, i.e., global optimum. As a result, the goal of RISC is to find a series of assignments in a given duration to achieve a global optimal spatial and temporal coverage, instead of obtaining the local optimal spatial and temporal coverage. We show our framework in Figure 7.

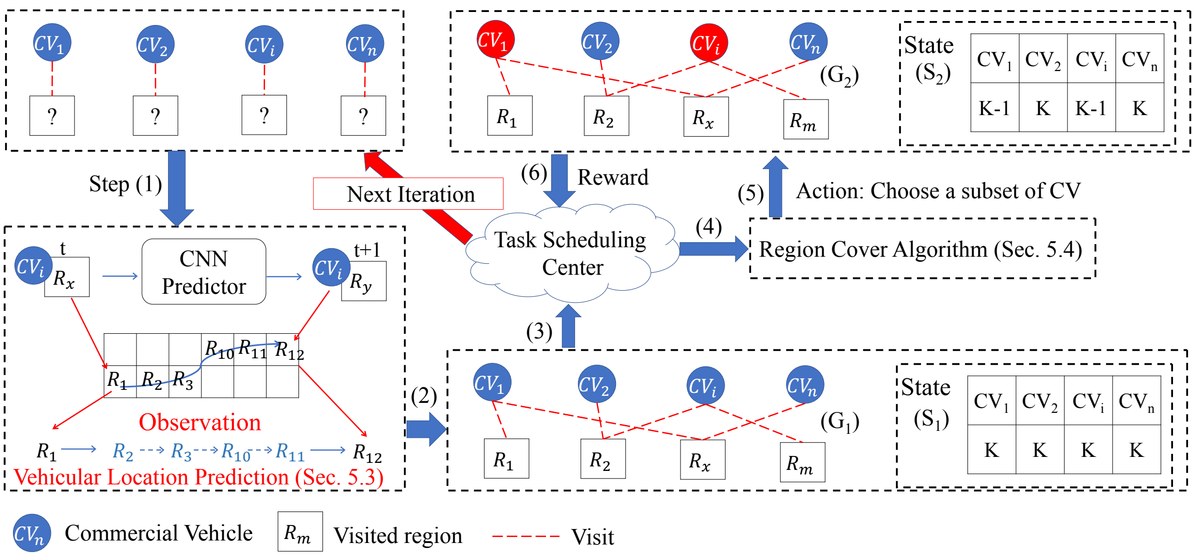


Fig. 7. RISC Framework

Step (1)&(2): As shown in the step (thick arrow) (1) and step (2) of Figure 7, in the initial iteration, based on the current and previous locations of all commercial vehicles (i.e., from CV_1 to CV_n), RISC predicts their future locations by utilizing the CNN-based (i.e., Convolutional Neural Network) location prediction model. The detail of this location prediction model is introduced in section 5.2. Based on the prediction locations of all commercial vehicles, we have the possible visit road segment of these vehicles. This relationship between the commercial vehicles and their possible visit road segments in the near future is represented as a bipartite graph, such as the example shown in the bipartite graph G_1 in the right bottom box of Figure 7. In the example, CV_1 will visit road segment R_1 and R_x based on the predicted location.

Step (3)&(4): We define the data uploading capability of all the commercial vehicles as a state, which is quantified by the remaining available time this vehicle can be used as a sensor. The state tables in Figure 7 gives two example of a state, i.e., S_1 (the table in the bottom right) and S_2 (the table in the top right). As shown in the examples, the initial capability of a commercial vehicle is K, and it will be reduced by 1 when the vehicle is assigned as a sensor in a time slot by an assignment. In step (3) and step (4), based on the bipartite graph and the state S_1, the task scheduling center utilizes a vertex cover algorithm to obtain a set of possible assignments that choose which subset of CV nodes to cover all R nodes (all the road segments) as many as possible in the bipartite graph G_1. The details of the vertex cover algorithm are given in section 5.3.

Step (5): We define an action a is an assignment that chooses one possible minimal subset obtained from the vertex cover algorithm. In each time slot, for each candidate vehicle we exclude it in the vertex cover algorithm and obtain one candidate action. We define A as the set of all the candidate actions. In step (5), RISC chooses an action from A for sensing, i.e., commercial vehicles CV_1 and CV_i in G_2. In the meantime, the state is updated, which is S_2 in the example.

Step (6): In step 6, RISC gets the reward by the selected action and feeds to the Bellman equation 1. We define a reward function F as $F(s_i, a_i, s_i') = N_{a_i}$, where N_{a_i} is the number of newly covered road segments based on the action a_i; s_i and s_i' is the current state and the next state after action a_i selected (i.e., an assignment). After step (6), RISC will go to the next iteration. In the next iteration of this action, the commercial vehicles will start at the predicted locations. In one iteration, there are multiple actions. RISC will try all possible actions, and get different rewards and different next state. In other words, one action in an iteration will create a branch of the next iteration, and in the next iteration, it is possible that actions will create their branch of iterations. RISC will explore all the possible sequences of actions and find out the sequence with the highest reward. This set of iterations will continue for K times (where K is set to 10) to simulate the future dynamics and then stop.
Based on the following Bellman equation, the goal of our MDP is to obtain a policy π that satisfies that

$$V^{\pi}(s) = \max_{a \in A}[\sum_{s' \in S} T(s, a, s')(F(s, a, s') + \lambda V^{*}(s'))] \tag{1}$$

where s is the current state; s' is the possible state after action a; $V^{*}(s')$ is the optimal reward in the next iteration of state s'; V^{π} is the final optimal reward in this time slot; π is the optimal sequence actions. T is the function that indicates the probability that state s transfers to state s' by action a. $\lambda \in [0, 1]$ is the discount factor, indicating the difference in the importance of future rewards and current rewards. In our setting, λ is the average predictability of assigned commercial vehicles obtained from the predictability estimation. The detail of predictability estimation is shown in section 5.4 With this formulation, we apply PIA to obtain π. In particular, in each iteration, the commercial vehicles will not get assigned for sensing in practice. RISC simulates their movements in the future K time slots based on the CNN-based location prediction model. It had been proved that MDP can make the optimal policy (assignment) [30] under the deterministic environment, e.g., the correct predicted future locations of the commercial vehicles. Therefore, based on an assumption that the prediction of locations for all commercial vehicles is correct in these K time slots, RISC obtains the optimal sequence of assignments for future K time slots.

Remarks: Real-Time Online Aspects on Iterative Scheduling: The MDP scheduling Framework utilizes the vehicular real-time location prediction model and the vertex cover algorithm to obtain the action space and the reward function. However, due to the dynamic nature of commercial vehicles and unpredictable events in the city, it is challenging to predict their future locations without errors for the next K slots. Therefore, in the current slot, to guarantee the optimal sensing coverage in an online fashion, RISC only executes the first optimal assignment for the current slow, and even it has a sequence of assignments for the next K slots based on the simulation. In the next slot, RISC will start this whole process again based on the updated locations of commercial vehicles in real time to obtain a sequence of assignments for the following K slots, but still only executes the first assignment of this newly obtained optimal sequence of assignments.

5.2 Vehicular Location Prediction

We utilize a CNN-based model to predict vehicular mobility patterns by the offline training based on the historical location observations of a candidate commercial vehicle that could be able to be used for urban sensing. We then predict its next location based on the current location and the previous locations. We formulate this vehicular next location prediction problem as a state prediction problem in deep learning and utilize a CNN-based model[25] to solve this problem. Since the commercial vehicles are required to upload their GPS information to the cloud around every 30 seconds, RISC could obtain each commercial vehicle's previous locations and current locations. In our setting, we define a spatial-temporal record as one state. For the spatial granularity, we divide Shenzhen into 1200×600 grids, and each of them is $100\text{m} \times 100\text{m}$. For the temporal granularity, we divide one day into 1,440-time slots where each time slot is 1 minute. With this setting, for a commercial vehicle, RISC utilizes a k-length sequence of location records from R_{t-k} to R_t as the input of the CNN model, where k is the number of the previous k time slots and t is the current time slot. With this input, the CNN model predicts the location R_{t+1}, which is the next location of the commercial vehicle in the next future time slot.

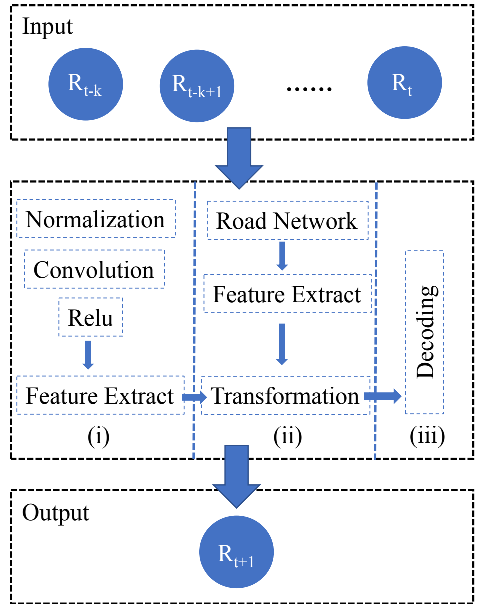


Fig. 8. CNN Framework

Fig. 8 shows the design of our CNN network with three phases. **(i) Encoding Phase**: this phase contains one normalization layer and four convolution layers. In our setting, we represent one spatiotemporal record of a vehicle as a vector that includes the location of the vehicle, the time of day, the day of the week, the weather, of the collected record. One vector of a vehicle is defined as the state of the vehicle in the slots. **(ii) Transformation Phase**: this phase contains three conditional transformation layers. Accepting the encoded features of the previous input states and the extracted features of the road network, it generates a high-level feature for the prediction of the next state. **(iii) Decoding Phase**: this phase contains two fully-connected layers. It predicts the next state and then maps the predicted high-level features to a spatial-temporal vector. With the predicted next location of the commercial vehicles, it can infer which CV nodes will connect to R nodes.

5.3 Vertex Cover Algorithm

Given the predicted location of commercial vehicles, this component aims to obtain the possible minimal subsets of commercial vehicles to cover maximal road segments covered by using all the commercial vehicles. As shown at the beginning of this section, in one slot, we have a bipartite graph, in which the vertices include two sides, i.e., the commercial vehicles and the road segments. In the example at the bottom of Figure 7, we show an example of the vertex cover problem. In this bipartite graph, there are four vertices in the vehicle side and four vertices in the road segment side. The objective of the sensing tasks assignment algorithm is to find the possible minimal subsets that have edges connecting to all vertices in the road segment side, namely the subset containing CV_1 and CV_i in the table of the example. This could be regarded as a minimal set cover problem. In this formulation, we define a road segment as r and all the possible passed road segments as the universe U of the set cover problem where $U = r_1, r_2, \ldots, r_m$, and one vehicle as one set s whose elements are the road segment it will pass. We denote the set of all the commercial vehicles as S, where $S = s_1, s_2, \ldots, s_n$. We utilize a state-of-the-art method[16] to solve this classic minimal set cover problem.

5.4 Vehicular Predictability Estimation

We estimate the predictability of the commercial vehicles based on their historical traces. Based on the historical GPS data of a commercial vehicle, we first divide them into different trips by utilizing a stay point detection algorithm [14], where each trip has an origin and a destination. Then we cluster them into different patterns by utilizing a density-based clustering [47] to obtain a set of clusters as with different mobility patterns, based on the origin-destination information and the temporal information of these trips. For the historical GPS data of each cluster, we apply the CNN-based location prediction on these locations and compute the average accuracy of prediction for each cluster. In the real-time setting, we classify the observed partial traces of a commercial vehicle into one of its clustered patterns based on the distances between traces[9] and use the accuracy of the specific cluster as the predictability used for λ in MDP.

6 FIELD STUDY

China Mobile	
Fee	Volume
50 RMB	1 GB
70 RMB	2 GB
100 RMB	3 GB
130 RMB	4 GB
180 RMB	6 GB
280 RMB	11 GB

Fig. 9. Mobile Fee

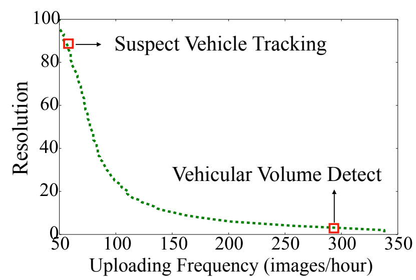


Fig. 10. Tradeoff

We introduce a preliminary test to verify that if the volume of real-time uploading sensor data is acceptable for the cellular premium plans. We set up a dashcam in a vehicle and collect the image data while driving on the road. The dashcam is a 1080p HD car dashboard camera HP F500G with 32GB storage and 24 FPS. During the test, the HP F500G dashcam collects video data while the car is active for a few days. The data collected by the dashcam during one hour is about 5GB on average. This is a significant volume, which exceeds the limited volume of many cellular premium plans. As shown in Figure 9, in Shenzhen, the basic plan of China mobile is 1 GB per month with a 50 RMB monthly fee. Even the ultimate plan provides 11 GB data, and it can only support two hours of data uploading. As a result, if all the commercial vehicles (more than 31 thousand vehicles) are used as mobile sensors, it would cost 31,000 dollars to upload one-hour sensing data. On the other hand, although for some cellular service providers in some countries, e.g., AT&T in the US, the 4G data is unlimited (but the uploading speed will be lower after the initial 5GB), 5GB per hour per vehicles is still a big burden for the storage in the cloud. To enable the real-time collected data uploading, some techniques could be applied, e.g., data compression [26] or a lower data uploading frequency. We study the trade-off between the compressed data resolution and the frequency of data uploading with a given limited volume in one hour and show the curve in Figure 10. In our study, we assume the vehicles use the cellular premium plan of China Mobile, i.e., 1 GB per month, and limit the total duration of data uploading in one day is 1 hour, leading 34 MB per hour in the test. From the result, we found that, if we compress the data with high resolution, e.g., 95, in one hour the vehicle can only upload about 60 images, whereas if we compress the data with low resolution, e.g., 5, then within one hour the vehicle can upload 300 images. The choice of the compression resolution and data uploading frequency depends on the application based on the uploaded data. For example, based on the collected high-resolution images, the plate and appearance of most of the vehicles could be recognized by some computer vision algorithms, which could be used for the suspect vehicle tracking. For the low-resolution images, even the plates of vehicles are hard to be recognized due to the loss of resolution, the number of vehicles still could be identified, which could be used for the vehicular volume count.

7 EVALUATION

To evaluate the performance of RISC in the urban traffic sensing, we compare RISC with a state-of-the-art baseline on a metric in terms of coverage of road segment with changing some factors.

7.1 Evaluation Methodology

7.1.1 Evaluation Setting.

- **Baselines:** (1) We implement a greedy method, DEGR[23], from the multi-skill resource-constrained project scheduling problem, which regards the sensing coverage of road segments as the task, the commercial vehicles as the resources, and the sequence of the visited road segments as the skills of the commercial vehicles. (2) We also implement a crowdsensing algorithm HSF-ST[49], which aims to maximize the sensing coverage and data utility. In our setting, the data utility of a commercial vehicle is replaced with the data uploading capability, which is calculated as $\frac{M}{N} \times K$, where M is the available uploading capability (How much data it can upload under the constraint) of the commercial vehicle; N is the total uploading capability of the commercial vehicle; K is the number of road segments visited by the commercial vehicle. HSF-ST chooses commercial vehicles with higher coverage and higher data uploading capability first. We use DEGR and HSF-ST as the baselines and compare RISC with them in the evaluation.
- **Metrics:** As the description in section 5.2, we divide Shenzhen into a 1200×600 grid and map the road network into this grid. Given a time slot, let R be the set of the cells the road network mapped to, R_v be the set of the cells the assigned commercial vehicles passed, the coverage of road segment is $\frac{|R_V|}{|R|}$. In our simulation, we do not consider the physical structure of the road, e.g., the number of lanes in the road, etc. When a vehicle passes a mapped cell, we assume the cell is covered. In our simulation, we only use the spatial-temporal information of the commercial vehicles from GPS records to simulate sensing activities instead of performing the actual sensing. When an assigned commercial vehicle is located at a covered cell, we assume this vehicle collects traffic data by its sensors in this cell. A few existing studies are utilizing similar approaches to simulate sensing activities such as [49][6][21]. As a result, we use the coverage of the cells with collected traffic data to measure the performance of our sensing scheduling. The higher coverage means better performance.
- **Factors:** To investigate the performance of RISC in different situations, we evaluate RISC with different factors, including the time of a day (ToD), the evaluating areas, and the limitation of uploading capability. Also, we implement RISC with single fleets to compare the performance on sensing for different fleets.

7.2 Evaluation Result

7.2.1 Economical Efficiency. To study costs reduced by the RISC, we compare the number of vehicles used for sensing by RISC and the number of all the active commercial vehicles of each time slot in Figure 11 and use the ratio between these two numbers as a metric. We found that the implementation of RISC in the whole city of Shenzhen reduces about 50% of vehicles for sensing on average, while that of the implementation of RISC on the CDB (central business district) region of Shenzhen is nearly 75% on average. This is because, in the CBD region, the resource of the public transportation system is much more sufficient than that in the suburban area. The results show that RISC has a significant impact on the reduction in the burden of urban-scale vehicular sensing, especially in the downtown area.

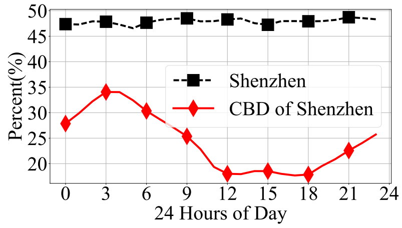


Fig. 11. Reduced Cost

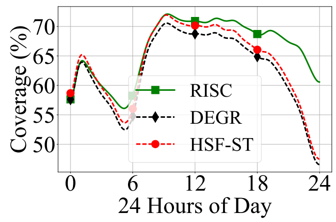


Fig. 12. RISC in Urban

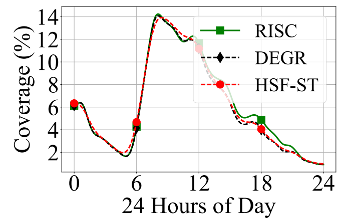


Fig. 13. RISC in Suburban

7.2.2 *Impact of the Time of Day.* We implement RISC, DEGR, and HSF-ST with the data of the Futian District of Shenzhen, i.e., the urban area in Shenzhen, and show the resultant Figure 12. We found that RISC outperforms DEGR and HSF-ST in most time slots, especially in the evening. This is because DEGR and HSF-ST only consider the optimal assignment in the current time slot. They may reduce the capability of data uploading of some commercial vehicles. The reduced capability of these commercial vehicles becomes essential when they pass some unique road segments that can be only covered by them. This leads to lower coverage in the later time slots. We also found that there exists a peak from 12:00 AM to 2:00 AM due to the operation of a late-night bus route. Compared with DEGR and HSF-ST, RISC has a better performance in the urban region and has been less impacted by the time of the day because of its future consideration of the data uploading capability.

7.2.3 *Impact of Sensing Areas.* To evaluate the performance of RISC in the suburban area, we implement RISC, DEGR, and HSF-ST with the data in the Baoan District of Shenzhen. Figure 13 shows that all of them have a significant performance decrease in the later time slots. This is because, in the suburban area, the number of available commercial vehicles is much smaller than that in the urban area. Even RISC considers the capability of the commercial vehicles in the future time slot, the small number of commercial vehicles limits its performance. Nevertheless, RISC has a better performance than DEGR and HSF-ST.

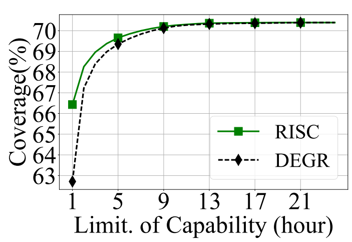


Fig. 14. Limit. of Cap.

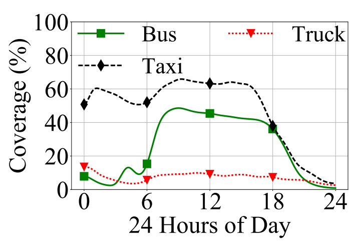


Fig. 15. Impact of Fleets

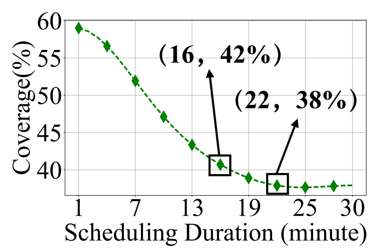


Fig. 16. Robustness

7.2.4 *Impact of Limitation of Uploading Capability.* The limitation of the data uploading duration is an essential factor for both RISC and DEGR. With a longer uploading capability, the commercial vehicles could be assigned as the sensors for more times. To study the impact of this limitation, we implement RISC and DEGR on data in Futian District, changing the limitation from 1 hour to 24 hours. The resultant Figure 14 shows RISC is better than DEGR. With the increase of the data uploading duration, the difference between the two methods decreases. Especially after 11 hours, there is almost no difference. This is because, with the increase of uploading capability, the drawback of DEGR is not evident since those commercial vehicles who pass the unique road segments that can be only covered by them in the later time slots can still upload their sensor data.

7.2.5 *Impact of Fleets.* To explore the performance of RISC on homogeneous fleets and compare the performances of different commercial fleets on traffic sense, we implement RISC on different individual fleets, and show the result in Figure 15. The result shows RISC-taxi has a better performance than the RISC-bus and RISC-truck due to its larger number of vehicles and more random mobility patterns. Also, since trucks are concentrated on the highway and the truck, the coverage of RISC-truck is the least one. Besides, we found in the evening, and the coverage drops significantly. This is because most of the vehicles are out of the capability of data uploading. Compared the performance of RISC implemented with the homogeneous fleet and the heterogeneous fleets, we found that the heterogeneous method is better than the homogeneous method since an individual fleet might be concentrated on some specific road segments, which might cause bias in the sensing task.

7.2.6 Robustness. In the previous evaluation, RISC obtains an assignment to assign sensing tasks to commercial vehicles every minute based on the predicted locations of commercial vehicles within 1 minute. However, the duration of the scheduling is a significant factor for the robustness of RISC. To explore robustness, we implement RISC by utilizing the predicted locations within a duration, changing from 1 minute to 30 minutes, as shown in Figure 16. We found that the average coverage decreases when the duration increases. However, the decrease becomes smaller after 16 minutes, and finally, after 22 minutes, the coverage is kept at around 38%. This is because, within the scheduling duration, RISC predicts a sequence of commercial vehicles in the future duration. From this predicted location information, RISC obtains the optimal local assignment. Therefore, even RISC causes some errors in the prediction of locations of the commercial vehicles, RISC still guarantees real-time sensing coverage.

8 APPLICATION: SUSPECT VEHICLE TRACKING

Based on RISC, there are some potential applications to improve urban efficiency and security, e.g., (i) the suspect vehicle tracking, which infers a trace of a specific vehicle, e.g., hit-and-run vehicles, (ii) the estimation of travel time between two regions by utilizing the regular vehicles captured by the commercial vehicles in two regions, (iii) the real-time traffic situation monitor, etc. To verify the performance of RISC on those applications, we take the suspect vehicle tracking as an example, and we evaluate it with the real regular vehicle GPS data in Shenzhen in terms of two metrics introduced in the follows. We envision there exists a sensor in commercial vehicles that could be used to detect the nearby vehicles, e.g., all commercial vehicles in Shenzhen have been equipped with dash cams, which have been used to capture front-view for security and insurance purposes. We utilize this kind of commercial vehicles as mobile sensors for vehicular sensing.

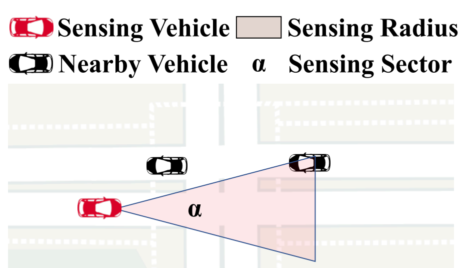


Fig. 17. Sensing Scenario

8.1 Sensing Scenario

We give an example of sensing scenarios in Figure 17. In the example, the sensing vehicle is equipped with a dash-cam with 60 to 120 degrees, which would capture the vehicles in front of it. As Figure 17 shows, due to the limitation of sensing angles, the sensing vehicle on the right side would only sense a black vehicle and cannot capture another closer vehicle if a 60-degree camera was used. This sensing method is transparent to the sensing vehicles. Note that we use the dash-cam with 60-120 degrees (dependent on cameras) as the example of sensor-based sensing, but other sensing devices in these two kinds of methods would have different configurations. For example, the 360 degrees lidar would also probe the nearby vehicles with 360 degrees. Therefore, to make our sensing model more generic, we build a unified framework for vehicular sensing. Given a vehicle equipped with sensing devices, we quantify its sensing capability with the following two metrics.

Sensing Mileage Rate: Given a trip of an equipped vehicle, let d_t be its driving distance during time slot t, I_t be its indicator in time slot i where $I_i = t$ if there exists other vehicles nearby and $I_t = 0$ if not. The Sensing Mileage Rate is $\frac{\sum_t I_t d_t}{\sum_t d_t}$. In the rest of this paper, we use **mileage** for short of Sensing Mileage Rate.
Sensing Gap: Given a trip of an equipped vehicle, if there are no vehicles close to it during a sequence of time slots, this period is called a sensing gap. We use the length of the time slots as the metric to measure the sensing gap. In the rest of this paper, we use **gap** for Sensing Gap.

To evaluate the performance of the suspect vehicle tracking application with the two metrics, i.e., mileage and the gap, we implement RISC with the changing of some essential factors, including the time of day, the angle of sensing sectors, the radius of sensing sectors, and the sensing fleets. In the rest of the evaluation, we implement RISC and DEGR in the Futian District of Shenzhen.

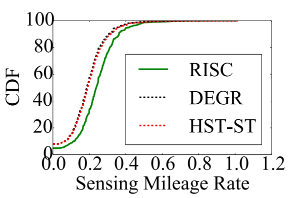


Fig. 18. Sensing Mileage

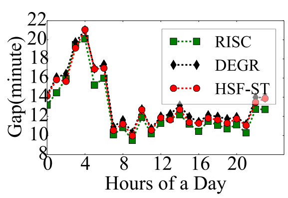


Fig. 19. Sensing Gap

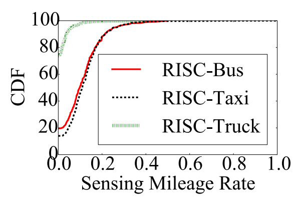


Fig. 20. Top-N Fleet on Mileage

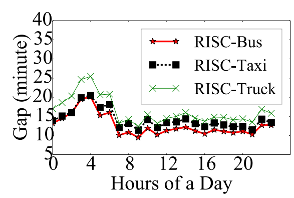


Fig. 21. Top-N Fleet on Gap

8.2 Sensing Mileage Rate

We evaluate RISC on the performance on the sensing mileage rate (defined previously) by comparing it to DEGR and HSF-ST. Figure 18 shows the results of these methods. In this evaluation, we set the sensing radius is 100 meters, and the sensing angle is 360 degrees. In Figure 18, we show the CDF of mileage for each system, and the lower curve has a better performance. Based on the results, we found that RISC has a better performance than DEGR and HSF-ST. In particular, more than 70% of regular vehicles could be captured with a probability larger than 20% by using RISC. Compared with RISC, both DEGR, half of the regular vehicles are captured with a probability of more than 20%, as same as HSF-ST.

8.3 Sensing Gaps

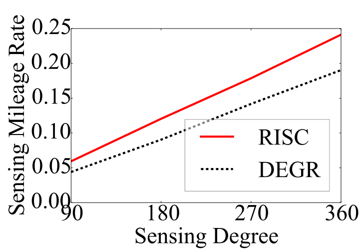
We also evaluate RISC on the performance of the gap by comparing it to DEGR and HSF-ST. We show the performance on the gap of these systems over 24 hours in a day in Figure 19 where a lower curve has a better performance. Compared with RISC, DEGR and HSF-ST have larger gaps on average. In particular, the gap in implementing DEGR is 1 minute longer than that of using RISC, especially during the early morning and the evening. This might be because, during those hours, the number of commercial vehicles is less than other times, using the DEGR and HSF-ST method might drain out the budget of uploading of some vehicles quickly, which cannot be used in the later time. Therefore, on the performance of the gap, RISC still has a better performance than DEGR and HSF-ST with the same configuration of commercial sensing vehicles.

8.4 Performances of Top-N Fleets

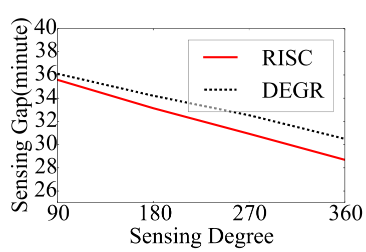
To study the impact of different fleets on the performance of urban vehicular sensing, we implement RISC with the three commercial fleets individually, i.e., RISC-Bus, RISC-Taxi, and RISC-Truck. We evaluate these three systems on the two metrics and show their performances in Figure 20 and Figure 21. In Figure 20 about the mileage performance, the lower the curve, the better the system. Taxi-RISC turns out to have the best performance among these systems, and it outperforms Bus-RISC slightly and Truck-RISC significantly. Figure 21 about the gap performance, the lower the curve, the better the system. Similarly, Taxi-RISC also has the best performance and Truck-RISC has a significant drawback compared with the other two systems. The difference is caused by the distinct mobility patterns of these three fleets. The bus fleet has fixed routes and is most predictable. However, due to the randomness of regular vehicles and the limitation of the size of buses, the bus fleet has a smaller mileage area than that of the taxi fleet. Different from the bus fleet, the taxi fleet has a more similar pattern with the regular vehicles and similar size with vehicles, leading to its best performance on the vehicular sensing. For the truck-RISC, due to its operating features, it will concentrate on the main roads, especially the highways. Therefore, except in some particular regions, e.g., the industrial regions, the truck fleet has the worst performance on sensing regular vehicles.

8.5 Impact of Factors

To study the impact of different factors on the performances of mileage and gap, we measure RISC with different configurations, including the degree of the sensing sectors, and the radius of the sensing sectors.
The degree of sensing sector: We envision a regular vehicle that could be detected by a commercial vehicle if the regular vehicle is in the sensing sector of the commercial vehicle, e.g., the regular vehicle is in the front of the dashcam of the commercial vehicle. The degree of the sector will impact the sensing area of the dashcam. To measure the impact of the sector degree, we implement RISC and DEGR with four choices of sector degrees, (i.e., 90, 180, 270, and 360) and show the results in Figure 22. In Figure 22a, the X-axis is the degree of the sector, and the Y-axis is the average mileage. We found that the sensing mileage of RISC changes from 25% to 5% when the degree of sectors changes from 360 degrees to 90 degrees. Similarly, the difference in mileage of DEGR from 360 degrees to 90 degrees is from 17% to 4%. In Figure 22b, the average gap of RISC with 360 degrees is shorter than 29 minutes; whereas that of RISC with 90 degrees is more than 35 minutes. From the results in Figure 22, we found that the degree of sensing sectors has a very limited effect on both RISC and DEGR.



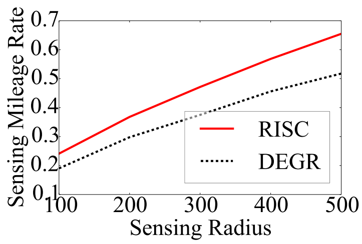
(a) Mileage



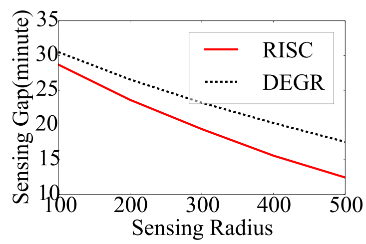
(b) Gap

Fig. 22. Impact of Sector Degree

The radius of the sensing sector: Another important factor is the radius of the sensing sector. A larger radius enables commercial vehicles to probe more vehicles. To evaluate the impact of the radius of the sensing sector on the performance of mileage and gap, we implement RISC and DEGR with five different radii, e.g., 100m (meters), 200m, 300m, 400m, and 500m. Figure 23a reveals that RISC has a more than 60% of mileage after the radius is larger than 400m. In addition, we found that, even with a 400m radius, the mileage of DEGR is still around 40%. In the performance of the gap, as shown in Figure 23b, changing the sensing radius from 100m to 500m reduces the gap around 15 minutes, where that of DEGR is 13 minutes.



(a) Mileage



(b) Gap

Fig. 23. Impact of Sensing Radius

9 DISCUSSION

Lessons Learned: As shown in the evaluation, with the real-time updated information, RISC utilizes less than 50% commercial vehicles for sensing to achieve similar real-time sensing coverage by utilizing all commercial vehicles. In addition, with a more frequent assignment updating rate, RISC is able to achieve higher sensing coverage. Since the sampling rate of the location data of commercial vehicles is around 20 seconds per record, it is possible to set the RISC with a high frequent information updating rate to achieve better sensing coverage.
Limitations: A major limitation of our RISC is to require a large number of commercial vehicles to be installed with some vehicular sensors, e.g., dash cameras, and have the data uploading capability. We believe this is a reasonable assumption since large cities have already had this kind of infrastructure installed due to accounting and security purposes, e.g., NYC [42], Beijing [31], Shenzhen.
Privacy Protection: While vehicle sensing in aggregation has the potential for significant social benefits, e.g., to reduce traffic jams and increase public security, we have to protect the privacy of drivers and vehicles involved. Benefited from the reduction of insurance premiums, the drivers of the regular vehicles consent to upload their data. All vehicular data analyzed are anonymized and hashed by service providers, and a randomized serial number replaces all IDs identifiable for a particular vehicle during our analyses.

10 CONCLUSION

In this paper, we conduct the first study on the resource-constrained vehicular sensing with heterogeneous fleets based on a large-scale multi-modal dataset. In particular, we design a vehicular sensing task scheduling system called RISC to quantify the mobility patterns of various commercial vehicles a unified framework and provide the optimal assignment to schedule their vehicular sensing under their sensing and communication constraints with a guarantee of real-time sensing coverage. We implement and evaluate our RISC system in Shenzhen based on one-month detailed data from three commercial fleets and design a potential application, i.e., suspect vehicle tracking, simulated with GPS data of a regular vehicle fleet. The evaluation shows that compared to two state-of-the-art solutions, we improve sensing coverage by 10% on average. More importantly, we provide a few insights based on our case study, which has the potential to offer some guidance for some real-world applications based on urban-scale real-time vehicular sensing.

ACKNOWLEDGMENTS

This work is partially supported by NSF 1849238 and 1932223. We thank the associate editor and reviewers for the detailed and insightful feedback for this work

REFERENCES

[1] 2019. *business insider*. https://www.businessinsider.com/global-transport-use-will-double-by-2040-as-china-and-india-gdp-balloon
[2] 2019. *California DMV*. https://www.dmv.ca.gov/portal/dmv
[3] 2019. *Here Mobility*. https://mobility.here.com/smart-city-car-connected-intelligent-integrated
[4] 2019. *Moter vehicle Wikipedia*. https://en.wikipedia.org/wiki/Motor_vehicle#cite_note-9
[5] 2019. Shenzhen Airport Shuttle Bus and City-Bus. (2019). https://www.travelchinaguide.com/cityguides/guangdong/shenzhen/airport-shuttle-bus.htm
[6] Fabian Bock, Yuri Attanasio, and Sergio Di Martino. 2017. Spatio-temporal road coverage of probe vehicles: a case study on crowd-sensing of parking availability with taxis. In *The Annual International Conference on Geographic Information Science*. Springer, 165–184.
[7] Beijing Statistical Bureau. 2017. Beijing statistical yearbook.
[8] Paul Y Cao, Gang Li, Adam C Champion, Dong Xuan, Steve Romig, and Wei Zhao. 2017. On human mobility predictability via wlan logs. In *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 1–9.
[9] Philippe Cudre-Mauroux, Eugene Wu, and Samuel Madden. 2010. Trajstore: An adaptive storage system for very large trajectory data sets. In *Data Engineering (ICDE), 2010 IEEE 26th International Conference on*. IEEE, 109–120.
[10] Jiaxin Ding, Jie Gao, and Hui Xiong. 2015. Understanding and modelling information dissemination patterns in vehicle-to-vehicle networks. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 41.
[11] Zhihan Fang, Boyang Fu, Zhou Qin, Fan Zhang, and Desheng Zhang. 2020. PrivateBus: Privacy Identification and Protection in Large-Scale Bus WiFi Systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–23.
[12] Zhihan Fang, Yu Yang, Shuai Wang, Boyang Fu, Zixing Song, Fan Zhang, and Desheng Zhang. 2019. MAC: Measuring the Impacts of Anomalies on Travel Time of Multiple Transportation Systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 2 (2019), 1–24.
[13] Raghu Ganti, Iqbal Mohomed, Ramya Raghavendra, and Anand Ranganathan. 2011. Analysis of data from a taxi cab participatory sensor network. In *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*. Springer, 197–208.
[14] Raghu Ganti, Iqbal Mohomed, Ramya Raghavendra, and Anand Ranganathan. 2012. Analysis of Data from a Taxi Cab Participatory Sensor Network *(Mobile and Ubiquitous Systems: Computing, Networking, and Services)*.
[15] Yi Gao, Wei Dong, Kai Guo, Xue Liu, Yuan Chen, Xiaojin Liu, Jiajun Bu, and Chun Chen. 2016. Mosaic: A low-cost mobile sensing system for urban air quality monitoring. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 1–9.
[16] Lukasz Golab, Flip Korn, Feng Li, Barna Saha, and Divesh Srivastava. 2015. Size-constrained weighted set cover. In *Data Engineering (ICDE), 2015 IEEE 31st International Conference on*. IEEE, 879–890.
[17] Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, and Ken Wood. 2006. SenseCam: A retrospective memory aid. In *International Conference on Ubiquitous Computing*. Springer, 177–193.
[18] Sibren Isaacman, Richard Becker, Ramón Cáceres, Margaret Martonosi, James Rowland, Alexander Varshavsky, and Walter Willinger. 2012. Human mobility modeling at metropolitan scales. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*. Acm, 239–252.
[19] Minho Kim, Jaemin Lim, Hyunwoo Yu, Kiyeon Kim, Younghoon Kim, and Suk-Bok Lee. 2017. ViewMap: Sharing Private In-Vehicle Dashcam Videos.. In *NSDI*. 163–176.
[20] Yazhi Liu, Jianwei Niu, and Xiting Liu. 2016. Comprehensive tempo-spatial data collection in crowd sensing using a heterogeneous sensing vehicle selection method. *Personal and Ubiquitous Computing* 20, 3 (2016), 397–411.
[21] Xufei Mao, Xiaohua Xu, ShaoJie Tang, and Xiang-Yang Li. 2012. Providing and finding k-road-coverage efficiently in wireless sensor networks. *Wireless Communications and Mobile Computing* 12, 12 (2012), 1053–1065.
[22] Sean P Meyn. 1997. The policy iteration algorithm for average reward Markov decision processes with general state space. *IEEE Trans. Automat. Control* 42, 12 (1997), 1663–1680.
[23] Paweł B Myszkowski, Łukasz P Olech, Maciej Laszczyk, and Marek E Skowroński. 2018. Hybrid differential evolution and greedy algorithm (DEGR) for solving multi-skill resource-constrained project scheduling problem. *Applied Soft Computing* 62 (2018), 1–14.
[24] Diala Naboulsi and Marco Fiore. 2013. On the instantaneous topology of a large-scale urban vehicular network: the cologne case. In *Proceedings of the fourteenth ACM international symposium on Mobile ad hoc networking and computing*. ACM, 167–176.
[25] Junhyuk Oh, Xiaoxiao Guo, Honglak Lee, Richard L Lewis, and Satinder Singh. 2015. Action-conditional video prediction using deep networks in atari games. In *Advances in Neural Information Processing Systems*. 2863–2871.
[26] Majid Rabbani. 2002. JPEG2000: Image compression fundamentals, standards and practice. *Journal of Electronic Imaging* 11, 2 (2002), 286.
[27] Sawsan Abdul Rahman, Azzam Mourad, May El Barachi, and Wael Al Orabi. 2018. A novel on-demand vehicular sensing framework for traffic condition monitoring. *Vehicular Communications* 12 (2018), 165–178.

[28] Joshua E Siegel, Dylan C Erb, and Sanjay E Sarma. 2017. A survey of the connected vehicle landscape—architectures, enabling technologies, applications, and development areas. *IEEE Transactions on Intelligent Transportation Systems* 19, 8 (2017), 2391–2406.
[29] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-Laszlo Barabasi. 2010. Limits of Predictability in Human Mobility *(Science)*.
[30] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.
[31] Yongxin Tong, Yuqiang Chen, Zimu Zhou, Lei Chen, Jie Wang, Qiang Yang, Jieping Ye, and Weifeng Lv. 2017. The simpler the better: a unified approach to predicting original taxi demands based on large-scale online platforms. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1653–1662.
[32] Elisabeth Uhlemann. 2016. Connected-vehicles applications are emerging [connected vehicles]. *IEEE Vehicular Technology Magazine* 11, 1 (2016), 25–96.
[33] Yang Wang, Wuji Chen, Wei Zheng, He Huang, Wen Zhang, and Hengchang Liu. 2017. Tracking Hit-and-Run Vehicle with Sparse Video Surveillance Cameras and Mobile Taxicabs. In *2017 IEEE International Conference on Data Mining (ICDM)*. IEEE, 495–504.
[34] Xiaoyang Xie, Yu Yang, Zhihan Fang, Guang Wang, Fan Zhang, Fan Zhang, Yunhuai Liu, and Desheng Zhang. 2018. coSense: Collaborative Urban-Scale Vehicle Sensing Based on Heterogeneous Fleets. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (2018), 196.
[35] Xiaoyang Xie, Fan Zhang, and Desheng Zhang. 2018. PrivateHunt: Multi-Source Data-Driven Dispatching in For-Hire Vehicle Systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 45.
[36] Xiaoyang Xie, Fan Zhang, and Desheng Zhang. 2019. Understanding real-time interaction in heterogeneous vehicular sensing. In *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*. 450–451.
[37] Susu Xu, Xinlei Chen, Xidong Pi, Carlee Joe-Wong, Pei Zhang, and Hae Young Noh. 2019. Vehicle dispatching for sensing coverage optimization in mobile crowdsensing systems. In *2019 18th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 311–312.
[38] Bo Yang, Xiaofei Xu, Feng Tan, and Dong Ho Park. 2011. An utility-based job scheduling algorithm for cloud computing considering reliability factor. In *2011 International Conference on Cloud and Service Computing*. IEEE, 95–102.
[39] Yu Yang, Xiaoyang Xie, Zhihan Fang, Fan Zhang, Yang Wang, and Desheng Zhang. 2019. VeMo: Enabling Transparent Vehicular Mobility Modeling at Individual Levels with Full Penetration. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.
[40] Yu Yang, Fan Zhang, and Desheng Zhang. 2018. SharedEdge: GPS-free fine-grained travel time estimation in state-level highway systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–26.
[41] Changxi You, Jianbo Lu, Dimitar Filev, and Panagiotis Tsiotras. 2018. Highway traffic modeling and decision making for autonomous vehicle using reinforcement learning. In *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 1227–1232.
[42] Jia Yu and Mohamed Sarwat. 2017. Indexing the Pickup and Drop-Off Locations of NYC Taxi Trips in PostgreSQL–Lessons from the Road. In *International Symposium on Spatial and Temporal Databases*. Springer, 145–162.
[43] Desheng Zhang, Jun Huang, Ye Li, Fan Zhang, Chengzhong Xu, and Tian He. 2014. Exploring human mobility with multi-source data at extremely large metropolitan scales. In *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 201–212.
[44] Desheng Zhang, Fan Zhang, and Tian He. 2016. MultiCalib: National-scale traffic model calibration in real time with multi-source incomplete data. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 19.
[45] Desheng Zhang, Juanjuan Zhao, Fan Zhang, and Tian He. 2015. coMobile: Real-time human mobility modeling at urban scale using multi-view learning. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 40.
[46] Dong Zhao, Huadong Ma, Liang Liu, and Xiang-Yang Li. 2015. Opportunistic coverage for urban vehicular sensing. *Computer Communications* 60 (2015), 71–85.
[47] Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. 2014. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5, 3 (2014), 38.
[48] Pengfei Zhou, Yuanqing Zheng, and Mo Li. 2012. How Long to Wait? Predicting Bus Arrival Time With Mobile Phone Based Participatory Sensing. *IEEE Trans. Mob. Comput.* 13 (2012), 1228–1241.
[49] Qiuxi Zhu, Md Yusuf Sarwar Uddin, Nalini Venkatasubramanian, and Cheng-Hsin Hsu. 2018. Spatiotemporal scheduling for crowd augmented urban sensing. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 1997–2005.