Pricing-aware Real-time Charging Scheduling and Charging Station Expansion for Large-scale Electric Buses

GUANG WANG, ZHIHAN FANG, and XIAOYANG XIE, Rutgers University, USA
SHUAI WANG, Southeast University, USA
HUIJUN SUN, School of Traffic and Transportation, Beijing Jiaotong University, China
FAN ZHANG, Shenzhen Beidou Intelligent Technology Co., Ltd., China
YUNHUAI LIU, Beijing Institute of Big Data Research and Peking University, China
DESHENG ZHANG, Rutgers University, USA

We are witnessing a rapid growth of electrified vehicles due to the ever-increasing concerns on urban air quality and energy security. Compared to other types of electric vehicles, electric buses have not yet been prevailingly adopted worldwide due to their high owning and operating costs, long charging time, and the uneven spatial distribution of charging facilities. Moreover, the highly dynamic environment factors such as unpredictable traffic congestion, different passenger demands, and even the changing weather can significantly affect electric bus charging efficiency and potentially hinder the further promotion of large-scale electric bus fleets. To address these issues, in this article, we first analyze a real-world dataset including massive data from 16,359 electric buses, 1,400 bus lines, and 5,562 bus stops. Then, we investigate the electric bus network to understand its operating and charging patterns, and further verify the necessity and feasibility of a real-time charging scheduling. With such understanding, we design *busCharging*, a pricing-aware real-time charging scheduling system based on Markov Decision Process to reduce the overall charging and operating costs for city-scale electric bus fleets, taking the time-variant electricity pricing into account. To show the effectiveness of *busCharging*, we implement it with the real-world data from Shenzhen, which includes GPS data of electric buses, the metadata of all bus lines and bus stops, combined with data of 376 charging stations for electric buses. The evaluation results show that *busCharging* dramatically reduces the charging cost by 23.7% and 12.8% of electricity usage simultaneously. Finally, we design a scheduling-based charging station expansion strategy to verify our *busCharging* is also effective during the charging station expansion process.

CCS Concepts: • **Applied computing** → **Transportation**; • **Information systems** → *Location based services*; *Data analytics*;

Additional Key Words and Phrases: Electric bus, charging scheduling, data driven, charging pattern, MDP

This work is partially supported by NSF IIS 1849238, NSF CNS 1932223, NSF 1951890, NSF 1952096, NSF 2003874. Guang also thanks his wonderful wife Jie for her unstinting support.

Authors' addresses: G. Wang, Z. Fang, X. Xie, and D. Zhang, Department of Computer Science, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ, 08854-8019; emails: guang.wang@rutgers.edu, {zhihan.fang, xx88, desheng.zhang}@cs.rutgers.edu; S. Wang, Department of Computer Science, Southeast University, Key Lab of Computer Network and Information Integration, School of Computer Science and Engineering, Nanjing, Jiangsu, 210096, China; email: shuaiwang@seu.edu.cn; H. Sun, School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China; email: hjsun1@bjtu.edu.cn; F. Zhang, Shenzhen Beidou Intelligent Technology Co., Ltd. Shenzhen, Guangdong, 518000, China; email: zhangfan@siat.ac.cn; Y. Liu, Beijing Institute of Big Data Research and Peking University, Beijing, 100000, China; email: yunhuai.liu@pku.edu.cn.

2157-6904/2020/11-ART13 $15.00
https://doi.org/10.1145/3428080

ACM Reference format:
Guang Wang, Zhihan Fang, Xiaoyang Xie, Shuai Wang, Huijun Sun, Fan Zhang, Yunhuai Liu, and Desheng Zhang. 2020. Pricing-aware Real-time Charging Scheduling and Charging Station Expansion for Large-scale Electric Buses. *ACM Trans. Intell. Syst. Technol.* 12, 1, Article 13 (November 2020), 26 pages.
https://doi.org/10.1145/3428080

1 INTRODUCTION

With the growing concerns over air quality [16, 63, 65–68] and energy security [58, 59], more and more countries, e.g., U.S. and China, have started to promote electric vehicles to improve energy efficiency and reduce emissions [10]. It is reported that the worldwide sales of electric vehicles have been nearly quadrupled since 2014, and half of the vehicle sales will be electric vehicles by 2027 [34].

As one of the most common transportation modes, transit buses play an important role in people's daily lives [56, 57, 60, 61]. Due to the long daily travel distance and high-frequency services, electric buses (e-bus) have greater potentials to reduce the nitrogen oxides and carbon dioxide emissions [10, 58, 59] compared to other types of electric vehicles, e.g., electric taxis (e-taxis) [6, 40, 45, 46] and electric private vehicles (e-pvs) [23].

Yet, to date, e-buses have not been extensively adopted worldwide because of the following distinctive characteristics: (i) lack of spacious charging infrastructures for large-scale e-bus fleets, e.g., large charging stations with enough parking spaces and sufficient charging stations [31]; (ii) high purchase costs due to the relatively new technologies, e.g., a BYD e-bus will cost about $263,000 [31, 42] in 2017, which is 2.5 times of a diesel bus; (iii) relatively high operating costs caused by the charging fees (i.e., electricity) compared to one-time costs of infrastructure construction and ownership. For example, the yearly charging cost for one e-bus is about $18,000 [37] in Shenzhen, resulting in the day-to-day charging costs being one of the key concerns that hinder the e-buses to fully release their potential [10].

Many works have been done on reducing the charging costs for e-taxis and e-pvs [6, 24, 39], and some other works [2, 8, 26, 35, 36, 49, 50] have built the theoretical models and simulations for e-buses. However, few works have been conducted on the data-driven modeling and optimization for real-world e-bus fleets charging. More importantly, approaches for e-taxis [6, 24, 39, 44, 47] can hardly be directly applied to the e-buses because of the following fundamentally different features: (i) The charging activities of e-taxis are directly related to the income of e-taxi drivers, so their charging and routes will be incentive-based; whereas e-bus drivers are not. (ii) The charging activities of e-taxis are mostly distributed and flexible. An e-taxi driver can decide when and where to charge, while the e-bus network is based on centralized operating and charging management. These two key differences lead to different charging incentives and optimization goals [23].

E-buses are centrally managed with fixed timetables, which makes it possible to design offline charging schedules and operating strategies. However, such offline strategies are not always optimal. The high dynamics of the real-world environments bring great challenges for optimal solutions, making e-buses very different from the flexible e-taxis and e-pvs. Such dynamics may include the unexpected break-downs of e-buses, unpredictable traffic congestion, time-variant electricity rates, as well as the changing weather/temperature and the traffic-light conditions [37]. For example, both too-hot or too-cold weather will require the e-buses to operate their air conditioners, which drain their energy quickly. These dynamic factors will lead to both (i) the non-deterministic departure and arrival time for each e-bus and (ii) unpredictable State of Charge (SOC) (i.e., the remaining battery level) of e-buses when they arrive at terminals. Note that the fixed timetables only require e-buses to leave the terminal on time. As a result, the off-line charging schedule and operation strategies—i.e., all e-bus lines follow the predesigned and fixed operating and charging patterns—may be far from the high operating and charging efficiency. For example, some e-buses may wait too long in the charging stations for available charging points, while other charging stations have a lot of unoccupied charging points.

To address these real-time issues, we develop *busCharging*, a data-driven real-time charging scheduling system for large-scale e-bus fleets based on real-world data. The dataset is obtained from the Chinese city Shenzhen, a pilot city that promotes e-buses in China. Shenzhen has electrified 100% of its public transit buses and became the first and the only city with a full e-bus network in the world. Moreover, Shenzhen also has the largest installation base of e-buses, e.g., 16,359 e-buses [29], in December 2017. Based on Shenzhen's e-bus data, we perform a set of data-driven analyses to understand the behaviors of e-bus fleets and then design a data-driven real-time scheduling strategy to reduce the overall charging cost of the fleets. Our key contributions are as follows:

- To our best knowledge, we are the first to conduct the city-scale data-driven investigation of the real-time scheduling for electric bus fleets. Our investigation has two key features based on four-year real-world e-bus data, including (i) the largest e-bus fleet in the world with more than 16,000 e-buses; (ii) the largest number of e-bus charging stations and charging points, e.g., 376 charging stations for e-buses. Such a large-scale data-driven investigation enables us to identify the real-world e-bus operating and charging issues, which are challenging to reveal by using simulation studies.
- We design a data-driven real-time charging scheduling system called *busCharging* taking factors of the e-bus daily and per-charge operating distances, charging spatial-temporal and cost distribution, and charging station utilization rate into account. *busCharging* is based on a thorough data-driven analysis, which reveals some novel insights including the e-bus adoption process, e-bus demand/supply, charging network distribution, charging activities distribution, charging cost, and so on. To our knowledge, this is the first time that such detailed analyses were performed for a large-scale e-bus fleet to support data-driven scheduling.
- Given these data-driven insights, *busCharging* is designed based on the Markov Decision Process (MDP) by considering the status of the e-bus fleet and the contextual factors. In particular, we consider both revenues and charging costs, along with timetables and time-of-use electricity prices to schedule e-buses among different bus lines. Then, we utilize an effective solution to solve this MDP problem, and we also theoretically investigate the real-time features of our MDP-based scheduling by analyzing its time complexity. Finally, we elaborate on how our scheduling strategy guarantees timetables for the real-time requirement.
- We implement and evaluate *busCharging* based on the real-world data in Shenzhen. The results show *busCharging* reduces 23.7% of the overall charging cost and 12.8% of the electricity usage. We also design a scheduling-based charging station upgrading strategy to verify our *busCharging* is also effective, even enlarging existing charging stations. Besides, some lessons were learned and experiences were reported, which are helpful for other cities to promote and optimize their e-bus fleets.

The rest of the article is organized as follows: Section 2 introduces our dataset and conducts the detailed analyses. Section 3 presents the design and implementation of *busCharging*. Section 4 evaluates the performance of *busCharging*. The lessons learned and related works are summarized in Sections 5 and 6. We conclude this article in Section 7.

Table 1. Specifications of One Type e-bus in Shenzhen

Model	Battery	Length	Charging Duration	Maximum Speed	Maximum Distance
BYD K9	324 kWh	12 m	3h	90 km/h	250 km

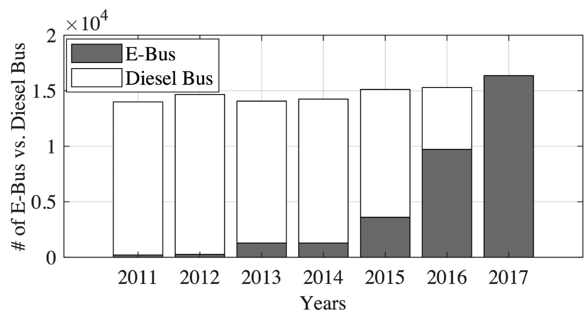


Fig. 1. Number of e-buses and diesel buses.

2 *BUSCHARGING*: DATASETS AND ANALYSES

In this section, we first describe our large-scale real-world dataset generated from the Shenzhen e-bus fleet. Based on the dataset, we then comprehensively investigate the operating patterns, charging patterns, and cost patterns of the Shenzhen e-bus fleet to motivate our electricity pricing-aware real-time charging scheduling.

2.1 Data Description

Our *busCharging* is based on large-scale e-bus datasets obtained from Shenzhen, which is the fourth-largest city in the Chinese mainland. The time span of these datasets is from the year 2014 to 2018, during which Shenzhen has experienced very rapid growth of e-buses, e.g., the percentage of e-buses among all buses (i.e., e-buses + diesel buses) has increased from 7.8% to 100% as shown in Figure 1. One of the most popular e-bus vehicle models in Shenzhen is BYD K9, whose specifications are shown in Table 1. The battery capacity and the maximum per-charge distance of BYD K9 are 324 kWh and 250 km, respectively [7]. The four-year datasets include five different types of data, and the details are shown as follows:

- **GPS Data** include 1.92 TB historical and real-time GPS records of all buses in Shenzhen from July 2014 to May 2018. Each GPS record includes 19 fields describing the status of a bus, e.g., the bus ID, the time-stamps, the bus line ID, the GPS coordinates (i.e., longitude and latitude), the direction, the current speed, and the total mileage (i.e., odometer data). The GPS data are collected by an onboard device with a cellular connection in real-time.
- **Bus Stop Data** include all bus stops' information of 1,400 bus lines (including inbound and outbound directions) with 5,562 unique bus stops. For each bus stop, there are seven key fields including the route ID, the line direction, the stop name, the GPS location, and so on.
- **Bus Transaction Data** include all transaction records of passengers' trip fares. The average daily number of passengers taking buses using smartcards is about 2.4M, and 5M in total. Each transaction has six key fields including the route ID, the line direction, the station ID, the station name, and the GPS location.

- **Bus Charging Station Data** include the station names, the station IDs, the GPS locations, and the number of charging points in each station. There are 376 e-bus charging stations in Shenzhen as of the end of 2017.
- **Electricity Rate Data** include the time-variant electricity pricing within 24 hours of Shenzhen. Shenzhen adopts the time-of-use rating, which breaks up 24 hours of a day into several intervals and charges a different price for each interval [38]. The rates in Shenzhen are divided into three types, i.e., off-peak prices (low rates), semi-peak prices (medium rates, also called flat rates), and peak prices (high rates), and the corresponding electricity rates are 0.049 $/kWh, 0.121 $/kWh, and 0.173 $/kWh, respectively. The time-variant electricity pricing in Shenzhen is shown in Figure 2.

Based on these data, we perform an intensive data-driven analysis to understand the operating, charging, and cost patterns of the Shenzhen e-bus fleet. The details are shown below.

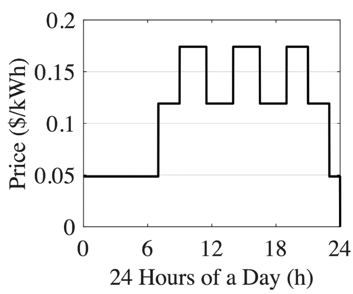


Fig. 2. Time-variant industrial electricity prices in Shenzhen.

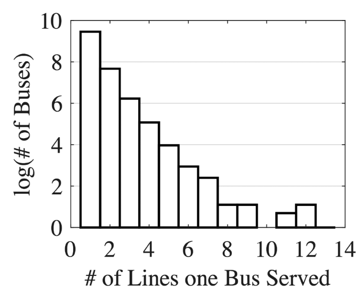


Fig. 3. The number of lines served for each bus in Shenzhen.

2.2 Operating Patterns

A visualization of Shenzhen e-bus network in 2018 is shown as Figure 4, where the yellow and red parts stand for the bus lines across Shenzhen. The yellow parts mean there are more bus passengers on these lines, and the red parts mean there are fewer passengers on these lines.

We found that the bus lines reach the most remote rural areas, although the highest density of lines is in the downtown area. We also found that the bus in Shenzhen may serve different lines at different times of a day, e.g., rush hour and non-rush hour. The number of lines served by each bus can be seen from Figure 3. We found that about one-fifth of buses in Shenzhen serve for more than one fixed line during different times, e.g., The *Bus A* served for *line 1* can be borrowed to *line 2* if it is in peak hours of *line 2*. In addition, the lengths of different lines vary a lot, e.g., the shortest line is only 1 km and the longest line is over 110 km, and the average and variance of the length of all lines are 9.7 km and 6.3 km, respectively, which potentially indicates it is feasible for us to schedule e-buses to serve other lines.

In Figure 5, we investigate the daily number of passengers on each bus line. We found that 21% of bus lines have over 6,000 passengers per day, and the maximum number of passengers that a single bus line carries up to 23,200; such huge amount of passengers on bus lines pose a huge challenge for charging scheduling, since we need to satisfy passengers' traveling demand. Hence, guaranteeing the timetables of all bus lines is necessary when we change their charging plans. In addition, we found when the passenger demand becomes higher, the departure interval between two e-buses will be shorter from observing the operation data. Figure 6 shows the number of e-buses and the number of passengers in Shenzhen during 24 hours of a day. We found that at different times of a day, the number of passengers and e-buses is also proportional, which means

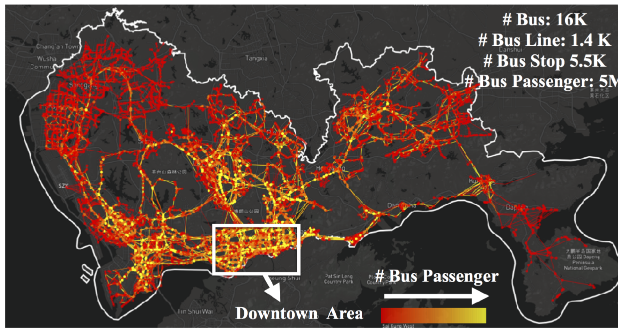


Fig. 4. Station and line distribution of the Shenzhen e-bus network.

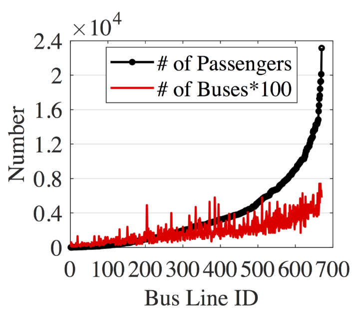


Fig. 5. Passengers & buses (lines).

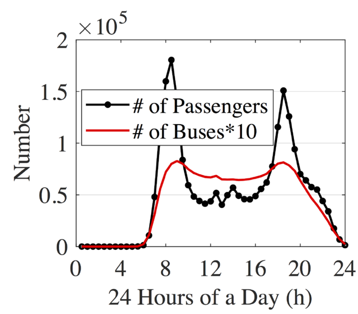


Fig. 6. Passengers & buses (time).

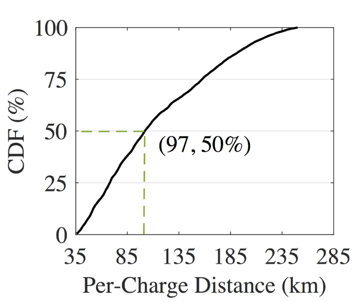


Fig. 7. Distance between two charges.

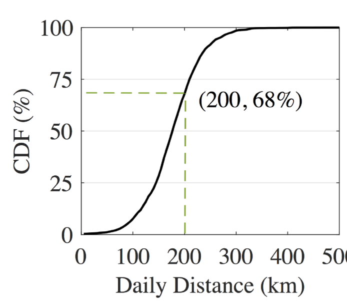


Fig. 8. Daily operating distance.

we can have some e-buses charge during off-rush hours and then serve other bus lines during rush hours.

We further investigate the operating distances between two charges (i.e., per-charge distance) and the daily operating distances of Shenzhen e-buses. The Cumulative Distribution Function (CDF) of operating distances between two charges of all e-buses is shown in Figure 7. We found that about 50% of e-buses operate no more than 97 km between two charges, even though their maximum operating distance is around 250 km, which is caused by many real-world factors, e.g., the availability of charging points, the electricity rates, and the range anxiety.

Further, we show the CDF of the total daily operating distance of each e-bus in Figure 8. We found that the daily operating distances of 68% of e-buses are less than 200 km, which is the maximum practical distance that most e-buses would travel before a charge. In addition, 32% of

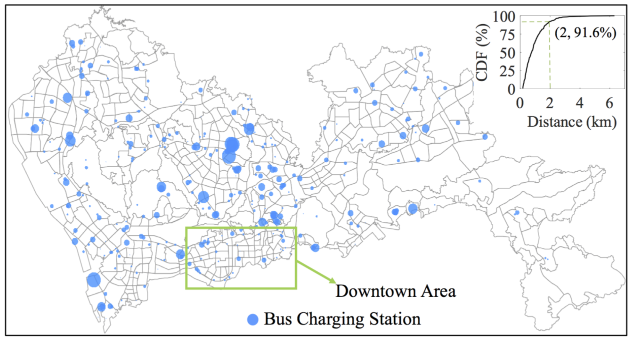


Fig. 9. Spatial distribution of Shenzhen e-bus charging network.

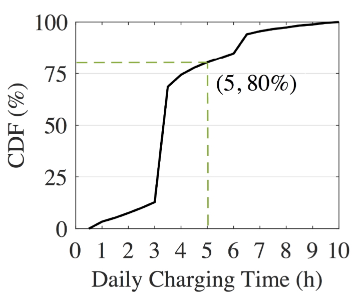


Fig. 10. Daily total charging time.

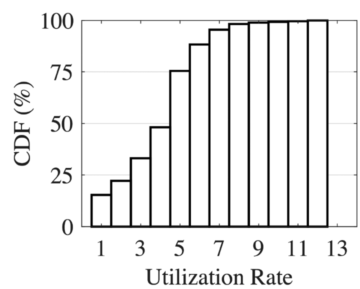


Fig. 11. Daily station utility rate.

e-buses travel more than 200 km per day, which means they need at least two charges per day. Considering the per-charge distance of 50% of e-buses' is no more than 97 km, it could be concluded that most e-buses need to charge at least twice per day.

2.3 Charging Patterns

We utilize long-term e-bus data and charging station data to fully understand the overall e-buses charging patterns in Shenzhen. The spatial distribution of the e-bus charging stations is shown in Figure 9, where the sizes of the circles indicate the number of charging points in each station, which means a larger circle stands for more charging points in this charging station. We found that most large stations are located in suburban areas, which is because there are more available and cheaper land resources in these areas than in downtown areas. We also investigate the distances between two charging stations. As shown in the right upper corner of Figure 9, 91.6% of charging stations have at least one neighbor charging station within 2 km, which indicates Shenzhen has a well-connected e-bus charging station network.

We further investigate the daily charging time of each e-bus. As shown in Figure 10, we found only 13% of e-buses spend less than 3 hours for charging each day, and the charging time of over 80% of e-buses is no more than 5 hours, which indicates that most e-buses spend 3–5 hours for at least two charges each day.

To study the e-bus charging station effectiveness, we define the daily Charging Station Utilization Rate *UR* at a station s_i in a day as follows:

$$UR(s_i) = \frac{CE(s_i)}{CP(s_i)}, \tag{1}$$

(a) E-buses in Shenzhen (b) Charging station in Shenzhen

Fig. 12. Electric buses and charging stations in Shenzhen.

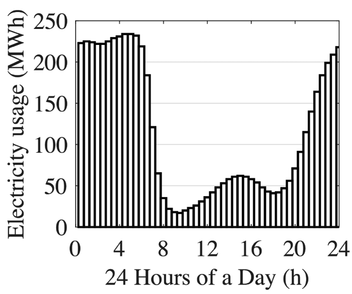


Fig. 13. Electricity usage.

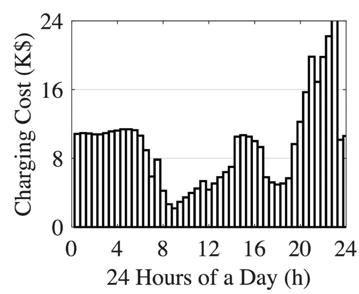


Fig. 14. Charging cost distribution.

where $CE(s_i)$ is the daily total number of charging events in the station s_i; $CP(s_i)$ is the number of charging points in station s_i.

Figure 11 shows the daily charging station utilization rate distribution of e-bus charging stations in Shenzhen. We found that the utilization rates of 50% of charging stations are no more than 4. However, there are also some charging stations with very high utilization rates, e.g., 12. These unbalanced utilization rates lead to charging resources waste in some stations while very crowded in some other stations. Besides, we have conducted a series of field studies in 2018 to investigate the charging patterns of e-buses in Shenzhen. Figure 12 shows the e-buses and a charging station we visited in Shenzhen.

2.4 Cost Patterns

Based on these charging station utilization rates, we further study charging start time and charging distribution over time of day to understand the electricity usage and charging costs. As shown in Figure 13, the highest electricity usage for e-bus charging occurs in the off-peak hours, accounting for 63% of the total electricity usage. However, there is still 13.6% of electricity usage during peak hours. Although the percentage of electricity usage in the peak hours and flat hours is much lower than the usage in the off-peak hours, the usage in these two durations can cause more charging costs than the peak-hour charging due to the time-of-use pricing strategy. Comparing Figure 14 to Figure 13, we found that the charging cost distribution is different from the electricity usage distribution. In particular, we found the electricity cost gap between 8:00–20:00 and 20:00–8:00 in Figure 14 is much smaller than the electricity usage gap between the same two periods in Figure 13. This indicates that even though the bus fleet does not charge much from 8:00–20:00, the costs are almost as high as 20:00–8:00, during which the e-buses charge substantially. **As a result, it motivates us to ask a question: Can we reduce the charging cost of large-scale e-bus fleets by scheduling more e-buses to charge in off-peak hours?**, i.e., further increasing the electricity usage gap between 8:00–20:00 and 20:00–8:00, since a small ratio of usage decrease during peak hours will result in a huge cost decrease.

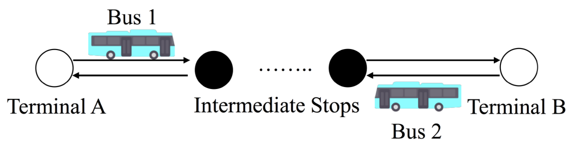


Fig. 15. Existing bus operating and charging patterns.

3 *BUSCHARGING*: SCHEDULING DESIGN

In this section, we first present our problem formulation by showing the existing operating and charging scheduling of Shenzhen e-buses. Then, we present our *busCharging* scheduling idea. Next, we introduce the detailed design of our *busCharging* in terms of the scheduling formulation, scheduling design, and scheduling complexity analyses. Finally, we describe the timetable guarantee of our scheduling for the real-time guarantee.

3.1 Problem Formulation

3.1.1 Existing Operating and Charging Patterns. Figure 15 shows the operating patterns of buses in Shenzhen. In Shenzhen, a bus line generally has two terminals, e.g., Terminal *A* and Terminal *B* in Figure 15. Based on the timetable for this line, a *Bus 1* travels from Terminal *A* to Terminal *B* (or from *B* to *A*) through some intermediate bus stops and then goes back to the *Terminal A* (or *B*). During the same time, multiple buses are serving for this line with the same or different directions, e.g., *Bus 2*. For e-buses, they need to charge when their SOC decreases to a predefined low threshold. Besides, different from other electric vehicle types (e.g., e-taxis and e-pvs), e-buses generally make charging decisions (i.e., to charge or to continue to serve) only after they arrive at bus terminals. This is caused by their operational feature, i.e., they normally cannot charge in the middle of a trip with bus passengers onboard.

As a result, based on this feature, the charging stations in Shenzhen are usually deployed near terminals or at terminals, which also addresses the parking issues, since e-buses need lots of space for parking compared to taxis and private vehicles. Moreover, in Shenzhen, nearly half of the e-buses in the fleets charge at their closest terminals (e.g., the terminals at which they just arrived), which may potentially decrease the charging efficiency of the entire e-bus charging network due to the unbalanced charging points placement (e.g., fewer charging points in the downtown and more in the suburb, as shown in Figure 9), resulting in no charging points available when e-buses arrive at their charging stations but many unoccupied charging points in some other charging stations.

3.1.2 busCharging *Operating and Charging Patterns.* In this work, we focus on the real-time charging scheduling problem for e-bus fleets, considering the overall operational cost and revenues of all e-buses in the fleet, instead of individual vehicles, given its centralized management mode. Compared to conventional diesel buses, e-buses are less flexible due to their limited operating ranges and reliance on the charging infrastructures, which makes it challenging to schedule the e-bus fleet to operate and charge compared to conventional diesel buses, especially for a large-scale fleet. In addition, the lower flexibility of e-buses potentially makes it necessary to (i) purchase extra e-buses for contingency plans to cover additional ranges on routes or (ii) redesign the lines to accommodate e-buses. These two actions have been taken in Shenzhen. Moreover, the time-variant electricity rates compared with the 24-hour stable diesel price also make the charging issues of e-bus fleets more complicated if we consider the refueling costs.

In *busCharging*, we consider different real-world factors for e-buses, including time-variant electricity rates and scheduling between different lines/extra e-buses. Weather conditions, congestion, time of day, and demographic features as contexts are also implicitly considered, since we leverage both the historical and real-time GPS records to predict energy consumption to serve a particular line with detailed route information. Based on previous research [43, 55, 64], both periodic congestion and static demographic features remain stable for the same spatial-temporal combination, e.g., for the same road segment during the same time slot of different days. As a result, our historical GPS records implicitly contain the period congestion and basic demographic features.

Our **key idea** for *busCharging* is that we schedule some e-buses to serve other bus lines that share the same terminals with them when they arrive at their terminals based on some real-world factors. These factors we considered include: (i) the real-time SOC of e-buses; (ii) the availability of charging points in charging stations; (iii) the expected energy consumption of different lines, which is related to both lengths and travel time related to traffic, which means e-buses have to be able to arrive at the scheduled terminals based on their current SOC; (iv) the expected charging cost at particular terminals for different time slots of a day, which is related to both charging point availability (e.g., staying without charging if no point is available) and electricity prices; (v) our final consideration for the scheduling is that we have to guarantee the timetables of all lines, which we will clarify in Section 3.3.

Based on this key idea, we make streaming scheduling decisions after each e-bus arrives at a terminal and drops off all passengers, since all e-buses arrive at terminals in an online fashion [23]. In particular, *busCharging* has four potential scheduling modes for an e-bus when it is in a terminal based on the four factors we discussed in the last paragraph: (i) stay at the terminal but not charge; (ii) charge at the terminal immediately; (iii) keep serving the current line, i.e., go back to the terminal it came from; (iv) serve another line, i.e., go to a terminal of a different line. To make our scheduling easier to guarantee timetables of all bus lines, we should make the number of buses to serve other lines as little as possible, so we set e-buses to prioritize serving for the original lines over serving the new lines.

3.1.3 Scheduling Objective. The objective of our *busCharging* is to optimize the e-bus fleet by reducing the overall operational cost and increasing the profits by collecting fares for serving passengers, which can be formulated as follows:

$$F_s - C_c = \sum_{t \in 24h} \sum_{n=1}^{N_{eb}} \left(F_n^t - R^t \cdot C_n^t \right), \tag{2}$$

where F_s is the collected fare for serving passengers of the whole e-bus fleet; C_c is the charging cost for operating the e-bus fleet; N_{eb} is the number of e-buses in the fleet; F_n^t is the fare collected by the nth e-bus during time t by serving passengers for a particular line; R^t is the electricity rate at time t; C_n^t stands for the electricity consumed (i.e., charged) by the nth e-bus during time t. As Equation (2) shows, the overall optimization objective depends on three items, e.g., the real-time electricity rates, the energy charged by each e-bus, and the fares collected by each e-bus. Compared to the taxi and private vehicles, bus passenger demand is more stable and the fare is flatter in Shenzhen [25], so the expected fare can be obtained from historical data given a time slot and a bus line. As a result, we focus on deciding on scheduling e-buses to serve which lines and when to charge for reducing the overall charging cost.

An intuitive idea to achieve this goal is to have enough buses and to schedule all the e-buses to serve their original lines and to charge only during the off-peak hours, i.e., 23:00–7:00. However, given the limited number of buses in Shenzhen and a large number of lines, as indicated by Figure 8, around 32% of e-buses cannot accomplish the daily operating task with one charge during nights alone. Note that for practical consideration, we only focus on scheduling existing buses

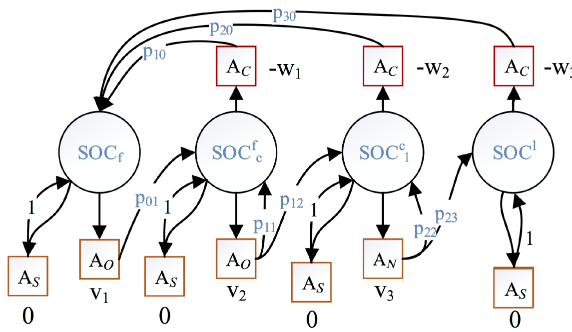


Fig. 16. Markov decision process for charging scheduling.

for improving the charging efficiency, instead of adding new buses in this article. We present the detailed charging scheduling algorithm in the next Section 3.2.

3.2 Charging Scheduling

Since e-buses arrive at bus terminals in an online fashion, we schedule the charging task for e-buses one by one. One of our major technical contributions in this article is that we formulate the e-bus charging scheduling problem as an MDP problem to reduce the charging cost and effectively solve this problem.

An MDP is a discrete-time state transition system that aims to find an optimal policy to maximize the expected utility. Formally, an MDP is defined as a 5-tuple (S, A, T, R, β) [32]. The MDP framework of *busCharging* is shown as Figure 16.

- S is a set of states. For the charging scheduling scenario, we define four different states according to the SOC of e-buses. As Figure 16 shows, the four states in *busCharging* are (i) SOC_f, which indicates an e-bus is at the Full SOC; (ii) SOC^f_c, which indicates the SOC of an e-bus is lower than the Full SOC but higher than an SOC with which it can serve the Current line, i.e., go back to the original terminal; (iii) SOC^c_l, which indicates the SOC of an e-bus is lower than the required SOC to serve the Current line but higher than a mandatory charging threshold, i.e., it may still have SOC to serve other lines sharing the same terminal with it; (iv) SOC^l, which indicates the SOC of an e-bus is lower than the mandatory charging threshold, i.e., it needs to charge and cannot serve any lines.
- A is a set of actions. In *busCharging*, there are four actions: (i) A_S: Staying at this terminal but not to charge; (ii) A_O: going back to the Original terminal, i.e., serving the current line; (iii) A_N: serving a New line; (iv) A_C: Charging at this terminal. For the charging scheduling, the results of the former three actions follow the energy non-increasing principle, i.e., the energy in the next state does not exceed the current state; whereas, for the state A_C, i.e., charging, it increases the SOC of e-buses.
- T is a state transition matrix, which consists of the probability transition from one state to another state by taking an action. For example, $\text{T}(\text{SOC}_f, \text{A}_O, \text{SOC}^f_c) = \text{P}^{A_O}_{SOC_f \to SOC^f_c} = \text{p}_{01}$ means the probability of an e-bus transferring from the full battery capacity SOC_f to SOC^f_c by serving its own line (i.e., action A_O) is p_{01}.
- R is a reward function. For each action, the scheduling strategy will generate a corresponding reward. In the charging scheduling problem, if an e-bus stays at the terminal (i.e., A_S), it will have no passengers (i.e., no revenue) and also have no energy consumption (i.e., no

cost), so the current reward is 0. When e-buses serve for passengers (i.e., the action A_O or A_N), they will have passenger fares for their lines, so they will have a positive reward V_i, which is based on which line they serve. But if they take the action A_C, there is a charging cost for electricity, so the reward is negative, which is denoted as the -W_j in Figure 16, where $1 \leq j \leq 3$. The different negative reward values depend on the real-time electricity rates and the current SOC, as well as their full battery capacity.

- β is the discount factor, which captures the fact that an immediate reward might be worth more than the same reward in the future. The value of β is generally selected from [0,1), so the final expected utility will be convergent and bounded to a finite number. β is set to 0 if and only if we do not consider the future reward.

Definition 2. a **policy** π is defined as a distribution over actions given states, which gives the e-bus an action to execute at each state to maximize the expected utility:

$$\pi\left(a|s\right) = P\left[A = a|S = s\right], \tag{3}$$

where $a \in$ A(s) = $\{A_S, A_O, A_N, A_C\}$ and $s \in \{SOC_f, SOC_c^f, SOC_l^c, SOC^l\}$.

Definition 3. a **utility** of a state for a given policy is defined as $U^\pi(s)$, which can be formulated as:

$$U^\pi(s) = E\left[\sum_{t=0}^{\infty} \beta^t \cdot R(s_t)|\pi, s_0 = s\right], \tag{4}$$

where s_0 is the initial state; s_t is the state of the e-bus after executing the policy π for t actions; β^t is the discount after t actions; and $R(s_t)$ is the immediate reward at each state.

The objective of the data-driven charging scheduling strategy is to derive an optimal policy π^* that achieves the maximum utility $U^*\left(s\right)$ for all states, which is formulated as the Bellman Equation [3] as Equation (5):

$$\begin{aligned} U^*\left(s\right) &= \max_{\pi} U^\pi\left(s\right) \\ &= \max_{a \in A(s)} \left[R_{sa} + \beta \cdot \sum_{s'} P_{ss'}^a \cdot U^*\left(s'\right)\right], \end{aligned} \tag{5}$$

where R_{sa} is the immediate reward after taking the action a in the state s, which is also the operational revenue or charging cost after taking different actions. $\sum_{s'} P_{ss'}^a \cdot U^*\left(s'\right)$ is the expected future utility. As Equation (5) shows, the immediate reward R_{sa}, the discount factor β, and the transition probability $P_{ss'}^a$ are required to obtain the best policy.

As shown in Equation (2), the objective of our *busCharging* is to minimize the overall charging cost of the e-bus fleet and maximize fares collected for serving passengers. The fares would be reduced if the e-buses do not keep the timetable, e.g., some passengers may take taxis. In this work, we envision that the fares are maximized if the timetables of all lines are kept as much as possible, which is ensured by our scheduling policy, which we will explain in Section 3.3. As a result, the objective of Equation (2) is equivalent to the objective of Equation (5), which indicates we can solve the e-bus charging scheduling problem by considering both charging cost and revenues of collected fares. In the next section, we will theoretically investigate if this MDP-based scheduling can satisfy the real-time requirement by studying its scheduling complexity.

Scheduling Complexity: There are two common approaches to solve the MDP optimization problem, i.e., value iteration and policy iteration [9]. There are n unknowns needed to solve in the equations when there are n states. If we leverage the value iteration, we need to use *max* in the equations, which is nonlinear, resulting in complexity of $O(m \cdot n^2)$ for each iteration for

m actions. However, the policy iteration has the operation $\sum$ instead of the operation *max* in the equations, which implies that the equations are linear. Thus, solving these n linear equations is with a complexity of $O(n^3)$. Since we have four states and four actions in our scheduling process, the time complexity of the two approaches is similar for each iteration. However, the policy iteration searches a finite policy space instead of an uncountably infinite value space, indicating the policy iteration converges much faster than the value iteration. Now, we have shown that the policy iteration satisfies the real-time requirement for scheduling. In the worst case, the policy iteration for *busCharging* needs to search for all the policy space, which is limited by the number of our states and the limited number of lines sharing the same terminals. Then the total time cost of the policy iteration for *busCharging* is linear, which can be easily realized at the second level for a normal PC, so the policy iteration–based MDP is fast enough for the real-time charging scheduling requirement. Therefore, we leverage the policy iteration to find the best charging scheduling policy for e-buses.

The policy iteration algorithm for our e-bus charging scheduling is shown in Algorithm 1.

ALGORITHM 1: Policy iteration algorithm for e-bus charging scheduling

Input: The set of all states $S = \{\mathrm{SOC}_f, \mathrm{SOC}_c^f, \mathrm{SOC}_l^c, \mathrm{SOC}^l\}$
The set of all actions $A = \{\mathrm{A}_S, \mathrm{A}_O, \mathrm{A}_N, \mathrm{A}_C\}$
State transition function $P_{ss'}^a$
Reward function R_{sa}
Output: Optimal policy π^*
Initialize: Pick an arbitrary policy π'
repeat
  $\pi \leftarrow \pi'$
  Policy evaluation: solve the linear system
  $U^\pi(s) = R_{sa} + \beta \cdot \sum_{s'} P_{ss'}^a \cdot U(s')$;
  Policy improvement: for each $s' \in S$
  $\pi'(s) \leftarrow \arg\max_{a \in A} \left(R_{sa} + \beta \cdot \sum_{s'} P_{ss'}^a \cdot U(s')\right)$
until $\pi = \pi'$;
return π *as* π^*

3.3 Timetable Guarantee

Since keeping the timetable is critical for the bus fleet operation, our algorithm considers the timetable constraint and effectively guarantees timetables for e-bus lines when performing the scheduling. Our scheduling can be regarded as a priority-based scheduling, and the priority is decided by the utilities that different strategies can achieve, which means a higher utility strategy has a higher priority. As follows, we show how our scheduling algorithm guarantees timetables for bus lines with an example:

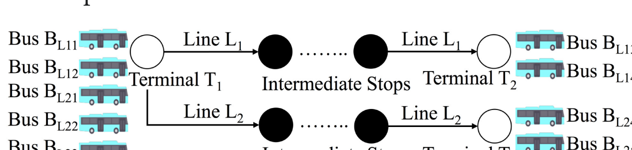


Fig. 17. Scheduling with timetable guarantee.

As shown in Figure 17, the line L_1 and line L_2 share the same terminal T_1, and there are many e-buses serving for different lines, e.g., B_{L11} for L_1 and B_{L21} for L_2. There are three different scenarios when *busCharging* makes scheduling:

- If the next e-bus B_{L11} for line L_1 can have a better scheduling to serve another line, e.g., L_2, compared to the original line L_1 and another e-bus B_{L12} or B_{L13} is available for L_1, then we schedule B_{L11} to serve another line (e.g., L_2) and have B_{L12} or B_{L13} to take the place of B_{L11}. We utilize the charging deadline to break the tie if multiple e-buses were available, which means our scheduling strategy will choose the bus with the highest SOC to replace B_{L11}. In this way, the timetable of L_1 is guaranteed.Similarly, we utilize the energy consumption to break the tie if multiple e-bus lines were available, which means our scheduling strategy will choose the line with the lowest energy consumption to serve. Although the selected line does not necessarily reduce electricity consumption, it can guarantee the reachability and timetable of this line, and it will have positive effects on charging cost reduction due to the time-variant charging pricing.
- If we cannot find the next bus B_{L12} and B_{L13} for L_1 to guarantee the timetable of B_{L11}, we find all other lines that share the terminals with L_1, i.e., L_2 in this example, for an extra backup bus for L_1 (which do not affect their timetables). If there is an e-bus available from B_{L21}, B_{L22}, and B_{L23}, then the timetable of L_1 can be guaranteed while we schedule B_{L11} for other lines.
- If there are no available buses from all other lines (e.g., L_2) to guarantee the departure timetable of B_{L11}, we keep bus B_{L11} on L_1 to guarantee L_1's timetable. E-buses prioritize serving the original lines over new lines when they can achieve the same performance by serving different lines, which is also for guaranteeing the timetables of all bus lines. For example, at the beginning of a day, all e-buses have enough energy to serve current lines and other lines so they will keep their current lines. In this case, the timetable is always kept.

The timetable guarantee is for satisfying the real-time requirement, which we used as the constraint of the MDP model. That is to say, all scheduling decisions should guarantee the timetable for serving passengers. In particular, the timetable guarantee guides the state and action choices, e.g., if serving the current line or serving another line that shares the same terminal with the current line will be decided by the timetable and SOC of e-buses.

4 EVALUATION

In this section, we extensively evaluate the performance of our *busCharging* based on the massive e-bus data from Shenzhen by four different metrics, i.e., the temporal distribution of charging events, the spatial distribution of charging events, the joint spatiotemporal distribution of charging events, and, the most important, electricity usage and charging cost. In addition, we also discuss the impact of our scheduling on bus drivers.

4.1 Experimental Setup

Data Management: Due to the data-driven nature of our *busCharging*, we introduce how we manage our multi-source data related to e-buses as follows: In this project, we are working with Shenzhen Transportation Committee, and we utilize various data processing frameworks. The streaming data from Shenzhen e-buses require significant efforts for efficient management, querying, and processing.

We employ a high-performance cluster with Spark for data processing. The details are given as follows: (i) 12 HP machines with 2 Tesla K80c each; (ii) 10 Dell machines with 4 Tesla K80c each; (iii) 4 Xeon E5-2650 with a half TB memory each; (iv) A series of 800 GB SSD and 15 TB of spinning-disk spaces; (v) 2 PB additional disk space. Due to the large size of our bus data, we performed a detailed cleaning process to filter out duplicate, error, and incomplete GPS/transaction data. More importantly, the key challenge in bus transaction data processing, compared to bus GPS data processing, is to protect the privacy of smartcard users and ensure the utility of the models at the same time. We will briefly mention our privacy consideration in Section 5.

Evaluation Data: In this evaluation part, as introduced in Section 2, we utilize one-week of GPS records generated by 16,359 e-buses from January 13th–19th, 2018, in the Chinese city Shenzhen. More than 69.6M GPS records are generated by the e-bus fleet during this period. In addition to GPS data, the evaluation dataset includes the static data of 376 e-bus charging stations, 1,400 bus lines, and 5,562 bus stops.

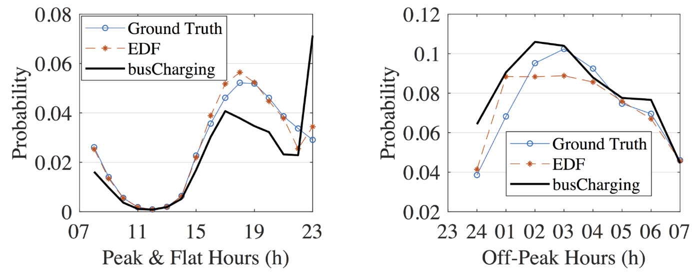
Parameter Setting: In our charging scheduling strategy *busCharging*, four parameters are needed to decide, as follows:

- **Four Different States**: For SOC^l, we decide it based on the real-world interaction with Shenzhen e-bus drivers. Based on our field studies in Shenzhen, drivers of e-buses are very conservative, and they normally stop serving passengers and go to find charging points if the SOC declines to 30%. As a result, we set the SOC^l to be 30%. Based on the expected real-time traffic and current line serving, we can calculate the energy consumption for serving the current line and other lines for each bus, which gives us the SOC_c^f and SOC_l^c, since we have SOC^l.
- **Discount Factor β**: We empirically choose the discount factor β as 0.9 to guarantee the convergence of the algorithm similar to many previous works [1, 14, 17].
- **Immediate Reward R_{sa}**: For the immediate reward, the V_i $(1 \leq i \leq 3)$ is the expected revenue by serving different lines, which are calculated based on historical bus passenger demand; W_j $(1 \leq j \leq 3)$ is the expected charging cost, which is calculated based on the time of scheduling and the real-time SOC of e-buses.
- **Transition Probability $P^a_{ss'}$**: For the state transition probability, if an e-bus takes action A_S, it will stay in the same state, so the probability is set to 1. Since there is also no electricity consumption and operational profits, the reward is also set to 0 when taking action A_S. If there is only one possible transition, the probability is also 1, since the sum of the transition probabilities is 1, so p_{01} is 1 in Figure 16. Other transition probabilities are calculated in real time based on policies given in Section 3.2.

Baseline Setting: To show the effectiveness of our real-time charging scheduling for e-bus fleets, we compare the performance of our *busCharging* with Ground Truth and another real-time scheduling method called Earliest-deadline-first (EDF), which is a common method for electric vehicle charging scheduling [6, 23, 41]. In EDF, they schedule the e-buses with the earliest timetable deadline to charge first. Further, we leverage four metrics to understand the performance of *busCharging*. They are (i) the temporal distribution of charging events, (ii) the utilization rates of charging stations, (iii) the occupation rates of charging stations, and (iv), most importantly, the electricity usage and charging cost.

4.2 Temporal Distribution of Charging Events

Figure 18(a) and Figure 18(b) show the distribution of the charging start time of *busCharging*, EDF, and Ground Truth in different electricity rates durations, i.e., different time slots of a day. We found that more charging events happen in 23:00–7:00 under *busCharging*, resulting in fewer charging events during the daytime. Especially during 18:00–21:00, the gap between *busCharging* and the



(a) Peak & flat hour (b) Off-peak hour

Fig. 18. Charging start time of e-buses.

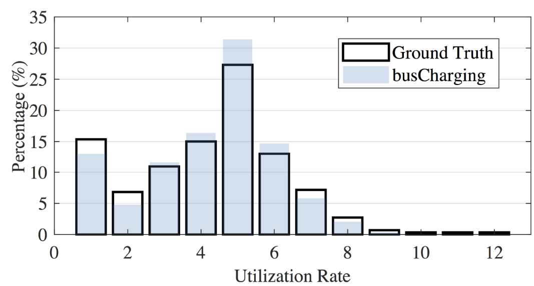


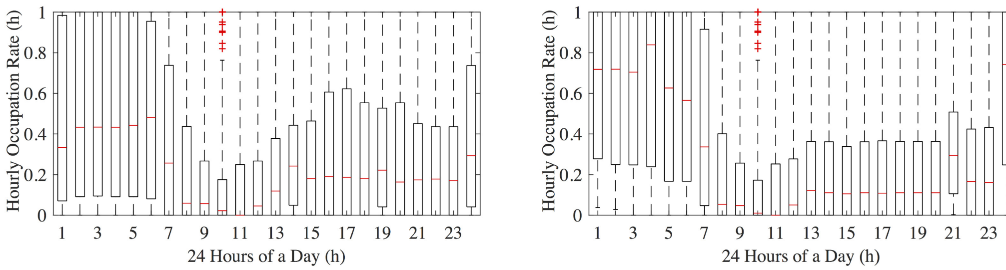
Fig. 19. Comparison of the utilization rate.

Ground Truth is more obvious. This is because *busCharging* schedules some e-buses that have been charged from 12:00–16:00 to serve other lines to replace some other e-buses, leaving these e-buses to charge after 23:00 for a lower charging electricity rate. While, during 16:00–18:00, more e-buses need to charge after the long-time operation, which results in the increase of charging events under EDF. Overall, *busCharging* has more charging events during the late night to early morning, i.e., off-peak hours, which potentially lead to lower charging costs.

4.3 Utilization of Charging Stations

We first leverage the *Charging Station Utilization Rate* that we defined in Section 2 to describe the spatial distribution of e-bus charging events. As shown in Figure 19, there are fewer charging stations with too high or too low utilization rates under *busCharging* compared to the Ground Truth. In particular, the number of charging stations with the utilization rates between 3–6 accounts for 73% under *busCharging*, which is 7% less than the Ground Truth. Since EDF does not change the charging locations of e-buses, the utilization rate of EDF is the same as the Ground Truth, so we do not show its performance. The more balanced utilization of the charging infrastructure can effectively reduce the under-utilization and the overcrowded utilization phenomenon among all e-bus charging stations.

The reason why *busCharging* balances the charges between stations is that the e-buses can be scheduled to serve for different lines under *busCharging*, which increases the flexibility of scheduling and potentially leads to better performance; for example, if an e-bus arrives at the terminal and it needs to charge, but no charging points are available at this charging station, and its SOC is not enough for serving the current line, i.e., go back to the original terminal. In this case, *busCharging* schedules the e-bus to serve another shorter line with lower expected energy consumption, e.g., a shorter distance or a less-congested route. Hence, *busCharging* can potentially improve the



(a) Ground truth (b) busCharging

Fig. 20. Hourly occupation rates.

charging efficiency of two charging stations at the same time, i.e., reducing the utilization rate of one charging station and increasing the utilization rate of the other to balance their utilization rates.

Even though the 7% of improvement in charge station utilization rate seems not to be significant in terms of percentage, but the rebalanced 9,613 charging activities (137,332*7%) will potentially improve the uneven charging infrastructure deployment situation caused by some real-world problems, e.g., unavailable land resources for charging stations at some places, so we believe this spatial improvement can make a difference to improve the charging efficiency of the city-scale electric bus fleet.

4.4 Charging Station Occupation Rate

In addition to the spatial distribution of charging events in the e-bus charging network, we also define the hourly **Charging Station Occupation Rate** as a joint metric to describe the temporal and spatial distribution of charging events among all charging stations, which stands for the average occupation time in one hour of all charging points in each station. The Charging Station Occupation Rate in a station s_i is expressed as:

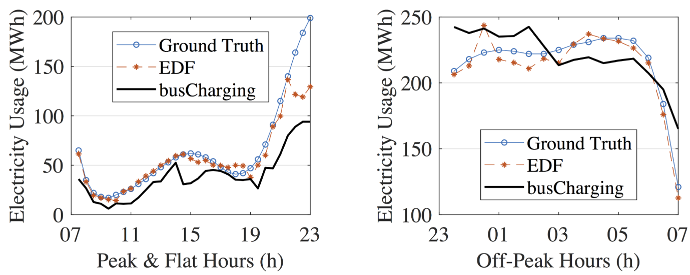
$$OR\left(s_i\right) = \frac{\sum_{j=1}^{N_i} CT\left(s_i^j\right)}{N_i}, \tag{6}$$

where $CT(s_i^j)$ is the hourly occupation time of the jth charging point s_i^j in the station s_i; N_i is the total number charging points in station s_i.

Figure 20(a) and Figure 20(b) show the box-plots of the hourly occupation rates of all e-bus charging stations in Shenzhen under current scheduling (i.e., ground truth) and our *busCharging* scheduling. We found that the occupation rates of the e-bus charging network increase from 0:00–6:00 under *busCharging* scheduling, which is the off-peak hours of electricity pricing. The occupation rates of the e-bus charging network decrease during 14:00–20:00 under *busCharging*, which is in the peak and flat hours of Shenzhen electricity pricing, which indicates that our *busCharging* can make more charges happen in low electricity pricing hours and has the potential to reduce charging costs for the e-bus fleet.

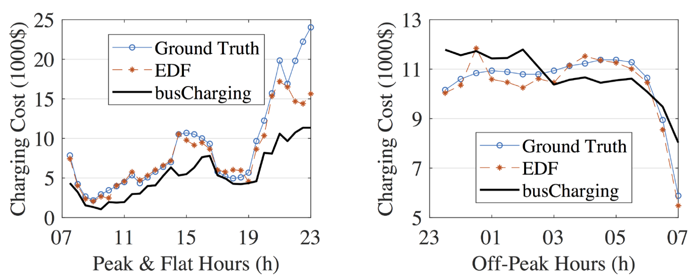
4.5 Electricity Usage & Charging Cost

Figure 21(a) and Figure 21(b) show the electricity usage for refueling e-buses in different electricity rates durations. We found *busCharging* has more electricity usage during the midnight to early morning, especially during 23:00–23:59. This is because *busCharging* schedules some e-buses with enough energy to serve for other lines before this duration and then charge during this period,



(a) Peak & flat hour (b) Off-peak hour

Fig. 21. Electricity usage for e-bus charging.



(a) Peak & flat hour (b) Off-peak hour

Fig. 22. Charging cost for e-bus charging.

resulting in more electricity usage during this period. In addition, *busCharging* also reduces the electricity usage during daytime peak hours, e.g., 14:00–16:00. EDF increases the electricity usage from 16:00–18:00 due to the high charging demand in this duration. In total, our *busCharging* reduces 12.8% (701 MWh) and 8.2% of electricity energy for the e-bus fleet in Shenzhen per day compared with the Ground Truth and EDF.

Figure 22(a) and Figure 22(b) show the charging cost distribution. We found that the cost gap between *busCharging* and Ground Truth/EDF is more significant than the charging start time gap in Figure 18(a) during the daytime. This is because the electricity rates in the daytime are much higher than the price of off-peak hours at late night. Even though *busCharging* causes a slightly higher charging cost for the fleet during 23:00–2:00, it reduces 23.7% ($106,870) and 17.8% of the overall charging cost per day for the e-bus fleet compared with the Ground Truth and EDF, which indicates *busCharging* can potentially reduce 39M dollars for the Shenzhen e-bus fleet per year based on the current Shenzhen e-bus budget.

4.6 Impacts on E-bus Drivers

Since bus drivers serve for bus companies instead of themselves, they will follow the scheduling decisions made by our system. A practical issue is that some e-bus drivers need to sacrifice their time to serve other bus lines, resulting in extra working hours, which may cause these drivers to feel unfairly treated and make complaints. In this part, we quantify the impacts on drivers due to extra work for serving other lines.

Figure 23 shows the number of influenced bus lines and the number of e-buses that are scheduled to serve other lines during each day of the week. We found that the number of influenced lines ranges from 195 (on Sunday) to 258 (on Thursday), which accounts for about 25% of the total

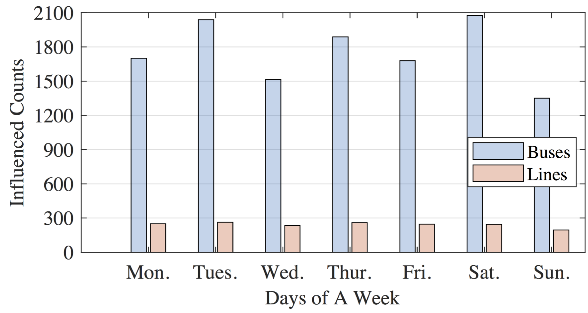


Fig. 23. Number of influenced buses and lines.

number of operating bus lines each day. The number of buses that served other lines ranges from 1,350 (on Sunday) to 2,075 (on Saturday), which accounts for about 15% of the total number of operating buses in the day. In this case, about 15% of e-bus drivers need to do extra work each day.

(i) One possible solution could be that the e-bus operators provide some economic compensation for those drivers who make extra efforts for better charging efficiency. The current average wage of Shenzhen bus drivers is about $780 per month [12], which means about $3.25 per hour. If operators compensate a two-hour salary to these drivers, the maximum extra cost would no more than $13,500 per day. Compared to the $106,870 saving for e-bus charging, bus operators can still save $93,370 per day by adopting our *busCharging* scheduling strategy. (ii) In addition, it is expected that the shared autonomous electric vehicles will shape the human mobility of tomorrow [4, 5]. Under this setting, no extra efforts for drivers to serve other bus lines, and all work will be performed by the autonomous electric vehicles, so our *busCharging* scheduling strategy may be more effective and attract more attention in the future.

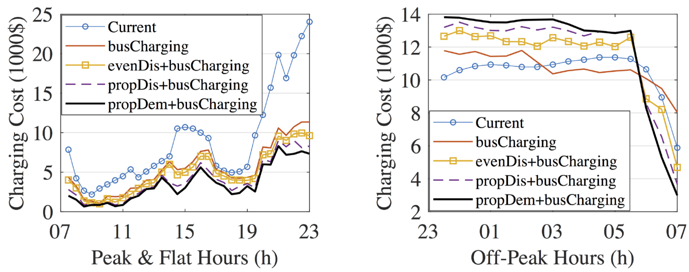
4.7 Scheduling-based Charging Station Expansion

Since the E-bus promotion is a long-term process, the number of charging points for them will increase with the maturity of the charging infrastructure. In this case, a key issue is how to assign the new charging points to the existing charging stations/bus terminals. To further verify if our *busCharging* is effective if more charging points are deployed, we provide a potential application of our *busCharging* by designing a scheduling-based optimal charging station expansion mechanism, which also has the potential to help us to assign charging points to each station.

It should be noted that we do not consider to deploy new charging stations in other locations, since most existing charging points are deployed in bus terminals to reduce operation costs, so it may potentially cause extremely high costs for operators if deploying charging stations in other places. In addition, deploying new charging stations is another parallel topic of this work, so we leave this for future work.

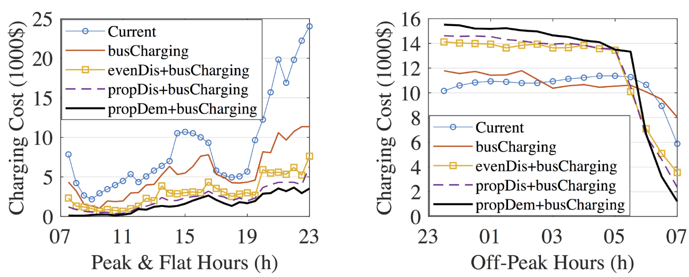
In this part, we focus on expanding the existing charging stations by adding more charging points, which can further reduce the charging operation costs and satisfy large-scale e-bus fleets' charging demand. In particular, we propose a *busCharging*-based charging station expansion mechanism, which means we decide which deployment strategy is better if considering our *busCharging* scheduling.

For simplification, we compare three widely used charging station expansion methods, i.e., (i) evenly distribute new charging points to existing stations (evenDis); (ii) distribute new charging points proportionally with the current distribution (propDis); (iii) distribute new charging points proportionally with the charging demand in each station (propDem). After expanding



(a) Peak & flat hour (b) Off-peak hour

Fig. 24. Charging cost with 10% more charging points.



(a) Peak & flat hour (b) Off-peak hour

Fig. 25. Charging cost with 30% more charging points.

charging stations by using each strategy, we then utilize our *busCharging* to conduct the charging scheduling. We conduct experiments on adding 10% and 30% of new charging points to show the performance of different approaches.

From Figure 24, we found that the propDis and propDem combined with our *busCharging* can achieve good performance for reducing charging costs if we add 10% more charging points, even though the propDem + *busCharging* can achieve the largest charging cost reduction, since more charges will happen during off-peak hours under this strategy. More specifically, propDem + *busCharging* with 10% more charging points can reduce the daily charging cost of the Shenzhen e-bus fleet by $148,201.

Figure 25 shows the performance of different strategies with an extra 30% of charging points. We found with more charging points deployed, the charging cost will further reduce during peak & flat hours, especially from 19:00–23:00. The reason is that more charging points can be utilized to serve e-buses during 23:00–5:00 if more charging points are deployed proportionally with the charging demand in each station. More specifically, propDem + *busCharging* with 30% more charging points can reduce the daily charging cost of the Shenzhen e-bus fleet by $199,051.

We have also compared the performance of the three-station expansion strategy without our *busCharging* and with our *busCharging*. Since we found the performance with our *busCharging* is much worse than that with our *busCharging*, we do not show their performance in Figure 24 and Figure 25. In summary, we found more charging points deployed in existing charging stations combined with our *busCharging* charging scheduling can significantly reduce the charging cost of the Shenzhen e-bus fleet, which means the current charging points may not be sufficient for the Shenzhen e-bus fleet. We also found that distributing new charging points proportionally with the charging demand in each station may be better compared to other strategies.

（a）Charging station at 23:05 （b）Charging station at 17:06

Fig. 26. Charging station status in different times.

This finding indicates that our *busCharging* is not only efficient for current charging stations but also effective during the charging station expansion process, which can also provide guidelines for charging station expansion.

5 LESSONS LEARNED

Based on our results, we have been conducting a few rounds of field studies to verify the patterns we found. In particular, we have been communicating with bus drivers, fleet managers, and charging station operators to fully understand the potential impacts of our study and the implication of our scheduling. We summarize a few lessons we learned from the project and field studies regarding the e-bus fleet in Shenzhen below.

Data Issues: Some unexpected lessons we learned are related to data issues. In our field studies, we have been communicating with various interested parties for data collection, data quality, and data management. We summarize a few key insights as follows: (i) Since the Shenzhen e-bus fleet is operated by three companies, the data formats and access policies are very different. It takes us a long time to prepare the bus GPS data to understand the current operating patterns. (ii) Further, Shenzhen bus-fare data are managed separately by a smartcard company, and the fare data regarding smartcards have personal information, e.g., cellphone numbers and addresses. The company staff at the smartcard company has removed all this personal information from the smartcard data to help us understand the bus passenger demand anonymously without privacy concerns. We have reported many data issues to the fleet management team, and some of these problems have been addressed after we informed them and other issues are also in process.

Additional Drivers Exclusively for Charging: The final unexpected lesson is the labor-intensive charging operation in the night. Figure 26 shows a detailed charging station setting in our field studies and their status at two different times of a day. We can see that all charging points are occupied by e-buses at 23:05; whereas only one e-bus was charging at 17:06. This is because the electricity rate between 23:00–7:00 is much lower than the rate at 17:00. However, given limited charging points due to high costs, e.g., \$ 80,000 for a charging point deployment, e-buses need to be moved around before or after charging, but the regular bus drivers will be off-duty after 23:00 and before 7:00. Hence, the Shenzhen e-bus network hires additional drivers just for moving buses from late night to early morning for e-bus charging. In particular, one of the three e-buses operating companies in Shenzhen has 750 additional drivers just for moving e-buses before or after charging. By considering this labor cost factor, it may be more reasonable to charge more buses during the daytime, since regular drivers can move the buses without cost for additional drivers. However, based on our interactions with the fleet management team, hiring additional drivers is a

Fig. 27. Charging point status during the charging process.

Table 2. Categories of Related Work

Scheduling	Small-Scale	City-Scale
Decentralized	[28, 30, 35, 36]	[6, 24, 39, 51, 52]
Centralized	[18–20, 22, 23, 53]	*busCharging*

short-term and temporary issue and they expected a more efficient charging scheduling strategy like our *busCharging* to address this issue.

Charging Issues: Normally, e-buses will be fully charged during nights for satisfying the daily operation. However, some e-buses still need to recharge during the day due to their long daily operation time and distance, but the break time in the noon is not enough for a full charge, so these e-buses will not be fully charged during day time. In addition, the charging speed from low battery capacity to 80% is fast, so it is easy for e-buses to be recharged to a high energy capacity at noon instead of being fully charged. For example, as shown in Figure 27, we found that the remaining charging time of the e-bus is 52 minutes when the SOC is 56%, and the output current and output voltage are 198.5 A and 554.7 V, respectively. At that time, the e-bus has been charged 35 minutes and 64.1 kWh. When the remaining time is 51 minutes, the SOC of the e-bus becomes 57%.

Implementation in Different Cities: The bus networks in different cities typically have different operating patterns due to geographic and demographic features, so it is extremely significant to implement *busCharging* in different cities. Currently, we are trying to obtain e-bus data from other cities for dual-city modeling. However, since only Shenzhen has such a large-scale e-bus network, it is challenging to find such a similar-scale e-bus fleet for a parallel study currently. One potential direction we are exploring is designing transfer learning models to transfer the knowledge (e.g., operating patterns, charging patterns) from the Shenzhen e-bus network to bus fleets in other cities for a "what if" investigation. For example, what if all traditional diesel buses in Beijing or New York City were replaced with e-buses, how much will it cost, and how to schedule e-buses in these cities to charge. It opens some very interesting research directions.

6 RELATED WORK

There are two charging management modes for electric vehicles (EVs), i.e., decentralized and centralized ones [19, 23]. For existing research, some works have been done based on large-scale real-world data, while others utilize a small real-world dataset [18, 19] or simulated data [13, 15, 54, 62]. Based on these two criteria, we divide the EV research into four different categories, as shown in Table 2.

6.1 Decentralized Scheduling

Small-Scale Scheduling: The decentralized charging scheduling of EVs has been widely studied by many researchers, but most of them are based on small-scale data or simulated data. Reference [30] developed a reservation recommendation algorithm for EVs considering the shortest distance and shortest waiting time. Reference [21] designed a distributed market mechanism to improve the economic and temporal efficiency of EV demand response. Mou et al. [28] tried to flatten the load curve of low-voltage transformers, while satisfying each consumer's requirement for their EVs to be charged to the required level by the specified time. They first formulated this problem as a convex optimization problem and then proposed a decentralized water-filling-based algorithm to solve it. Gan et al. [11] proposed a decentralized algorithm to optimally schedule EV charging. The algorithm exploited the elasticity of electric vehicle loads to fill the valleys in electric load profiles. Wen et al. [48] proposed the application of the convex relaxation optimization method to solve the EV charging selection problem and then developed a distributed optimization algorithm to solve this problem in a decentralized manner. Reference [35] presented a distributed power schedule framework based on the Game Theory to obtain an optimal schedule for online EVs. These works are based on small-scale data and theoretical models, which are difficult to capture the dynamics of real-world large-scale EVs operating and charging patterns.

City-scale Scheduling: Reference [24] developed a charging infrastructure deployment and charging point placement framework to minimize the overall charging time of e-taxis. Reference [39] designed a charging recommendation system for e-taxis to reduce their total charging time for drivers. PickaChu [51] provided a charging point deployment mechanism that maximizes the probability of picking up passengers for e-taxis and minimizes the deployment cost. Ma et al. [27] established a framework for EV charging coordination that facilitates the tradeoff between total generation cost and the local costs associated with overloading and battery degradation. A decentralized approach is proposed to solve the resulting large-scale optimization problem involving each PEV minimizing their charging cost with respect to a forecast price profile while taking into account local grid and battery effects. All these works are based on a distributed charging nature. However, in this work, we consider a centralized scheduling model for higher efficiency.

6.2 Centralized Scheduling

Small-scale Scheduling: There are also some existing works for the centralized scheduling of EVs based on small-scale data or experimental simulations. Reference [19] proposed a population-based heuristic approach to minimize the total charging costs, which is executed on a 20-bus test system. Reference [18] introduced and analyzed the electric transit bus system in a research institute campus. Reference [22] proposed an effective charging rate control algorithm to optimize the social welfare of EVs. REC [6] developed a real-time EV charging scheduling framework for e-taxi fleets, which informs each e-taxi driver at runtime when and where to have a charge. Reference [33] presented an off-board real-time SOC estimation technique for a centralized battery management system. However, these small systems cannot fully reveal the complexity and advantage of centralized scheduling for city-scale fleets.

City-scale Scheduling: Different from the existing work, our work addresses a practical charging issue for e-bus fleets by centralized scheduling in a setting of city-scale fleets. To our best knowledge, *busCharging* is the first work of city-scale data-driven investigation on studying the real-time charging scheduling for large-scale e-bus fleets. Such a data-driven investigation enables us to identify the real-world e-bus operating and charging issues, which are challenging to reveal using simulated data-based studies, small-scale data, or under a decentralized setting.

7 CONCLUSION

In this article, we conduct, to the best of our knowledge, the first study called *busCharging* for data-driven, real-time charging scheduling for large-scale e-bus fleets based on a real-world dataset in Shenzhen, which includes the data from 16,359 e-buses, 1,400 bus lines, and 376 charging stations. In *busCharging*, we consider various real-world factors based on the long-term data including e-bus daily and per-charge operating distances, charging spatial-temporal distributions, charging station utilization rates, and time-variant electricity rates, and so on. We formulate the electricity pricing-aware charging scheduling problem into a Markov Decision Process, and then we effectively solve it by the policy iteration for the real-time requirement. More importantly, we have shown that with its effective scheduling, *busCharging* outperforms the ground truth by 23.7% and outperforms the baseline method by 17.8% regarding the total charging cost. For the immediate benefit, *busCharging* can reduce the operating cost for the Shenzhen e-bus network with its data-driven real-time scheduling. For the long-term benefit, our results in *busCharging* may be used for other cities to promote e-buses for green public transportation.

REFERENCES

[1] Aijun Bai, Feng Wu, and Xiaoping Chen. 2015. Online planning for large Markov decision processes with hierarchical decomposition. *ACM Trans. Intell. Syst. Technol.* 6, 4 (2015), 1–28.

[2] Mikołaj Bartłomiejczyk. 2018. Driving performance indicators of electric bus driving technique: Naturalistic driving data multicriterial analysis. *IEEE Trans. Intell. Transport. Syst.* 20, 4 (2018), 1442–1451.

[3] Richard Bellman. 2013. *Dynamic Programming*. Courier Corporation.

[4] Lawrence D. Burns. 2013. A vision of our transport future. *Nature* 497, 7448 (2013), 181–182.

[5] T. Donna Chen, Kara M. Kockelman, and Josiah P. Hanna. 2016. Operations of a shared, autonomous, electric vehicle fleet: Implications of vehicle & charging infrastructure decisions. *Transport. Res. Part A: Polic. Pract.* 94 (2016), 243–254.

[6] Zheng Dong, Cong Liu, Yanhua Li, Jie Bao, Yu Gu, and Tian He. 2017. Rec: Predictable charging scheduling for electric taxi fleets. In *Proceedings of the IEEE Real-time Systems Symposium (RTSS'17)*. IEEE, 287–296.

[7] Build Your Dreams. 2018. Whole vehicle technologies and battery technology of BYD K9. Retrieved from http://en.byd.com/la/auto/ebus.html.

[8] Bowen Du, Yongxin Tong, Zimu Zhou, Qian Tao, and Wenjun Zhou. 2018. Demand-aware charger planning for electric vehicle sharing. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 1330–1338.

[9] Eugene A. Feinberg and Adam Shwartz. 2012. *Handbook of Markov Decision Processes: Methods and Applications*. Springer Science & Business Media.

[10] Bloomberg New Energy Finance. 2018. Electric Buses in Cities: Driving Towards Cleaner Air and Lower CO_2. Retrieved from http://www.c40.org/c40_research.

[11] Lingwen Gan, Ufuk Topcu, and Steven H. Low. 2012. Optimal decentralized protocol for electric vehicle charging. *IEEE Trans. Power Syst.* 28, 2 (2012), 940–951.

[12] Lei Gong. 2018. Wages of Shenzhen Bus Drivers. Retrieved from https://www.douban.com/note/683752483/.

[13] Azwirman Gusrialdi, Zhihua Qu, and Marwan A. Simaan. 2017. Distributed scheduling and cooperative control for charging of electric vehicles at highway service stations. *IEEE Trans. Intell. Transport. Syst.* 18, 10 (2017), 2713–2727.

[14] John Holler, Risto Vuorio, Zhiwei Qin, Xiaocheng Tang, Yan Jiao, Tiancheng Jin, Satinder Singh, Chenxi Wang, and Jieping Ye. 2019. Deep reinforcement learning for multi-driver vehicle dispatching and repositioning problem. In *Proceedings of the IEEE International Conference on Data Mining (ICDM'19)*. IEEE, 1090–1095.

[15] Luyang Hou, Chun Wang, and Jun Yan. 2020. Bidding for preferred timing: An auction design for electric vehicle charging station scheduling. *IEEE Trans. Intell. Transport. Syst.* 21, 8 (2020), 3332–3343.

[16] Hsun-Ping Hsieh, Shou-De Lin, and Yu Zheng. 2015. Inferring air quality for station location recommendation based on urban big data. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 437–446.

[17] Emil B. Iversen, Juan M. Morales, and Henrik Madsen. 2014. Optimal charging of an electric vehicle using a Markov decision process. *Appl. Energy* 123 (2014), 1–12.

[18] Young Jae Jang, Eun Suk Suh, and Jong Woo Kim. 2016. System architecture and mathematical models of electric transit bus system utilizing wireless power transfer technology. *IEEE Syst. J.* 10, 2 (2016), 495–506.

[19] Qi Kang, JiaBao Wang, MengChu Zhou, and Ahmed Chiheb Ammari. 2016. Centralized charging strategy and scheduling algorithm for electric vehicles under a battery swapping scenario. *IEEE Trans. Intell. Transport. Syst.* 17, 3 (2016), 659–669.

[20] Eugene Kim, Jinkyu Lee, and Kang G. Shin. 2015. Modeling and real-time scheduling of large-scale batteries for maximizing performance. In *Proceedings of the IEEE Real-time Systems Symposium*. IEEE, 33–42.

[21] Fanxin Kong and Xue Liu. 2015. Distributed deadline and renewable aware electric vehicle demand response in the smart grid. In *Proceedings of the IEEE Real-time Systems Symposium (RTSS'15)*. IEEE, 23–32.

[22] Fanxin Kong, Xue Liu, Zhonghao Sun, and Qinglong Wang. 2016. Smart rate control and demand balancing for electric vehicle charging. In *Proceedings of the International Conference on Cyber-physical Systems (ICCPS'16)*. 4.

[23] Fanxin Kong, Qiao Xiang, Linghe Kong, and Xue Liu. 2016. On-line event-driven scheduling for electric vehicle charging via park-and-charge. In *Proceedings of the IEEE Real-time Systems Symposium (RTSS'16)*. IEEE, 69–78.

[24] Yanhua Li, Jun Luo, Chi-Yin Chow, Kam-Lam Chan, Ye Ding, and Fan Zhang. 2015. Growing the charging station network for electric vehicles with trajectory data analytics. In *Proceedings of the IEEE 31st International Conference on Data Engineering (ICDE'15)*. IEEE, 1376–1387.

[25] Ying Li, Changjie Zhan, Martin de Jong, and Zofia Lukszo. 2016. Business innovation and government regulation for the promotion of electric vehicle use: Lessons from Shenzhen, China. *J. Clean. Product.* 134 (2016), 371–383.

[26] Chen Liu, Ke Deng, Chaojie Li, Jianxin Li, Yanhua Li, and Jun Luo. 2016. The optimal distribution of electric-vehicle chargers across a city. In *Proceedings of the IEEE 16th International Conference on Data Mining (ICDM'16)*. IEEE, 261–270.

[27] Zhongjing Ma, Suli Zou, Long Ran, Xingyu Shi, and Ian A. Hiskens. 2016. Efficient decentralized coordination of large-scale plug-in electric vehicle charging. *Automatica* 69 (2016), 35–47.

[28] Yuting Mou, Hao Xing, Zhiyun Lin, and Minyue Fu. 2015. Decentralized optimal demand-side management for PHEV charging in a smart grid. *IEEE Trans. Smart Grid* 6, 2 (2015), 726–736.

[29] People's Daily Online. 2017. Shenzhen becomes world's first city with all-electric public transportation. Retrieved from http://en.people.cn/n3/2017/1228/c90000-9309683.html.

[30] Chan Jung Park, Junghoon Lee, Gyung Leen Park, and Jung Suk Hyun. 2014. Development of reservation recommendation algorithms for charging electric vehicles in smart-grid cities. *Int. J. Smart Home* 8, 1 (2014), 113–122.

[31] People.cn. 2018. Shenzhen is Connected by Electric Buses. Retrieved from http://society. people.com.cn/n1/2018/0508/c1008-29970131.html.

[32] Martin L. Puterman. 2014. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons.

[33] Omid Rahbari, Noshin Omar, Peter Van Den Bossche, and Joeri Van Mierlo. 2018. A centralized state of charge estimation technique for electric vehicles equipped with lithium-ion batteries in smart grid environment. In *Proceedings of the IEEE International Conference on Industrial Technology (ICIT'18)*. IEEE, 1721–1725.

[34] Viktor Irle Roland Irle, José Pontes. 2018. The electric vehicle world sales database. Retrieved from http://www.ev-volumes.com/.

[35] Ankur Sarker, Zhuozhao Li, William Kolodzey, and Haiying Shen. 2017. Opportunistic energy sharing between power grid and electric vehicles: A game theory-based pricing policy. In *Proceedings of the IEEE 37th International Conference on Distributed Computing Systems (ICDCS'17)*. IEEE, 1197–1207.

[36] Ankur Sarker, Haiying Shen, and John A. Stankovic. 2018. MORP: Data-driven multi-objective route planning and optimization for electric vehicles. *Proc. ACM Interact., Mob., Wear. Ubiq. Technol.* 1, 4 (2018), 162.

[37] Sohu. 2016. A Survey of Shenzhen Electric Buses. Retrieved from https://m.sohu.com /n/465555062/.

[38] Beia Spiller. 2015. All Electricity Is Not Priced Equally: Time-Variant Pricing. Retrieved from http://blogs.edf.org/energyexchange/2015/01/27/all-electricity-is-not-priced-equally-time-variant-pricing-101/.

[39] Zhiyong Tian, Taeho Jung, Yi Wang, Fan Zhang, Lai Tu, Chengzhong Xu, Chen Tian, and Xiang Yang Li. 2016. Real-time charging station recommendation system for electric-vehicle taxis. *IEEE Trans. Intell. Transport. Syst.* 17, 11 (2016), 3098–3109.

[40] Guang Wang, Xiuyuan Chen, Fan Zhang, Yang Wang, and Desheng Zhang. 2019. Experience: Understanding long-term evolving patterns of shared electric vehicle networks. In *Proceedings of the 25th Annual International Conference on Mobile Computing and Networking*. 1–12.

[41] Guang Wang, Wenzhong Li, Jun Zhang, Yingqiang Ge, Zuohui Fu, Fan Zhang, Yang Wang, and Desheng Zhang. 2019. sharedCharging: Data-driven shared charging for large-scale heterogeneous electric vehicle fleets. *Proc. ACM Interact., Mob., Wear. Ubiq. Technol.* 3, 3 (2019), 1–25.

[42] Guang Wang, Xiaoyang Xie, Fan Zhang, Yunhuai Liu, and Desheng Zhang. 2018. bCharge: Data-driven real-time charging scheduling for large-scale electric bus fleets. In *Proceedings of the IEEE Real-Time Systems Symposium (RTSS'18)*. IEEE, 45–55.

[43] Guang Wang, Tianhua Xu, Tao Tang, Tangming Yuan, and Haifeng Wang. 2017. A Bayesian network model for prediction of weather-related failures in railway turnout systems. *Expert Syst. Applic.* 69 (2017), 247–256.

[44] Guang Wang and Desheng Zhang. 2019. Poster: Understanding long-term mobility and charging evolving of shared EV networks. In *Proceedings of the 25th Annual International Conference on Mobile Computing and Networking*. 1–3.

[45] Guang Wang, Fan Zhang, Huijun Sun, Yang Wang, and Desheng Zhang. 2020. Understanding the long-term evolution of electric taxi networks: A longitudinal measurement study on mobility and charging patterns. *ACM Trans. Intell. Syst. Technol.* 11, 4 (2020), 1–27.

[46] Guang Wang, Fan Zhang, and Desheng Zhang. 2019. tCharge-A fleet-oriented real-time charging scheduling system for electric taxi fleets. In *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*. 440–441.

[47] Guang Wang, Yongfeng Zhang, Zhihan Fang, Shuai Wang, Fan Zhang, and Desheng Zhang. 2020. FairCharge: A data-driven fairness-aware charging recommendation system for large-scale electric taxi fleets. *Proc. ACM Interact., Mob., Wear. Ubiq. Technol.* 4, 1 (2020), 1–25.

[48] Chao-Kai Wen, Jung-Chieh Chen, Jen-Hao Teng, and Pangan Ting. 2012. Decentralized plug-in electric vehicle charging selection algorithm in power systems. *IEEE Trans. Smart Grid* 3, 4 (2012), 1779–1789.

[49] Yanhai Xiong, Jiarui Gan, Bo An, Chunyan Miao, and Ana L. C. Bazzan. 2015. Optimal electric vehicle charging station placement. In *Proceedings of the 24th International Conference on Artificial Intelligence.* AAAI Press, 2662–2668.

[50] Yanyan Xu, Serdar Çolak, Emre C. Kara, Scott J. Moura, and Marta C. González. 2018. Planning for electric vehicle needs by coupling charging profiles with urban mobility. *Nat. Ener.* 3, 6 (2018), 484–493.

[51] Li Yan, Haiying Shen, Zhuozhao Li, Ankur Sarker, John A. Stankovic, Chenxi Qiu, Juanjuan Zhao, and Chengzhong Xu. 2018. Employing opportunistic charging for electric taxicabs to reduce idle time. *Proc. ACM Interact., Mob., Wear. Ubiq. Technol.* 2, 1 (2018), 47.

[52] Li Yan, Haiying Shen, Juanjuan Zhao, Chengzhong Xu, Feng Luo, and Chenxi Qiu. 2017. CatCharger: Deploying wireless charging lanes in a metropolitan road network through categorization and clustering of vehicle traffic. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM'17).* IEEE, 1–9.

[53] Chao Yang, Wei Lou, Junmei Yao, and Shengli Xie. 2017. On charging scheduling optimization for a wirelessly charged electric bus system. *IEEE Trans. Intell. Transport. Syst.* 19, 6 (2017), 1814–1826.

[54] Zonggen Yi and Peter H. Bauer. 2018. Energy aware driving: Optimal electric vehicle speed profiles for sustainability in transportation. *IEEE Trans. Intell. Transport. Syst.* 20, 3 (2018), 1137–1148.

[55] Jing Yuan, Yu Zheng, Xing Xie, and Guangzhong Sun. 2011. Driving with knowledge from the physical world. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, 316–324.

[56] Desheng Zhang, Jun Huang, Ye Li, Fan Zhang, Chengzhong Xu, and Tian He. 2014. Exploring human mobility with multi-source data at extremely large metropolitan scales. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking.* ACM, 201–212.

[57] Desheng Zhang, Juanjuan Zhao, Fan Zhang, and Tian He. 2015. coMobile: Real-time human mobility modeling at urban scale using multi-view learning. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems.* ACM, 40.

[58] Fuzheng Zhang, David Wilkie, Yu Zheng, and Xing Xie. 2013. Sensing the pulse of urban refueling behavior. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing.* ACM, 13–22.

[59] Fuzheng Zhang, Nicholas Jing Yuan, David Wilkie, Yu Zheng, and Xing Xie. 2015. Sensing the pulse of urban refueling behavior: A perspective from taxi mobility. *ACM Trans. Intell. Syst. Technol.* 6, 3 (2015), 37.

[60] Jun Zhang, Dayong Shen, Lai Tu, Fan Zhang, Chengzhong Xu, Yi Wang, Chen Tian, Xiangyang Li, Benxiong Huang, and Zhengxi Li. 2017. A real-time passenger flow estimation and prediction method for urban bus transit systems. *IEEE Trans. Intell. Transport. Syst.* 18, 11(2017), 3168–3178.

[61] Jun Zhang, Xin Yu, Chen Tian, Fan Zhang, Lai Tu, and Chengzhong Xu. 2014. Analyzing passenger density for public bus: Inference of crowdedness and evaluation of scheduling choices. In *Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC'14).* IEEE, 2015–2022.

[62] Yongmin Zhang, Pengcheng You, and Lin Cai. 2018. Optimal charging scheduling by pricing for EV charging station with dual charging modes. *IEEE Trans. Intell. Transport. Syst.* 20, 9 (2018), 3386–3396.

[63] Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. 2014. Urban computing: Concepts, methodologies, and applications. *ACM Trans. Intell. Syst. Technol.* 5, 3 (2014), 38.

[64] Yu Zheng, Yukun Chen, Quannan Li, Xing Xie, and Wei-Ying Ma. 2010. Understanding transportation modes based on GPS data for web applications. *ACM Trans. Web* 4, 1 (2010), 1.

[65] Yu Zheng, Furui Liu, and Hsun-Ping Hsieh. 2013. U-air: When urban air quality inference meets big data. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, 1436–1444.

[66] Yu Zheng and Xing Xie. 2011. Learning travel recommendations from user-generated GPS traces. *ACM Trans. Intell. Syst. Technol.* 2, 1 (2011), 2.

[67] Yu Zheng, Xiuwen Yi, Ming Li, Ruiyuan Li, Zhangqing Shan, Eric Chang, and Tianrui Li. 2015. Forecasting fine-grained air quality based on big data. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, 2267–2276.

[68] Julie Yixuan Zhu, Chao Zhang, Huichu Zhang, Shi Zhi, Victor O. K. Li, Jiawei Han, and Yu Zheng. 2017. pg-Causality: Identifying spatiotemporal causal pathways for air pollutants with urban big data. *IEEE Trans. Big Data* 4, 4 (2017), 571–585.

Received November 2019; revised August 2020; accepted October 2020