Reinforcement Learning based Scheduling for Cooperative EV-to-EV Dynamic Wireless Charging

Li Yan*, Haiying Shen*, Liuwang Kang*, Juanjuan Zhao† and Chengzhong Xu‡
*Department of Computer Science, University of Virginia, USA
†Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China
‡State Key Lab of IoTSC and Dept of Computer Science, University of Macau, China
Email: {ly4ss, hs6ms, lk2sa}@virginia.edu, jj.zhao@siat.ac.cn, czxu@um.edu.mo

***Abstract*—Previous Electric Vehicle (EV) charging scheduling methods and EV route planning methods require EVs to spend extra waiting time and driving burden for a recharge. With the advancement of dynamic wireless charging for EVs, Mobile Energy Disseminator (MED), which can charge an EV in motion, becomes available. However, existing wireless charging scheduling methods for wireless sensors, which are the most related works to the deployment of MEDs, are not directly applicable for the scheduling of MEDs on city-scale road networks. We present *MobiCharger*: a *Mobi*le wireless *Charger* guidance system that determines the number of serving MEDs, and the optimal routes of the MEDs periodically (e.g., every 30 minutes). Through analyzing a metropolitan-scale vehicle mobility dataset, we found that most vehicles have routines, and the temporal change of the number of driving vehicles changes during different time slots, which means the number of MEDs should adaptively change as well. Then, we propose a Reinforcement Learning based method to determine the number and the driving route of serving MEDs. Our experiments driven by the dataset demonstrate that *MobiCharger* increases the medium state-of-charge and the number of charges of all EVs by 50% and 100%, respectively.**

I. Introduction

Due to limited battery capacity [1], [2], the continuous running of EVs is a major problem. Many EV charging scheduling methods that aim to optimize the charging efficiency of EVs in stationary charging stations have been proposed [1], [3]–[7]. Generally, these methods focus on utilizing various statistical models (e.g., Hidden Markov Chain for describing charging demand) and controlling methods (e.g., multi-objective optimization) to recommend the optimal target charging stations to EVs to minimize charging time cost (i.e., time wasted in driving and waiting until charging complete), based on the status information of charging stations such as location and charger availability. However, due to the limited number of stationary charging stations in a city, EVs may still need to wait for a long time during charging peak hours [8].

To overcome the problem, many EV route planning methods have been proposed [8]–[11]. The main idea of these works is to design routing methods to give routing recommendation to an EV, which satisfies multiple energy constraints of the EV. For example, the EV's charging time cost must be shorter than a certain threshold, the EV's energy consumption and travel time should be minimized. Although these methods can generate the optimal path to a the destination of an EV considering multiple constraints, the EV still need to compromise its original driving route and spend extra driving burden in fulfilling its charging requests.

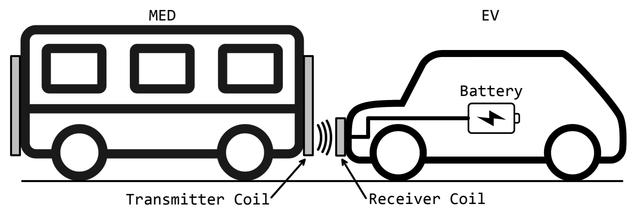

Fig. 1. Illustration of MED charging.

Recently, Mobile Energy Disseminator (MED) [12]–[14], which can charge an EV in motion, becomes available. As shown in Figure 1, an MED is equipped with several Transmitter Coils and drives independently as a mobile charging station. This charging approach is defined as *cooperative EV-to-EV (V2V) dynamic wireless charging* [15]. If an MED can proactively encounter and charge an EV, the EV will not need to spend extra time to drive to a charging station to receive a recharge and its state-of-charge (SoC) can always be maintained via encountering with different MEDs during its driving. Deploying more MEDs in the road network can satisfy the demands of more EVs on roads but generates more monetary cost and also higher traffic on roads, and vice versa. Thus, an interesting problem is *how to consider the mobility records of the EVs (e.g., historical trajectories, SoC) and the limited battery capacity of MEDs to minimize the total number of MEDs, maximize the number of encountering EVs of all the MEDs per unit time (e.g., 15 minutes), and meanwhile ensure that each EV can maintain its SoC above 0?*

In Wireless Sensor Network (WSN) research area, many traditional works have been proposed to optimize the scheduling and routing of mobile chargers that move and replenish the sensors [16]–[21]. However, most of these methods are only applicable for small-scale static WSNs with around hundreds of sensors nodes due to their lack of effective methods on describing the dynamic change of city-scale vehicle traffic (e.g., driving trajectory, vehicle density) or adapting to the real-time change of the number of driving EVs (i.e., change of charging demands) when guiding the mobile chargers.

To solve the problem, we propose *MobiCharger*, a *Mobi*le wireless *Charger* guidance system that determines the number of serving MEDs, and the optimal routes of the MEDs periodically (e.g., every 30 minutes). First, we analyzed a long-term metropolitan-scale vehicle mobility dataset that records the trajectories of 15,610 taxicabs, which consist of 6,510 EVs, and 12,386 customized transit service vehicles (e.g., Uber, Lyft). In this paper, if an EV's ratio of driving a trajectory at around a specific time (the variance of the trajectory start times is no higher than a threshold (e.g., 1/2 hour)) during a time period (e.g., 30 days) is higher than a threshold (e.g., 20%), we define this trajectory the EV's *routine*. Our data analysis observations are as follows:

(1) We found that most taxicabs and customized transit service vehicles have routines, which is especially conspicuous for private vehicles due to people's daily routines [22], [23]. The combination of the vehicles' current driving trajectory (i.e., short-term mobility information) and routines (i.e., long-term mobility information) can be utilized to estimate the density of EVs that the MEDs can potentially offer charging.
(2) We found that the number of driving vehicles changes during different time slots (i.e., 30-minute-long), which means the number of required MEDs should also adaptively change.

The observations support the design of *MobiCharger*. *MobiCharger* combines the arrival probabilities of multiple vehicles to estimate the vehicle density of the given road segment during specific time durations from both long-term and available short-term mobility information. Based on the estimated vehicle density, *MobiCharger* schedules MED deployment (i.e., determining the number of MEDs and the driving route of each MED) by using Reinforcement Learning (RL). The inputs to the RL model include the status of each MED (SoC and reachable road segments) and the outputs include the guidance on where to drive to serve EVs for each MED. The reward is determined based on the number of serving MEDs, the number of EVs that all MEDs can possibly charge, and the estimated total distance that MEDs need to drive given certain assigned road segments. In summary, our contributions include:

(1) We analyze a metropolitan-scale taxicab dataset to gain insights on vehicle driving trajectories and temporal change of the number of driving EVs on the road network, which serve as the foundation for *MobiCharger* design.
(2) We propose *MobiCharger*, a *Mobi*le wireless *Charger* guidance system that determines the number of serving MEDs, and the optimal routes of the MEDs periodically (e.g., 30 minutes) that minimizes the total number of MEDs, maximizes the number of encountering EVs of all the MEDs, and meanwhile ensures that each EV can maintain non-zero SoC all the time.
(3) We conduct an extensive trace-driven evaluation of *MobiCharger*. Compared with the previous methods, *MobiCharger* increases the medium state-of-charge and the number of charges of all EVs by 50% and 100%, respectively.

In our knowledge, *MobiCharger* is the first to optimize the driving paths of MEDs to minimize the total number of MEDs, maximize the number of encountering EVs of all the MEDs, and meanwhile ensure that each EV has sufficient SoC all the time. 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 *MobiCharger*. Section V presents performance evaluation results. Section VI concludes the paper with remarks on our future work.

II. Related Work

Charging Scheduling of EVs. Many works that focus on optimizing the scheduling of EVs have been proposed. Based on the analysis of charging and discharging processes between power grid and EVs, Yu *et al.* [6] proposed a coalitional game to evenly distribute charge load. Malandrino *et al.* [7] formulated a game theory based method to optimize the charging load considering EVs' charging positions, charging time and charger availability. Tian *et al.* [5] proposed to use each EV's historical charging events, real-time trajectories and current traffic state to recommend charging stations with the minimal time cost. Tan *et al.* [4] developed a distributed optimization algorithm to optimize the demand side management problem for the future smart grid with EVs. Kang *et al.* [1] proposed a charging strategy by considering optimal charging priority and charging location based on electric price. Bashash *et al.* [3] proposed a convex quadratic programming framework for the charge pattern optimization of EVs under time-varying electricity prices. However, since multiple EVs may simultaneously compete for the same charging station during charging peak hours, some EVs may have to spend extra time to wait for their turn, which greatly reduces the charging service efficiency [8]. Cooperative V2V dynamic wireless charging can potentially mitigate this problem, but an MED guidance system that minimizes the total number of serving MEDs, maximizes the number of the MEDs' encountering EVs, and meanwhile prevents each EV from SoC exhaustion is necessary.

Route Planning for EVs. Many works EV route planning methods have been proposed. Eisner *et al.* [9] proposed to consider the EV's battery capacity in the calculation of edge costs (e.g., energy consumption, distance) to optimize the EVs' driving route. Sachenbacher *et al.* [10] further proposed to dynamically adjust the edge cost according to the specific parameter of an EV (e.g., weight, aerodynamic resistance) in generating the optimal route. Schneider *et al.* [8] proposed to combine customer appearance time, customer demand, battery capacity constraints and charging station locations, to minimize the EV's energy consumption. Sarker *et al.* [11] utilized Autoregressive Integrated Moving Average (ARIMA) model to predict traffic volume, and determined the optimal route that minimizes energy consumption, travel time, charging cost, and range anxiety. However, they focus on generating the optimal path from the EV's current position to the EV's destination, which creates extra driving burden on the EVs. Cooperative V2V dynamic wireless charging can potentially mitigate this problem, but an MED guidance system described above is needed, which is the focus of this paper.

III. Metropolitan-Scale Trace Data Analysis

A. Dataset Description and Definitions

In this analysis, we use the data recorded from Jan 1, 2015 to Dec 31, 2015 for measurement, which include:

(1) **Taxicab Dataset.** This dataset records the status (e.g., timestamp, GPS position, velocity, occupancy) of 15,610 taxi-

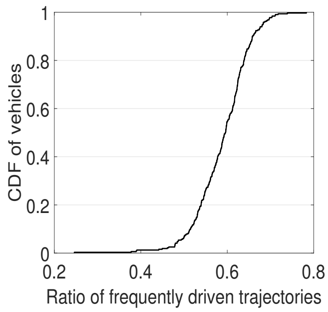

Fig. 2. Distribution of the ratios of frequently driven trajectories.

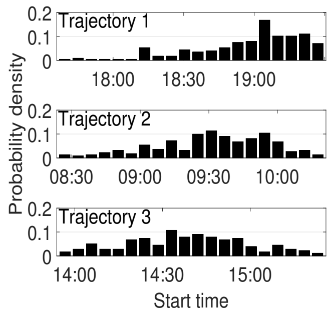

Fig. 3. Distribution of the start times of three vehicle trajectories.

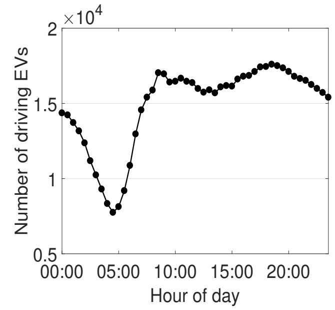

Fig. 4. Number of driving EVs over time.

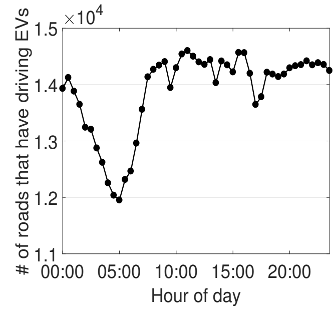

Fig. 5. Number of road segments that have driving EVs over time.

cabs. 6,510 of them are electric taxicabs. This dataset is used for analyzing the mobility characteristics of EVs.

(2) **Dada Car Dataset.** This dataset records the status (e.g., timestamp, position, velocity) of 12,386 electric Dada Cars (a customized transit service similar to UberPool). This dataset is also used for analyzing the mobility characteristics of EVs.

(3) **Charging Station Dataset.** This dataset records the GPS position and number of chargers of 81 existing plug-in charging stations in Shenzhen. This dataset is used for simulating the recharge of MEDs in experiment.

(4) **Road Map.** The road map of Shenzhen is obtained from OpenStreetMap [24]. We use a bounding box with a south-west coordinate (lat=22.4450, lon=113.7130), and a north-east coordinate (lat=22.8844, lon=114.5270) to crop the data.

Definition 1: Vehicle Trajectory. A vehicle trajectory is a sequence of N^e time-ordered landmarks, where each landmark is represented by a latitude and a longitude.

Definition 2: Vehicle Routine. A vehicle's trajectory is a routine if the vehicle's probability of driving the trajectory at around a specific time (the standard deviation of the start times is no higher than the threshold (i.e., 35 minutes in our analysis)) during a time period (i.e., 365 days in our analysis) is higher than a threshold (i.e., 20% in our analysis).

B. Dataset Analysis

1) **Observation 1**: Existence of Vehicle Driving Routines*:* The Cumulative Density Function (CDF) of the calculated ratios of of all the vehicles is illustrated in Figure 2. We can see that for about 80% of the vehicles, more than 55% and maximally about 77% of their trajectories are their frequently driven trajectories. This result implies that most vehicles do frequently drive several similar trajectories in different days. Besides, we are also interested in the start time of the frequently driven trajectories because knowing the trajectories and their start time can exactly tell us where and when the vehicles are likely to appear. Driven by this motivation, we collected the distribution of the start times of each frequently driven trajectory, and randomly selected the results of three trajectories for illustration. Figure 3 shows the histogram of the start times of each selected trajectory. The first trajectory (Trajectory 1) is generally driven at around 19:00 every day, the second trajectory (Trajectory 2) is generally driven at around 09:30 every day, and the third trajectory (Trajectory 3) is generally driven at around 14:30 every day. We can see that most start times of Trajectory 1 vary within 20 minutes around 19:00 and most start times of Trajectory 3 vary within 30 minutes around 14:30. In contrast, the start times of Trajectory 2 vary over a wider range (as long as 1 hour) around 09:30. This means that the start times of Trajectory 1 and Trajectory 3 vary less than those of Trajectory 2.

2) **Observation 2**: Temporal Change of the Number of Driving EVs*:* Figure 4 shows the calculated average number of driving EVs in each hour of a day. We can see that during the time slots between around 02:00 and 05:00, the number of driving EVs significantly dropped to less than 8,000. This is primarily because that during these time slots, the human transit demand for taxicabs and Dada cars greatly decreased. Some taxicabs and Dada cars chose to stop running during these time slots to save cost. Then the number of driving EVs quickly increased to around 18,000 at around 09:30, which corresponds to the "rush hour" in the morning of a day. At around 12:00, there is a slight drop-down on the number of driving EVs due to reduced human transit demand at noon. Then the number of driving EVs increased to another peak at around 19:30, which corresponds to the "rush hour" in the evening of a day. These results demonstrate that the number of EVs that the MEDs need to support to keep running on the road network varies during different time slots.

We measured the number of road segments that have driving EVs during each time slot in each day throughout the 365 days. Figure 5 shows the average number of road segments versus hour of day. We can see that the change of the number of road segments that have driving EVs is generally similar as the change of the number of driving EVs in Figure 4. However, there are several conspicuous drops of the number of road segments that have driving EVs at around 10:00, 13:00 and 17:00. This may be because that there are not many passengers requesting transit during these time slots, so some taxicabs and Dada cars chose to wait at some places with high likelihood of passenger appearance to save driving cost [25]. From these results, we can conclude that in addition to adjusting the number of the serving MEDs, we should also optimize the driving route of the MEDs.

IV. SYSTEM DESIGN

A. Framework of MobiCharger

1. Vehicle mobility information derivation. First, we clean the raw positions by filtering out the positions that are not in the Shenzhen area. Then, we reorder the cleaned positions by timestamp and map them on the *Roadmap with Landmarks and Road Segments* to generate the *Trajectories Represented in Landmarks* of EVs as explained in Section III-A.

2. **EV Traffic Density Estimation** (Section IV-B). Following the first stage, we combine the EVs' current *Trajectories Represented in Landmarks* and routines to complete *EV Traffic Density Estimation* for each road segment of the road network.
3. **Reinforcement Learning based MED Routing** (Section IV-C). Based on the real-time status of EVs, we apply the *EV Traffic Density Estimation* to determine the road segments with new changes of EV traffic. Then we train and utilize the *Reinforcement Learning based MED Routing* to decide the place each MED should drive to as a response to the traffic change, and the place can be the originally planned road segment, a new road segment or an MED's nearest parking lot.

B. EV Traffic Density Estimation

Most previous vehicle density estimation works [26], [27] fail to estimate the movement of individual vehicles (i.e., arrival time on any position of the vehicle's driving route). Since the future vehicle density estimation accuracy of a road segment is determined by the arrival estimation of the vehicles that will drive through the road segment in the near future (e.g., 5 minutes), accurate estimation of each individual vehicle's movement will help generate a more accurate vehicle density estimation. Therefore, we need a more accurate method to estimate vehicle density in the near future. As defined in Definition 2, each EV's current trajectory demonstrates the EV's approximate position in the near future (e.g., 15 minutes). If we combine the trajectories of all the EVs, we may be able to deduce the density of EVs at each road segment during specific time slots. However, even if a vehicle's trajectory from its origin to destination is determined, the EV's actual time spent on traveling through the road segments in the trajectory may vary within a certain range due to the dynamic change of traffic status. That is, the travel time of a road segment may follow a certain statistical distribution. If we can properly model the statistical distribution of the travel time, we can estimate each EV's probability of appearing on the road segments included in its current trajectory. Then by summing up all the EVs' probabilities of appearing on the road segment during specific time slots, we can predict the density of EVs on each road segment during the time slots. The key problem here is: *how to model the statistical distribution of road segment travel time and utilize the model to predict the future EV density of each road segment?*

It has been confirmed by several previous methods that the Gamma distribution can be utilized to model the distribution of the travel time of a road segment with a sufficiently high accuracy [28], [29]. We build our EV density estimation method based on the modeling results of these methods. Suppose the Gamma distribution of the travel time of a road segment e_i is represented as $T_i \sim \Gamma(\kappa_{e_i}, \theta_{e_i})$, where parameter κ_{e_i} determines the shape of Γ, and θ_{e_i} determines the scale of Γ. Specifically, the travel time distribution of each road segment can be obtained through calculating the EVs' travel time reflected in historical trajectory data. According to the characteristics of Gamma distribution [28], [30], the parameters κ_{e_i} and θ_{e_i} are computed with the mean travel time $E[T_i] = \mu_i$ and the travel time variance $Var[T_i] = \sigma_i^2$ using the relationship among the mean $E[T_i]$, the variance $Var[T_i]$, κ_{e_i}, and θ_{e_i} such that: $E[T_i] = \kappa_{e_i}\theta_{e_i}$, and $Var[T_i] = \kappa_{e_i}\theta_{e_i}^2$ for $T_i, \kappa_{e_i}, \theta_{e_i} > 0$.

Therefore, the parameter θ_{e_i} can be computed as the road segment's mean travel time divided by the road segment's travel time variance: $\theta_{e_i} = \frac{Var[T_i]}{E[T_i]} = \frac{\sigma_i^2}{\mu_i}$. Similarly, the parameter κ_{e_i} can be computed as the road segment's mean travel time divided by the calculated parameter θ_{e_i}: $\kappa_{e_i} = \frac{E[T_i]}{\theta_{e_i}} = \frac{\mu_i}{\theta_{e_i}} = \frac{\mu_i^2}{\sigma_i^2}$.

Based on the previous works that utilize vehicle trajectories for estimating vehicle driving delays [28], we suppose that the travel times of the road segments included in the trajectory are statistically independent random variables following the Gamma distribution. Thus, we estimate the total travel time of the trajectory of the k^{th} EV by summing up the travel times of the road segments included in the trajectory from the EV's current position p_k to e_i, i.e., $T_{ki} = \sum_{j=p_k}^{e_i} T_i$. The mean and variance of the statistical distribution of T_{ki} can be computed as: $E[T_{ki}] = \sum_{i=1}^{N^e} E[T_i] = \sum_{i=1}^{N^e} \mu_i$ and $Var[T_{ki}] = \sum_{i=1}^{N^e} Var[T_i] = \sum_{i=1}^{N^e} \sigma_i^2$

Thus, the statistical distribution of T_{ki} can be modeled as $T_{ki} \sim \Gamma(K_i, \Theta_i)$, where the parameters K_i and Θ_i are computed as in κ_{e_i} and θ_{e_i}. Then, we can use this $\Gamma(K_i, \Theta_i)$ to estimate the appearance probabilities of all the EVs at a road segment during a specific time duration. Specifically, given current time t^c and a specific time duration with start time t^s and end time t^e, which is denoted as $[t^s, t^e]$, (e.g., [12:00,12:05]), an EV's appearance probability at e_i during $[t^s, t^e]$ is measured by estimating the EV's travel time to the end of road segment e_i:

$$P(t^s - t^c \leqslant T_{ki} \leqslant t^e - t^c) = F(t^e - t^c; K_i, \Theta_i) - F(t^s - t^c; K_i, \Theta_i) \quad (1)$$

where T_{ki} is the actual travel time of the trajectory connecting current position p_k and e_i, and $F(\cdot)$ is the CDF of the Gamma distribution with parameters K_i and Θ_i. The EV density of each e_i in the time duration is estimated by summing up the appearance probabilities of all the EVs that will pass e_i during the time duration:

$$d_{e_i}^v = \sum_{j=1}^{N_i^v} P_j(t^s - t^c \leqslant T_i \leqslant t^e - t^c), \quad (2)$$

where N_i^v is the number of EVs whose trajectories demonstrate that they will pass e_i during $[t^s, t^e]$.

C. Reinforcement Learning based MED Routing

The data analysis result in Section III-B2 demonstrates that the number of driving EVs has rise and fall during different time slots and changes in real time. Thus, we need to solve an important problem: *how to efficiently optimizes the driving route of MEDs according to the short-term change of some EVs' trajectories?*

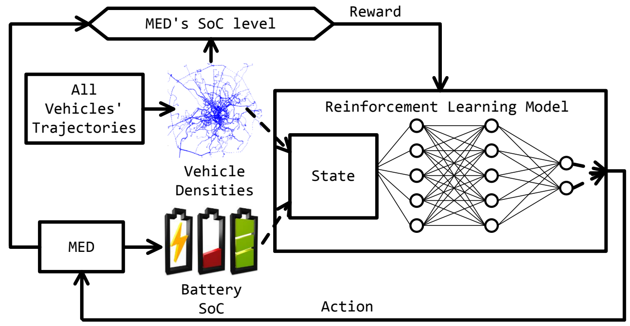

Fig. 6. Training of reinforcement learning model.

We propose a Reinforcement Learning (RL) based routing model to determine the action of all the MEDs, which include MEDs driving on the road network and the MEDs standing by in the parking lots, based on the real-time EV traffic status. We define an *action* as the driving decisions of all the MEDs, which include serving MEDs and out-of-service MEDs. The driving decision of an MED is which road segment the MED will drive to or driving to the MED's nearest parking lot. The MEDs chosen to take action will drive to their respective road segments to serve EVs or the nearest parking lots to stand by. The RL model will be utilized for MED routing only when the real-time EV traffic has significant change.

Generally, given a current state s_t describing the status of the road network (e.g., current positions of each MED and each EV), the RL model outputs an action a_t that maximizes the reward. We define the *reward* caused by an action of the MEDs as the weighted sum of the estimated total number of EVs that can be served by the MEDs, the sum of the MEDs' service latency and the number of serving MEDs. We use the MED's travel time of the route connecting its current position and the end position on its destination road segment as the service latency of the MED. As long as the RL model's policy $\pi : s_t \mapsto a_t$ is optimized, we can use it to output the optimal action that maximizes the reward for guiding the movement of the MEDs. Driven by this reward, the RL model will choose the minimum number of MEDs nearby the road segments with new short-term EV traffic change for service and determine the chosen MEDs' driving routes. The chosen MEDs will serve the new EVs that they encounter when driving in their routes. The rest MEDs won't take any action, but keep following their original driving routes or drive to their nearest parking lots to stand by. In the following, we introduce the state, action and reward of our RL model.

1) State Set: Since the RL model is expected to output the driving route for each MED, which maximizes the total number of EVs that can be encountered by all the MEDs, we take into account the estimated number of encountering EVs of each MED given a certain destination road segment for the MED. For example, for the k^{th} MED (denoted by m_k), we first determine the shortest trajectory connecting the MED's current position (denoted by p_k) and the end of each road segment in the road network ($e_j \in E$) using the Dijkstra algorithm [31], which is denoted as $\Phi_{kj} = \{p_k, \ldots, e_j\}$. We then calculate the number of encountering EVs in each trajectory. Based on MED m_k's current SoC and the energy consumption of a trajectory, we can find the road segments that MED m_k cannot reach. That means MED m_k will not be chosen to drive to these road segments, and we set the number of encountering EVs for these road segments to 0. That is, driving to such a road segment produces no reward in the RL model. To calculate the number of encountering EVs in a trajectory, we calculate the sum of the EV densities of the road segments in the trajectory. Specifically, for each road segment e_i in the trajectory, we estimate the general arrival timestamp range $[t_i^s, t_i^e]$ of the MED. For each e_i, we utilize Equation (2) to calculate the EV density $d_{e_i}^v$ during the arrival timestamp range $[t_i^s, t_i^e]$. Note that the EV density is calculated based on reported short-term mobility information of all the EVs, so dynamic change of the EV traffic has been reflected in their short-term mobility information because it shows the actual driving route of the EVs. Finally, we use the sum of all the EV densities of the road segments $\mathcal{D}_k = \sum_{e_i \in \Phi_{kj}} d_{e_i}^v$ as the k^{th} number of encountering EVs if the MED drives to e_j.

Moreover, since an MED's current position determines its possible service latency to other candidate road segments and we expect the RL model to choose the serving MEDs that have the shortest service latency, we also need to include current position of each MED in formulating the state set. Thus, the state set S consists of current position of each MED and each MED's estimated numbers of EVs that the MED may encounter in driving its trajectory, which is defined as:

$$s = (p_k, \mathcal{D}_k \mid \forall\, m_k \in \mathcal{M}) \in S \tag{3}$$

where s denotes the state. p_k is current position of the k^{th} MED. $\mathcal{D}_k = \{\sum_{e_i \in \Phi_{kj}} d_{e_i}^v | e_j \in E\}$ is the estimated potential number of EVs that the k^{th} MED can charge if it drives to e_j. $e_j \in E$ means that $\mathcal{D}_k$ is calculated for the total set of road segments E in the road network. m_k denotes the k^{th} MED. M denotes the set of all the MEDs.

2) Action Set: The action output by the RL model determines where to drive to for each MED. Specifically, the action of the k^{th} MED, which is denoted as x_k, has three possibilities: driving to a destination road segment (denoted by $x_k = e_j \in E$), driving to the parking lot nearest to its current position ($x_k = 1$), or follow its original driving decision ($x_k = 0$). If $x_k = e_j \in E$ or $x_k = 1$, we determine the k^{th} MED's shortest driving route that connects its current position and the end of the destination road segment e_j or the nearest parking lot by using an existing routing method (e.g., Dijkstra algorithm [31]). If $x_k = 0$, the MED will not change its current driving decision, but keep following its original driving route. Thus, the action set A is defined as:

$$a = (x_k \mid \forall\, m_k \in \mathcal{M}) \in A, \tag{4}$$

where a denotes the action vector that indicates where to drive to for each MED. x_k denotes the action of the k^{th} MED. m_k represents the k^{th} MED. M represents the set of all the MEDs.

3) Reward: We expect the MEDs to cover the road segments with new EV traffic as early as possible. That is, the MEDs' service latency is expected to be minimized. Therefore, the main goal of the RL model is to maximize the total number of EVs served by all the serving MEDs, and meanwhile minimize the number of the serving MEDs and the sum of all the MEDs' service latency. The reward function resulted by the x_k of all the MEDs is defined as the weighted sum of the total number of potential EVs that the MEDs can serve, the sum of the MEDs' service latency and the number of serving MEDs, which is formulated as:

$$r(s_t, a_t, s_{t+1}) = \alpha N^v - \beta T^d - \gamma N^m \tag{5}$$

where s_t is the state at current time; s_{t+1} is the next state caused by the action a_t; $|\mathcal{M}|$ is the maximum number of MEDs that can be used for charging service; The reward is determined by N^v, T^d and N^m. N^v is the total number of EVs that can be encountered by all the MEDs. Under a certain action and current state, the more EVs that can be encountered by the MEDs, the higher reward the action will result in. T^D is the service latency of the MEDs. Under a certain action and current state, the shorter driving latency that all the serving MEDs suffer from, the higher reward the action will result in. N^m denotes the number of serving MEDs. The fewer MEDs are chosen to serve, the higher reward the MEDs will contribute. We use the constants α, β and γ ($\alpha, \beta, \gamma \in [0, 1]$ and $\alpha+\beta+\gamma = 1$) as the weights of the metrics N^v, T^d and N^m, respectively.

The total number of encountering EVs metric is the total number of EVs that all the MEDs will encounter by taking different driving decisions. When determining where to drive to for each MED, we do not expect that too many MEDs from driving the same road segments during the same time because this reduces the utilization efficiency of the MEDs' service ability. Therefore, we globally estimate the total number of EVs that can be encountered by all the MEDs on the road network. Specifically, we first determine the union set of the road segments covered by all the MEDs' driving routes, which is denoted as $E' = \bigcup_{m_k \in \mathcal{M}} \Phi_{kj}$. Then, we calculate the sum of the EV densities of all the road segments in E' as the total number of EVs that can be encountered by the MEDs. That is, this metric is calculated as:

$$N^v = \sum_{e_i \in E'} d^v_{e_i}, \tag{6}$$

where e_i denotes a road segment included in E'. Note that N^v is calculated from the state s_t and the action of choosing certain MEDs to drive to e_i. The more EVs encountered by the MEDs, the higher reward that the action will result in.

The MED service latency metric is the sum of the service latencies of all the MEDs $T^d = \sum_{k=1}^{|\mathcal{M}|} T^d_k$, where T^d_k is the service delay of the k^{th} MED. Specifically, the k^{th} MED's service latency to e_j if it is chosen to drive to the end of the road segment e_j ($x_k = e_j \in E$), which is denoted with $t_{kj} = \sum_{e_i \in \Phi_{kj}} \frac{l_{e_i}}{v_i}$, where l_{e_i} is the length of e_i, and v_i is the speed limit of e_i. Note that if the k^{th} MED takes no action ($x_k = 0$), its service latency is the driving time of its original driving route Φ_{kd}, which connects its current position (p_k) and the end of its original destination road segment (e_d). That is, its service latency is calculated as: $t_{kd} = \sum_{e_i \in \Phi_{kd}} \frac{l_{e_i}}{v_i}$. or it drives to the nearest parking lot ($x_k = 1$), the effect of the k^{th} MED's service latency is 0. Thus, T^d_k is calculated as:

$$T^d_k = \begin{cases} t_{kj}, & \text{if } x_k = e_j \in E \\ t_{kd}, & \text{if } x_k = 0 \\ 0, & \text{if } x_k = 1, \end{cases} \tag{7}$$

where t_{kj} and t_{kd} are determined based on the k^{th} MED's current position, which is from the state s_t, and the new destination road segment or original destination road segment indicated by the k^{th} MED's driving decision (x_k), which is from the action a_t. The shorter driving latency the MEDs will have, the higher reward the MEDs will contribute.

The MED serving metric is the total number of MEDs chosen to continue service $N^m = \sum_{k=1}^{|\mathcal{M}|} b^m_k$, where b^m_k denotes whether the k^{th} MED is chosen to continue its service. If $x_k = e_j \in E$ or $x_k = 0$, the k^{th} MED is chosen to drive to a new destination road segment to cover the new short-term EV traffic change or take no action but to continue their original driving route, respectively. In both cases, the k^{th} MED will continue its service (i.e., $b^m_k = 1$). If the k^{th} MED is chosen to drive to the nearest parking lot ($x_k = 1$), the k^{th} MED will stop its service (i.e., $b^m_k = 0$). Therefore, b^m_k is calculated as:

$$b^m_k = \begin{cases} 1, & \text{if } x_k = e_j \in E \text{ or } 0 \\ 0, & \text{if } x_k = 1, \end{cases} \tag{8}$$

where x_k is the k^{th} MED's driving decision, which is from the action a_t. The fewer serving MEDs are required, the higher reward the MEDs will contribute because fewer serving MEDs will result in fewer cost.

4) Obtaining the Optimal Policy: We use the Deep Neural Network (DNN) to obtain the optimal policy as in [32]. Given a state s_t, the optimal policy π^* is defined as the map $\pi^* : s_t \mapsto a_t$ that maximizes the reward received by taking the corresponding action a_t. To discover the optimal action strategy that maximizes the reward under various states, we need historical movement data of MEDs for offline training of the RL model. However, there is no historical movement data of MEDs for training directly. Therefore, we collect movement data of MEDs through simulating the random movement of MEDs for a long time. Specifically, we record the generated MED routes, the positions of each MED and the change of EV densities of all the road segments with a short recording period (e.g., 1 minute). After recording for a relatively long time (e.g., 30 days), which is defined as collection period, we use the MED movement data collected in the collection period to train the RL model. Then, after every collection period, we repeat the training process with the newly collected MED movement data. Note that in the first collection period, the RL model is not yet trained, we will only rely on collecting the historical MED movement data generated from the simulated random movement of MEDs. In the later collection periods, the historical MED movement data will be collected from the output of the RL model.

V. Performance Evaluation

A. Comparison Methods

We compare the performance of *MobiCharger* (*MC* in short) in maintaining the EVs' SoC with a representative EV charging scheduling method [5] (*Schedule* in short), a representative Multi-Objective Route Planning method [11] (*MORP* in short), and our previous work for optimizing the charging efficiency of EVs [33] (*WPT-Opt* in short). Specifically, *Schedule* takes into account the historical charging preference, real-time trajectory of each EV, current road traffic state and charging station availability to schedule target charging stations for the EVs in order to minimize the charging time of the EVs (i.e., charger seeking time before reaching the target charging station plus the waiting time at the target charging station). *MORP* considers the change of traffic volume and determines the optimal route that minimizes energy consumption, travel time, charging monetary cost on the way, and range anxiety for EVs. *WPT-Opt* utilizes game theory to maximally avoid congestion of EVs while minimizing the EVs' charging time cost to their target chargers. To make the methods comparable, we assume that they all use wireless chargers with the same charging rate. The difference is that in *Schedule* and *MORP*, there are only charging stations equipped with wireless chargers. While in *MC*, there are only MEDs equipped with wireless chargers. The total number of chargers in the three methods are the same, and the distribution of chargers follows the existing charging stations in Shenzhen.

B. Experiment Settings

We suppose that every EV starts driving with a random SoC above a lower bound $SoC_{\min}$ at the beginning of a day. The SoC lower bound $SoC_{\min}$ is set to 20%. The battery capacities of the EVs follow a uniform distribution from 65 kWh to 85 kWh, which is the common battery capacity of the EVs currently in use in Shenzhen [5]. The charging rate of a charging infrastructure (in charging station or MED) is set to 150 kW [34]. The energy consumption rate of an EV is a 0.425 kWh/km [25]. We use the historical EV trajectory data from July, 2014 to June, 2015 to train the RL model that determines the route of the MEDs with the maximum reward. Specifically, we utilize Flow [35], which is a vehicle traffic simulation framework with the integration of deep RL, to implement the RL based MED routing method. Flow utilizes SUMO [36] to simulate the states and actions of EVs and utilizes DNN to train the optimal route of the MEDs with the maximum reward. We use a Deep Q Network with the following settings: 3 layers with 512 neurons in each layer; the activation function is Rectified Linear Unit (ReLU); the learning rate is 10^{-4}; the size of relay buffer is 10^5; the size of a batch sampled from replay buffer for training is 32; the discount factor is 0.99. We use SUMO [36] to simulate the movement of 1,000 EVs for 24 hours on Shenzhen's road network, and directly use the positions of existing charging stations in Shenzhen.

To collect training data, we simulate the movement of multiple MEDs to explore as many states and actions as possible to obtain the optimal policy. Specifically, 1) we suppose that each MED's initial state starts from a road segment randomly selected from current cruising graph $\widetilde{G}$ with $SoC = 1$; 2) we simulate the movement of each MED from one road segment to another road segment randomly selected from its candidate road segments based on its SoC, which means the MEDs transfer from one state to another state by taking different actions. 3) we utilize the historical vehicle trajectory information at each time when the MEDs take an action in the simulation to calculate the reward. The RL model calculates the cumulative reward resulted by each action at each state (i.e., defined as the Q value of the actions). After exploring over all the possible states and actions that appear during the training process, the RL model finds the optimal policy that maximizes the Q value for each possible state, starting from the initial state. The metrics we measured are:

- *The SoC of EVs*: In each time slot throughout a day, we measure the SoC of each EV, and calculate the medium, the 5^{th} percentile and the 95^{th} percentile values of all the EVs' SoC in each time slot.
- *The number of charges of EVs*: For each EV, we measure its number of received charges throughout a day. Then, we measure the CDF of the number of charges of all the EVs.
- *Energy supply overhead*: For each charger, we measure its amount of energy supplied to EVs in each time slot throughout a day. Then, we calculate the sum of all the chargers' energy supplied to EVs in each time slot.
- *Average vehicle flow rate*. We define the vehicle flow rate of a road segment as the number of vehicles driving through the road segment per unit time [31]. We measure the average vehicle flow rate of all road segments per hour in a day.

C. Experimental Results

1) The SoC of EVs: Figure 7 shows the medium, 5^{th} percentile and 95^{th} percentile values of all the EVs' SoC after every two hours in a day under different methods. We can see that the results of *MobiCharger* is much more stable than the other methods. The medium values generally follow: *MC*>*WPT-Opt*≈*Schedule*≈*MORP*.

In *Schedule*, *MORP* and *WPT-Opt*, all the EVs kept driving until their SoC is about to be exhausted. Therefore, from 00:00 to 04:00, most EVs' SoC kept dropping. By the time 04:00, most EVs began to run out of SoC (<20%) and drove to a charging station for a full recharge, which makes their medium SoCs return to around 0.8 at around 04:00. However, after 04:00, different EVs consume different amount of SoC and exhausted their SoC at different times, which causes the EVs' medium SoCs in these two methods to be very low. The reason behind this is that except for driving to a charging station, the EVs in the three methods do not have alternative method to replenish their SoC.

In contrast, the EVs' medium SoCs in *MC* remain high and are relatively more stable throughout the day. Generally, *MC* increases the medium SoC of all EVs by 50% during all time slots. This is because that the driving route of MEDs in *MC* fully considers the future trajectories and routines of the EVs and can create abundant charging opportunities for the EVs. To illustrate that the MEDs in *MC* proactively contributes much charging opportunity for the EVs, we further measure the number of charges of all the EVs throughout the day.

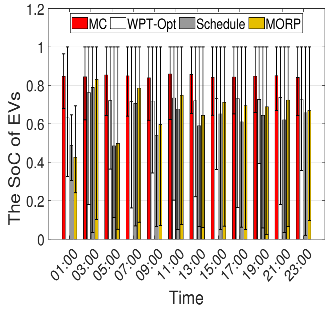

Fig. 7. EVs' SoC.

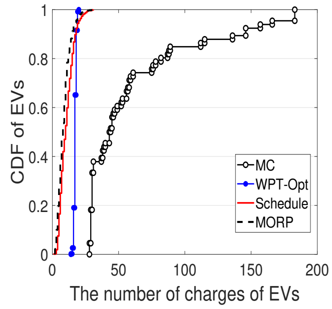

Fig. 8. Distribution of the number of charges of EVs.

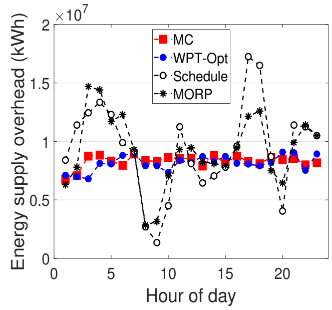

Fig. 9. Total energy supply overhead of all chargers per hour.

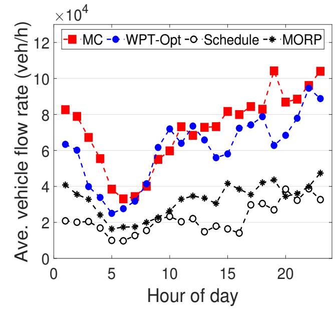

Fig. 10. Average vehicle flow rate of all road segments per hour.

2) The Number of Charges of EVs: Figure 8 shows the CDF of the numbers of charges of all the EVs under different methods. We can see that the results generally follow: *MC>WPT-Opt≈Schedule≈MORP*.

The results of *Schedule*, *MORP* and *WPT-Opt* are quite similar to each other. Around 80% of the EVs were charged for at least 5 times during a day. This is because that the EVs in these methods follow the "drive-exhaust-recharge" pattern, so the number of charges is equal to the number of SoC exhaustion of the EVs. Different EVs have different numbers of charges depending their different energy consumption.

In contrast, the numbers of charges of most EVs in *MC* are much higher than those in *Schedule* and *MORP*. Around 80% of the EVs received more than 25 charges in *MC*. Generally, *MC* increases the number of charges of EVs by almost 100%. This is because that the RL based routing method adjusts the MEDs to better cover the changed EV traffic in real time, which increases the numbers of charges of most EVs. These results show that the routing actions generated by *MC* can enable the MEDs to effectively cover the EV traffic. Combining the SoC results illustrated in Figure 7, we can see that *MC* can maintain the EVs' SoC above a high level through letting the MEDs proactively adapt to the EV traffic.

3) Energy Supply Overhead: Figure 9 shows the overall energy supply overhead of different methods under different hours throughout a day. The results follow: *MC≈WPT-Opt>Schedule≈MORP*.

From previous works [5], we know that the electric taxicabs and Dada cars in Shenzhen prefer to charge at around 03:00-05:00, 11:00-13:00, 16:00-18:00 and 20:00-22:00 (i.e., the EVs' charging time pattern). We can see that the energy supply overhead in *Schedule* and *MORP* is highly correlated with the EVs' charging time pattern. The energy supply overhead reaches the peak values at around 03:00-05:00, 09:00-13:00, 16:00-18:00 and 20:00-22:00, which matches the charging time pattern of the EVs in Shenzhen. This is because that these EVs simultaneously charge at the charging stations and create much energy supply overhead during these time slots. It is also worth noticing that during the busy hours when the EVs are driving (i.e., 06:00-11:00, 14:00-15:00 and 19:00-21:00), the energy supply overhead reaches the valley values. The results in *Schedule* and *MORP* illustrate that all the EVs follow the "drive-exhaust-recharge" pattern, which causes they exhaust and recharge at around the same time. Therefore, they follow almost the same charging time pattern, which may cause them to compete for the limited charging stations simultaneously.

In contrast, the energy supply overhead in *MC* and *WPT-Opt* is much more stable than that in *Schedule* and *MORP*. The energy supply overhead of these two methods only slightly rises at around 03:00-05:00, 09:00-13:00, 16:00-18:00 and 20:00-22:00 and slightly falls at around 06:00-11:00, 14:00-15:00 and 19:00-21:00, which is generally consistent with the results in *Schedule* and *MORP*. This is because that in *MC*, the MEDs have proactively charged many EVs even if their SoC did not exhaust. Thus, most EVs did not need to drive to nearby charging stations to receive a full recharge. While in *WPT-Opt*, the charging demand of EVs is evenly scheduled to avoid congestion at the chargers and on the way to the chargers, so the charging overhead is correspondingly evened. As a result, the energy supply overhead generated by the EVs during specific time slots is handled by the MEDs.

4) Average Vehicle Flow Rate: Figure 10 shows the average vehicle flow rate of all the road segments per hour under different methods. We can see the results follow: *MC>WPT-Opt>MORP>Schedule*. These results are generally consistent with the change of energy supply overhead of *Schedule* and *MORP* illustrated in Figure 9.

In *Schedule*, there is no specific method to avoid traffic congestion when generating the driving route for EVs. Therefore, during the time slots with high energy supply overhead, the vehicle flow rate is affected and decreases due to the EVs' competition for charging stations. In comparison, the routing of *MORP* considers the change of EV traffic, which effectively avoids the traffic congestion when generating the driving route for EVs. However, its average vehicle flow rate during time slots with high energy supply overhead is still lower than that in *MC*. In contrast, the average vehicle flow rate of *WPT-Opt* during the time slots when many EVs looked for recharge (i.e., 03:00-05:00, 09:00-13:00, 16:00-18:00 and 20:00-22:00) remains quite high. This is because that the game theory based approach maximally avoided the congestion at the chargers and on the way to the chargers.

We can observe that the vehicle flow rate of *MC* is higher than *WPT-Opt* during some time slots. This is because that the EVs have sufficiently high SoC during these time slots and do not need to compete for the limited charging stations, which may cause traffic congestion on the way to the charging stations. Therefore, the EVs can drive their original driving paths without traffic congestion, which results in the high average vehicle flow rate in *MC*. The exceptions happen at the time slots with busy traffic (i.e., 06:00-11:00, 14:00-15:00 and 19:00-21:00), which is due to the fact that the road network is congested by too many driving EVs during these time slots.

VI. Conclusion

Dynamic wireless charging for EVs enables an MED to charge an EV in motion. The deployment of MEDs adaptive to the change of the number of driving EVs is essential for maintaining the SoC of EVs. Our proposed *MobiCharger* is the first to optimize the driving paths of MEDs to minimize the total number of MEDs, maximize the number of encountering EVs of all the MEDs, and meanwhile ensure that each EV has sufficient SoC all the time. The design of *MobiCharger* is based on the observations obtained through analyzing a metropolitan-scale vehicle mobility dataset. We utilize the combination of EVs' current trajectories and the EVs' routines to estimate the density of EVs and the cruising graph that the MEDs should cover. Then, we develop an online method that utilizes RL to adjust the driving route of MEDs when the real-time vehicle traffic changes. We conducted trace-driven experiments on SUMO to verify the performance of *MobiCharger*. Compared with previous methods, *MobiCharger* increases the medium state-of-charge of all EVs by 50% during all time slots, and the number of charges of EVs by almost 100%.

Acknowledgements

This research was supported in part by U.S. NSF grants NSF-1827674, CCF-1822965, OAC-1724845, Microsoft Research Faculty Fellowship 8300751, The Science and Technology Development Fund, Macau SAR (File no. 0015/2019/AKP).

References

[1] Q. Kang, J. Wang, M. Zhou, and A. C. Ammari, "Centralized charging strategy and scheduling algorithm for electric vehicles under a battery swapping scenario," *IEEE TITS*, vol. 17, no. 3, 2016.

[2] C. Qiu, A. Sarker, and H. Shen, "Power distribution scheduling for electric vehicles in wireless power transfer systems," in *Proc. of SECON*, 2017.

[3] S. Bashash and H. K. Fathy, "Cost-optimal charging of plug-in hybrid electric vehicles under time-varying electricity price signals," *IEEE TITS*, vol. 15, no. 5, 2014.

[4] Z. Tan, P. Yang, and A. Nehorai, "An optimal and distributed demand response strategy with electric vehicles in the smart grid," *IEEE TSG*, vol. 5, no. 2, 2014.

[5] Z. Tian, T. Jung, Y. Wang, F. Zhang, L. Tu, C. Xu, C. Tian, and X.-Y. Li, "Real-time charging station recommendation system for electric-vehicle taxis," *IEEE TITS*, vol. 17, no. 11, 2016.

[6] R. Yu, J. Ding, W. Zhong, Y. Liu, and S. Xie, "PHEV charging and discharging cooperation in V2G networks: A coalition game approach," *IEEE ITJ*, vol. 1, no. 6, 2014.

[7] F. Malandrino, C. Casetti, and C.-F. Chiasserini, "A holistic view of its-enhanced charging markets," *IEEE TITS*, vol. 16, no. 4, 2015.

[8] M. Schneider, A. Stenger, and D. Goeke, "The electric vehicle-routing problem with time windows and recharging stations," *Transportation Science*, vol. 48, no. 4, 2014.

[9] J. Eisner, S. Funke, and S. Storandt, "Optimal route planning for electric vehicles in large networks," in *Proc. of AAAI*, 2011.

[10] M. Sachenbacher, M. Leucker, A. Artmeier, and J. Haselmayr, "Efficient energy-optimal routing for electric vehicles," in *Proc. of AAAI*, 2011.

[11] A. Sarker, H. Shen, and J. A. Stankovic, "MORP: Data-driven multi-objective route planning and optimization for electric vehicles," *Proc. of IMWUT*, vol. 1, no. 4, 2018.

[12] S. Moschoyiannis, L. Maglaras, J. Jiang, F. Topalis, and A. Maglaras, "Dynamic wireless charging of electric vehicles on the move with mobile energy disseminators," *IJACSA*, vol. 6, no. 6, 2015.

[13] D. Kosmanos, L. A. Maglaras, M. Mavrovouniotis, S. Moschoyiannis, A. Argyriou, A. Maglaras, and H. Janicke, "Route optimization of electric vehicles based on dynamic wireless charging," *IEEE Access*, vol. 6, 2018.

[14] L. A. Maglaras, F. V. Topalis, and A. L. Maglaras, "Cooperative approaches for dymanic wireless charging of electric vehicles in a smart city," in *Proc. of ENERGYCON*, 2014.

[15] R. Zhang, X. Cheng, and L. Yang, "Flexible energy management protocol for cooperative ev-to-ev charging," *IEEE TITS*, 2018.

[16] H. Dai, X. Wu, L. Xu, G. Chen, and S. Lin, "Using minimum mobile chargers to keep large-scale wireless rechargeable sensor networks running forever," in *Proc. of ICCCN*, 2013.

[17] A. Madhja, S. Nikoletseas, and T. P. Raptis, "Hierarchical, collaborative wireless charging in sensor networks," in *Proc. of WCNC*, 2015.

[18] S. Guo, C. Wang, and Y. Yang, "Joint mobile data gathering and energy provisioning in wireless rechargeable sensor networks," *IEEE TMC*, vol. 13, no. 12, 2014.

[19] L. He, L. Kong, Y. Gu, J. Pan, and T. Zhu, "Evaluating the on-demand mobile charging in wireless sensor networks," *IEEE TMC*, vol. 14, no. 9, 2015.

[20] L. Xie, Y. Shi, Y. T. Hou, W. Lou, H. D. Sherali, and S. F. Midkiff, "Multi-node wireless energy charging in sensor networks," *IEEE TON*, vol. 23, no. 2, 2015.

[21] Y. Shi, L. Xie, Y. T. Hou, and H. D. Sherali, "On renewable sensor networks with wireless energy transfer," in *Proc. of INFOCOM*, 2011.

[22] J. Zhao, Q. Qu, F. Zhang, C. Xu, and S. Liu, "Spatio-temporal analysis of passenger travel patterns in massive smart card data," *IEEE TITS*, vol. 18, no. 11, 2017.

[23] J. Zhao, C. Tian, F. Zhang, C. Xu, and S. Feng, "Understanding temporal and spatial travel patterns of individual passengers by mining smart card data," in *Proc. of ITSC*, 2014.

[24] "Openstreetmap," www.openstreetmap.org, 2020, accessed February, 2020.

[25] L. Yan, H. Shen, Z. Li, A. Sarker, J. A. Stankovic, C. Qiu, J. Zhao, and C. Xu, "Employing opportunistic charging for electric taxicabs to reduce idle time," *ACM IMWUT*, vol. 2, no. 1, 2018.

[26] J. Zhao, F. Zhang, L. Tu, C. Xu, D. Shen, C. Tian, X.-Y. Li, and Z. Li, "Estimation of passenger route choice pattern using smart card data for complex metro systems," *IEEE TITS*, vol. 18, no. 4, 2016.

[27] D. Deng, C. Shahabi, U. Demiryurek, L. Zhu, R. Yu, and Y. Liu, "Latent space model for road networks to predict time-varying traffic," in *Proc. of SIGKDD*, 2016.

[28] J. Jeong, S. Guo, Y. Gu, T. He, and D. H. Du, "TSF: Trajectory-based statistical forwarding for infrastructure-to-vehicle data delivery in vehicular networks," in *Proc. of ICDCS*, 2010.

[29] D. S. Berry and D. M. Belmont, "Distribution of vehicle speeds and travel times," in *Proc. BSMSP*, 1951.

[30] M. H. DeGroot and M. J. Schervish, *Probability and Statistics*. Pearson Education, 2012.

[31] Y. Zheng, "Trajectory data mining: An overview," *ACM TIST*, vol. 6, no. 3, 2015.

[32] H. Mao, R. Netravali, and M. Alizadeh, "Neural adaptive video streaming with pensieve," in *Proc. of SIGCOMM*, 2017.

[33] L. Yan and H. Shen, "Optimizing in-motion wireless charging service efficiency for electric vehicles: A game theoretic approach," in *Proc. of MASS*, 2019.

[34] 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.

[35] C. Wu, A. Kreidieh, K. Parvate, E. Vinitsky, and A. M. Bayen, "Flow: Architecture and benchmarking for reinforcement learning in traffic control," *arXiv*, 2017.

[36] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent development and applications of SUMO - Simulation of Urban MObility," *International Journal On Advances in Systems and Measurements*, vol. 5, no. 3&4, 2012.