CatCharger: Deploying Wireless Charging Lanes in a Metropolitan Road Network through Categorization and Clustering of Vehicle Traffic

Li Yan*, Haiying Shen*, Juanjuan Zhao†, Chengzhong Xu†, Feng Luo‡ and Chenxi Qiu§
*Department of Computer Science, University of Virginia, USA
†Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China
‡School of Computing, Clemson University, USA
§College of Information Sciences and Technology, Pennsylvania State University, USA
Email: {ly4ss, hs6ms}@virginia.edu, {jj.zhao, cz.xu}@siat.ac.cn, luofeng@clemson.edu, czq3@psu.edu

***Abstract*—The future generation of transportation system will be featured by electrified public transportation. To fulfill metropolitan transit demands, electric vehicles (EVs) must be continuously operable without recharging downtime. Wireless Power Transfer (WPT) techniques for in-motion EV charging is a solution. It however brings up a challenge: how to deploy charging lanes in a metropolitan road network to minimize the deployment cost while enabling EVs' continuous operability. In this paper, we propose *CatCharger*, which is the first work that handles this challenge. From a metropolitan-scale dataset collected from multiple sources of vehicles, we observe the diversity of vehicle passing speed and daily visit frequency (called traffic attributes) at intersections (i.e., landmarks), which are important factors for charging lane deployment. To select landmarks for deployment, we first group landmarks with similar traffic attribute values using the entropy minimization clustering method, and choose better candidate landmarks from each group suitable for deployment. To determine the deployment locations from the candidate landmarks, we infer the expected vehicle residual energy at each landmark using a Kernel Density Estimator fed by the vehicles' mobility, and formulate and solve an optimization problem to minimize the total deployment cost while ensuring a certain level of expected residual energy of EVs at each landmark. Our trace-driven experiments demonstrate the superior performance of *CatCharger* over other methods.**

I. Introduction

In recent decades, due to the foreseen depletion of fossil fuels, the world is actively competing against its vanishing in various energy-related directions. Gasoline-based vehicles, as the major contributor of greenhouse gases, have become the top concern, which need alternatives. With the support of government and auto companies, many techniques of electric vehicles (EVs) have been developed ranging from the EV itself to supportive infrastructures [1].

Because of the limit of the battery capacity, the driving range of EVs is limited (e.g., 100 miles) [1]. Hence, EVs must be recharged frequently during driving time. In plug-in station charging, EVs stop and get plugged in the charging points of the charging stations to get recharged. Frequently stopping to get recharged from charging stations wastes time and becomes an obstacle for wide adoption of EVs. To fulfill metropolitan transit demands, EVs must be continuously operable without recharging downtime [2]. Recently, the world has seen a surge in Wireless Power Transfer (WPT) techniques for in-motion EV charging as a solution. It can charge an EV as long as the EV is moving over a charging lane installed in the road [3], [4]. It however brings up a new challenge: how to deploy charging lanes (i.e., determine the locations and lane lengths) in a metropolitan road network to minimize the deployment cost while enabling EVs to be continuously operable on the roads. By operable, we mean that an EV's residual energy measured by State of Charge (SoC) (i.e., percentage of stored energy) is non-zero. However, *there have been no previous works that handle this challenge*.

There have been multiple works proposed for optimally deploying plug-in charging stations [5]–[11]. Generally, such methods can be categorized into charging demand based method and traffic flow based method. In the charging demand based methods, vehicle charging demands at different locations are modeled [5]–[7], and the locations of the charging stations are determined to maximally meet the inferred demands on the road network with the minimum deployment cost. However, the inferred demands may not be sufficiently accurate to reflect the real demands. Traffic flow based methods conduct fine-grained analysis of traffic flow measured by the number of vehicles passing certain origin-destination pairs during a unit time [9], [12]. The methods aim to maximally cover the traffic flows with the minimum deployment cost through proper placement of the charging stations under several constraints (e.g., battery capacity, traffic flow) so that the EVs have accessibility to charging stations within their driving ranges [8]–[11].

However, these methods cannot be applied to deploying wireless charging lanes because of their different charging approaches. The amount of energy an EV received from a wireless charging lane depends on the EV's passing speed and the length of the charging lane. Slower speed or longer charging lane leads to more received energy and vice versa, while a longer charging lane costs more to deploy. Therefore, locations with slower driving speeds are better options for deploying charging lanes in order to reduce the deployment cost. However, the driving speed is not considered in planning plug-in charging stations, which therefore cannot be used for wireless charging lane deployment.

Although there have been some researches on formulating and solving an optimization problem for deploying wireless charging lanes to support the SoC of buses driving on a single determined route with the minimum deployment cost [3], the challenge of deploying wireless charging lanes in a metropolitan road network with different sources of traffic and many roads has not been studied, which however is much more formidable. The locations of charging lanes need to cover the traffic maximally, while the charging lane lengths need to be minimized by considering driving speed [13], [14]. Also, in order to enable the deployed charging lanes to support public

transit services including taxicab, bus, and customized transit vehicles (e.g., UberPool) in a metropolitan city, the datasets for traffic analysis for the deployment task must be from these different sources of vehicles and cover a metropolitan-scale road network for a certain period of time [15], [16].

To handle the challenge, we propose *CatCharger*, a method using Categorization and clustering of vehicle traffic attributes (i.e., driving speed, visit frequency at a location) to determine the deployment of wireless Chargers considering metropolitan-scale charging demands of multiple sources of EVs. Since intersections have more traffic than the other positions [17], [18], we extract intersections (i.e., landmarks) as candidate locations. *CatCharger* is a heuristic method consisting of four steps: "categorization-clustering-extraction-deployment".

Specifically, we first analyze a metropolitan-scale dataset consisting of multiple sources of vehicles in the city of Shenzhen, China to extract traffic statistics. It records the status (e.g., timestamp, GPS position, speed) of 15,610 taxicabs, 14,262 buses, 12,386 customized transit vehicles every 30 seconds for one month (July 1 – July 31, 2015). We measured vehicles' average passing speed in each landmark (which determines the lane length required for fully charging an EV [3]), and average daily vehicle visit frequency in each landmark (which reflects the charging service capability), and observed that they are widely distributed. Our dataset analytical results lay the foundation of the design of *CatCharger*.

We utilize an entropy-based algorithm [19], which is effective in clustering items with similar attributes, to group the landmarks with similar average passing speed and visit frequency. Within each group, we rank each landmark (based on the average daily visit frequency and the estimated lane length required for fully charging an EV) to evaluate the landmark's suitability for installing a charging lane. Based on the average rank of each group, we select groups that are more suitable for deployment. We then select the top ranked landmarks from each selected group as candidate locations.

To keep EVs operable at any position in the road network, we aim to ensure that the EVs have a certain level of residual energy when they arrive at each landmark. This residual energy level enables an EV to move to its nearest charging lane. Therefore, we need to infer the EVs' traffic (i.e., the probability of reaching each landmark in the road network) in order to estimate the expectation of the EVs' residual energy when they arrive at each landmark given deployed charging lanes. Since the vehicles' mobility is not uniformly distributed at all landmarks, we use the Kernel Density Estimator (KDE) [20] to effectively model the traffic of vehicles at each landmark. Finally, we formulate and solve an optimization problem to minimize the total cost of deploying charging lanes while ensuring that EVs have a certain level of expected residual energy at each landmark.

We have conducted extensive trace-driven experiments to show the effectiveness of *CatCharger* in handling the challenge compared with the random location selection method and the method that chooses the landmarks covering the most traffic. *To our knowledge, this paper is the first work for wireless charging lane deployment considering multiple sources of vehicle mobility in a metropolitan scenario.* The remainder of the paper is organized as follows. Section II presents literature review. Section III presents our metropolitan dataset measurement results. Section IV presents the detailed design of *CatCharger*. Section V presents trace-driven evaluations. Section VI concludes the paper with remarks on future work.

II. Related Work

Plug-in charging station deployment. Several previous works [5]–[7] deploy charging stations based on the demands deduced from models (e.g., queueing theory, driver preference, and parking patterns). Bae *et al.* [5] proposed to deploy charging stations through analyzing the spatial and temporal dynamics of charging demand profiles at potential positions using the fluid dynamic model. Zheng *et al.* [6] formulated an optimization problem trying to maximize the number of EVs charged while minimizing the life cycle cost of all the stations. Eisel *et al.* [7] aimed at dealing with drivers' range anxiety (i.e., fear of being unable to reach destination due to insufficient charging opportunities) by transforming the drivers' preference in charging into planning of stations.

Further, several traffic flow based charging station deployment algorithms have been proposed [8]–[11]. Lam *et al.* [8] formulated the station placement as a vertex cover problem, proved its NP-hardness and proposed four solutions. Wang *et al.* [9] determined constraints (e.g., driving range, traffic volume) from EV traffic statistics, and formulated and solved a multi-objective location optimization problem to maximize the coverage of EV traffic. Sánchez-Martín *et al.* [10] proposed to deploy charging stations at the positions with many parking events and suitable parking time length with the minimum deployment cost to offer EVs enough charging opportunities. Yao *et al.* [11] formulated a problem trying to minimize deployment cost to maximize the covered EV traffic flow.

The methods for deploying plug-in charging stations cannot be used for deploying wireless charging lanes due to their different charging approaches as explained previously.

Wireless power transfer for EVs. In 2006, Karalis *et al.* [21] from MIT introduced a resonant coupler that wirelessly transmits a large amount of power to EVs at low frequencies. Jang *et al.* [3] formulated an optimization problem, which considers battery capacity and charging lane length as constraints, to deploy wireless charging lanes to maintain the SoC of buses on a single determined route with the minimum cost. Sarker *et al.* [22] developed a wireless power transfer system for balancing the SoC of EVs in a charging lane in the urban scenario. However, the problem of deploying wireless charging lanes in a metropolitan road network considering different sources of traffic and many roads has not been studied. *As far as we know, this work is the first to handle this problem.*

III. Metropolitan-Scale Dataset Measurement

A. Dataset Description and Data Processing System

Our datasets for Shenzhen record the status (e.g., timestamp, position, speed) of vehicles for one month (July 1 – 31, 2015), with a recording period less than 30 seconds, which include:

(1) **Taxicab Dataset.** It is collected by the Shenzhen Transport Committee, which records the status (e.g., timestamp, position, speed) of 15,610 taxicabs. The daily size of the uploaded data is around 2GB.

(2) **Bus Dataset.** It is also collected by the Shenzhen Transport Committee, which records the status of 14,262 buses (e.g., timestamp, GPS position).
(3) **Dada bus Dataset.** It is provided by the Dada Bus corporation (a customized transit service similar to UberPool), which records the status (e.g., timestamp, position, speed) of 12,386 reserved service buses.
(4) **Road Map.** The road map of Shenzhen is obtained from OpenStreetMap [23]. According to the municipal information of Shenzhen [24], we use a bounding box with coordinate $(lat = 22.4450, lon = 113.7130)$ as the south-west corner, and coordinate $(lat = 22.8844, lon = 114.5270)$ as the north-east corner, which covers an area of around 2,926km^2, to crop the road map data.

For efficient management of such large datasets, we utilized a 34 TB Hadoop Distributed File System (HDFS) [25] on a cluster consisting of 10 nodes, each of which is equipped with 16 cores and 64 GB RAM [26]–[28]. For data processing, we utilized Apache Spark [29], which is a fast in-memory cluster computing system, alongside the Hadoop cluster.

B. Important Issues

There are two main issues that need to be addressed in handling the charging lane deployment challenge:

(1) **Reducing charging lane length.** The charging lanes need to be as short as possible in order to reduce the deployment cost, while still enabling EVs to be fully charged when they pass a lane. To select locations for charging lane deployment to achieve this objective is non-trivial.
(2) **Reducing the number of deployed charging lanes.** The problem of determining the locations of the charging lanes on a metropolitan road network to maintain the continuous operability of the EVs on roads, while minimizing the number of deployed charging lanes, is non-trivial.

1) Vehicle Velocity at Charging Lanes Matters: The amount of energy transmitted to an EV from a wireless charging lane (E) equals: $E = L \cdot r/v$, where L denotes the length of the charging lane, r denotes its energy supply rate, and v denotes the vehicle's speed passing through the charging lane.

Since EVs with different battery capacities may pass a charging lane, to ensure that any EV can be fully charged when it passes a charging lane, the charging lane must be able to supply an EV E_{max} amount of energy, which is the maximum battery capacity of EVs. Then, an EV with energy demand less than E_{max} still can be fully charged when it passes the charging lane. Therefore, when a landmark i with average passing speed $\overline{v}_i$ is chosen to deploy a charging lane, its length is determined to meet the above condition:

$$L_i = \frac{E_{max}}{r}\overline{v}_i. \quad (1)$$

Note $\frac{E_{max}}{r}$ is a constant, so the charging lane length (L_i) is directly determined by vehicle average passing speed ($\overline{v}_i$). Since a longer charging lane leads to higher deployment cost [3], [4], [14], the charging lanes should be placed at the positions with the slowest passing speed. Then, the charging lane has the shortest length that can fully charge passing EVs.

2) Vehicle Visit Frequency and Multi-source Vehicle Traffic Matter: To keep the EVs operable at any location in the city without downtime, the placement of charging lanes must cover the majority of the EV traffic. Therefore, the selection of the charging positions should also consider EV visit frequency. Meanwhile, to support the operability of all EV-based public transit services, considering a single source of vehicle traffic may generate bias and we must consider all sources of vehicle traffic. Our datasets meet this requirement.

C. Dataset Analysis

A road network is essentially a directed graph, in which nodes represent intersections and edges represent road segments [30]. The movement records of a vehicle are continuous. We first generate the driving trajectory of each vehicle. We view a vehicle has finished its previous trajectory if it stops at a location for more than 10 minutes. Thus the stopping locations cut the movement records into multiple trajectories. Since vehicles usually change movement at intersections, we map each position record to its nearest landmark (in Euclidean distance). Then, a vehicle trajectory can be represented by a sequence of landmarks [31]. We define vehicle trajectory as:

Definition 1. *A vehicle n_i's trajectory is a sequence of time-ordered spatial positions, $Tr_i : \{(\boldsymbol{p_0}, t_0), (\boldsymbol{p_1}, t_1), \ldots, (\boldsymbol{p_m}, t_m)\}$, where each position is represented by a latitude and a longitude $\boldsymbol{p_j} = (lat_j, lon_j)$.*

Through measurement, we found that the range and the average of vehicle visit frequency at a landmark are $[0/day, 96, 637/day]$ and $3,840/day$, and the range and the average of vehicle passing speed in a landmark are $[0km/h, 142km/h]$ and $20km/h$. Figure 1 shows the distribution of landmarks (black dots) whose vehicle visit frequency is higher than $10^4/day$, and vehicle passing speed is lower than $60km/h$. The territory of Shenzhen consists of 7 functional regions (e.g., commercial, residential). We can see that each region has several candidate landmarks with both high vehicle visit frequency and slow passing speed.

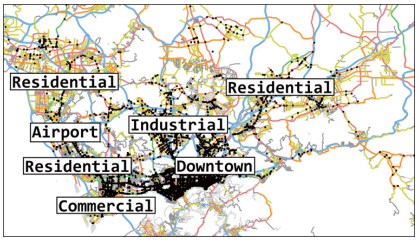


Fig. 1: Distribution of potential positions for placing charging lanes.

Figure 2 shows the cumulative distribution function (CDF) of average vehicle passing speed and average vehicle visit frequency per day of each landmark. Figure 3 plots the density distribution of vehicle passing speed with respect to (w.r.t.) vehicle visit frequency to illustrate the distribution of positions with both slow vehicle passing speed and high vehicle visit frequency. In Figure 2, we see that the landmarks with vehicle visit frequency higher than $10^4/day$ only take less than 25% of all the landmarks, and the landmarks with vehicle passing speed less than $60km/h$ take up about 80% of all the landmarks. In Figure 3, we can see the landmarks with

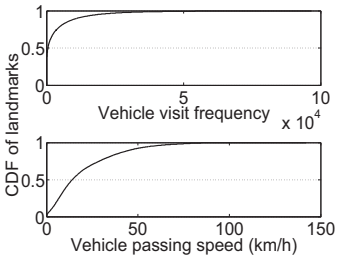


Fig. 2: Average vehicle passing speed & daily vehicle visit frequency.

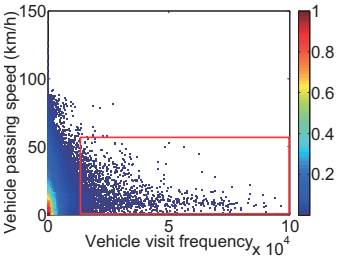


Fig. 3: Density scatter of vehicle passing speed w.r.t. vehicle visit frequency.

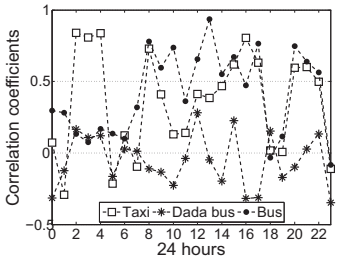


Fig. 4: Correlation between each vehicle source and public transit traffic.

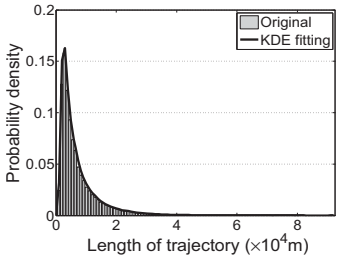


Fig. 5: Distribution of trajectory lengths & estimation of KDE.

both low vehicle passing speed ($60km/h$) and high vehicle visit frequency ($10^4/day$) take up a small portion within the square circle. The above observations motivate us to find an innovative method to properly extract candidate charging lane placement positions considering the diversity in vehicle passing speed and visit frequency, and their distribution in different regions. The details will be elaborated in Section IV-C.

We further compare the mobility of each vehicle source with the total public transit mobility to demonstrate the necessity of using multiple sources of vehicle mobility in collecting traffic statistics. We define an activity of a vehicle as a position change in the vehicle's trajectory. Then, we calculate the average number of the activities of each kind of vehicles during each hour throughout a day for one month. Next, we use the Pearson correlation coefficient [32] to measure their respective correlation to the total movement activity of public transit vehicles (i.e., bus+taxi+Dada bus), as shown in Figure 4. The result shows that during morning hours (i.e., 00:00∼06:00), the activity of taxis is more correlated with the public transit mobility than bus and Dada bus, which means that the taxis play the main role in public transit service during this period of time. This is because most bus lines and customized transits are not in service during this period. Starting from 07:00, the correlation between the bus and the public transit mobility is higher than the others. This is because the buses are in service after this time point, so they represent the public transit mobility the most. Since the operation of Dada buses is driven by crowdsourced requests, so they do not operate throughout 24 hours. Therefore, it has low correlation to the public transit traffic at many time points in the figure. We see that at around 12:00, 13:00 and 18:00, the correlation of Dada buses to the public transit mobility is close to the others. It means that during these times, Data buses provide a majority service to meet the public transit demands, and the mobility of the Dada buses must be considered in measuring the public transit demands that the charging lanes need to satisfy.

Considering that the vehicles' trajectories reflect their traffic between different locations [20], and the length of a trajectory determines the energy consumption, we calculated the length of the trajectories of each vehicle in one month. The distribution of the collected trajectory lengths is shown in Figure 5. We can see that most of the trajectories are less than 10,000 meters. The long trajectories are mostly generated by buses, as they drive continuously on scheduled routes when in service. However, the distribution of the trajectory lengths cannot be simply modeled using a parametric distribution. Since KDE is a non-parametric method to estimate the probability density function of a random variable, we feed the lengths of the trajectories to the KDE model to infer the vehicles' probability of reaching each landmark in the road network. The curve in Figure 5 represents the distribution fitting result from the KDE. The KDE is a function of the trajectory length. Based on a trajectory's length, we can calculate an EV's residual energy after it drives through the trajectory. Then, using the probability of reaching each landmark from the KDE, we can estimate the expectation of residual energy of EVs at each landmark on the road network given deployed charging lanes. Then, we can formulate an optimization problem that aims to maintain the expected residual energy above a threshold at each landmark with the minimum charging lane deployment cost.

Summary: We observed that different landmarks have different vehicle passing speeds and vehicle visiting frequencies. We conclude that the determination of wireless charging lane positions needs to: (i) consider vehicle passing speed since it determines the deployment cost of the charging lane required for fully charging EVs, (ii) consider vehicle visit frequency since it determines the landmark's capability of serving charging demands, and (iii) comprehensively analyze vehicle mobility from various types of vehicles to ensure that the deployed charging lanes can meet the charging demands from various vehicles. This conclusion motivates us to design a novel approach using multi-source vehicle mobility to extract candidate positions suitable for placing wireless charging lanes, and properly choose the positions so that they can meet all the charging demands of EVs with the minimum deployment cost. We find a solution for the charging lane placement challenge as follows:

Solution: Given a road network comprised of a set of landmarks LM, and trajectory datasets of multiple sources of vehicles $\{Tr\}$, we first extract candidate positions from LM that have both slow passing speed and high visit frequency (i.e., short length of charging lane required for fully charging an EV and high capability for serving charging demands). We then further select positions to place charging lanes to ensure that the expected residual energy of EVs at each landmark is no less than a threshold with the minimum deployment cost.

IV. System Design of CatCharger

A. Framework of CatCharger

As shown in Figure 6, the *CatCharger* consists of following three stages (highlighted as three dashed boxes):

1. **Vehicle mobility normalization** (Section IV-B). First, we need to apply the *Data Cleaning* on the vehicle datasets. Then, based on *OpenStreetMap*, we extract all intersections

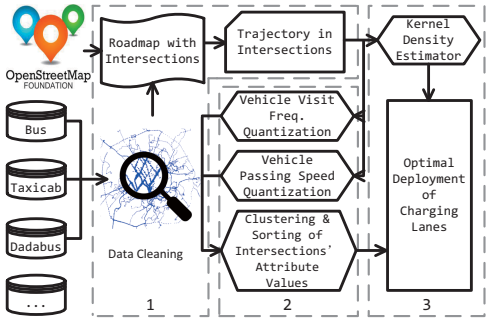


Fig. 6: Framework of CatCharger.

(landmarks) and generate the *Roadmap with Intersections*. Finally, by mapping each position record to respective nearest intersection (in Euclidean distance), we represent a vehicle's mobility by a *Trajectory in Intersections*.

2. **Charging lane location candidate extraction** (Section IV-C). With the data output from the first stage, we apply the *Vehicle Visit Frequency Quantization* and the *Vehicle Passing Speed Quantization* to generate the traffic attribute values for each intersection. Then, we apply the *Clustering & Sorting of the Intersections' Attribute Values* to extract the intersections with both high vehicle visit frequency and short required charging lane length.
3. **Charging lane location determination** (Section IV-D). We first use the lengths of the trajectories to build the *Kernel Density Estimator* (KDE), which is used to estimate the vehicles' traffic at different landmarks. Then we formulate an optimization problem to solve the wireless charging lane deployment problem, and its solution outputs the locations and lane lengths for *Optimal Deployment of Charging Lanes*.

B. Vehicle Mobility Normalization

The original movement records of vehicles are mixed with noises (e.g., records with duplicated GPS position, timestamp and etc., and records with GPS positions out of the area range of Shenzhen), so we first need to clean the datasets by removing the duplicated records. Moreover, as a road network can be abstracted into intersections and road segments [30], we map each position record to its nearest intersection in Euclidean distance. Finally, the original position records generated by the vehicles in a time period (Figure 7 for 07:00–07:30 on July 1, 2015) are normalized to sequences of landmarks (Figure 8).

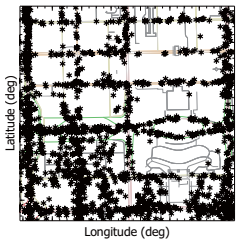


Fig. 7: Original mobility.

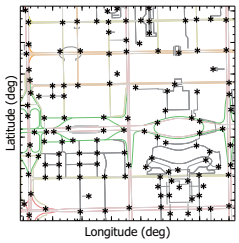


Fig. 8: Normalized mobility.

C. Charging Lane Location Candidate Extraction

Vehicle visit frequency on different landmarks varies in different regions [33]. For example, the vehicle visit frequency of a landmark in Downtown is generally much higher than a landmark in an Industrial region. The deployment of charging lanes still needs to consider the charging demands in Industrial areas as well in order to support the traffic of the EVs on the entire road network. For this purpose, considering that the landmarks belonging to the same region usually have similar attribute values (i.e., average vehicle passing speed, daily vehicle visit frequency), we cluster such landmarks to one group. Since the landmarks in each group have similar attribute values, they are almost equally important in deploying location selection. We then choose groups with better attribute values for deploying charging lanes. Next, from each group, we further extract the landmarks with the best attribute values from different regions. Finally, the extracted candidate landmarks are the ones with high suitability for deploying charging lanes, and geographically distributed in different functional regions.

In the following, we first introduce how *CatCharger* categorizes attribute values, and then clusters the landmarks using the entropy minimization based clustering method.

1) Categorization of Original Mobility Data: As previously indicated, we need to consider two attributes of the landmarks (i.e., vehicle passing speed, and vehicle visit frequency) in the deployment. But directly clustering the landmarks with different numerical attributes is non-trivial because it is not easy to define "closeness" between the attributes based on the Euclidean distance. For example, given three landmarks: $A(50km/h, 10000/day)$, $B(50km/h, 9000/day)$ and $C(100km/h, 10000/day)$. With K-means using Euclidean distance, C is more similar to A than B, though actually A is more similar to B because they require the same charging lane length and have high vehicle visit frequency.

To handle this problem, we propose to properly categorize the attribute value range into several intervals, with each interval representing a range of speed (v) and visit frequency (f). Jang *et al.* [3] used a speed variance of $5km/h$ in determining the charging lane length, so we use it in categorizing speeds. As for vehicle visit frequency at landmarks, since there are around 51,000 vehicles being considered, we use 1,000 as the categorization interval. Each interval has an ID. Thus, the attributes are categorized into IDs like:

$$\begin{aligned} &v: <0,\ 0\sim 5km/h>, <1,\ 5\sim 10km/h>, \ldots, \\ &f: <0,\ 0\sim 1000/day>, <1,\ 1000\sim 2000/day>, \ldots. \end{aligned} \tag{2}$$

Each attribute has the form <attribute ID, description>. For example, the attributes of a landmark with an average vehicle passing speed of $3km/h$ and a daily average vehicle visit frequency of $1500/day$ are represented as $\{0, 1\}$.

2) Clustering of Landmarks: After categorization, each landmark is described with two attribute IDs. *CatCharger* clusters the landmarks with the most similar attribute values into one group. We use entropy, which measures categorical disorder (i.e., dissimilarity of attribute IDs within a group) [19] for clusering. Let's take the attribute of vehicle passing speed as an example. Suppose a group has two landmarks with attribute IDs $\{0\}$ and $\{1\}$, respectively. F is a discrete random variable representing an attribute (e.g., average vehicle passing speed), $A(F)$ is the set of the attribute IDs of F in a group (e.g., 0, 1), and $p(f)$ is the probability function of F, namely the ratio of the attribute ID in the group (e.g., 0.5). The entropy of the attribute $H(F)$ within the group is defined as:

$$H(F) = -\sum_{f \in A(F)} p(f)\log_2(p(f)), \tag{3}$$

where $-\log_2(p(f))$ measures the dissimilarity of the attribute in the group. The entropy of the two landmarks in the example is $\frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2 = 1$. Higher dissimilarity between two landmarks' attribute IDs leads to a larger entropy. Since each landmark has two attributes, the entropy of a cluster C_i can be calculated as the sum of the entropies of the two attributes:

$$H(C_i) = H_i(F_0) + H_i(F_1). \tag{4}$$

Suppose all candidate landmarks LM are clustered into k clusters $\mathcal{C} = \{C_0, \ldots, C_{k-1}\}$. To measure the quality of the clustering, we use the weighted sum of the entropies of all clusters as the expected entropy resulted from the clustering. The weight for each cluster is calculated as $\frac{|C_i|}{|LM|}$, where $|\cdot|$ means the number of landmarks in the set. Thus, the expected entropy is calculated by:

$$\overline{H}(\mathcal{C}) = \sum_{k=0}^{k-1} \frac{|C_i|}{|LM|} H(C_i). \tag{5}$$

Given a set of landmarks for clustering, we first find all possible clustering arrangements, and then choose the one with the minimum expected entropy. As a result, the optimal clustering strategy renders clusters whose member landmarks have the least dissimilar attribute IDs between each other. Unfortunately, such a clustering strategy is difficult to execute because it is NP-complete [34]. Then, *CatCharger* instead follows a heuristic method introduced in [19] to approximate the best solution. The steps of the landmark clustering are as follows:

(i) *Initialization:* To cluster landmarks into k groups, we must start with k most dissimilar landmarks. But directly extracting such k landmarks from the entire set of landmarks is non-trivial. To handle this problem, we take a sample S from the set of landmarks LM ($|S| \ll |LM|$). In S, we enumeratively calculate the entropy generated by each pair of landmarks, and place the two landmarks that generate the maximum entropy in two clusters (C_0, C_1) as the two starting clusters. Then, the remaining $k-2$ starting landmarks will be incrementally found as the ones that are most dissimilar with the already determined ones.

(ii) *Incremental clustering:* After the initialization, the remaining $|LM| - k$ landmarks will be clustered to the respective starting landmark that renders the minimum total expected entropy (Equation (5)) one by one.

The major problems with such heuristic clustering include: i) how to select the sample S, ii) how to determine the number of clusters k, and iii) incrementally clustering the landmarks may deteriorate the clustering quality. For i), we randomly select $\gamma\%$ (e.g., 10%) of the landmarks from every functional region of Shenzhen, and combine them as the sample because each region needs several charging positions to support the EV traffic. For ii), within the sample, we follow the algorithm developed in [35] to find the most suitable k that results in the maximum difference in entropy changing rate, of which complexity is $O(|S|^2)$. As for iii), we repeat the clustering steps (in which landmarks are randomly picked) for several times and choose the result with the minimum entropy.

3) Extracting Top Ranked Landmarks from Clusters: Note the required length of charging lane i (L_i) can be calculated by Equation (1) based on passing speed. Since the shorter charging lane a landmark requires, and the higher vehicle visit frequency the landmark has, the more suitable it is for placing a charging lane. Then, we define rank of landmark $\boldsymbol{lm_i} \in C_j$:

$$R(\boldsymbol{lm_i}) = \frac{\log(f_i)}{L_i}, \tag{6}$$

where f_i is the vehicle visit frequency at $\boldsymbol{lm_i}$. We use logarithmic value of f_i because f_i is generally much larger than L_i. To ensure the suitability of selected landmarks, we need to remove landmarks with low ranks. For this purpose, we calculate the average rank of each group, and then remove groups with ranks lower than a threshold. Next, we order the landmarks in each group in decreasing order of the rank. In one group, if there are several landmarks in one region, we remove the low-rank landmarks. Finally, we select the top ranked $\eta\%$ (e.g., 10%) of the landmarks from each group, and use them as the candidates for charging lane deployment, which are denoted as $\widetilde{LM} = \{\boldsymbol{lm_0}, \boldsymbol{lm_1}, \ldots, \boldsymbol{lm}_{|\widetilde{LM}|}\}$.

D. Charging Lane Location Determination

To determine the deployment plan on the selected candidate locations, we first use the KDE, which is fed with vehicle mobility, to infer the EVs' expected residual energy at each landmark given that certain landmarks are installed with charging lanes. Then, we formulate an optimization problem that aims to minimize the total cost of deployment while ensuring that the EVs can have a certain level of expected residual energy when they arrive at each landmark. This residual energy level enables an EV to move to its nearest charging lane.

1) Inferring Expected Residual Energy: KDE can be used to describe the vehicles' probability of reaching a landmark on the road network given a source landmark. Also, the residual energy of a vehicle is a function of the distance from the vehicle's source landmark to the destination landmark. Then, the expected residual energy of a vehicle at a landmark in the road network can be calculated. We present the details below.

Since vehicles' mobility patterns imply their traffic at certain locations [20], we feed the vehicles' trajectories into a KDE model to infer the probability density function (PDF) of the distribution of the trajectory lengths as in Equation (7), namely the trip lengths that need to be supported.

$$\hat{f}_h(d) = \frac{1}{mh}\sum_{i=0}^{m-1} K(\frac{d-d_i}{h});\ -\infty < d < \infty, \tag{7}$$

where m is the number of sample trajectories, d_i is the length of the i^{th} trajectory, and h is the smoothing parameter influencing the estimation accuracy of the KDE and is determined according to the MISE criterion [36]. $K(\cdot)$ is the kernel function whose value decays with the increasing of d. It is set to the Gaussian function as in Equation (8) based on [37], [38].

$$K(\frac{d-d_i}{h}) = \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{(d-d_i)^2}{2h^2}\right]. \tag{8}$$

Suppose the vehicles' energy consumption rate per meter is a constant c, and the minimum battery capacity of the EVs considered is E_{min}. Suppose an EV gets fully charged at a charging lane, its residual energy (in percentage) at a location, which is d away from the charging lane, can be estimated as

$$SOC(d) = \begin{cases} \frac{E_{min} - cd}{E_{min}}, & \text{if } E_{min} \geqslant cd \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

The reason we consider the minimum battery capacity here is that when d increases, the residual energy of the vehicle with the minimum battery capacity, $E_{min} - cd$, will first become negative. Therefore, as long as the deployment plan can enable the EVs with the minimum battery capacity to be operable, the other EVs will also be operable. Thus, given a binary integer x_i to denote whether a candidate landmark $\boldsymbol{lm_i} \in \widetilde{LM}$ is installed with a charging lane or not, the expected SOC of EVs at a landmark $\boldsymbol{lm_j} \in LM$ in the road network is:

$$\overline{SOC}(\boldsymbol{lm_j}) = \sum_{i=0}^{|\widetilde{LM}|-1} \hat{f}(d_{i,j}) SOC(d_{i,j}) x_i, \quad (10)$$

where $d_{i,j}$ is the shortest route distance from $\boldsymbol{lm_i}$ to $\boldsymbol{lm_j}$.

2) Formulating Optimization Problem: Our objective is to minimize the total deployment cost through properly selecting landmarks from $\widetilde{LM}$ to install charging lanes while ensuring that at each landmark, the expected residual energy of an EV is higher than a threshold α (e.g., 20%). The threshold is determined so that an EV use the residual energy to reach the nearest charging lane. Thus, the optimization problem can be formulated as below:

$$\begin{aligned} \text{minimize} \quad & \sum_{\boldsymbol{lm_i} \in \widetilde{LM}} \omega_0 x_i L_i \\ \text{subject to} \quad & \overline{SOC}(\boldsymbol{lm_j}) \geqslant \alpha, \forall\, \boldsymbol{lm_j} \in LM \\ & x_i \in \{0,1\}, \forall\, \boldsymbol{lm_i} \in \widetilde{LM} \end{aligned} \quad (11)$$

where ω_0 is a constant representing the cost of deploying a unit length of charging lane. Note the reason we filter candidate landmarks by their attribute values of vehicles' passing speed and visit frequency is that *CatCharger* does not consider the landmarks that require a too long charging lane to fully charge an EV or have low vehicle visit frequency. Therefore, the binary integers for the non-candidate landmarks are 0, namely $x_i = 0, \forall\ \boldsymbol{lm_i} \in LM \setminus \widetilde{LM}$. By using b_{ij} to represent $\hat{f}(d_{i,j})\overline{SOC}(d_{i,j})$ in Equation (10), the optimization problem (11) is finally simplified into the form of a Binary Integer Programming (BIP) problem. We refer to an existing toolbox (e.g. GLPK [39], Cbc [40]) to obtain the integer-feasible solution to the problem.

V. Performance Evaluation

A. Experiment Settings

We used our Shenzhen datasets to drive our experiments. We built a trace-driven simulator with Apache Spark 1.5.2 [29]. Since there are no previous methods that handle the wireless charging lane deployment in a road network, we created two methods to compare with *CatCharger*: random placement (denoted by *Random*), and a method that maximally covers traffic flows (denoted by *MaxFlow*) [11].

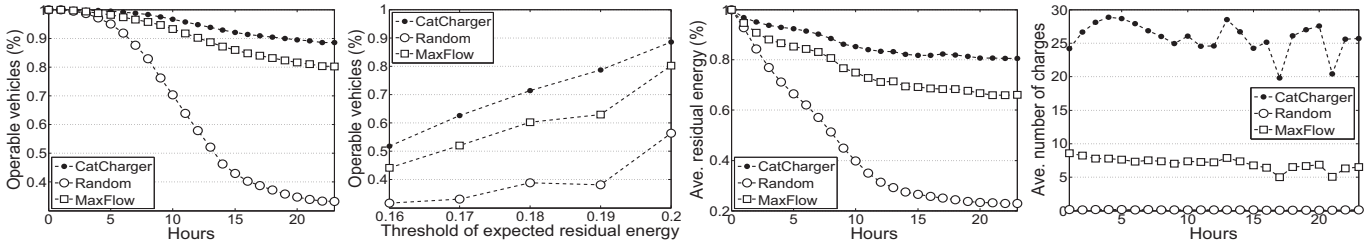
In simulation, the battery capacities of the EVs follow a uniform distribution ranging from 5kWh to 10kWh. We suppose every vehicle starts driving with full energy in battery at the beginning of a day. The energy supply rate of a charging lane is 150kW [3]. The unit price of a charging lane is $\$100/m$ [3]. In *CatCharger*, the length of a charging lane is calculated by Equation (1). Since *Random* and *MaxFlow* do not have methods to determine the charging lane length, we suppose they deploy a 5km-long charging lane (maximum length in *CatCharger*) at each charging landmark, which can fully charge the EVs with the battery capacity smaller than 10kWh and the passing speed slower than 65km/h. For fair comparison, the deployment cost in *Random* and *MaxFlow* is the same as *CatCharger*. In *Random*, the locations for placing charging lanes are chosen randomly from the collection of landmarks. *MaxFlow* is for charging station deployment and we use it for charging lane deployment. We choose the landmark that covers the most traffic sequentially until the deployment cost is reached. *MaxFlow* is a traffic flow based method. Since traffic flow based methods can more accurately estimate the charging demands than the charging demand based methods [8], we do not include a charging demand based method for comparison. We set the values of the parameters (γ, η and α) as indicated previously.

B. Experimental Results

From total 26,036 landmarks, *CatCharger* chose 922 landmarks to deploy charging lanes, while *Random* and *MaxFlow* chose 228. Since *CatCharger* places most of the charging lanes at positions with short required lane lengths, while *Random* and *MaxFlow* use the same deployment cost and set the length of each charging lane to the longest length in *CatCharger*, so they result in much fewer charging lanes.

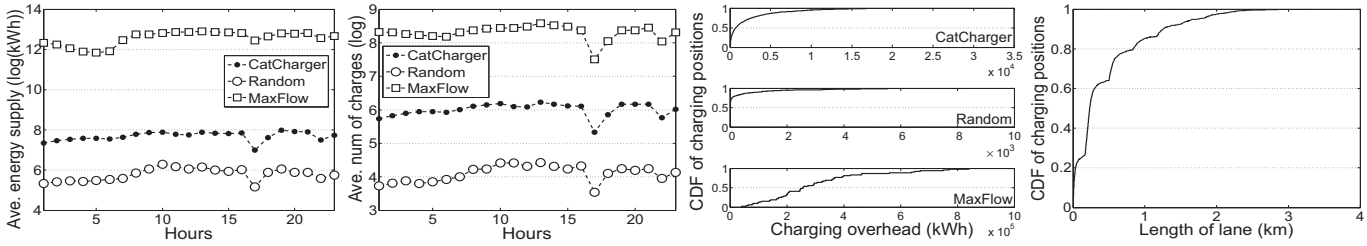
1) Average Ratio of Operable Vehicles: Figure 9(a) shows the average ratios of operable vehicles (SOC>0%) in each hour in a day during the month. Figure 9(b) shows the ratios of operable vehicles resulted from different residual energy thresholds. In both figures, the ratios follow: *CatCharger*>*MaxFlow*>*Random*.

Figure 9(a) shows that at the beginning of a day, the dropping rate of the ratio of operable vehicles is slow because most of the vehicles are not in service and their batteries remain full. Starting from 06:00, the result drops dramatically because the EVs start driving on the road network, which consumes much battery. In *Random*, during the time between 06:00 and 14:00, the dropping rate is almost linear with the change of time, which means that many vehicles cannot be charged during this time. This observation demonstrates that *Random* cannot cover the traffic of most vehicles since it does not consider the mobility of vehicles in determining charging positions. After 15:00, the ratio of operable vehicles gradually stabilizes at around 30%. This is because the buses and Dada buses gradually stop service so their energy levels do not change anymore, while taxicabs are still driving on roads and get charged randomly. As for *MaxFlow*, since it deploys charging lanes at the landmarks with the most traffic flows, and each vehicle passing a charging lane can be fully charged, it can keep most vehicles operable most of the time. During the time between 06:00 and 14:00, the ratio of operable vehicles drops 15%. This is because *MaxFlow* does not consider maintaining the operability of the EVs. When the EVs drive to some landmarks not frequently visited by vehicles, they may not be able to get recharged. Similar to *Random*, after 15:00, the metric gradually stabilizes at around



(a) Operable vehicles over time. (b) Operable vehicles versus threshold. (c) Average residual energy. (d) Avg number of charges of vehicles.

Fig. 9: Performance in supporting EV charging demands.



(a) Energy supply overhead. (b) Num of charges of charging lanes. (c) Distribution of supply overhead. (d) Distribution of charging lane length.

Fig. 10: Performance in distributing energy supply overhead.

80% due to the same reasons. *CatCharger* can keep most of the vehicles operable throughout the day. The stabilized ratio of operable vehicles stays at around 90% in the end of the day. *CatCharger* chooses landmarks that have both high average daily vehicle visit frequency and require short required charging lane length. Thus given the same objective of minimizing the total deployment cost, each charging position of *CatCharger* consumes low cost because it generally has shorter lane length. Therefore, *CatCharger* can offer more charging opportunities (positions) than the other two methods. These positions may not necessarily be the most frequently visited ones, but they altogether can support the continuous operability of the vehicles in the road network.

Figure 9(b) shows that the ratio of operable vehicles of *CatCharger* increases linearly with the increase of the threshold of expected residual energy, since a higher residual energy guarantee increases the probability that an EV can be operable at any position. A higher expected residual energy threshold results in a higher deployment cost of the deployment solution. As a result, as the allowed deployment cost increases, the ratio of operable vehicles in *Random* and *MaxFlow* increases. The increase rate of *CatCharger* is higher than *Random* and *MaxFlow*, which means that *CatCharger* can more effectively plan the positions and lengths to maintain the highest ratio of operable vehicles given a deployment budget.

2) Average Residual Energy of Vehicles: Figure 9(c) shows the average residual energy of vehicles under different hours throughout a day. The results follow *CatCharger*>*MaxFlow*>*Random*. The relationship between the methods and the changing trends of this metric are similar to those in Figure 9(a) due to the same reasons.

3) Average Number of Charges of Vehicles: Figure 9(d) shows the average number of charges of all vehicles under different hours throughout a day. The results follow *CatCharger*>*MaxFlow*>*Random*. We see that *CatCharger* can offer the most charging opportunities to vehicles due to the same reasons as in Section V-B1. Note that the result of *Random* is almost 0. This is because only a small portion of the EVs can receive charging opportunities from the deployment of *Random*.

4) Performance in Distributing Supply Overhead: Figure 10(a) shows the average energy supply overhead (i.e., amount of transferred energy) per landmark (in logarithmic scale) under different hours in a day. Figure 10(b) shows the average number of charges per landmark (in logarithmic scale) under different hours in a day. In both figures, the results follow *MaxFlow*$\gg$*CatCharger*>*Random*. *Random* has the lowest supply overhead because it fails to cover most of the vehicle traffic. *MaxFlow* suffers from a much larger average supply overhead. This is because *MaxFlow* aims to place charging lanes at the landmarks most frequently visited by vehicles. But the popular positions generally concentrate on a few areas. Moreover, the higher cost of a charging lane in *MaxFlow* results in fewer charging positions. Since *CatCharger* tries to deploy short charging lanes by considering vehicle passing speed, it leads to more charging lanes with a certain deployment cost. Also, it tries to cover most vehicle traffic. Then, vehicles in *CatCharger* can be more frequently charged at more landmarks, resulting in lower average energy supply overhead per landmark.

Figure 10(c) shows the CDF of the energy supply overhead over all charging lanes. Figure 10(d) shows the CDF of the length of the charging lanes deployed in *CatCharger*, and *Random* and *MaxFlow* are not included since they have the same charging lane length. In Figure 10(c), we see that the distribution of energy supply overhead in *CatCharger* is more balanced than the others. In Figure 10(d), we see that most of the charging lanes have lengths shorter than 1,000m, while the other methods have 5,000m lengths in order to fully charge an EV without considering passing speed. From Figure 10(c), we see that in *CatCharger*, 80% of the charging lanes only need to supply less than 5,000kWh energy. This can be verified with Figure 10(d), in which 80% of the charging lanes have length shorter than 0.8km. In *Random*, most of the charging

lanes have 0kWh supply overhead. While in *MaxFlow*, around 75% of the charging stations have supply overhead higher than 200,000kWh, which is caused by the fewer charging positions. These observations illustrate that *CatCharger* can better balance the distribution of the energy supply overhead among charging lanes while satisfying the charging demands of EVs.

VI. Conclusion

The rapid development of vehicular WPT techniques brings up a new challenge of deploying wireless charging lanes in a metropolitan road network that support the continuous movement of vehicles with minimum deployment cost. Previous methods for deploying plug-in station are not qualified due to different charging approaches. Previous methods for deploying wireless charging lanes cannot handle the challenge in metropolitan scale. Our proposed *CatCharger* is the first work that tackles this challenge. Our analytical results on a dataset consisting of the mobility records of all public transit vehicles in the city of Shenzhen, China lay the foundation of the design of *CatCharger*. Using an entropy minimization based method, we conduct categorization and clustering on the intersections (landmarks), and extract the candidate positions for placing charging lanes that have low vehicle passing speed (hence short charging lanes) and high vehicle visit frequency (hence high covered traffic). Then by using KDE to model vehicle mobility and to estimate the residual energy of EVs at a landmark, we formulate an optimization problem to minimize the total deployment cost while ensuring the continuous operability of the vehicles on roads. We conducted trace-driven experiments to verify the superior performance *CatCharger* over other methods. In the future, we plan to consider more human activities that affect the movement of public transit vehicles (e.g., pickup requests).

Acknowledgements

This research was supported in part by U.S. NSF grants ACI-1661378 and CNS-1254006, Microsoft Research Faculty Fellowship 8300751, China National Basic Research Program under grant No. 2015CB352400, Research Program of Shenzhen under Grant JSGG20150512145714248, KQCX2015040111035011, CYZZ20150403111012661, and NSFC under grant U1401258.

References

[1] L. Yan, H. Shen, S. Li, and Y. Huang, "Electrical vehicle charging station placement determination based on real world vehicle trace," in *Proc. of IOV*, 2016.

[2] K. C. Dey, L. Yan, X. Wang, Y. Wang, H. Shen, M. Chowdhury, L. Yu, C. Qiu, and V. Soundararaj, "A review of communication, driver characteristics, and controls aspects of cooperative adaptive cruise control (CACC)," *IEEE TITS*, vol. 17, no. 2, 2016.

[3] Y. J. Jang, E. S. Suh, and J. W. Kim, "System architecture and mathematical models of electric transit bus system utilizing wireless power transfer technology," *Systems Journal*, 2015.

[4] S. Jeong, Y. J. Jang, and D. Kum, "Economic analysis of the dynamic charging electric vehicle," *TPE*, 2015.

[5] S. Bae and A. Kwasinski, "Spatial and temporal model of electric vehicle charging demand," *TSG*, 2012.

[6] Y. Zheng, Z. Y. Dong, Y. Xu, K. Meng, J. H. Zhao, and J. Qiu, "Electric vehicle battery charging/swap stations in distribution systems: comparison study and optimal planning," *TPS*, 2014.

[7] M. Eisel, J. Schmidt, and L. Kolbe, "Finding suitable locations for charging stations," in *Proc. of IEVC*, 2014.

[8] A. Lam, Y.-W. Leung, and X. Chu, "Electric vehicle charging station placement: Formulation, complexity, and solutions," *TSG*, 2014.

[9] G. Wang, Z. Xu, F. Wen, and K. P. Wong, "Traffic-constrained multi-objective planning of electric-vehicle charging stations," *TPD*, 2013.

[10] P. Sanchez-Martin, G. Sanchez, and G. Morales-Espana, "Direct load control decision model for aggregated ev charging points," *TPS*, 2012.

[11] W. Yao, J. Zhao, F. Wen, Z. Dong, Y. Xue, Y. Xu, and K. Meng, "A multi-objective collaborative planning strategy for integrated power distribution and electric vehicle charging systems," *TPS*, 2014.

[12] Y. Zhao, Y. Zhang, T. Yu, T. Liu, X. Wang, X. Tian, and X. Liu, "CityDrive: A map-generating and speed-optimizing driving system," in *Proc. of INFOCOM*, 2014.

[13] L. Yan, H. Shen, and K. Chen, "MobiT: a distributed and congestion-resilient trajectory based routing algorithm for vehicular delay tolerant networks," in *Proc. of IoTDI*, 2017.

[14] O. C. Onar, J. M. Miller, S. L. Campbell, C. Coomer, C. White, L. E. Seiber *et al.*, "A novel wireless power transfer for in-motion ev/phev charging," in *Proc. of APEC*, 2013.

[15] J. Liu, H. Shen, and X. Zhang, "A survey of mobile crowdsensing techniques: A critical component for the internet of things," in *Proc. of ICCCN*, 2016.

[16] L. Yan and H. Shen, "TOP: vehicle trajectory based driving speed optimization strategy for travel time minimization and road congestion avoidance," in *Proc. of MASS*, 2016.

[17] Y. Wu, Y. Zhu, and B. Li, "Trajectory improves data delivery in vehicular networks," in *Proc. of INFOCOM*, 2011.

[18] F. Xu, S. Guo, J. Jeong, Y. Gu, Q. Cao, M. Liu, and T. He, "Utilizing shared vehicle trajectories for data forwarding in vehicular networks," in *Proc. of INFOCOM*, 2011.

[19] D. Barbará, Y. Li, and J. Couto, "Coolcat: an entropy-based algorithm for categorical clustering," in *Proc. of CIKM*, 2002.

[20] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and pois," in *Proc. of SIGKDD*, 2012.

[21] A. Karalis, J. D. Joannopoulos, and M. Soljačić, "Efficient wireless non-radiative mid-range energy transfer," *Annals of Physics*, 2008.

[22] A. Sarker, C. Qiu, H. Shen, A. Gil, J. Taiber, M. Chowdhury, J. Martin, M. Devine, and A. Rindos, "An efficient wireless power transfer system to balance the state of charge of electric vehicles," in *Proc. of ICPP*, 2016.

[23] "Openstreetmap," http://www.openstreetmap.org/, accessed July, 2016.

[24] "Shenzhen wiki," https://en.wikipedia.org/wiki/Shenzhen, accessed July 24, 2016.

[25] Z. Li and H. Shen, "Designing a hybrid scale-up/out hadoop architecture based on performance measurements for high application performance," in *Proc. of ICPP*, 2015.

[26] C. Qiu, H. Shen, and L. Chen, "Probabilistic demand allocation for cloud service brokerage," in *Proc. of INFOCOM*, 2016.

[27] L. Chen, H. Shen, and S. Platt, "Cache contention aware virtual machine placement and migration in cloud datacenters," in *Proc. of ICNP*, 2016.

[28] B. Wu and H. Shen, "Analyzing and predicting news popularity on twitter," *Int J. Information Management*, vol. 35, no. 6, 2015.

[29] "Apache spark 1.5.2," http://spark.apache.org/, accessed July 24, 2016.

[30] Y. Zheng, "Trajectory data mining: an overview," *TIST*, 2015.

[31] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang, "T-drive: driving directions based on taxi trajectories," in *Proc. of GIS*, 2010.

[32] K. Pearson, "Note on regression and inheritance in the case of two parents," in *Proc. of the Royal Society of London*, 1895.

[33] D. Zhang, J. Huang, Y. Li, F. Zhang, C. Xu, and T. He, "Exploring human mobility with multi-source data at extremely large metropolitan scales," in *Proc. of MobiCom*, 2014.

[34] R. G. Michael and S. J. David, "Computers and intractability: a guide to the theory of NP-completeness," *W. H. Freeman*, 1979.

[35] K. Chen and L. Liu, "The 'Best K' for entropy-based categorical data clustering," in *Proc. of SSDBM*, 2005.

[36] M. P. Wand and M. C. Jones, *Kernel smoothing*. CRC Press, 1994.

[37] R. Li, G. Rose, and M. Sarvi, "Using automatic vehicle identification data to gain insight into travel time variability and its causes," *JTRB*, 2006.

[38] R. Li, H. Chai, and J. Tang, "Empirical study of travel time estimation and reliability," *Mathematical Problems in Engineering*, 2013.

[39] "Glpk," https://www.gnu.org/software/glpk/, accessed July 24, 2016.

[40] "Cbc," https://projects.coin-or.org/Cbc, accessed July 24, 2016.